RADIO ENGINEERS' HANDBOOK

OTHER BOOKS

by FREDERICK EMMONS TERMAN

***

Radio Engineering

Fundamentals of Radio

Measurements in Radio Engineering

RADIO ENGINEERS' HANDBOOK

BY

FREDERICK EMMONS TERMAN, Sc.D.
Professor of Electrical Engineering and Dean, School of Engineering, Stanford University, Past President, Institute of Radio Engineers

FIRST EDITION
TENTH IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK AND LONDON
1943

RADIO ENGINEERS' HANDBOOK

PREFACE

The purpose of the "Radio Engineers' Handbook" is to provide a reference book summarizing the body of engineering knowledge that is the basis of radio and electronics.

Particular effort has been made to achieve thoroughness. Some two thousand technical articles were reviewed during the preparation of the manuscript, and the book in its final form presents in an organized way the more important contributions made to the art by this vast body of technical literature. In the selection of material an attempt has been made to avoid the common failing of most handbooks, which instead of summarizing *all* of the important facts about a topic, and bringing together formulas and procedures useful in actual design, too often devote most of the available space to introductory material, such as found in good textbooks, and to descriptions of particular equipments.

The text is well documented, there being about fifteen hundred footnote references to published articles. Practically all the pertinent engineering literature in English that bears on any particular subject is referred to in connection with the text on that subject. The reader who wishes to make an intensive study of a particular topic is thus in most cases spared the necessity of engaging in a time-consuming search through many indexes to find out what has been published. At the same time, the essential information, formulas, etc., are insofar as possible included in this book. The reader, therefore, does not need to have an entire library within arm's reach in order to obtain the answer to most problems, and yet knows just where to go for further information.

Attention is called to certain sections of the book which present in one place material previously widely scattered. Thus the formulas and curves on skin effect, inductance, mutual inductance, and capacity in Section 2 represent the most complete collections ever presented at one place. The same is true of the transmission-line equations in Section 3, the formulas for field patterns and radiation resistance of antennas in Section 11, the treatment of ground-wave and ultra-high-frequency propagation in Section 10, etc. Similarly, the important subject of network theory is covered in a straightforward and comprehensive way that, it is hoped, will widen the understanding of such important matters as Foster's reactance theorem, attenuation and phase equalizers, lattice and ladder filters of various types, impedance matching and insertion loss, the relation between attenuation and phase shift, etc. The treatment on

electron optics is noteworthy in that it is the first summary of the subject that has appeared in handbook form, and also because it presents the most complete collection of data on electron lenses that has ever been published.

The "Radio Engineers' Handbook" is unique among handbooks in that it is essentially a one-man job. This has made it possible to coordinate different sections of the book very closely. Thus, related subjects in different parts of the book are treated from the same viewpoint and are fully cross-referenced, duplication is avoided, and there is no trouble from gaps such as inevitably appear when the writer of one section is not familiar with the details of all of the other sections. At the same time, the one-man approach has necessarily limited the range of topics covered to the basic engineering principles which the radio man thinks of as constituting radio engineering. Related communication fields, such as wire communication, as well as the physics aspects of radio and electronics, including particularly the physics literature and special applications to problems in physics, are covered only very incompletely. This is not because these things are not important, but rather because what one man can do in one book is inevitably limited. Within the scope of the book, the coverage will be found to be almost complete, however.

There is one omission that the inquiring reader will wonder about, namely, the lack of a section on television. This was to be Section 14, but as the author became deeply involved in war work early in 1942, it became impossible to follow the original plans without delaying publication indefinitely. Television material does appear in sections dealing with amplifiers, antennas, etc., but the coordinated presentation of television systems is regrettably missing.

In selecting the material to go into "Radio Engineers' Handbook" it has of course been necessary to avoid any material that even hints at the new developments that have evolved from the needs of war. Certain topics were in fact deliberately deemphasized to avoid indicating the directions in which the new developments are being made. Accordingly, even when readers are disappointed in not finding descriptions and explanations of the new techniques evolved from the war effort, they should not feel that the author is oblivious to them. Quite the contrary is the case.

Many people and organizations have helped in one way and another to bringing the "Radio Engineers' Handbook" into being. The most important contribution was made by Dr. Karl Spangenberg, the author's associate at Stanford University, who wrote Section 4 on Vacuum Tubes and Electronics, and who was largely responsible for the material on wave guides and cavity resonators in Section 3 and the collection of formulas giving inductance, mutual inductance, and capacity in Section 2.

Acknowledgment is also due Dr. W. Y. Pan and Donald MacQuivey, who prepared most of the illustrations, and to W. S-Y Yeh, T. H. Pi, C. K. Chang, Robert Soderman, O. G. Villard, Jr., Dr. E. A. Yunker, Clark Cahill, Robert Barnes, and others who assisted in various ways at one time or another. Thanks are also due the International Standard Electric Corporation for making available the data on electron lenses that appears in Section 4, and to the General Radio Company, Driver Harris Co., RCA Manufacturing Co., Inc., Ohmite Manufacturing Company, and Phelps Dodge Copper Products Corporation for permission to reproduce various charts and tables.

In a work of this type some mistakes in equations, curves, etc., will inevitably occur no matter how much care is taken. Where these are found, the author would appreciate their being called to his attention so that subsequent printings may be corrected.

FREDERICK EMMONS TERMAN.

CAMBRIDGE, MASSACHUSETTS,
May, 1943.

CONTENTS

PAGE

SECTION 13

MEASUREMENTS

SECTION 1

TABLES, MATHEMATICAL RELATIONS, AND UNITS

TABLES

1. Decibel Table

Db	Current and voltage ratio		Power ratio		Db	Current and voltage ratio		Power ratio	
	Gain	Loss	Gain	Loss		Gain	Loss	Gain	Loss
0.1	1.01	0.989	1.02	0.977	8.0	2.51	0.398	6.31	0.158
0.2	1.02	0.977	1.05	0.955	8.5	2.66	0.376	7.08	0.141
0.3	1.03	0.966	1.07	0.933	9.0	2.82	0.355	7.94	0.126
0.4	1.05	0.955	1.10	0.912	9.5	2.98	0.335	8.91	0.112
0.5	1.06	0.944	1.12	0.891	10.0	3.16	0.316	10.00	0.100
0.6	1.07	0.933	1.15	0.871	11.0	3.55	0.282	12.6	0.079
0.7	1.08	0.923	1.17	0.851	12.0	3.98	0.251	15.8	0.063
0.8	1.10	0.912	1.20	0.832	13.0	4.47	0.224	19.9	0.050
0.9	1.11	0.902	1.23	0.813	14.0	5.01	0.199	25.1	0.040
1.0	1.12	0.891	1.26	0.794	15.0	5.62	0.178	31.6	0.032
1.1	1.13	0.881	1.29	0.776	16.0	6.31	0.158	39.8	0.025
1.2	1.15	0.871	1.32	0.759	17.0	7.08	0.141	50.1	0.020
1.3	1.16	0.861	1.35	0.741	18.0	7.94	0.126	63.1	0.016
1.4	1.17	0.851	1.38	0.724	19.0	8.91	0.112	79.4	0.013
1.5	1.19	0.841	1.41	0.708	20.0	10.00	0.100	100.0	0.010
1.6	1.20	0.832	1.44	0.692	25.0	17.8	0.056	3.16×10^{2}	3.16×10^{-3}
1.7	1.22	0.822	1.48	0.676	30.0	31.6	0.032	10^{3}	10^{-3}
1.8	1.23	0.813	1.51	0.661	35.0	56.2	0.018	3.16×10^{3}	3.16×10^{-4}
1.9	1.24	0.803	1.55	0.646	40.0	100.0	0.010	10^{4}	10^{-4}
2.0	1.26	0.794	1.58	0.631	45.0	177.8	0.006	3.16×10^{4}	3.16×10^{-5}
2.2	1.29	0.776	1.66	0.603	50.0	316	0.003	10^{5}	10^{-5}
2.4	1.32	0.759	1.74	0.575	55.0	562	0.002	3.16×10^{5}	3.16×10^{-6}
2.6	1.35	0.741	1.82	0.550	60.0	1,000	0.001	10^{6}	10^{-6}
2.8	1.38	0.724	1.90	0.525	65.0	1,770	0.0006	3.16×10^{6}	3.16×10^{-7}
3.0	1.41	0.708	1.99	0.501	70.0	3,160	0.0003	10^{7}	10^{-7}
3.2	1.44	0.692	2.09	0.479	75.0	5,620	0.0002	3.16×10^{7}	3.16×10^{-8}
3.4	1.48	0.676	2.19	0.457	80.0	10,000	0.0001	10^{8}	10^{-8}
3.6	1.51	0.661	2.29	0.436	85.0	17,800	0.00006	3.16×10^{8}	3.16×10^{-9}
3.8	1.55	0.646	2.40	0.417	90.0	31,600	0.00003	10^{9}	10^{-9}
4.0	1.58	0.631	2.51	0.398	95.0	56,200	0.00002	3.16×10^{9}	3.16×10^{-10}
4.2	1.62	0.617	2.63	0.380	100.0	100,000	0.00001	10^{10}	10^{-10}
4.4	1.66	0.603	2.75	0.363	105.0	178,000	0.000006	3.16×10^{10}	3.16×10^{-11}
4.6	1.70	0.589	2.88	0.347	110.0	316,000	0.000003	10^{11}	10^{-11}
4.8	1.74	0.575	3.02	0.331	115.0	562,000	0.000002	3.16×10^{11}	3.16×10^{-12}
5.0	1.78	0.562	3.16	0.316	120.0	1,000,000	0.000001	10^{12}	10^{-12}
5.5	1.88	0.531	3.55	0.282	130.0	3.16×10^{6}	3.16×10^{-7}	10^{13}	10^{-13}
6.0	1.99	0.501	3.98	0.251	140.0	10^{7}	10^{-7}	10^{14}	10^{-14}
6.5	2.11	0.473	4.47	0.224	150.0	3.16×10^{7}	3.16×10^{-8}	10^{15}	10^{-15}
7.0	2.24	0.447	5.01	0.199	160.0	10^{8}	10^{-8}	10^{16}	10^{-16}
7.5	2.37	0.422	5.62	0.178	170.0	3.16×10^{8}	3.16×10^{-9}	10^{17}	10^{-17}

2. Natural Sines and Cosines*

NOTE.—For cosines use right-hand column of degrees and lower line of tenths.

Deg	°0.0	°0.1	°0.2	°0.3	°0.4	°0.5	°0.6	°0.7	°0.8	°0.9	
0°	0.0000	0.0017	0.0035	0.0052	0.0070	0.0087	0.0105	0.0122	0.0140	0.0157	89
1	0.0175	0.0192	0.0209	0.0227	0.0244	0.0262	0.0279	0.0297	0.0314	0.0332	88
2	0.0349	0.0366	0.0384	0.0401	0.0419	0.0436	0.0454	0.0471	0.0488	0.0506	87
3	0.0523	0.0541	0.0558	0.0576	0.0593	0.0610	0.0628	0.0645	0.0663	0.0680	86
4	0.0698	0.0715	0.0732	0.0750	0.0767	0.0785	0.0802	0.0819	0.0837	0.0854	85
5	0.0872	0.0889	0.0906	0.0924	0.0941	0.0958	0.0976	0.0993	0.1011	0.1028	84
6	0.1045	0.1063	0.1080	0.1097	0.1115	0.1132	0.1149	0.1167	0.1184	0.1201	83
7	0.1219	0.1236	0.1253	0.1271	0.1288	0.1305	0.1323	0.1340	0.1357	0.1374	82
8	0.1392	0.1409	0.1426	0.1444	0.1461	0.1478	0.1495	0.1513	0.1530	0.1547	81
9	0.1564	0.1582	0.1599	0.1616	0.1633	0.1650	0.1668	0.1685	0.1702	0.1719	80°
10°	0.1736	0.1754	0.1771	0.1788	0.1805	0.1822	0.1840	0.1857	0.1874	0.1891	79
11	0.1908	0.1925	0.1942	0.1959	0.1977	0.1994	0.2011	0.2028	0.2045	0.2062	78
12	0.2079	0.2096	0.2113	0.2130	0.2147	0.2164	0.2181	0.2198	0.2215	0.2232	77
13	0.2250	0.2267	0.2284	0.2300	0.2317	0.2334	0.2351	0.2368	0.2385	0.2402	76
14	0.2419	0.2436	0.2453	0.2470	0.2487	0.2504	0.2521	0.2538	0.2554	0.2571	75
15	0.2588	0.2605	0.2622	0.2639	0.2656	0.2672	0.2689	0.2706	0.2723	0.2740	74
16	0.2756	0.2773	0.2790	0.2807	0.2823	0.2840	0.2857	0.2874	0.2890	0.2907	73
17	0.2924	0.2940	0.2957	0.2974	0.2990	0.3007	0.3024	0.3040	0.3057	0.3074	72
18	0.3090	0.3107	0.3123	0.3140	0.3156	0.3173	0.3190	0.3206	0.3223	0.3239	71
19	0.3256	0.3272	0.3289	0.3305	0.3322	0.3338	0.3355	0.3371	0.3387	0.3404	70°
20°	0.3420	0.3437	0.3453	0.3469	0.3486	0.3502	0.3518	0.3535	0.3551	0.3567	69
21	0.3584	0.3600	0.3616	0.3633	0.3649	0.3665	0.3681	0.3697	0.3714	0.3730	68
22	0.3746	0.3762	0.3778	0.3795	0.3811	0.3827	0.3843	0.3859	0.3875	0.3891	67
23	0.3907	0.3923	0.3939	0.3955	0.3971	0.3987	0.4003	0.4019	0.4035	0.4051	66
24	0.4067	0.4083	0.4099	0.4115	0.4131	0.4147	0.4163	0.4179	0.4195	0.4210	65
25	0.4226	0.4242	0.4258	0.4274	0.4289	0.4305	0.4321	0.4337	0.4352	0.4368	64
26	0.4384	0.4399	0.4415	0.4431	0.4446	0.4462	0.4478	0.4493	0.4509	0.4524	63
27	0.4540	0.4555	0.4571	0.4586	0.4602	0.4617	0.4633	0.4648	0.4664	0.4679	62
28	0.4695	0.4710	0.4726	0.4741	0.4756	0.4772	0.4787	0.4802	0.4818	0.4833	61
29	0.4848	0.4863	0.4879	0.4894	0.4909	0.4924	0.4939	0.4955	0.4970	0.4985	60°
30°	0.5000	0.5015	0.5030	0.5045	0.5060	0.5075	0.5900	0.5105	0.5120	0.5135	59
31	0.5150	0.5165	0.5180	0.5195	0.5210	0.5225	0.5240	0.5255	0.5270	0.5284	58
32	0.5299	0.5314	0.5329	0.5344	0.5358	0.5373	0.5388	0.5402	0.5417	0.5432	57
33	0.5446	0.5461	0.5476	0.5490	0.5505	0.5519	0.5534	0.5548	0.5563	0.5577	56
34	0.5592	0.5606	0.5621	0.5635	0.5650	0.5664	0.5678	0.5693	0.5707	0.5721	55
35	0.5736	0.5750	0.5764	0.5779	0.5793	0.5807	0.5821	0.5835	0.5850	0.5864	54
36	0.5878	0.5892	0.5906	0.5920	0.5934	0.5948	0.5962	0.5976	0.5990	0.6004	53
37	0.6018	0.6032	0.6046	0.6060	0.6074	0.6088	0.6101	0.6115	0.6129	0.6143	52
38	0.6157	0.6170	0.6184	0.6198	0.6211	0.6225	0.6239	0.6252	0.6266	0.6280	51
39	0.6293	0.6307	0.6320	0.6334	0.6347	0.6361	0.6374	0.6388	0.6401	0.6414	50°
40°	0.6428	0.6441	0.6455	0.6468	0.6481	0.6494	0.6508	0.6521	0.6534	0.6547	49
41	0.6561	0.6574	0.6587	0.6600	0.6613	0.6626	0.6639	0.6652	0.6665	0.6678	48
42	0.6691	0.6704	0.6717	0.6730	0.6743	0.6756	0.6769	0.6782	0.6794	0.6807	47
43	0.6820	0.6833	0.6845	0.6858	0.6871	0.6884	0.6896	0.6909	0.6921	0.6934	46
44	0.6947	0.6959	0.6972	0.6984	0.6997	0.7009	0.7022	0.7034	0.7046	0.7059	45
	°1.0	°0.9	°0.8	°0.7	°0.6	°0.5	°0.4	°0.3	°0.2	°0.1	Deg

* From "Standard Handbook for Electrical Engineers," 7th ed.

2. Natural Sines and Cosines.*—(*Concluded*)

Deg	°0.0	°0.1	°0.2	°0.3	°0.4	°0.5	°0.6	°0.7	°0.8	°0.9	
45	0.7071	0.7083	0.7096	0.7108	0.7120	0.7133	0.7145	0.7157	0.7169	0.7181	44
46	0.7193	0.7206	0.7218	0.7230	0.7242	0.7254	0.7266	0.7278	0.7290	0.7302	43
47	0.7314	0.7325	0.7337	0.7349	0.7361	0.7373	0.7385	0.7396	0.7408	0.7420	42
48	0.7431	0.7443	0.7455	0.7466	0.7478	0.7490	0.7501	0.7513	0.7524	0.7536	41
49	0.7547	0.7559	0.7570	0.7581	0.7593	0.7604	0.7615	0.7627	0.7638	0.7649	40°
50°	0.7660	0.7672	0.7683	0.7694	0.7705	0.7716	0.7727	0.7738	0.7749	0.7760	39
51	0.7771	0.7782	0.7793	0.7804	0.7815	0.7826	0.7837	0.7848	0.7859	0.7869	38
52	0.7880	0.7891	0.7902	0.7912	0.7923	0.7934	0.7944	0.7955	0.7965	0.7976	37
53	0.7986	0.7997	0.8007	0.8018	0.8028	0.8039	0.8049	0.8059	0.8070	0.8080	36
54	0.8090	0.8100	0.8111	0.8121	0.8131	0.8141	0.8151	0.8161	0.8171	0.8181	35
55	0.8192	0.8202	0.8211	0.8221	0.8231	0.8241	0.8251	0.8261	0.8271	0.8281	34
56	0.8290	0.8300	0.8310	0.8320	0.8329	0.8339	0.8348	0.8358	0.8368	0.8377	33
57	0.8387	0.8396	0.8406	0.8415	0.8425	0.8434	0.8443	0.8453	0.8462	0.8471	32
58	0.8480	0.8490	0.8499	0.8508	0.8517	0.8526	0.8536	0.8545	0.8554	0.8563	31
59	0.8572	0.8581	0.8590	0.8599	0.8607	0.8616	0.8625	0.8634	0.8643	0.8652	30°
60°	0.8660	0.8669	0.8678	0.8686	0.8695	0.8704	0.8712	0.8721	0.8729	0.8738	29
61	0.8746	0.8755	0.8763	0.8771	0.8780	0.8788	0.8796	0.8805	0.8813	0.8821	28
62	0.8829	0.8838	0.8846	0.8854	0.8862	0.8870	0.8878	0.8886	0.8894	0.8902	27
63	0.8910	0.8918	0.8926	0.8934	0.8942	0.8949	0.8957	0.8965	0.8973	0.8980	26
64	0.8988	0.8996	0.9003	0.9011	0.9018	0.9026	0.9033	0.9041	0.9048	0.9056	25
65	0.9063	0.9070	0.9078	0.9085	0.9092	0.9100	0.9107	0.9114	0.9121	0.9128	24
66	0.9135	0.9143	0.9150	0.9157	0.9164	0.9171	0.9178	0.9184	0.9191	0.9198	23
67	0.9205	0.9212	0.9219	0.9225	0.9232	0.9239	0.9245	0.9252	0.9259	0.9265	22
68	0.9272	0.9278	0.9285	0.9291	0.9298	0.9304	0.9311	0.9317	0.9323	0.9330	21
69	0.9336	0.9342	0.9348	0.9354	0.9361	0.9367	0.9373	0.9379	0.9385	0.9391	20°
70°	0.9397	0.9403	0.9409	0.9415	0.9421	0.9426	0.9432	0.9438	0.9444	0.9449	19
71	0.9455	0.9461	0.9466	0.9472	0.9478	0.9483	0.9489	0.9494	0.9500	0.9505	18
72	0.9511	0.9516	0.9521	0.9527	0.9532	0.9537	0.9542	0.9548	0.9553	0.9558	17
73	0.9563	0.9568	0.9573	0.9578	0.9583	0.9588	0.9593	0.9598	0.9603	0.9608	16
74	0.9613	0.9617	0.9622	0.9627	0.9632	0.9636	0.9641	0.9646	0.9650	0.9655	15
75	0.9659	0.9664	0.9668	0.9673	0.9677	0.9681	0.9686	0.9690	0.9694	0.9699	14
76	0.9703	0.9707	0.9711	0.9715	0.9720	0.9724	0.9728	0.9732	0.9736	0.9740	13
77	0.9744	0.9748	0.9751	0.9755	0.9759	0.9763	0.9767	0.9770	0.9774	0.9778	12
78	0.9781	0.9785	0.9789	0.9792	0.9796	0.9799	0.9803	0.9806	0.9810	0.9813	11
79	0.9816	0.9820	0.9823	0.9826	0.9829	0.9833	0.9836	0.9839	0.9842	0.9845	10°
80°	0.9848	0.9851	0.9854	0.9857	0.9860	0.9863	0.9866	0.9869	0.9871	0.9874	9
81	0.9877	0.9880	0.9882	0.9885	0.9888	0.9890	0.9893	0.9895	0.9898	0.9900	8
82	0.9903	0.9905	0.9907	0.9910	0.9912	0.9914	0.9917	0.9919	0.9921	0.9923	7
83	0.9925	0.9928	0.9930	0.9932	0.9934	0.9936	0.9938	0.9940	0.9942	0.9943	6
84	0.9945	0.9947	0.9949	0.9951	0.9952	0.9954	0.9956	0.9957	0.9959	0.9960	5
85	0.9962	0.9963	0.9965	0.9966	0.9968	0.9969	0.9971	0.9972	0.9973	0.9974	4
86	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982	0.9983	0.9984	0.9985	3
87	0.9986	0.9987	0.9988	0.9989	0.9990	0.9990	0.9991	0.9992	0.9993	0.9993	2
88	0.9994	0.9995	0.9995	0.9996	0.9996	0.9997	0.9997	0.9997	0.9998	0.9998	1
89	0.9998	0.9999	0.9999	0.9999	0.9999	1.000	1.000	1.000	1.000	1.000	0°
	°1.0	°0.9	°0.8	°0.7	°0.6	°0.5	°0.4	°0.3	°0.2	°0.1	Deg

* From "Standard Handbook for Electrical Engineers," 7th ed.

3. Natural Tangents and Cotangents*

NOTE.—For cotangents use right-hand column of degrees and lower line of tenths.

Deg	°0.0	°0.1	°0.2	°0.3	°0.4	°0.5	°0.6	°0.7	°0.8	°0.9	
0°	0.0000	0.0017	0.0035	0.0052	0.0070	0.0087	0.0105	0.0122	0.0140	0.0157	89
1	0.0175	0.0192	0.0209	0.0227	0.0244	0.0262	0.0279	0.0297	0.0314	0.0332	88
2	0.0349	0.0367	0.0384	0.0402	0.0419	0.0437	0.0454	0.0472	0.0489	0.0507	87
3	0.0524	0.0542	0.0559	0.0577	0.0594	0.0612	0.0629	0.0647	0.0664	0.0682	86
4	0.0699	0.0717	0.0734	0.0752	0.0769	0.0787	0.0805	0.0822	0.0840	0.0857	85
5	0.0875	0.0892	0.0910	0.0928	0.0945	0.0963	0.0981	0.0998	0.1016	0.1033	84
6	0.1051	0.1069	0.1086	0.1104	0.1122	0.1139	0.1157	0.1175	0.1192	0.1210	83
7	0.1228	0.1246	0.1263	0.1281	0.1299	0.1317	0.1334	0.1352	0.1370	0.1388	82
8	0.1405	0.1423	0.1441	0.1459	0.1477	0.1495	0.1512	0.1530	0.1548	0.1566	81
9	0.1584	0.1602	0.1620	0.1638	0.1655	0.1673	0.1691	0.1709	0.1727	0.1745	80°
10°	0.1763	0.1781	0.1799	0.1817	0.1835	0.1853	0.1871	0.1890	0.1908	0.1926	79
11	0.1944	0.1962	0.1980	0.1998	0.2016	0.2035	0.2053	0.2071	0.2089	0.2107	78
12	0.2126	0.2144	0.2162	0.2180	0.2199	0.2217	0.2235	0.2254	0.2272	0.2290	77
13	0.2309	0.2327	0.2345	0.2364	0.2382	0.2401	0.2419	0.2438	0.2456	0.2475	76
14	0.2493	0.2512	0.2530	0.2549	0.2568	0.2586	0.2605	0.2623	0.2643	0.2661	75
15	0.2679	0.2698	0.2717	0.2736	0.2754	0.2773	0.2792	0.2811	0.2830	0.2849	74
16	0.2867	0.2886	0.2905	0.2924	0.2943	0.2962	0.2981	0.3000	0.3019	0.3038	73
17	0.3057	0.3076	0.3096	0.3115	0.3134	0.3153	0.3172	0.3191	0.3211	0.3230	72
18	0.3249	0.3269	0.3288	0.3307	0.3327	0.3346	0.3365	0.3385	0.3404	0.3424	71
19	0.3443	0.3463	0.3482	0.3502	0.3522	0.3541	0.3561	0.3581	0.3600	0.3620	70°
20°	0.3640	0.3659	0.3679	0.3699	0.3719	0.3739	0.3759	0.3779	0.3799	0.3819	69
21	0.3839	0.3859	0.3879	0.3899	0.3919	0.3939	0.3959	0.3979	0.4000	0.4020	68
22	0.4040	0.4061	0.4081	0.4101	0.4122	0.4142	0.4163	0.4183	0.4204	0.4224	67
23	0.4245	0.4265	0.4286	0.4307	0.4327	0.4348	0.4369	0.4390	0.4411	0.4431	66
24	0.4452	0.4473	0.4494	0.4515	0.4536	0.4557	0.4578	0.4599	0.4621	0.4642	65
25	0.4663	0.4684	0.4706	0.4727	0.4748	0.4770	0.4791	0.4813	0.4834	0.4856	64
26	0.4877	0.4899	0.4921	0.4942	0.4964	0.4986	0.5008	0.5029	0.5051	0.5073	63
27	0.5095	0.5117	0.5139	0.5161	0.5184	0.5206	0.5228	0.5250	0.5272	0.5295	62
28	0.5317	0.5340	0.5362	0.5384	0.5407	0.5430	0.5452	0.5475	0.5498	0.5520	61
29	0.5543	0.5566	0.5589	0.5612	0.5635	0.5658	0.5681	0.5704	0.5727	0.5750	60°
30°	0.5774	0.5797	0.5820	0.5844	0.5867	0.5890	0.5914	0.5938	0.5961	0.5985	59
31	0.6009	0.6032	0.6056	0.6080	0.6104	0.6128	0.6152	0.6176	0.6200	0.6224	58
32	0.6249	0.6273	0.6297	0.6322	0.6346	0.6371	0.6395	0.6420	0.6445	0.6469	57
33	0.6494	0.6519	0.6544	0.6569	0.6594	0.6619	0.6644	0.6669	0.6694	0.6720	56
34	0.6745	0.6771	0.6796	0.6822	0.6847	0.6873	0.6899	0.6924	0.6950	0.6976	55
35	0.7002	0.7028	0.7054	0.7080	0.7107	0.7133	0.7159	0.7186	0.7212	0.7239	54
36	0.7265	0.7292	0.7319	0.7346	0.7373	0.7400	0.7427	0.7454	0.7481	0.7508	53
37	0.7536	0.7563	0.7590	0.7618	0.7646	0.7673	0.7701	0.7729	0.7757	0.7785	52
38	0.7813	0.7841	0.7869	0.7898	0.7926	0.7954	0.7983	0.8012	0.8040	0.8069	51
39	0.8098	0.8127	0.8156	0.8185	0.8214	0.8243	0.8273	0.8302	0.8332	0.8361	50°
40°	0.8391	0.8421	0.8451	0.8481	0.8511	0.8541	0.8571	0.8601	0.8632	0.8662	49
41	0.8693	0.8724	0.8754	0.8785	0.8816	0.8847	0.8878	0.8910	0.8941	0.8972	48
42	0.9004	0.9036	0.9067	0.9099	0.9131	0.9163	0.9195	0.9228	0.9260	0.9293	47
43	0.9325	0.9358	0.9391	0.9424	0.9457	0.9490	0.9523	0.9556	0.9590	0.9623	46
44	0.9657	0.9691	0.9725	0.9759	0.9793	0.9827	0.9861	0.9896	0.9930	0.9965	45
	°1.0	°0.9	°0.8	°0.7	°0.6	°0.5	°0.4	°0.3	°0.2	°0.1	Deg

* From "Standard Handbook for Electrical Engineers," 7th ed.

3. Natural Tangents and Cotangents.*—(*Concluded*)

Deg	°0.0	°0.1	°0.2	°0.3	°0.4	°0.5	°0.6	°0.7	°0.8	°0.9	
45	1.0000	1.0035	1.0070	1.0105	1.0141	1.0176	1.0212	1.0247	1.0283	1.0319	44
46	1.0355	1.0392	1.0428	1.0464	1.0501	1.0538	1.0575	1.0612	1.0649	1.0686	43
47	1.0724	1.0761	1.0799	1.0837	1.0875	1.0913	1.0951	1.0990	1.1028	1.1067	42
48	1.1106	1.1145	1.1184	1.1224	1.1263	1.1303	1.1343	1.1383	1.1423	1.1463	41
49	1.1504	1.1544	1.1585	1.1626	1.1667	1.1708	1.1750	1.1792	1.1833	1.1875	40°
50°	1.1918	1.1960	1.2002	1.2045	1.2088	1.2131	1.2174	1.2218	1.2261	1.2305	39
51	1.2349	1.2393	1.2437	1.2482	1.2527	1.2572	1.2617	1.2662	1.2708	1.2753	38
52	1.2799	1.2846	1.2892	1.2938	1.2985	1.3032	1.3079	1.3127	1.3175	1.3222	37
53	1.3270	1.3319	1.3367	1.3416	1.3465	1.3514	1.3564	1.3613	1.3663	1.3713	36
54	1.3764	1.3814	1.3865	1.3916	1.3968	1.4019	1.4071	1.4124	1.4176	1.4229	35
55	1.4281	1.4335	1.4388	1.4442	1.4496	1.4550	1.4605	1.4659	1.4715	1.4770	34
56	1.4326	1.4882	1.4938	1.4994	1.5051	1.5108	1.5166	1.5224	1.5282	1.5340	33
57	1.5399	1.5458	1.5517	1.5577	1.5637	1.5697	1.5757	1.5818	1.5880	1.5941	32
58	1.6003	1.6066	1.6128	1.6191	1.6255	1.6319	1.6383	1.6447	1.6512	1.6577	31
59	1.6643	1.6709	1.6775	1.6842	1.6909	1.6977	1.7045	1.7113	1.7182	1.7251	30°
60°	1.7321	1.7391	1.7461	1.7532	1.7603	1.7675	1.7747	1.7820	1.7893	1.7966	29
61	1.8040	1.8115	1.8190	1.8265	1.8341	1.8418	1.8495	1.8572	1.8650	1.8728	28
62	1.8807	1.8887	1.8967	1.9047	1.9128	1.9210	1.9292	1.9375	1.9458	1.9542	27
63	1.9626	1.9711	1.9797	1.9883	1.9970	2.0057	2.0145	2.0233	2.0323	2.0413	26
64	2.0503	2.0594	2.0686	2.0778	2.0872	2.0965	2.1060	2.1155	2.1251	2.1348	25
65	2.1445	2.1543	2.1642	2.1742	2.1842	2.1943	2.2045	2.2148	2.2251	2.2355	24
66	2.2460	2.2566	2.2673	2.2781	2.2889	2.2998	2.3109	2.3220	2.3332	2.3445	23
67	2.3599	2.3673	2.3789	2.3906	2.4023	2.4142	2.4262	2.4383	2.4504	2.4627	22
68	2.4751	2.4876	2.5002	2.5129	2.5257	2.5386	2.5517	2.5649	2.5782	2.5916	21
69	2.6051	2.6187	2.6325	2.6464	2.6605	2.6746	2.6889	2.7034	2.7179	2.7326	20°
70°	2.7475	2.7625	2.7776	2.7929	2.8083	2.8239	2.8397	2.8556	2.8716	2.8878	19
71	2.9042	2.9208	2.9375	2.9544	2.9714	2.9887	3.0061	3.0237	3.0415	3.0595	18
72	3.0777	3.0961	3.1146	3.1334	3.1524	3.1716	3.1910	3.2106	3.2305	2.2506	17
73	3.2709	3.2914	3.3122	3.3332	3.3544	3.3759	3.3977	3.4197	3.4420	3.4646	16
74	3.4874	3.5105	3.5339	3.5576	3.5816	3.6059	3.6305	3.6554	3.6806	3.7062	15
75	3.7321	3.7583	3.7848	3.8118	3.8391	3.8667	3.8947	3.9232	3.9520	3.9812	14
76	4.0108	4.0408	4.0713	4.1022	4.1335	4.1653	4.1976	4.2303	4.2635	4.2972	13
77	4.3315	4.3662	4.4015	4.4374	4.4737	4.5107	4.5483	4.5864	4.6252	4.6646	12
78	4.7046	4.7453	4.7867	4.8288	4.8716	4.9152	4.9594	5.0045	5.0504	5.0970	11
79	5.1446	5.1929	5.2422	5.2924	5.3435	5.3955	5.4486	5.5026	5.5578	5.6140	10°
80°	5.6713	5.7297	5.7894	5.8502	5.9124	5.9758	6.0405	6.1066	6.1742	6.2433	9
81	6.3138	6.3859	6.4596	6.5350	6.6122	6.6912	6.7720	6.8548	6.9395	7.0264	8
82	7.1154	7.2066	7.3002	7.3962	7.4947	7.5958	7.6996	7.8062	7.9158	8.0285	7
83	8.1443	8.2636	8.3863	8.5126	8.6427	8.7769	8.9152	9.0579	9.2052	9.3572	6
84	9.5144	9.677	9.845	10.02	10.20	10.39	10.58	10.78	10.99	11.20	5
85	11.43	11.66	11.91	12.16	12.43	12.71	13.00	13.30	13.62	13.95	4
86	14.30	14.67	15.06	15.46	15.89	16.35	16.83	17.34	17.89	18.46	3
87	19.08	19.74	20.45	21.20	22.02	22.90	23.86	24.90	26.03	27.27	2
88	28.64	30.14	31.82	33.69	35.80	38.19	40.92	44.07	47.74	52.08	1
89	57.29	63.66	71.62	81.85	94.49	114.6	143.2	191.0	286.5	573.0	0°
	°1.0	°0.9	°0.8	°0.7	°0.6	°0.5	°0.4	°0.3	°0.2	°0.1	Deg

* From "Standard Handbook for Electrical Engineers," 7th ed.

4. Common Logarithms of Numbers*

N	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396

* From "Standard Handbook for Electrical Engineers," 7th ed.

4. Common Logarithms of Numbers.*—*(Concluded)*

N	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

* From "Standard Handbook for Electrical Engineers," 7th ed.

5. Natural, Napierian, or Hyperbolic Logarithms*

N	0	1	2	3	4	5	6	7	8	9
0	$-\infty$	0.0000	0.6931	1.0986	1.3863	1.6094	1.7918	1.9459	2.0794	2.1972
10	2.3026	2.3979	2.4849	2.5649	2.6391	2.7081	2.7726	2.8332	2.8904	2.9444
20	2.9957	3.0445	3.0910	3.1355	3.1781	3.2189	3.2581	3.2958	3.3322	3.3673
30	3.4012	3.4340	3.4657	3.4965	3.5264	3.5553	3.5835	3.6109	3.6376	3.6636
40	3.6889	3.7136	3.7377	3.7612	3.7842	3.8067	3.8286	3.8501	3.8712	3.8918
50	3.9120	3.9318	3.9512	3.9703	3.9890	4.0073	4.0254	4.0431	4.0604	4.0775
60	4.0943	4.1109	4.1271	4.1431	4.1589	4.1744	4.1897	4.2047	4.2195	4.2341
70	4.2485	4.2627	4.2767	4.2905	4.3041	4.3175	4.3307	4.3438	4.3567	4.3694
80	4.3820	4.3944	4.4067	4.4188	4.4308	4.4427	4.4543	4.4659	4.4773	4.4886
90	4.4998	4.5109	4.5218	4.5326	4.5433	4.5539	4.5643	4.5747	4.5850	4.5951
100	4.6052	4.6151	4.6250	4.6347	4.6444	4.6540	4.6634	4.6728	4.6821	4.6913
110	4.7005	4.7095	4.7185	4.7274	4.7362	4.7449	4.7536	4.7622	4.7707	4.7791
120	4.7875	4.7958	4.8040	4.8122	4.8203	4.8283	4.8363	4.8442	4.8520	4.8598
130	4.8675	4.8752	4.8828	4.8903	4.8978	4.9053	4.9127	4.9200	4.9273	4.9345
140	4.9416	4.9488	4.9558	4.9628	4.9698	4.9767	4.9836	4.9904	4.9972	5.0039
150	5.0106	5.0173	5.0239	5.0304	5.0370	5.0434	5.0499	5.0562	5.0626	5.0689
160	5.0752	5.0814	5.0876	5.0938	5.0999	5.1059	5.1120	5.1180	5.1240	5.1299
170	5.1358	5.1417	5.1475	5.1533	5.1591	5.1648	5.1705	5.1761	5.1818	5.1874
180	5.1930	5.1985	5.2040	5.2095	5.2149	5.2204	5.2257	5.2311	5.2364	5.2417
190	5.2470	5.2523	5.2575	5.2627	5.2679	5.2730	5.2781	5.2832	5.2883	5.2933
200	5.2983	5.3033	5.3083	5.3132	5.3181	5.3230	5.3279	5.3327	5.3375	5.3423
210	5.3471	5.3519	5.3566	5.3613	5.3660	5.3706	5.3753	5.3799	5.3845	5.3891
220	5.3936	5.3982	5.4027	5.4072	5.4116	5.4161	5.4205	5.4250	5.4293	5.4337
230	5.4381	5.4424	5.4467	5.4510	5.4553	5.4596	5.4638	5.4681	5.4723	5.4765
240	5.4806	5.4848	5.4889	5.4931	5.4972	5.5013	5.5053	5.5094	5.5134	5.5175
250	5.5215	5.5255	5.5294	5.5334	5.5373	5.5413	5.5452	5.5491	5.5530	5.5568
260	5.5607	5.5645	5.5683	5.5722	5.5759	5.5797	5.5835	5.5872	5.5910	5.5947
270	5.5984	5.6021	5.6058	5.6095	5.6131	5.6168	5.6204	5.6240	5.6276	5.6312
280	5.6348	5.6384	5.6419	5.6454	5.6490	5.6525	5.6560	5.6595	5.6630	5.6664
290	5.6699	5.6733	5.6768	5.6802	5.6836	5.6870	5.6904	5.6937	5.6971	5.7004
300	5.7038	5.7071	5.7104	5.7137	5.7170	5.7203	5.7236	5.7268	5.7301	5.7333
310	5.7366	5.7398	5.7430	5.7462	5.7494	5.7526	5.7557	5.7589	5.7621	5.7652
320	5.7683	5.7714	5.7746	5.7777	5.7807	5.7838	5.7869	5.7900	5.7930	5.7961
330	5.7991	5.8021	5.8051	5.8081	5.8111	5.8141	5.8171	5.8201	5.8230	5.8260
340	5.8289	5.8319	5.8348	5.8377	5.8406	5.8435	5.8464	5.8493	5.8522	5.8551
350	5.8579	5.8608	5.8636	5.8665	5.8693	5.8721	5.8749	5.8777	5.8805	5.8833
360	5.8861	5.8889	5.8916	5.8944	5.8972	5.8999	5.9026	5.9054	5.9081	5.9108
370	5.9135	5.9162	5.9189	5.9216	5.9243	5.9269	5.9296	5.9322	5.9349	5.9375
380	5.9402	5.9428	5.9454	5.9480	5.9506	5.9532	5.9558	5.9584	5.9610	5.9636
390	5.9661	5.9687	5.9713	5.9738	5.9764	5.9789	5.9814	5.9839	5.9865	5.9890
400	5.9915	5.9940	5.9965	5.9989	6.0014	6.0039	6.0064	6.0088	6.0113	6.0137
410	6.0162	6.0186	6.0210	6.0234	6.0259	6.0283	6.0307	6.0331	6.0355	6.0379
420	6.0403	6.0426	6.0450	6.0474	6.0497	6.0521	6.0544	6.0568	6.0591	6.0615
430	6.0638	6.0661	6.0684	6.0707	6.0730	6.0753	6.0776	6.0799	6.0822	6.0845
440	6.0868	6.0890	6.0913	6.0936	6.0958	0.0981	6.1003	6.1026	6.1048	6.1070
450	6.1092	6.1115	6.1137	6.1159	6.1181	6.1203	6.1225	6.1247	6.1269	6.1291
460	6.1312	6.1334	6.1356	6.1377	6.1399	6.1420	6.1442	6.1463	6.1485	6.1506
470	6.1527	6.1549	6.1570	6.1591	6.1612	6.1633	6.1654	6.1675	6.1696	6.1717
480	6.1738	6.1759	6.1779	6.1800	6.1821	6.1841	6.1862	6.1883	6.1903	6.1924
490	6.1944	6.1964	6.1985	6.2005	6.2025	6.2046	6.2066	6.2086	6.2106	6.2126

* From "Standard Handbook for Electrical Engineers," 7th ed.

NOTE 1: Moving the decimal point n places to the right (or left) in the number is equivalent to adding (or subtracting) n times 2.3026.

NOTE 2:

$$\log_e x = 2.3026 \log_{10} x$$
$$\log_{10} x = 0.4343 \log_e x$$
$$\log_e 10 = 2.3026$$
$$\log_{10} e = 0.4343$$

n	$n \times 2.3026$
1	2.3026 = 0.6974–3
2	4.6052 = 0.3948–5
3	6.9078 = 0.0922–7
4	9.2103 = 0.7897–10
5	11.5129 = 0.4871–12
6	13.8155 = 0.1845–14
7	16.1181 = 0.8819–17
8	18.4207 = 0.5793–19
9	20.7233 = 0.2767–21

5. Natural, Napierian, or Hyperbolic Logarithms.*—(*Concluded*)

N	0	1	2	3	4	5	6	7	8	9
500	6.2146	6.2166	6.2186	6.2206	6.2226	6.2246	6.2265	6.2285	6.2305	6.2324
510	6.2344	6.2364	6.2383	6.2403	6.2422	6.2442	6.2461	6.2480	6.2500	6.2519
520	6.2538	6.2558	6.2577	6.2596	6.2615	6.2634	6.2653	6.2672	6.2691	6.2710
530	6.2729	6.2748	6.2766	6.2785	6.2804	6.2823	6.2841	6.2860	6.2879	6.2897
540	6.2916	6.2934	6.2953	6.2971	6.2989	6.3008	6.3026	6.3044	6.3063	6.3081
550	6.3099	6.3117	6.3135	6.3154	6.3172	6.3190	6.3208	6.3226	6.3244	6.3261
560	6.3279	6.3297	6.3315	6.3333	6.3351	6.3368	6.3386	6.3404	6.3421	6.3439
570	6.3456	6.3474	6.3491	6.3509	6.3256	6.3544	6.3561	6.3578	6.3596	6.3613
580	6.3630	6.3648	6.3665	6.3682	6.3699	6.3716	6.3733	6.3750	6.3767	6.3784
590	6.3801	6.3818	6.3835	6.3852	6.3869	6.3886	6.3902	6.3919	6.3936	6.3953
600	6.3969	6.3986	6.4003	6.4019	6.4036	6.4052	6.4069	6.4085	6.4102	6.4118
610	6.4135	6.4151	6.4167	6.4184	6.4200	6.4216	6.4232	6.4249	6.4265	6.4281
620	6.4297	6.4313	6.4329	6.4345	6.4362	6.4378	6.4394	6.4409	6.4425	6.4441
630	6.4457	6.4473	6.4489	6.4505	6.4520	6.4536	6.4552	6.4568	6.4583	6.4599
640	6.4615	6.4630	6.4646	6.4661	6.4677	6.4693	6.4708	6.4723	6.4739	6.4754
650	6.4770	6.4785	6.4800	6.4816	6.4831	6.4846	6.4862	6.4877	6.4892	6.4907
660	6.4922	6.4938	6.4953	6.4968	6.4983	6.4998	6.5013	6.5028	6.5043	6.5058
670	6.5073	6.5088	6.5103	6.5117	6.5132	6.5147	6.5162	6.5177	6.5191	6.5206
680	6.5221	6.5236	6.5250	6.5265	6.5280	6.5294	6.5309	6.5323	6.5338	6.5352
690	6.5367	6.5381	6.5396	6.5410	6.5425	6.5439	6.5453	6.5468	6.5482	6.5497
700	6.5511	6.5525	6.5539	6.5554	6.5568	6.5582	6.5596	6.5610	6.5624	6.5639
710	6.5653	6.5667	6.5681	6.5695	6.5709	6.5723	6.5737	6.5751	6.5765	6.5779
720	6.5793	6.5806	6.5820	6.5834	6.5848	6.5862	6.5876	6.5889	6.5903	6.5917
730	6.5930	6.5944	6.5958	6.5971	6.5985	6.5999	6.6012	6.6026	6.6039	6.6053
740	6.6067	6.6080	6.6093	6.6107	6.6120	6.6134	6.6147	6.6161	6.6174	6.6187
750	6.6201	6.6214	6.6227	6.6241	6.6254	6.6267	6.6280	6.6294	6.6307	6.6320
760	6.6333	6.6346	6.6359	6.6373	6.6386	6.6399	6.6412	6.6425	6.6438	6.6451
770	6.6464	6.6477	6.6490	6.6503	6.6516	6.6529	6.6542	6.6554	6.6567	6.6580
780	6.6593	6.6606	6.6619	6.6631	6.6644	6.6657	6.6670	6.6682	6.6695	6.6708
790	6.6720	6.6733	6.6746	6.6758	6.6771	6.6783	6.6796	6.6809	6.6821	6.6834
800	6.6846	6.6859	6.6871	6.6884	6.6896	6.6908	6.6921	6.6933	6.6946	6.6958
810	6.6970	6.6983	6.6995	6.7007	6.7020	6.7032	6.7044	6.7056	6.7069	6.7081
820	6.7093	6.7105	6.7117	6.7130	6.7142	6.7154	6.7166	6.7178	6.7190	6.7202
830	6.7214	6.7226	6.7238	6.7250	6.7262	6.7274	6.7286	6.7298	6.7310	6.7322
840	6.7334	6.7346	6.7358	6.7370	6.7382	6.7393	6.7405	6.7417	6.6429	6.7441
850	6.7452	6.7464	6.7476	6.7488	6.7499	6.7511	6.7523	6.7534	6.7546	6.7558
860	6.7569	6.7581	6.7593	6.7604	6.7616	6.7627	6.7639	6.7650	6.7662	6.7673
870	6.7685	6.7696	6.7708	6.7719	6.7731	6.7742	6.7754	6.7765	6.7776	6.7788
880	6.7799	6.7811	6.7822	6.7833	6.7845	6.7856	6.7867	5.7878	6.7890	6.7901
890	6.7912	6.7923	6.7935	6.7946	6.7957	6.7968	6.7979	6.7991	6.8002	6.8013
900	6.8024	6.8035	6.8046	6.8057	6.8068	6.8079	6.8090	6.8101	6.8112	6.8123
910	6.8134	6.8145	6.8156	6.8167	6.8178	6.8189	6.8200	6.8211	6.8222	6.8233
920	6.8244	6.8255	6.8265	6.8276	6.8287	6.8298	6.8309	6.8320	6.8330	6.8341
930	6.8352	6.8363	6.8373	6.8384	6.8395	6.8405	6.8416	6.8427	6.8437	6.8448
940	6.8459	6.8469	6.8480	6.8491	6.8501	6.8512	6.8522	6.8533	6.8544	6.8554
950	6.8565	6.8575	6.8586	6.8596	6.8607	6.8617	6.8628	6.8638	6.8648	6.8659
960	6.8669	6.8680	6.8690	6.8701	6.8711	6.8721	6.8732	6.8742	6.8752	6.8763
970	6.8773	6.8783	6.8794	6.8804	6.8814	6.8824	6.8835	6.8845	6.8855	6.8865
980	6.8876	6.8886	6.8896	6.8906	6.8916	6.8926	6.8937	6.8947	6.8957	6.8967
990	6.8977	6.8987	6.8997	6.9007	6.9017	6.9027	6.9037	6.9047	6.9057	6.9068

* From "Standard Handbook for Electrical Engineers," 7th ed.

6. Hyperbolic Sines* $\sinh x = (e^x - e^{-x})/2$

x	0	1	2	3	4	5	6	7	8	9	Av. diff.
0.0	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0701	0.0801	0.0901	100
0.1	0.1002	0.1102	0.1203	0.1304	0.1405	0.1506	0.1607	0.1708	0.1810	0.1911	101
0.2	0.2013	0.2115	0.2218	0.2320	0.2423	0.2526	0.2629	0.2733	0.2837	0.2941	103
0.3	0.3045	0.3150	0.3255	0.3360	0.3466	0.3572	0.3678	0.3785	0.3892	0.4000	106
0.4	0.4108	0.4216	0.4325	0.4434	0.4543	0.4653	0.4764	0.4875	0.4986	0.5098	110
0.5	0.5211	0.5324	0.5438	0.5552	0.5666	0.5782	0.5897	0.6014	0.6131	0.6248	116
0.6	0.6367	0.6485	0.6605	0.6725	0.6846	0.6967	0.7090	0.7213	0.7336	0.7461	122
0.7	0.7586	0.7712	0.7838	0.7966	0.8094	0.8223	0.8353	0.8484	0.8615	0.8748	130
0.8	0.8881	0.9015	0.9150	0.9286	0.9423	0.9561	0.9700	0.9840	0.9981	1.012	138
0.9	1.027	1.041	1.055	1.070	1.085	1.099	1.114	1.129	1.145	1.160	15
1.0	1.175	1.191	1.206	1.222	1.238	1.254	1.270	1.286	1.303	1.319	16
1.1	1.336	1.352	1.369	1.386	1.403	1.421	1.438	1.456	1.474	1.491	17
1.2	1.509	1.528	1.546	1.564	1.583	1.602	1.621	1.640	1.659	1.679	19
1.3	1.698	1.718	1.738	1.758	1.779	1.799	1.820	1.841	1.862	1.883	21
1.4	1.904	1.926	1.948	1.970	1.992	2.014	2.037	2.060	2.083	2.106	22
1.5	2.129	2.153	2.177	2.201	2.225	2.250	2.274	2.299	2.324	2.350	25
1.6	2.376	2.401	2.428	2.454	2.481	2.507	2.535	2.562	2.590	2.617	27
1.7	2.646	2.674	2.703	2.732	2.761	2.790	2.820	2.850	2.881	2.911	30
1.8	2.942	9.973	3.005	3.037	3.069	3.101	3.134	3.167	3.200	3.234	33
1.9	3.268	3.303	3.337	3.372	3.408	3.443	3.479	3.516	3.552	3.589	36
2.0	3.627	3.665	3.703	3.741	3.780	3.820	3.859	3.899	3.940	3.981	39
2.1	4.022	4.064	4.106	4.148	4.191	4.234	4.278	4.322	4.367	4.412	44
2.2	4.457	4.503	4.549	4.596	4.643	4.691	4.739	4.788	4.837	4.887	48
2.3	4.937	4.988	5.039	5.090	5.142	5.195	5.248	5.302	5.356	5.411	53
2.4	5.466	5.522	5.578	5.635	5.693	5.751	5.810	5.869	5.929	5.989	58
2.5	6.050	6.112	6.174	6.237	6.300	6.365	6.429	6.495	6.561	6.627	64
2.6	6.695	6.763	6.831	6.901	6.971	7.042	7.113	7.185	7.258	7.332	71
2.7	7.406	7.481	7.557	7.634	7.711	7.789	7.868	7.948	8.028	8.110	79
2.8	8.192	8.275	8.359	8.443	8.529	8.615	8.702	8.790	8.879	8.969	87
2.9	9.060	9.151	9.244	9.337	9.431	9.527	9.623	9.720	9.819	9.918	96
3.0	10.02	10.12	10.22	10.32	10.43	10.53	10.64	10.75	10.86	10.97	11
3.1	11.08	11.19	11.30	11.42	11.53	11.65	11.76	11.88	12.00	12.12	12
3.2	12.25	12.37	12.49	12.62	12.75	12.88	13.01	13.14	13.27	13.40	13
3.3	13.54	13.67	13.81	13.95	14.09	14.23	14.38	14.52	14.67	14.82	14
3.4	14.97	15.12	15.27	15.42	15.58	15.73	15.89	16.05	16.21	16.38	16
3.5	16.54	16.71	16.88	17.05	17.22	17.39	17.57	17.74	17.92	18.10	17
3.6	18.29	18.47	18.66	18.84	19.03	19.22	19.42	19.61	19.81	20.01	19
3.7	20.21	20.41	20.62	20.83	21.04	21.25	21.46	21.68	21.90	22.12	21
3.8	22.34	22.56	22.79	23.02	23.25	23.49	23.72	23.96	24.20	24.45	24
3.9	24.69	24.94	25.19	25.44	25.70	25.96	26.22	26.48	26.75	27.02	26
4.0	27.29	27.56	27.84	28.12	28.40	28.69	28.98	29.27	29.56	29.86	29
4.1	30.16	30.47	30.77	31.08	31.39	31.71	32.03	32.35	32.68	33.00	32
4.2	33.34	33.67	34.01	34.35	34.70	35.05	35.40	35.75	36.11	36.48	35
4.3	36.84	37.21	37.59	37.97	38.35	38.73	39.12	39.52	39.91	40.31	39
4.4	40.72	41.13	41.54	41.96	42.38	42.81	43.24	43.67	44.11	44.56	43
4.5	45.00	45.46	45.91	46.37	46.84	47.31	47.79	48.27	48.75	49.24	47
4.6	49.74	50.24	50.74	51.25	51.77	52.29	52.81	53.34	53.88	54.42	52
4.7	54.97	55.52	56.08	56.64	57.21	57.79	58.37	58.96	59.55	60.15	58
4.8	60.75	61.36	61.98	62.60	63.23	63.87	64.51	65.16	65.81	67.47	64
4.9	67.14	67.82	68.50	69.19	69.88	70.58	71.29	72.01	72.73	73.46	71
5.0	74.20										

If $x > 5$, $\sinh x = e^x/2$ correct to four significant figures.

* From Lionel S. Marks, "Mechanical Engineers' Handbook."

7. Hyperbolic Cosines* $\cosh x = (e^x + e^{-x})/2$

x	0	1	2	3	4	5	6	7	8	9	Av. diff.
0.0	1.000	1.000	1.000	1.000	1.001	1.001	1.002	1.002	1.003	1.004	1
0.1	1.005	1.006	1.007	1.008	1.010	1.011	1.013	1.014	1.016	1.018	2
0.2	1.020	1.022	1.024	1.027	1.029	1.031	1.034	1.037	1.039	1.042	3
0.3	1.045	1.048	1.052	1.055	1.058	1.062	1.066	1.069	1.073	1.077	4
0.4	1.081	1.085	1.090	1.094	1.098	1.103	1.108	1.112	1.117	1.122	5
0.5	1.128	1.133	1.138	1.144	1.149	1.155	1.161	1.167	1.173	1.179	6
0.6	1.185	1.192	1.198	1.205	1.212	1.219	1.226	1.233	1.240	1.248	7
0.7	1.255	1.263	1.271	1.278	1.287	1.295	1.303	1.311	1.320	1.329	8
0.8	1.337	1.346	1.355	1.365	1.374	1.384	1.393	1.403	1.413	1.423	10
0.9	1.433	1.443	1.454	1.465	1.475	1.486	1.497	1.509	1.520	1.531	11
1.0	1.543	1.555	1.567	1.579	1.591	1.604	1.616	1.629	1.642	1.655	13
1.1	1.669	1.682	1.696	1.709	1.723	1.737	1.752	1.766	1.781	1.796	14
1.2	1.811	1.826	1.841	1.857	1.872	1.888	1.905	1.921	1.937	1.954	16
1.3	1.971	1.988	2.005	2.023	2.040	2.058	2.076	2.095	2.113	2.132	18
1.4	2.151	2.170	2.189	2.209	2.229	2.249	2.269	2.290	2.310	2.331	20
1.5	2.352	2.374	2.395	2.417	2.439	2.462	2.484	2.507	2.530	2.554	23
1.6	2.577	2.601	2.625	2.650	2.675	2.700	2.725	2.750	2.776	2.802	25
1.7	2.828	2.855	2.882	2.909	2.936	2.964	2.992	3.021	3.049	3.078	28
1.8	3.107	3.137	3.167	3.197	3.228	3.259	3.290	3.321	3.353	3.385	31
1.9	3.418	3.451	3.484	3.517	3.551	3.585	3.620	3.655	3.690	3.726	34
2.0	3.762	3.799	3.835	3.873	3.910	3.948	3.987	4.026	4.065	4.104	38
2.1	4.144	4.185	4.226	4.267	4.309	4.351	4.393	4.436	4.480	4.524	42
2.2	4.568	4.613	4.658	4.704	4.750	4.797	4.844	4.891	4.939	4.988	47
2.3	5.037	5.087	5.137	5.188	5.239	5.290	5.343	5.395	5.449	5.503	52
2.4	5.557	5.612	5.667	5.723	5.780	5.837	5.895	5.954	6.013	6.072	58
2.5	6.132	6.193	6.255	6.317	6.379	6.443	6.507	6.571	6.636	6.702	64
2.6	6.769	6.836	6.904	6.973	7.042	7.112	7.183	7.255	7.327	7.400	70
2.7	7.473	7.548	7.623	7.699	7.776	7.853	7.932	8.011	8.091	8.171	78
2.8	8.253	8.335	8.418	8.502	8.587	8.673	8.759	8.847	8.935	9.024	86
2.9	9.115	9.206	9.298	9.391	9.484	9.579	9.675	9.772	9.869	9.968	95
3.0	10.07	10.17	10.27	10.37	10.48	10.58	10.69	10.79	10.90	11.01	11
3.1	11.12	11.23	11.35	11.46	11.57	11.69	11.81	11.92	12.04	12.16	12
3.2	12.29	12.41	12.53	12.66	12.79	12.91	13.04	13.17	13.31	13.44	13
3.3	13.57	13.71	13.85	13.99	14.13	14.27	14.41	14.56	14.70	14.85	14
3.4	15.00	15.15	15.30	15.45	15.61	15.77	15.92	16.08	16.25	16.41	16
3.5	16.57	16.74	16.91	17.08	17.25	17.42	17.60	17.77	17.95	18.13	17
3.6	18.31	18.50	18.68	18.87	19.06	19.25	19.44	19.64	19.84	20.03	19
3.7	20.24	20.44	20.64	20.85	21.06	21.27	21.49	21.70	21.92	22.14	21
3.8	22.36	22.59	22.81	23.04	23.27	23.51	23.74	23.98	24.22	24.47	23
3.9	24.71	24.96	25.21	25.46	25.72	25.98	26.24	26.50	26.77	27.04	26
4.0	27.31	27.58	27.86	28.14	28.42	28.71	29.00	29.29	29.58	29.88	29
4.1	30.18	30.48	30.79	31.10	31.41	31.72	32.04	32.37	32.69	33.02	32
4.2	33.35	33.69	34.02	34.37	34.71	35.06	35.41	35.77	36.13	36.49	35
4.3	36.86	37.23	37.60	37.98	38.36	38.75	39.13	39.53	39.93	40.33	39
4.4	40.73	41.14	41.55	41.97	42.39	42.82	43.25	43.68	44.12	44.57	43
4.5	45.01	45.47	45.92	46.38	46.85	47.32	47.80	48.28	48.76	49.25	47
4.6	49.75	50.25	50.75	51.26	51.78	52.30	52.82	57.35	57.89	54.43	52
4.7	54.98	55.53	56.09	56.65	57.22	57.80	58.38	58.96	59.56	60.15	58
4.8	60.76	61.37	61.99	62.61	63.24	63.87	64.52	65.16	65.82	66.48	64
4.9	67.15	67.82	68.50	69.19	69.89	70.59	71.30	72.02	72.74	73.47	71
5.0	74.21										

If $x > 5$, $\cosh x = e^x/2$ correct to four significant figures.

* From Lionel S. Marks, "Mechanical Engineers' Handbook."

8. Hyperbolic Tangents*

$\tanh x = (e^x - e^{-x})/(e^x + e^{-x}) = \sinh x/\cosh x$

x	0	1	2	3	4	5	6	7	8	9	Av. diff.
0.0	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0599	0.0699	0.0798	0.0898	100
0.1	0.0997	0.1096	0.1194	0.1293	0.1391	0.1489	0.1587	0.1684	0.1781	0.1878	98
0.2	0.1974	0.2070	0.2165	0.2260	0.2355	0.2449	0.2543	0.2636	0.2729	0.2821	94
0.3	0.2913	0.3004	0.3095	0.3185	0.3275	0.3364	0.3452	0.3540	0.3627	0.3714	89
0.4	0.3800	0.3885	0.3969	0.4053	0.4137	0.4219	0.4301	0.4382	0.4462	0.4542	82
0.5	0.4621	0.4700	0.4777	0.4854	0.4930	0.5005	0.5080	0.5154	0.5227	0.5299	75
0.6	0.5370	0.5441	0.5511	0.5581	0.5649	0.5717	0.5784	0.5850	0.5915	0.5980	67
0.7	0.6044	0.6107	0.6169	0.6231	0.6291	0.6352	0.6411	0.6469	0.6527	0.6584	60
0.8	0.6640	0.6696	0.6751	0.6805	0.6858	0.6911	0.6963	0.7014	0.7064	0.7114	52
0.9	0.7163	0.7211	0.7259	0.7306	0.7352	0.7398	0.7443	0.7487	0.7531	0.7574	45
1.0	0.7616	0.7658	0.7699	0.7739	0.7779	0.7818	0.7857	0.7895	0.7932	0.7969	39
1.1	0.8005	0.8041	0.8076	0.8110	0.8144	0.8178	0.8210	0.8243	0.8275	0.8306	33
1.2	0.8337	0.8367	0.8397	0.8426	0.8455	0.8483	0.8511	0.8538	0.8565	0.8591	28
1.3	0.8617	0.8643	0.8668	0.8693	0.8717	0.8741	0.8764	0.8787	0.8810	0.8832	24
1.4	0.8854	0.8875	0.8896	0.8917	0.8937	0.8957	0.8977	0.8996	0.9015	0.9033	20
1.5	0.9052	0.9069	0.9087	0.9104	0.9121	0.9138	0.9154	0.9170	0.9186	0.9202	17
1.6	0.9217	0.9232	0.9246	0.9261	0.9275	0.9289	0.9302	0.9316	0.9329	0.9342	14
1.7	0.9354	0.9367	0.9379	0.9391	0.9402	0.9414	0.9425	0.9436	0.9447	0.9458	11
1.8	0.9468	0.9478	0.9488	0.9498	0.9508	0.9518	0.9527	0.9536	0.9545	0.9554	9
1.9	0.9562	0.9571	0.9579	0.9587	0.9595	0.9603	0.9611	0.9619	0.9626	0.9633	8
2.0	0.9640	0.9647	0.9654	0.9661	0.9668	0.9674	0.9680	0.9687	0.9693	0.9699	6
2.1	0.9705	0.9710	0.9716	0.9722	0.9727	0.9732	0.9738	0.9743	0.9748	0.9753	5
2.2	0.9757	0.9762	0.9767	0.9771	0.9776	0.9780	0.9785	0.9789	0.9793	0.9797	4
2.3	0.9801	0.9805	0.9809	0.9812	0.9816	0.9820	0.9823	0.9827	0.9830	0.9834	4
2.4	0.9837	0.9840	0.9843	0.9846	0.9849	0.9852	0.9855	0.9858	0.9861	0.9863	3
2.5	0.9866	0.9869	0.9871	0.9874	0.9876	0.9879	0.9881	0.9884	0.9886	0.9888	2
2.6	0.9890	0.9892	0.9895	0.9897	0.9899	0.9901	0.9903	0.9905	0.9906	0.9908	2
2.7	0.9910	0.9912	0.9914	0.9915	0.9917	0.9919	0.9920	0.9922	0.9923	0.9925	2
2.8	0.9926	0.9928	0.9929	0.9931	0.9932	0.9933	0.9935	0.9936	0.9937	0.9938	1
2.9	0.9940	0.9941	0.9942	0.9943	0.9944	0.9945	0.9946	0.9947	0.9949	0.9950	1
3.0	0.9951	0.9959	0.9967	0.9973	0.9978	0.9982	0.9985	0.9988	0.9990	0.9992	4
4.0	0.9993	0.9995	0.9996	0.9996	0.9997	0.9998	0.9998	0.9998	0.9999	0.9999	1
5.0	0.9999	If $x > 5$, $\tanh x = 1.0000$ to four decimal places.									

* From Lionel S. Marks, "Mechanical Engineers' Handbook."

9. Degrees and Minutes Expressed in Radians*

Degrees						Hundredths				Minutes	
1°	0.0175	61°	1.0647	121°	2.1118	0°.01	0.0002	0°.51	0.0089	1′	0.0003
2	0.0349	62	1.0821	122	2.1293	0 .02	0.0003	0 .52	0.0091	2′	0.0006
3	0.0524	63	1.0996	123	2.1468	0 .03	0.0005	0 .53	0.0093	3′	0.0009
4	0.0698	64	1.1170	124	2.1642	0 .04	0.0007	0 .54	0.0094	4′	0.0012
5°	0.0873	65°	1.1345	125°	2.1817	0 .05	0.0009	0 .55	0.0096	5′	0.0015
6	0.1047	66	1.1519	126	2.1991	0 .06	0.0010	0 .56	0.0098	6′	0.0017
7	0.1222	67	1.1694	127	2.2166	0 .07	0.0012	0 .57	0.0099	7′	0.0020
8	0.1396	68	1.1868	128	2.2340	0 .08	0.0014	0 .58	0.0101	8′	0.0023
9	0.1571	69	1.2043	129	2.2515	0 .09	0.0016	0 .59	0.0103	9′	0.0026
10°	0.1745	70°	1.2217	130°	2.2689	0°.10	0.0017	0°.60	0.0105	10′	0.0029
11	0.1920	71	1.2392	131	2.2864	0 .11	0.0019	0 .61	0.0106	11′	0.0032
12	0.2094	72	1.2566	132	2.3038	0 .12	0.0021	0 .62	0.0108	12′	0.0035
13	0.2269	73	1.2741	133	2.3213	0 .13	0.0023	0 .63	0.0110	13′	0.0038
14	0.2443	74	1.2915	134	2.3387	0 .14	0.0024	0 .64	0.0112	14′	0.0041
15°	0.2618	75°	1.3090	135°	2.3562	0 .15	0.0026	0 .65	0.0113	15′	0.0044
16	0.2793	76	1.3265	136	2.3736	0 .16	0.0028	0 .66	0.0115	16′	0.0047
17	0.2967	77	1.3439	137	2.3911	0 .17	0.0030	0 .67	0.0117	17′	0.0049
18	0.3142	78	1.3614	138	2.4086	0 .18	0.0031	0 .68	0.0119	18′	0.0052
19	0.3316	79	1.3788	139	2.4260	0 .19	0.0033	0 .69	0.0120	19′	0.0055
20°	0.3491	80°	1.3963	140°	2.4435	0°.20	0.0035	0°.70	0.0122	20′	0.0058
21	0.3665	81	1.4137	141	2.4609	0 .21	0.0037	0 .71	0.0124	21′	0.0061
22	0.3840	82	1.4312	142	2.4784	0 .22	0.0038	0 .72	0.0126	22′	0.0064
23	0.4014	83	1.4486	143	2.4958	0 .23	0.0040	0 .73	0.0127	23′	0.0067
24	0.4189	84	1.4661	144	2.5133	0 .24	0.0042	0 .74	0.0129	24′	0.0070
25°	0.4363	85°	1.4835	145°	2.5307	0 .25	0.0044	0 .75	0.0131	25′	0.0073
26	0.4538	86	1.5010	146	2.5482	0 .26	0.0045	0 .76	0.0133	26′	0.0076
27	0.4712	87	1.5184	147	2.5656	0 .27	0.0047	0 .77	0.0134	27′	0.0079
28	0.4887	88	1.5359	148	2.5831	0 .28	0.0049	0 .78	0.0136	28′	0.0081
29	0.5061	89	1.5533	149	2.6005	0 .29	0.0051	0 .79	0.0138	29′	0.0084
30°	0.5236	90°	1.5708	150°	2.6180	0°.30	0.0052	0°.80	0.0140	30′	0.0087
31	0.5411	91	1.5882	151	2.6354	0 .31	0.0054	0 .81	0.0141	31′	0.0090
32	0.5585	92	1.6057	152	2.6529	0 .32	0.0056	0 .82	0.0143	32′	0.0093
33	0.5760	93	1.6232	153	2.6704	0 .33	0.0058	0 .83	0.0145	33′	0.0096
34	0.5934	94	1.6406	154	2.6878	0 .34	0.0059	0 .84	0.0147	34′	0.0099
35°	0.6109	95°	1.6581	155°	2.7053	0 .35	0.0061	0 .85	0.0148	35′	0.0102
36	0.6283	96	1.6755	156	2.7227	0 .36	0.0063	0 .86	0.0150	36′	0.0105
37	0.6458	97	1.6930	157	2.7402	0 .37	0.0065	0 .87	0.0152	37′	0.0108
38	0.6632	98	1.7104	158	2.7576	0 .38	0.0066	0 .88	0.0154	38′	0.0111
39	0.6807	99	1.7279	159	2.7751	0 .39	0.0068	0 .89	0.0155	39′	0.0113
40°	0.6981	100°	1.7453	160°	2.7925	0°.40	0.0070	0°.90	0.0157	40′	0.0116
41	0.7156	101	1.7628	161	2.8100	0 .41	0.0072	0 .91	0.0159	41′	0.0119
42	0.7330	102	1.7802	162	2.8274	0 .42	0.0073	0 .92	0.0161	42′	0.0122
43	0.7505	103	1.7977	163	2.8449	0 .43	0.0075	0 .93	0.0162	43′	0.0125
44	0.7679	104	1.8151	164	2.8623	0 .44	0.0077	0 .94	0.0164	44′	0.0128
45°	0.7854	105°	1.8326	165°	2.8798	0 .45	0.0079	0 .95	0.0166	45′	0.0131
46	0.8029	106	1.8500	166	2.8972	0 .46	0.0080	0 .96	0.0168	46′	0.0134
47	0.8203	107	1.8675	167	2.9147	0 .47	0.0082	0 .97	0.0169	47′	0.0137
48	0.8378	108	1.8850	168	2.9322	0 .48	0.0084	0 .98	0.0171	48′	0.0140
49	0.8552	109	1.9024	169	2.9496	0 .49	0.0086	0 .99	0.0173	49′	0.0143
50°	0.8727	110°	1.9199	170°	2.9671	0°.50	0.0087	1°.00	0.0175	50′	0.0145
51	0.8901	111	1.9373	171	2.9845	....	.....	....	.....	51′	0.0148
52	0.9076	112	1.9548	172	3.0020	....	.....	....	.....	52′	0.0151
53	0.9250	113	1.9722	173	3.0194	....	.....	....	.....	53′	0.0154
54	0.9425	114	1.9897	174	3.0369	....	.....	....	.....	54′	0.0157
55°	0.9599	115°	2.0071	175°	3.0543	....	.....	....	.....	55′	0.0160
56	0.9774	116	2.0246	176	3.0718	....	.....	....	.....	56′	0.0163
57	0.9948	117	2.0420	177	3.0892	....	.....	....	.....	57′	0.0166
58	1.0123	118	2.0595	178	3.1067	....	.....	....	.....	58′	0.0169
59	1.0297	119	2.0769	179	3.1241	....	.....	....	.....	59′	0.0172
60°	1.0472	120°	2.0944	180°	3.1416	....	.....	....	.....	60′	0.0175

Arc 1° = 0.01745. Arc 1′ = 0.0002909. Arc 1″ = 0.000004848.
1 radian = 57°.296 = 57° 17′.75 = 57° 17′ 44″.81.
* From Lionel S. Marks, "Mechanical Engineers' Handbook."

10. Radians Expressed in Degrees*

										Interpolation	
0.01	0°.57	0.64	36°.67	1.27	72°.77	1.90	108°.86	2.53	144°.96	0.0002	0°.01
0.02	1°.15	0.65	37°.24	1.28	73°.34	1.91	109°.43	2.54	145°.53	0.0004	0°.02
0.03	1°.72	0.66	37°.82	1.29	73°.91	1.92	110°.01	2.55	146°.10	0.0006	0°.03
0.04	2°.29	0.67	38°.39	1.30	74°.48	1.93	110°.58	2.56	146°.68	0.0008	0°.05
0.05	2°.86	0.68	38°.96	1.31	75°.06	1.94	111°.15	2.57	147°.25	0.0010	0°.06
0.06	3°.44	0.69	39°.53	1.32	75°.63	1.95	111°.73	2.58	147°.82	0.0012	0°.07
0.07	4°.01	0.70	40°.11	1.33	76°.20	1.96	112°.30	2.59	148°.40	0.0014	0°.08
0.08	4°.58	0.71	40°.68	1.34	76°.78	1.97	112°.87	2.60	148°.97	0.0016	0°.09
0.09	5°.16	0.72	41°.25	1.35	77°.35	1.98	113°.45	2.61	149°.54	0.0018	0°.10
0.10	5°.73	0.73	41°.83	1.36	77°.92	1.99	114°.02	2.62	150°.11	0.0020	0°.11
0.11	6°.30	0.74	42°.40	1.37	78°.50	2.00	114°.59	2.63	150°.69	0.0022	0°.13
0.12	6°.88	0.75	42°.97	1.38	79°.07	2.01	115°.16	2.64	151°.26	0.0024	0°.14
0.13	7°.45	0.76	43°.54	1.39	79°.64	2.02	115°.74	2.65	151°.83	0.0026	0°.15
0.14	8°.02	0.77	44°.12	1.40	80°.21	2.03	116°.31	2.66	152°.41	0.0028	0°.16
0.15	8°.59	0.78	44°.69	1.41	80°.79	2.04	116°.88	2.67	152°.98	0.0030	0°.17
0.16	9°.17	0.79	45°.26	1.42	81°.36	2.05	117°.46	2.68	153°.55	0.0032	0°.18
0.17	9°.74	0.80	45°.84	1.43	81°.93	2.06	118°.03	2.69	154°.13	0.0034	0°.19
0.18	10°.31	0.81	46°.41	1.44	82°.51	2.07	118°.60	2.70	154°.70	0.0036	0°.21
0.19	10°.89	0.82	46°.98	1.45	83°.08	2.08	119°.18	2.71	155°.27	0.0038	0°.22
0.20	11°.46	0.83	47°.56	1.46	83°.65	2.09	119°.75	2.72	155°.84	0.0040	0°.23
0.21	12°.03	0.84	48°.13	1.47	84°.22	2.10	120°.32	2.73	156°.42	0.0042	0°.24
0.22	12°.61	0.85	48°.70	1.48	84°.80	2.11	120°.89	2.74	156°.99	0.0044	0°.25
0.23	13°.18	0.86	49°.27	1.49	85°.37	2.12	121°.47	2.75	157°.56	0.0046	0°.26
0.24	13°.75	0.87	49°.85	1.50	85°.94	2.13	122°.04	2.76	158°.14	0.0048	0°.28
0.25	14°.32	0.88	50°.42	1.51	86°.52	2.14	122°.61	2.77	158°.71	0.0050	0°.29
0.26	14°.90	0.89	50°.99	1.52	87°.09	2.15	123°.19	2.78	159°.28	0.0052	0°.30
0.27	15°.47	0.90	51°.27	1.53	87°.66	2.16	123°.76	2.79	159°.86	0.0054	0°.31
0.28	16°.04	0.91	52°.14	1.54	88°.24	2.17	124°.33	2.80	160°.43	0.0056	0°.32
0.29	16°.62	0.92	52°.71	1.55	88°.81	2.18	124°.90	2.81	161°.00	0.0058	0°.33
0.30	17°.19	0.93	53°.29	1.56	89°.38	2.19	125°.48	2.82	161°.57	0.0060	0°.34
0.31	17°.76	0.94	53°.86	1.57	89°.95	2.20	126°.05	2.83	162°.15	0.0062	0°.36
0.32	18°.33	0.95	54°.43	1.58	90°.53	2.21	126°.62	2.84	162°.72	0.0064	0°.37
0.33	18°.91	0.96	55°.00	1.59	91°.10	2.22	127°.20	2.85	163°.29	0.0066	0°.38
0.34	19°.48	0.97	55°.58	1.60	91°.67	2.23	127°.77	2.86	163°.87	0.0068	0°.39
0.35	20°.05	0.98	56°.15	1.61	92°.25	2.24	128°.34	2.87	164°.44	0.0070	0°.40
0.36	20°.63	0.99	56°.72	1.62	92°.82	2.25	128°.92	2.88	165°.01	0.0072	0°.41
0.37	21°.20	1.00	57°.30	1.63	93°.39	2.26	129°.49	2.89	165°.58	0.0074	0°.42
0.38	21°.77	1.01	57°.87	1.64	93°.97	2.27	130°.06	2.90	166°.16	0.0076	0°.44
0.39	22°.35	1.02	58°.44	1.65	94°.54	2.28	130°.63	2.91	166°.73	0.0078	0°.45
0.40	22°.92	1.03	59°.01	1.66	95°.11	2.29	131°.21	2.92	167°.30	0.0080	0°.46
0.41	23°.49	1.04	59°.59	1.67	95°.68	2.30	131°.78	2.93	167°.88	0.0082	0°.47
0.42	24°.06	1.05	60°.16	1.68	96°.26	2.31	132°.35	2.94	168°.45	0.0084	0°.48
0.43	24°.64	1.06	60°.73	1.69	96°.83	2.32	132°.93	2.95	169°.02	0.0086	0°.49
0.44	25°.21	1.07	61°.31	1.70	97°.40	2.33	133°.50	2.96	169°.60	0.0088	0°.50
0.45	25°.78	1.08	61°.88	1.71	97°.98	2.34	134°.07	2.97	170°.17	0.0090	0°.52
0.46	26°.36	1.09	62°.45	1.72	98°.55	2.35	134°.65	2.98	170°.74	0.0092	0°.53
0.47	26°.93	1.10	63°.03	1.73	99°.12	2.36	135°.22	2.99	171°.31	0.0094	0°.54
0.48	27°.50	1.11	63°.60	1.74	99°.69	2.37	135°.79	3.00	171°.89	0.0096	0°.55
0.49	28°.07	1.12	64°.17	1.75	100°.27	2.38	136°.36	3.01	172°.46	0.0098	0°.56
0.50	28°.65	1.13	64°.74	1.76	100°.84	2.39	136°.94	3.02	173°.03		
0.51	29°.22	1.14	65°.32	1.77	101°.41	2.40	137°.51	3.03	173°.61		
0.52	29°.79	1.15	65°.89	1.78	101°.99	2.41	138°.08	3.04	174°.18		
0.53	30°.37	1.16	66°.46	1.79	102°.56	2.42	138°.66	3.05	174°.75		
0.54	30°.94	1.17	67°.04	1.80	103°.13	2.43	139°.23	3.06	175° 33		
0.55	31°.51	1.18	67°.61	1.81	103°.71	2.44	139°.80	3.07	175°.90		
0.56	32°.09	1.19	68°.18	1.82	104°.28	2.45	140°.37	3.08	176°.47		
0.57	32°.66	1.20	68°.75	1.83	104°.85	2.46	140°.95	3.09	177°.04		
0.58	33°.23	1.21	69°.33	1.84	105°.42	2.47	141°.52	3.10	177°.62		
0.59	33°.80	1.22	69°.90	1.85	106°.00	2.48	142°.09	3.11	178°.19		
0.60	34°.38	1.23	70°.47	1.86	106°.57	2.49	142°.67	3.12	178°.76		
0.61	34°.95	1.24	71°.05	1.87	107°.14	2.50	143°.24	3.13	179°.34		
0.62	35°.52	1.25	71°.62	1.88	107°.72	2.51	143°.81	3.14	179°.91		
0.63	36°.10	1.26	72°.19	1.89	108°.29	2.52	144°.39	3.15	180°.48		

Multiples of π

1	3.1416	180°
2	6.2832	360°
3	9.4248	540°
4	12.5664	720°
5	15.7080	900°
6	18.8496	1080°
7	21.9911	1260°
8	25.1327	1440°
9	28.2743	1620°
10	31.4159	1800°

* From Lionel S. Marks, "Mechanical Engineers' Handbook."

11. Exponentials*

(e^n and e^{-n})

n	e^n	Diff.	n	e^n	Diff.	n	e^n	n	e^{-n}	Diff.	n	e^{-n}	n	e^{-n}
0.00	1.000	10	0.50	1.649	16	1.0	2.718†	0.00	1.000	−10	0.50	0.607	1.0	0.368
0.01	1.010	10	0.51	1.665	17	1.1	3.004	0.01	0.990	−10	0.51	0.600	1.1	0.333
0.02	1.020	10	0.52	1.682	17	1.2	3.320	0.02	0.980	−10	0.52	0.595	1.2	0.301
0.03	1.030	11	0.53	1.699	17	1.3	3.669	0.03	0.970	− 9	0.53	0.589	1.3	0.273
0.04	1.041	10	0.54	1.716	17	1.4	4.055	0.04	0.961	−10	0.54	0.583	1.4	0.247
0.05	1.051	11	0.55	1.733	18	1.5	4.482	0.05	0.951	− 9	0.55	0.577	1.5	0.223
0.06	1.062	11	0.56	1.751	17	1.6	4.953	0.06	0.942	−10	0.56	0.571	1.6	0.202
0.07	1.073	10	0.57	1.768	18	1.7	5.474	0.07	0.932	− 9	0.57	0.566	1.7	0.183
0.08	1.083	11	0.58	1.786	18	1.8	6.050	0.08	0.923	− 9	0.58	0.560	1.8	0.165
0.09	1.094	11	0.59	1.804	18	1.9	6.686	0.09	0.914	− 9	0.59	0.554	1.9	0.150
0.10	1.105	11	0.60	1.822	18	2.0	7.389	0.10	0.905	− 9	0.60	0.549	2.0	0.135
0.11	1.116	11	0.61	1.840	19	2.1	8.166	0.11	0.896	− 9	0.61	0.543	2.1	0.122
0.12	1.127	12	0.62	1.859	19	2.2	9.025	0.12	0.887	− 9	0.62	0.538	2.2	0.111
0.13	1.139	11	0.63	1.878	18	2.3	9.974	0.13	0.878	− 9	0.63	0.533	2.3	0.100
0.14	1.150	12	0.64	1.896	20	2.4	11.02	0.14	0.869	− 8	0.64	0.527	2.4	0.0907
0.15	1.162	12	0.65	1.916	19	2.5	12.18	0.15	0.861	− 9	0.65	0.522	2.5	0.0821
0.16	1.174	11	0.66	1.935	19	2.6	13.46	0.16	0.852	− 8	0.66	0.517	2.6	0.0743
0.17	1.185	12	0.67	1.954	20	2.7	14.88	0.17	0.844	− 9	0.67	0.512	2.7	0.0672
0.18	1.197	12	0.68	1.974	20	2.8	16.44	0.18	0.835	− 8	0.68	0.507	2.8	0.0608
0.19	1.209	12	0.69	1.994	20	2.9	18.17	0.19	0.827	− 8	0.69	0.502	2.9	0.0550
0.20	1.221	13	0.70	2.014	20	3.0	20.09	0.20	0.819	− 8	0.70	0.497	3.0	0.0498
0.21	1.234	12	0.71	2.034	20	3.1	22.20	0.21	0.811	− 8	0.71	0.492	3.1	0.0450
0.22	1.246	13	0.72	2.054	21	3.2	24.53	0.22	0.803	− 8	0.72	0.487	3.2	0.0408
0.23	1.259	12	0.73	2.075	21	3.3	27.11	0.23	0.795	− 8	0.73	0.482	3.3	0.0369
0.24	1.271	13	0.74	2.096	21	3.4	29.96	0.24	0.787	− 8	0.74	0.477	3.4	0.0334
0.25	1.284	13	0.75	2.117	21	3.5	33.12	0.25	0.779	− 8	0.75	0.472	3.5	0.0302
0.26	1.297	13	0.76	2.138	22	3.6	36.60	0.26	0.771	− 8	0.76	0.468	3.6	0.0273
0.27	1.310	13	0.77	2.160	21	3.7	40.45	0.27	0.763	− 7	0.77	0.463	3.7	0.0247
0.28	1.323	13	0.78	2.181	22	3.8	44.70	0.28	0.756	− 8	0.78	0.458	3.8	0.0224
0.29	1.336	14	0.79	2.203	23	3.9	49.40	0.29	0.748	− 7	0.79	0.454	3.9	0.0202
0.30	1.350	13	0.80	2.226	22	4.0	54.60	0.30	0.741	− 8	0.80	0.449	4.0	0.0183
0.31	1.363	14	0.81	2.248	22	4.1	60.34	0.31	0.733	− 7	0.81	0.445	4.1	0.0166
0.32	1.377	14	0.82	2.270	23	4.2	66.69	0.32	0.726	− 7	0.82	0.440	4.2	0.0150
0.33	1.391	14	0.83	2.293	23	4.3	73.70	0.33	0.719	− 7	0.83	0.436	4.3	0.0136
0.34	1.405	14	0.84	2.316	24	4.4	81.45	0.34	0.712	− 7	0.84	0.432	4.4	0.0123
0.35	1.419	14	0.85	2.340	23	4.5	90.02	0.35	0.705	− 7	0.85	0.427	4.5	0.0111
0.36	1.433	15	0.86	2.363	24			0.36	0.698	− 7	0.86	0.423		
0.37	1.448	14	0.87	2.387	24	5.0	148.4	0.37	0.691	− 7	0.87	0.419	5.0	0.00674
0.38	1.462	15	0.88	2.411	24	6.0	403.4	0.38	0.684	− 7	0.88	0.415	6.0	0.00248
0.39	1.477	15	0.89	2.435	25	7.0	1097.	0.39	0.677	− 7	0.89	0.411	7.0	0.000912
0.40	1.492	15	0.90	2.460	24	8.0	2981.	0.40	0.670	− 6	0.90	0.407	8.0	0.000335
0.41	1.507	15	0.91	2.484	25	9.0	8103.	0.41	0.664	− 7	0.91	0.403	9.0	0.000123
0.42	1.522	15	0.92	2.509	26	10.0	22026.	0.42	0.657	− 6	0.92	0.399	10.0	0.000045
0.43	1.537	16	0.93	2.535	25	$\pi/2$	4.810	0.43	0.651	− 7	0.93	0.395		
0.44	1.553	15	0.94	2.560	26	$2\pi/2$	23.14	0.44	0.644	− 6	0.94	0.391		
0.45	1.568	16	0.95	2.586	26	$3\pi/2$	111.3	0.45	0.638	− 7	0.95	0.387	$\pi/2$	0.208
0.46	1.584	16	0.96	2.612	26	$4\pi/2$	535.5	0.46	0.631	− 6	0.96	0.383	$2\pi/2$	0.0432
0.47	1.600	16	0.97	2.638	26	$5\pi/2$	2576	0.47	0.625	− 6	0.97	0.379	$3\pi/2$	0.00898
0.48	1.616	16	0.98	2.664	27	$6\pi/2$	12392	0.48	0.619	− 6	0.98	0.375	$4\pi/2$	0.00187
0.49	1.632	17	0.99	2.691	27	$7\pi/2$	59610	0.49	0.613	− 6	0.99	0.372	$5\pi/2$	0.000388
0.50	1.649		1.00	2.718		$8\pi/2$	286751	0.50	0.607		1.00	0.368	$6\pi/2$	0.000081

$e = 2.71828$. $1/e = 0.367879$. $\log_{10} e = 0.4343$. $1/(0.4343) = 2.3026$.

* From Lionel S. Marks, "Mechanical Engineers' Handbook."

† NOTE: Do not interpolate in this column.

12. Sine Integral Si(x)*

$$Si(x) = \int_0^x \frac{\sin u}{u}\, du$$

x	$Si(x)$	x	$Si(x)$	x	$Si(x)$	x	$Si(x)$	x	$Si(x)$	x	$Si(x)$
0.0	0.00000	5.0	1.54993	10.0	1.65835	15.0	1.61819	20.0	1.54824	25.0	1.53148
0.1	0.09994	5.1	1.53125	10.1	1.65253	15.1	1.62226	20.1	1.55289	50.0	1.55162
0.2	0.19956	5.2	1.51367	10.2	1.64600	15.2	1.62575	20.2	1.55767		
0.3	0.29850	5.3	1.49732	10.3	1.63883	15.3	1.62865	20.3	1.56253		
0.4	0.39646	5.4	1.48230	10.4	1.63112	15.4	1.63093	20.4	1.56743		
0.5	0.49311	5.5	1.46872	10.5	1.62294	15.5	1.63258	20.5	1.57232		
0.6	0.58813	5.6	1.45667	10.6	1.61439	15.6	1.63359	20.6	1.57714		
0.7	0.68122	5.7	1.44620	10.7	1.60556	15.7	1.63396	20.7	1.58186		
0.8	0.77210	5.8	1.43736	10.8	1.59654	15.8	1.63370	20.8	1.58641		
0.9	0.86047	5.9	1.43018	10.9	1.58743	15.9	1.63280	20.9	1.59077		
1.0	0.94608	6.0	1.42469	11.0	1.57831	16.0	1.63130	21.0	1.59489		
1.1	1.02869	6.1	1.42087	11.1	1.56927	16.1	1.62921	21.1	1.59873		
1.2	1.10805	6.2	1.41871	11.2	1.56042	16.2	1.62657	21.2	1.60225		
1.3	1.18396	6.3	1.41817	11.3	1.55182	16.3	1.62339	21.3	1.60543		
1.4	1.25623	6.4	1.41922	11.4	1.54356	16.4	1.61973	21.4	1.60823		
1.5	1.32468	6.5	1.42179	11.5	1.53571	16.5	1.61563	21.5	1.61063		
1.6	1.38918	6.6	1.42582	11.6	1.52835	16.6	1.61112	21.6	1.61261		
1.7	1.44959	6.7	1.43121	11.7	1.52155	16.7	1.60627	21.7	1.61415		
1.8	1.50582	6.8	1.43787	11.8	1.51535	16.8	1.60111	21.8	1.61525		
1.9	1.55778	6.9	1.44570	11.9	1.50981	16.9	1.59572	21.9	1.61590		
2.0	1.60541	7.0	1.45460	12.0	1.50497	17.0	1.59014	22.0	1.61608		
2.1	1.64870	7.1	1.46443	12.1	1.50088	17.1	1.58443	22.1	1.61582		
2.2	1.68763	7.2	1.47509	12.2	1.49755	17.2	1.57865	22.2	1.61510		
2.3	1.72221	7.3	1.48644	12.3	1.49501	17.3	1.57285	22.3	1.61395		
2.4	1.75249	7.4	1.49834	12.4	1.49327	17.4	1.56711	22.4	1.61238		
2.5	1.77852	7.5	1.51068	12.5	1.49234	17.5	1.56146	22.5	1.61041		
2.6	1.80039	7.6	1.52331	12.6	1.49221	17.6	1.55598	22.6	1.60806		
2.7	1.81821	7.7	1.53611	12.7	1.49287	17.7	1.55070	22.7	1.60536		
2.8	1.83210	7.8	1.54894	12.8	1.49430	17.8	1.54568	22.8	1.60234		
2.9	1.84219	7.9	1.56167	12.9	1.49647	17.9	1.54097	22.9	1.59902		
3.0	1.84865	8.0	1.57419	13.0	1.49936	18.0	1.53661	23.0	1.59546		
3.1	1.85166	8.1	1.58637	13.1	1.50292	18.1	1.53264	23.1	1.59168		
3.2	1.85140	8.2	1.59810	13.2	1.50711	18.2	1.52909	23.2	1.58772		
3.3	1.84808	8.3	1.60928	13.3	1.51188	18.3	1.52600	23.3	1.58363		
3.4	1.84191	8.4	1.61981	13.4	1.51716	18.4	1.52339	23.4	1.57945		
3.5	1.83313	8.5	1.62960	13.5	1.52291	18.5	1.52128	23.5	1.57521		
3.6	1.82195	8.6	1.63857	13.6	1.52905	18.6	1.51969	23.6	1.57097		
3.7	1.80862	8.7	1.64665	13.7	1.53352	18.7	1.51863	23.7	1.56676		
3.8	1.79333	8.8	1.65379	13.8	1.54225	18.8	1.51810	23.8	1.56262		
3.9	1.77650	8.9	1.65993	13.9	1.54917	18.9	1.51810	23.9	1.55860		
4.0	1.75820	9.0	1.66504	14.0	1.55621	19.0	1.51863	24.0	1.55474		
4.1	1.73874	9.1	1.66908	14.1	1.56330	19.1	1.51967	24.1	1.55107		
4.2	1.71837	9.2	1.67205	14.2	1.57036	19.2	1.52122	24.2	1.54762		
4.3	1.69732	9.3	1.67393	14.3	1.57733	19.3	1.52324	24.3	1.54444		
4.4	1.67583	9.4	1.67473	14.4	1.58414	19.4	1.52572	24.4	1.54154		
4.5	1.65414	9.5	1.67446	14.5	1.59072	19.5	1.52863	24.5	1.53897		
4.6	1.63246	9.6	1.67316	14.6	1.59702	19.6	1.53192	24.6	1.53672		
4.7	1.61101	9.7	1.67084	14.7	1.60296	19.7	1.53357	24.7	1.53484		
4.8	1.58998	9.8	1.66757	14.8	1.60851	19.8	1.53954	24.8	1.53333		
4.9	1.56956	9.9	1.66338	14.9	1.61360	19.9	1.54378	24.9	1.53221		

* From P. O. Pedersen, "Radiation from a Vertical Antenna over Flat Perfectly Conducting Earth," G. E. C. Gad, Copenhagen.

13. $S_1(x)$ and Cosine Integral $Ci(x)$*

$$S_1(x) = \log_e x + 0.5772 - Ci(x) = \int_0^x \frac{1 - \cos x}{x}\, dx$$

$$Ci(x) = -\int_x^\infty \frac{\cos u}{u}\, du = \log_e x + 0.5772 - Si(x)$$

x	$S_1(x)$	x	$S_1(x)$	x	$S_1(x)$	x	$S_1(x)$	x	$S_1(x)$	x	$S_1(x)$
0.0	0.00000	5.0	2.37669	10.0	2.92527	15.0	3.23899	20.0	3.52853	25.0	3.80295
0.1	0.00249	5.1	2.38994	10.1	2.94327	15.1	3.25090	20.1	3.53173	50.0	4.49486
0.2	0.00998	5.2	2.40113	10.2	2.96050	15.2	3.26308	20.2	3.53535		
0.3	0.02241	5.3	2.41044	10.3	2.97688	15.3	3.27552	20.3	3.53946		
0.4	0.03973	5.4	2.41801	10.4	2.99234	15.4	3.28814	20.4	3.54402		
0.5	0.06185	5.5	2.42402	10.5	3.00688	15.5	3.30087	20.5	3.54905		
0.6	0.08866	5.6	2.42866	10.6	3.02045	15.6	3.31363	20.6	3.55456		
0.7	0.12002	5.7	2.43210	10.7	3.03300	15.7	3.32641	20.7	3.56049		
0.8	0.15579	5.8	2.43452	10.8	3.04457	15.8	3.33911	20.8	3.56687		
0.9	0.19578	5.9	2.43610	10.9	3.05514	15.9	3.35167	20.9	3.57368		
1.0	0.23981	6.0	2.43704	11.0	3.06467	16.0	3.36401	21.0	3.58085		
1.1	0.28766	6.1	2.43749	11.1	3.07323	16.1	3.37612	21.1	3.58840		
1.2	0.33908	6.2	2.43764	11.2	3.08083	16.2	3.38790	21.2	3.59629		
1.3	0.39384	6.3	2.43766	11.3	3.08749	16.3	3.39932	21.3	3.60446		
1.4	0.45168	6.4	2.43770	11.4	3.09322	16.4	3.41032	21.4	3.61288		
1.5	0.51233	6.5	2.43792	11.5	3.09814	16.5	3.42088	21.5	3.62155		
1.6	0.57549	6.6	2.43847	11.6	3.10225	16.6	3.43096	21.6	3.63037		
1.7	0.64088	6.7	2.43947	11.7	3.10561	16.7	3.44050	21.7	3.63935		
1.8	0.70820	6.8	2.44106	11.8	3.10828	16.8	3.44947	21.8	3.64842		
1.9	0.77713	6.9	2.44335	11.9	3.11038	16.9	3.45788	21.9	3.65751		
2.0	0.84739	7.0	2.44643	12.0	3.11190	17.0	3.46568	22.0	3.66662		
2.1	0.91865	7.1	2.45040	12.1	3.11301	17.1	3.47288	22.1	3.67568		
2.2	0.99060	7.2	2.45534	12.2	3.11370	17.2	3.47945	22.2	3.68465		
2.3	1.06295	7.3	2.46130	12.3	3.11412	17.3	3.48543	22.3	3.69348		
2.4	1.13540	7.4	2.46834	12.4	3.11429	17.4	3.49077	22.4	3.70216		
2.5	1.20764	7.5	2.47649	12.5	3.11436	17.5	3.49553	22.5	3.71059		
2.6	1.27939	7.6	2.48577	12.6	3.11437	17.6	3.49969	22.6	3.71879		
2.7	1.35038	7.7	2.49619	12.7	3.11438	17.7	3.50330	22.7	3.72670		
2.8	1.42035	7.8	2.50775	12.8	3.11453	17.8	3.50639	22.8	3.73427		
2.9	1.48903	7.9	2.52044	12.9	3.11484	17.9	3.50895	22.9	3.74153		
3.0	1.55620	8.0	2.53423	13.0	3.11540	18.0	3.51107	23.0	3.74838		
3.1	1.62163	8.1	2.54906	13.1	3.11628	18.1	3.51276	23.1	3.75483		
3.2	1.68511	8.2	2.56491	13.2	3.11754	18.2	3.51404	23.2	3.76089		
3.3	1.74646	8.3	2.56171	13.3	3.11924	18.3	3.51500	23.3	3.76651		
3.4	1.80552	8.4	2.59938	13.4	3.12142	18.4	3.51568	23.4	3.77170		
3.5	1.86211	8.5	2.61786	13.5	3.12414	18.5	3.51610	23.5	3.77644		
3.6	1.91613	8.6	2.63704	13.6	3.12745	18.6	3.51633	23.6	3.78072		
3.7	1.96745	8.7	2.65686	13.7	3.13134	18.7	3.51645	23.7	3.78459		
3.8	2.01600	8.8	2.67721	13.8	3.13587	18.8	3.51648	23.8	3.78801		
3.9	2.06170	8.9	2.69799	13.9	3.14104	18.9	3.51648	23.9	3.79101		
4.0	2.10449	9.0	2.71909	14.0	3.14688	19.0	3.51660	24.0	3.79360		
4.1	2.14438	9.1	2.74042	14.1	3.15338	19.1	3.51661	24.1	3.79582		
4.2	2.18131	9.2	2.76186	14.2	3.16054	19.2	3.51685	24.2	3.79767		
4.3	2.21535	9.3	2.78332	14.3	3.16835	19.3	3.51727	24.3	3.79917		
4.4	2.24648	9.4	2.80468	14.4	3.17677	19.4	3.51790	24.4	3.80036		
4.5	2.27479	9.5	2.82583	14.5	3.18583	19.5	3.51879	24.5	3.80129		
4.6	2.30033	9.6	2.84669	14.6	3.19545	19.6	3.52002	24.6	3.80197		
4.7	2.32317	9.7	2.86713	14.7	3.20564	19.7	3.52156	24.7	3.80243		
4.8	2.34344	9.8	2.88712	14.8	3.21630	19.8	3.52348	24.8	3.80271		
4.9	2.36124	9.9	2.90651	14.9	3.22746	19.9	3.52578	24.9	3.80288		

* From P. O. Pedersen, "Radiation from a Vertical Antenna over Flat Perfectly Conducting Earth." G. E. C. Gad, Copenhagen.

MATHEMATICAL RELATIONS

14. Trigonometric Relations

$$\sin x = \frac{A}{B}$$

$$\cos x = \frac{C}{B}$$

$$\tan x = \frac{A}{C} = \frac{\sin x}{\cos x}$$

$$\sec x = \frac{B}{C} = \frac{1}{\cos x}$$

$$\operatorname{cosec} x = \frac{B}{A} = \frac{1}{\sin x}$$

$$\cot x = \frac{C}{A} = \frac{1}{\tan x} = \frac{\cos x}{\sin x}$$

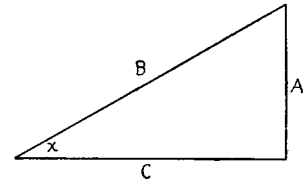


FIG. 1.

$$\sin(-x) = -\sin x$$
$$\cos(-x) = \cos x$$
$$\tan(-x) = -\tan x$$

$$\sin\left(\frac{\pi}{2} - x\right) = \cos x \qquad \sin\left(\frac{\pi}{2} + x\right) = \cos x$$

$$\cos\left(\frac{\pi}{2} - x\right) = \sin x \qquad \cos\left(\frac{\pi}{2} + x\right) = -\sin x$$

$$\tan\left(\frac{\pi}{2} - x\right) = \cot x \qquad \tan\left(\frac{\pi}{2} + x\right) = -\cot x$$

$$\sin(\pi - x) = \sin x \qquad \sin(\pi + x) = -\sin x$$
$$\cos(\pi - x) = -\cos x \qquad \cos(\pi + x) = -\cos x$$
$$\tan(\pi - x) = -\tan x \qquad \tan(\pi + x) = \tan x$$

$$\left.\begin{aligned}\sin(x + 2\pi n) &= \sin x\\ \cos(x + 2\pi n) &= \cos x\\ \tan(x + 2\pi n) &= \tan x\end{aligned}\right\} (n \text{ a positive or negative integer})$$

$$\sin(x + y) = \sin x \cos y + \cos x \sin y$$
$$\sin(x - y) = \sin x \cos y - \cos x \sin y$$
$$\cos(x + y) = \cos x \cos y - \sin x \sin y$$
$$\cos(x - y) = \cos x \cos y + \sin x \sin y$$

$$\tan(x + y) = \frac{\tan x + \tan y}{1 - \tan x \tan y}$$

$$\tan(x - y) = \frac{\tan x - \tan y}{1 + \tan x \tan y}$$

$$\sin 2x = 2 \sin x \cos x \qquad \cos 2x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$$

$$\tan 2x = \frac{2 \tan x}{1 - \tan^2 x} \qquad \cot 2x = \frac{(\cot^2 x - 1)}{2 \cot x}$$

$$\sin 3x = 3 \sin x - 4 \sin^3 x \qquad \cos 3x = 4\cos^3 x - 3 \cos x$$

$$\sin\frac{1}{2}x = \left[\frac{1}{2}(1 - \cos x)\right]^{\frac{1}{2}} = \frac{1}{2}(1 + \sin x)^{\frac{1}{2}} - \frac{1}{2}(1 - \sin x)^{\frac{1}{2}}$$

$$\cos\frac{1}{2}x = \left[\frac{1}{2}(1 + \cos x)\right]^{\frac{1}{2}} = \frac{1}{2}(1 + \sin x)^{\frac{1}{2}} + \frac{1}{2}(1 - \sin x)^{\frac{1}{2}}$$

$$\tan\frac{1}{2}x = \frac{(1 - \cos x)^{\frac{1}{2}}}{(1 + \cos x)^{\frac{1}{2}}} = \frac{1 - \cos x}{\sin x} = \frac{\sin x}{1 + \cos x}$$

$$\sin^2 x = \frac{1}{2}(1 - \cos 2x) \qquad \cos^2 x = \frac{1}{2}(1 + \cos 2x)$$

$$\sin^3 x = \frac{1}{4}(3 \sin x - \sin 3x) \qquad \cos^3 x = \frac{1}{4}(\cos 3x + 3 \cos x)$$

$$\sin x \sin y = \frac{1}{2}\cos(x - y) - \frac{1}{2}\cos(x + y)$$

$$\cos x \cos y = \frac{1}{2}\cos(x - y) + \frac{1}{2}\cos(x + y)$$

$$\sin x \cos y = \frac{1}{2}\sin(x - y) + \frac{1}{2}\sin(x + y)$$

$$\sin x + \sin y = 2 \sin \frac{1}{2}(x + y) \cos \frac{1}{2}(x - y)$$

$$\sin x - \sin y = 2 \cos \frac{1}{2}(x + y) \sin \frac{1}{2}(x - y)$$

$$\cos x + \cos y = 2 \cos \frac{1}{2}(x + y) \cos \frac{1}{2}(x - y)$$

$$\cos x - \cos y = -2 \sin \frac{1}{2}(x + y) \sin \frac{1}{2}(x - y)$$

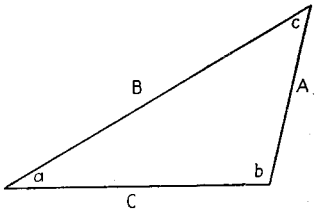


FIG. 2.

$$\tan x + \tan y = \frac{\sin(x + y)}{\cos x \cos y} \qquad \tan x - \tan y = \frac{\sin(x - y)}{\cos x \cos y}$$

$$\sin^2 x - \sin^2 y = \sin(x + y) \sin(x - y)$$
$$\cos^2 x - \cos^2 y = -\sin(x + y) \sin(x - y)$$
$$\cos^2 x - \sin^2 y = \cos(x + y) \cos(x - y)$$

$$\sin^2 x + \cos^2 x = 1 \qquad \sec^2 x - \tan^2 x = 1$$

In any triangle (Fig. 2):

$$\frac{A}{\sin a} = \frac{B}{\sin b} = \frac{C}{\sin c} \qquad \text{(law of sines)}$$

$$A^2 = B^2 + C^2 - 2BC \cos a \qquad \text{(law of cosines)}$$

$$\frac{A + B}{A - B} = \frac{\tan \frac{1}{2}(a + b)}{\tan \frac{1}{2}(a - b)} = \frac{\sin a + \sin b}{\sin a - \sin b} \qquad \text{(law of tangents)}$$

$$a + b + c = 180^\circ \qquad A = B \cos c + C \cos b$$
$$B = C \cos a + A \cos c \qquad C = A \cos b + B \cos a$$

15. Properties of Hyperbolic Functions

$$\sinh x = \frac{\epsilon^x - \epsilon^{-x}}{2} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

$$\cosh x = \frac{\epsilon^x + \epsilon^{-x}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

$$\tanh x = \frac{\sinh x}{\cosh x}$$

$$\sinh(-x) = -\sinh x$$
$$\cosh(-x) = \cosh x$$
$$\cosh^2 x = 1 + \sinh^2 x$$

$$\frac{d(\sinh x)}{dx} = \cosh x$$

$$\frac{d(\cosh x)}{dx} = \sinh x$$

$$\int \sinh x \cdot dx = \cosh x$$
$$\int \cosh x \cdot dx = \sinh x$$
$$\sinh(x \pm j\pi) = -\sinh x$$
$$\cosh(x \pm j\pi) = -\cosh x$$
$$\sinh(x \pm j2\pi) = \sinh x$$
$$\cosh(x \pm j2\pi) = \cosh x$$

$$\sinh (x \pm j\pi/2) = \pm j \cosh x$$
$$\cosh (x \pm j\pi/2) = \pm j \sinh x$$
$$\tanh (x \pm j\pi n) = \tanh x$$
$$\tanh (x \pm jn\pi/2) = \frac{1}{\tanh x}$$
$$\sinh jb = j \sin b$$
$$\cosh jb = \cos b$$
$$\tanh jb = j \tan b$$
$$\sin jb = j \sinh b$$
$$\cos jb = \cosh b$$
$$\sinh (a \pm jb) = \sinh a \cos b \pm j \cosh a \sin b$$
$$\cosh (a \pm jb) = \cosh a \cos b \pm j \sinh a \sin b$$
$$\sinh \left(\frac{\theta}{2}\right) = \sqrt{\frac{\cosh \theta - 1}{2}}$$
$$\tanh \left(\frac{\theta}{2}\right) = \frac{\cosh \theta - 1}{\sinh \theta}$$
$$\sinh (\theta_1 \pm \theta_2) = \sinh \theta_1 \cosh \theta_2 \pm \cosh \theta_1 \sinh \theta_2$$
$$\cosh (\theta_1 \pm \theta_2) = \cosh \theta_1 \cosh \theta_2 \pm \sinh \theta_1 \sinh \theta_2$$

16. Series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \qquad [x^2 < \infty]$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \qquad [x^2 < \infty]$$
$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \frac{62x^9}{2835} + \cdots \qquad \left[x^2 < \frac{1}{4}\pi^2\right]$$
$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$
$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots$$
$$\epsilon^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)}{2!}x^2 \pm \frac{n(n-1)(n-2)}{3!}x^3 + \cdots \qquad (x^2 < 1)$$
$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)a^{n-2}b^2}{2!} + \cdots \qquad (b^2 < a^2)$$
$$J_0(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 4^2} - \frac{x^6}{2^2 4^2 6^2} + \cdots$$
$$J_n(x) = \frac{x^n}{2^n n!}\left\{1 - \frac{x^2}{2(2n+2)} + \frac{x^4}{2 \cdot 4(2n+2)(2n+4)} - \frac{x^6}{2 \cdot 4 \cdot 6(2n+2)(2n+4)(2n+6)} + \cdots\right\} \qquad (n \text{ an integer})$$

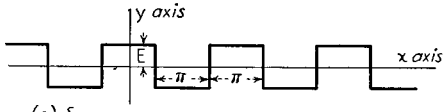
17. Harmonic Composition of Some Common Periodic Waves

Square wave (Fig. 3*a*):

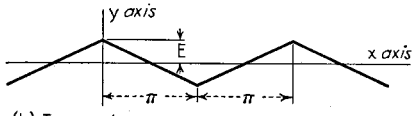
$$y = \frac{4}{\pi}E\left(\cos x - \frac{1}{3}\cos 3x + \frac{1}{5}\cos 5x - \frac{1}{7}\cos 7x + \cdots\right)$$

Triangular wave (Fig. 3*b*):

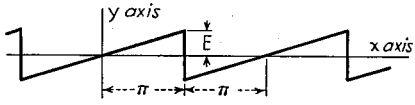
$$y = \frac{8}{\pi^2}E\left(\cos x + \frac{1}{9}\cos 3x + \frac{1}{25}\cos 5x + \cdots\right)$$



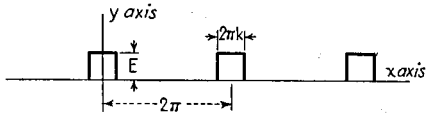
(a) Square wave



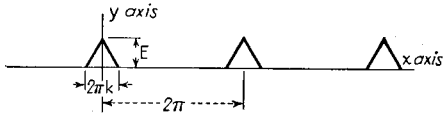
(b) Triangular wave



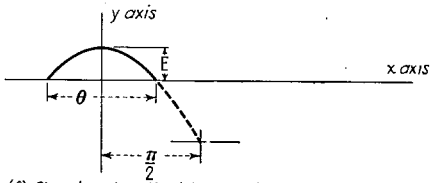
(c) Saw-tooth wave



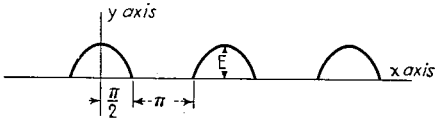
(d) Short square pulse



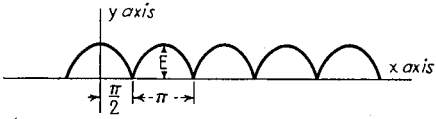
(e) Short triangular pulse



(f) Short pulse that is a section of a sine wave



(g) Half-wave rectifier output



(h) Full-wave rectifier output

FIG. 3.—Periodic pulses of various types.

Sawtooth wave (Fig. 3*c*):

$$y = \frac{2}{\pi} E \left(\sin x - \frac{1}{2} \sin 2x + \frac{1}{3} \sin 3x - \frac{1}{4} \sin 4x + \cdots \right)$$

Short square pulse (Fig. 3*d*):

$$y = E \left\{ k + \frac{2}{\pi} \left(\sin k\pi \cos x + \frac{1}{2} \sin 2k\pi \cos 2x + \frac{1}{3} \sin 3k\pi \cos 3x \cdots + \frac{1}{n} \sin nk\pi \cos nx \cdots \right) \right\}$$

The relative values of the coefficients of cos nx are plotted in Fig. 4 as a function of nk for the case $k << 1$, *i.e.*, for a short pulse.. Under these conditions the coefficients of cos nx are proportional to $(\sin nk\pi)/nk\pi$.

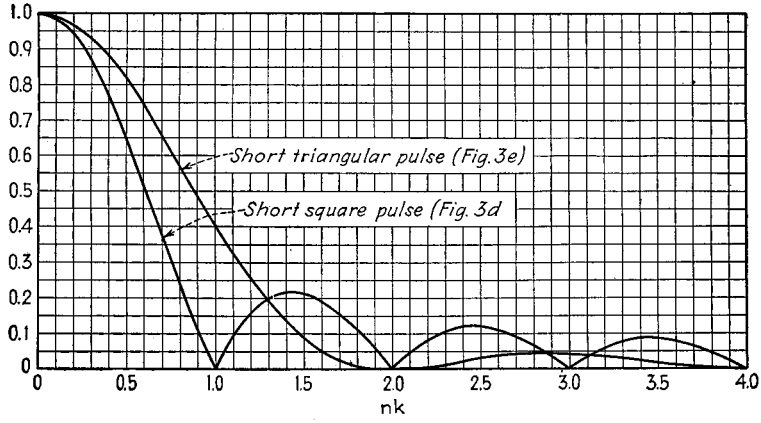


FIG. 4.—Relative amplitude of harmonics for waves of Fig. 3*d* and *e* for $k << 1$.

Short triangular pulse (Fig. 3*e*):

$$y = E \left[\frac{k}{2} + \sum_{n=1} \left\{ \frac{2}{n\pi} \sin n\pi k - \frac{2}{n^2\pi^2 k} \left(n\pi k \sin n\pi k - 2 \sin^2 \frac{n\pi k}{2} \right) \right\} \cos nx \right]$$

The relative values of the coefficients of cos nx are plotted in Fig. 4 as a function of nk, for $k << 1$, *i.e.*, for a short pulse.

Short pulse that is a section of a sine wave (Fig. 3*f*):

$$y = \frac{E}{\pi \left(1 - \cos \frac{\theta}{2} \right)} \left[\left(\sin \frac{\theta}{2} - \frac{\theta}{2} \cos \frac{\theta}{2} \right) + \left(\frac{\theta}{2} - \sin \frac{\theta}{2} \cos \frac{\theta}{2} \right) \cos x + \cdots \left(\frac{\sin (n+1) \frac{\theta}{2}}{n+1} + \frac{\sin (n-1) \frac{\theta}{2}}{n-1} - \frac{2 \sin n \frac{\theta}{2} \cos \frac{\theta}{2}}{n} \right) \cos nx + \cdots \right]$$

Half-wave rectifier output (Fig. 3*g*):

$$y = \frac{1}{\pi} E \left(1 + \frac{\pi}{2} \cos x + \frac{2}{3} \cos 2x - \frac{2}{15} \cos 4x + \frac{2}{35} \cos 6x \cdots (-1)^{\frac{n}{2}+1} \frac{2}{n^2 - 1} \cos nx \cdots \right) \qquad (n \text{ even})$$

Full-wave rectifier output (Fig. 3*h*):

$$y = \frac{2}{\pi} E \left(1 + \frac{2}{3}\cos 2x - \frac{2}{15}\cos 4x + \frac{2}{35}\cos 6x \right.$$
$$\left. \cdots (-1)^{\frac{n}{2}+1} \frac{2}{n^2 - 1}\cos nx \cdots \right) \qquad (n \text{ even})$$

UNITS

18. Units of Length

1 kilometer (km) = 1,000 meters
1 decimeter (dm) = 0.1 meter
1 centimeter (cm) = 0.01 meter
1 millimeter (mm) = 0.001 meter
1 micron (μ) = 0.000,001 meter = 0.001 millimeter
1 millimicron (mμ) = 0.000,000,001 meter = 0.001 micron
1 angstrom (A) = 0.000,000,1 millimeter
= 0.0001 micron
= 0.1 millimicron
1 statute mile = 1,760 yards = 5,280 feet
1 foot = $\frac{1}{3}$ yard = 12 inches
1 inch = $\frac{1}{36}$ yard = $\frac{1}{12}$ foot
1 mil = 1/1,000 inch
1 U.S. nautical mile, 1 sea mile, 1 geographical mile = 1,853.248 meters = 6,080.20 feet
1 international nautical mile = 1,852 meters
= 6,076.10 feet
1 fathom = 6 feet

19. Conversion Table for Units of Length

Centimeters	Meters	Kilometers	Inches	Feet	Miles
1	0.01	10^{-5}	0.3937	0.03281	0.6214×10^{-5}
2.540	0.02540	2.54×10^{-5}	1	0.0833	0.1578×10^{-4}
100	1	0.001	39.37	3.281	0.0006214
100,000	1,000	1	39,370	3,281	0.6214
30.48	0.3048	30.48×10^{-5}	12	1	0.0001894
160,935	1,609	1.609	63,360	5,280	1

NOTE: All numbers in the same horizontal column represent the same length.

20. Systems of Electrical Units

Electromagnetic Cgs Units.—The electromagnetic system of cgs units (often abbreviated emu) results if one uses centimeters, grams, and seconds, and then arbitrarily assumes that the magnetic permeance μ of a centimeter cube in a vacuum is unity. These units are sometimes designated by the prefix *ab*, such as abohm, abvolt, etc.

Electrostatic Cgs Units.—The electrostatic system of cgs units (sometimes abbreviated esu) results if one uses centimeters, grams, and seconds, and then arbitrarily assumes that the capacity across the faces of a centimeter cube is unity. These units are sometimes designated by the prefix *stat*, such as statohm, statvolt, etc.

"Practical" Electrical Units.—This term refers to the ordinary system of volts, amperes, etc., used in practice. These units have more convenient values than the electromagnetic and electrostatic systems.

Relationship between Systems of Units.—The fundamental relation between the cgs electromagnetic units and the cgs electrostatic units, as determined experimentally, is

$$1 \text{ abfarad} = 8.9878 \times 10^{20} \text{ statfarads}$$
$$\cong 9 \times 10^{20} \text{ statfarads}$$

The fundamental relation between the cgs electromagnetic system of units and the practical system is

$$1 \text{ abcoulomb} = 10 \text{ coulombs}$$
$$1 \text{ erg} = 10^{-7} \text{ watt-seconds (or joules)}$$

The erg is the unit of energy in the cgs electromagnetic system and the watt-second (or joule) that of the practical system.

It is convenient to note that when an equation is expressed in electromagnetic units one can convert to practical units by substituting 30 ohms for the velocity of light (which will appear either directly or indirectly in the relation and if it does not appear directly can be inserted by multiplying by $\frac{c}{3 \times 10^{10}} = 1$).

21. Relations between the Practical and Corresponding Cgs Units

Quantity	Symbol	Practical unit	Cgs electromagnetic unit	Cgs electrostatic unit
Emf	E	Volt $= 10^8$ abvolts Volt $= 3.\dot{3} \times 10^{-3}$ statvolt	Abvolt $= 10^{-8}$ volt Abvolt $= 3.\dot{3} \times 10^{-11}$ statvolt	Statvolt $= 300$ volts very nearly Statvolt $= 3 \times 10^{10}$ abvolts
Resistance	R	Ohm $= 10^9$ abohms Ohm $= 1.1 \times 10^{-12}$ statohm	Abohm $= 10^{-9}$ ohm Abohm $= 1.\dot{1} \times 10^{-21}$ statohm	Statohm $= 9 \times 10^{11}$ ohms Statohm $= 9 \times 10^{20}$ abohms
Current	I	Ampere $= 10^{-1}$ abampere Ampere $= 3 \times 10^9$ statamperes	Abampere $= 10$ amperes Abampere $= 3 \times 10^{10}$ statamperes	Statampere $= 3.\dot{3} \times 10^{-10}$ ampere Statampere $= 3.\dot{3} \times 10^{-11}$ abampere
Charge	Q	Coulomb $= 10^{-1}$ abcoulomb Coulomb $= 3 \times 10^9$ statcoulombs	Abcoulomb $= 10$ coulombs Abcoulomb $= 3 \times 10^{10}$ statcoulombs	Statcoulomb $= 3.\dot{3} \times 10^{-10}$ coulomb Statcoulomb $= 3.\dot{3} \times 10^{-11}$ abcoulomb
Capacity	C	Farad $= 10^{-9}$ abfarad Farad $= 9 \times 10^{11}$ statfarads	Abfarad $= 10^9$ farads Abfarad $= 9 \times 10^{20}$ statfarads	Statfarad $= 1.\dot{1} \times 10^{-12}$ farad $= 1.\dot{1}\ \mu\mu f$ Statfarad $= 1.\dot{1} \times 10^{-21}$ abfarad
Inductance	L	Henry $= 10^9$ abhenrys Henry $= 1.\dot{1} \times 10^{-12}$ stathenry	Abhenry $= 10^{-9}$ henry Abhenry $= 1.\dot{1} \times 10^{-21}$ stathenry	Stathenry $= 9 \times 10^{11}$ henrys Stathenry $= 9 \times 10^{20}$ abhenrys
Energy	W	Joule $= 10^7$ ergs	Erg $= 10^{-7}$ joule	Erg $= 10^{-7}$ joule
Power	P	Watt $= 10^7 \frac{\text{ergs}}{\text{sec}}$	$\frac{\text{Erg}}{\text{Sec}} = 10^{-7}$ watt	$\frac{\text{Erg}}{\text{Sec}} = 10^{-7}$ watt

NOTE: A dot over a number indicates a recurring decimal.

SECTION 2

CIRCUIT ELEMENTS

RESISTANCE AND RESISTORS

1. Specific Resistance and Temperature Coefficient of Resistance.—The resistance R of a conductor of length l and cross-sectional area A is

$$R = \rho \frac{l}{A} \tag{1}$$

where ρ is termed the *specific resistance*, or *resistivity*, and is a property of the conductor material and the temperature.

The specific resistance ρ can be expressed in terms of the resistance of a wire one mil in diameter and one foot long (ohms per circular mil foot), in which case the length in Eq. (1) is in feet and the area A in circular mils.[1] The specific resistance is also often expressed in terms of the resistance existing between opposite faces of a cube one centimeter on a side (ohms per cm cube), in which case it is termed the *volume resistivity*, and corresponds to the length l in Eq. (1), expressed in centimeters, and the area A, in square centimeters. Values of specific resistance for a number of common materials are given in Table 1.

The resistance of a conductor depends upon the temperature as well as the material and form. To a first approximation, the change in resistance produced by a variation in temperature is proportional to the temperature increment, provided this increment is not too large. Hence the relation between the resistances R_{T_1} and R_{T_2} of a conductor at two temperatures T_1 and T_2, respectively, is

$$R_{T_2} = R_{T_1}[1 + \alpha(T_2 - T_1)] \tag{2}$$

where α is the fractional change in resistance per degree of temperature rise, and is termed the *temperature coefficient of resistance*. Values of temperature coefficient for a number of common materials are given in Table 1.

2. Resistance Wire.—Although the usefulness of copper is due largely to its low specific resistance, other materials, particularly certain alloys, are important because of their high specific resistance. The most important of these resistance alloys for radio purposes are listed in Table 1, together with their principal properties. A wire table giving further properties of some of these resistance alloys also appears below.

The nichrome group of resistance alloys listed in Table 1 is capable of operating at high temperatures, and is the wire generally employed for rheostats, heating elements, enamel-coated resistance tubes, etc.

Advance and the related resistance alloys are characterized by zero temperature coefficient of resistance and a high thermoelectric voltage against copper. This type of wire is used for precision purposes and also for thermocouples. Its maximum working temperature is appreciably lower than that of nichrome, but considerably higher than manganin.

[1] A circular mil is the area a circle one mil (0.001 inch) in diameter. The area of a wire in circular mils is hence the square of the diameter when the diameter is expresesd in mils.

TABLE 1

Material	Resistivity		Temperature coefficient of resistance per °C at 20°C	Remarks
	Ohms per cir mil ft at 20°C	Microhms per cm cube at 20°C		
Aluminum	17	2.828	0.0049	
Brass	45	7.5	0.002–0.007	
Chromium	16	2.7		
Carbon (graphite)	400–1,100	33–185	−0.0006 to −0.0012	
Copper	10.37	1.724	0.00393	
Iron	59	9.8	0.006	
Mercury	575	95.8	0.00089	
Molybdenum	34	5.7	0.0033	
Nickel	60	10	0.005	Max. working temp. = 500°C
Phosphor bronze	70	11.5	0.004	
Silver	9.8	1.629	0.00381	
Tantalum	93	15.5	0.0031	
Tin	69	11.5	0.0042	
Tungsten	33	5.51	0.0045	
Zinc	36	5.9	0.0035	
Resistance alloys:				
Nichrome, Nichrome I–V, Chromel A-C, etc.	650	108	0.0002	Max. working temp. = 1100°C. Thermal emf against copper = 22 μv per °C
Advance, Constantan, Copel, Ideal, $I_a - I_a$, etc.	295	49	0	Max. working temp. = 500°C. Thermal emf against copper = 43 μv per °C
Manganin	290	48	±0.00001	Thermal emf against copper = 2 μv per °C
Ohmax	1,000	167	−0.00035	Max. working temp. = 500°C
Radiohm	800	133	0.0007	
Monel	255	43	0.0019	Max. working temp. = 425°C
German silver	185	31	0.00027	
Midohm	180	30	0.00018	

NOTE: The foregoing are typical values. Actual values in the case of elements will depend on purity, heat treatment, etc., and also upon exact composition in the case of alloys.

Manganin is used primarily for precision resistors. It has negligible temperature coefficient of resistance at ordinary room temperatures, and also an extremely small thermoelectric coefficient against copper. The resistance of manganin is affected appreciably by mechanical strains such as are introduced by winding, but when these are relieved by baking the completed resistance units for 24 hours at a temperature of 120°C, the resistance is extremely stable in its characteristics, provided that the resistance is not subsequently overheated.

Ohmax is an aluminum-bearing alloy having a high specific resistance, but is available only in the smaller sizes (10 mils and less). It is used where high resistance is required in a small space, as in radio rheostats. It has a negative temperature coefficient of resistance and is sometimes employed in combination with nichrome to

give zero over-all temperature coefficient in precision resistors requiring high resistance in such a small space that manganin is not suitable.

Radiohm is similar to ohmax, but has a slightly lower specific resistance, is available in a greater range of sizes, and has a positive temperature coefficient of resistivity.

Midohm is a copper-nickel alloy having a lower temperature coefficient of resistivity than other alloys of similar specific resistance.

3. Wire Tables.

TABLE 2.—COPPER-WIRE TABLE, STANDARD ANNEALED COPPER
American Wire Gage (B & S)

Gage No.	Diameter, mils at 20°C	Cross section at 20°C		Ohms per 1,000 ft at 20°C (= 68°F)	Lb. per 1,000 ft	Ft per lb	Ft per ohm at 20°C (= 68°F)	Ohms per lb at 20°C (= 68°F)	Gage No.
		Circular mils	Sq in.						
0000	460.0	211,600.0	0.1662	0.04901	640.5	1.561	20,400.0	0.00007652	0000
000	409.6	167,800.0	0.1318	0.06180	507.9	1.968	16,180.0	0.0001217	000
00	364.8	133,100.0	0.1045	0.07793	402.8	2.482	12,830.0	0.0001935	00
0	324.9	105,500.0	0.08289	0.09827	319.5	3.130	10,180.0	0.0003076	0
1	289.3	83,690.0	0.06573	0.1239	253.3	3.947	8,070.0	0.0004891	1
2	257.6	66,370.0	0.05213	0.1563	200.9	4.977	6,400.0	0.0007778	2
3	229.4	52,640.0	0.04134	0.1970	159.3	6.276	5,075.0	0.001237	3
4	204.3	41,740.0	0.03278	0.2485	126.4	7.914	4,025.0	0.001966	4
5	181.9	33,100.0	0.02600	0.3133	100.2	9.980	3,192.0	0.003127	5
6	162.0	26,250.0	0.02062	0.3951	79.46	12.58	2,531.0	0.004972	6
7	144.3	20,820.0	0.01635	0.4982	63.02	15.87	2,007.0	0.007905	7
8	128.5	16,510.0	0.01297	0.6282	49.98	20.01	1,592.0	0.01257	8
9	114.4	13,090.0	0.01028	0.7921	39.63	25.23	1,262.0	0.01999	9
10	101.9	10,380.0	0.008155	0.9989	31.43	31.82	1,001.0	0.03178	10
11	90.74	8,234.0	0.006467	1.260	24.92	40.12	794.0	0.05053	11
12	80.81	6,530.0	0.005129	1.588	19.77	50.59	629.6	0.08035	12
13	71.96	5,178.0	0.004067	2.003	15.68	63.80	499.3	0.1278	13
14	64.08	4,107.0	0.003225	2.525	12.43	80.44	396.0	0.2032	14
15	57.07	3,257.0	0.002558	3.184	9.858	101.4	314.0	0.3230	15
16	50.82	2,583.0	0.002028	4.016	7.818	127.9	249.0	0.5136	16
17	45.26	2,048.0	0.001609	5.064	6.200	161.3	197.5	0.8167	17
18	40.30	1,624.0	0.001276	6.385	4.917	203.4	156.6	1.299	18
19	35.89	1,288.0	0.001012	8.051	3.899	256.5	124.2	2.065	19
20	31.96	1,022.0	0.0008023	10.15	3.092	323.4	98.50	3.283	20
21	28.46	810.1	0.0006363	12.80	2.452	407.8	78.11	5.221	21
22	25.35	642.4	0.0005046	16.14	1.945	514.2	61.95	8.301	22
23	22.57	509.5	0.0004002	20.36	1.542	648.4	49.13	13.20	23
24	20.10	404.0	0.0003173	25.67	1.223	817.7	38.96	20.99	24
25	17.90	320.4	0.0002517	32.37	0.9699	1,031.0	30.90	33.37	25
26	15.94	254.1	0.0001996	40.81	0.7692	1,300.0	24.50	53.06	26
27	14.20	201.5	0.0001583	51.47	0.6100	1,639.0	19.43	84.37	27
28	12.64	159.8	0.0001255	64.90	0.4837	2,067.0	15.41	134.2	28
29	11.26	126.7	0.00009953	81.83	0.3836	2,607.0	12.22	213.3	29
30	10.03	100.5	0.00007894	103.2	0.3042	3,287.0	9.691	339.2	30
31	8.928	79.70	0.00006260	130.1	0.2413	4,145.0	7.685	539.3	31
32	7.950	63.21	0.00004964	164.1	0.1913	5,227.0	6.095	857.6	32
33	7.080	50.13	0.00003937	206.9	0.1517	6,591.0	4.833	1,364.0	33
34	6.305	39.75	0.00003122	260.9	0.1203	8,310.0	3.833	2,168.0	34
35	5.615	31.52	0.00002476	329.0	0.09542	10,480.0	3.040	3,448.0	35
36	5.000	25.00	0.00001964	414.8	0.07568	13,210.0	2.411	5,482.0	36
37	4.453	19.83	0.00001557	523.1	0.06001	16,660.0	1.912	8,717.0	37
38	3.965	15.72	0.00001235	659.6	0.04759	21,010.0	1.516	13,860.0	38
39	3.531	12.47	0.000009793	831.8	0.03774	26,500.0	1.202	22,040.0	39
40	3.145	9.888	0.000007766	1,049.0	0.02993	33,410.0	0.9534	35,040.0	40

4. Resistance at Radio Frequencies—Skin Effect.—In circuits carrying alternating currents, particularly high-frequency currents, the power loss is often greater than where a direct current of the same value is being carried. This is because of such

TABLE 3.—RESISTANCE WIRE TABLE

B & S No.	Diam., in.	Nichrome		Advance		Manganin		B & S No.
		Ohms per ft at 68°F (20°C)	Weight per 1,000 ft bare wire, lb	Ohms per ft at 68°F (20°C)	Weight per 1,000 ft bare wire, lb	Ohms per ft at 68°F (20°C)	Weight per 1,000 ft bare wire, lb	
10	0.102	0.06248	29.60	0.02826	31.35	.....	.....	10
11	0.091	0.07849	23.65	0.03550	25.06	.....	.....	11
12	0.081	0.09907	18.67	0.04481	19.77	.....	.....	12
13	0.072	0.1255	14.75	0.05672	15.62	.....	.....	13
14	0.064	0.1588	11.65	0.07178	12.35	.....	.....	14
15	0.057	0.2000	9.240	0.09049	9.788	0.0893	9.054	15
16	0.051	0.2499	7.400	0.1130	7.836	0.1115	7.249	16
17	0.045	0.3209	5.760	0.1452	6.100	0.1432	5.642	17
18	0.040	0.4062	4.550	0.1837	4.822	0.1813	4.460	18
19	0.036	0.5015	3.690	0.2270	3.906	0.2238	3.613	19
20	0.032	0.6347	2.915	0.2871	3.085	0.2832	2.854	20
21	0.0285	0.8002	2.294	0.3619	2.430	0.3570	2.248	21
22	0.0253	1.017	1.807	0.4557	1.946	0.4531	1.771	22
23	0.0226	1.272	1.453	0.5756	1.539	0.5678	1.424	23
24	0.0201	1.609	1.139	0.7277	1.218	0.7178	1.116	24
25	0.0179	2.029	0.9110	0.9176	0.9655	0.9051	0.8931	25
26	0.0159	2.571	0.7190	1.163	0.7619	1.147	0.7047	26
27	0.0142	3.228	0.5740	1.458	0.6076	1.438	0.5620	27
28	0.0126	4.090	0.4540	1.852	0.4786	1.826	0.4427	28
29	0.0113	5.090	0.3570	2.302	0.3785	2.271	0.3501	29
30	0.0100	6.500	0.2845	2.940	0.3014	2.900	0.2788	30
31	0.0089	8.206	0.2253	3.722	0.2387	3.662	0.2208	31
32	0.0080	10.16	0.1821	4.594	0.1929	4.531	0.1784	32
33	0.0071	12.90	0.1434	5.833	0.1519	5.754	0.1405	33
34	0.0063	16.37	0.1129	7.408	0.1196	7.305	0.1106	34
35	0.0056	20.72	0.08922	9.375	0.09451	9.247	0.08742	35
36	0.0050	26.00	0.07113	11.76	0.07534	11.60	0.06969	36
37	0.0045	32.09	0.05758	14.52	0.06100	14.32	0.05642	37
38	0.0040	40.62	0.04552	18.37	0.04822	18.13	0.04460	38
39	0.0035	53.06	0.03482	24.00	0.03689	23.67	0.03412	39
40	0.0031	67.63	0.02731	30.60	0.02893	30.18	0.02676	40
	0.00275	85.98	0.02128	38.88	0.02254	38.36	0.02085	
	0.00250	104.00	0.01775	47.04	0.01880	46.40	0.01739	
	0.00225	128.5	0.01423	58.07	0.01507	57.31	0.01394	
	0.00200	162.5	0.01138	73.50	0.01205	72.50	0.01115	
	0.00175	212.4	0.008706	96.00	0.009222			
	0.00150	288.9	0.006396	130.70	0.006774			
	0.0014	331.6	0.005577					
	0.0013	384.6	0.004809					
	0.0012	451.4	0.004097					
	0.0011	537.2	0.003443					
	0.001	650.0	0.002845					

* This table is reproduced by permission from *Booklet R-36* of the Driver-Harris Company.

factors as dielectric hysteresis, eddy currents, skin effect, etc. As a result, it is customary, in dealing with alternating-current circuits, to consider that the effective or equivalent circuit resistance is that quantity which when multiplied by the square of the current equals the power dissipated in the circuit.

Skin Effect.—At high frequencies the current carried by a conductor is not uniformly distributed over the conductor cross section, as is the case with direct currents, but rather tends to be concentrated near the surface. This action, termed *skin effect*, is a result of magnetic flux lines that circle part but not all of the conductor. Those parts of the cross section which are circled by the largest number of flux lines have higher inductance than other parts of the conductor, and hence a greater reactance. The result is a redistribution of current over the cross section in such a way as to

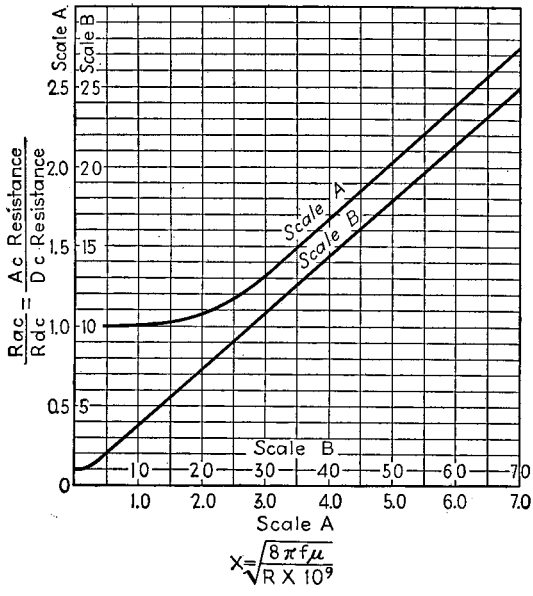


FIG. 1.—Resistance ratio in isolated round wires as a function of parameter x defined by Eq. (3).

cause those parts of the conductor having the highest reactance, *i.e.*, those parts nearer the center, to carry the least current. With a round wire this causes the current density to be maximum at the surface and least at the center. With a square bar the greatest concentration of current is at the corners, with the flat sides coming next and the center carrying the least current. With a flat strip the current density is greatest at the edges, considerable at the flat surfaces, and again least in the center. In every case it will be noticed that the current is so distributed as to cause those parts of the cross section that are inclosed by the greatest number of flux lines to carry the least current.

The redistribution of current over a conductor cross section that is associated with skin effect causes the ratio R_{ac}/R_{dc} of effective alternating-current resistance to direct-current resistance to be greater than unity, because in the presence of skin effect, portions of the conductor are not fully effective in carrying current. Along with this increase in resistance, there is also a decrease in the inductance of the conductor because of the fact that the redistribution of current is always of such a character

as to make the flux linkages, and hence the inductance, less than with a uniform current distribution. The magnitude of these effects on inductance and a-c resistance increases with frequency, conductivity, magnetic permeability, and size of the the conductor. This is because a high frequency increases the difference in reactance resulting from the different inductance of various current paths, whereas greater conductivity makes the same difference in reactance more effective in modifying current distribution, and a greater magnetic permeability increases the flux.

In dealing with skin-effect problems, it is often convenient to make use of the principle of similitude, which states that in the case of nonmagnetic materials, if a conductor or combination of conductors has a given ratio R_{ac}/R_{dc} at a certain value of f/R_{dc}, then, if the relative shapes are not altered, changes in the frequency, conductor, resistivity, and size that do not alter f/R_{dc} will not alter the resistance ratio R_{ac}/R_{dc}.[1]

Resistance Ratio of Isolated Conductors.—The ratio R_{dc}/R_{ac} of alternating-current to direct-current resistance for straight cylindrical wires is given in Fig. 1 and

TABLE 4.—RATIO OF ALTERNATING-CURRENT RESISTANCE TO DIRECT-CURRENT RESISTANCE FOR A SOLID ROUND WIRE

x	$\frac{R_{ac}}{R_{dc}}$	x	$\frac{R_{ac}}{R_{dc}}$	x	$\frac{R_{ac}}{R_{dc}}$
0	1.0000	5.2	2.114	14.0	5.209
0.5	1.0003	5.4	2.184	14.5	5.386
0.6	1.0007	5.6	2.254	15.0	5.562
0.7	1.0012	5.8	2.324	16.0	5.915
0.8	1.0021	6.0	2.394	17.0	6.268
0.9	1.0034	6.2	2.463	18.0	6.621
1.0	1.005	6.4	2.533	19.0	6.974
1.1	1.008	6.6	2.603	20.0	7.328
1.2	1.011	6.8	2.673	21.0	7.681
1.3	1.015	7.0	2.743	22.0	8.034
1.4	1.020	7.2	2.813	23.0	8.387
1.5	1.026	7.4	2.884	24.0	8.741
1.6	1.033	7.6	2.954	25.0	9.094
1.7	1.042	7.8	3.024	26.0	9.447
1.8	1.052	8.0	3.094	28.0	10.15
1.9	1.064	8.2	3.165	30.0	10.86
2.0	1.078	8.4	3.235	32.0	11.57
2.2	1.111	8.6	3.306	34.0	12.27
2.4	1.152	8.8	3.376	36.0	12.98
2.6	1.201	9.0	3.446	38.0	13.69
2.8	1.256	9.2	3.517	40.0	14.40
3.0	1.318	9.4	3.587	42.0	15.10
3.2	1.385	9.6	3.658	44.0	15.81
3.4	1.456	9.8	3.728	46.0	16.52
3.6	1.529	10.0	3.799	48.0	17.22
3.8	1.603	10.5	3.975	50.0	17.93
4.0	1.678	11.0	4.151	60.0	21.47
4.2	1.752	11.5	4.327	70.0	25.00
4.4	1.826	12.0	4.504	80.0	28.54
4.6	1.899	12.5	4.680	90.0	32.07
4.8	1.971	13.0	4.856	100.0	35.61
5.0	2.043	13.5	5.033	∞	∞

[1] See H. B. Dwight, Skin Effect on Tubular and Flat Conductors, *Trans. A.I.E.E.*, Vol. 37, p. 1379, 1918.

Table 4[1] in terms of a parameter x defined by the equation

$$\left.\begin{aligned} x &= \pi d \sqrt{\frac{2\mu f}{\rho \times 10^9}} = \sqrt{\frac{8\pi\mu f A}{\rho \times 10^9}} \\ &= \sqrt{\frac{8\pi\mu f}{R \times 10^9}} \quad = 0.1585 \sqrt{\frac{\mu f_{mc}}{R}} \end{aligned}\right\} \tag{3}$$

where A = wire area, sq cm.
d = wire diameter, cm.
μ = permeability of conductor ($\mu = 1$ for air).
ρ = specific resistivity in ohms per cm cube.
f and f_{mc} = frequency in cycles and megacycles, respectively.
R = resistance to direct current for 1 cm of conductor, ohms.

In the case of copper wire, $\mu = 1$, and $\rho = 1.724 \times 10^{-6}$, so that Eq. (3) becomes

$$x \text{ (for copper)} = 0.271 d_m \sqrt{f_{mc}} \tag{4}$$

where d_m is the diameter in mils. The largest diameter that is permissible at various frequencies for a resistance ratio not exceeding 1.01 is tabulated in Table 5 for several common materials.

TABLE 5.—LARGEST PERMISSIBLE WIRE DIAMETER IN MILS FOR SKIN-EFFECT RATIO OF 1.01 OR LESS

Frequency, kc	Nichrome	Advance and manganin	Copper
100	104.5	70.2	14.0
200	74.0	49.6	9.9
500	46.8	31.4	6.3
1,000	33.0	22.2	4.4
2,000	23.4	15.7	3.1
5,000	14.7	9.9	2.0
10,000	10.4	7.0	1.4
20,000	7.4	5.0	1.0
50,000	4.7	3.1	0.6

For a resistance ratio of 1.1, multiply the preceding diameters by 1.78. For a resistance ratio of 1.001, multiply the preceding diameters by 0.55.

In the case of isolated tubular conductors, the resistance ratio is always closer to unity than for a solid conductor of the same outside diameter. This is because the center of a solid wire does not do its full share in the carrying of current; so if the center is removed to form a tube the resistance ratio will be improved. However, removing the center to form a tube increases the d-c resistance sufficiently so that the a-c resistance of the tube is greater than for the corresponding solid wire, even though the resistance ratio is less. Values of resistance ratio for isolated nonmagnetic tubular conductors are given in Fig. 2.[2]

A conductor consisting of a flat rectangular strip (ribbon) will have a lower resistance ratio than a solid round wire of the same cross section when the frequency is

[1] From *Bur. Standards Circ.* 74.

For the derivation of the resistance ratio of a round wire, see A. E. Kennelly, F. A. Laws, and P. H. Pierce, Experimental Researches on Skin Effect in Conductors, *Trans. A.I.E.E.*, Vol. 34, p. 1953, 1915; or see L. F. Woodruff, "Electric Power Transmission and Distribution," 2d ed., p. 54, Wiley, New York, 1938.

[2] The analysis of the tubular conductor case is given by Woodruff, *op. cit.*, and also by H. B. Dwight, A Precise Method of Calculating Skin Effect in Isolated Tubes, *A.I.E.E. Jour.*, Vol. 42, p. 827, August, 1923.

high enough to make the resistance ratio appreciable, but has a higher resistance ratio than would be obtained by forming the ribbon into a tube of the same cross-sectional area and the same wall thickness. This is because the current in a ribbon tends to

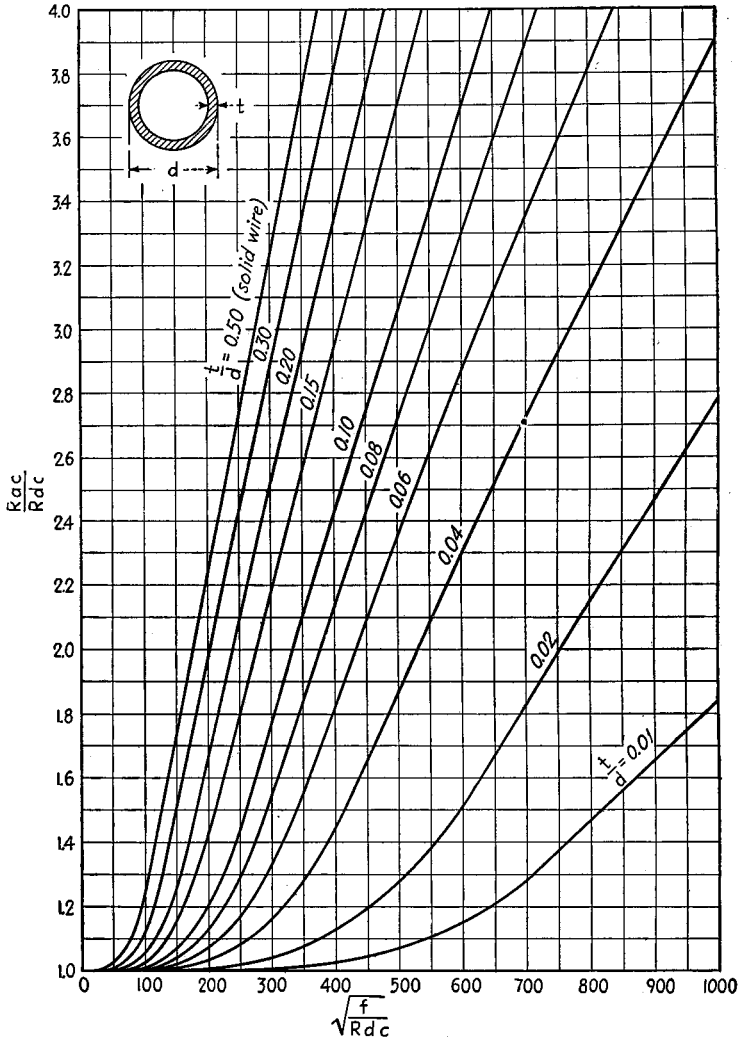


FIG. 2.—Resistance ratio of isolated nonmagnetic tubular conductors in terms of parameter $\sqrt{f/R_{dc}}$, where f is the frequency in cycles, and R_{dc} the direct-current resistance per thousand feet.

concentrate more at the edges than at the sides, since the edges are circled by the fewest flux lines. Information on the resistance ratio of rectangular conductors of various shapes is given in Fig. 3.[1] It is seen that for shape ratios exceeding about

[1] Most of the results in Fig. 3 are based on experimental data given by S. J. Haefner, Alternating Current Resistance of Rectangular Conductors, *Proc. I.R.E.*, Vol. 25, p. 434, April, 1937.

2:1, the ribbon has lower a-c resistance than a solid round wire of the same cross-section except when the resistance ratio approaches unity.

Skin Depth and Resistance at Very High Frequencies.—At frequencies so great that the resistance ratio is large, substantially all of the current carried by the conductor is concentrated very close to the surface. Under these conditions, the a-c resistance offered by the conductor is approximately the same as the d-c resistance of a hollow

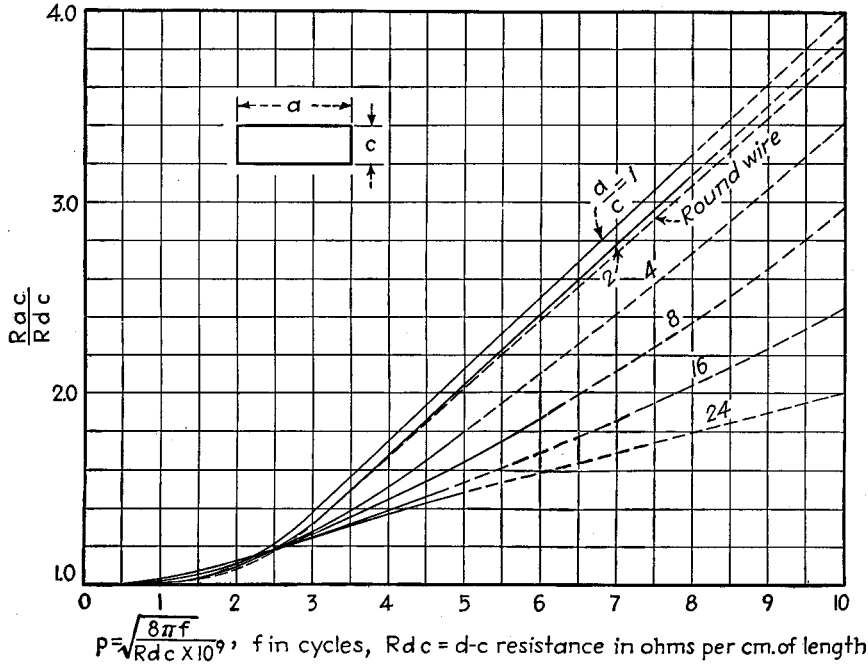


FIG. 3.—Resistance ratio of rectangular conductors in terms of parameter p, involving the frequency f in cycles, and R_{dc} the resistance in ohms per centimeter of length. These curves are a combination of experimental and calculated results, with the dotted portions representing low-frequency results extrapolated to join on with the theoretical results for high frequencies.

conductor having the same external shape as the actual conductor but having a thickness equal to

$$\text{Skin depth, cm} = \epsilon = \frac{1}{2\pi}\sqrt{\frac{\rho \times 10^9}{\mu f}} = 5{,}033\sqrt{\frac{\rho}{\mu f}} \qquad (5)$$

where ρ is in ohms per cm cube and f is in cycles. This thickness, termed the *skin depth*, is a rough measure of the current penetration since the current density drops off to a very small value at a distance corresponding to several skin depths. For copper at 20°C, the skin depth is $6.62/\sqrt{f}$ cm, or $2.61/\sqrt{f}$ inches, where f is in cycles. The skin depth of copper is given graphically in Fig. 4.

It will be noted that where the skin depth controls the a-c resistance, a-c resistance is directly proportional to the square root of frequency and inversely proportional to the conductor size (or perimeter).

The skin-depth concept can be used to calculate the radio-frequency resistance of conductors of any shape *provided the conductor thickness is everywhere considerably greater than the skin depth and provided the radius of curvature of the conductor surface*

is likewise everywhere appreciably greater than the skin depth and does not vary too rapidly around the periphery. Under these conditions, the high-frequency resistance of a copper conductor is

$$\text{High-frequency resistance, ohms per centimeter} = \frac{261\sqrt{f} \times 10^{-9}}{P} \qquad (6)$$

where P is the perimeter in centimeters and f is in cycles. This equation is also plotted in Fig. 4. Equation (6) can be expected to hold reasonably well for resistance ratios exceeding 2, and quite closely when the ratio is of the order of 5 or more. If the radius of curvature varies too rapidly around the periphery, the a-c resistance is greater than given by Eq. (6) by a factor that depends on the shape. It will be noted

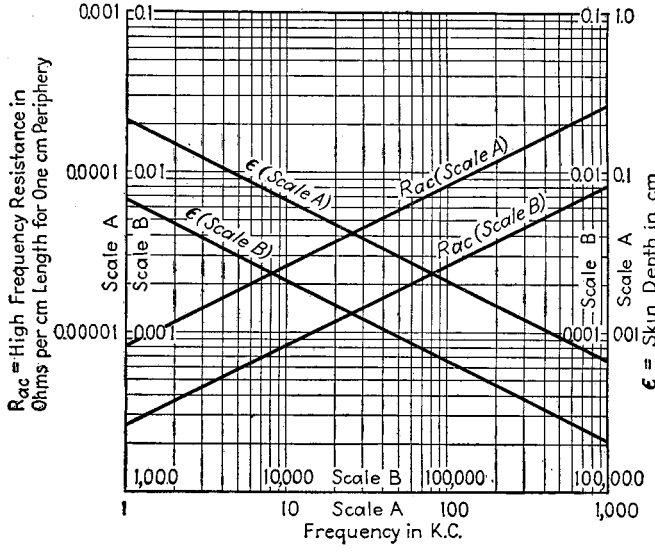


FIG. 4.—Skin depth and high-frequency resistance of copper at frequencies so high that substantially the entire current is concentrated near the conductor surface.

that when Eq. (6) holds, the alternating-current resistance is directly proportional to the square root of the frequency, and inversely proportional to the conductor size (or perimeter).

In the special case where the conductor is a round wire or tube, one has $P = \pi d$, where d is the outside diameter in centimeters, and Eq. (6) becomes

$$\left.\begin{matrix}\text{High-frequency resistance of wire or tube} \\ \text{at high frequencies, ohms per cm}\end{matrix}\right\} = \frac{83.2\sqrt{f}}{d} \times 10^{-9} \qquad (7)$$

In the case of a rectangular conductor, the radius of curvature varies so rapidly at the corners that Eq. (6) must be modified as follows:

$$\text{High-frequency resistance of ribbon, ohms per centimeter} = K\frac{261\sqrt{f}}{2(a+c)} \times 10^{-9} \qquad (8)$$

where a and c are the width and thickness, respectively, in centimeters, f is in cycles, and K is a constant determined by the ratio a/c, and given by Fig. 5.[1] Equation (8)

[1] This figure is calculated from formulas derived by J. D. Cockroft, Skin Effect in Rectangular Conductors at High Frequency, *Proc. Roy. Soc. (London)*, Vol. 122, No. A790, p. 533, Feb. 4, 1929.

will give reasonably accurate results when the thickness is appreciably greater than twice the skin depth.

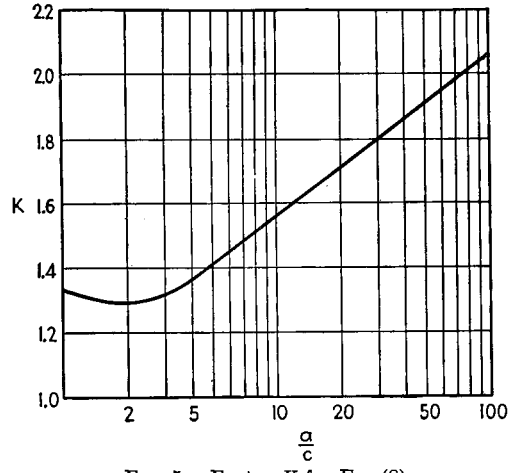


FIG. 5.—Factor K for Eq. (8).

Proximity Effect.—When two or more adjacent conductors are carrying current, the current distribution in one conductor is affected by the magnetic flux produced by the adjacent conductors, as well as the magnetic flux produced by the current in the conductor itself. This effect is termed *proximity effect*, and ordinarily causes the resistance ratio to be greater than in the case of simple skin effect. Proximity effect is very important in radio-frequency inductance coils, and is sometimes a factor in other cases.

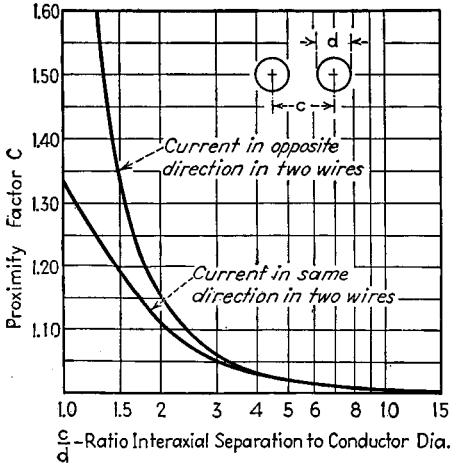


FIG. 6.—Proximity factor in parallel cylinders or tubes for very high frequencies.

Formulas for accurately calculating the proximity effect have been worked out for only very simple cases. For a two-wire line employing round or tubular wires and operating at frequencies so high that the skin depth is small, the factor C by which proximity effect increases the resistance of the conductor is given in Fig. 6.[1] When the frequency is low enough so that the skin depth is not negligible compared with other dimensions involved, the proximity effect will be less than that given by Fig. 6 and can be calculated according to formulas available in the literature.[2]

[1] The curve for currents in the opposite direction is calculated by formulas given by Sallie Pero Mead, Wave Propagation over Parallel Tubular Conductors: The Alternating Current Resistance, *Bell System Tech. Jour.*, Vol. 4, p. 327, April, 1925. The curve applying to currents flowing in the same direction is from S. Butterworth, On the Alternating Current Resistance of Solenoidal Coils, *Proc. Roy. Soc. (London)*, Vol. 107*A*, p. 693, 1925.

[2] See Mead, *op. cit.*; also, H. B. Dwight, Proximity Effect in Wires and Thin Tubes, *Trans. A.I.E.E.*, Vol. 42, p. 850, 1923.

The case of two flat wound ribbons with parallel flat sides adjacent to each other and very close together has not been worked out, but experiments indicate that the proximity effect in this case, assuming the ribbons carry equal currents flowing in the same direction, is very small and sometimes even negative (*i.e.*, may actually reduce the total resistance to alternating currents).

Litz Wire.—The resistance ratio of a conductor can be made to approach unity very closely, at least at the lower radio frequencies, by use of a conductor consisting of a large number of strands of fine wire that are insulated from each other except at the ends where the various wires are connected in parallel. Such a conductor is termed a *litz* or Litzendraht conductor.

In order that a litz conductor may be effective, the strands must be woven or transposed in such a way that each strand occupies all possible positions in the cable to approximately the same extent. In this way the total flux linkages surrounding the various individual strands of the cable, when totaled over an appreciable length, will be substantially the same for all strands even though at any one spot the strands near the center are circled by more flux than the strands that are near the surface at that particular point. The result of this equalizing the flux linkages, and hence reactances, of the individual strands is to cause the current to divide uniformly between strands. The resistance ratio then tends to approximate unity.

The ratio of a-c to d-c resistance of an isolated litz wire is[1]

$$\frac{\text{Resistance to alternating currents}}{\text{Resistance to direct currents}} = H + k\left(\frac{nd_s}{d_0}\right)^2 G \qquad (8a)$$

where H = resistance ratio of individual strand when isolated, as obtained from Table 18 (or Table 4), with x evaluated for the *diameter of the individual strand.*

G = constant taking into account the proximity of neighboring wires, and given by Table 18.

n = number of strands in cable.

d_s = diameter of individual strand.

d_0 = diameter of cable.

k = constant depending on n, given in the following table.

n =	3	9	27	Infinity
k =	1.55	1.84	1.92	2.0

Practical litz conductors are very effective at frequencies below 500 kc, but as the frequency becomes higher the benefits obtained are less. This is because irregularities in stranding and capacity between the individual strands result in failure to realize the ideal conditions. As a consequence, litz wire becomes progressively less effective as the frequency is increased and seldom is useful at frequencies greater than about 2 mc.

5. Types of Resistors Most Frequently Used in Communication.—Resistors used in communication work are commonly divided into wire-wound and composition types according to whether the resistance element is metallic, or is a conducting compound, respectively.

Fixed Wire-wound Resistors.—Fixed wire-wound resistors consist of resistance wire, normally nichrome or an equivalent type, wound on an insulating form. Where

[1] S. Butterworth, Effective Resistance of Inductance Coils at Radio Frequency, *Exp. Wireless and Wireless Eng.*, Vol. 3, p. 483, August, 1926.

appreciable power is to be dissipated, the form is either a ceramic tube, or a metal sheet covered with asbestos. In the former case the resistance wire is protected against mechanical injury and corrosion by a cement or vitreous enamel coating. Power resistors of these types are available in a large variety of sizes and mounting arrangements, and are also available with intermediate taps. The power ratings

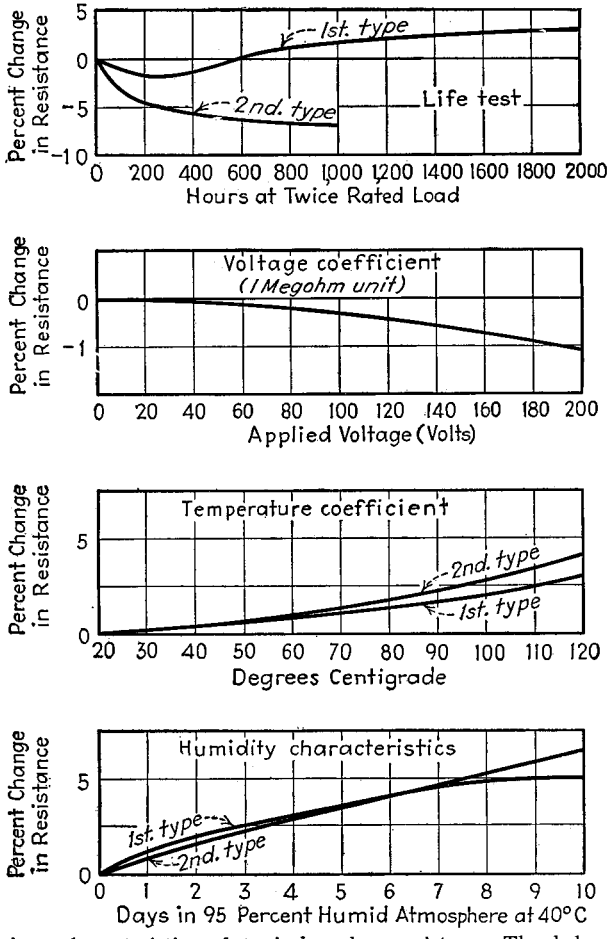


FIG. 7.—Various characteristics of typical carbon resistors. The behavior will vary greatly with ohmage, wattage rating, manufacturer, etc.

depend upon the size, method of construction, and freedom with which air is allowed to circulate around the resistor.

Resistors that are called upon to dissipate only a few watts commonly employ insulated wire wound on a bakelite strip, or on a molded or ceramic spool. The spool-type resistor is widely used for meter multipliers and similar applications, while the strip type is employed for the center-tapped resistance across filaments, etc.

Composition Resistors.—The commonest type of composition resistor is the "carbon" resistor, widely used in radio work. This obtains its resistance by combining powdered carbon or graphite with a synthetic resin bond and an inert material such as

talc, in the proper proportions. Another type of composition resistor makes use of a conducting filament that is painted or deposited on a glass rod or other insulator. The "metalized" resistors are of this type.

Composition resistors are made in ratings from ¼ watt to several watts dissipation, and can be obtained with a variety of terminal arrangements. The conducting filament (or metalized) type and also many of the carbon types are inclosed in an insulating tube. With carbon resistors, the resistance material is, however, often simply formed in the shape of a stick, provided with terminals, and then painted.

Composition resistors differ from wire-wound resistors in that the resistance is less stable with time, drops somewhat when large voltages are applied, and has a large negative temperature coefficient of resistance.

Typical performance curves of composition resistors are shown in Fig. 7. Standard procedures for making these and other tests on composition resistors have been worked out by the Radio Manufacturers Association.

Color Code.—Composition resistors normally have their resistance value indicated by a color code established by the Radio Manufacturers Association. In this code, the color of the body of the resistor designates the first digit. The color at the end of the resistor designates the second digit, and the color of the dot (or band) in the center of the resistor indicates the number of ciphers after the first two digits. The colors used to designate different numbers are as follows:

0	Black	5	Green
1	Brown	6	Blue
2	Red	7	Violet
3	Orange	8	Gray
4	Yellow	9	White

Thus a resistor with a black body, red end, and brown band would be 20 ohms, whereas one with a green body, black end, and yellow band would be 500,000 ohms.

When direct current is passed through a composition resistor, there are minute irregularities in the voltages appearing across the resistor. These are the result of random variations in the resistance due to intermittent contacts between conducting granules, and cause small audio-frequency (and also radio-frequency) voltages to appear across the resistor terminals.[1] As a result of this "noise," it is not permissible to use most types of composition resistors as the coupling resistance in the first stages of high gain amplifiers.

Variable Resistors Used in Communication Work–Volume Controls.—When a resistor is to be used as a rheostat, potentiometer, or volume control, an adjustable connection to the resistance element is required. The construction depends upon the power that must be dissipated. An arrangement very widely used in laboratory work to handle powers in excess of 25 watts makes use of resistance wire wound on a ceramic metal tube provided with a sliding contact. The wire is ordinarily bare and is oxidized to provide insulation between turns. In many cases all the wire except that over which the slider operates is coated with vitreous enamel to provide mechanical protection and additional insulation. Another arrangement employs a rotating contact arm, with the resistance wire wound on a circular ceramic core or on an asbestos-covered metal strip.

Where the power to be dissipated is only a few watts, the resistance element is generally wound on a fiber strip that is bent into a circular arc and mounted so that a slider operates over an exposed edge. The wire is usually enameled, with the

[1] C. J. Christensen and G. L. Pearson, Spontaneous Resistance Fluctuations in Carbon Microphones and Other Granular Resistances, *Bell System Tech. Jour.*, Vol. 15, p. 181, April, 1936.

enamel being cleaned off where the slider touches. Wire-wound variable resistors of these types are available up to about 100,000 ohms, and are capable of dissipating up to about 15 watts, according to size.

When a variable resistance of very low ohmage is required, it is customary to use a slide wire. In its simplest form this consists of a disk that has a single resistance wire mounted on its periphery and is rotated in contact with a slider.

Resistors used for volume control and similar applications need dissipate only negligible power, but often must have very high resistance. They must also generally be tapered, *i.e.*, the variation of resistance with rotation must be nonlinear. Such resistors may be either of the wire-wound or composition types. The wire-wound type is suitable for total resistances up to about 50,000 ohms, and, although more expensive than the composition type, has greater durability, is able to dissipate more power, and is less likely to become noisy. Wire-wound resistors can be tapered by varying the shape of the form upon which the wire is wound to give a turn length that varies for different parts of the resistor, by changing the winding pitch, and by using a different size of wire in different parts of the resistor.

Composition-type volume controls in a typical case utilize a resistance element consisting of either an absorbent paper upon which a conducting solution is painted or a bakelite form upon which the conducting compound is deposited. The desired taper is obtained by the shape of the pattern formed by the conducting paint, the number of coats of paint, and the thickness and composition of this paint. Contact with the rotating element can be made without introducing excessive wear by means of a roller, a floating disk, or very fine sliding wires. Composition-type volume controls are inexpensive, provide high resistances, and make possible almost any conceivable taper. Their chief disadvantage is lack of durability and low dissipation.

6. Inductance and Capacity Effects Associated with Resistors and Resistor Behavior at High Frequencies.—Every resistor has a certain amount of inductance and capacity associated with it. The inductance arises because whenever current flows through a conductor, magnetic fields are produced. Capacity is inevitably present because of the capacity between terminals, between various parts of the resistor and terminals, and between parts of the resistor. As a result, a resistor can be represented by the equivalent circuit shown in Fig. 8, in which L is an equivalent inductance that takes into account the magnetic fields produced by the current flowing through the resistor and C is the equivalent lumped capacity which when connected in shunt with the resistance terminals will act in the same way as the actual distributed capacities associated with the resistor. This is a low Q parallel resonant circuit, the properties of which are discussed in Sec. 3, Par. 2.

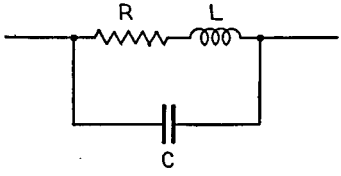


FIG. 8.—Equivalent circuit of resistor, showing inductive and capacitive effects associated with a resistor.

When effects from the inductance are to be minimized, it is necessary that $\omega L << R$ and that $1/\omega C >> R$. Under these conditions, the resistor can be thought of as consisting of a resistance R_{eq} in series with an equivalent inductance L_{eq}, for which

$$\text{Equivalent resistance} = R_{eq} = R[1 + \omega^2 C(2L - R^2 C)] \quad (9)$$

$$\text{Equivalent inductance} = L_{eq} = L - R^2 C \quad (10)$$

These equations show that if $2L = R^2 C$, then the equivalent resistance R_{eq} equals the actual resistance, and no error arises in the equivalent series resistance as a result of the associated inductance and capacity (provided $R >> \omega L$ and $R << 1/\omega C$) The resistance will, however, under these conditions, act as though it has an equivalent

series inductance of $-L$, *i.e.*, will have a capacitive phase angle. When, on the other hand, $L = R^2C$, then the equivalent reactance associated with the resistance becomes zero, but the resistance between terminals drops as the frequency increases according to the equation $R_{eq} = R[1 + \omega^2 LC]$. When $L > R^2C$, as tends to be the case with low resistances, an inductive impedance is obtained, whereas if $L < R^2C$, capacitive effects predominate.

The amount by which the power-factor angle of a resistor departs from unity is termed the *phase angle*. When the phase angle is small, its value is to a good approximation given by the equation

$$\text{Phase angle, radians} = \frac{\omega(L - R^2C)}{R} \tag{11}$$

The merit of a resistance unit from the point of view of freedom from reactance effects is also sometimes expressed in terms of the *time constant* ($= L_{eq}/R_{eq}$), which may be as low as 10^{-8} when care is taken to minimize capacitive and inductive effects.

Phase-angle Characteristics of Commercial Wire-wound Resistors.—The ordinary wire-wound resistors, rheostats, and potentiometers used in commercial equipment normally have reasonably low phase angles at audio frequencies, even though designed on the basis of current-carrying capacity. Some typical values are given in Table 6. Although the phase angle varies widely according to the method of construction, it is so small at audio frequencies that the reactances associated with the resistance have negligible effect upon the impedance. These commercial resistance devices are satisfactory for such applications as voltmeter multipliers, attenuators, and voltage dividers intended for audio-frequency service. It will be noted that when the resistance is low, the impedance tends to be inductive, whereas high-resistance units have a capacitive phase angle.

Characteristics of Composition Resistors at High Frequencies.—Composition resistors, because of their small physical size and short current paths in proportion to resistance, have negligible series inductance. The principal reactive effect is accordingly

TABLE 6.—PHASE ANGLES OF TYPICAL COMMERCIAL WIRE-WOUND RESISTORS

Description	Resistance, ohms	Phase angle at 1,000 cycles
Wire-wound potentiometer	200	6′ lag
Wire-wound potentiometer	360	8′ lag
Wire-wound potentiometer	27	9′ lag
Wire-wound potentiometer	50,000	24′ lead
50-watt vitreous enamel tube	100	14′ lag
50-watt vitreous enamel tube	10	26′ lag
200-watt vitreous enamel tube	100,000	2°54′ lead
1-watt wire wound	18,000	6′ lead
1-watt wire wound	3,500	1′ lag
Service test box	50,000	1°40′ lead

produced by the electrostatic field existing between the terminal electrodes and between electrodes and the resistance element itself. Approximate analysis indicates that the effects produced by these electrostatic fields associated with the resistor are equivalent to a lumped shunting capacity that increases with the length and diameter of the resistance unit.[1]

[1] G. W. O. Howe, The Behavior of High Resistances at High Frequencies, *Wireless Eng.*, Vol. 12, p 291, June, 1935.

The effective resistance of a composition resistor, considered independently of the equivalent shunting capacity, is influenced by two factors. Skin effect tends to make the resistance increase with frequency, whereas the capacity between conducting granules separated by the inert material in the mix tends to cause the resistance to drop with increasing frequency.[1] With carbon resistors, the second effect is far greater than the first, and the effective resistance falls off very greatly at high frequencies. The extent of this falling off depends upon the composition of the resistor, and in general tends to be greatest with units having high resistance. The behavior of a particular carbon resistor at high frequencies is shown in Fig. 9.

Composition resistors of the filament or metalized type normally have a resistance that is much less affected by frequency than carbon resistors. This is because, first, the conductor is in the form of a thin tubular film so that skin effect is small, and,

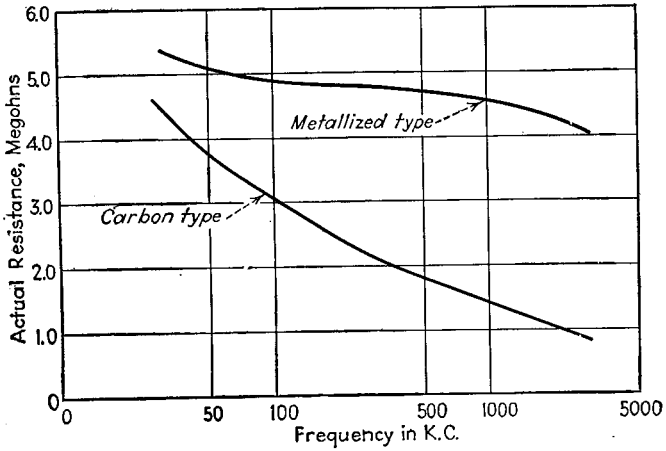


FIG. 9.—Effect of frequency on the resistance of typical metalized and carbon resistors, showing the superior performance of the former type and the very pronounced falling off in resistance with increase in frequency in the case of the carbon resistor.

second, the small conductor cross section permits the use of higher conductivity mix containing proportionately less inert material than carbon resistors. In a typical metalized resistor, the two opposing influences on the resistance are of the same order of magnitude, with one or the other predominating according to the frequency or con-

TABLE 7.—PERCENTAGE CHANGE IN EFFECTIVE RESISTANCE FROM DIRECT-CURRENT VALUES

I.R.C. Ceramic-type Resistors (1-watt Size)

Nominal value, ohms	30 mc	50 mc	100 mc	200 mc	250 mc
10,000	0	0	− 4	− 2	0
20,000	0	....	− 9	−10	−7
50,000	− 7	−11	−16	−10	−6
100,000	− 9	−12	−18	−11	−3
200,000	−17	−17	−23	−17	−4

[1] O. S. Puckle, The Behavior of High Resistances at High Frequencies, *Wireless Eng.*, Vol. 12, p. 303, June, 1935.

struction. Some experimental data on the high-frequency behavior of metalized resistors are shown in Table 7.[1] The excellent behavior at extremely high frequencies is apparent.

Nonreactive Wire-wound Resistors.—Reactance effects associated with wire-wound resistors can be minimized by special winding arrangements.

The inductance of a resistor is determined primarily by the number of turns of wire and the area inclosed by the individual turns. To keep the inductance low, each turn should inclose the minimum possible area, and the wire should have as many ohms per foot of length as possible so that the length required to obtain the desired resistance will be small. In addition, it is desirable that adjacent turns carry current in opposite directions so that the residual inductance of an individual turn is neutralized by the effect of adjacent turns. A low capacitive reactance associated with a

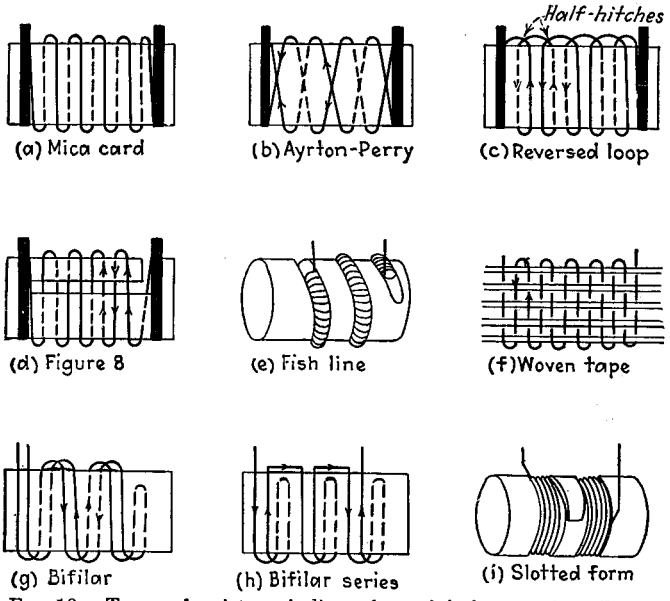


FIG. 10.—Types of resistor windings that minimize reactive effects.

resistor is obtained by arranging the winding in such a way that adjacent turns of wire have a low potential difference between them and are as far apart as possible.

Methods that can be used to minimize the reactive effects associated with a resistor are shown in Fig. 10. The mica-card type of resistor uses a single-layer winding on a thin mica form provided with copper end strips that serve as terminals and reinforcing. A low inductance can be obtained by making the card very thin and using small wire to give a high resistance per turn.[2] The Ayrton-Perry type of resistor is constructed by winding a spaced layer of insulated wire on a thin strip, after which a second wire is wound in the opposite direction between turns of the first winding. The two windings are connected in parallel and thereby produce practically zero resultant magnetic effect. The distributed capacity is low, because adjacent turns have very little potential difference between them. The reversed-loop winding obtains low inductive effects by making a half hitch at the end of each turn and thus reversing the direction of current in adjacent turns. The winding of Fig. 10*d* accom-

[1] Data by J. M. Miller and B. Salzberg, Measurement of Admittances at Ultra-high Frequencies, *RCA Rev.*, Vol. 3, p. 486, April, 1939.

[2] The inductance of windings on cards can be calculated by Eq. (41) of this section.

plishes substantially the same result in a different way. It should be noted that in this winding the wire is passed through the slot only on the *alternate* times it comes to the slot. The fish-line type of resistor consists of a fine resistance wire wound over a silk cord that serves as a core, and the resulting "fish line" is then space-wound on a cylindrical form. The tape resistor is made by weaving the resistance wire into a fabric in which the wire serves as the weft and silk thread functions as the warp.[1] The bifilar winding has negligible inductance, but the capacity is relatively large, because the beginning and end of the resistance are close together. This capacity effect can be minimized to some extent by subdividing the total resistance into several bifilar sections, as shown in Fig. 10*h*. The slotted-form winding will have moderately low capacity because of the subdivision of the winding, and the inductance can be kept moderately low by reversing the direction of the winding in adjacent slots and by using small wire to keep down the required size and the number of turns.

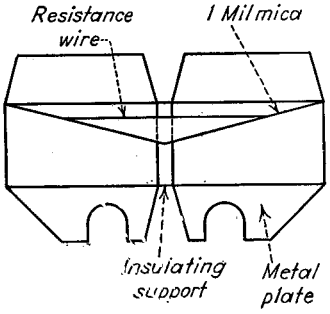


FIG. 11.—Fixed resistor having low reactive effects at very high frequencies.

The mica-card, reversed-loop, and figure-eight types of resistors can be made to have very low phase angles and are the types used in radio-frequency attenuators. The mica-card, fish-line, and woven-tape types of construction are commonly used in decade resistance boxes designed to have a low phase angle at radio frequencies, particularly for the high-resistance units. The Ayrton-Perry winding is also suitable for use at radio frequencies, particularly for resistances up to several thousand ohms. The simple bifilar winding is suitable at radio frequencies only for resistances so low that capacity effects are of no importance. The slotted type of construction is used in very high resistance units where only moderately low phase angle is essential, as in the case of voltmeter multipliers.

A radically different arrangement that is suitable for providing a fixed resistance having low reactance effects at very high frequencies is illustrated in Fig. 11. Here the resistance element is a very short, straight length of fine wire in which the inductance is minimized by mounting this wire against a metal plate with only a thin mica sheet for insulation. The shielding effect of the metal at high frequencies is such as to prevent the penetration of magnetic flux into the metal, with a consequent virtual elimination of the magnetic field around the wire. Capacity is still present, but can be minimized by using a wire of such small diameter and high resistivity that only a short length is required. This type of construction is particularly suitable for resistances of the order of five to two hundred ohms, and has good resistance and reactance characteristics to frequencies exceeding 10 mc. The power-dissipating ability is reasonably large in spite of the small dimensions because of the proximity of the metal plates.[2]

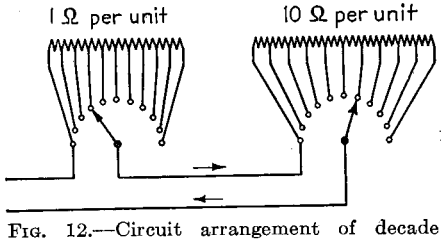


FIG. 12.—Circuit arrangement of decade resistance box.

[1] For a discussion of various types of weaves that may be used, see L. Behr and R. E. Tarpley, Design of Resistors for Precise High-frequency Measurements, *Proc. I.R.E.* Vol. 20, p. 1101, July, 1932.

[2] Further information on this type of resistance is given by D. B. Sinclair, Type 663 Resistor—A Standard for Use at High Frequencies, *Gen. Rad. Exp.*, Vol. 13, p. 6, January, 1939.

Decade Resistance Boxes.—A decade resistance box consists of a number of individual resistance units associated with switches so that the total resistance is adjustable in decade increments. Such boxes are used primarily in laboratory work and in test instruments. The usual circuit arrangement of a decade resistance box is shown in Fig. 12, where ten resistance units are connected in series to form a decade controlled by an eleven-point switch. The use of ten instead of nine resistance units provides an overlap between decades and is extremely desirable.

The phase angle of a resistance box depends upon the characteristics of the individual resistance units and upon the wiring, the shielding, and the capacities introduced by the switches. When it is desired to keep the phase angle small, the resistance units must employ one of the nonreactive types of windings illustrated in Fig. 10, the switch and resistance units should be as compact as possible, and the wiring must be carefully arranged. In particular, the outgoing and return wires should be placed reasonably close together, as illustrated in Fig. 12 in order to minimize inductance effects. Even with the utmost precautions it is always found, however, that the inductance introduced by the connecting wires, and the extra shunting capacity produced by the leads and the switch, cause the phase angle of the resistance box to be much larger than that of the individual resistance units. The characteristics of the individual decade resistance assemblies of a particular type available commercially are given in Table 8.[1]

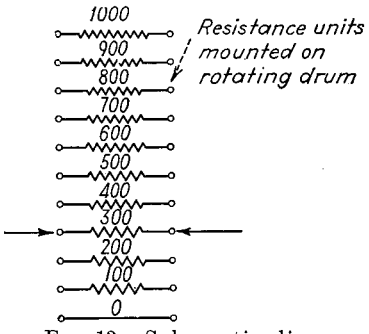


FIG. 13.—Schematic diagram showing decade resistance box which is adjusted by rotating a drum to connect the appropriate resistance unit.

The effect of the leads and switch can be minimized by rotating the resistance elements instead of the switch in the manner indicated schematically in Fig. 13. In this way, it is possible to keep the leads as short as possible, and also to reduce the capacity effects, because the unused resistance units are disconnected. Furthermore, by proper design, the inductance and capacity effects can be made independent of the value of resistance, which is of advantage in some cases.

TABLE 8.—CHARACTERISTICS OF COMMERCIAL DECADE UNITS FOR MAXIMUM SETTING OF DECADE

Decade, ohms per step	Accuracy of adjustment, per cent	Type of winding	Percentage error in resistance			Percentage error in impedance		
			Frequency, kc			Frequency, kc		
			100	1,000	5,000	100	1,000	5,000
0.1	1.0	Bifilar	0.1%	5 %	....	0.7%		
1.0	0.25	Ayrton-Perry	0	1	....	0.2		
10	0.1	Ayrton-Perry	0	0.5	11%	0	2 %	
100	0.1	Ayrton-Perry	0	0.3	4	0	0.3	5%
1,000	0.1	Unifilar on mica	− 0.1	−11	....	− 0.1	−6	
10,000	0.1	Unifilar on mica	−12	.....	....	−10		

The best resistance boxes have inclosed switches to protect the contacts from dust and are mounted in electrostatically shielding containers. In this way the resistance

[1] Further and more detailed information on the behavior of decade resistance boxes is given by D. B. Sinclair, Radio-frequency Characteristics of Decade Resistors, *Gen. Rad. Exp.*, Vol. 15, p. 1, December, 1940.

and switch assembly are protected from electrostatic couplings with neighboring objects. The shield may be connected to either terminal of the resistance box or may be left floating. The phase angle will vary depending upon the way in which the shield is handled.[1]

Inductance-compensated Resistors.—In certain types of measurements, particularly in the determination of small inductances at audio frequencies, it is necessary that the equivalent series inductance of a resistance unit be independent of the resistance setting, although at the same time the inductance need not be zero. This result can be obtained by compensating for the inductance of the resistances switched in and out by switching an equivalent inductance in and out of the circuit. An arrangement of this type is illustrated in Fig. 14*a*, where the opposite ends of the switch control, respectively, resistance and inductance windings. The arrangement is so designed that as the switch is rotated the total inductance stays absolutely constant, and the resistance is increased or decreased by the desired increments.

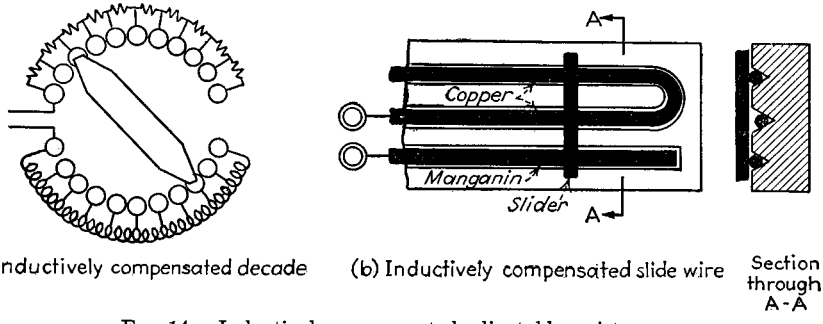


FIG. 14.—Inductively compensated adjustable resistors.

Inductance compensation can also be used with slide wires, a typical arrangement being shown in Fig. 14*b*. Here the moving slider causes a length of manganin wire to be replaced by a length of copper in an equivalent position, so that the resulting inductance is constant at all times.

Load Resistors for Absorbing Radio-frequency Power.—Resistors used as radio-frequency loads, *i.e.*, as dummy antennas, present a particularly difficult problem, since here one desires a nonreactive unit capable of dissipating appreciable wattage. When the power is of the order of fifty watts or less, several satisfactory arrangements are available. One consists of a bifilar resistance element supported on mica and mounted in a glass bulb filled with inert gas, preferably hydrogen.[2] Another arrangement consists of a mica-card type of unit mounted between two large aluminum castings that are for the purpose of conducting away the heat. Both these arrangements give excellent phase-angle characteristics.

Where larger amounts of power are to be handled, various expedients are used. In most of these the reactance is eliminated by tuning, and the dissipated power is evaluated by a calorimetric or photometric method, or by measuring the equivalent circuit resistance at the frequency involved. Another possibility is to use a metalized

[1] For a detailed discussion of the problem of shielding resistance units, see J. G. Ferguson, Shielding in High Frequency Measurements, *Trans. A.I.E.E.*, Vol. 48, p. 1286, October, 1929.

[2] The use of hydrogen as a cooling medium is very effective because of the high heat conductivity of this gas. Thin wires in hydrogen will dissipate as much as 1,000 times as much heat at a given temperature as in a vacuum. Furthermore, with small wires, the heat dissipated at a given temperature does not change greatly with wire diameter, thus permitting a wide change in resistance without changing dissipative ability and wire length. Detailed information on the theory and design of resistors of this type is given by E. G. Linder, The Use of Gas-filled Lamps as High Dissipation, High-frequency Resistors Especially for Power Measurements, *RCA Rev.*, Vol. 4, p. 83, July, 1939.

type of resistor immersed in cooling water.[1] In this way the rating can be increased to 50 to 80 times that for air, and is of the order of 100 watts per square inch of surface. The power being dissipated can be determined from the rate of flow and temperature rise of the cooling water. By making the resistor the central conductor of a concentric line shorted at the receiving end, the reactance at the input terminals can by suitable design proportions be made zero.

INDUCTANCE AND MUTUAL INDUCTANCE

7. Self-inductance.—Self-inductance is that property of electrical circuits which causes them to oppose a change in the current flowing through them. The opposition to the change in current manifests itself in the form of a back emf that is developed when the current is changed. A circuit is said to have an inductance of one henry if it develops a back emf of one volt when the current through it changes at the rate of one ampere per second.

By virtue of the way in which the back emf in a circuit is related to the time rate of change of the magnetic field, inductance is related to the number of flux linkages per unit current in the circuit. A flux linkage represents one flux line encircling the circuit current once. Flux linkages may be fractional as well as integral. The relation between inductance, total flux linkages, and current is given by

$$\text{Inductance } L, \text{ henrys} = \frac{\text{flux linkages}}{\left\{\begin{matrix}\text{current producing}\\ \text{the flux, amperes}\end{matrix}\right\}} \times 10^{-8} \tag{12}$$

Inductance is also related to the energy stored in the magnetic field by the relation

$$L = 2\frac{W}{I^2} \text{ henrys} \tag{13}$$

in which W is the energy stored in the field in watt-seconds.

Units and Dimensions.—The practical unit of inductance is the *henry.* This is such a large unit of inductance that it is often more convenient to express inductance in terms of millihenrys (10^{-3} henrys) or microhenrys (10^{-6} henrys).

Inductance has the dimensions of length. It will be found that in all inductance formulas the inductance is proportional to a linear dimension and to the square of the number of turns when the shape of the inductance is kept constant.

Inductance Formulas.—For various geometrical configurations it is possible to calculate the number of flux lines set up or the energy stored in the magnetic field for an assumed current. The inductance formulas that follow have been calculated from Eqs. (12) or (13) in this manner. The collection of inductance formulas given here will be found complete enough to satisfy the ordinary needs of engineers and physicists. In general the formulas are accurate to better than 0.5 per cent. For more extensive collections, the reader is referred to treatises on the subject.[2]

Because of the way in which the current distribution within a wire changes with frequency, the inductance of a circuit changes somewhat with frequency. Of most interest are the values that the inductance assumes at very low frequencies when the current distribution through the wire cross section is uniform, and the limiting value that the inductance assumes as the frequency approaches infinity. This latter inductance is always less than the low-frequency inductance, since the redistribution

[1] See G. H. Brown and J. W. Conklin, Water-cooled Resistors for Ultrahigh Frequencies, *Electronics*, Vol. 14, p. 24, April, 1941.

[2] E. B. Rosa and F. W. Grover, Formulas and Tables for the Calculation of Mutual and Self Inductance, *Bur. Standards Bull.*, Vol. 8, No. 1, pp. 1–237, Jan. 1, 1912.

Radio Instruments and Measurements, *Bur. Standards Circ.* 74, pp. 242–296.

J. Hak, "Eisenlose Drosselspulen [Air-core Inductances]," Verlag K. F. Koehler, Leipzig, 1938.

of current at high frequencies is always of such a character as to reduce the flux linkages. The difference between the low-frequency and high-frequency inductance is usually not great, however.

8. Self-inductance of Straight Conductors.—The self-inductance of an unclosed circuit is here taken as the inductance of the unclosed circuit considered as a part of a closed circuit. The total self-inductance of the closed circuit is equal to the sum of the self-inductances of all of its parts plus the sum of the mutual inductances of each one of the component parts on every other part.

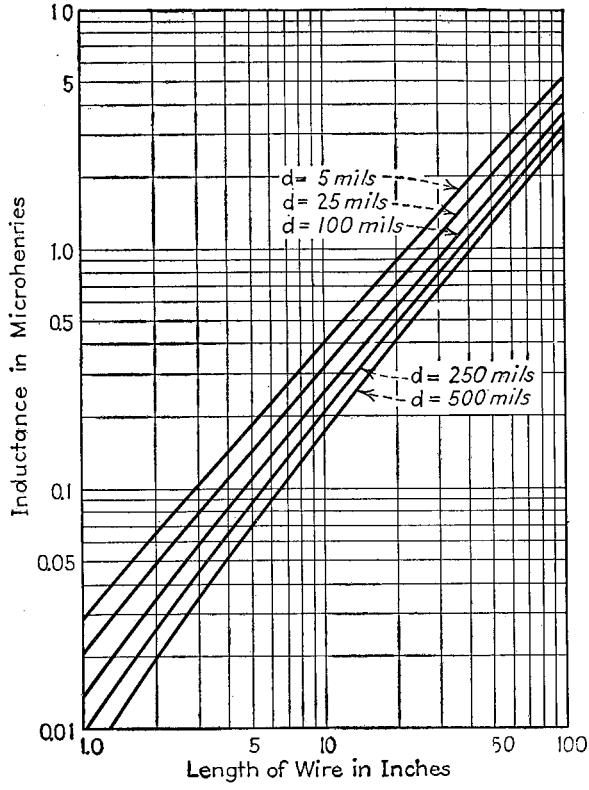


FIG. 15.—Self-inductance of a straight round wire at high frequencies.

Straight Round Wires.—The self-inductance of a round straight wire of length l, diameter d, and permeability μ is

$$L_0 = 0.00508l\left(2.303 \log_{10}\frac{4l}{d} - 1 + \frac{\mu}{4}\right) \text{ microhenrys} \qquad (14)$$

where L_0 is the low-frequency inductance and the dimensions are in inches. For wires of nonmagnetic materials such as copper, this becomes

$$L_0 = 0.00508l\left(2.303 \log_{10}\frac{4l}{d} - 0.75\right) \text{ microhenrys} \qquad (15)$$

For such very short lengths of wire that $l < 100d$, a term $d/2l$ should be inserted within the parentheses above.

As frequency increases, the inductance of a straight round wire assumes a limiting value of

$$L = 0.00508l\left(2.303\log_{10}\frac{4l}{d} - 1\right) \text{ microhenrys} \tag{16}$$

Limiting values of inductance for various wire lengths and diameters are shown in

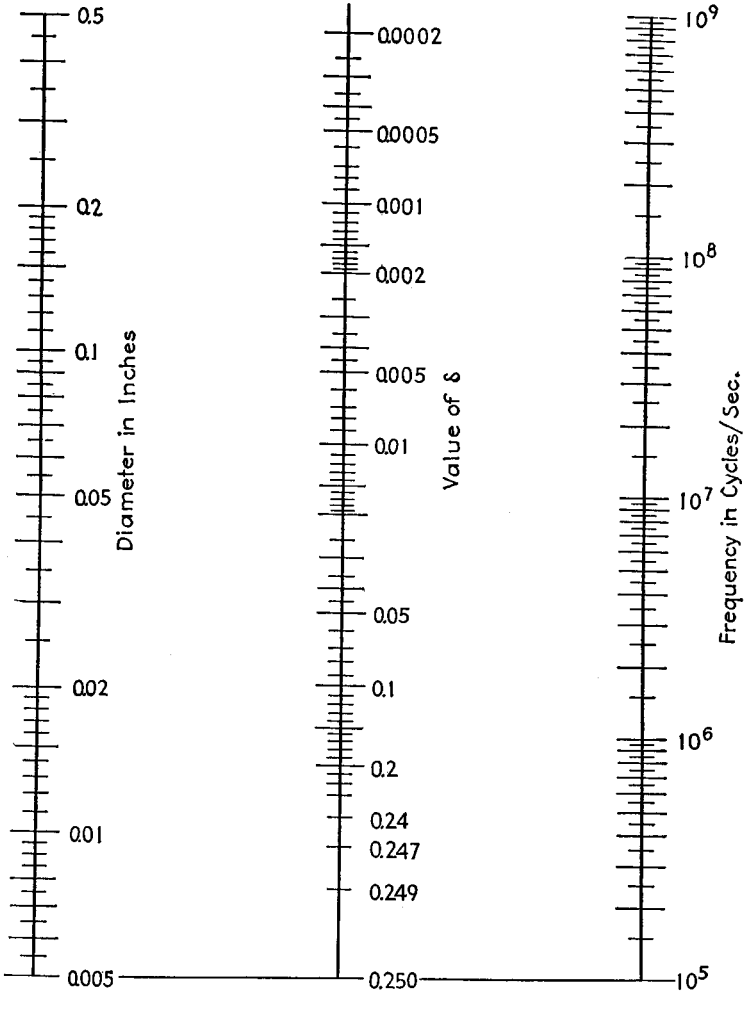


FIG. 16.—Skin-effect correction factor δ as a function of wire diameter and frequency.

Fig. 15. For intermediate frequencies, the inductance is

$$L = 0.00508l\left(2.303\log_{10}\frac{4l}{d} - 1 + \mu\delta\right) \text{ microhenrys} \tag{17}$$

where δ is a skin-effect factor that depends upon the same quantity x defined in Eq. (3).

$$x = \left(d\sqrt{\frac{\mu f}{\rho}}\right) \times \text{constant} \tag{18}$$

where f is the frequency, ρ is the volume resistivity, and μ the magnetic permeability. Values of δ for copper wire, as determined by d and f, may be obtained from the nomogram of Fig. 16. To obtain from this chart the value of δ for materials other than copper, an equivalent value of frequency, $f' = f\rho_c/\rho$ should be used instead of the actual value of frequency f, where ρ_c is the resistivity of copper. The factor δ approaches the value of 0.25 for low frequencies. For high frequencies, δ is inversely proportional to the square root of frequency, and approaches the value zero as the frequency becomes very large.

The changes of inductance with frequency are comparatively small. Skin effect at very high frequencies decreases inductance from about 6 per cent for short wires to about 2 per cent for long wires.

Return Circuit of Two Parallel Round Wires.—The self-inductance of a pair of parallel round wires carrying current in opposite directions, each having a length l, a diameter d, and spaced a distance D from one another (dimensions in inches), is

$$L = 0.01016l\left(2.303\log_{10}\frac{2D}{d} - \frac{D}{l} + \mu\delta\right) \text{ microhenrys} \tag{19}$$

This neglects the inductance of the connecting link between the ends of the wires. As before, δ is a skin-effect factor that may be determined from Fig. 16.

Grounded Horizontal Round Wire.—The self-inductance of a wire of diameter d, length l, height above ground h, placed horizontally, with the earth used as a return circuit, is

$$\begin{aligned} L = {} & 0.001170l\left[\log_{10}\left(\frac{l + \sqrt{l^2 + d^2/4}}{l + \sqrt{l^2 + 4h^2}}\right) + \log_{10}\frac{4h}{d}\right] \\ & + 0.00508\left[\sqrt{l^2 + 4h^2} - \sqrt{l^2 + d^2/4} + \mu l\delta - 2h + \frac{d}{2}\right] \text{ microhenrys} \end{aligned} \tag{20}$$

The dimensions are in inches, and δ has the same meaning as above. In most applications the diameter is very small compared with the length of the wire, and the preceding expression may be simplified to the following forms:

For $2h/l \leqq 1$:

$$L = 0.00508l\left(2.303\log_{10}\frac{4h}{d} - P + \mu\delta\right) \text{ microhenrys} \tag{21}$$

For $l/2h \leqq 1$:

$$L = 0.00508l\left(2.303\log_{10}\frac{4l}{d} - Q + \mu\delta\right) \text{ microhenrys} \tag{22}$$

Here P and Q are functions of $2h/l$ and $l/2h$ respectively, which may be determined with sufficient accuracy from the values in Table 9.

Further simplification is possible when the length is much greater than the height. The inductance per *foot* for this case is

$$L = 0.1404\log_{10}\frac{4h}{d} + 0.061\,\mu\delta \text{ microhenrys per foot} \tag{22a}$$

Grounded Round Wires in Parallel.—The formulas given here are for n parallel wires evenly spaced in a plane. The resultant expression depends upon the self-inductance of each wire and the mutual inductances between the wires. Exact expressions for this case are quite complicated. The following are approximate expressions accurate to about 1 per cent.

$$L = l\left[\frac{L_1 + (n-1)M_1}{n} - 0.00254k\right] \text{ microhenrys} \tag{23}$$

where l is the length in inches, k is a function of the number of wires n, as tabulated in Table 10, L_1 is the inductance per inch of a single grounded wire, as determined from the Eqs. (21) and (22) in the previous section, and M_1 is the mutual inductance per unit length between two adjacent wires, as determined from the following:

For $2h/l \leqq 1$:

$$M_1 = \frac{M}{l} = 0.00508\left(2.303 \log_{10}\frac{2h}{D} - P + \frac{D}{l}\right) \tag{24}$$

For $1/2h \leqq 1$:

$$M_1 = \frac{M}{l} = 0.00508\left(2.303 \log_{10}\frac{2l}{D} - Q + \frac{D}{l}\right) \tag{25}$$

in which P and Q may be found from the values in Table 9.

TABLE 9.—VALUES OF P AND Q IN EQS. (21), (22), (24), AND (25)

$2h/l$	P	$l/2h$	Q
0.0	0.0000	0.0	1.0000
0.1	0.0975	0.1	1.0499
0.2	0.1900	0.2	1.0997
0.3	0.2778	0.3	1.1489
0.4	0.3608	0.4	1.1975
0.5	0.4393	0.5	1.2452
0.6	0.5136	0.6	1.2918
0.7	0.5840	0.7	1.3373
0.8	0.6507	0.8	1.3819
0.9	0.7139	0.9	1.4251
1.0	0.7740	1.0	1.4672

Straight Rectangular Bar.—The inductance at low frequencies of a straight rectangular bar of length l, width b, and thickness c inches is

$$L_0 = 0.00508l\left(2.303 \log_{10}\frac{2l}{b+c} + 0.5 + 0.2235\frac{b+c}{l}\right) \tag{26}$$

The last term may be neglected for $l > 50(b + c)$.

TABLE 10.—VALUES OF k IN EQ. (23)

n	k
2	0
3	0.308
4	0.621
5	0.906
6	1.18
7	1.43
8	1.66
9	1.86
10	2.05
11	2.22
12	2.37
13	2.51
14	2.63
15	2.74
16	2.85
17	2.95
18	3.04
19	3.14
20	3.24

Inductance of Concentric Cable.—The low-frequency inductance is

$$L_0 = 0.140 \log_{10} \frac{r_2}{r_1} + 0.015 \text{ microhenrys per foot} \tag{27}$$

where r_1 is the radius of the outside of the inner conductor and r_2 is the radius of the inner side of the outer conductor.

Return Circuit of Two Parallel Rectangular Bars.—The low-frequency inductance of two parallel rectangular bars of length l, width b, breadth c, and center spacing D inches is

$$L_0 = 0.01016l \left(2.303 \log_{10} \frac{D}{b+c} + 1.5 - \frac{D}{l} + 0.2235 \frac{b+c}{l}\right) \text{ microhenrys} \tag{28}$$

The accuracy can be increased slightly by taking D as the geometrical mean distance between the wires.[1]

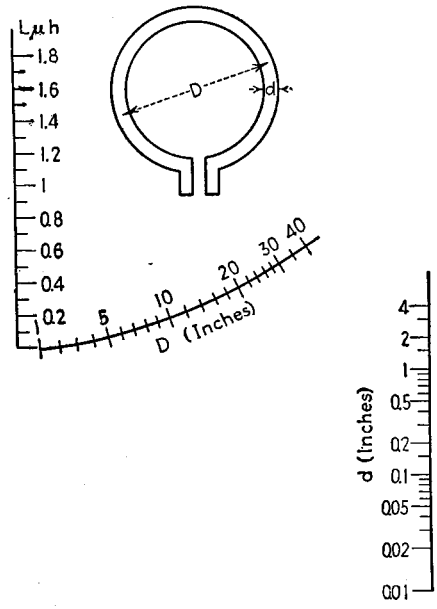


FIG. 17.—High-frequency inductance of a circular loop.

9. Self-inductance of Single-turn Loops. *Circular Ring of Circular Section.* The inductance of a circular loop of diameter D of wire of diameter d inches is

$$L = 0.01595D \left(2.303 \log_{10} \frac{8D}{d} - 2 + \mu\delta\right) \text{ microhenrys} \tag{29}$$

In Fig. 17 is given a nomogram from which the limiting value of inductance at high frequencies may be obtained for any loop and wire diameter. The nomogram may be extended by making use of the fact that inductance is proportional to linear dimension.

Circular Ring of Tubular Section.—The low-frequency inductance of a tube of inner diameter d_1 and outer diameter d_2, bent into a circle of mean diameter D inches, is

$$L_0 = 0.01595D \left[2.303 \log_{10} \frac{8D}{d_2} - 1.75 - \frac{d_1^2}{2(d_2^2 - d_1^2)} + 2.303 \frac{d_1^4}{(d_2^2 - d_1^2)^2} \log_{10} \frac{d_2}{d_1}\right] \text{ microhenrys} \tag{30}$$

For infinite frequencies this assumes the value

$$L_\infty = 0.01595D \left(2.303 \log_{10} \frac{8D}{d_2} - 2\right) \text{ microhenrys} \tag{31}$$

Square of Round Wire.—The inductance of a square of side s of wire of diameter d inches is

$$L = 0.02032s \left(2.303 \log_{10} \frac{2s}{d} + \frac{d}{2s} - 0.774 + \mu\delta\right) \text{ microhenrys} \tag{32}$$

[1] For a discussion of geometrical mean distance, see E. B. Rosa and F. W. Grover, *Bur. Standards Bull.*, Vol. 8, p. 166, 1912. For parallel rectangular bars, the geometrical mean distance is slightly greater than the center distance. The difference is inappreciable when the cross-sectional dimensions are small compared to the center spacing.

Rectangle of Round Wire.—The inductance of a rectangle of sides s_1 and s_2 and diagonal g of wire of diameter d inches is

$$L = 0.02339 \left[(s_1 + s_2) \log_{10} \frac{4s_1s_2}{d} - s_1 \log_{10} (s_1 + g) - s_2 \log_{10} (s_2 + g) \right] + 0.01016 \left[\mu\delta(s_1 + s_2) + 2\left(g + \frac{d}{2}\right) - 2(s_1 + s_2) \right] \text{microhenrys} \quad (33)$$

Rectangle of Rectangular Wire.—The low-frequency inductance of a rectangle of sides s_1 and s_2 and diagonal g of rectangular wire of thickness b and width c, as shown in Fig. 18, with dimensions in inches, is

$$L_0 = 0.02339 \left[(s_1 + s_2) \log_{10} \frac{2s_1s_2}{b + c} - s_1 \log_{10} (s_1 + g) - s_2 \log_{10} (s_2 + g) \right] + 0.01016 \left[2g - \frac{s_1 + s_2}{2} + 0.447(b + c) \right] \text{microhenrys} \quad (34)$$

Simplified Loop Formula for Regular Figures.—Examination of the foregoing loop-inductance equations shows them to be of the same general form regardless of the shape. If l is the total perimeter of the figure, then the inductance at high frequencies can be written as

$$L_\infty = 0.00508l \left(2.303 \log_{10} \frac{4l}{d} - \theta \right) \text{microhenrys} \quad (35)$$

The dimensions are in inches. The quantity θ is a constant that depends upon the shape of the loop.[1] Values of θ for the commonest regular figures are given in Table 11.

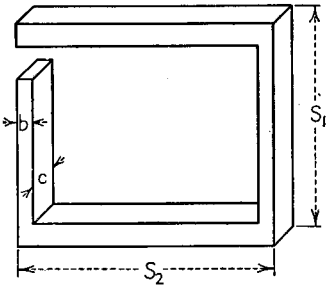


FIG. 18.—Rectangular loop of rectangular cross section.

The above formula is fairly accurate when the perimeter of the figure is large compared to the wire diameter. For ordinary loop shapes it is correct to within 0.5 per cent.

TABLE 11.—VALUES OF θ FOR USE IN EQ. (35)

Circle	$\theta = 2.451$
Regular octagon	$\theta = 2.561$
Regular hexagon	$\theta = 2.636$
Regular pentagon	$\theta = 2.712$
Square	$\theta = 2.853$
Equilateral triangle	$\theta = 3.197$
Isosceles right-angled triangle	$\theta = 3.332$

10. Inductance of Single-layer Coils. *Single-layer Solenoid of Round Wire.*—The low-frequency inductance of a single-layer solenoid is given by[2]

$$L_0 = Fn^2d \text{ microhenrys} \quad (36)$$

[1] For the development of the preceding expression, see V. I. Bashenoff, Inductance of Oddly Shaped Loops, *Proc. I.R.E.*, Vol. 15, p. 1013, December, 1927. For an extension of Bashenoff's work, see P. L. Kalantaroff and V. I. Worobieff, Self Inductance of Polygonal Circuits, *Proc. I.R.E.*, Vol. 24, p. 1585, December, 1936, which tabulates constants for 51 polygonal shapes, including many with reentrant angles.

[2] This is a rearrangement of the well-known Nagaoka formula, which is frequently written in the form $L_0 = \dfrac{2.54 \times 0.03948 \left(\frac{d}{2}\right)^2 n^2}{l} K$, where K is a constant depending upon the ratio of diameter d to length l. Factor F in Eq. (36) is $\left(2.54 \times 0.03948 \times \frac{d}{4l} K\right)$, which is a function only of d/l. See E. B. Rosa and F. W. Grover, *Bur. Standards Bull.*, Vol. 8, p. 119, 1912.

TABLE 12.—VALUES OF F FOR USE IN EQ. (36)

Diameter / Length	F	Difference	Diameter / Length	F	Difference
0.00	0.0000000	+0.0004972	3.00	0.03228	+0.00049
0.02	0.0004972	+0.0004888	3.10	0.03277	+0.00048
0.04	0.000986	+0.000481	3.20	0.03325	+0.00047
0.06	0.001467	+0.000472	3.30	0.03372	+0.00045
0.08	0.001939	+0.000465	3.40	0.03417	+0.00044
0.10	0.002404	+0.000457	3.50	0.03461	+0.00042
0.12	0.002861	+0.000449	3.60	0.03503	+0.00042
0.14	0.003310	+0.000442	3.70	0.03545	+0.00041
0.16	0.003752	+0.000434	3.80	0.03586	+0.00040
0.18	0.004186	+0.000428	3.90	0.03626	+0.00039
0.20	0.004614	+0.000420	4.00	0.03665	+0.00038
0.22	0.005034	+0.000413	4.10	0.03703	+0.00037
0.24	0.005447	+0.000407	4.20	0.03740	+0.00036
0.26	0.005854	+0.000400	4.30	0.03776	+0.00035
0.28	0.006254	+0.000393	4.40	0.03811	+0.00035
0.30	0.006647	+0.000387	4.50	0.03846	+0.00034
0.32	0.007034	+0.000381	4.60	0.03880	+0.00033
0.34	0.007415	+0.000375	4.70	0.03913	+0.00033
0.36	0.007790	+0.000369	4.80	0.03946	+0.00032
0.38	0.008159	+0.000363	4.90	0.03978	+0.00031
0.40	0.008522	+0.000358	5.00	0.04009	+0.00061
0.42	0.008883	+0.000352	5.20	0.04070	+0.00058
0.44	0.009232	+0.000346	5.40	0.04128	+0.00057
0.46	0.009578	+0.000341	5.60	0.04185	+0.00055
0.48	0.009919	+0.000336	5.80	0.04240	+0.00053
0.50	0.01026	+0.00081	6.00	0.04293	+0.00051
0.55	0.01107	+0.00079	6.20	0.04344	+0.00050
0.60	0.01186	+0.00076	6.40	0.04394	+0.00049
0.65	0.01262	+0.00073	6.60	0.04443	+0.00046
0.70	0.01335	+0.00071	6.80	0.04489	+0.00045
0.75	0.01406	+0.00068	7.00	0.04534	+0.00044
0.80	0.01474	+0.00066	7.20	0.04578	+0.00043
0.85	0.01540	+0.00064	7.40	0.04621	+0.00043
0.90	0.01604	+0.00062	7.60	0.04664	+0.00041
0.95	0.01666	+0.00060	7.80	0.04705	+0.00040
1.00	0.01726	+0.00058	8.00	0.04745	+0.00095
1.05	0.01784	+0.00056	8.50	0.04840	+0.00091
1.10	0.01840	+0.00055	9.00	0.04931	+0.00085
1.15	0.01895	+0.00053	9.50	0.05016	+0.00081
1.20	0.01948	+0.00052	10.00	0.05097	
1.25	0.02000	+0.00050	10.0	0.05097	+0.00151
1.30	0.02050	+0.00049	11.0	0.05248	+0.00138
1.35	0.02099	+0.00048	12.0	0.05386	+0.00128
1.40	0.02147	+0.00046	13.0	0.05514	+0.00119
1.45	0.02193	+0.00045	14.0	0.05633	+0.00110
1.50	0.02238	+0.00044	15.0	0.05743	+0.00102
1.55	0.02282	+0.00043	16.0	0.05845	+0.0095
1.60	0.02325	+0.00042	17.0	0.05940	+0.0090
1.65	0.02367	+0.00041	18.0	0.06030	+0.0086
1.70	0.02408	+0.00040	19.0	0.06116	+0.0081
1.75	0.02448	+0.00039	20.0	0.06197	+0.00151
1.80	0.02487	+0.00038	22.0	0.06348	+0.00138
1.85	0.02525	+0.00037	24.0	0.06486	+0.00127
1.90	0.02562	+0.00037	26.0	0.06613	+0.00119
1.95	0.02599	+0.00036	28.0	0.06732	+0.00112
2.00	0.02635	+0.00069	30.0	0.0684	+0.0024
2.10	0.02704	+0.00067	35.0	0.0708	+0.0022
2.20	0.02771	+0.00065	40.0	0.0730	+0.0020
2.30	0.02836	+0.00062	45.0	0.0750	+0.0016
2.40	0.02898	+0.00059	50.0	0.0766	+0.0028
2.50	0.02957	+0.00058	60.0	0.0794	+0.0025
2.60	0.03015	+0.00056	70.0	0.0819	+0.0022
2.70	0.03071	+0.00054	80.0	0.0841	+0.0019
2.80	0.03125	+0.00052	90.0	0.0850	+0.0017
2.90	0.03177	+0.00051	100.0	0.0877	

where n is the number of turns, F a quantity that depends upon the ratio of the diameter to the length of the coil, and d is the diameter of the coil in inches. The quantity F is tabulated for a wide range of values of the ratio of (diameter/length) in Table 12 and plotted in Fig. 19.

Where great accuracy is required, a correction factor may be applied to Eq. (36) to take account of the fact that the coil is wound of spaced round wires rather than with a uniform current sheet.[1] This correction rarely exceeds 0.5 per cent, is greatest for widely spaced turns, and increases with the number of turns.

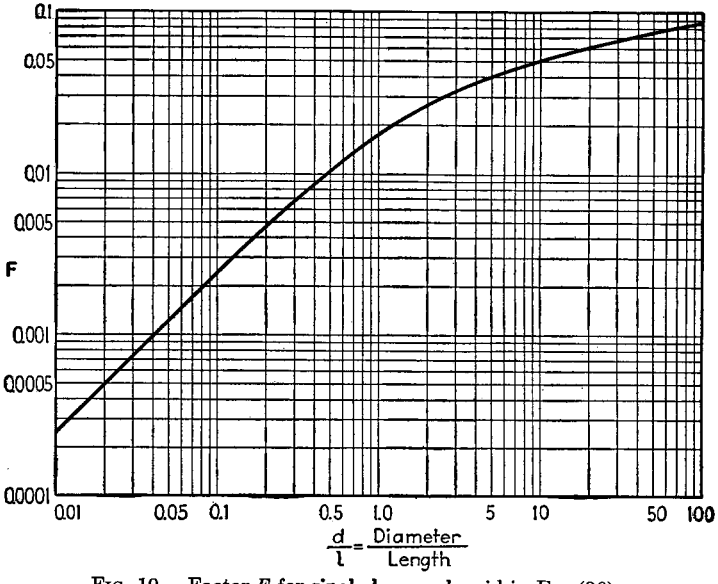


FIG. 19.—Factor F for single-layer solenoid in Eq. (36).

A simple *approximate* formula for the low-frequency inductance of a single-layer solenoid is[2]

$$L_0 = \frac{r^2n^2}{9r + 10l} \text{ microhenrys} \tag{37}$$

where r is the radius of the coil and l its length in inches. This formula is accurate to within one per cent for $l > 0.8r$, *i.e.*, if the coil is not too short.

Single-layer Solenoid of Wire of Rectangular Cross Section.—The low-frequency inductance is most conveniently given by

$$L_0 = Fn^2d - \frac{0.01596n^2dc}{l}(0.693 + B_s) \tag{38}$$

where n = number of turns of the winding.

d = diameter of the coil to center of winding, in.

l = length of the coil, in.

c = radial thickness of the wire strip, in.

F = function of $\frac{\text{diameter}}{\text{length}}$ given in Table 12 and Fig. 19.

B_s = function of $\frac{\text{length}}{\text{radial thickness } c \text{ of wire}}$ given in Table 13.

[1] Radio Instruments and Measurements, *Bur. Standards Circ.* 74, p. 253.

[2] H. A. Wheeler, Simple Inductance Formulas for Radio Coils, *Proc. I.R.E.*, Vol. 16, p. 1398, October, 1928.

It will be recognized that the first term gives the inductance of the equivalent solenoid of round wire and that the second is a correction for the fact that the wires are not round.

TABLE 13.—VALUES OF B_s FOR USE IN EQ. (38)

l/c	B_s	l/c	B_s	l/c	B_s
1	0.0000	11	0.2844	21	0.3116
2	0.1202	12	0.2888	22	0.3131
3	0.1753	13	0.2927	23	0.3145
4	0.2076	14	0.2961	24	0.3157
5	0.2292	15	0.2991	25	0.3169
6	0.2446	16	0.3017	26	0.3180
7	0.2563	17	0.3041	27	0.3190
8	0.2656	18	0.3062	28	0.3200
9	0.2730	19	0.3082	29	0.3209
10	0.2792	20	0.3099	30	0.3218

Single-layer Polygonal Coil.—The inductance of a single-layer polygonal coil can be calculated from the formula for a single-layer solenoid, Eq. (36), by assuming that the polygonal coil is equivalent to a single-layer solenoid whose diameter is equal to the mean of the radii of the circumscribed and inscribed circles of the polygon. Thus if a solenoid of N sides has a circumscribed diameter d, then the equivalent solenoid diameter is

$$d_0 = d \cos^2 \frac{\pi}{2N} \tag{39}$$

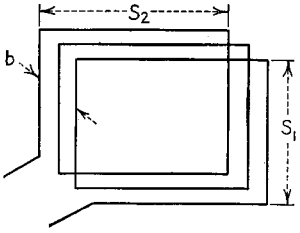


FIG. 20.—Single-layer rectangular coil.

The use of this equivalent diameter generally gives results accurate to within one per cent. The results become more accurate as the number of sides of the polygon increases.

Single-layer Rectangular Coil.—The low-frequency inductance is

$$L_0 = pn^2(G + H) \text{ microhenrys} \tag{40}$$

where p = perimeter, $2(s_1 + s_2)$.
s_1 = length of short side, in.
s_2 = length of long side, in.
b = axial length of the coil, in.
n = number of turns.
G = factor determined by s_1/s_2 and b/s_2 given in Fig. 21.
H = factor determined by n and $\dfrac{\text{diameter of wire}}{\left\{\begin{matrix}\text{spacing between} \\ \text{adjacent turns}\end{matrix}\right\}}$ and given in Fig. 21.

The foregoing expression is accurate only when the distance between turns is small compared to the short side of the coil.

Single-layer Square Coil.—This is conveniently obtained by setting $s_1 = s_2$ in the formula, Eq. (40), of the previous section.

Winding Very Long Compared with Smallest Cross-sectional Dimension.—The low-frequency inductance of such a winding, irrespective of the shape of the cross section, is

$$L_0 = 0.0319 \frac{An^2}{l} \text{ microhenrys} \tag{41}$$

where A is the area of the cross section in square inches, l is the length of the coil in inches, and n is the number of turns.

Toroidal Coil of Circular Cross Section.—The low-frequency inductance of a toroidal coil of n turns with a winding of circular cross section of diameter d and a

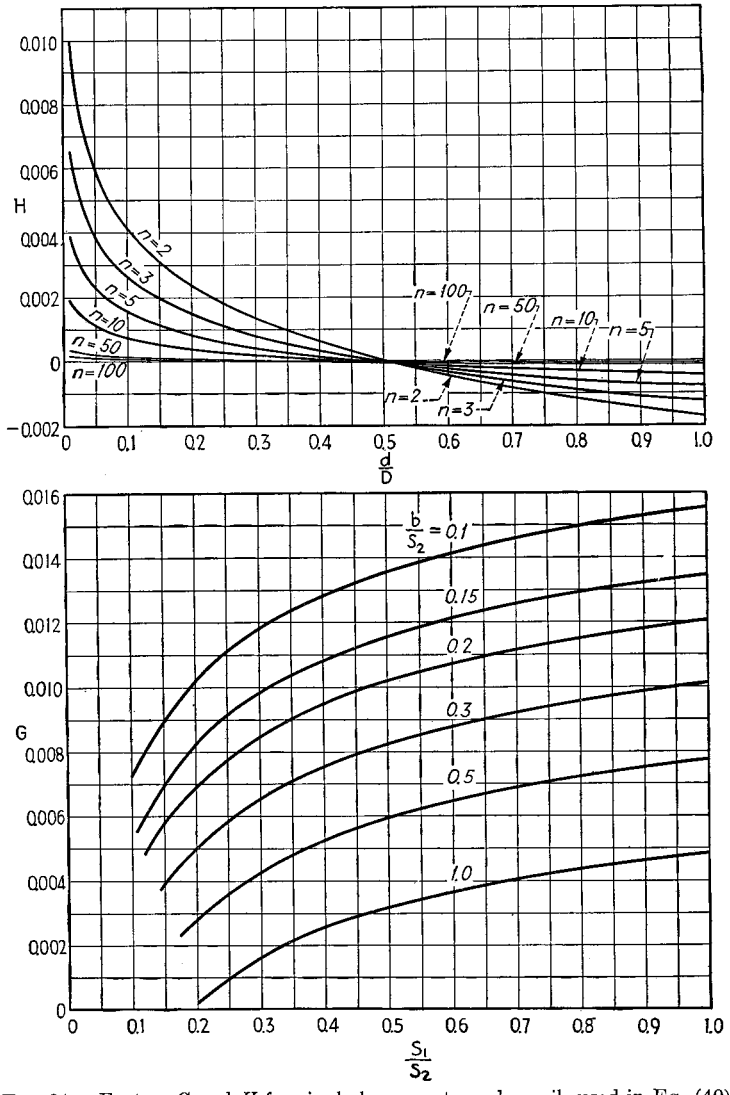


FIG. 21.—Factors G and H for single-layer rectangular coil, used in Eq. (40).

radius of revolution $D/2$ inches, as shown in Fig. 22, is

$$L_0 = 0.01595n^2[D - (D^2 - d^2)^{1/2}] \text{ microhenrys} \tag{42}$$

For toroidal coils in which the turn diameter is much less than the toroid diameter (*i.e.*, $d < 0.1D$), the inductance is *approximately*

$$L_0 = 0.007975 \frac{d^2n^2}{D} \text{ microhenrys} \tag{43}$$

Toroidal Coil of Rectangular Cross Section.—When the axial depth of the winding is h and the inner and outer diameters d_1 and d_2 inches, respectively, as shown in Fig. 23, the low-frequency inductance is

$$L_0 = 0.01170n^2h \log_{10} \frac{d_2}{d_1} \text{ microhenrys} \tag{44}$$

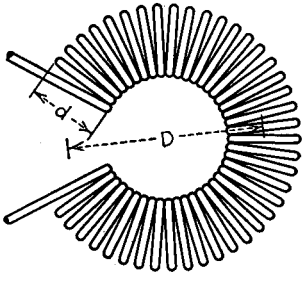


FIG. 22.—Toroidal coil of circular cross section showing notation of Eq. (42).

Flat Spiral.—The low-frequency inductance of a flat spiral of *round wire* of mean radius a and radial depth of winding c inches, as shown in Fig. 24, is conveniently expressed by

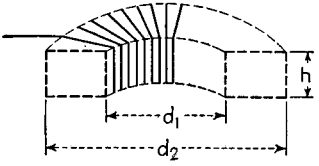


FIG. 23.—Toroidal coil of rectangular cross section illustrating notation of Eq. (44).

$$L_0 = an^2K \text{ microhenrys} \tag{45}$$

in which n is the number of turns and K is a quantity depending upon a/c as given in Fig. 24.

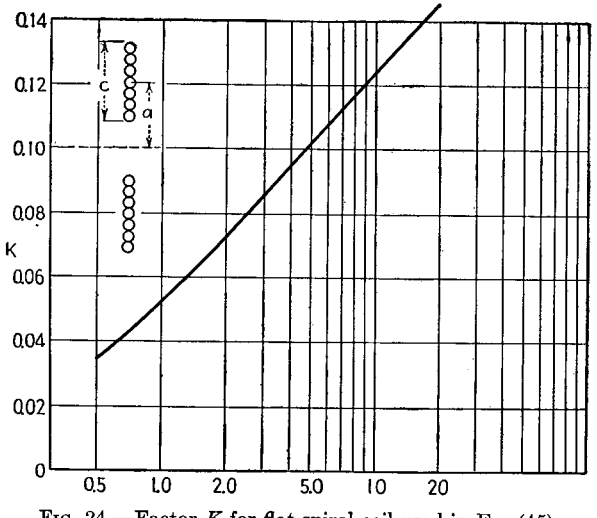


FIG. 24.—Factor K for flat spiral coil used in Eq. (45).

An *approximate* formula for this type of inductance is[1]

$$L_0 = \frac{a^2n^2}{8a + 11c} \text{ microhenrys} \tag{46}$$

This formula is accurate to within 5 per cent for $c > 0.2a$.

[1] See Wheeler, *loc. cit.*

The inductance of a flat spiral of *flat ribbon* is most readily found by using Eq. (51) for a short solenoidal multilayer coil. This gives results that are, in general, small by less than one per cent.

TABLE 14.—VALUES OF CORRECTION TERM A IN EQ. (48)

d/D	A	Difference	d/D	A	Difference
1.00	0.557	−0.051	0.20	−1.053	−0.051
0.95	0.506	−0.054	0.19	−1.104	−0.054
0.90	9.452	−0.057	0.18	−1.158	−0.057
0.85	0.394	−0.061	0.17	−1.215	−0.061
0.80	0.334	−0.065	0.16	−1.276	−0.064
0.75	0.269	−0.069	0.15	−1.340	−0.069
0.70	0.200	−0.074	0.14	−1.409	−0.074
0.65	0.126	−0.080	0.13	−1.483	−0.080
0.60	0.046	−0.087	0.12	−1.563	−0.087
0.55	−0.041	−0.095	0.11	−1.650	−0.096
0.50	−0.136	−0.041	0.10	−1.746	−0.105
0.48	−0.177	−0.043	0.09	−1.851	−0.118
0.46	−0.220	−0.044	0.08	−1.969	−0.133
0.44	−0.264	−0.047	0.07	−2.102	−0.154
0.42	−0.311	−0.048	0.06	−2.256	−0.173
0.40	−0.359	−0.052	0.05	−2.439	−0.223
0.38	−0.411	−0.054	0.04	−2.662	−0.288
0.36	−0.465	−0.057	0.03	−2.950	−0.405
0.34	−0.522	−0.061	0.02	−3.355	−0.693
0.32	−0.583	−0.064	0.01	−4.048	
0.30	−0.647	−0.069			
0.28	−0.716	−0.074			
0.26	−0.790	−0.080			
0.24	−0.870	−0.087			
0.22	−0.957	−0.096			

Flat Rectangular Coil.—The low-frequency inductance of a flat rectangular coil whose average dimensions are s_1 and s_2 with an average diagonal $g = \sqrt{s_1^2 + s_2^2}$ wound of n complete turns of wire of diameter d with a pitch of winding D, as shown in Fig. 25, is

$$L_0 = 0.02339n^2\left[(s_1 + s_2)\log_{10}\frac{2s_1s_2}{nD} - s_1\log_{10}(s_1 + g) - s_2\log_{10}(s_2 + g)\right]$$

$\therefore$

$$+ 0.01016n^2\left(2g - \frac{s_1 + s_2}{2} + 0.447nD\right) - 0.01016n(s_1 + s_2)(A + B) \text{ microhenrys} \tag{47}$$

The dimensions are in inches. In the preceding, A and B are constants depending upon the wire spacing and number of turns, respectively. Values of A corresponding to various values of d/D are given in Table 14. Values of B corresponding to various values of n are given in Table 15.

Flat Square Coil.—If $s = s_1 = s_2$ are set in the previous section, the expression for inductance reduces to

$$L_0 = 0.02032n^2s\left[2.303\log_{10}\frac{s}{nD} + 0.2235\frac{nD}{s} + 0.726\right] - 0.02032ns(A + B) \text{ microhenrys} \tag{48}$$

TABLE 15.—VALUES OF CORRECTION TERM B IN EQ. (48)

Number of turns, n	B	Number of turns, n	B	Number of turns, n	B
1	0.000	15	0.286	80	0.326
2	0.114	20	0.297	90	0.327
3	0.166	25	0.304	100	0.328
4	0.197	30	0.308	150	0.331
5	0.218	35	0.312	200	0.333
6	0.233	40	0.315	300	0.334
7	0.244	45	0.317	400	0.335
8	0.253	50	0.319	500	0.336
9	0.260	60	0.322	700	0.336
10	0.266	70	0.324	1,000	0.336

where s is the average length of side, n is the number of turns, d is the diameter of the wire, D is the distance between turns, and A and B are constants that may be determined from Tables 14 and 15.

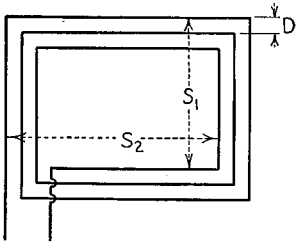


FIG. 25.—Flat rectangular coil.

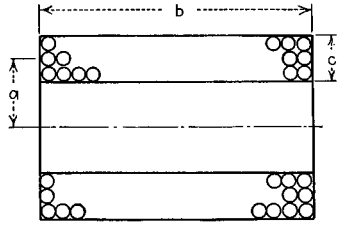


FIG. 26.—Long multiple-layer coil of solenoidal form.

11. Inductance of Multiple-layer Coils. *Long Solenoidal Form.*—The low-frequency inductance of a long multiple-layer coil of solenoidal form, as shown in Fig. 26, where b and c are the axial length and radial thickness of the winding, both measured to the outside of the wires, and a is the mean radius, is given by

$$L_0 = Fn^2d - \frac{0.03193n^2ac}{b}(0.693 + B_s) \text{ microhenrys} \quad (49)$$

The dimensions are in inches. It will be recognized that the first term of this expression is the equation for the inductance of a single-layer solenoid given in Eq. (36), F being the factor given in Table 12. The second term is a correction for the thickness of the winding with the quantity B_s the same quantity given in Table 13, except that b/c is substituted for l/c in the table. The foregoing expression is generally accurate to within 0.5 per cent.

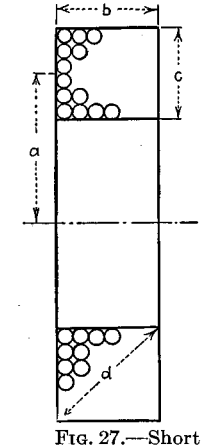


FIG. 27.—Short multiple-layer coil of solenoidal form.

If the insulation thickness is an appreciable fraction of the wire thickness, the following correction should be added:

$$\Delta L = 0.03193an\left(2.303 \log_{10}\frac{D}{d} + 0.155\right) \text{ microhenrys} \quad (50)$$

where D is the distance between wire centers and d is the diameter of the bare wire in inches. This incremental correction increases as the insulation thickness increases.

Short Solenoidal Form.—The inductance of the multiple layer coil of the form shown in Fig. 27, in which b and c are the axial length and radial thickness of the winding measured to the outside of the winding, d the diagonal of the winding cross section, and a the mean radius, may be expressed in several ways. If n is the number

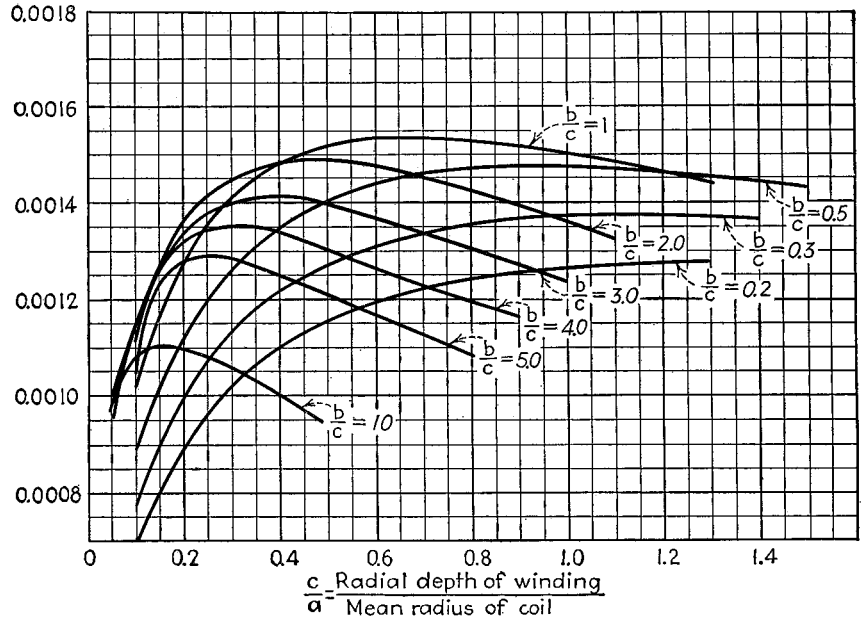


FIG. 28.—Factor I for use in Eq. (51) applying to multiple-layer coil of solenoidal form.

of turns, l the total length of the wire, and D the distance between centers of adjacent wires, then the low-frequency inductance may be expressed by (dimensions in inches)

$$L_0 = \frac{l^{5/3}}{D^{2/3}} I \text{ microhenrys} \tag{51}$$

or

$$L_0 = an^2 J \text{ microhenrys} \tag{52}$$

where I and J are the factors given in Figs. 28 and 29, respectively, in terms of c/a and b/c. It will be observed that the maximum inductance is obtained from a given length of wire when the cross section of the winding is square and the side of the cross section is 0.662 times the mean coil radius.

Where greater accuracy is desired than may be obtained from the curves of Figs. 28 and 29, the following formulas may be used:

For $b > c$:

$$L_0 = 0.03193an^2\left[2.303\left(1 + \frac{b^2}{32a^2} + \frac{d^2}{96a^2}\right)\log_{10}\frac{8a}{d} - y_1 + \frac{b^2}{16a^2}y_2\right] \tag{53}$$

where y_1 and y_2 are constants depending upon b/c and given in Table 16 below.

For $b < c$:

$$L_0 = 0.03193an^2\left[2.303\left(1 + \frac{b^2}{32a^2} + \frac{c^2}{96a^2}\right)\log_{10}\frac{8a}{d} - y_1 + \frac{c^2}{16a^2}y_3\right] \text{ microhenrys} \quad (54)$$

where y_1 and y_3 are constants depending upon b/c, as given in Table 16 below. The foregoing expressions are accurate to within 0.1 per cent. When the insulation

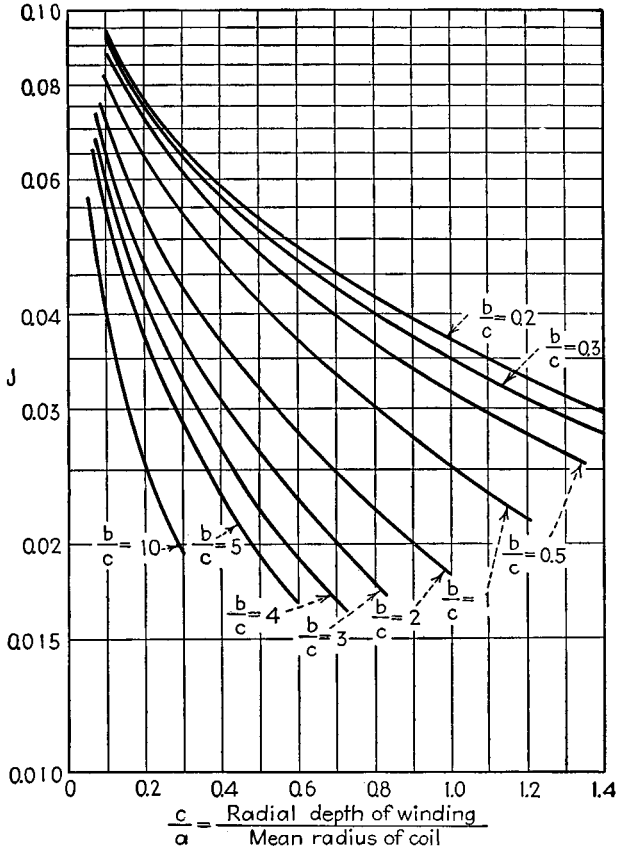


FIG. 29.—Factor J for use in Eq. (52) applying to multiple-layer coil of solenoidal form.

thickness of the wire is an appreciable fraction of the wire diameter, the following correction should be added:

$$\Delta L = +0.03193an\left(2.303 \log \frac{D}{d} + 0.155\right) \text{ microhenrys} \quad (55)$$

where D is the distance between adjacent wire centers and d is the diameter of the bare wire in inches.

A simple *approximate* formula for a short multiple-layer inductance of solenoidal form[1] is

$$L_0 = \frac{0.8a^2n^2}{6a + 9b + 10c} \text{ microhenrys} \quad (56)$$

[1] See Wheeler, *loc. cit.*

where the symbols have the significance indicated in Fig. 27 and the dimensions are in inches. This formula is accurate to within one per cent when the terms of the denominator are of about the same size, *i.e.*, for a coil of about the shape shown in Fig. 27.

Circular Coil of Circular Cross Section.—The low-frequency inductance of such a coil is

$$L_0 = 0.01595Dn^2 \left(2.303 \log \frac{8D}{d} - 1.75\right) \text{ microhenrys} \qquad (57)$$

TABLE 16.—VALUES OF y_1, y_2, AND y_3 FOR USE IN EQS. (53) AND (54)

b/c or c/b	y_1	Difference	c/b	y_2	Difference	b/c	y_3	Difference
0	0.5000	0.0253	0	0.125	0.002	0	0.597	0.002
0.025	0.5253	0.0237						
0.05	0.5490	0.0434	0.05	0.127	0.005	0.05	0.599	0.003
0.10	0.5924	0.0386	0.10	0.132	0 010	0.10	0.602	0.006
0.15	0.6310	0.0342	0.15	0.142	0.013	0.15	0.608	0.007
0.20	0.6652	0.0301	0.20	0.155	0.016	0.20	0.615	0.009
0.25	0.6953	0.0266	0.25	0.171	0.020	0.25	0.624	0.009
0.30	0.7217	0.0230	0.30	0.192	0.023	0.30	0.633	0.010
0.35	0.7447	0.0198	0.35	0.215	0.027	0.35	0.643	0.011
0.40	0.7645	0.0171	0.40	0.242	0.031	0.40	0.654	0.011
0.45	0.7816	0.0144	0.45	0.273	0.034	0.45	0.665	0.012
0.50	0.7960	0.0121	0.50	0.307	0.037	0.50	0.667	0.013
0.55	0.8081	0.0101	0.55	0.344	0.040	0.55	0.690	0.012
0.60	0.8182	0.0083	0.60	0.384	0.043	0.60	0.702	0.013
0.65	0.8265	0.0066	0.65	0.427	0.047	0.65	0.715	0.014
0.70	0.8331	0.0052	0.70	0.474	0.049	0.70	0.729	0.013
0.75	0.8383	0.0039	0.75	0.523	0.053	0.75	0.742	0.014
0.80	0.8422	0.0029	0.80	0.576	0.056	0.80	0.756	0.015
0.85	0.8451	0.0019	0.85	0.632	0.059	0.85	0.771	0.015
0.90	0.8470	0.0010	0.90	0.690	0.062	0.90	0.786	0.015
0.95	0.8480	0.0003	0.95	0.752	0.064	0.95	0.801	0.015
1.00	0.8483	.....	1.00	0.816	.....	1.00	0.816	

where D is the mean diameter of the turns and d is the diameter of the circular winding cross section in inches. This will be recognized as being simply n^2 times the expression for a single turn of wire of diameter d. The expression neglects the space occupied by the insulation between wires.

Rectangular Coil of Rectangular Cross Section.—The low frequency inductance may be represented by

$$L_0 = pn^2G \text{ microhenrys} \qquad (58)$$

where p is to be defined as being the perimeter of the rectangle in inches, n is the number of turns, and G is a quantity depending upon the ratio s_1/s_2 of the length of the short side to the length of the long side, as given in Fig. 21. The quantity b/s_2 appearing in this figure is taken to be the circumference of the winding cross section divided by twice the length of the longer side (*i.e.*, for b use b = circumference/$2S_2$).

Where greater accuracy is desired than may be obtained from the preceding formula, the following equation should be used. As before, let s_1 and s_2 be the short and long sides of the rectangle, respectively. Let t and w be the thickness and width of the cross section of the winding and g. the diagonal of the coil in inches. Let n represent the number of turns. Then

$$L_0 = 0.02339(s_1 + s_2)n^2\left[\log_{10}\frac{2s_1s_2}{t+w} - \frac{s_1}{s_1+s_2}\log_{10}(s_1+g) - \frac{s_2}{s_1+s_2}\log_{10}(s_2+g)\right] + 0.01016(s_1+s_2)n^2\left[2\left(\frac{g}{s_1+s_2}\right) - \frac{1}{2} + 0.447\frac{(t+w)}{(s_1+s_2)}\right] \text{ microhenrys} \quad (59)$$

If the insulation thickness is an appreciable fraction of the wire diameter, the correction of Eq. (55) should be added.

Square Coil of Rectangular Cross Section.—The inductance of a square coil may be obtained by setting $s_1 = s_2 = s$ in Eqs. (58) and (59). With this substitution, Eq. (59) simplifies to

$$L_0 = 0.04678sn^2\left[\log_{10}\frac{2s^2}{t+w} - \log_{10} 2.414s\right] + 0.02032sn^2\left[0.914 + 0.2235\frac{t+w}{s}\right] \text{ microhenrys} \quad (60)$$

12. Mutual Inductance and Coefficient of Coupling.—When the relative position of two inductances is such that lines of flux from one inductance link with turns of the other, the two inductances are said to be inductively coupled, and mutual inductance exists between them. Mutual inductance may be defined in terms of the number of flux linkages in the second coil per unit current in the first coil, or vice versa. In practical units, the relation is

$$M = \frac{\left\{\begin{array}{c}\text{flux linkages in second coil}\\ \text{produced by current in first coil}\end{array}\right\}}{\text{current in first coil}} 10^{-8} \quad (61a)$$

$$= \frac{\left\{\begin{array}{c}\text{flux linkages in first coil}\\ \text{produced by current in second coil}\end{array}\right\}}{\text{current in second coil}} 10^{-8} \quad (61b)$$

where M is the mutual inductance in henrys.

Mutual inductance may also be defined as the voltage induced in the second circuit when the current in the first circuit is changing at a unit rate. If the current flowing in the first circuit is sinusoidal, then the voltage induced in the second circuit is

$$E_2 = -j\omega M I_1. \quad (62)$$

The negative of the ratio of the voltage induced in the second circuit to the current in the first circuit is called the *mutual impedance.* That is

$$Z_m = j\omega M = \frac{-E_2}{I_1} \quad (63)$$

where Z_m is the mutual impedance and the quantity $-j$ indicates that the induced voltage lags the current by 90°. The concept of mutual impedance can be generalized to give the mutual impedance between any two circuits coupled in any fashion no matter how complex. The generalized mutual impedance is the vector ratio $-E_2/I_1$, where E_2 is the voltage appearing across the opened terminals of the second circuit and I_1 is the current flowing in the first.

Coefficient of Coupling.—The maximum value of mutual inductance that can exist between two coils of inductance L_1 and L_2 is $\sqrt{L_1L_2}$, which occurs when all the flux of one coil links with all the turns of the other. The ratio of the mutual inductance actually present to the maximum possible value that can occur is called the *coefficient of coupling* and is written as

$$k = \frac{M}{\sqrt{L_1L_2}} \quad (64)$$

The coefficient of coupling k is a dimensionless quantity having a maximum value of 1.

Closely coupled coils usually have a value of k of 0.5 or greater. Coils with a coefficient of 0.01 or less are said to be loosely coupled.

Circuits may be coupled capacitively or resistively as well as inductively. In the general case the coefficient of coupling is the ratio of the mutual impedance, as defined

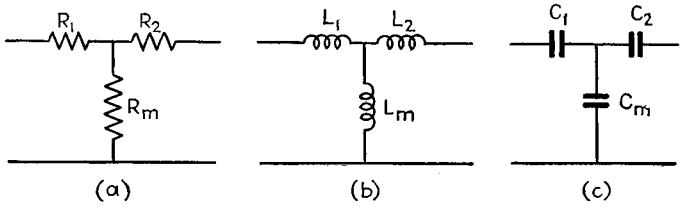


FIG. 30.—Directly coupled circuits.

above, to the square root of the product of the total self-impedances in the coupled circuits of the same nature as the coupling element. Thus, in general

$$k = \frac{Z'_m}{(Z'_1 Z'_2)^{1/2}} \tag{65}$$

where the primes indicate that all impedances are of the same kind. For circuits coupled resistively, as in Fig. 30*a*, the coefficient of coupling is

$$k = \frac{R_m}{\sqrt{(R_1 + R_m)(R_2 + R_m)}} \tag{66}$$

Similarly, for circuits directly coupled by a common inductance, as shown in Fig. 30*b*, the coefficient of coupling is

$$k = \frac{L_m}{\sqrt{(L_1 + L_m)(L_2 + L_m)}} \tag{67}$$

For circuits that are directly coupled by a common capacity, as shown in Fig. 30*c*, the coefficient of coupling is

$$k = \frac{\sqrt{C_1 C_2}}{\sqrt{(C_1 + C_m)(C_2 + C_m)}} \tag{68}$$

Combinations of Inductances Involving Mutual Inductances.—The total self-inductance of combinations of inductances depends upon the self- and mutual inductances involved. The total inductance resulting from two inductances in *series* is

$$L = L_1 + L_2 \pm 2M \tag{69}$$

where L_1 and L_2 are the self-inductances of the two coils and M is the mutual inductance between them. In general the total self-inductance of a number of inductances in series is the sum of the self-inductances of all the components plus the algebraic sum of the mutual inductances of each one of the component parts to all the other parts.

The total inductance resulting when two inductances are in *parallel* is

$$L = \frac{L_1 L_2 - M^2}{L_1 + L_2 \mp 2M} \tag{70}$$

13. Mutual Inductance between Straight Conductors. *Two Parallel Wires Side by Side.*—The mutual inductance between two straight parallel wires of the same length, as shown in Fig. 31, is given by

$$M = 0.00508l \left(2.303 \log_{10} \frac{2l}{D} - 1 + \frac{D}{l}\right) \text{ microhenrys} \tag{71}$$

where l is the length of the wire in inches and D is the distance between the wires. The formula assumes that the length of the wires is much greater than their spacing and that the effect of the cross section is unappreciable.

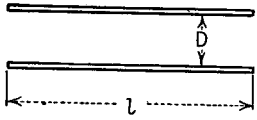


FIG. 31.—Two parallel wires side by side.

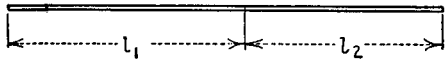


FIG. 32.—Two wire segments in same line, end-to-end.

Two Wire Segments in the Same Line End to End.—If the wire lengths are l_1 and l_2 in inches, as shown in Fig. 32, the mutual inductance is

$$M = 0.005850 \left(l_1 \log_{10} \frac{l_1 + l_2}{l_1} + l_2 \log_{10} \frac{l_1 + l_2}{l_2} \right) \text{ microhenrys} \tag{72}$$

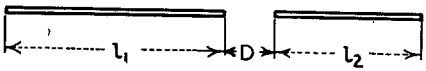


FIG. 33.—Two wire segments in same line, ends separated.

Two Wire Segments in the Same Line, Ends Separated.—If the wire lengths are l_1 and l_2 in inches and the ends are separated a distance D, as in Fig. 33, mutual inductance is

$$\begin{aligned} M = 0.005850[&(l_1 + l_2 + D) \log_{10} (l_1 + l_2 + D) + D \log_{10} D \\ &- (l_1 + D) \log_{10} (l_1 + D) - (l_2 + D) \log_{10} (l_2 + D)] \text{ microhenrys} \end{aligned} \tag{73}$$

Two Parallel Segments.—The mutual inductance for this case is

$$\begin{aligned} M = 0.00294 \left[l_1 \log_{10} \left(\frac{R_{1,46}}{R_{1,35}} \right) + l_2 \log_{10} \left(\frac{R_{1,46}}{R_{2,35}} \right) + e \log_{10} \left(\frac{R_{1,46}}{R_{1,36}} \times \frac{R_{2,35}}{R_{2,46}} \right) \right] \\ - 0.00254(r_{14} - r_{13} + r_{23} - r_{24}) \text{ microhenrys} \end{aligned} \tag{74}$$

where

$$R_{k,nm} = \frac{r_{kn} + r_{km}}{r_{kn} - r_{km}}$$

where

$$r_{14} = \sqrt{a^2 + (l_1 + l_2 + e)^2}$$
$$r_{16} = l_1 - l_2 + e, \text{ etc.}$$

and the other symbols are the lengths and distances indicated in Fig. 34 in inches.

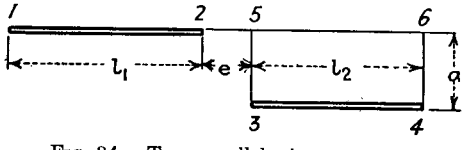


FIG. 34.—Two parallel wire segments.

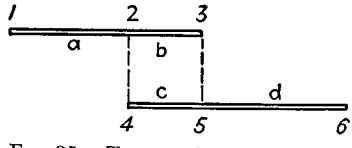


FIG. 35.—Two overlapping parallel wire segments.

The formula of Eq. (74) holds for $e = 0$, but not when one wire overlaps the other. When the wires overlap, as in Fig. 35

$$M = M_{a,cd} + M_{bc} + M_{bd} \text{ microhenrys} \tag{75}$$

where $M_{a,cd}$ is calculated by Eq. (74), using $e = 0$, using the length 1–2 for l_1, and using the length 4–6 for l_2. Similarly, for M_{bd}, use $e = 0$, length 2–3 for l_1, length 5–6 for l_2. M_{bc} is calculated from Eq. (71).

Two Parallel Symmetrically Placed Wires.—For the configuration of Fig. 36, the mutual inductance is

$$M = 0.00508 \left[4.606\, l_1 \log_{10} \left(\frac{l_1 + l_2 + \sqrt{(l_1 + l_2)^2 + D^2}}{D} \right) + 2.303(l_1 + l_2) \log_{10} \left(\frac{l_1 + l_2 + \sqrt{(l_1 + l_2)^2 + D^2}}{l_2 - l_1 + \sqrt{(l_2 - l_1)^2 + D^2}} \right) + \sqrt{(l_1 - l_2)^2 + D^2} - \sqrt{(l_1 + l_2)^2 + D^2} \right] \text{ microhenrys} \quad (76)$$

where the dimensions are in inches.[1]

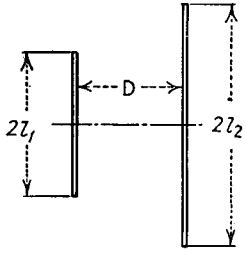


FIG. 36.—Two parallel symmetrically placed wires.

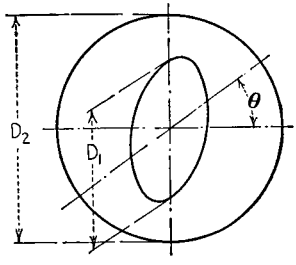


FIG. 38.—Two concentric circles.

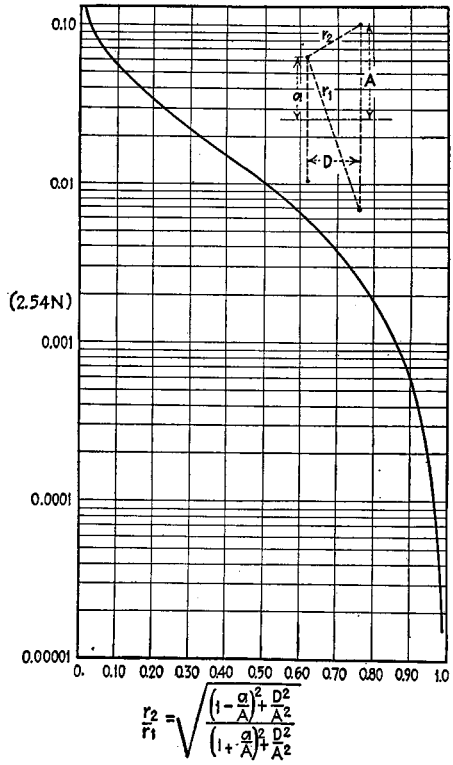


FIG. 37.—Factor $2.54N$ for use in Eq. (77) applying to two parallel coaxial circles.

14. Mutual Inductance between Single-turn Coils. *Two Parallel Coaxial Circles.* The mutual inductance between two parallel coaxial circles having the configuration shown in Fig. 36 is given by

$$M = 2.54\, N \sqrt{Aa} \text{ microhenrys} \quad (77)$$

where A is the radius of the large circle in inches, a is the radius of the small circle, and N depends upon r_1/r_2, as given in Fig. 37 and Table 17. The quantities r_1 and r_2 are, respectively, the longest and the shortest distances between points on the circum-

[1] For formulas for wire segments in positions other than given here, see Hak, *op. cit.*, p. 81, 1938; V. I. Bashenoff, Mutual Inductance of Circuits of Arbitrary Form, *Elektrotech. u. Machinenbau*, p. 1027, 1929.

ferences of the two circles. Their ratio in terms of the circle radii and spacing is

$$\frac{r_2}{r_1} = \sqrt{\frac{\left(1 - \frac{a}{A}\right)^2 + \frac{D^2}{A^2}}{\left(1 + \frac{a}{A}\right)^2 + \frac{D^2}{A}}} \tag{78}$$

Two Concentric Circles.—The mutual inductance of any two concentric circles of diameters D_2 and D_1 turned so that the angle between their planes is θ as in Fig. 38 may be expressed by

$$M = 0.00254 D_2 \phi \text{ microhenrys} \tag{79}$$

where D_2 is the diameter of the larger circle in inches and ϕ depends upon the ratio of diameters D_2/D_1, and the angle θ, as given by the curves of Fig. 39. For a ratio

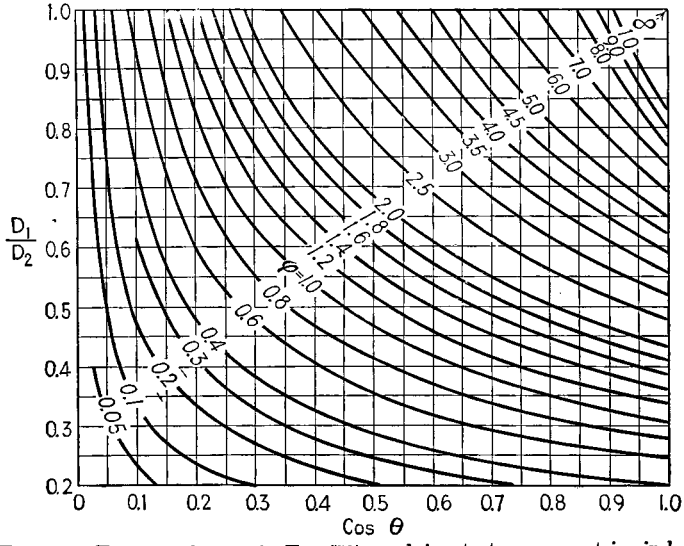


FIG. 39.—Factor ϕ for use in Eq. (79) applying to two concentric circles.

of diameters of 0.55, the mutual inductance is almost exactly proportional to the angle between the planes of the circles.[1]

Formulas are also available for the mutual inductance between two circles of any size in any arbitrary relative position.[2] These formulas tend to be quite complex, being usually given as a series expansion of spherical harmonics.

Two Equal Parallel Coaxial Rectangles.—Two equal parallel coaxial rectangles having the configuration shown in Fig. 40, with sides of length l_1 and l_2 in inches and spaced a distance D, have a mutual inductance given by

$$\begin{aligned} M = 0.02339 \Bigg[& l_1 \left(\log_{10} \frac{l_1 + \sqrt{l_1^2 + D^2}}{l_1 + \sqrt{l_1^2 + l_2^2 + D^2}} \times \frac{\sqrt{l_1^2 + D^2}}{D} \right. \\ & \left. + l_2 \left(\log_{10} \frac{l_2 + \sqrt{l_2^2 + D}}{l_2 + \sqrt{l_1^2 + l_2^2 + D^2}} \times \frac{\sqrt{l_1^2 + D^2}}{D} \right) \right] \\ & + 0.02032 \left(\sqrt{l_1^2 + l_2^2 + D^2} - \sqrt{l_1^2 + D^2} - \sqrt{l_2^2 + D_2^2} + D \right) \text{ microhenrys} \end{aligned} \tag{80}$$

[1] Hak, *op. cit.*, p. 78, 1938.

[2] Hak, *op. cit.*, p. 77; Chester Snow, Mutual Inductance of Any Two Circles, *Bur. Standards Jour. Research*, Vol. 1, p. 531, October, 1928.

TABLE 17.—VALUES OF N FOR USE IN EQ. (77)

r_2/r_1	N	Difference	r_2/r_1	N	Difference
.....	.....	.....	0.30	0.008844	−0.000341
0.000	∞	.....	0.31	0.008503	−0.000328
0.010	0.05016	−0.00120	0.32	0.008175	−0.000314
0.011	0.04897	−0.00109	0.33	0.007861	−0.000302
0.012	0.04787	−0.00100	0.34	0.007559	−0.000290
0.013	0.04687	−0.00093	0.35	0.007269	−0.000280
0.014	0.04594	−0.00087	0.36	0.006989	−0.000270
0.015	0.04507	−0.00081	0.37	0.006720	−0.000260
0.016	0.04426	−0.00148	0.38	0.006460	−0.000249
0.018	0.04278	−0.00132	0.39	0.006211	−0.000241
0.020	0.04146	−0.00119	0.40	0.005970	−0.000232
0.022	0.04027	−0.00109	0.41	0.005738	−0.000225
0.024	0.03918	−0.00100	0.42	0.005514	−0.000217
0.026	0.03818	−0.00093	0.43	0.005297	−0.000210
0.028	0.03725	−0.00086	0.44	0.005087	−0.000202
0.030	0.03639	−0.00081	0.45	0.004885	−0.000195
0.032	0.03558	−0.00076	0.46	0.004690	−0.000189
0.034	0.03482	−0.00071	0.47	0.004501	−0.000183
0.036	0.03411	−0.00068	0.48	0.004318	−0.000178
0.038	0.03343	−0.00064	0.49	0.004140	−0.000171
0.040	0.03279	−0.00061	0.50	0.003969	−0.000166
0.042	0.03218	−0.00058	0.51	0.003803	−0.000160
0.044	0.03160	−0.00055	0.52	0.003643	−0.000156
0.046	0.03105	−0.00053	0.53	0.003487	−0.000150
0.048	0.03052	−0.00051	0.54	0.003337	−0.000146
0.050	0.03001	−0.00226	0.55	0.003191	−0.000141
0.060	0.02775	−0.00191	0.56	0.003050	−0.000137
0.070	0.02584	−0.00164	0.57	0.002913	−0.000133
0.080	0.02420	−0.00144	0.58	0.002780	−0.000128
0.090	0.02276	−0.00128	0.59	0.002652	−0.000125
0.100	0.02148	−0.00116	0.60	0.002527	−0.000120
0.11	0.02032	−0.00104	0.61	0.002407	−0.000117
0.12	0.01928	−0.00096	0.62	0.002290	−0.000113
0.13	0.01832	−0.00089	0.63	0.002177	−0.000109
0.14	0.01743	−0.00082	0.64	0.002068	−0.000106
0.15	0.01661	−0.00075	0.65	0.001962	−0.000103
0.16	0.01586	−0.00071	0.66	0.001859	−0.000099
0.17	0.01515	−0.00066	0.67	0.001760	−0.000096
0.18	0.01449	−0.00062	0.68	0.001664	−0.000093
0.19	0.01387	−0.00059	0.69	0.001571	−0.000090
0.20	0.01328	−0.00055	0.70	0.001481	−0.000087
0.21	0.01273	−0.00052	0.71	0.001394	−0.000084
0.22	0.01221	−0.00050	0.72	0.001310	−0.000081
0.23	0.01171	−0.00047	0.73	0.001228	−0.000078
0.24	0.01124	−0.00045	0.74	0.001150	−0.000076
0.25	0.010792	−0.000425	0.75	0.0010741	−0.0000731
0.26	0.010366	−0.000408	0.76	0.0010010	−0.0000704
0.27	0.009958	−0.000388	0.77	0.0000931	−0.0000680
0.28	0.009570	−0.000371	0.78	0.0000863	−0.0000643
0.29	0.009199	−0.000355	0.79	0.0000797	−0.0000628

TABLE 17.—VALUES OF N FOR USE IN EQ. (77).—(*Continued*)

r_2/r_1	N	Difference	r_2/r_1	N	Difference
0.80	0.0007345	−0.0000604	0.950	0.00008107	−0.00000494
0.81	0.0006741	−0.0000579	0.952	0.00007613	−0.00000482
0.82	0.0006162	−0.0000555	0.954	0.00007131	−0.00000470
0.83	0.0005607	−0.0000531	0.956	0.00006661	−0.00000458
0.84	0.0005076	−0.0000507	0.958	0.00006202	−0.00000446
0.85	0.0004569	−0.0000484	0.960	0.00005756	−0.00000436
0.86	0.0004085	−0.0000460	0.962	0.00005320	−0.00000421
0.87	0.0003625	−0.0000437	0.964	0.00004899	−0.00000409
0.88	0.0003188	−0.0000413	0.966	0.00004490	−0.00000397
0.89	0.0002775	−0.0000389	0.968	0.00004093	−0.00000383
0.90	0.0002386	−0.0000365	0.970	0.00003710	−0.00000370
0.91	0.0002021	−0.0000341	0.972	0.00003340	−0.00000356
0.92	0.0001680	−0.0000316	0.974	0.00002984	−0.00000341
0.93	0.0001364	−0.0000290	0.976	0.00002643	−0.00000327
0.94	0.0001047	−0.0000263	0.978	0.00002316	−0.00000312
0.95	0.00008107	−0.00002351	0.980	0.00002004	−0.00000296
0.96	0.00005756	−0.00002046	0.982	0.00001708	−0.00000278
0.97	0.00003710	−0.00001706	0.984	0.00001430	−0.00000262
0.98	0.00002004	−0.00001301	0.986	0.00001168	−0.00000242
0.99	0.00000703	−0.00000703	0.988	0.00000926	−0.00000223
1.00	0.00000		0.990	0.00000703	−0.00000201
			0.992	0.00000502	−0.00000177
			0.994	0.00000326	−0.00000148
			0.996	0.00000177	−0.00000115
			0.998	0.00000062	−0.00000062

Two Equal Parallel Squares.—If l is the side of each of two squares spaced a distance D inches, then Eq. (80) becomes

$$M = 0.04678 \left(l \log_{10} \frac{l + \sqrt{l^2 + D^2}}{l + \sqrt{2l^2 + D^2}} \times \frac{\sqrt{l^2 + D^2}}{D} \right) + 0.02032 \left(\sqrt{2a^2 + D^2} - 2\sqrt{a^2 + D^2} + D \right) \text{ microhenrys} \quad (81)$$

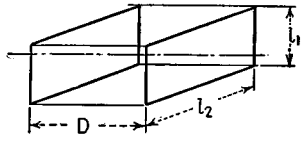


FIG. 40.—Two equal parallel coaxial rectangles.

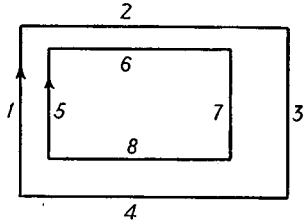


FIG. 41.—Coplanar rectangles with parallel sides.

Coplanar Rectangles with Parallel Sides.—If the two rectangles have their sides numbered as in Fig. 41, then

$$M = M_{15} + M_{26} + M_{37} + M_{48} - M_{17} - M_{28} - M_{35} - M_{46} \text{ microhenrys} \quad (82)$$

If the rectangles are concentric

$$M_{15} = M_{37}, \qquad M_{26} = M_{48}, \qquad M_{17} = M_{35}, \qquad M_{28} = M_{46},$$

so that

$$M = 2(M_{15} + M_{26}) - 2(M_{17} + M_{28}) \text{ microhenrys} \tag{83}$$

If the figures are concentric squares

$$M = 4(M_{15} - M_{17}) \tag{84}$$

The separate terms are calculated by Eq. (76) if the sides are symmetrically placed. If they are not symmetrically placed, the terms are calculated by Eqs. (76) and (74).

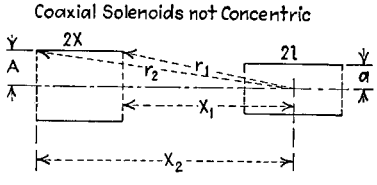


FIG. 42.—Coaxial solenoids not concentric.

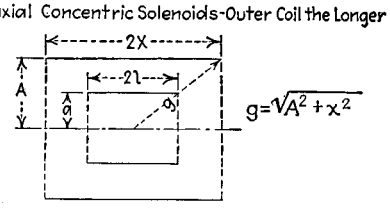


FIG. 43.—Coaxial concentric solenoids, outer coil the longer.

15. Mutual Inductance between Single-layer Solenoids. *Coaxial Coils Not Concentric.*—When the configuration is as shown in Fig. 42, the mutual inductance is given by

$$M = 0.02505 \frac{a^2 A^2 n_1 n_2}{4lx} (K_1 k_1 + K_3 k_3 + K_5 k_5) \text{ microhenrys} \tag{85}$$

where a = smaller radius, measured from the axis of the coil to the center of the wire, in.

A = larger radius, measured in the same way, in.

$2l$ = length of coil of smaller radius = number of turns times pitch of winding, in.

$2x$ = length of coil of larger radius, measured in the same way, in.

n_1 and n_2 = total number of turns on the two coils.

$$K_1 = \frac{2}{A^2}\left(\frac{x_2}{r_2} - \frac{x_1}{r_1}\right), \qquad k_1 = 2l$$

$$x_1 = D - x, \qquad r_1 = \sqrt{x_1^2 + A^2}$$

$$x_2 = D + x, \qquad r_2 = \sqrt{x_2^2 + A^2}$$

$$K_3 = \frac{1}{2}\left(\frac{x_1}{r_1^5} - \frac{x_2}{r_2^5}\right), \qquad k_3 = a^2 l\left(3 - \frac{4l^2}{a^2}\right)$$

$$K_5 = -\frac{A^2}{8}\left[\frac{x_1}{r_1^9}\left(3 - \frac{4x_1^2}{A^2}\right) - \frac{x_2}{r_2^9}\left(3 - \frac{4x_2^2}{A^2}\right)\right]$$

$$k_5 = a^4 l\left(\frac{5}{2} - 10\frac{l^2}{a^2} + 4\frac{l^2}{a^4}\right)$$

D = axial distance between centers of coils, in.

Coaxial Concentric Coils—Outer Coil the Longer.—For coils having the configuration shown in Fig. 43 and for symbols having the significance shown (measured in inches), the mutual inductance is given by

$$M = 0.0501 \frac{a^2 n_1 n_2}{g}\left[1 + \frac{A^2 a^2}{8g^4}\left(3 - 4\frac{l^2}{a^2}\right)\right] \text{ microhenrys} \tag{86}$$

A simple *approximate* expression for the coefficient of coupling for coils having the configuration of Fig. 43 is given by

$$k = \frac{a^2 l}{A^2 x} \tag{87}$$

Coaxial Concentric Solenoids—Outer Coil the Shorter.—When the coils have the configuration shown in Fig. 44, mutual inductance is again given by Eq. (86), provided g is defined as indicated in Fig. 44 and all dimensions are in inches.

Discussion of Formulas.—In the preceding formulas for the mutual inductance between solenoids, the expressions involve differences between quantities of nearly the same size. To obtain a high degree of accuracy it is hence necessary to carry out the calculations with a considerable number of significant figures. Other formulas are also available for calculating the mutual inductance of solenoids in which the calculations are not so exacting but involve a greater number of terms.[1] Where great accuracy is not demanded and many calculations are to be made, time can be saved by using a combined graphical and computational method. Formulas are also available for other cases of mutual inductance between solenoids than those given here.[2] These are invariably quite complex, involving elliptic integrals or series expansions with spherical harmonics.

16. Mutual Inductance between Multilayer Coils. *Two Coaxial Circular Coils of Small Rectangular Cross Section.*—For coils having the approximate configuration shown in Fig. 45, the mutual inductance is given by

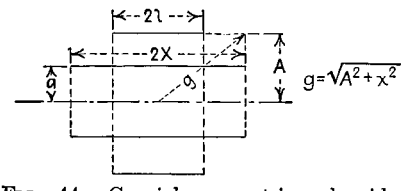


FIG. 44.—Coaxial concentric solenoids, outer coil the shorter.

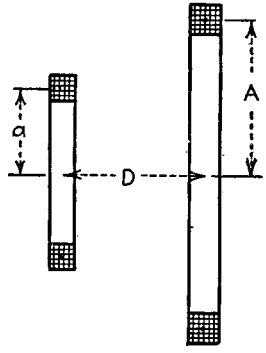


FIG. 45.—Two coaxial circular coils of small rectangular cross section.

$$M = n_1 n_2 M_0 \text{ microhenrys} \tag{88}$$

where n_1 and n_2 are the number of turns of the two coils and M_0 is the mutual inductance of the coaxial circles located at the centers of the cross sections of the two windings. Thus if a is the mean radius in inches of the small coil and A is the mean radius of the large coil and D the distance between their planes, the value of M_0 is that computed by Eq. (77) and Table 17 for the two coaxial circles thus defined. This formula is best suited for coils with square cross section and gives results in error by less than 0.5 per cent, even with relatively large cross-sectional dimensions, except when the coils are close together.

Two Coaxial Circular Coils of Rectangular Cross Section.—For coils whose cross-sectional dimensions are relatively large, compared to the coil diameters, as shown in Fig. 46, the mutual inductance is given by

$$M = \frac{n_1 n_2}{6}\Big[(M_{12'} + M_{13'} + M_{14'} + M_{15'}) + (M_{1'2} + M_{1'3} + M_{1'4} + M_{1'5}) - 2M_{11'}\Big] \text{ microhenrys} \tag{89}$$

[1] H. B. Dwight and F. W. Grover, Some Series Formulas for Mutual Inductance of Solenoids, *Elec. Eng.*, Vol. 56, p. 347, March, 1937; F. W. Grover, Tables for the Calculation of the Mutual Inductance of Any Two Coaxial Single-layer Coils, *Proc. I.R.E.*, Vol. 21, p. 1039, July, 1933.

[2] F. W. Grover, Mutual Inductance of Coils not Coaxial, *Communications*, Vol. 18, p. 10, August, 1938. See also Chester Snow, Mutual Inductance of Concentric Solenoids, *Bur. Standards Jour. Research*, Vol. 1, p. 531, 1928.

where the various terms represent the mutual inductance between the coaxial circles corresponding to the subscripts as indicated in Fig. 46. The separate terms are calculated by Eq. (77) and Table 17.[1]

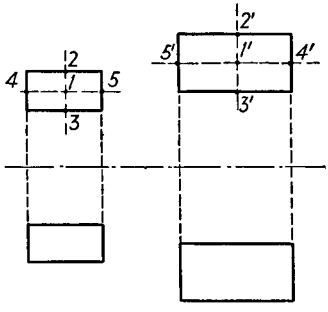


FIG. 46.—Two coaxial circular coils of rectangular cross section.

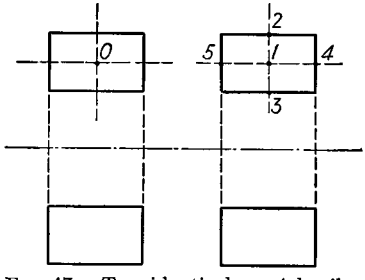


FIG. 47.—Two identical coaxial coils.

Two Identical Coaxial Coils.—When the coils are identical and of rectangular cross section, as shown in Fig. 47, the mutual inductance is given by

$$M = \frac{n^2}{3}(M_{02} + M_{03} + M_{04} + M_{05} - M_{01}) \text{ microhenrys} \qquad (90)$$

where n is the number of turns per coil and the separate terms are the mutual inductances between the coaxial circles corresponding to the subscripts, as indicated in the figure. As before, the separate terms are calculated by Eq. (77) and Table 17.

Two Coils of Equal Diameter.—When the coils are of equal diameter and radial depth but of different lengths, as shown in Fig. 48, the mutual inductance is given by

$$M = \frac{n_1 n_2}{6}(2M_{11'} + M_{14'} + M_{15'} + M_{1'4} + M_{1'5}) \text{ microhenrys} \qquad (91)$$

where n_1 and n_2 are the number of turns of the two coils and the M terms are the mutual inductances between the coaxial circles corresponding to the subscripts as indicated in the figure as calculated by Eq. (77) and Table 17.

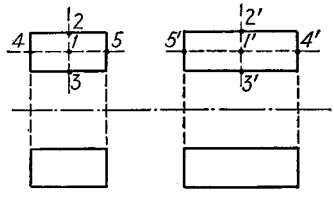


FIG. 48.—Two coaxial coils of equal diameter.

COILS WITH AIR CORES

17. Types of Windings Used in Air-cored Coils.—Most air-cored coils are single or multilayer arrangements wound on cylindrical forms. Other types occasionally used are toroids, flat spirals (or disks), and single or multilayer windings on square, hexagonal, etc., forms.

Single-layer coils are used to provide inductances for resonant circuits at broadcast and higher frequencies and for radio-frequency choke coils at very high frequencies. Multilayer coils are used when a large inductance is to be obtained in a small space. Such coils are employed in resonant circuits at broadcast and lower frequencies, as radio-frequency choke coils and as inductance standards for audio frequencies.

[1] Hak, *op. cit.*, p. 61, 1938.

18. Losses in Air-cored Coils at Radio Frequencies. *Method of Expressing Losses.*—Coil losses can be expressed in terms of the equivalent series resistance R, in terms of power factor of the coil, or in terms of the ratio $\omega L/R = Q$. The coil power factor and Q are related by the equation

$$\text{Power factor} = \frac{1}{\sqrt{Q^2 + 1}} \tag{92}$$

When $Q > 5$, the power factor can be taken as $1/Q$.

Factors Contributing to Coil Losses.—The equivalent series resistance of a coil is greater than the d-c resistance at all but the very lowest frequencies, as a result of:[1] (1) skin effect in the conductor that modifies the current distribution over the cross section of the individual conductor; (2) proximity effect with respect to adjacent conductors in the coil, which produces an additional modification in the current distribution; (3) losses in dielectrics in the vicinity of the coil, particularly in the coil form; (4) eddy-current losses in neighboring metal objects. These effects tend to increase with frequency, with the result that the ratio $\omega L/R = Q$ varies only slowly with variations in frequency.

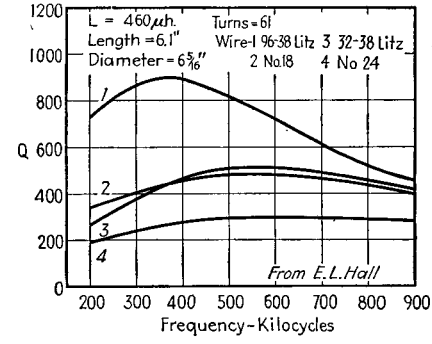
Q of Single-layer Coils.—The Q of a single-layer coil depends upon the frequency, size, and shape of the coil, and such constructional details as the number of turns, size, and character of the wire, etc. With a given coil, the Q, as observed experimentally, varies somewhat with frequency and goes through a broad maximum at a frequency that is usually either in or near the working frequency range of the coil. With a given inductance and coil diameter, the Q is maximum when the ratio (length)/(diameter) is of the order of ½:1, although the relation is not at all critical. With everything fixed except the size of wire, it is usually found that maximum Q is obtained with wire somewhat smaller than the largest that could be used in the available space. When the size of a coil is increased while maintaining the inductance and shape constant, the Q will be roughly proportional to square root of the coil diameter, provided the wire diameter is always maintained at the optimum value.

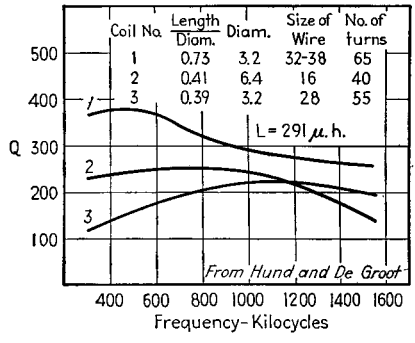
Edgewise-wound strip and ribbon conductors are sometimes used in large single-layer and pancake coils, respectively. Such conductors can be expected to give lower Q than solid wire of the same surface area, but if the dimensions are large and the strip or ribbon of ample cross section, the losses will still be low.

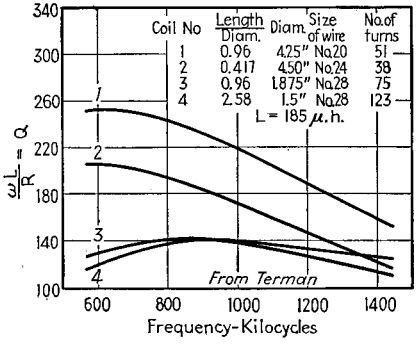
Litz wire gives a higher Q than solid wire at the lower radio frequencies, while at very high frequencies solid wire is as good as or better than litz. The frequency at which litz loses its advantage depends upon circumstances, but is normally of the order of 1,000 to 2,000 kc. The proper size of litz wire to use depends upon the frequency and coil construction, and, in general, does not have the same cross section as does the corresponding solid wire of optimum proportions. The effect of broken or insulated strands in a coil wound with litz wire is far less than the loss in d-c conductivity.

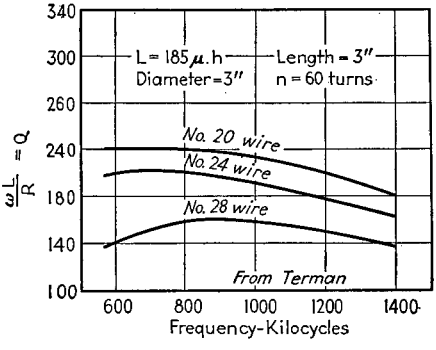
The form, binder, etc., used with single-layer coils have relatively little effect upon the losses, except possibly at very high frequencies, provided the dielectrics involved do not absorb moisture. Thus bakelite, polystyrene, isolantite, etc., are quite satisfactory, whereas untreated paper or cardboard is not. Experiments indicate that there is no unusual advantage, even at very high frequencies, of using types of construction such as "basket-weave" coils, in which the amount of dielectric associated with the coil is reduced to the barest minimum.

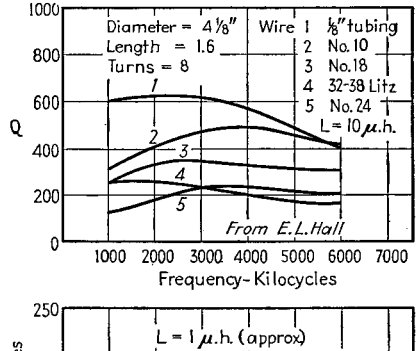
[1] Considerable attention is sometimes paid to coil proportions that have the lowest possible d-c resistance. Such considerations are, however, of little significance at radio frequencies because of skin and proximity effect. Their only usefulness is in connection with inductances for low frequencies, such as standard inductances for audio-frequency bridge work.

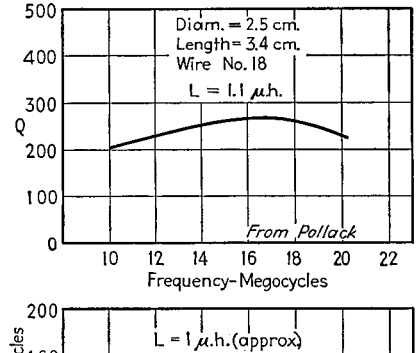


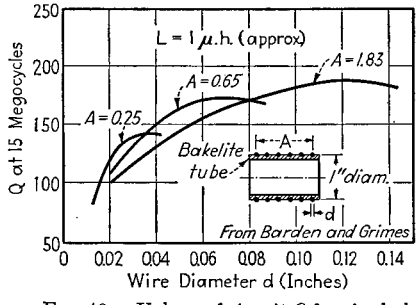












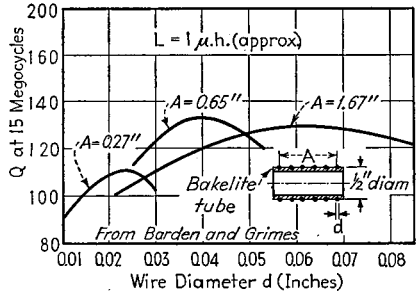


FIG. 49.—Values of circuit Q for single-layer coils designed for different frequency ranges, as reported by various investigators. All of these data were measured on the assumption that the tuning condenser had zero losses (*i.e.*, the condenser losses are charged against the coil).

Experimental data giving Q as a function of frequency for a variety of single-layer coils are given in Fig. 49. These data have been obtained from various sources, and are representative of what can be expected.[1,2]

Q of Multilayer Coils.—The losses in a multilayer coil depend upon the frequency, size, and shape of the coil, the arrangement of the winding, the size and kind of wire, etc., just as in the single-layer coil. With a given coil, the Q varies only slowly with frequency and goes through a broad maximum at a frequency that is usually in or near the working range of the coil. The Q increases with the size of the coil providing the optimum wire size and arrangement is maintained, but the effect of size is not as

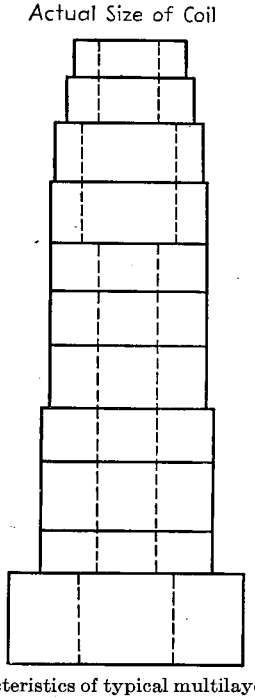


FIG. 50.—Characteristics of typical multilayer coils such as used at intermediate frequencies.

great as might be expected, and is frequently overshadowed by other factors. The exact shape and arrangement of the winding space is not critical, and satisfactory multilayer coils can be built in such widely different arrangements as a thin winding having two or three layers of many turns, and a disk or pancake winding having many layers, each of only a few turns. Good efficiency will be obtained provided the proportions do not depart too greatly from the relation $3t + 2b = D$, where t is the radial thickness of the winding, b the axial length, and D the outer diameter.

[1] Values given by the curves involve in nearly every case the assumption that the condenser used to tune the coil to resonance when the measurements were being made had zero losses. As a result, the true coil losses are probably somewhat less than those presented in the figures, particularly toward the high-frequency end of the useful range of the coil.

[2] D. Pollack, The Design of Inductances for Frequencies between 4 and 25 Megacycles, *Elec. Eng.*, Vol. 56, p. 1169, September, 1937; D. Grimes and W. Barden, Coil Design for Short Wave Receivers, *Electronics*, Vol. 7, p. 174, June, 1934; F. E. Terman, "Radio Engineering," 2d ed., p. 38; A. J. Palermo and F. W. Grover, Study of the High-frequency Resistance of Single Layer Coils, *Proc. I.R.E.*, Vol. 18, p. 2041, December, 1930; E. L. Hall, Resistance of Conductors of Various Types and Sizes as Windings of Single-layer Coils at 150 to 6,000 Kilocycles, *Bur. Standards Tech. Paper* 330, October, 1926; August Hund and H. B. DeGroot, Radio-frequency Resistance and Inductance of Coil used in Radio Reception, *Bur. Standards Tech. Paper* 298, October, 1925.

The optimum size of wire ordinarily has a wire diameter that gives a cross section of copper considerably less than the cross section of the winding space, corresponding to a not too closely packed winding. Litz gives a higher Q than solid conductors at broadcast and lower frequencies, which are the frequencies where multilayer coils find their principal use. The size and number of the strands in litz for best results depends upon the frequency, physical size of the coil, etc., and ranges from three strands of No. 40 or five strands of No. 41 wire for very compact coils intended for broadcast and intermediate frequencies, to proportionately larger and more numerous strands as the physical size of the coil is increased.

Data on typical multilayer coils of the type used in radio receivers are shown in Fig. 50, and indicate the type of behavior that can be expected.[1]

Q of Toroidal Coils.—The Q's obtained with toroidal coils are always appreciably lower than for ordinary single-layer or multilayer coils of corresponding physical dimensions. In fact, an ordinary radio-frequency coil that is well designed and enclosed in a shield will have at least as high a Q and no more external magnetic field than a toroid of corresponding dimensions, and will also have zero external electrostatic field.

Maximum Q with given outer diameter is obtained in single-layer toroids of circular cross section wound with the best size of wire when the diameter of the winding section is approximately 0.35 times the over-all diameter. With single-layer toroids of rectangular cross section wound with the optimum size of wire, the optimum radial depth of winding is approximately 0.3 times the outside diameter, and the optimum height is approximately 0.4 times the outside diameter.[2]

The losses of toroidal coils can be reduced by using a single-layer winding and a relatively flat conductor so curved that the flat side follows exactly the surface of the toroid, as in Fig. **52.**[3] The maximum Q of such a toroid with circular cross section occurs at $r/R = 0.71$ for fixed outer diameter. For rectangular cross section, the optimum condition for given height h and outer diameter are $r_1/r_2 = 0.279$. The Q of such coils is theoretically proportional to the size and to the square root of the frequency. With large dimensions and very high frequencies, Q's of the order of 1,000 to 10,000 are theoretically realizable. There are practical difficulties of construction, however.

19. Calculation of Copper Loss in Coils.—The most important loss in coils is the copper loss as influenced by skin and proximity effects. The problem of deriving formulas for calculating coil losses when skin and proximity effects are involved has been attacked by several investigators. The most extensive work of this character is that of Butterworth[4], who considered a coil to be composed of a number of short cylindrical parallel sections of conductor, and then determined the losses in each of these cylinders as a result of the magnetic fields produced by the current in each cylinder and in the neighboring cylinders. The results of these calculations are summarized below.

Copper Losses in Single-layer Solenoids.—For not too closely wound single-layer solenoids employing solid wire, one has

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = H + u\left(\frac{d_0}{c}\right)^2 G \tag{93}$$

[1] D. Grimes and W. S. Barden, A Study of Litz Wire Coils, *Electronics*, Vol. 6, p. 303, November; p. 342, December, 1933.

[2] S. Butterworth, The High-frequency Resistance of Toroidal Coils, *Exp. Wireless and Wireless Eng.*, Vol. 6, p. 13, January, 1929.

[3] F. E. Terman, Some Possibilities for Low Loss Coils, *Proc. I.R.E.*, Vol. 23, p. 1069, September, 1935.

[4] S. Butterworth, Effective Resistance of Inductance Coils at Radio Frequencies, *Exp. Wireless and Wireless Eng.*, Vol. 3, p. 203, April, 1926; p. 302, May, 1926; p. 417, July, 1926; p. 483, August, 1926; On the Alternating Current Resistance of Solenoidal Coils, *Proc. Roy. Soc. (London)*, Vol. 107, p. 693, 1925. A summary of Butterworth's work is given by B. B. Austin, Effective Resistance of Inductance Coils at Radio Frequency, *Wireless Eng. and Exp. Wireless*, Vol. 11, p. 12, January, 1934.

TABLE 18.—VALUES OF THE FUNCTIONS H AND G

d = diameter of wire, cm; ρ = resistivity, cgs units; f = frequency, cycles per sec; $x = \pi d\sqrt{2f/\rho}$. For copper of resistivity 1,700 cgs units $x = 0.1078\, d\sqrt{f}$

x	H	G	x	H	G	x	H	G	x	H	G
0.0	1.000	$x^4/64$	2.5	1.175	0.2949	5.0	2.043	0.755	10.0	3.799	1.641
0.1	1.000		2.6	1.201	0.3184	5.2	2.114	0.790	11.0	4.151	1.818
0.2	1.000		2.7	1.228	0.3412	5.4	2.184	0.826	12.0	4.504	1.995
0.3	1.000		2.8	1.256	0.3632	5.6	2.254	0.861	13.0	4.856	2.171
0.4	1.000		2.9	1.286	0.3844	5.8	2.324	0.896	14.0	5.209	2.348
0.5	1.000	0.00097	3.0	1.318	0.4049	6.0	2.394	0.932	15.0	5.562	2.525
0.6	1.001	0.00202	3.1	1.351	0.4247	6.2	2.463	0.967	16.0	5.915	2.702
0.7	1.001	0.00373	3.2	1.385	0.4439	6.4	2.533	1.003	17.0	6.268	2.879
0.8	1.002	0.00632	3.3	1.420	0.4626	6.6	2.603	1.038	18.0	6.621	3.056
0.9	1.003	0.01006	3.4	1.456	0.4807	6.8	2.673	1.073	19.0	6.974	3.233
1.0	1.005	0.01519	3.5	1.492	0.4987	7.0	2.743	1.109	20.0	7.328	3.409
1.1	1.008	0.02196	3.6	1.529	0.5160	7.2	2.813	1.144	21.0	7.681	3.586
1.2	1.011	0.03059	3.7	1.566	0.5333	7.4	2.884	1.180	22.0	8.034	3.763
1.3	1.015	0.04127	3.8	1.603	0.5503	7.6	2.954	1.216	23.0	8.388	3.940
1.4	1.020	0.0541	3.9	1.640	0.5673	7.8	3.024	1.251	24.0	8.741	4.117
1.5	1.026	0.0691	4.0	1.678	0.5842	8.0	3.094	1.287	25.0	9.094	4.294
1.6	1.033	0.0863	4.1	1.715	0.601	8.2	3.165	1.322	30.0	10.86	5.177
1.7	1.042	0.1055	4.2	1.752	0.618	8.4	3.235	1.357	40.0	14.40	6.946
1.8	1.052	0.1265	4.3	1.789	0.635	8.6	3.306	1.393	50.0	17.93	8.713
1.9	1.064	0.1489	4.4	1.826	0.652	8.8	3.376	1.428	60.0	21.46	10.48
2.0	1.078	0.1724	4.5	1.863	0.669	9.0	3.446	1.464	70.0	25.00	12.25
2.1	1.094	0.1967	4.6	1.899	0.686	9.2	3.517	1.499	80.0	28.54	14.02
2.2	1.111	0.2214	4.7	1.935	0.703	9.4	3.587	1.534	90.0	32.07	15.78
2.3	1.131	0.2462	4.8	1.971	0.720	9.6	3.657	1.570	100.0	35.61	17.55
2.4	1.152	0.2708	4.9	2.007	0.738	9.8	3.728	1.605			
2.5	1.175	0.2949	5.0	2.043	0.755	10.0	3.799	1.641	Large	$\frac{\sqrt{2x}+1}{4}$	$\frac{\sqrt{2x}-1}{8}$

TABLE 19.—VALUES OF u FOR SINGLE-LAYER SOLENOIDS OF MANY TURNS (SPACED)

D = diameter of coil
b = length of coil
u_1 = contribution of radial component of field
u_2 = contribution of axial component of field

b/D	u_1	u_2	$u = u_1 + u_2$
0.0	3.29	0.00	3.29
0.2	3.13	0.50	3.63
0.4	2.83	1.23	4.06
0.6	2.51	1.99	4.50
0.8	2.22	2.71	4.93
1.0	1.94	3.35	5.29
2	1.11	5.47	6.58
4	0.51	7.23	7.74
6	0.31	8.07	8.38
8	0.21	8.52	8.73
10	0.17	8.73	8.90
∞	0.00	9.87	9.87

where H = resistance ratio of wire when isolated, as given by Table 4 (or Table 18).
G = proximity effect factor, given by Table 18.
u = a constant given by Table 19 or 20.

$$d_0/c = \frac{\text{wire diameter}}{\text{spacing between centers of adjacent turns}}.$$

This equation holds for single-layer solenoids having many turns provided u is obtained from Table 19 and $d_0/c \lesssim 0.6$ (*i.e.*, turns not too closely spaced[1]). It also applies to coils having a few spaced turns provided the length is small compared with the diameter and u is obtained from Table 20.

TABLE 20.—VALUES OF u FOR SINGLE-LAYER SOLENOIDS OF FEW TURNS (SPACED)

No. of turns.....	2	4	6	8	10	12	16	24	32	∞
u.....	1.00	1.80	2.16	2.37	2.51	2.61	2.74	2.91	3.00	3.29

TABLE 21.—VALUES OF α, β, γ IN EQ. (94)*

d_0/c	$x = 1$			$x = 2$			$x = 3$		
	α	β	γ	α	β	γ	α	β	γ
1.0	1.01	1.02	0.96	1.09	1.34	0.67	1.31	2.29	0.49
0.9	1.00	1.02	0.97	1.06	1.29	0.72	1.20	1.99	0.55
0.8	....	1.02	0.98	1.04	1.23	0.78	1.13	1.73	0.62
0.7	....	1.02	0.98	1.02	1.18	0.83	1.08	1.52	0.68
0.6	....	1.01	0.99	1.00	1.13	0.87	1.04	1.36	0.75
0.5	....	1.01	0.99	....	1.09	0.91	1.02	1.24	0.82
0.4	....	1.01	0.99	....	1.06	0.94	1.01	1.14	0.88
0.3	....	1.00	1.00	....	1.04	0.97	1.00	1.06	0.93
0.2	....	....	....	....	1.01	0.99	....	1.03	0.97
0.1	....	....	....	....	1.00	1.00	....	1.01	0.99

d_0/c	$x = 4$			$x = 5$			$x = \infty$		
	α	β	γ	α	β	γ	α	β	γ
1.0	1.43	3.61	0.43	1.50	4.91	0.41	1.71	inf.	0.35
0.9	1.30	2.75	0.49	1.37	3.39	0.46	1.55	12.45	0.39
0.8	1.21	2.12	0.55	1.25	2.48	0.53	1.41	4.83	0.44
0.7	1.12	1.71	0.62	1.15	1.94	0.60	1.27	2.87	0.52
0.6	1.07	1.51	0.70	1.09	1.60	0.68	1.16	2.03	0.60
0.5	1.03	1.32	0.78	1.04	1.37	0.76	1.08	1.59	0.69
0.4	1.02	1.19	0.85	1.02	1.22	0.84	1.03	1.33	0.78
0.3	1.00	1.10	0.91	1.00	1.11	0.90	1.01	1.17	0.87
0.2	....	1.04	0.96	....	1.05	0.96	1.00	1.07	0.94
0.1	....	1.01	0.99	....	1.01	0.99	....	1.02	0.98

* x has the same meaning as in Tables 4 and 18.

[1] At frequencies such that H does not differ greatly from unity, the formula holds even when the spacing is less.

With closely wound coils having very many turns of solid wire, the resistance ratio for any spacing between turns, even very close spacing, is

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = \alpha H + (\beta u_1 + \gamma u_2) G \left(\frac{d_0}{c}\right)^2 \tag{94}$$

where α, β, and γ depend on (d_0/c) and are given by Table 21, u_1 and u_2 depend upon the ratio of length to diameter of the coil, as in Table 19, and the remaining notation is as in Eq. (93).

At very high frequencies, Eqs. (93) and (94) reduce to the following:

For Eq. (93), assuming copper wire:

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = 38 \sqrt{f_{mc}} d_0 \left[1 + \frac{1}{2} u \left(\frac{d_0}{c}\right)^2\right] \tag{93a}$$

For Eq. (94), assuming copper wire:

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = H\phi \tag{94a}$$

Here d_0 is the wire diameter in centimeters. The factor ϕ is given in Table 22, and depends upon the ratio of coil length to diameter and upon the closeness of the winding.

When litz wire is used, the resistance ratio of single-layer coils of many turns not too closely spaced is

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = H + \left[k + u \left(\frac{d_0}{c}\right)^2\right] \left(\frac{d_s}{d_0}\right)^2 n^2 G \tag{95}$$

where d_0 = diameter of cable.
c = spacing between centers of adjacent turns.
d_s = diameter of individual strand of cable.
n = number of strands in cable.
k = constant depending on n, given in Table 23.
G and H are the same as above, except that x in Table 18 is evaluated for the *diameter of the individual strand.*

TABLE 22.—VALUES OF FACTOR ϕ IN EQ. (94a)

$\frac{d_0}{c}$	$\frac{b}{D} = \frac{\text{coil length}}{\text{coil diameter}}$											
	0	0.2	0.4	0.6	0.8	1.0	2	4	6	8	10	∞
1.0	∞	.....	.....	.....	.....	.....	....	....	....	....	....	3.41
0.9	18.2	17.5	16.1	14.6	13.2	11.9	8.02	5.27	4.39	3.96	3.78	3.11
0.8	6.49	6.32	5.96	5.57	5.23	4.89	3.91	3.20	3.04	2.97	2.92	2.82
0.7	3.59	3.53	3.43	3.29	3.17	3.07	2.74	2.61	2.51	2.51	2.50	2.52
0.6	2.36	2.35	2.32	2.29	2.26	2.23	2.16	2.15	2.14	2.16	2.16	2.22
0.5	1.73	1.74	1.75	1.75	1.75	1.76	1.77	1.85	1.85	1.86	1.86	1.93
0.4	1.38	1.39	1.41	1.42	1.44	1.45	1.49	1.56	1.57	1.59	1.60	1.65
0.3	1.16	1.19	1.21	1.22	1.22	1.24	1.28	1.34	1.34	1.35	1.36	1.39
0.2	1.07	1.08	1.08	1.09	1.10	1.10	1.13	1.16	1.16	1.17	1.17	1.19
0.1	1.02	1.02	1.03	1.03	1.03	1.03	1.04	1.04	1.04	1.04	1.04	1.05

TABLE 23

No. of strands, n	3	9	27	∞
k	1.55	1.84	1.92	2

Calculation of Copper Losses in Multilayer Coils.—Butterworth has also derived a formula for the a-c/d-c resistance ratio of multilayer coils having many turns not too compactly wound. For solid wire one has

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = H + \frac{1}{4}\left(\frac{Kbm}{D}\right)^2\left(\frac{d_0}{c}\right)^2 G \tag{96}$$

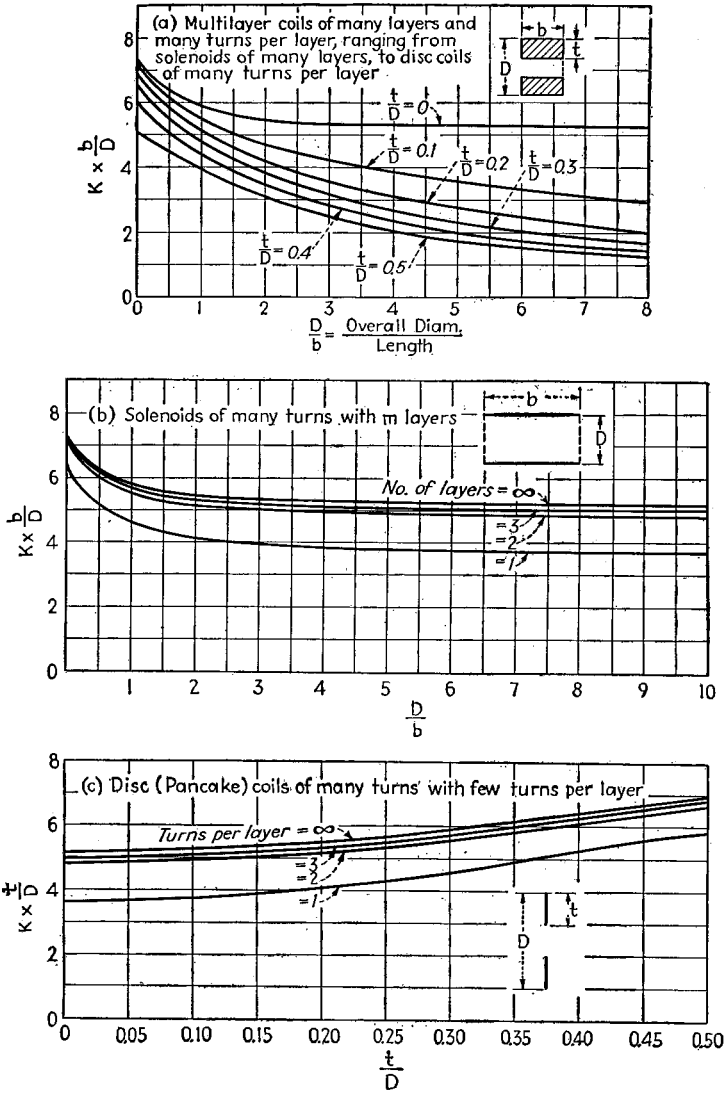


FIG. 51.—Values of constant K in Eqs. (96) and (97) for multilayer coils.

where K = a constant depending upon the type of winding, obtainable from Fig. 51.
b = axial length of winding.
D = over-all diameter.
m = number of layers.

$$d_0/c = \frac{\text{diameter of wire}}{\text{spacing between centers of adjacent turns in same layer}}.$$

H and G = constants given by Table 18.

For litz, one has

$$\frac{\text{A-c resistance}}{\text{D-c resistance}} = H + \left[k + \frac{1}{4}\left(\frac{Kbm}{D}\right)^2 \left(\frac{d_0}{c}\right)^2\right]\left(\frac{d_s}{d_0}\right)^2 n^2 G \qquad (97)$$

The notation in Eq. (97) is the same as in Eqs. (95) and (96).

Equations (96) and (97) apply to all shapes of coils, ranging from solenoids in one extreme ($t/D \to 0$) to disk or pancake coils in the other extreme ($D/b \to \infty$). The only restriction is that there must be many turns and these must be reasonably well spaced (*i.e.*, not too compactly wound). In the particular case of a single-layer solenoid, the value of K is such that Eqs. (96) and (97) reduce to Eqs. (93) and (95).

Calculation of Copper Losses in Toroidal Coils.—The resistance ratio of toroidal coils is given by Eq. (96) when solid round wire is used, and by Eq. (97) when litz wire is employed, provided that the following modified definitions are used.[1]

D = over-all diameter of toroid.

$b = \pi D$ = outer periphery of winding section.

c = spacing between centers of adjacent turns of same layer at outer periphery.

K = a constant given, depending upon the type of toroid and the proportions, as given in Table 24.

TABLE 24.—VALUES OF K FOR SINGLE-LAYER TOROIDS

Circular cross section		Rectangular cross section				
			K			
Diam. of winding / Over-all diam.	K	Radial depth / Over-all diam.	Axial length / Over-all diam.			
			0.125	0.250	0.375	0.500
0.05	2.110					
0.10	2.243					
0.15	2.409	0.1	2.25	2.25	2.26	2.26
0.20	2.602	0.2	2.65	2.67	2.69	2.70
0.25	2.913	0.3	3.37	3.47	3.54	3.58
0.30	3.327	0.4	5.25	5.68	5.96	6.14
0.35	3.978					
0.40	5.180					
0.45	8.341					

In the case of single-layer toroids wound with conductor of the type illustrated in Fig. 52, the Q obtained with optimum proportions for Fig. 52*a* is[2]

$$Q \text{ for optimum } \frac{r_2}{r_1} = \frac{84.3\sqrt{f}\, r_2}{\left(1 + 0.556\frac{r_2}{h}\right)} \qquad (98)$$

[1] S. Butterworth, The High Frequency Resistance of Toroidal Coils, *Exp. Wireless and Wireless Eng.*, Vol. 6, p. 13 January, 1929.

[2] Terman, *loc. cit.* See also G. Reber, Optimum Design of Toroidal Inductances, *Proc. I.R.E.*, Vol. 23, p. 1056, September, 1935.

For a toroid of circular cross section (Fig. 52*b*),

$$Q \text{ for optimum } \frac{r_2}{r_1} = 52.1b\sqrt{f} \tag{99}$$

The dimensions in these equations are in centimeters, and the frequency f is in megacycles.

Practical Usefulness of Formulas for Calculating Copper Losses.—Equations (93) to (99) take into account only the copper losses resulting from skin and proximity effects. At frequencies somewhat less than that giving maximum Q, these losses represent nearly the entire loss of energy in the coil. At higher frequencies, dielectric losses become important and lower the Q.[1]

The chief value of the formulas for calculating losses is in indicating the relative effect that such factors as size, shape, wire size, and other proportions have upon

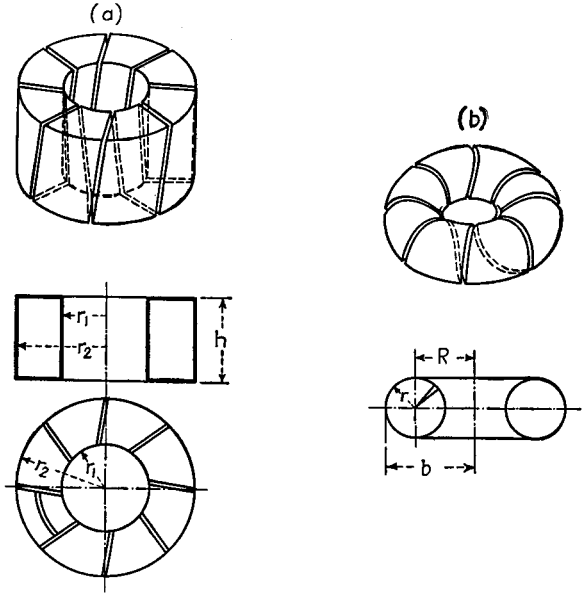


FIG. 52.—Toroidal coils wound with strip conductor that is curved to follow the surface of revolution formed by the toroid.

the coil Q. These formulas will satisfactorily compare the merit of different proposed designs, and indicate which are most promising, even under conditions where the absolute accuracy of the formulas is not particularly high.

The formulas are especially useful in indicating the size of wire for winding a given form. In the case of a single-layer solenoid in which Eq. (93*a*) applies, the wire diameter giving maximum Q when the length, diameter, and number of turns is fixed is

$$\frac{\text{Optimum diameter}}{\left\{\begin{matrix}\text{Spacing between centers}\\ \text{of adjacent turns}\end{matrix}\right\}} = \sqrt{\frac{2}{\mu}} \tag{100}$$

where μ is given by Table 19.

[1] A comparison of calculated and observed coil losses is given by W. Jackson, Measurements of the High-frequency Resistance of Single-layer Solenoids, *Jour. I.E.E.*, Vol. 80, p. 844, 1937; see also Wireless Section, *I.E.E.*, Vol. 12, p. 133, June, 1937.

20. Distributed Capacity of Coils.—The voltage difference that exists between different parts of a coil produces an electrostatic field in the air and in the dielectric near the coil. The effect of the resulting storage of electrostatic energy upon the coil behavior is to a good approximation equivalent to the effect produced by a small capacity shunted across the terminals of the coil. Such a hypothetical shunting capacity is termed the *distributed capacity* of the coil. Under practical conditions, there are also capacities between the coil terminals, and between lead wires, which increase the total effective shunting capacity.[1]

The distributed capacity of an isolated single-layer solenoid is approximately proportional to the diameter of the coil and decreases slowly with increasing length. The distributed capacity is largely independent of the number of turns if there are a considerable number of turns and these are not too closely spaced. Very close spacing increases the capacity somewhat, however, particularly if there are only a few turns.

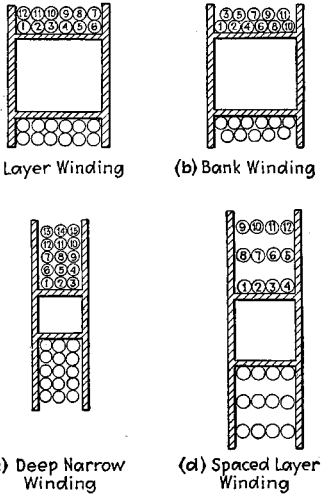


FIG. 53.—Several types of multilayer windings. The numbers in the drawing indicate the order of the turns.

Multilayer coils tend to have higher distributed capacity than do single-layer coils. In order to avoid excessive distributed capacity, the winding of a multilayer coil must be arranged so that the difference of potential between adjacent turns is only a small fraction of the total voltage applied to the coil. Thus a simple two-layer winding, as shown in Fig. 53*a*, will have very high distributed capacity, whereas with a winding in which the turns are wound as indicated in Fig. 53*b* (bank winding), the distributed capacity will be only moderate. Other means of minimizing distributed capacity consist of employing a narrow winding of many layers with few turns per layer, as shown in Fig. 53*c*, or the use of spaced layers, as in Fig. 53*d*. The common universal or honeycomb-type multilayer coil has a reasonably low distributed capacity, since it is equivalent to a coil of many layers with few turns per layer.

Dielectric in the field of a coil, such as the form upon which the coil is wound, the insulation of the wire, etc., increase the distributed capacity. Metal objects such as a shield, metal panel, etc., near a coil increase the distributed capacity, particularly if the distance from the coil is of the same order of magnitude as the coil dimensions or is less.

Effects of Distributed Capacity.—The presence of distributed capacity introduces losses, because some of the electrostatic energy storage involves solid dielectrics, which are never perfect. Dielectric losses have an effect equivalent to adding a resistance in series with the inductance coil according to the equation

$$\left.\begin{array}{l}\text{Equivalent series}\\ \text{resistance representing}\\ \text{dielectric losses}\end{array}\right\} = \tau\omega^3 L^2 C \qquad (101)$$

where $\omega = 2\pi \times$ frequency.

L = true coil inductance.

[1] The effective distributed capacity will also depend to some extent upon the current distribution in the coil, and will in general be larger when the coil is shunted with a large external tuning capacity than when the coil is resonated with its self-capacity.

C = distributed capacity.
τ = power factor of distributed capacity.

The resistance given by Eq. (101) is a part of the coil resistance R in Fig. 54*a*. Experiments indicate that the dielectric losses are not over 20 per cent of the total loss in the working range of the coil, unless the form or wire insulation has absorbed moisture.[1] When hygroscopic forms are used, and when the wire has cotton insulation, it is advisable to impregnate the coil with wax, shellac, or lacquer to prevent the absorption of moisture.

The presence of the distributed capacity causes a partial resonance that modifies the apparent resistance and reactance of the coil as viewed from the terminals. The apparent series resistance L_{eq} and R_{eq} that appear to be present when the coil is

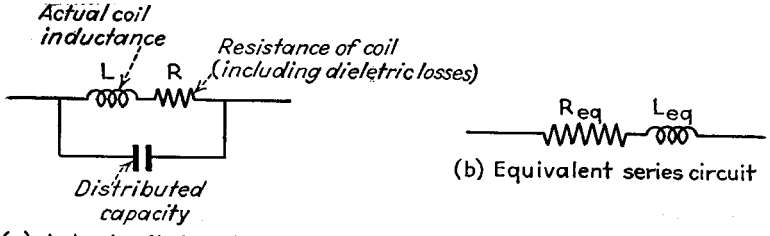


FIG. 54.—Actual circuit of coil with distributed capacity, together with equivalent series circuit.

viewed from its terminals (see Fig. 54*b*) are related to the true resistance and inductance according to the equations

$$\left.\begin{array}{l}\text{Apparent inductance } L_{eq} \text{ of} \\ \text{coil with distributed capacity}\end{array}\right\} = \frac{L}{1-\gamma^2} \tag{102}$$

$$\left.\begin{array}{l}\text{Apparent resistance } R_{eq} \text{ of} \\ \text{coil with distributed capacity}\end{array}\right\} = \frac{R}{(1-\gamma^2)^2} \tag{103}$$

where L and R are the true coil inductance and resistance, respectively, and γ is the ratio of actual frequency to the frequency at which the coil inductance would be resonant with the distributed capacity.

21. Temperature Coefficient of Inductance.[2]—The inductance of a coil varies with temperature as a result of (1) changes in dimensions with temperature and (2) changes in current distribution in the wire as a result of the change of wire resistivity with temperature modifying the skin and proximity effects.

The relation between the dimensions of a coil and the temperature depend upon several factors. In the ideal case of a freely suspended coil expanding according to the coefficient of linear expansion of the metal, the temperature coefficient of inductance will be the same as the temperature coefficient of expansion of the metal. Actually, however, the mechanical strains existing in the conductor ordinarily cause added changes in dimensions, with the result that the temperature coefficient of such a coil is

[1] W. Jackson, Dielectric Losses in Single-layer Coils at Radio Frequencies, *Exp. Wireless and Wireless Eng.*, Vol. 5, p. 255, May, 1928.

[2] H. A. Thomas, The Stability of Inductance Coils for Radio Frequencies, *Jour. I.E.E.*, Vol. 77, p. 702, 1935; also Wireless Section, *Jour. I.E.E.*, Vol. 11, p. 44, March, 1936; D. A. Bell, Temperature Coefficient of Inductance, *Wireless Eng.*, Vol. 16, p. 240, May, 1939; H. A. Thomas, The Dependence on Frequency of the Temperature-coefficient of Inductance of Coils, *Jour. I.E.E.*, Vol. 84, p. 101, 1939; also, Wireless Section, *Jour. I.E.E.*, Vol. 14, p. 19, March, 1939; E. B. Moullin, The Temperature Coefficient of Inductances for Use in a Valve Generator, *Proc. I.R.E.*, Vol. 26, p. 1385, November, 1938; Janusz Groszkowski, The Temperature Coefficient of Inductance, *Proc. I.R.E.*, Vol. 25, p. 448, April, 1937; Janusz Groszkowski, Temperature Coefficient of Inductance, *Wireless Eng.*, Vol. 12, p. 650, December, 1935.

usually a number of times the coefficient of linear expansion. When the coil is wound on a form instead of being freely suspended, the changes of dimension tend to be controlled by the form, but even here it is found that the temperature coefficient of inductance will normally be considerably greater than the expansion coefficient of either the form or the wire.

The change in wire resistivity with temperature modifies the skin and proximity effect and so causes the inductance to vary. Inasmuch as skin and proximity effects depend upon the frequency, this causes the temperature coefficient of inductance to depend likewise upon the frequency. The contribution to the temperature coefficient resulting from resistance changes is low at low frequencies when the skin and proximity effect are small, and is also low at very high frequencies, where these effects are very large. The change in inductance is maximum at a frequency between the extremes, where the skin and proximity effects are only moderate and the current distribution is particularly sensitive to resistance changes.

It is frequently found that the variation of coil inductance with temperature is noncyclical; *i.e.*, if the coil undergoes temperature cycles the inductance will not follow the temperature exactly, but, rather, will vary irregularly and will not return to its initial value when the temperature returns to normal. This is apparently the result of mechanical changes such as result from slippage of the wire over the form, and permanent changes in physical dimensions occasioned by the relieving of initial stress.

Practical Inductance Coils Having Low Temperature Coefficients.—Several methods have been suggested for designing coils to have a low and stable temperature coefficient. Probably the best arrangement is a single-layer solenoid with flat conductor wound directly on an insulating form having a low coefficient of expansion. The conductor can be a flat strip, but is preferably a conducting film deposited directly on the surface. If the radial depth of the flat conductor is very small and if the turns are well spaced axially, then the temperature coefficient of inductance will be of the same order of magnitude as that of the form, will be independent of frequency, and will be stable over repeated temperature cycles. Temperature coefficients of inductance of 5 to 15 parts per million per°C can be obtained in this way.

Another possible method of construction is to anneal the turns to relieve initial stress, and then mount them so that the coefficient of expansion in an axial direction is controlled by material having a relatively large temperature coefficient, while the coefficient of expansion in a radial direction is fixed by material having a low expansion coefficient. By suitable choice of the relative coefficients, it is then possible to obtain either zero or any desired temperature coefficient. Such coils must, however, be very carefully constructed or their behavior will not repeat on successive temperature cycles.[1]

Small inductances having low temperature coefficient can be constructed by employing a single-layer coil in which the conductor is a copper- or silver-plated invar (or nilvar) wire wound on a ceramic form having a low temperature coefficient of expansion. It is possible in this way to obtain an inductance coefficient of about one part in a million per°C.[2]

22. Practical Air-cored Coils. *Coils for Resonant Circuits in Radio Receivers.*—Coils for resonant circuits of radio receivers must be compact, and at the same time have a moderate to high Q. At frequencies in excess of 1,500 kc, the coils are practically always single-layer solenoids using spaced turns of solid wire wound on insu-

[1] For further information on such coils, see W. H. F. Griffiths, Recent Improvements in Air-cored Inductances, *Wireless Eng.*, Vol. 19, p. 8, January, 1942.

[2] S. W. Seeley and E. I. Anderson, UHF Oscillator Frequency-stability Considerations, *RCA Rev.*, Vol. 5, p. 77, July, 1940.

lating tubes. Coils for broadcast frequencies (550 to 1,500 kc) are usually also single-layer solenoids, although bank and universal windings are sometimes employed. At frequencies below 550 kc, the coils for resonant circuits are nearly always multilayer, usually of the universal or similar type.[1] Litz wire is generally preferred for these frequencies, although solid wire is often used because of lower cost.

Disk, toroidal, and other types of coils are practically never used for resonant circuits in receivers. At broadcast and lower frequencies, coils with finely subdivided iron cores are sometimes used (see Par. 27).

Coils for Resonant Circuits of Radio Transmitters.—Coils for transmitters are subjected to high voltage stresses and must handle large amounts of reactive circulating energy. This requires relatively large spacings in order to obtain the required voltage insulation and heat-dissipating ability. The size of coils for transmitters tends to be directly proportional to the power and inversely proportional to frequency.

In short-wave transmitters, it is customary to use coils wound with solid wire on ceramic forms when the power is low, and to use coils of copper tubing supported from the terminals for high power. In high-power broadcast transmitters, coils may be

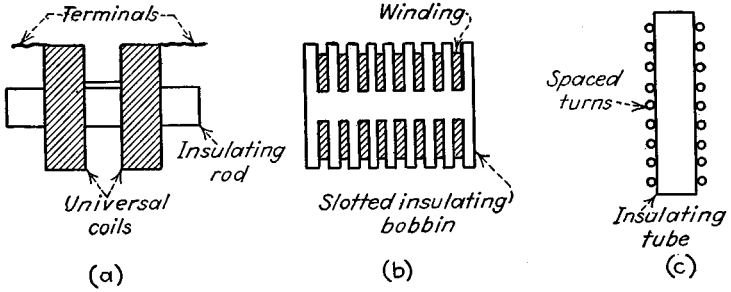


FIG. 55.—Typical examples of radio-frequency choke coils.

either large single-layer solenoids wound with copper tubing or edgewise wound strip. Single-layer disk or pancake coils using ribbon or solid conductors are also sometimes employed. When large powers are to be handled at still lower frequencies, it is customary to use a very large solenoid, either single or multilayer, with conductor consisting of large copper tubing, or litz, of large cross section.

Radio-frequency Choke Coils.—A radio-frequency choke coil is an inductance designed to offer a high impedance to alternating currents over the frequency range for which coil is to be used. This result is obtained by making the inductance high, and the distributed capacity low. Radio-frequency choke coils are commonly used at frequencies in the neighborhood of the natural resonant frequency of the coil. Thus the coils will normally offer a capacitive impedance to an applied voltage over at least a part of their working range.

The usual radio-frequency choke consists of two or more universal wound coils mounted some distance apart on an insulating rod, as shown in Fig. 55*a*. Another method of construction is to use a series of "pies" wound in deep narrow slots (Fig. 55*b*). A long single-layer solenoid, as in Fig. 56*c*, is sometimes used at very high frequencies.

The important characteristic of a radio-frequency choke coil is the variation of impedance with frequency. This depends upon the inductance and capacity of the

[1] The design of such multilayer windings is discussed by A. A. Joyner and V. D. Landon, Theory and Design of Progressive Universal Coils, *Communications*, Vol. 18, p. 5, September, 1938; A. W. Simon, Winding the Universal Coil, *Electronics*, Vol. 9, p. 22, October, 1936; L. M. Hershey, The Design of the Universal Winding, *Proc. I.R.E.*, Vol. 29, p. 442, August, 1941.

sections, the mutual inductance and mutual capacities between sections, the capacity to ground, and the lead arrangements. The behavior in a simple idealized case is illustrated in Fig. 56, in which the choke is considered as consisting of two inductances each with distributed capacities, but with mutual inductance and mutual capacities

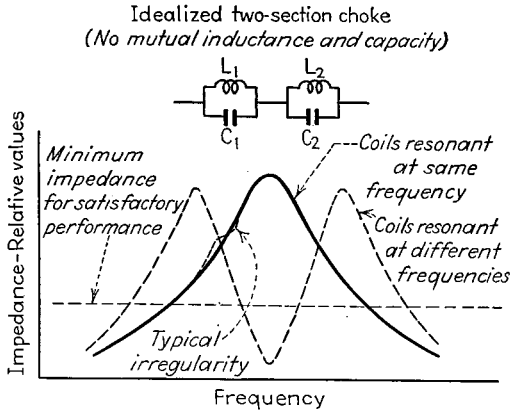


FIG. 56.—Impedance characteristics of idealized radio-frequency choke coils.

ignored. If the resonant frequencies of the two sections are the same, *i.e.*, if $L_1C_1 = L_2C_2$ in Fig. 56, then the choke acts as a single resonant circuit and has an impedance characteristic as shown by the solid line of Fig. 56. On the other hand, if the two sections have different resonant frequencies, the impedance will vary as a function of frequency in the manner shown by the dotted line. At a frequency some-

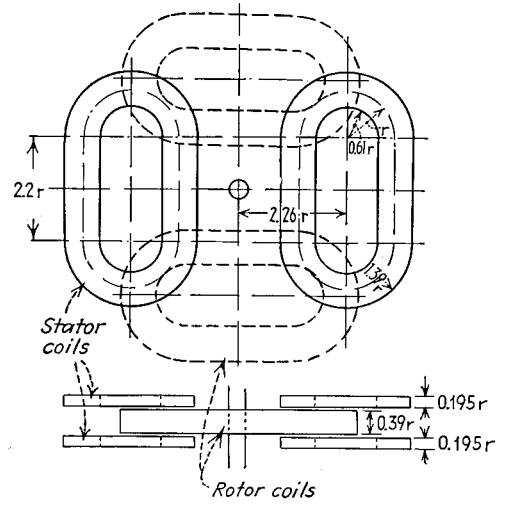


FIG. 57.—Brooks inductometer.

where between the two resonant frequencies, the choke has negligible impedance as a result of one section offering an inductive reactance of the correct magnitude to be in series resonance with the capacitive reactance of the second section. The actual situation existing in a practical radio-frequency choke coil is more complicated than

that of Fig. 56, because of mutual inductance and capacity effects and capacity to ground. The behavior is, however, of the same character, though more complex. By careful balancing of inductances, capacities, and couplings, it is possible to obtain performance approximating that corresponding to a single resonant frequency.[1]

Practical radio-frequency choke coils are designed either to have a single resonant frequency or to be proportioned so that the "holes" in the impedance characteristic occur at frequencies other than those for which the coil is to be used. Where the choke coil is to be shunted across the resonant circuit of a receiver that is tuned over a frequency band, it is necessary that the choke be very carefully proportioned to avoid even minor irregularities in the impedance curve, as shown by the dashed line in Fig. 56.

Variable Inductances.—The most common type of variable inductance is the ordinary variometer or inductometer. This consists of two systems of coils that are connected in series and rotated with respect to each other in order to vary the mutual inductance. The total inductance is $L_1 + L_2 \pm 2M$, and if the coefficient of coupling is large, the ratio of maximum to minimum inductance obtainable will be of the order of 10:1.

Two methods of constructing variable inductances are shown in Figs. 57 and 58. In the variometer, the rotating coil is wound on a spherical section, and mounted inside a corresponding fixed coil. In the Brooks inductometer, the same result is achieved in a slightly different way. The Brooks inductometer has the advantage that the calibration is not appreciably affected by axial displacements of the rotating coil and, further, that by proportioning the coils as shown in Fig. 57, the inductance varies linearly with the angle of rotation.[2]

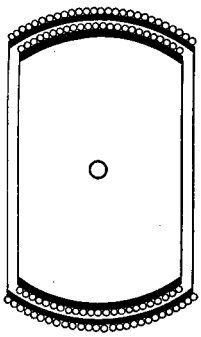


FIG. 58.—Variometer type of variable inductance.

The inductance of single-layer coils wound with relatively large wire can be varied by providing a wheel or trolley assembly that rides on the wire and provides a sliding contact that is adjusted by rotating the coil to vary the position of contact.[3] With proper mechanical design, such an arrangement provides a satisfactory and very fine control of the inductance.

Small variations in the inductance of coils used at radio frequencies can be obtained by rotating a copper disk or ring in the magnetic field of the coil. Such a copper object reduces the inductance in proportion to the magnetic flux lines that it intercepts, and so gives maximum inductance when the plane of the ring or disk is at right angles to the magnetic flux. The extra losses resulting from this method of controlling inductance are small, provided that the range of control is not excessive.

Fixed Inductances for Audio Frequencies.—Fixed inductances for such purposes as standards at audio frequencies are made of several types, as illustrated in Fig. 59. The conventional multilayer coil shown at Fig. 59*a* gives the largest inductance in proportion to the direct-current resistance, but it has the disadvantage of a large external field.[4] Inductances of this type will have minimum loss when proportioned so that $b = c$, and $c/a = 0.66$. The toroid has no external field but requires a very

[1] The design of radio-frequency choke coils for optimum performance is discussed by Harold A. Wheeler, The Design of Radio-frequency Choke Coils, *Proc. I.R.E.*, Vol. 24, p. 850, June, 1936; Herman P. Miller, Jr., Multi-band R-f Choke Coil Design, *Electronics*, Vol. 8, p. 254, August, 1935.

[2] For detailed design information on the Brooks inductometer see H. B. Brooks and F. C. Weaver, A Variable Self and Mutual Inductor, *Bur. Standards, Sci. Paper* 290.

[3] Paul Ware, A New System of Inductive Tuning, *Proc. I.R.E.*, Vol. 26, p. 308, March, 1939; L. E. Herborn, A Precise High-frequency Inductometer, *Bell Lab. Rec.*, Vol. 17, p. 351, July, 1939.

[4] The design of standard inductances of this type is thoroughly discussed by H. B. Brooks, Design of Standards of Inductance, and the Proposed Use of Model Reactors in the Design of Air-core and Iron-core Reactors, *Bur. Standards Jour. Research*, Vol. 7, p. 289, August, 1931.

large amount of wire in proportion to the inductance obtained, and so has high direct-current resistance. The double-D winding is intermediate in its characteristics between the toroid and the multilayer coil with respect to both external field and direct-current resistance. This type of coil has the advantage that the individual coils can be wound with a suitable number of turns, impregnated, taped, and mounted, all before final adjustment. The adjustment to give the exact value of inductance desired is then made by shifting the relative position of the two coils slightly with respect to each other before clamping them into final position.

Fixed inductances for use at radio frequencies should have low losses, good mechanical stability, and a minimum of dielectric loss. Litz wire is desirable for frequencies up to and including broadcast frequencies. The construction must be substantial, the form should be of good dielectric material, and the coil should be protected from moisture by suitable "dope." The distributed capacity should be low in order that the frequency at which the inductance is resonant with the distributed capacity will be as high as possible. In this way, the working range of frequencies over which the

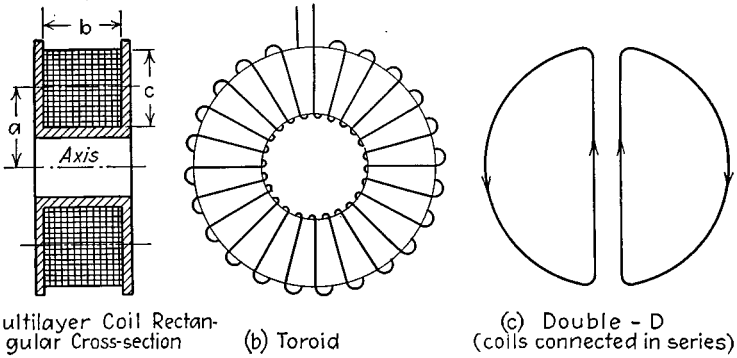


FIG. 59.—Types of fixed inductances used in audio-frequency work.

apparent and the true inductance are approximately the same will be as great as possible.

The temperature coefficient of inductance standards should be low. However, it is impractical to achieve a particularly low coefficient in multilayer and toroidal coils, as there is no material for the coil form that has the requisite low temperature coefficient combined with other desirable characteristics, such as stability, machinability, etc. Wood impregnated with beeswax and certain grades of porcelain are to be preferred, while bakelite and hard rubber are not recommended because of their high coefficients of expansion.

COILS WITH MAGNETIC CORES

23. Fundamental Properties of Magnetic Substances.—The relationship between flux density and magnetizing force in the core of a coil when the core is of magnetic material that has been previously demagnetized is shown in Fig. 60*a*, and is termed the saturation or magnetization curve. When such a core is carried through a complete cycle of magnetization, as when the magnetizing force is supplied by an alternating current, the relationship between flux density and magnetizing force follows a loop, as illustrated in Fig. 60*b*. This is termed the hysteresis loop. The residual flux B_r remaining after a positive magnetizing force has been removed is termed the *retentivity* of the material. The negative magnetizing force H_c required to reduce the flux to zero is termed the *coercive force.*

Permeability and Incremental Permeability.—The permeability of a magnetic material is defined as the ratio B/H, where B is the flux density in lines per square centimeter (gauss) and H the magnetizing force in oersteds (1 ampere turn per centimeter = 1.256 oersteds). With these units the permeability of air is unity. The

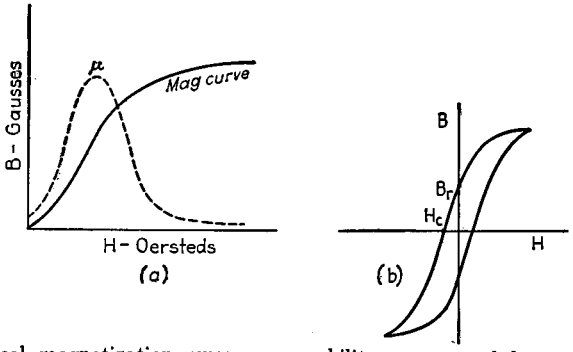


FIG. 60.—Typical magnetization curve, permeability curve, and hysteresis loop of a magnetic material.

permeability μ depends upon the flux density, and is in general small at low flux densities, maximum at moderate values of B, and quite small when the core is saturated, as shown in Fig. 60*a*.

The permeability at very low magnetizing forces is termed the *initial permeability*, and is of considerable importance in communication equipment. With ordinary magnetic materials, the initial permeability is very small.

The permeability offered to an alternating magnetizing force superimposed upon a d-c magnetizing force is termed *incremental permeability*, apparent permeability, or a-c permeability. Under such conditions, the cycle of magnetization follows a small displaced hysteresis such as is illustrated in Fig. 61; and the permeability to the alternating magnetization is determined by the slope of the line joining the tips of this displaced hysteresis loop.[1] The incremental permeability increases with increasing alternating flux densities (up to the point at which the core is saturated) and decreases with increasing d-c magnetization. The behavior in a typical case is illustrated in Fig. 62.

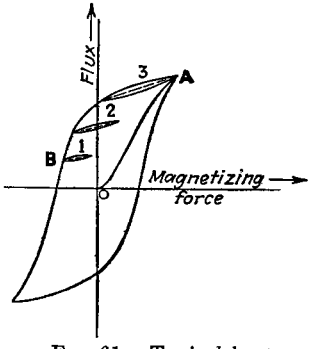


FIG. 61.—Typical hysteresis loop showing displaced minor hysteresis loops 1, 2, and 3, obtained when a small alternating current is superimposed on a direct-current magnetization. The effective permeability to the superimposed alternating current is proportional to the slope of the line joining the tips of the displaced hysteresis loop.

When both alternating and direct magnetizing forces are acting simultaneously, the permeability offered to the d-c magnetization in the low or moderate flux-density range is increased by a small amount of superimposed a-c excitation, and decreased by a large amount.[2] With large d-c magnetizations, the permeability to d-c is reduced by a superimposed alternating magnetization.

[1] Thomas Spooner, Permeability, *Trans. A.I.E.E.*, Vol. 43, p. 340, 1923.

[2] R. F. Edgar, Silicon Steel with A-C and D-C Excitation, *Trans. A.I.E.E.*, Vol. 53, p. 318, February, 1934.

Core Loss.—When a core is subjected to an alternating magnetization, there is a dissipation of energy in the core as a result of hysteresis and eddy currents. The total power loss in core materials used in transformers is often expressed in curves,

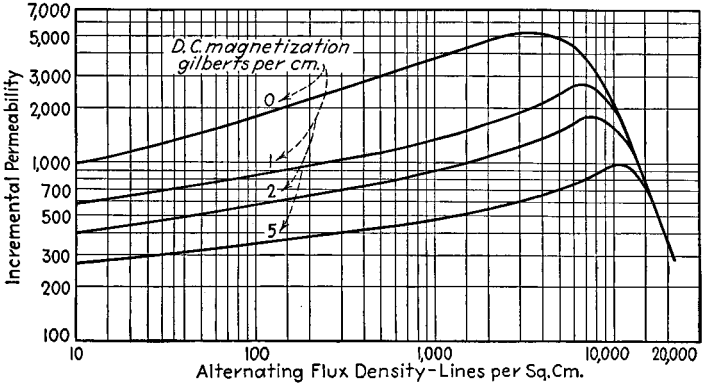


FIG. 62.—Incremental permeability of a typical sample of silicon steel as a function of alternating-current magnetization, for several values of superimposed direct current.

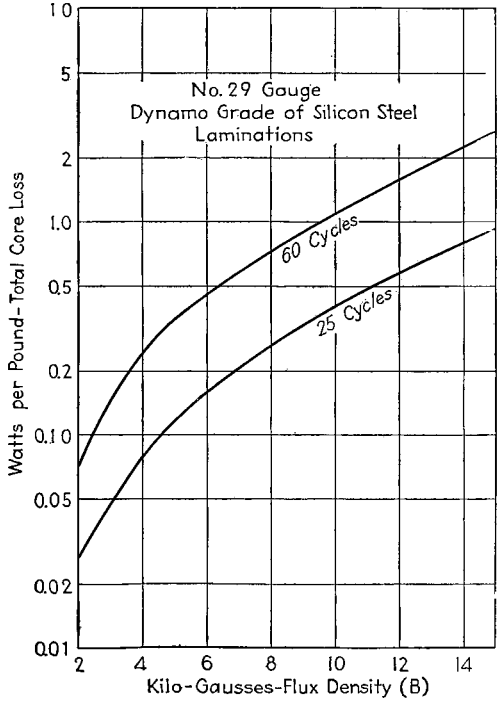


FIG. 63.—Core-loss curves of typical silicon steel sample.

such as those of Fig. 63, that give the watt loss per pound of material as a function of flux density for a given frequency. Such curves are of convenience in the design of power and audio transformers.

A coil having hysteresis and eddy-current losses can be represented by postulating series resistances R_h and R_e, respectively, in series with an inductance having no losses, together with a resistance R_c representing copper losses, as shown in Fig. 64*b*. In many cases, it is found convenient to represent the eddy-current losses as a resistance R_{sh} shunted across the inductance as in Fig. 64*c*, instead of as a series resistance. When the eddy-current losses are small, this shunting resistance is independent of frequency, whereas the series resistance R_e is proportional to the square of the frequency.

The eddy-current loss in a core is the result of currents induced in the core by the varying magnetic flux. The magnitude of the power loss per unit volume is proportional to the square of the frequency and to the square of the flux density, inversely proportional to the resistivity of the core material, and directly proportional to the square of the thickness (or diameter) of the individual core laminations (or particles). The eddy-current loss is not affected by d-c saturation of the core.

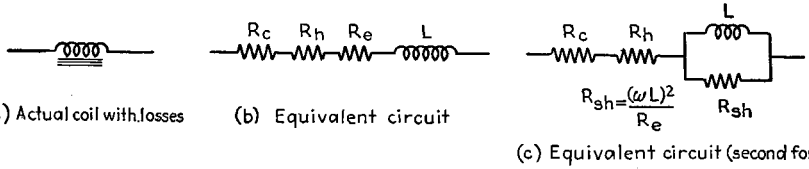


FIG. 64.—Equivalent circuits showing methods of representing hysteresis and eddy current losses in a coil having a magnetic core.

The equivalent series resistance R_e in Fig. 64*b* that can be considered as representing the eddy-current loss in a laminated core is[1]

$$R_e = 0.00658 \frac{t^2}{\rho_1} \mu_m f X_L \tag{104}$$

where t = thickness of the laminations, cm.
ρ_1 = resistivity of core material, microhm-cm.
μ_m = permeability of core corresponding to the maximum flux density produced by the magnetization ($\mu_m = 1$ for air).
X_L = inductive reactance of the coil with its magnetic core.

The value of the equivalent series resistance R_e for the case of a powdered core, on the assumption that the individual magnetic particles are of spherical shape is

$$R_e = \frac{0.00197 t^2}{\rho_1 k^{1/3}} \mu_m f X_L \tag{105}$$

where μ_m, f, ρ_1, and X_L have the same meaning as before, while t^2 is now defined as the mean square diameter of the spherical magnetic particles in cm and k is the fraction of the core volume that is occupied by the magnetic material. Equations (104) and (105) assume that the core is sufficiently subdivided so that magnetic skin effect is negligible.

Hysteresis is a form of magnetic friction, and for a given maximum flux density gives rise to a power loss per unit volume that is proportional to the frequency and to the area under the hysteresis loop. At flux densities sufficiently low that the permeability at the maximum flux density is not over 10 to 20 per cent greater than the initial permeability of the material, the hysteresis power loss per unit volume of most core materials is proportional to the fractional increase of the permeability over the initial value, and is approximately[2]

[1] V. E. Legg, Magnetic Measurements at Low Flux Densities using the Alternating Current Bridge, *Bell System Tech. Jour.*, Vol. 15, p. 39, January, 1936.

[2] This is a consequence of the Rayleigh hysteresis loop. See Legg, *loc. cit.*

$$R_h = 0.425 \left(1 - \frac{\mu_0}{\mu_m}\right) X_L \tag{106}$$

where μ_0 and μ_m are the initial permeability and the permeability corresponding to the maximum flux density, respectively, and X_L is the reactance of the coil with its iron core. With low magnetizing forces, the value of $\left(1 - \frac{\mu_0}{\mu_m}\right)$ is proportional to the magnetizing force, and hence to the peak flux density B_m. Under these conditions the equivalent resistance R_h representing hysteresis loss is proportional to B_m, and the power loss due to hysteresis[1] is proportional to B_m^3. When the flux density is moderately high, the hysteresis loss increases less rapidly with flux density. Thus, for silicon steel at flux densities such as used in ordinary transformers, the hysteresis power loss per volume is approximately proportional to $B_m^{1.6}$, corresponding to an equivalent resistance R_h that decreases slowly with increase in B_m.

Hysteresis loss is affected by d-c saturation and also by the presence of currents of other frequencies. The relations are relatively complicated but can, in general, be summarized as follows: With alternating flux densities below a critical value (13,000 to 14,000 gausses for silicon steel), the hysteresis loss rises with increasing d-c magnetization, when constant alternating flux density is assumed. With alternating flux densities greater than this critical value the hysteresis loss with constant alternating flux density decreases with superimposed d-c excitation.[2] When high- and low-frequency magnetizations are superimposed, the hysteresis loss to the high frequency is greater than it would be in the absence of the low-frequency magnetization, and the hysteresis loss of the latter is reduced. This action, termed *flutter effect*,[3] is of practical importance in loaded telephone lines carrying superimposed telegraph currents. The magnitude of the flutter effect is proportional to the area of the normal hysteresis loop and to the frequency of the high-frequency current.

Harmonics and Cross-modulation Caused by Magnetic Cores.[4]—The nonlinear relation existing between flux and magnetizing force in magnetic cores results in the production of harmonics when a sine-wave magnetizing force is applied to the coil, and also to cross modulation when two or more frequencies are applied.

With sine-wave excitation, the distortion components are in the form of odd harmonics, of which the third is the most important. With high maximum flux densities, the principal cause of such harmonic generation is the variation of permeability with flux density. With very low flux densities, hysteresis is the principal factor accounting for harmonic generation.

When the magnetizing force consists of a sine wave upon which a d-c current is superimposed, there results a second-harmonic distortion component having an amplitude depending upon the core material and upon the direct and alternating magnetizing forces involved.

When two alternating magnetizations are superimposed upon a core, the result is to produce combination frequencies (cross-modulation), the magnitude of which is dependent upon the amplitude of the magnetizing forces involved and upon the magnetic material. It has been determined experimentally that the third harmonic produced with a sinusoidal magnetizing force is an index of the tendency toward cross-modulation, which, accordingly, depends primarily upon hysteresis effects at low flux densities and upon the variation of permeability at high flux densities.

[1] There is also a component of hysteresis loss at low flux densities that is proportional to B_m^2. This is termed a "residual loss" and is of importance only at very low frequencies. For further discussion, see Legg, *loc. cit.*

[2] See Edgar, *loc. cit.*

[3] W. Fondiller and W. H. Martin, Hysteresis Effects with Varying Superimposed Magnetizing Forces, *Trans. A.I.E.E.*, Vol. 40, p. 553, 1921.

[4] Eugene Peterson, Harmonic Production in Ferromagnetic Materials at Low Frequencies and Low Flux Densities, *Bell System Tech. Jour.*, Vol. 7, p. 762, October, 1928.

Permalloy is many times better than silicon steel in freedom from harmonic generation and cross-modulation, while perminvar is many times better than permalloy.

24. Magnetic Materials Used in Communication.[1] *Silicon Steel.*[2]—The most widely used magnetic material in communication work is silicon steel. Silicon steel is available in various grades and types, according to the particular properties to be

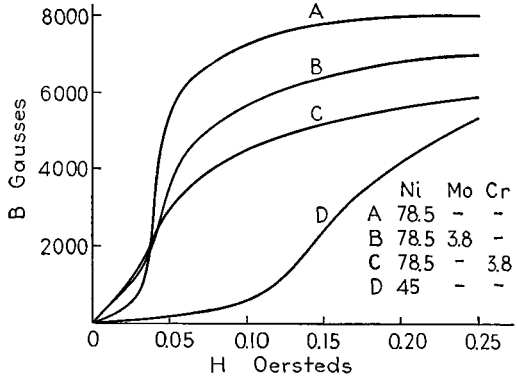


FIG. 65.—Typical magnetization curves of various nickel-iron alloys.

emphasized. It is used for power transformers, filter chokes, and as a core material for most audio-frequency, interstage, and output transformers.

Magnetic Alloys of Iron, Nickel, and Cobalt (Permalloys, Perminvar, and Permendur).[3]—Alloys of iron, nickel, and cobalt have some remarkable properties that are very useful in specialized applications. The characteristics of these alloys depend upon their composition and heat treatment, and a wide variety of properties are obtainable.

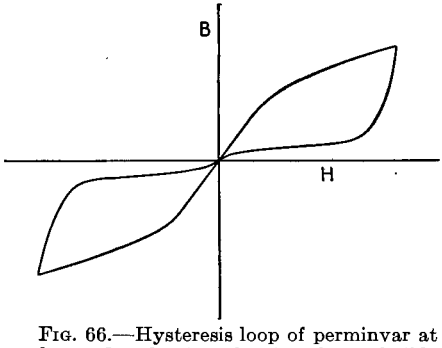


FIG. 66.—Hysteresis loop of perminvar at moderate flux density, showing the negligible coercive force and residual magnetization.

An important class of alloys of this type are those consisting of nickel and iron in the range 45 to 80 per cent nickel, with small amounts of chromium and molybdenum added in some cases. Such alloys go by such names as *Permalloy, Hipernik, Electric metal, Mumetal,* etc., according to the exact composition and the manufacturer. When properly heat-treated, they are all characterized by high maximum permeability and high initial permeability. The flux density at which saturation occurs is, however, the same as or even slightly less than for silicon steel. Typical magnetization curves are shown in Fig. 65. The addition of molybdenum to permalloy increases the resistivity, and hence reduces the eddy-current losses, without serious sacrifice in the low hysteresis loss which permalloy has. Molybdenum per-

[1] V. E. Legg, Survey of Magnetic Materials and Applications in the Telephone Systems, *Bell System Tech. Jour.*, Vol. 18, p. 438, July, 1939.

[2] C. H. Crawford and E. J. Thomas, Silicon Steel in Communication Equipment, *Elec. Eng.*, Vol. 54, p. 1348, December, 1935.

[3] G. W. Elmen, Magnetic Alloys of Iron, Nickel, and Cobalt, *Elec. Eng.*, Vol. 54, p. 1292 December, 1935.

malloy accordingly gives low losses at relatively high frequencies provided the flux densities are not too great.

Alloys of nickel, iron, and cobalt are termed perminvars, and when suitably heat-treated are characterized by substantially constant permeability and extremely low hysteresis loss at low flux densities. At medium flux densities, the hysteresis loop is severely constricted in the middle, as shown in Fig. 66, so that although the area of the hysteresis loop is not zero, the coercive force and retentivity are very small. At high flux densities, this constriction disappears, and the hysteresis loop has normal shape. Typical characteristics of perminvar are shown in Fig. 67.

Iron and cobalt in equal parts yield a useful magnetic alloy termed *permendur*. Its outstanding property is high permeability at high flux densities, corresponding to a saturation flux density higher than that of most materials. Permendur also has higher incremental permeability than most magnetic materials at relatively high d-c magnetizing forces. A magnetization curve of permendur is shown in Fig. 67.

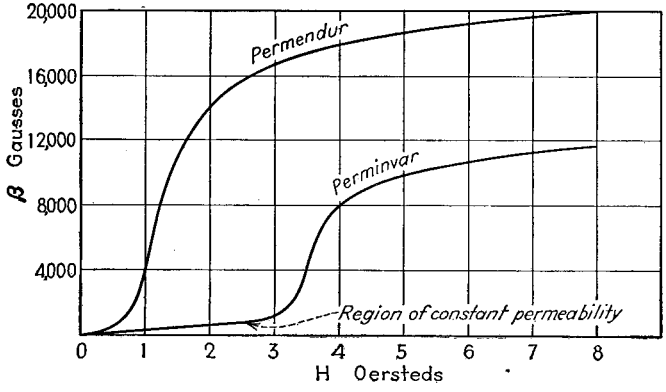


FIG. 67.—Typical magnetization curves of permendur and perminvar, showing the high flux density at saturation of the former and the constant permeability of low flux densities of the latter.

Permanent Magnets.[1]—The merit of a magnetic substance for use as a permanent magnet can be expressed in terms of the coercive force H_c the retentivity B_r, and the maximum value of the energy product BH of the demagnetization curve. The value of the energy product BH gives the relative energy available to sustain flux in the external magnetic circuit per unit volume of magnetic material. For most efficient utilization of the magnetic material, it is accordingly necessary to make the cross section of the magnet such that the desired flux will be obtained in the external circuit with a flux density in the permanent magnet corresponding to the flux giving maximum energy product. The demagnetizing force corresponding to the flux density existing in the magnet represents the magnetizing force per unit length of magnet available to force this flux through the reluctance of the magnetic circuit external to the permanent magnet. As a result of these factors, the total cross section of the permanent magnet is determined by the total flux required in the external circuit, and the length of the magnet is determined by the reluctance of the circuit.

The most desirable materials for permanent magnets are listed in Table 25, and have characteristics such as shown in Fig. 68 in the demagnetization portion of their

[1] C. S. Williams, Permanent Magnet Materials, *Elec. Eng.*, Vol. 55, p. 19, January, 1936; K. L. Scott, Magnet Steels and Permanent Magnets—Relationship among Their Magnetic Properties, *Bell System Tech. Jour.*, Vol. 11, p. 383, July 1932.

hysteresis curves.[1] Alnico and the related Alnic and Alnico-cu, are the best compromises between high coercive force and high retentivity. These alloys are inexpensive, but have the disadvantage that they must be cast to shape and finished by grinding, since they are so brittle that drilling and machining are impossible. The oxide material has the highest coercive force. It is also the least expensive and has an

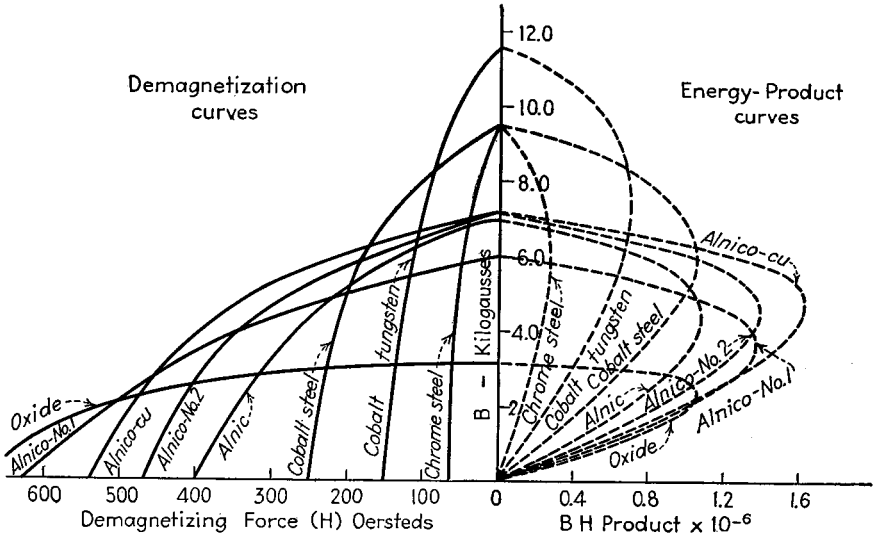


FIG. 68.—Demagnetization curves for permanent magnet materials.

unusually low specific gravity, but is even weaker mechanically than Alnico, since it consists of a sintered mass of oxides. Cobalt steel has a very high retentivity but low coercive force, is machinable, but is relatively expensive.

TABLE 25

Name	Component elements*	Coercive force, oersteds	Retentivity, gausses
Alnic	Al-Ni-Fe	100–600	8 000–4,000
Alnico	Al-Ni-Co-Fe	450–650	7,000–6,000
Alnico-cu	Al-Ni-Co-Cu-Fe	550	7,000
Cobalt steel	Co-Fe	150–250	10,000
Cobalt-tungsten	Co-W-Fe	30–150	13,500–11,000
Chrome	Cr-Fe	60	10,000
Oxide	CoO, Fe_2O_3, Fe_3O_4	300–1,100	6,000–1,500

* Ni = nickel; Al = aluminum; Fe = iron; Co = cobalt; W = tungsten

25. Calculation of Properties of Iron-cored Coils. *Inductance.*—The inductance of an iron-cored coil subjected to an alternating magnetization corresponding to a maximum a-c flux density B_{max} in the iron is as follows:

[1] Another permanent magnet material called Vicalloy has recently been announced, and promises to be of commercial importance. It is composed of vanadium, iron, and cobalt, and is described in *Bell Lab. Rec.*, Vol. 19, p. 36, September, 1940.

Dimensions in centimeters:

$$\text{Inductance in henrys} = L = \frac{1.256N^2}{\left(\frac{l_i}{\mu a_i} + \frac{l_a}{a_a}\right)} \times 10^{-8} \qquad (107a)$$

Dimensions in inches:

$$\text{Inductance in henrys} = L = \frac{3.19N^2 \times 10^{-8}}{\left(\frac{l_i}{\mu a_i} + \frac{l_a}{a_a}\right)} \qquad (107b)$$

where N = number of turns.
l_i = length of magnetic circuit in iron.
l_a = length of magnetic circuit in air.
μ = permeability of magnetic material evaluated at the appropriate maximum alternating flux density B_{max}.
a_i = cross section of iron (not including the area of insulation between laminations).
a_a = effective cross section of air gap.

The effective cross section a_a of the air gap will in general be somewhat greater than the cross section a_i of the iron because of fringing flux, and because the insulation between laminations makes the effective cross section of iron less than the cross section calculated from the external dimensions of the core.

The relationship between the maximum alternating flux density in the core and the rms value of a sinusoidal voltage is

$$\text{Voltage (rms)} = 4.44fNAB_{max} \times 10^{-8} \qquad (108)$$

where f = frequency in cycles per sec.
N = number of turns.
A = net area of core.
B_{max} = maximum flux density in core in lines per unit area.

Incremental Inductance.—The incremental inductance, sometimes called also the apparent, or a-c, inductance, is also given by Eq. (107) for any given alternating flux density B_{max} provided the permeability μ appearing in the equation is interpreted to mean the incremental permeability evaluated at the appropriate a-c and d-c flux densities from data such as are presented in Fig. 62.

The value of d-c flux density B_0 existing in an iron core having a winding of N turns carrying a d-c current I_0 can be determined from the fact that it is necessary to satisfy the relation.

$$1.256NI_0 = H_i l_i + B_0 \left(\frac{a_i}{a_a}\right) l_a \qquad (109)$$

The notation is the same as above, with the addition that H_i represents the magnetizing force in oersteds per centimeter required to support a d-c flux density of B_0 lines per square centimeter in the iron. The value of d-c flux density B_0 that satisfies Eq. (109) can be obtained by the cut-and-try process of calculating the value of the right-hand side of Eq. (109) for different assumed flux densities B_0 until a value is obtained that satisfies the equation. An alternative method is to employ the graphical solution shown in Fig. 69. Here B_0 is found as the intersection of the BH magnetization curve of the magnetic material, with a straight line crossing the horizontal axis at

$$H = 1.256NI/l_i$$

and the vertical axis at $B = 1.256NI/[l_a(a_i/a_a)]$.

Leakage Inductance.—The leakage inductance existing between two coils wound on the same core depends upon the geometry of the windings. The leakage induc-

tance is independent of the character of the core material or of the d-c magnetization, since the paths of the leakage flux are largely in air, and the core contributes negligible reluctance to them.

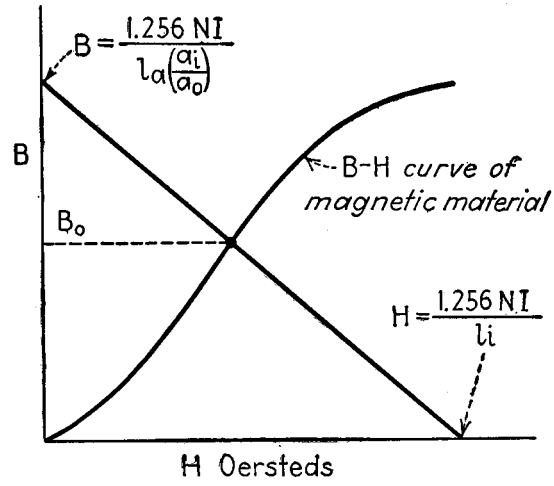


FIG. 69.—Graphical method of determining the flux density B_0 existing in a magnetic core having a series air gap and subjected to a known magnetizing force.

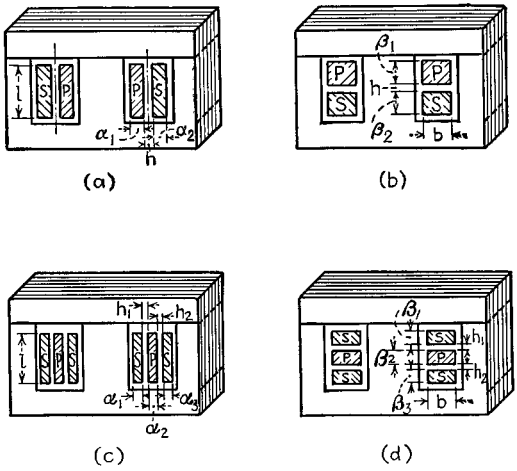


FIG. 70.—Typical coil arrangements in audio transformers.

When the secondary is wound on the outside of the primary, as shown in Fig. 70*a*, the total leakage inductance of the windings, referred to the primary side, is[1]

$$L'_p = \frac{2.9CN_p^2}{l}\left(h + \frac{\alpha_1 + \alpha_2}{3}\right) \times 10^{-8} \text{ henry} \qquad (110)$$

where C is the length of the mean turn in inches, N_p the number of primary turns, and

[1] The total leakage inductance referred to the secondary side is obtained by replacing N_p in Eq. (110) with the number of secondary turns N_s. The total leakage inductance L_p' referred to the primary

the remaining symbols have the meaning indicated in Fig. 70*a* (all measured in inches). This equation applies to coils wound on both circular and rectangular forms.

If the primary and secondary windings are arranged side by side as in Fig. 70*b*, the leakage inductance referred to the primary side is

$$L_p' = \frac{3.2CN_p^2}{b}\left(h + \frac{\beta_1 + \beta_2}{3}\right) \times 10^{-8} \text{ henry} \qquad (111)$$

where the notation is as in Fig. 70*b*, or is as above.

The leakage inductance can be reduced by dividing the primary and secondary coils into sections and interspersing these. Simple arrangements of this type are shown in Figs. 70*c* and 70*d*, and have leakage inductances referred to the primary that are

For Fig. 70*c*:

$$L_p' = \frac{0.72CN_p^2}{l}\left(h_1 + h_2 + \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}\right) \times 10^{-8} \text{ henry} \qquad (112)$$

For Fig. 70*d*:

$$L_p' = \frac{0.8CN_p^2}{b}\left(h_1 + h_2 + \frac{\beta_1 + \beta_2 + \beta_3}{3}\right) \times 10^{-8} \text{ henry} \qquad (113)$$

Sectionalization of windings, although greatly reducing the leakage inductance, simultaneously increases the winding capacities. The reduction achieved in leakage inductance is accordingly not all net gain.

In push-pull output transformers, the primary is frequently sectionalized to maintain symmetry. When this is done, the leakage reactance between each half of the primary and the secondary can be kept at a minimum by dividing the secondary into two identical parts, as in Fig. 71*a*, and connecting these in *parallel*.

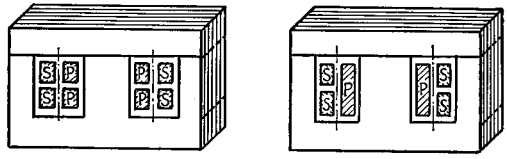


FIG. 71.—Transformers with divided windings.

Distributed and Other Capacities of Iron-cored Coils.—The distributed capacity of multilayer windings such as commonly wound on iron cores is made up primarily of the layer-to-layer capacities.[1] The turn-to-turn capacities can be neglected unless there are only a few layers. When the coil is wound in the normal manner so that the voltage between one end of the adjacent layers is twice the voltage per layer and the voltage between layers at the other end of the same layers is zero, the equivalent distributed capacity C_s that can be considered as shunted across the terminals of the

side is $L_p'' + L_s''/n^2$, where n is the step-up ratio of the transformer and L_p'' and L_s'' are the leakage inductances of the individual primary and secondary windings, respectively.

In the textbooks where the leakage inductance for Fig. 70*a* is calculated for an idealized case, the numerical constant in the equation is found to be 3.2 instead of 2.9. This latter value is obtained by introducing an empirical modifying factor of 0.9 to give more accurate agreement with observed results, as recommended by J. G. Story, Design of Audio-frequency Input and Intervalve Transformers, *Wireless Eng.*, Vol. 15, p. 69, February, 1938.

[1] Glenn Koehler, The Design of Transformers for Audio-frequency Amplifiers with Preassigned Characteristics, *Proc. I.R.E.*, Vol. 16, p. 1742, December, 1928.

winding is

$$C_s = \frac{0.30PlK}{dT} \mu\mu f \tag{114}$$

where P = mean circumference of coil turns, in.
l = axial length of coil, in.
d = distance between copper of adjacent layers, in.
T = number of layers.
K = average dielectric constant of insulation between layers, in. For paper-insulated wax-impregnated wire, K is approximately 3.

It is to be noted that Eq. (114) gives the distributed capacity of the coil referred to its own terminals. The equivalent value of the same distributed capacity referred to

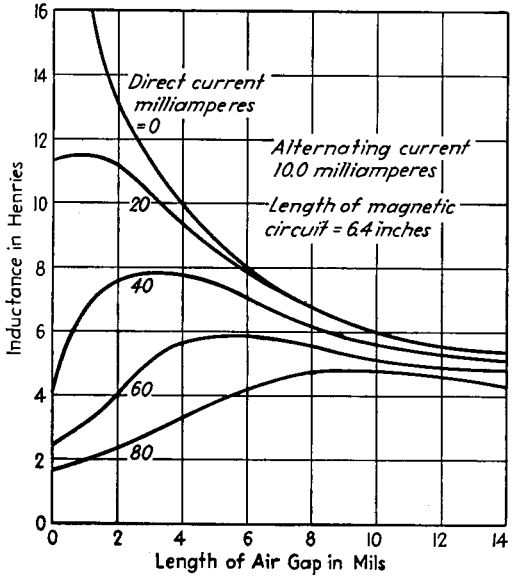


FIG. 72.—Example of incremental inductance as a function of air-gap length with various d-c magnetizing currents.

another winding on the core is obtained by converting in accordance with the inverse of the square of the turn ratio.

The coil capacity may be reduced by splitting the coil into two or more sections as in Fig. 71, each coil having the same number of layers as the single coil, but fewer turns per layer. In such an arrangement the capacity of each section can be determined by Eq. (114), and the capacities of the various sections then considered as acting in series. Thus subdivision into two equal sections quarters the distributed capacity. Subdivision will, however, increase the capacity of the winding to ground, and so may not improve the performance of the coil.

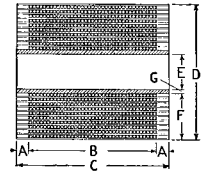
26. Design of Coils and Transformers with Magnetic Cores.—The design of coils with magnetic cores is normally carried out by a cut-and-try process. A typical procedure is to start by assuming a core cross section on the basis of previous experience. The space available for winding is then allotted to the various coils to be wound on the core, and a tentative selection is made of wire size and winding details in accordance with the insulation required, current to be carried, flux densities desired,

the space available, etc. The performance of the resulting reactor or transformer is then calculated, and the tentative design modified as required.

Winding Data.—The coils of chokes, transformers, etc., using magnetic cores are normally layer-wound by machine. The insulation employed depends upon the size of wire and the voltage between turns and between layers. In the case of small coils using relatively large wire, cotton-covered insulation is often sufficient. With very small wire, such as is used in audio-frequency transformers, the bulk of cotton or silk insulation is proportionately excessive, and the coils usually use enameled wire with paper insulation between layers in many cases.

TABLE 26*a*.—TURNS POSSIBLE PER SQUARE INCH OF WINDING SPACE, USING ONE LAYER OF PAPER OF PROPER THICKNESS AS SHOWN
Table Based on Average Commercial Practice

B & S gage	Plain enamel	Layer insulation	S.C.E.	D.C.C.	S.C.C.	S.S.E.	S.S.C.	D.S.C.
10	70	.010 in. Kraft	56	56	64			
11	86	.010 in. Kraft	72	64	72			
12	105	.010 in. Kraft	90	90	90			
13	126	.010 in. Kraft	111	111	120			
14	162	.010 in. Kraft	143	125	143			
15	191	.010 in. Kraft	180	143	168	Sq in. of winding space = $B \times F$		
16	238	.010 in. Kraft	208	180	208			
17	310	.007 in. Kraft	255	238	270			
18	396	.007 in. Kraft	304	288	340			
19	473	.007 in. Kraft	378	340	396			
20	621	.005 in. Kraft	460	418	504			
21	750	005 in. Kraft	572	504	621			
22	952	.005 in. Kraft	696	621	750			
23	1,140	.005 in. Kraft	832	725	980			
24	1,596	.0022 in. Glassine	1,120	928	1,258			
25	1,974	.0022 in. Glassine	1,330	1,054	1,517	1,596	1,800	1,517
26	2,392	.0022 in. Glassine	1,596	1,258	1,800	1,974	2,200	1,800
27	3,009	.0022 in. Glassine	1,886	1,440	2,156	2,392	2,640	2,150
28	3,894	.0015 in. Glassine	2,397	1,720	2,646	2,964	3,355	2,538
29	4,745	.0015 in. Glassine	2,750	1,978	3,074	3,591	4,020	3,074
30	5,904	.0015 in. Glasine	3,300	2,254	3,648	4,340	4,884	3,648
31	7,189	.0015 in. Glassine	3,712	2,548	4,216	5,025	5,757	4,216
32	8,888	.0013 in. Glassine	4,414	2,860	4,891	6,068	6,864	4,891
33	11,172	.0013 in. Glassine	5,100	3,132	5,688	7,280	8,245	5,688
34	14,336	.001 in. Glassine	6,150	3,658	6,715	8,910	10,085	6,794
35	17,324	.001 in. Galsine	6,880	3,965	7,728	10,272	12,051	7,728
36	21,330	.001 in. Glassine	7,840	4,352	8,722	12,168	13,860	8,722
37	25,404	.001 in. Glassine				14,112	15,182	10,379
38	31,878	.001 in. Glassine				16,184	18,270	11,845
39	42,020	.0007 in. Glassine				19,140	22,594	14,022
40	51,168	.0007 in. Glassine				21,855	25,992	15,600
41	62,746	.0007 in. Glassine						
42	75,392	.0005 in. Glassine						
43	91,800	.0005 in. Glassine						
44	105,534	.0005 in. Glassine						



A—margins*
B—winding length
C—over-all length
D—outside diameter
E—inside diameter
F—build-up
G—tube

* Margins are determined by wire sizes. The following are the minimum allowances for commercial windings:

Wire gage:
18–29 1/8 in. each end
30–37 3/32 in. each end
38–44 1/16 in. each end

Data useful in the winding of coils are given in Tables 26*a* to 26*d*.[1]

Some Factors Involved in the Design of Inductances Carrying Direct Current.—There is a particular air gap for which the incremental or a-c inductance is maximum when the core material and dimensions, the number of turns, and the d-c magnetizing current are fixed. As the d-c magnetizing current is increased, the optimum air gap becomes longer and the corresponding incremental (a-c) inductance less. The behavior in a typical case is shown in Fig. 72.[2]

This best length of air gap can be obtained by plotting curves of LI^2/V as a function of NI/l for various percentage air-gap lengths, as shown in Fig. 73, where L is the

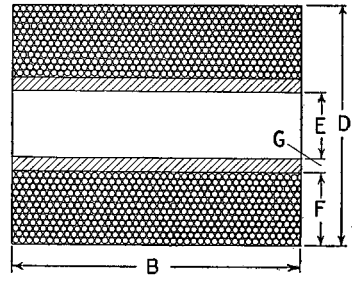
[1] Data from *Inca Bull.* 3, Phelps Dodge Copper Products Corporation, Inca Mfg. Division.

[2] H. M. Turner, Inductance as Effected by the Initial Magnetic State, Air-gap, and Superimposed Currents, *Proc. I.R.E.*, Vol. 17, p. 1822, October, 1929.

TABLE 26*b*.—RANDOM WOUND COILS SUCH AS DYNAMIC SPEAKER FIELDS USING ENAMELED WIRE

Table Based on Average Commercial Practice

B & S gage	Turns possible per linear in. of winding space*		Turns possible per sq in. of cross section	Ohms possible per cu in. of winding space
10	9		85	0.00716
11	10		106	0.0112
12	11		131	0.01733
13	12		160	0.02673
14	14		209	0.0440
15	15		251	0.0666
16	17	90%	319	0.1068
17	19		99	0.168
18	22		519	0.275
19	24		631	0.424
20	27		794	0.672
21	30		987	1.053
22	34		1,255	1.689
23	31		1,525	2.59
24	35		1,933	4.14
25	39		2,428	6.55
26	44		3,080	10.48
27	49		3,891	17.6
28	56		5,001	27.3
29	62		6,200	42.4
30	69		7,797	67.1
31	77	75%	9,800	106.2
32	86		12,210	166.9
33	98		15,740	271.5
34	109		19,520	425
35	121		24,680	677
36	136		31,280	1,082
37	150		38,500	1,681
38	170		49,350	2,720
39	192		63,400	4,370
40	243		67,500	5,910
41	274		85,800	9,370
42	304	85%	104,900	14,410
43	340		130,900	23,300
44	369		153,850	33,300



Linear in. of winding space = B

Sq in. of winding space = $B \times F$

Cu in. of winding space = $[(D^2 \times 0.7854) - (E + 2G)^2 \times 0.7854]B$

No insulation used between layers on random wound coils.

* Percentages give normal space factor.

incremental (a-c) inductance, I the direct current in the winding, V the volume of the core, N the number of turns, and l the length of the magnetic circuit.[1] The envelope of these curves, shown dotted in Fig. 73, gives the maximum incremental inductance per unit volume obtainable for a particular current I, with any given d-c magnetizing force per unit length NI/l, and corresponds to a percentage air gap given

[1] C. R. Hanna, Design of Reactances and Transformers which Carry Direct Current, *Trans. A.I.E.E.*, Vol. 46, p. 155, 1927. An alternative method of design is given by John Minton and Ioury G. Maloff, Design Methods for Soft Magnetic Materials in Radio, *Proc. I.R.E.*, Vol. 17, p. 1021, June, 1929.

by the particular curve tangent to the envelope at the point in question. Information such as shown in Fig. 73 is particularly useful where a large number of cores are to be designed employing the same material, since then it is worth while to calculate or measure the special curves required for the particular material involved. Where only a single core is to be designed, it is simpler to determine the optimum air gap by assuming various values and calculating the behavior of each case.

TABLE 26c.—OHMS POSSIBLE PER CUBIC INCH IN LAYER INSULATED COILS*
Table Based on Average Commercial Practice

B & S gage	Plain enamel	S.C.E.	D.C.C.	S.C.C.	S.S.E.	S.S.C.	D.S.C.
10	0.00584	0.00466	0.00466	0.00534			
11	0.00904	0.00756	0.00672	0.00756			
12	0.0139	0.0119	0.0119	0.0119			
13	0.0210	0.01852	0.01852	0.02003			
14	0.0341	0.0301	0.0263	0.0301			
15	0.05075	0.04775	0.03895	0.0446			
16	0.0797	0 0696	0.06025	0.0696			
17	0.1305	0.1074	0.1002	0.1136			
18	0.21	0.1614	0.153	0.1805			
19	0.318	0.254	0.2285	0.266			
20	0.526	0.3895	0.354	0.426			
21	0.800	0.610	0.5375	0.6625			
22	1.275	0.936	0.835	1.09			
23	1.935	1.412	1.23	1.613			
24	3.41	2.395	1.983	2.685			
25	5.325	3.59	2.84	4.09	4.31	4.86	4.09
26	8.140	5.43	4.28	6.125	6.725	7.48	6.125
27	12.79	8.10	6.18	9.25	10.82	11.35	9.25
28	21.05	12.95	9.30	14.33	16.05	18.15	13.71
29	32.81	18.80	13.50	21.00	24.50	27.43	20.99
30	50.90	28.40	19.40	31.42	37.35	42.05	31.40
31	77.95	40.25	27.65	45.70	54.50	62.50	45.70
32	121.4	60.40	39.10	66.80	82.95	93.75	66.80
33	192.5	87.95	54.15	98.05	125.60	142.20	98.05
34	312.3	133.75	79.50	146.25	194.25	219.00	147.75
35	475	188.65	108.85	212.00	280.65	330.15	212.00
36	729	271.50	150.50	301.50	421.00	479.00	301.50
37	1,110	.....	.....	.....	616	662	452
38	1,750	.....	.....	.....	882	1,002	649
39	2,920	.....	.....	.....	1,330	1,568	974
40	4,480	.....	.....	.....	1,913	2,269	1,365
41	6,860						
42	10,390						
43	16,380						
44	22,850						

* This table based on using one layer of paper insulation of proper thickness.

In the design of filter reactors, the allowable d-c voltage drop in the winding is frequently the limiting factor rather than the heating of the winding. Under these conditions, a larger core and winding will be required to obtain a desired inductance than would otherwise be the case.

Factors Involved in the Design of Small Transformers.—In the design of transformers, the first step is to determine the volt-ampere ratings of the windings. In the case of

TABLE 26*d*.—TURNS POSSIBLE PER LINEAR INCH OF WINDING SPACE IN LAYER INSULATED COILS*

Table Based on Average Commercial Practice

B & S gage	Plain enamel†	S.C.E.†	D.C.C.†	S.C.C.†	S.S.E.†	S.S.C.†	D.S.C.†
10	9	8	8	8			
11	10	9	8	9			
12	11	10	10	10			
13	12	11	11	12			
14	14	13	12	13			
15	15	14	13	14			
16	17	16	15	16			
17	19	17	17	18			
18	22 } 90 %	19 } 90 %	18 } 90 %	20 } 90 %			
19	24	21	20	22			
20	27	23	22	24			
21	30	26	24	27			
22	34	29	27	30			
23	38	32	29	35			
24	42	35	32	37			
25	47	38	34	41	42	45	41
26	52	42	37	45	47	50	45
27	59	46	40	49	52	55	49
28	66 } 89 %	51 } 89 %	43 } 89 %	54 } 89 %	57 } 89 %	61 } 89 %	54 } 89 %
29	73	55	46	58	63	67	58
30	82	60	49	64	70	74	64
31	91	64	52	68	75	81	68
32	101	69	55	73	82	88	73
33	114 } 88 %	75 } 88 %	58 } 88 %	79 } 88 %	91 } 88 %	97 } 88 %	79 } 88 %
34	128	82	62	85	99	106	86
35	142	86	65	92	107	117	92
36	158 } 87 %	92	68	98	117	125	98
37	174				126	134	107
38	198				136 } 87 %	145 } 87 %	115 } 87 %
39	220 } 86 %				145	158	123
40	246				155	171	130
41	274						
42	304 } 85 %						
43	340						
44	369						

* These figures are for paper layer insulated coils.

† The percentages represent the space factor under normal conditions.

small transformers used for heating filaments, etc., this is done on the assumption of 80 to 90 per cent efficiency and 90 per cent power factor in the primary and unity power factor on the secondaries. In transformers used with rectifiers, it is commonly necessary to increase the volt-ampere ratings, particularly of the secondary, because the utilization factor is appreciably less than unity. The next step in the design is to select a suitable core area. Experience with small transformers shows that, with ordinary silicon steel, the net area of iron in the core, in square inches, should be at least[1] $0.16\sqrt{\text{primary volt-amperes}}$. Where a conservative design is to be used, a somewhat larger core area than this minimum is recommended.

[1] The actual cross section of the core will be somewhat larger because of the insulation between laminations. The ratio of the area of iron to the gross area of the core, *i.e.*, the stacking factor, is commonly of the order of 0.9.

The third step in the design is to select a suitable maximum flux density, and then to calculate the turns per volt from the equation

$$\text{Turns per volt} = \frac{10^8}{4.44 f B_{max} A} \tag{115}$$

The notation is the same as in Eq. (108). The maximum flux density B_{max} should normally not exceed 75,000 lines per square inch for 60-cycle operation and 90,000 lines per square inch for 25-cycle operation. Lower flux densities will reduce the core losses and also the exciting current, but require more copper and iron.

The next step is the determination of the wire size and winding arrangements. It is customary to allow from 500 to 1,000 circular mils per ampere for transformers under 50 watts, and up to 1,500 to 1,750 circular mils per ampere for larger trans-

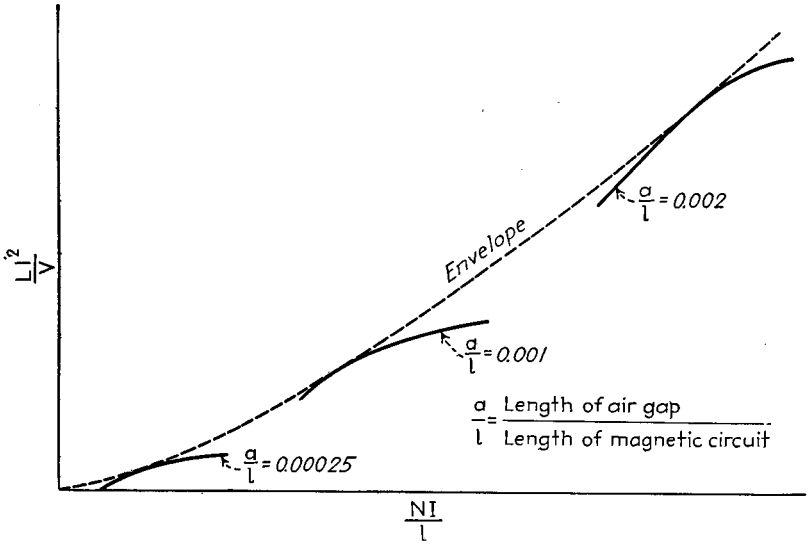


FIG. 73.—Graphical method of determining best air-gap length.

formers. The insulation between turns and layers, and between the winding and core are arranged in accordance with the voltages involved.

A core is now selected that will have sufficient window space to accommodate the winding to be used. The core is customarily composed of laminations of standard *E-I* punchings. This gives a shell-type core that is assembled in interleaved fashion to eliminate butt joints and minimize the air gap. Standard laminations of this type are available in a wide variety of shapes and in sizes large enough for power transformers of moderate power transmitters. With 60-cycle operation, the laminations are normally 29 gage (0.014 inches) silicon steel.

This completes the formal design of the transformer. The copper losses of the various windings can now be calculated from the currents and resistances of the windings, and the core loss can be evaluated from the volume of the core, the frequency, and the flux density, with the use of curves such as those of Fig. 63. The efficiency and the power that must be dissipated in the transformer can then be readily calculated. Voltage regulation can be determined from the resistances of the windings, and the leakage inductance as calculated by Eqs. (110) to (113). After the design has been completed and the important features of the performance cal-

culated, it is always desirable to examine the design as a whole. This may indicate that modifications will give an improved performance or a satisfactory performance at a lower cost.

27. Magnetic Cores in Audio- and Radio-frequency Coils.—The permeability of iron (and other magnetic materials) drops off slowly with increasing frequency, but the reduction is not great at broadcast and even short-wave frequencies. Thus tests with iron and nickel have indicated that the permeability does not drop to unity until the frequency is in the range 10^9 to 10^{11} cycles.[1] Magnetic cores can accordingly be employed in high-frequency coils. It is merely necessary that the core be subdivided sufficiently that the ratio R_e/X_L in Eqs. (104) and (105) are not too great.

If the core is not sufficiently subdivided (*i.e.*, if t^2f in Eqs. (104) and (105) is large), not only will the eddy-current losses be excessive, but also the magnetic flux produced by the eddy currents will be sufficient to modify the flux distribution over the cross section of the magnetic material. The net result is then to reduce flux density at the center of the laminations or particles, producing a form of *magnetic skin effect* or shielding that reduces the effective cross section of the core and lowers the inductance.[2]

Hysteresis does not introduce any special problems at high frequencies, since the ratio of hysteresis loss to inductive reactance is independent of frequency, *i.e.*, the power loss is proportional to the number of cycles per second.

Use of Laminated Cores at Audio and Radio Frequencies.—The silicon steel laminations of No. 29 gage (0.014 inches thick) used in 60-cycle transformers, also have reasonably low losses at the lower audio frequencies. Such laminations are also used in filter reactors for smoothing rectified 60-cycle currents, for the cores of audio-frequency input, interstage, and output transformers, and in some cases in wave filters having pass bands below 1,000 to 1,500 cycles. At higher frequencies it is necessary to use laminations thinner than 0.014 inches if excessive eddy current losses are to be avoided. Thicknesses as small as 2 mils have been used, and in this way magnetic cores can be obtained that have fairly low loss even at moderately high radio frequencies.[3] Thin laminations of this character were at one time used to a considerable extent in communication work, but have now been largely displaced by cores of compressed powdered magnetic material.

Coils Using Cores of Compressed Powdered Magnetic Material.[4]—The most effective way to employ magnetic material at the higher audio frequencies and at radio frequencies is to mix the magnetic material in the form of a very fine powder or dust

[1] For a discussion of the effect of frequency on permeability see G. W. O. Howe, Permeability at Very High Frequencies, *Wireless Eng.*, Vol. 16, p. 541, November, 1939.

[2] Analyses are available for calculating the extent of the magnetic shielding in the case of laminations of rectangular cross section. See K. L. Scott, Variation of Inductance of Coils Due to the Magnetic Shielding Effect of Eddy Currents in the Cores, *Proc. I.R.E.*, Vol. 18, p. 1750, October, 1930.

In practice there is a thin surface layer, normally about 1 mil thick, having lower permeability than the main body of the laminations. This surface film is apparently the result of the rolling operation used in producing the lamination, and is not affected by annealing. Its effect is important at high frequencies where the penetration of the flux is small, and at these frequencies the simple theory will accordingly fail to give correct results. Where the layer is important, it can be removed by etching.

For further discussion of magnetic skin effect, with particular reference to the oxide film layer, see E. Peterson and L. R. Wrathall, Eddy Currents in Composite Laminations, *Proc. I.R.E.*, Vol. 24, p. 275, February, 1936; M. Reed, An Experimental Investigation of the Theory of Eddy Currents in Laminated Cores of Rectangular Section, *Jour. I.E.E.*, Vol. 80, p. 567, 1937; also Wireless Section, *I.E.E.*, Vol. 12, p. 167, June, 1937; M. Reed, Losses in Ferromagnetic Laminae at Radio Frequencies, *Wireless Eng.*, Vol. 15, p. 263, May, 1938.

[3] Information on losses of thin silicon steel laminations at radio frequencies is given by Thomas Spooner, High Frequency Iron Losses, *A.I.E.E. Jour.*, Vol. 39, p. 809, September, 1920; Leon T. Wilson, The Behavior of Iron in Alternating Fields of between 100,000 and 1,500,000 cycles, *Proc. I.R.E.*, Vol. 9, p. 56, February, 1921.

[4] W. J. Polydoroff, Ferro-inductors and Permeability Tuning, *Proc. I.R.E.*, Vol. 21, p. 690, May, 1933; V. E. Legg and F. J. Given, Compressed Powdered Molyhdenum Permalloy for High-quality Inductance Coils, *Trans. A.I.E.E.*, Vol. 59, p. 1940; W. J. Shackelton and I. G. Barber, Compressed Powdered Permalloy, *Trans. A.I.E.E.*, Vol. 47, p. 429, April, 1928; Buckner Speed and G. W. Elmen, Magnetic Properties of Compressed Powdered Iron, *Trans. A.I.E.E.*, Vol. 40, p. 596, 1921.

with an insulating binder, and then compress the combination into cores of the appropriate form. In such cores the insulating binder surrounds the magnetic particles, thereby giving a very complete subdivision of the core, and low eddy-current losses.

The size of the magnetic particles and the amount of magnetic material in proportion to insulating binder control the effective permeability of the resulting core and also the frequency at which the coil Q is highest. The smaller the particles the less the effective permeability, and also the higher the optimum frequency. The effective permeabilities used in practice range from 75 to 150 for coils designed to be optimum at

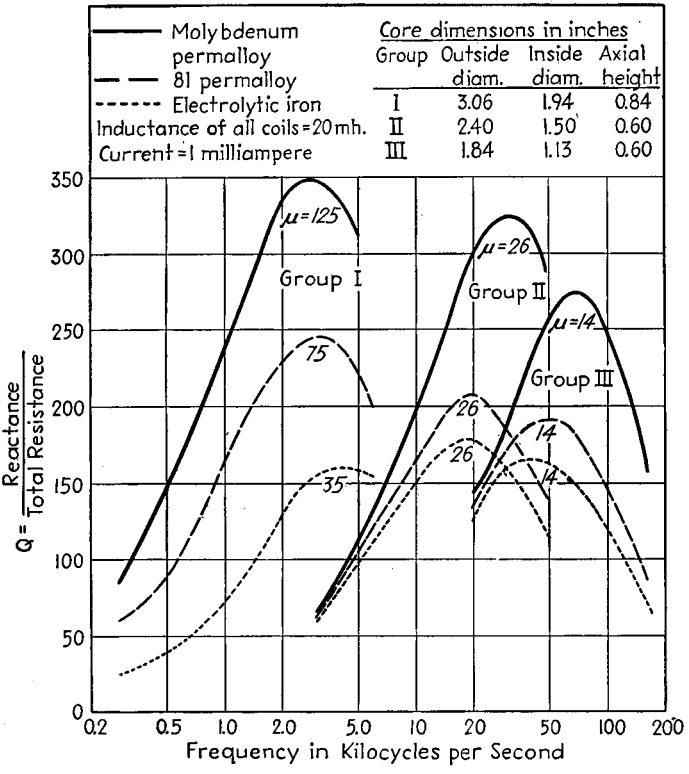


FIG. 74.—Variation of coil Q with frequency for several coils using toroidal dust cores.

audio frequencies, to values of 4 to 6 for coils intended for operation at frequencies of 500 to 1,500 kc.

The materials most frequently used in powdered cores are iron, permalloy, and molybdenum permalloy. Molybdenum permalloy has the lowest losses and hence gives the highest Q, while iron has the highest losses. The total core loss in any case will be quite small, however, provided that the core is used in its proper frequency range.

Coils employing powdered iron cores are often used in the intermediate-frequency and radio-frequency tuned circuits of radio receivers, up to and beyond the broadcast range.[1] For such radio-frequency application, the cores are commonly in the form of

[1] An excellent discussion of the properties of such cores and of the measurement of these properties is given by D. E. Foster and A. E. Newlon, Measurement of Iron Cores at Radio Frequencies, *Proc. I.R.E.*, Vol. 29, p. 266, May, 1941.

cylindrical slugs upon which the coil is wound. At intermediate (175 to 460 kc) and lower frequencies, the use of iron-dust cores makes it possible to realize a decidedly higher Q in a very small coil than is obtainable by the use of an air-cored coil of comparable size. In the case of intermediate-frequency transformers, magnetic cores also have the advantage that the resonant frequency can be controlled by adjusting the position of the core, thereby eliminating the trimmer condenser otherwise needed.

Toroidal coils using cores of compressed powdered molybdenum permalloy, permalloy, and iron-dust rings are widely used in telephone work for loading coils, in wave filters, etc. The losses of such coils are quite small, as shown by Fig. 74, with molybdenum permalloy the most desirable core material and permalloy next. Not only are the losses lower than with iron but also the inductance is less affected by the intensity of the magnetizing force, and the harmonics and cross-modulation are also lower than with iron. Another advantage is that a large d-c magnetization has less permanent effect upon the inductance.

CAPACITY AND CAPACITORS

28. Definitions of Capacity.—Any pair of electrical conductors that will store electrical charge when a difference in potential is applied to them constitutes a condenser. The defining relation for capacity may be written

$$C = \frac{Q}{E} \tag{116}$$

where C is the capacity in farads, Q is the charge in coulombs, and E is the potential in volts. Capacity may also be defined in terms of the energy stored in the electrostatic field about the condenser plates; thus

$$W = \tfrac{1}{2}CE^2 \tag{117}$$

where W is the energy stored in watt-seconds (joules), C is the capacity in farads, and E is the potential in volts. Since current is the time rate of change of charge, the current flowing into any condenser is proportional to the time rate of change of the potential across it. This relation is given by

$$I = C\frac{dE}{dt} \tag{118}$$

where I is the current in amperes, t is the time in seconds, and the other symbols have their previous significance.

Capacity has the dimensions of length when expressed in electrostatic units. As a result the capacities of similar condensers are proportional to the first power of their linear dimensions. Doubling all dimensions of a condenser doubles its capacity.

29. Condenser Losses and Their Representation.—A perfect condenser when discharged gives up all the electrical energy that was supplied to it in charging. Actual condensers never realize this ideal perfectly but, rather, dissipate some of the energy delivered to them. Most of the loss in ordinary condensers occurs in the dielectric. Other ways by which energy can be lost in a condenser are from the resistance of the leads and metal plates, from leakage resistance between plates, and as a result of corona.

The merit of a condenser from the point of view of freedom from losses is usually expressed in terms of the power factor or phase angle.[1] The power factor represents the fraction of the input volt-amperes that is dissipated in the condenser, while the phase angle is the angle by which the current flowing into the condenser fails to be 90°

[1] The merit of a condenser is occasionally expressed in terms of the condenser Q, which represents the ratio of the capacitive reactance of the condenser to the equivalent series resistance. For all practical purposes the condenser Q is the reciprocal for the power factor.

out of phase with the applied voltage. When the losses are low, as is practically always the case, the phase angle expressed in radians is equal to the power factor. The power factor (or phase angle) is a ratio representing the fraction of the input volt-amperes that is dissipated in the condenser, and for a given type of condenser tends to be independent of the applied voltage, of the condenser size, etc., and with ordinary dielectrics is substantially constant over wide frequency ranges.

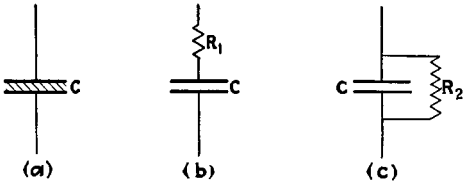


FIG. 75.—Representation of imperfect condenser by a perfect condenser of same capacity with series resistance, and by a perfect condenser with shunt resistance.

The effect of the condenser losses on the circuit in which the condenser is connected can be obtained by replacing the actual condenser by the combination of a perfect condenser of the same capacity with a resistance in series, as shown in Fig. 75*b* (or with a resistance in parallel, as in Fig. 75*c*). The value of the series (or shunt) resistance is so selected that the power factor of the combination of perfect condenser with its associated resistance is the same as the power factor of the actual condenser. The value of the resistances are

$$\text{Series resistance} = R_1 = \frac{\text{power factor}}{2\pi fC} \tag{119}$$

$$\text{Shunt resistance} = R_2 = \frac{1}{(2\pi fC)(\text{power factor})} \tag{120}$$

30. Dielectrics.—The insulating medium that separates the plates of a condenser is known as the *dielectric*, and plays an important role in determining the characteristics of the condenser. The presence of a dielectric other than a vacuum raises the capacity of the condenser in comparison to its capacity in the absence of the dielectric by a factor known as the *dielectric constant* (also sometimes called *specific inductive capacity*). The dielectric constants of the commonest insulating materials are given in Table 27, which also gives other properties of dielectrics.

Two very important classes of dielectrics listed in Table 27 are plastic and ceramic types. Some of the plastic dielectrics as, for example, bakelite, cellulose acetate, vinylite, etc., have very poor power factors. Others, notably polystyrene, have an unusually low power factor, which in some cases approaches that of quartz.

The more important ceramic dielectrics include steatite, cordierite, and titanium dioxide ceramics, and mycalex.[1] The term *steatite* refers to ceramic products that contain hydrous magnesium silicate (talc) as a predominant constituent of the fired body. Steatite materials combine low dielectric losses, high dielectric strength, and high insulation resistance. They are particularly suitable for general insulation purposes at high frequencies and high temperatures. It is possible to employ dry pressing in forming the product, so that parts can be manufactured in automatic presses to good accuracy and with economy. Cordierite is a complex magnesium aluminum silicate. Ceramics based upon it can be made that have an unusually low coefficient of thermal expansion, so that this material is particularly suitable for coil forms. The titanium dioxide ceramics are characterized by a high dielectric constant and low dielectric losses. This material finds its chief use as the dielectric of condensers and gives high mechanical stability and a high capacity in a relatively small volume. The temperature coefficient of capacity of such condensers depends upon the

[1] For further information on ceramic dielectrics, see Hans Thurnauer, Ceramic Insulating Materials, *Elec. Eng.*, Vol. 59, p. 451, November, 1940; M. D. Rigterink, Improved Ceramic Dielectric Materials, *Rev. Sci. Instruments*, Vol. 12, p. 527, November, 1941; W. G. Robinson, Ceramic Insulations for High-frequency Work, *Jour. I.E.E.*, Vol. 87, p. 570, 1940.

TABLE 27.—PROPERTIES OF DIELECTRICS*

	Dielectric constant	Power factor, per cent			Specific gravity	Tensile strength, lb. per sq in (multiply by 10^3)	Compressive strength, lb per sq in (multiply by 10^3)	Softens at °C	Stable at °C	Specific heat	Coefficient of linear expansion, parts in 10^6 per °C	Heat conductivity cgs	Machine ability	Water absorption, per cent in 24 hr	Remarks
		60 cycles	1 kc	1 mc											
Amber	2.9	....	.....	0.2	1.1	.....	.....	250	180	.....	44	.....	Very good	0	Natural petrified resin
Casein—moulded	6.4	....	.....	6	1.33	7	.....	177	165	.....	80	.....	Very good	4–9	
Cellulose acetate	6–8	7	.....	3–6	1.3	3	4	70	65	0.5	150	0.0005	Very good	4	Tenite, safety film—burns very slowly
Cellulose nitrate	4–7	5–9	5	5	1.5	3–6	.....	85	85	0.36	140	0.0003	Very good	2–3	Celluloid, pyralin, pyroxylin—burns rapidly
Cordierite ceramics	5–5.5	....	0.6–1	0.4–0.7	2.1	2–5	40–60	1,400	.....	.....	1–2	0.003	.....	0–1	Temp. coefficient of cap. = $+5 \times 10^{-4}$ per°C; used for coil forms
Fiber	4–5	6–9	5	5	1.3	10	25	130	95	.....	25	0.0011	Very good	30	
Glass—crown	6.2	....	1	.....	2.48	2–5	10–30	1,100	.....	0.161	8.9	0.0025	No	0	Window glass
Glass—flint	7	....	0.45	0.4	3.7	3–6	6–10	.....	.....	0.117	7.9	0.002	No	0	
Glass—pyrex	4.5	....	0.5	0.2	2.25	.....	40	600	520	0.2	3.2	0.0027	Very poor	0	
Methacrylic resin	2.8	3	2	2	1.19	8–9	12	135	90	0.45	70	0.00055	Very good	0.3	Lucite, plexiglass—slow burning
Mica—clear India	7–7.3	0.03	0.02	0.02	2.8	.....	.....	1,200	600	2.06	3–7	0.0018			
Mycalex	6–8	....	0.6	0.3	3.5	6–8	25–40	.....	350	0.22	8–9	0.0014	Poor	0.035	Mica and lead borate
Marble—white	7–9	....	.....	4	2.7	2	8–15	.....	.....	0.21	8–12	0.0015	Fair	Very high	
Phenol—pure	5	2	.....	1	1.3	5–11	15–30	.....	120	0.3	28	0.0004	Very good	0.15	Catalin, bakelite—burns very slowly
Phenol—yellow	5.3	2.5	1.4	0.7	1.9	5.5	.....	.....	130	.....	.....	.....	Poor	0.2	Low-loss bakelite—nearly nonburning
Phenol—black molded	5.5	8	6	3.5	1.35	7.5	30	.....	140	0.35	40	0.0005	Fair	0.3	Nearly nonburning
Phenol—paper base	5.5	6	5	3.5	1.35	10–15	30	.....	125	0.3	30	0.00065	Good	0.2–1	Nearly nonburning
Phenol—cloth base	5.6	5	5	5	1.38	11	35	.....	115	0.35	20	0.0005	Good	0.7	Nearly nonburning
Porcelain—wet process	6.5–7	2	1	0.6	2.4	3–6	30–50	1,610	1,050	0.25	4–5	0.0025	No	Low	
Porcelain—dry process	6.2–7.5	2	1	0.7	2.3	2–3	30–50	.....	1,050	0.26	3–4	0.0025	No	0.1–1	
Quartz—fused	4.2	0.03	0.03	0.03	2.21	7–10	200	1,430	1,150	0.18	0.45	0.0024	Very poor	0	SiO_2 conducts at 800°C
Rubber—hard	2–3	1	1	0.5–0.9	1.15	4–7	7	70	65	0.33	70–80	0.0004	Fair	0.02	Burns slowly
Slate	6–8	....	0.9	.....	2.8	5	15	.....	.....	0.22	10	0.005	Fair	High	
Steatite	6.1	1	0.4	0.3	2.5	8–10	50–100	1,500	1,000	.....	6–8	.....	No	0.02	Magnesium silicate—isolantite, lava—temp. coefficient of cap. = $+1.5 \times 10^{-4}$ per°C
Styrene (polymerized)	2.4–2.9	0.02	0.02	0.03	1.05	6–9	14	90	75	0.324	70	0.0004	Good	0.01	Victron, trolitul, polystyrene, amphenol—very slow burning
Sulphur	3–3.8	....	.....	.....	2.05	.....	.....	113	95	0.17	64	0.0006	.....	.....	Burns rapidly
Shellac	2.5–4	2.5	.....	0.9	1.1	0.9	7	85	75	.....	.....	0.0006	.....	0.1	Burns readily
Titanium dioxide	90–170	....	0.1	0.06	4–5	4	60	1,600	.....	.....	7–8	.....	No	0	Rutile—temp. coeff. of cap. = 6.5×10^{-4} per°C
Titanium dioxide magnesium titanate	12–18	....	0.1	0.02–0.1	3.1	8–12	70–85	1,400	.....	.....	6–10	0.010	.....	0	Coefficient of cap. = $\pm 0.5 \times 10^{-4}$
Titanium dioxide titanium—zirconium dioxides	40–60	....	0.1–0.3	0.05–0.10	3.6	4–7	40–80	1,500	.....	.....	6–7	.....	.....	0	Temp. coeff. of cap. = $\pm 0.5 \times 10^{-4}$
Urea—formaldehyde compounds	6–7	5	3.8	3	1.48	6–9	25–30	200	80	.....	.....	0.00017	Fair	0.4	Beetle, plaskon
Vinyl resins—unfilled	4	....	1.4	1.7	1.35	8–10	.....	.....	50	0.244	70	0.0005	Very good	0.15	Vinylite—nonburning

* Most of the data in this table are from the *Gen. Rad. Exp.* for June, 1939.

mixture, and values can be obtained that range from substantially zero to relatively large negative values. This gives the possibility of providing a condenser that will correct for the positive temperature coefficient of the remaining coils and condensers in a circuit.

Mycalex is composed of ground mica and lead borate glass mixed and fired. Metal inserts can be molded into mycalex pieces because of the relatively low firing temperature and the fact that there is no shrinkage. Mycalex can be machined, although with difficulty. It is used in radio transmitters for coil supports, control shafts, antenna insulators, and in other places where an insulator must be employed that is capable of withstanding high voltage.

Dielectric Losses—Loss Factor.—The merit of a dielectric from the point of view of losses can be expressed by the equivalent power factor or phase angle. When these terms are applied to a dielectric (as contrasted with a condenser), they are defined as the power factor (or phase angle) that would be obtained with a condenser having no losses other than those in the dielectric. Power factors of common dielectrics are given in Table 27.

The power factor (or phase angle) is independent of the size and shape of the dielectric or the applied voltage. It also tends to be independent of frequency, since the fraction of the energy lost during each cycle of charge and discharge is substantially independent of the number of cycles per second, over wide frequency ranges. The power factor of a dielectric depends not only upon the kind of dielectric involved but also upon the conditions of temperature and moisture under which it is used. In general, moisture and high temperature increase the power factor.

The power loss (and hence heating) per unit volume of dielectric is[1]

$$\left.\begin{array}{l}\text{Power loss in}\\ \text{watts per cu-}\\ \text{bic inch}\end{array}\right\} = 2\pi f G^2 K\tau \times 0.2244 \times 10^{-12} \qquad (121)$$

where f = frequency in cycles.
G = voltage gradient in dielectric, rms volts per in.
K = dielectric constant.
τ = power factor of dielectric.

The rate at which heat is generated in a dielectric is proportional to $K\tau$, which is termed the *loss factor* of the dielectric. The loss factor is the best single criterion of the ability of a solid insulating material to withstand high radio-frequency voltages.

31. Calculation of Condenser Capacity. *Parallel Plate Condenser.*—The capacity of a parallel plate condenser is

$$C = 0.2244K \frac{A}{d} \mu\mu\text{f} \qquad (122)$$

where K = dielectric constant of the dielectric between the plates.
A = area of one plate, sq in.
d = spacing of plates, in.

If the dimensions are in centimeters

$$C = 0.08842K \frac{A}{d} \mu\mu\text{f} \qquad (122a)$$

These expressions neglect the fringing flux lines at the edges of the plates and so give a result slightly less than the actual capacity.

[1] E. T. Hoch, Power Losses in Insulating Materials, *Bell System Tech. Jour.*, Vol. 1, p. 110, October, 1922.

Long Parallel Strips in Air.—The capacity between two parallel strips of great length compared with the width or spacing is given by the following approximate formulas:[1]

For $R = \dfrac{\text{width of strips}}{\text{separation of strips}} < 1$:

$$C = \frac{0.3063}{\log_{10}\left(\frac{4}{R}\right)} \mu\mu\text{f per inch} \tag{123}$$

This expression becomes more accurate as the value of R becomes smaller. For R less than 0.5, the formula gives results less than 2 per cent high, and for $R = 1$, the results are about 10 per cent high.

For $R > 1$:[2]

$$C = 0.2244R\left(1 + \frac{1}{\pi R}(1 + 2.303 \log_{10} 2\pi R)\right) \mu\mu\text{f per inch} \tag{124}$$

This expression gives values that are lower than the exact value, and whose accuracy increases as the value of R increases. The error is 4 per cent for $R = 2$ and less than 10 per cent for $R = 1$.

When the plates of the condenser are immersed in a dielectric of constant K (as if the dielectric were a liquid), the results obtained from Eqs. (123) and (124) must be multiplied by K. With solid dielectric between the plates but not projecting beyond the edges, these equations still hold, but with less accuracy. When the solid dielectric projects considerably beyond the edges of the plates, fair results can be obtained by multiplying values from Eqs. (123) and (124) by the dielectric constant.

Concentric Spheres.

$$C = 1.412K \frac{d_1 d_2}{d_2 - d_1} \mu\mu\text{f} \tag{125}$$

where d_1 = outer diameter of inside sphere, in.
d_2 = inner diameter of outside sphere, in.
K = dielectric constant of material between spheres.

Isolated Sphere.

$$C = 1.412d \mu\mu\text{f} \tag{126}$$

where d is the diameter of the sphere in inches and the sphere is assumed to be in free space (vacuum). This will be seen to correspond to the capacity of a set of concentric spheres with d_2 = infinity.

Isolated Disk.

$$C = 0.8992d \mu\mu\text{f} \tag{127}$$

where d is the diameter of the disk in inches. The disk is presumed to be of negligible thickness, and is assumed located in free space (vacuum).

Single Wire Parallel to Ground.—When the wire can be considered as infinitely long, then

$$C = \frac{7.354}{\log_{10}\left\{\frac{2h}{d}\left(1 + \sqrt{1 - \frac{1}{(2h/d)^2}}\right)\right\}} \mu\mu\text{f per foot} \tag{128}$$

where h = height of wire center above earth.
d = diameter of wire.

[1] The exact expression for this case involves elliptic integrals and can be solved only by trial and error. See H. B. Palmer, Capacitance of a Parallel Plate Capacitor by the Schwartz-Christoffel Transformation, *Trans. A.I.E.E.*, Vol. 56, No. 3, p. 363, March, 1937.

[2] A. E. H. Love, Some Electrostatic Distributions in Two Dimensions, *Proc. London Math. Soc.*, Vol. 22, p. 337, Mar. 8, 1923.

In the usual case, h is considerably greater than d, and Eq. (128) can be written as

$$C = \frac{7.354}{\log_{10}\frac{4h}{d}} \mu\mu f \text{ per foot} \tag{129}$$

Curves of capacity per inch as calculated by Eq. (129) for various wire sizes and heights above ground are given in Fig. 76. These curves are useful for determining wiring capacity in radio receivers.

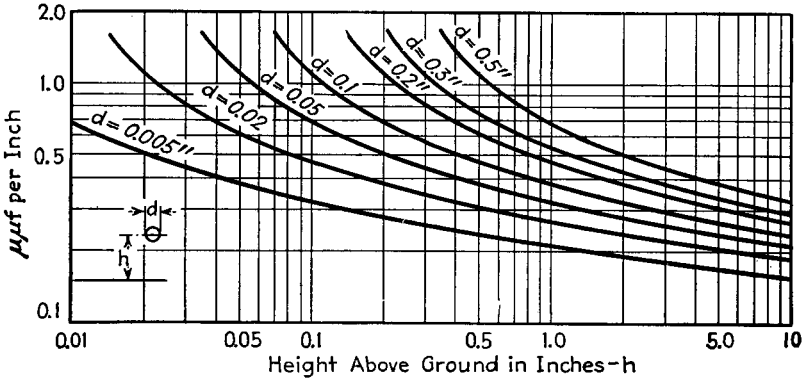


FIG. 76.—Capacity per inch of single-wire conductor of various diameters as a function of height above ground.

When the wire cannot be considered infinitely long, but still has a small diameter compared with the height, then[1]

$$C = \frac{7.354l}{\log_{10}\left(\frac{4h}{d}\right) - S} \mu\mu f \tag{130}$$

where l is the length in feet, h and d are the height and wire diameter, respectively, in the same units, and S is a constant given in Table 28.

TABLE 28.—VALUES OF THE CONSTANT S*

$2h/l$	S	$l/2h$	S	$l/2h$	S
0	0	1.00	0.336	0.50	0.541
0.1	0.042	0.95	0.350	0.45	0.576
0.2	0.082	0.90	0.364	0.40	0.617
0.3	0.121	0.85	0.379	0.35	0.664
0.4	0.157	0.80	0.396	0.30	0.721
0.5	0.191	0.75	0.414	0.25	0.790
0.6	0.223	0.70	0.435	0.20	0.874
0.7	0.254	0.65	0.457	0.15	0.990
0.8	0.283	0.60	0.482	0.10	1.155
0.9	0.310	0.55	0.510	0.05	1.445
1.0	0.336	0.50	0.541		

* The argument to be used is either $2h/l$ or $l/2h$, according to which is less than unity.

[1] Frederick W. Grover, Methods, Formulas, and Tables for the Calculation of Antenna Capacity, *Bur. Standards Sci. Paper* 568.

Capacity to Ground of Parallel Wires of the Same Height Joined Together.[1]—On the assumption that the wire diameter is small compared with the spacing between wires, then

$$C = \frac{7.36l}{F} \mu\mu f \tag{131}$$

where $F = \dfrac{P + (n-1)Q}{n} - S_n$.

$P = \log_{10} \dfrac{4h}{d} - S$.

$Q = \log_{10} \dfrac{2h}{D} - S$.

l = length of wire, ft (assumed same for all wires).
d = wire diameter, ft.
n = number of wires.
D = spacing between adjacent wires, ft (assumed same for all adjacent pairs).
h = height above the surface of the earth, ft.
S = constant given by Table 28.
S_n = constant given by Table 29.

This equation assumes that $(n-1)D \lesssim l/4$.

TABLE 29.—VALUES OF THE CONSTANT S_n

n	S_n	n	S_n	n	S_n	n	S_n
2	0	8	0.347	14	0.550	20	0.688
3	0.067	9	0.388	15	0.576	30	0.847
4	0.135	10	0.425	16	0.601	40	0.970
5	0.197	11	0.460	17	0.625	50	1.063
6	0.252	12	0.492	18	0.647	100	1.357
7	0.302	13	0.522	19	0.668		

For the case of two wires, Eq. (131) reduces to

$$C = \frac{14.73l}{\log_{10} \dfrac{4h}{d} + \log_{10} \dfrac{2h}{D} - 2S} \tag{132}$$

Capacity to Ground of a Horizontal Cage.[1]—When the wires are closely spaced around the circumference of the cage, the capacity to ground can be obtained very approximately by using Eq. (130), if the cage is considered to be a wire having a diameter equal to the diameter of the cage. The capacity obtained in this way is large. Accurate results are given by

$$C = \frac{7.36ln}{\log_{10} \dfrac{4h}{d} + \displaystyle\sum_{r=1}^{r=n-1} \left(\log_{10} \frac{2h}{D_r} + 0.434 \frac{D_r}{l} \right) - nS} \mu\mu f \tag{133}$$

where l = length of wires in cage, ft.
n = number of wires in cage.
d = diameter of wires in cage, ft.
h = height of cage center above ground, ft.
S = constant from Table 28.

[1] Grover, *loc. cit.*

D_r = distance between any given wire and another wire, ft. $= \delta \sin r\frac{\pi}{n}$, where δ is the diameter of the cage, ft.

Capacity to Ground of Vertical Wire and Cage Antennas.[1]—For a single vertical wire

$$C = \frac{7.36m}{\log_{10}\frac{2m}{d} - k} \mu\mu f \tag{134}$$

where m = length of vertical wire, ft.
h' = height of lower end of wire above ground, ft.
d = diameter of wire, ft.
k = constant given by Table 30 and depending on h'/m.

For accurate results, h'/m must be small.

The capacity to ground of a vertical cage of n wires of diameter d, length m, arranged as elements of a cylinder of diameter δ whose axis is vertical and whose lower end is at a height h' above ground is (dimensions in feet):

$$C = \frac{7.36nm}{\log_{10}\frac{2m}{d} + \sum_{r=1}^{r=n-1}\left(\log_{10}\frac{m}{D_r} + 0.434\frac{D_r}{m}\right) - nk} \mu\mu f \tag{135}$$

where k is obtained from Table 30 and

$$D_r = \delta \sin r\frac{\pi}{n} \tag{136}$$

TABLE 30.—VALUES OF THE CONSTANT k*

h'/m	k	h'/m	k	m/h'	k
....	.....	0.3	0.280	1.0	0.207
0.02	0.403	0.4	0.261	0.9	0.202
0.04	0.384	0.5	0.247	0.8	0.196
0.06	0.369	0.6	0.236	0.7	0.190
0.08	0.356	0.7	0.227	0.6	0.184
0.10	0.345	0.8	0.219	0.5	0.177
0.15	0.323	0.9	0.2125	0.4	0.170
0.20	0.305	1.0	0.207	0.3	0.162
0.25	0.291	...	.....	0.2	0.153
0.30	0.280	...	.....	0.1	0.144
....	.....	...	.....	0	0.133

* The argument is h'/m or m/h' according to which is less than unity.

Capacity to Ground of Single-wire T and Inverted L Antennas.[1]—The capacity in these cases can be determined approximately by adding the capacities calculated separately for the vertical and horizontal components. This gives a result somewhat high, however, because of mutual effects between the vertical and horizontal portions. More accurate results are obtainable by the formulas given below.

For a single-wire T having a horizontal length l at a height h above ground and a vertical wire of length m attached to the center of l, then, if the dimensions are in feet

$$C = \frac{7.36(l + m)}{U} \tag{137}$$

[1] Grover, *loc. cit.*

where

$$U = \frac{l}{l+m}\left(\log_{10}\frac{4h}{d} - S\right) + \frac{m}{l+m}\left(\log_{10}\frac{2m}{d} - k\right) + \frac{l+2m}{l+m}X \qquad (138)$$

The constants S and k are given by Tables 28 and 30; while X is given by Fig. 77. In using these tables, it is to be noted that the height h' of the lower end of the vertical above ground is $(h - m)$.

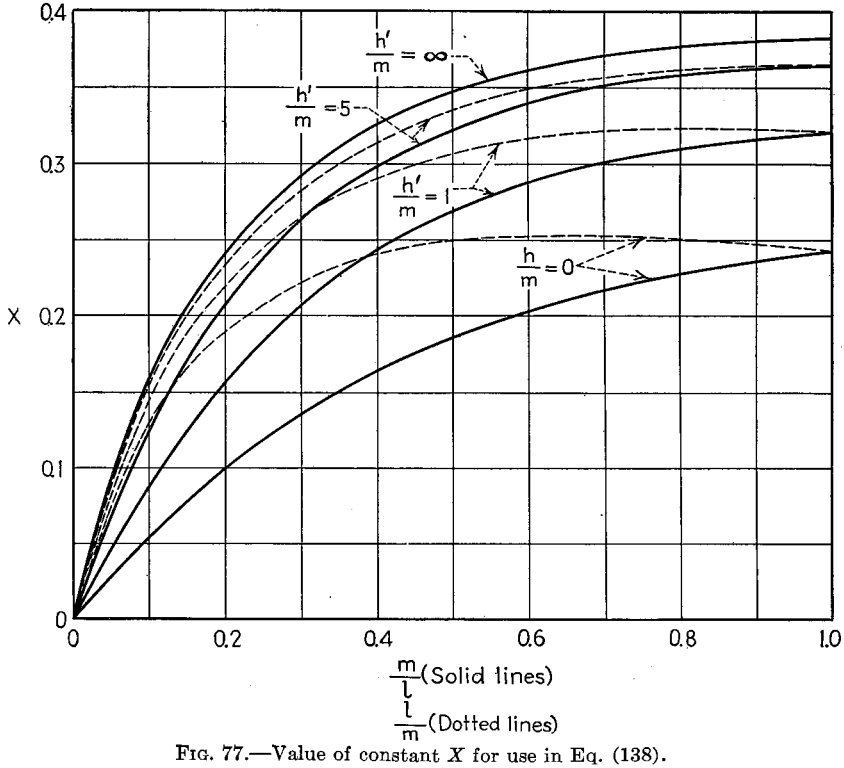


FIG. 77.—Value of constant X for use in Eq. (138).

The capacity to ground for the case of a single-wire inverted L antenna (*i.e.*, where the down-lead is attached to the end instead of being center of the horizontal portion) is also given by Eq. (138), provided U is defined as

$$U = \frac{l}{l+m}\left(\log_{10}\frac{4h}{d} - S\right) + \frac{m}{l+m}\left(\log_{10}\frac{2m}{d} - k\right) + X \qquad (139)$$

Capacity to Ground of Inclined Wire, Vertical Fan, etc.—The capacity to ground of a considerable variety of arrangements such as vertical cone, vertical fan, inclined wires, horizontal V, etc., are given in convenient form by F. W. Grover in the reference cited above.

Capacity between Two Horizontal Parallel Wires Remote from the Ground.—When the wires are infinitely long, the capacity per unit length is half that given by Eqs. (128) and (129), calculated by replacing h in these equations by $D/2$, where D is the spacing between wire centers. This gives

$$C = \frac{3.677}{\log_{10}\left\{\frac{D}{d}\left(1 + \sqrt{1 - \frac{1}{(D/d)^2}}\right)\right\}} \mu\mu\text{f per foot} \tag{140}$$

For the common case of d/D small, this becomes

$$C = \frac{3.677}{\log_{10}\frac{2D}{d}} \mu\mu\text{f per foot} \tag{141}$$

$$C = \frac{0.01941}{\log_{10}\left(\frac{2D}{d}\right)} \mu\text{f per mile} \tag{142}$$

The curves of Fig. 78 give the capacities for various wire sizes and spacings.

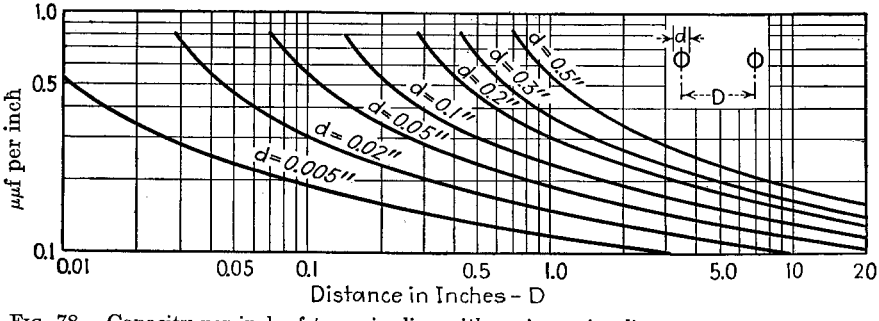


FIG. 78.—Capacity per inch of two-wire line with various wire diameters as a function of spacing.

The case for finite length, corresponding to Eq. (130), is

$$C = \frac{3.68l}{\log_{10}\frac{4h}{d} - S} \mu\mu\text{f} \tag{143}$$

where the notation has the same meaning as in Eq. (130).

Capacity between Two Parallel Wires in Presence of Ground.—In the usual case where d/l and d/D are very small, one has

$$C = \frac{3.677}{\log_{10}\frac{2D}{d}\sqrt{1 + \left(\frac{D}{2h}\right)^2}} \mu\mu\text{f per foot} \tag{144}$$

The presence of the ground increases the capacity for this case by the same amount that would result if the spacing were assumed to have an effective value of $D/\sqrt{1 + (D/2h)^2}$.

Capacity between Two Horizontal Parallel Wires One above the Other.—The capacity for this case may be determined approximately from Eq. (144) by taking for the value of h the average of the heights of the wires $\frac{h_1 + h_2}{2}$.

Capacity of Concentric Cable.—Let

D = inside diameter of the outside cylinder.

d = outside diameter of the inner conductor.

K = dielectric constant of material between conductors.

Then

$$C = \frac{7.354K}{\log_{10}\left(\frac{D}{d}\right)} \ \mu\mu\text{f per foot} \qquad (145)$$

The capacities of concentric cables with various inner and outer diameters are given in Fig. 79.

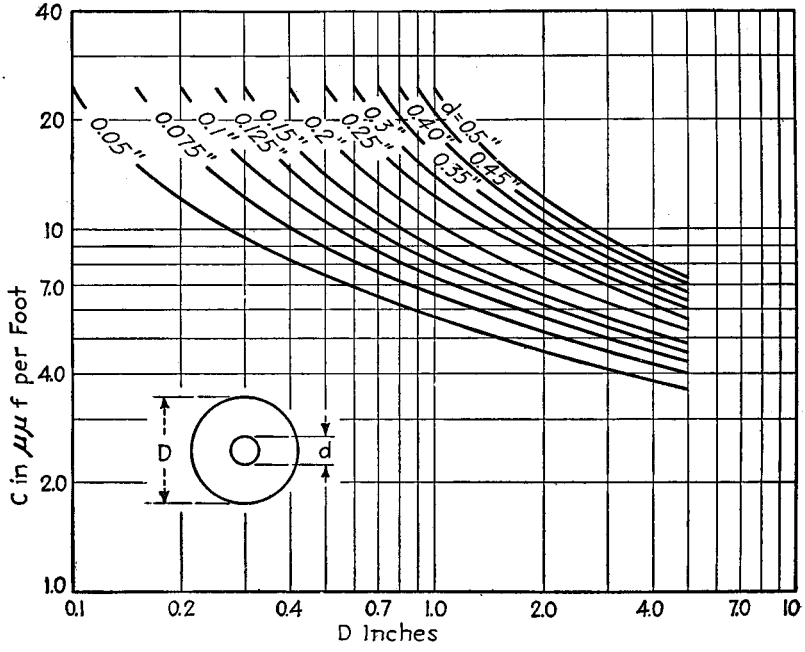


FIG. 79.—Capacity per foot of concentric lines with air dielectric.

32. Condensers with Air Dielectric.—An air-dielectric condenser consists of two assemblies of spaced plates held in fixed relationship to each other by insulating members so that the plates mesh but do not touch. Such condensers are normally

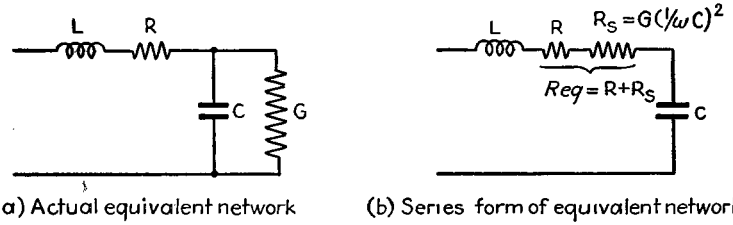


FIG. 80.—Equivalent electrical networks of a variable air condenser.

made so that one set of plates can be rotated to adjust the capacity. Variable air-dielectric condensers are characterized by very low losses, and by having reproducible capacity settings. They are used for adjusting the resonant frequency of tuned circuits in both transmitters and receivers.

Resistance and Inductance of Air-dielectric Condensers.—Even though air is a perfect dielectric with zero power factor, air condensers have losses because of the resistance of the plates, rods, washers, etc., and because of the losses in the dielectric used to mount the two sets of plates with respect to each other. Air condensers also have a small inductance as a consequence of the magnetic flux produced by the currents flowing through the condenser and its connections.

The actual condenser can be represented by the electrical network of Fig. 80*a* (or by the series equivalent of this, as given in Fig. 80*b*), where C is the capacity of the condenser.[1]

The condenser inductance L in Fig. 80 takes into account the magnetic effect of the currents in the condenser, and is to a first approximation independent of the capacity setting or the frequency. It depends upon the physical size of the condenser and the method of making connections to the two sets of condenser plates. This inductance can be reduced by making the condenser of small physical size and by bringing the connections to the centers of the plate assemblies, rather than to the ends as is usually done. The effect of the inductance is to cause the apparent capacity of the condenser, as observed at the terminals, to be greater than the actual capacity according to the equation

$$\text{Apparent capacity} = \frac{C}{1 - \omega^2 LC} \tag{146}$$

where C and L are the capacity and inductance in farads and henrys, respectively, and ω is 2π times the frequency in cycles. The difference between the apparent and actual capacities is negligible except at high frequencies, but then increases very rapidly. The order of magnitude that can be expected for L is indicated by a value of 0.06 μh for a physically large standard variable condenser having a maximum capacity of 1,000 $\mu\mu f$ and 0.0055 for a physically small (145 $\mu\mu f$ maximum) variable condenser having center-fed plates, and otherwise designed for low inductance.

The conductance G in Fig. 80*a* represents the losses in the solid dielectrics used for support, and is independent of the capacity setting but is directly proportional to the frequency. The conductance is determined by the type and amount of dielectric and the method of using it. The merit of a condenser with respect to freedom from dielectric losses can be expressed in terms of the quantity $R_s\omega C^2$, where R_s is the series resistance corresponding to G (see Fig. 80). For any given condenser this figure of merit is independent of frequency and capacity.

The resistance R represents the series resistance of the leads, washers, connecting rods, etc. It is substantially independent of the capacity setting, but increases with frequency as the result of skin effect, being proportional to the square root of the frequency at higher frequencies.

The ratio $R_{eq}/(1/\omega C)$ in Fig. 80 is termed the condenser power factor (the reciprocal of the condenser power factor is the condenser Q). The value of the power factor is determined primarily by the dielectric losses G at low frequencies and by the series resistance R at high frequencies. For a given variable condenser, the power factor is maximum at low capacity settings for low frequencies, whereas at high frequencies the power factor commonly goes through a minimum at something less than the maximum capacity. The behavior in some typical cases is shown in Fig. 81.[2] It will be noted that condenser power factors, although small, are by no means zero, particularly at small capacity settings and also at very high frequencies.

[1] R. F. Field and D. B. Sinclair, A Method for Determining the Residual Inductance and Resistance of a Variable Air Condenser at Radio Frequencies, *Proc. I.R.E.*, Vol. 24, p. 255, February, 1936.

[2] See Field and Sinclair, *loc. cit.;* also, W. Jackson, The Analysis of Air Condenser Loss Resistance, *Proc. I.R.E.*, Vol. 22, p. 957, August, 1934.

Law of Capacity Variation.—The plates of a variable air condenser can be shaped to give any desired law of capacity variation as a function of angle of rotation.[1] The types of condensers of greatest interest are condensers in which the capacity is a linear function of angle of rotation (termed straight-line capacity and abbreviated SLC); condensers in which the plates are shaped so that, when used to tune a coil to resonance, the wave length at resonance is a linear function of the angle of rotation (straight-line wave length, abbreviated SLW); condensers that, when used to tune an inductance, give a resonant frequency that is a linear function of the angle of rotation (straight-line frequency, abbreviated SLF); and, finally, condensers in which either the logarithm of capacity or the logarithm of frequency of resonance with a coil is a straight-line function of the angle of rotation. Practical condensers frequently use intermediate characteristics or combinations of these basic types. Thus the tuning

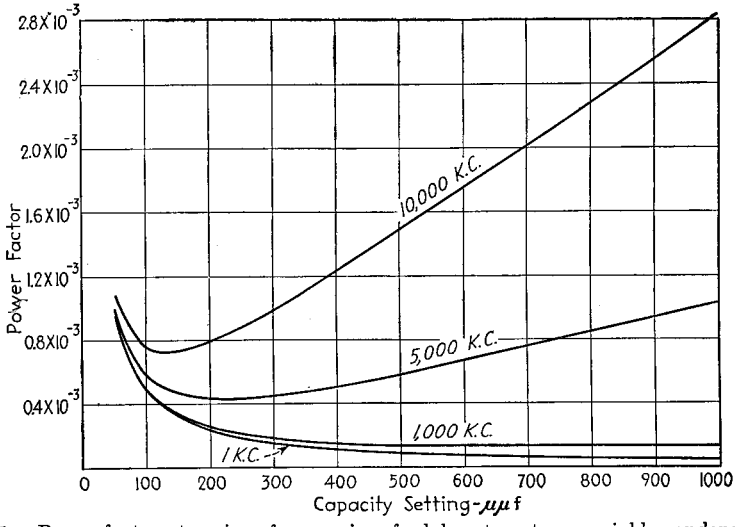


FIG. 81.—Power factor at various frequencies of a laboratory-type variable condenser as a function of capacity setting.

condensers used in a broadcast receiver usually approximate a straight-line wave-length law at low capacity settings and a straight-line frequency law at large capacity settings. Still other shapes are used to meet special requirements, as, for example, in beat-frequency oscillators.[2]

Straight-line capacity characteristics can be obtained by using semicircular rotating and stationary plates, as illustrated in Fig. 82. The straight-line wave-length characteristic is obtained by modifying the shape of the rotating plates according to the following equation:

$$r = \sqrt{\frac{5dK(K\theta + \lambda_{min})}{NL} + r_1^2} \text{ inches} \qquad (147)$$

[1] Condensers have also been designed with wedge-shaped plates so arranged that the thickness of the air gap varies with the position of the rotating plates, thereby introducing an additional factor modifying the law of capacity variation. Such condensers are used only to a limited extent, however, because the plates cannot be made of standard sheet material.

[2] Thus see E. R. Meissner, Logarithmic Scale for Beat-frequency Oscillator, *Proc. I.R.E.*, Vol. 17, p. 879, May, 1929; W. H. F. Griffiths, Notes on the Laws of Variable Air Condensers, *Exp. Wireless and Wireless Eng.*, Vol. 3, p. 14, January, and p. 743, December, 1926.

where d = separation between plates, in.

$K = \frac{\lambda_{max} - \lambda_{min}}{\pi}$.

N = number of air spaces in condenser.

r_1 = radius of cutout on stationary plates, in.

L = tuned inductance, microhenrys.

θ = angle of rotation, radians

λ = wave length, meters.

Similarly, a straight-line frequency characteristic is obtained with rotating plates cut so that the radius of the plates as a function of the angular position is given by[1]

$$r = \sqrt{\frac{4D^2}{nkK^2\left(\frac{D}{K\sqrt{C_0}} - \theta\right)^3} + r_1^2} \text{ cm} \qquad (148)$$

where $D = 1/2\pi\sqrt{L}$, a constant (L, henrys).

$K = \frac{f_{max} - f_{min}}{\pi}$, cycles per radian on dial.

n = number of air spaces in condenser.

$k = 10^{-11}/36\pi d$, a condenser constant.

C_0 = total minimum capacity when $\theta = 0$, farads.

θ = angle of rotation of rotor plates in radians.

r_1 = radius of cutout on stationary plates, cm.

d = air gap between plates, cm.

The approximate plate shape and the way in which the capacity varies with the angle of rotation in a typical case are shown in Fig. 82.

Temperature Coefficient of Capacity.—The principal factors that may cause the capacity of an air-dielectric condenser to change with temperature are (1) differential expansion of different portions of the condenser, particularly effects causing bending of the plates, (2) linear expansion that causes changes in surface areas, etc., (3) changes in dimensions and dielectric constant of the insulation used for mounting, (4) changes in residual stress with temperature that give rise to deformations.

The temperature coefficient of capacity of ordinary air-dielectric condensers may vary over a wide range. Thus values from +150 to −65 parts per million per °C have been observed for different designs. The behavior of ordinary condensers with respect to changes in temperature is also frequently noncyclic, *i.e.*, the capacity does not return to its original value when the temperature comes back to normal.

Air condensers with small and adjustable temperature coefficients, having cyclic behavior can be realized by proper design.[2] Such a result is facilitated by making the fixed and rotating plate assemblies of the same material throughout and annealing to relieve residual stress. The air gap should be large so that the capacity will not be critical with small deformations of the plates. The insulating supports should have low temperature coefficients of expansion and of dielectric constant, and possess cyclic behavior. Certain of the ceramics and plastics are the best. The exact temperature coefficient of such a condenser can be adjusted as desired by controlling the axial position of the rotating plates with respect to the fixed plates by means of metal rods of a second material having a different coefficient of expansion. In this way it is possi-

[1] H. C. Forbes, The Straight-line Frequency Variable Condenser, *Proc. I.R.E.*, Vol. 13, p. 507, August, 1925. See also O. C. Roos, Simplified SLF and SLW Design, *Proc. I.R.E.*, Vol. 14, p. 773, December, 1926; Griffiths, *loc. cit.*

[2] H. A. Thomas, The Development of a Small Variable Air Condenser Compensated for Rapid Changes of Temperature, *Jour. I.E.E.*, Vo.. 84, p. 495, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 157, June, 1939; The Electrical Stability of Condensers, *Jour. I.E.E.*, Vol. 79, p. 297, 1936; also, Wireless Section, *I.E.E.*, Vol. 11, p. 202, September, 1936.

ble to obtain zero temperature coefficient of capacity by starting with a proper initial displacement of the rotating plates away from the mid-position between fixed plates, or one can obtain either a positive or negative temperature coefficient as desired by using a greater or less displacement than that giving zero coefficient.[1]

Transmitting Condensers with Air Dielectric.—In transmitting condensers, the air gap between the plates must be in proportion to the voltage for which the condenser is rated, and care must be used in the design so that there will be no point at which electrostatic flux will concentrate unduly and cause even the slightest corona at rated

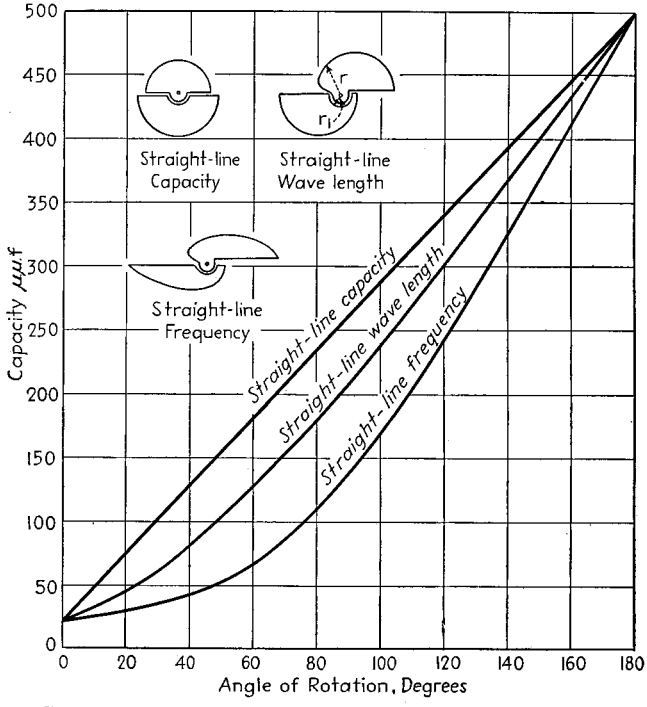


FIG. 82.—Characteristics of straight-line capacity, straight-line wavelength, and straight-line frequency condensers, showing capacity as a function of angle of rotation in typical cases, together with approximate shape of plates.

voltage. Furthermore, the solid insulation must have a low loss factor and be arranged so that the voltage gradient in it is low.

The dielectric strength of air is somewhat less at radio frequencies than at 60 cycles. Furthermore, a slight trace of corona such as would be of negligible consequence at 60 cycles, will cause intense ionization, accompanied by high power loss and possible breakdown when the frequency is high.[2]

The effective utilization of air as a dielectric in high-voltage condensers requires large spacings between plates and careful rounding of all sharp edges to avoid flux

[1] Another method of obtaining zero coefficient in a variable condenser involves a condenser in which all parts determining length are aluminum, and as are also all the stator plates, while half of the rotor plates are invar and half aluminum. See T. Slonczewski, High Accuracy Heterodyne Oscillators, *Bell System Tech. Jour.*, Vol. 19, p. 407, July, 1940.

[2] P. A. Ekstrand, Radio Frequency Spark-over in Air, *Proc. I.R.E.*, Vol. 28, p. 262; A. Alford and S. Pickles, Radio-frequency High Voltage Phenomena, *Trans. A.I.E.E.*, Vol. 59, p. 129, March, 1940; E. W. Seward, The Electric Strength of Air at High Frequencies, *Jour. I.E.E.*, Vol. 84, p. 288, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 31, March, 1939.

concentrations. In particular, the plates should be polished, have rounded edges, and have a thickness from one-third to one-fourth of the air gap.

Pressure and Vacuum Condensers.—The voltage rating of an air-dielectric condenser can be increased by increasing the air pressure, since the relationship between breakdown voltage and pressure at radio frequencies has the character shown in Fig. 83. Variable and fixed condensers based on this principle are commercially available[1] and are used in high-power radio transmitters at broadcast and higher frequencies for tuning tank circuits, as neutralizing condensers, and in antenna coupling networks. Practical compressed-gas condensers employ dry nitrogen from a tank to maintain the pressure, and operate at approximately 20 atm. This gives a voltage rating of four to five times that obtainable at atmospheric pressure.

Another means of increasing the voltage rating of an air-dielectric condenser is to place it in a vacuum comparable with that used in vacuum tubes. Under these condi-

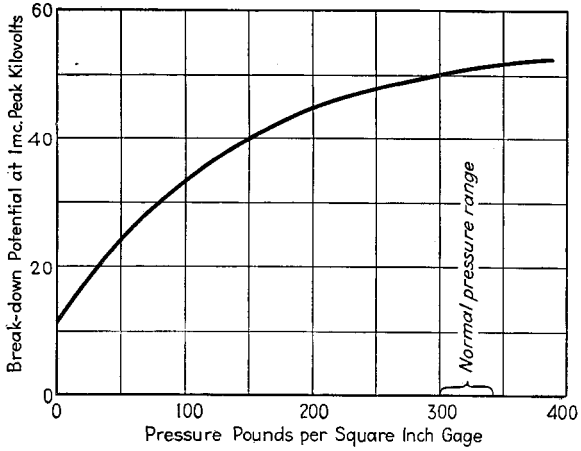


FIG. 83.—Breakdown voltage as a function of pressure at a frequency of one megacycle.

tions, there is so little gas present that ionization is not possible in appreciable amounts, and the dielectric strength is very high. Condensers of this type can be either fixed or variable. In the case of variable condensers, the adjustment is made from outside the vacuum by means of a copper bellows or flexible diaphragm.

33. Condensers with Solid Dielectric. *Types of Dielectrics Used.*—Solid dielectric is used in most fixed and in some adjustable condensers. The dielectrics most commonly employed are mica, paper, ceramics, and electrolytic films.

Mica is characterized by an unusually low power factor and high voltage strength. Mica condensers are commonly constructed by piling alternate sheets of mica and copper foil to form a stack that is held under pressure and impregnated. Such condensers are used for by-pass and blocking purposes and also whenever a fixed capacity of very low losses and appreciable capacity is required.[2]

Paper condensers normally consist of two strips of tin or aluminum foil insulated by paper and rolled into a bundle. This is vacuum-impregnated with oil or wax, and sealed against moisture. Paper condensers are characterized by relatively low cost, high capacity in proportion to voltage rating and size, and a power factor that is

[1] Pressure Capacitors, *Electronics*, Vol. 12, p. 16, April, 1939.

[2] The capacity of small mica condensers is frequently indicated by a series of three colored dots, which indicate the capacity in $\mu\mu f$ by the same color code used with resistors, described on page 39. The sequence in which the dots should be read is indicated by an arrow.

reasonably high (commonly of the order of 0.5 per cent). Such condensers are used primarily for by-pass and blocking purposes and for filter condensers in power-supply systems. The inexpensive type of paper condensers sealed in waxed cardboard containers deteriorate with time as a result of gradual penetration of moisture, and so have a limited life. This life is also dependent upon the d-c voltage applied to the condenser.[1]

Ceramic dielectrics are used to obtain controlled temperature coefficients. The ceramics used frequently contain titanium oxide (see Table 27), to give a very high dielectric constant. The power factor of such condensers can be quite low, and is frequently less than 0.001.

Electrolytic condensers are considered in Par. 34.

Power Factor, Resistance, and Reactance of Condensers.—Condensers with solid dielectric can be represented by the same equivalent electrical circuit shown in Fig. 80 for air condensers. The capacity C in this equivalent circuit is determined by the area, thickness, and dielectric constant of the dielectric. The inductance arises from flux linkages associated mainly with the leads. The loss in the condenser at low and moderate frequencies arises primarily from the conductance G, so the condenser power factor at such frequencies is equal to the power factor of the dielectric, and is independent of frequency.[2] However, as the frequency is increased the series resistance R begins to be of importance, and the power factor decreases. This series resistance tends to be proportional to the square root of frequency, making the power factor proportional to $f^{3/2}$ when the frequency f is so high that the series resistance is the principal source of loss.

Temperature Coefficient of Capacity.—Change in temperature can cause the capacity of condensers having solid dielectric to vary as a result of the effect of temperature on the dimensions, dielectric constant, and mechanical stress (and hence pressure) in the condenser. In ordinary mica and impregnated paper condensers, the effect of the temperature upon the pressures in the condenser assembly is the most important factor causing the capacity to change, and results in temperature coefficients of capacity that are large and noncyclic.

A low and very stable coefficient of temperature can be obtained by depositing the condenser electrodes either directly upon the opposite sides of a mica sheet or upon the inside and outside of a suitable ceramic tube. The metalized mica type of condenser gives a temperature coefficient of capacity of approximately +20 parts in a million per°C. The coefficient of the ceramic type of condenser can be controlled by the ceramic material employed, with values ranging all the way from about +200 to −800 parts in a million per°C. being available in commercial condensers.

Voltage Rating.—The voltage rating of a condenser at low frequencies is the voltage at which the dielectric will spark through. As the frequency is increased the voltage rating becomes less as a result of energy dissipated in the condenser. When the power factor of the dielectric accounts for most of the loss, as is the case at moderate radio frequencies, the voltage rating will be inversely proportional to the square root of the frequency. At higher frequencies, where the series resistance is also a factor, the permissible voltage decreases still more rapidly with frequency.

The voltage rating of a typical mica condenser at various frequencies is given in Table 31. These ratings assume that the voltage is applied continuously for a con-

[1] It has been stated that the life of an ordinary paper condenser such as is used in radio receivers is proportional to the fifth power of the applied voltage. This fact is sometimes used for accelerated life tests by operating the condenser at excess voltage and then using this rule to determine the significance of the results. Such tests neglect the time factor involved in moisture penetration, however, and so are only approximate.

[2] At quite low frequencies, there is usually a slight rise in the power factor as a result of leakage and absorption.

siderable length of time. When the voltage acts for a brief period in such a way that the condenser does not have time to exceed its allowable operating temperature, correspondingly higher voltages may be applied at the higher frequencies.

TABLE 31.—VOLTAGE RATING OF A PARTICULAR 0.001 μf MICA CONDENSER*

Frequency, kc	Rated effective voltage, volts	Rated current, amp.	Rated energy, kva
Direct current	10,000		
1	10,000	0.063	0.63
100	3,000	1.9	5.7
300	3,000	5.7	17.0
1,000	1,780	11.2	20.0
3,000	605	11.4	8.9
10,000	178	11.2	2.0

* I. G. Maloff, Mica Condensers in High Frequency Circuits, *Proc. I.R.E.*, Vol. 20, p. 647, April, 1932.

34. Electrolytic Condensers.[1]—An electrolytic condenser consists of positive and negative electrodes, usually of aluminum, placed in a suitable conducting liquid. The dielectric of such a condenser is provided by a thin insulating film formed directly on the positive electrode by suitable treatment.

Electrolytic condensers are characterized by a low cost per microfarad, a very large capacity in proportion to volume, a high power factor, and appreciable leakage. The characteristics such as capacity, leakage resistance, and power factor depend greatly upon the applied voltage, temperature, previous history of the condenser, etc. The ordinary electrolytic condensers used in radio work are polarized, so that the voltage on the anode plate must never be allowed to become negative even instantaneously. This limits the use of electrolytic condensers to circumstances where the alternating voltage present is less than the superimposed d-c voltage.

Electrolytic condensers are used as by-pass and filter condensers in radio receivers in parts of the circuit where a d-c potential is present and where a moderate leakage current can be tolerated.

Electrolytic Condensers.—Electrolytic condensers used in radio work are nearly always of the "dry" type. In such condensers, the positive and negative electrodes consist of aluminum foil separated by paper or gauze that is saturated with an electrolyte that is either a highly viscous liquid or a fudgelike solid. The entire assembly is wound in a roll and mounted in a waxed cardboard tube or box.[2] The anode is usually chemically etched to increase its effective area. The insulating film on the anode is formed chemically before assembly by passing the foil through suitable solutions and applying a proper sequence of voltages between the foil and the solution. After the condenser has been assembled any damage that may have occurred to the

[1] For further information, particularly on constructional details, see P. M. Deeley, "Electrolytic Capacitors," Cornell Dubilier Electric Corp., 1938; P. R. Coursey, "Electrolytic Condensers," 2d ed., Chapman and Hall, 1939.

[2] In the "wet" type of condenser, the anode consists of folded aluminum foil; the cathode electrode is usually an aluminum container lined inside with a perforated insulating material to prevent contact between cathode and anode. The electrolyte is a liquid, frequently moderately viscous.

Compared with the dry electrolytic condenser, the wet type has the disadvantages of higher power factor, narrower temperature range over which satisfactory operation can be obtained, and greater deterioration while lying idle. Furthermore, wet condensers must be provided with a vent to prevent excessive internal pressures from building up and must be mounted in a particular position to avoid leakage. The wet type is capable of giving voltage limiting action, since it can be designed to draw an excessively large leakage current when the voltage on the circuit becomes larger than normal. In such service, the wet type has both greater reliability and life and also possesses a more rapid rise of leakage current with voltage than the dry type.

anode film through handling, etc., is repaired by a reforming process that consists of applying voltage to the condenser under suitably controlled conditions.

Characteristics of Electrolytic Condensers.[1]—The capacity of an electrolytic condenser is influenced by the applied voltage, length and continuity of service, voltage at which the insulating film was formed on the anode, the temperature, and the frequency. In general the capacity decreases with age, increased forming voltage, increased temperature, and increased frequency. In condensers that must carry appreciable alternating currents, as, for example, the first condenser in a condenser-input rectifier filter system, there is a tendency for an insulating film to form on the cathode during use, with consequent excessive loss of capacity.

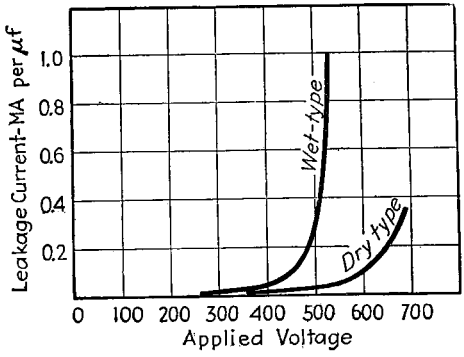


FIG. 84.—Leakage current in typical electrolytic condensers.

The leakage resistance of an electrolytic condenser depends upon the ratio of voltage applied to the voltage used in forming the anode film, and is affected by temperature, amount and conditions of use, etc. The leakage current varies with applied voltage in the manner shown in Fig. 84. It is quite small with applied voltages appreciably less than the forming voltage, but increasing rapidly in the region of the forming voltage, the rise being particularly great in the wet type. When voltage is applied to a condenser after a period of idleness, the leakage current is initially excessively large, but quickly drops to a normal value, as shown in Fig. 85. The leakage

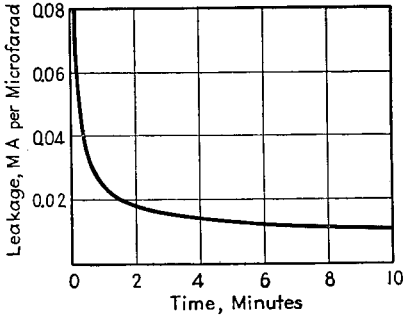


FIG. 85.—Curve illustrating typical manner in which leakage current varies with time when voltage is first applied to an electrolytic condenser that has been idle for some time.

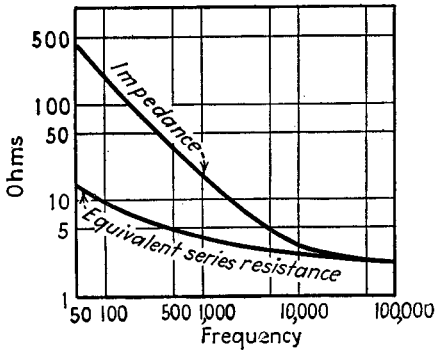


FIG. 86.—Impedance and equivalent series resistance of a typical electrolytic condenser having a normal capacity of 8 μf.

current for any particular operating voltage increases as the temperature becomes greater.

The power factor of electrolytic condensers is high, and increases with frequency and increasing applied voltage. The power factor may increase or decrease with

[1] A good summary of properties of wet and dry electrolytic condensers is given by P. R. Coursey and S. N. Ray, Electrolytic Condensers, *Jour. I.E.E.*, Vol. 85, p. 107, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 203, September, 1939.

temperature, depending upon the construction. At high frequencies the power factor becomes so high that the condenser acts essentially as a series resistance rather than as a capacity (see Fig. 86).

The maximum voltage that can be safely applied to an electrolytic condenser depends upon the voltage at which the insulating film was formed, with the rated voltage being a little less than the forming voltage. Rated voltages ranging all the way from approximately 25 to 550 volts peak are obtainable. When the applied voltage exceeds the rated value, the leakage current becomes excessive, as shown in Fig. 84, and finally scintillations are produced as a result of breaking down of the dielectric film.

The heating produced by the d-c leakage current and the alternating currents passing through an electrolytic condenser must not produce excessive temperature rise. This means in general that the compact types of condensers, such as those using etched foil, and particularly the dry type, are at a disadvantage as compared with bulkier constructions for certain types of service, as, for example, when a large alternating current is present.

The life of an electrolytic condenser depends upon the purity of the materials used in the condenser and upon the conditions of use, particularly the applied voltage and the temperature. Leakage is an indication of life, since leakage causes the electrolyte to be used up, particularly in the wet type, with a corresponding increase in condenser power factor. The life is accordingly increased by anything that reduces the leakage current, such as operating at appreciably below the forming voltage, and by avoiding high temperatures.

SHIELDING OF MAGNETIC AND ELECTROSTATIC FIELDS

35. Conducting Shields for Magnetic Fields.—Magnetic and electrostatic fields can be confined to restricted spaces, or can be prevented from entering a particular space, by the use of suitable shields. The most effective means of controlling magnetic fields at radio frequencies is by the use of shields consisting of a good conductor, such as copper or aluminum. Magnetic flux penetrates such a shield only with great difficulty, because as the flux cuts into the conducting material it produces eddy currents that oppose the penetration. To be most effective, such shields must completely inclose the space to be protected, and all joints should be well lapped to minimize the resistance offered to the eddy currents. Where the shielding must be extremely effective, as in signal generators, joints should be soldered wherever possible, and in other cases should be sufficiently good mechanically to be literally watertight. The thickness should be a number of times the skin depth of current penetration as calculated by Eq. (5), and so has a minimum value determined by the conductivity and frequency.

Shielded Coils.[1]—One of the most important examples of a conducting shield is the copper or aluminum can frequently placed around air-cored coils. Such shields serve to confine practically all of the magnetic field to the space within the can. They also reduce the effective inductance of the coil by restricting the cross-sectional area through which the flux can pass, and increase the losses as a result of the energy dissipated in the shield by the eddy currents. However, if the clearance between the coil and its shield is everywhere at least equal to the radius of the coil, the reduction in inductance and the decrease in coil Q produced by a copper or aluminum shield will not be serious.

The loss of inductance resulting from a cylindrical shield that is long compared with the length of the coil and the shield diameter is given approximately by the curves

[1] An excellent theoretical discussion of the effect of shielding on coil resistance and inductance is given by A. G. Bogle, The Effective Resistance and Inductance of Screened Coils, *Jour. I.E.E.*, Vol. 87, p. 299, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 221, September, 1940.

of Fig. 87.[1] The results also apply with satisfactory accuracy for finite lengths if the clearance between the ends of the shield and the coil is at least the coil radius, and the shield length exceeds the shield diameter. As these limiting conditions are approached, the actual inductance is slightly less than that calculated from Fig. 87.

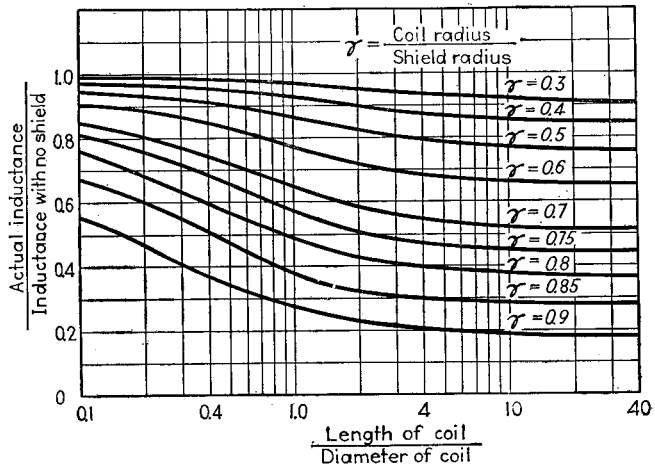


FIG. 87.—Curves giving effect of shield on the inductance on an air-cored coil.

Where Fig. 87 can be expected to be in error, the effect of the shield, including the end effects, can be more accurately taken into account by the equation[2]

$$\frac{\text{Actual inductance}}{\text{Inductance in absence of shield}} = \left[1 - \left(\frac{r_c}{r_s}\right)^2 \frac{l_c}{l_s}\frac{1}{K}\right] \tag{149}$$

where $\frac{l_c}{l_s} = \frac{\text{length of coil}}{\text{length of shield}}$.

$\frac{r_c}{r_s} = \frac{\text{radius of coil}}{\text{radius of shield}}$.

K = constant given by Fig. 88.

The resistance R_s that must be considered as added in series with the shielded coil in order to take into account the shield losses is given approximately by the equation

$$R_s = \frac{9.37 \times 10^{-4} N^2 r_c^4 \sqrt{f\rho}}{r^4} \tag{150}$$

where N = number of turns in coil.

r_c = radius of coil.

$r = \sqrt[3]{(\text{shield diameter})^2(\text{shield length})/8}$.

f = frequency in cycles.

ρ = resistivity of shield in ohms per cm cube ($= 1.724 \times 10^{-6}$ for copper).

[1] Graphic Determination of the Decrease in Inductance Produced by a Coil Shield, *R.C.A. Application Note* 48, June 12, 1935.

These curves can also be applied to the case of a square shielding can that is long compared with the coil length by assuming that the square can is equivalent to a cylindrical shield having a radius that is 0.6 times the length of one side of the can.

[2] Equations (149)–(151) are rearrangements of results given by G. W. O. Howe, The Effect of Screening Cans on the Effective Inductance and Resistance of Coils, *Wireless Eng.*, Vol. 11, p. 155, March, 1934. These equations are derived by replacing the actual shield by a sphere of diameter $2r$ that is the geometric mean of the three coordinate dimensions of the actual shield can. If these three dimensions do not differ too greatly, and if the shield clears the coil by a reasonable amount, then the results calculated for the hypothetical spherical shield approximate very closely those obtained with the actual shield.

This equation assumes that the thickness of the shield is appreciably greater than the skin depth as obtained from Eq. (5) or Fig. 4. When the thickness t in cm is not small compared with the skin depth, then

$$R_s = \frac{4.7N^2\rho}{t}\left(\frac{r_c}{r}\right)^4 \qquad (150a)$$

The effectiveness with which the shield surrounding a coil reduces the external field depends primarily upon the thickness t of the shield in relation to the skin depth ϵ.

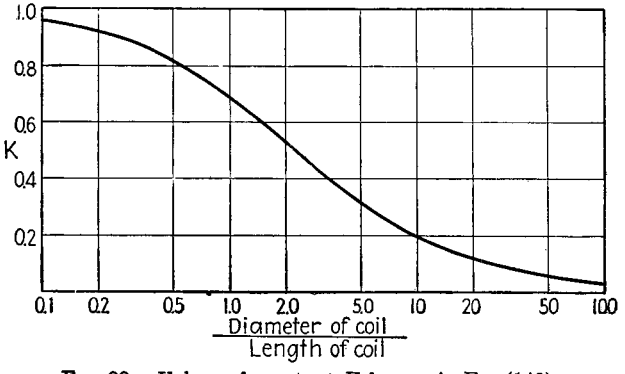


FIG. 88.—Values of constant K for use in Eq. (149).

Where $t/\epsilon > 2$, and where at the same time the shield is thin compared with its length or diameter, one has

$$\frac{\text{Magnetic field in absence of shielding}}{\text{Actual magnetic field}} = 0.24\frac{r}{\epsilon}e^{\frac{t}{\epsilon}} \qquad (151)$$

where r is as defined in Eq. (150), ϵ is the skin depth as calculated by Eq. (5), or obtained from Fig. 4, t is the thickness of the shield, and $e = 2.718$.

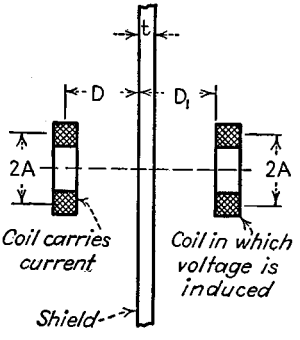


FIG. 89.—Diagram illustrating nomenclature in Eq. (152).

Planar Shield.—An important case of shielding exists where a flat sheet of conducting material is used to reduce the voltage induced in a second coil by current passing through a first coil. For the case illustrated in Fig. 89, where the two coils are coaxial and similar and the shield is a sheet placed at right angles to this axis, one has approximately[1]

$$\frac{\text{Induced voltage in absence of shield}}{\text{Induced voltage with shield}} = 0.0127At\gamma f \qquad (152)$$

where γ = conductivity of shield on basis that conductivity of standard annealed copper is 100.
A = mean radius of coil, cm.
t = thickness of shield, cm.
f = frequency in cycles.

Strictly speaking, Eq. (152) applies only where the shielding is appreciable and the shield is infinite in extent, but practically the results will not be greatly influenced by

[1] Samuel Levy, Electromagnetic Shielding Effect of an Infinite Plane Conducting Sheet Placed between Circular Coaxial Coils, *Proc. I.R.E.*, Vol. 24, p. 923, June, 1936.

finite dimensions, provided the length and width of the shield are considerably greater than $2A$ or $D + D_1$ in Fig. 89, whichever is larger.

Shields Immersed in Uniform Magnetic Field.—Shields are sometimes used to protect a small space from the effects of an external field. The effectiveness of spheres and of long cylinders of conducting material in providing an interior field-free space when the shield is placed in a uniform external field is given by[1]

$$\frac{\text{Field intensity in absence of shield}}{\text{Field in interior of shield}} = 0.707S\frac{r}{\epsilon}e^{\frac{t}{\epsilon}} \qquad (153)$$

where S = a constant that is one-third for spheres and one-half for cylinders.
ϵ = skin depth as given by Eq. (5) or Fig. 4.
t = thickness of shield.
r = radius of sphere or cylinder.
e = 2.718, the base of the Napierian logarithms.

In Eq. (153) it is assumed that $t/\epsilon > 2$. In the case of cylinders, external field is assumed to be parallel to the axis, and the cylinder is considered to be of infinite length. However, if the length is a number of times the diameter, the field at the midpoint of the cylinder will be substantially as calculated, even when the ends are open. The results calculated for the sphere can be used to give an approximate indication of the shielding of cubes and similar shapes by using an equivalent sphere inclosing the same volume as the actual shield.

36. Magnetic Shields.—Shields for unidirectional and low-frequency magnetic fields are made of magnetic material, preferably having a high initial permeability. Such shields act as low reluctance paths for the flux, thereby diverting the flux away from the space to be shielded.

The most important practical use of magnetic shields in communication is to minimize voltages induced in audio transformers by magnetic fields of power frequencies. This is accomplished by placing the transformer in a magnetic shield normally consisting of a rectangular box or a short cylinder with closed ends. The degree of protection obtained in this way depends upon the size and thickness of the shield, and upon whether the shield consists of a single layer of magnetic material or several concentric layers separated by air spaces. In the case of a single-layer shield, the effectiveness is approximately[2]

$$\frac{\text{Magnetic field in absence of shielding}}{\text{Magnetic field with shielding}} = 0.22\mu\left[1 - \left(1 - \frac{t}{r_0}\right)^3\right] \qquad (154)$$

where μ = initial permeability of shield.
r_0 = radius of sphere enclosing the same volume as the outer surface of the shield.
t = thickness of shield.

Equation (154) shows that the maximum possible shielding factor obtainable is 0.22μ and that approximately 50 per cent of this possibility is realized when the thickness of the shield is one-fifth of the radius of the equivalent sphere. Relatively thick single-layer shields are accordingly uneconomic.

When the degree of shielding required is greater than that obtainable from a single shield, two or three concentric magnetic shielding shells are employed, as shown in Fig. 90. When the total thickness of such a shield is greater than one-third the radius of the equivalent sphere, this subdivision of the magnetic material into concentric layers separated by air spaces gives more effective shielding than if the air spaces were filled solidly with magnetic material (see Fig. 90).

[1] Walter Lyons, Experiments on Electromagnetic Shielding at Frequencies between One and Thirty Kilocycles, *Proc. I.R.E.*, Vol. 21, p. 574, April, 1933.
[2] W. B. Ellwood, Magnetic Shields, *Bell Lab. Rec.*, Vol. 17, p. 93, November, 1938.

The effectiveness of concentric magnetic shields to alternating currents can be greatly increased by placing copper shields in the air spaces between the magnetic shields.[1] The behavior in a typical case with and without copper shields is shown in Fig. 91 as a function of frequency over the audio range. With unidirectional (d-c) fields, the copper has no effect upon shielding.

The effectiveness of magnetic shields depends primarily upon the initial permeability and is roughly proportional to μ^n where n is the number of concentric layers. Permalloy and similar magnetic materials having high initial permeability are,

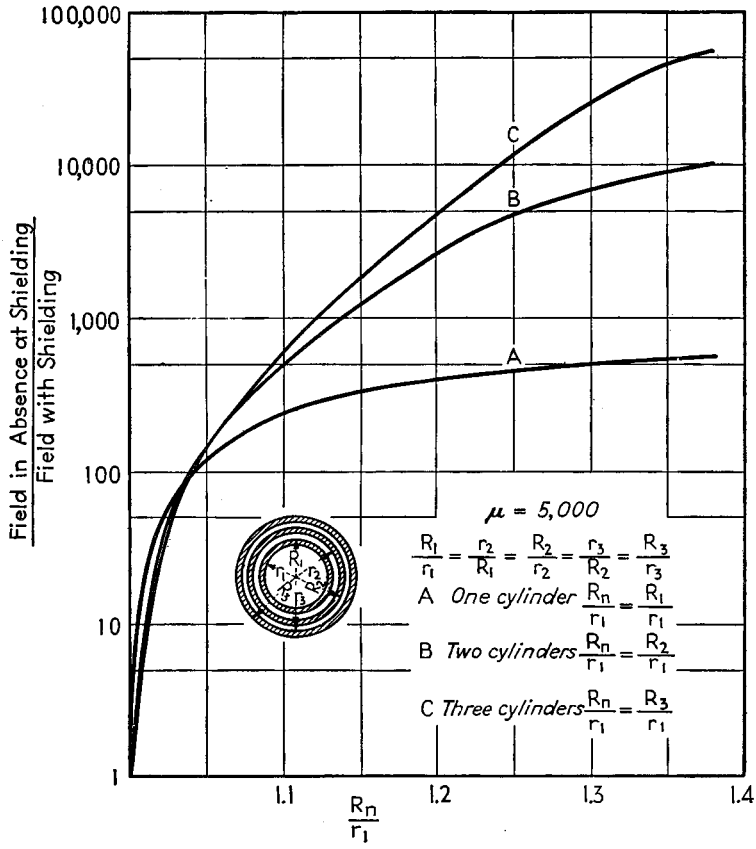


FIG. 90.—Performance of spherical magnetic shields in a typical case.

accordingly, vastly superior to ordinary cast iron or silicon steel, particularly in the case of multilayer shields.

A magnetic shield inclosing an air-cored coil causes the inductance to increase as a result of lowered reluctance offered to the flux paths by the high permeability material. The effective resistance of the shielded coil is also increased because of eddy-current and hysteresis losses in the iron.

37. Electrostatic Shields.—Any conductor can be used as an electrostatic shield. There are no particular requirements as to thickness or conductivity; it is necessary

[1] W. G. Gustafson, Magnetic Shielding of Transformers at Audio Frequencies, *Bell System Tech. Journ.*, Vol. 17, p. 416, July, 1938.

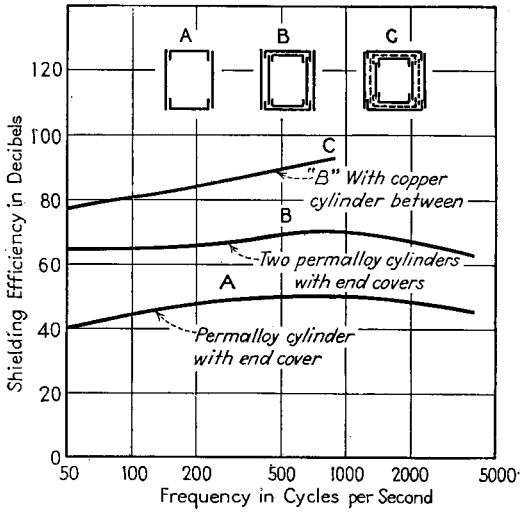


FIG. 91.—Curves showing effectiveness of various typical shielding arrangements for low-frequency fields.

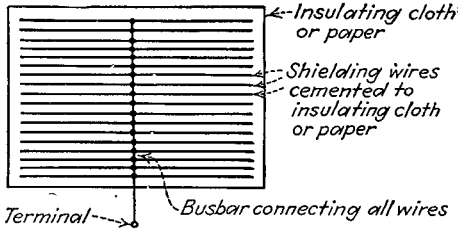


FIG. 92.—Electrostatic shield that does not affect magnetic flux.

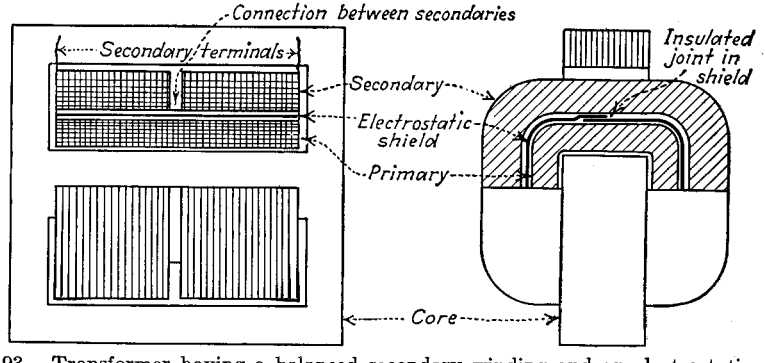


FIG. 93.—Transformer having a balanced secondary winding and an electrostatic shield between primary and secondary.

merely that the electrostatic fields terminate on the shield instead of passing through to the space being protected. To be most effective, electrostatic shields should be continuous closed surface, but even a screen will give fair shielding. The conducting and magnetic shields for magnetic flux discussed above therefore simultaneously provide electrostatic shielding along with the magnetic shielding.

Electrostatic Shielding without Magnetic Shielding.—Electrostatic shielding can be obtained without affecting magnetic flux by surrounding the space to be protected with a conducting nonmagnetic shield arranged to provide termination for electrostatic flux lines without providing closed loops of low resistance around which eddy currents can circulate. Such a shield is illustrated in Fig. 92.

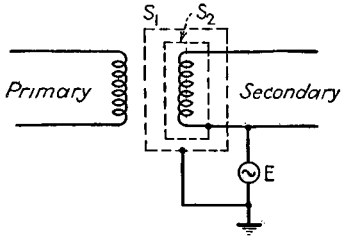


FIG. 94.—Schematic diagram of a double electrostatic shield.

Electrostatic shielding is frequently used in transformers to prevent electrostatic coupling between the various windings. Such a shield consists of a metal foil wrapped around the windings so as to inclose them completely but with an insulated lapped joint, as shown in Fig. 93, so that the foil does not act as a short-circuited turn. Such a shield replaces direct capacity coupling between coils by capacity from the individual coil to the shield (*i.e.*, to the ground).

Double electrostatic shields are sometimes required in transformers used in special applications. Thus in Fig. 94, if the secondary is at an a-c potential E above ground and it is absolutely necessary that this potential produce no voltage difference between the secondary terminals of the transformer, it is then necessary to employ a double shield around the secondary as shown. The outer shield S_1 is grounded, while the inner shield is insulated from S_1 and connected to one terminal of the secondary. In this way the capacity current flowing to ground as a result of the voltage E will not go through any part of either the secondary or primary winding, and so will not produce any voltage difference between secondary terminals.

SECTION 3

CIRCUIT THEORY

SIMPLE RESONANT CIRCUITS

1. Series Resonance.—A circuit consisting of a resistance, inductance, and capacity connected in series with a voltage applied as shown in Fig. 1 is termed a *series resonant circuit.* When the circuit resistance is low, as is normally the case, the current depends upon the frequency of the applied voltage in accordance with the manner shown in Fig. 2*a*. The current is a maximum and in phase with the applied voltage at the frequency, termed the resonant frequency, at which the inductive reactance and capacitive reactances are equal. At lower frequencies the current falls off and leads, while at higher frequencies it drops and lags.

The shape of the series resonance curve of current is determined by the ratio of the inductive reactance ωL to the circuit resistance R, *i.e.*, upon $\omega L/R$. This ratio is customarily represented by the symbol Q, and is called the *circuit Q*. Its numerical value commonly approximates the Q of the coil used in the resonant circuit because most of the loss in resonant circuits is usually in the coil, but will always be at least slightly less then the coil Q because of the condenser losses. The effect upon the resonance curve of changing the circuit resistance, and hence the circuit Q, is shown in Fig. 2*a*. Increasing the circuit resistance reduces the response in the vicinity of resonance, although producing virtually no effect at frequencies differing appreciably from the resonant frequency.

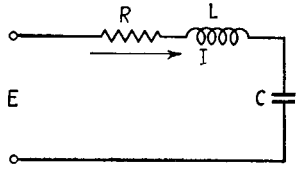
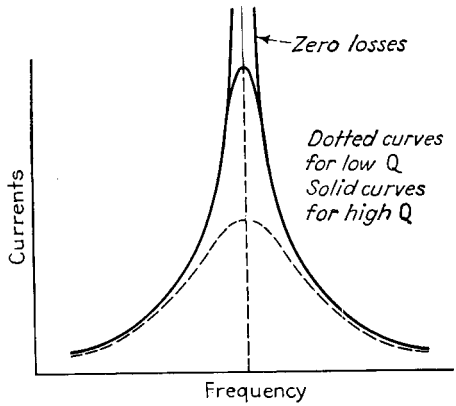
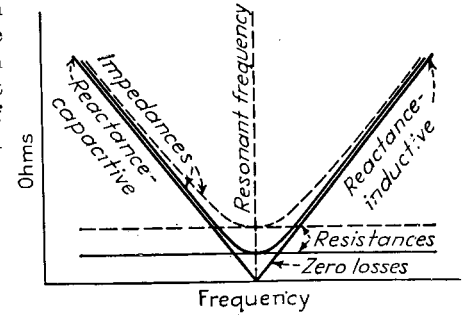


FIG. 1.—Series resonant circuit.



(a) Typical series resonance curves



(b) Corresponding impedance curves

FIG. 2.—Current and impedance of a series resonant circuit as a function of frequency.

The resistance and reactance components of the impedance of the series circuit vary with frequency as shown in Fig. 2*b*. In the small frequency range represented by the region around resonance, the circuit resistance is substantially constant. The reactance, on the other hand, varies substantially linearly from a relatively high capacitive value below resonance to a relatively high inductive value above resonance, and passes through zero at the resonant frequency.

General Equations of Series Resonance.—The equations given below use the following symbols:

E = voltage applied to circuit.
I = current flowing in circuit.
f = frequency in cycles.
$\omega = 2\pi f$.
$Q = \omega L/R$.
R = effective series resistance of tuned circuit.
L = inductance, henrys.
C = capacity, farads.
Z = impedance of series circuit.
θ = phase angle of impedance.
$\gamma = \dfrac{\text{actual frequency}}{\text{resonant frequency}}$.

Subscript $_0$ denotes values at resonant frequency.

$$\text{Resonant frequency} = \frac{1}{2\pi\sqrt{LC}} \tag{1}$$

$$\text{Circuit impedance} = Z = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{2a}$$

$$|Z| = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{2b}$$

$$\tan\theta = \frac{\omega L - \dfrac{1}{\omega C}}{R} = Q\left(1 - \frac{1}{\gamma^2}\right) \tag{2c}$$

$$I = \frac{E}{Z} = \frac{E}{R + j\left(\omega L - \dfrac{1}{\omega C}\right)} \tag{3}$$

$$\text{Voltage across } L = j\omega LI \tag{4a}$$

$$\text{Voltage across } C = \frac{I}{j\omega C} \tag{4b}$$

$$\text{Current at resonance} = I_0 = \frac{E}{R} \tag{5}$$

$$\text{Voltage across } L \text{ or } C \text{ at resonance} = EQ \tag{6}$$

Universal Resonance Curves.—Equation (3), when rearranged to express the current actually flowing in the circuit at resonance in terms of the circuit Q and the fractional detuning, gives the *universal resonance* curves shown in Fig. 3,[1] from which the exact resonance curve of any series circuit may be obtained without calculation when Q is known. These curves are extremely useful, because they are independent of the resonant frequency of the circuit and of the ratio of inductance to capacity.

The use of the universal resonance curves in practical calculations can be illustrated by two examples.

Example 1.—It is desired to know how many cycles one must be off resonance to reduce the current to one-half the value at resonance when the circuit has a Q of 125 and is resonant at 1,000 kc. Refer-

[1] For the derivation of these curves, see F. E. Terman, "Radio Engineering," 2d ed., p. 54, McGraw-Hill, New York, 1937.

ence to Fig. 3 shows that the response is reduced to 0.5 when $a = 0.86$. Hence

$$\text{Cycles off resonance} = \frac{0.86 \times 1{,}000}{125} = 6.88 \text{ kc}$$

The phase angle of the current as obtained from the curve is 60°.

Example 2.—With the same circuit as in the preceding example, it is desired to know what the response will be at a frequency 10,000 cycles below resonance. To solve this problem it is first necessary to determine a.

$$a = \left(\frac{10}{1{,}000}\right) \times 125 = 1.25$$

Reference to Fig. 3 shows that for $a = 1.25$, the response is reduced by a factor 0.365 and that the phase of the current is 68° leading.

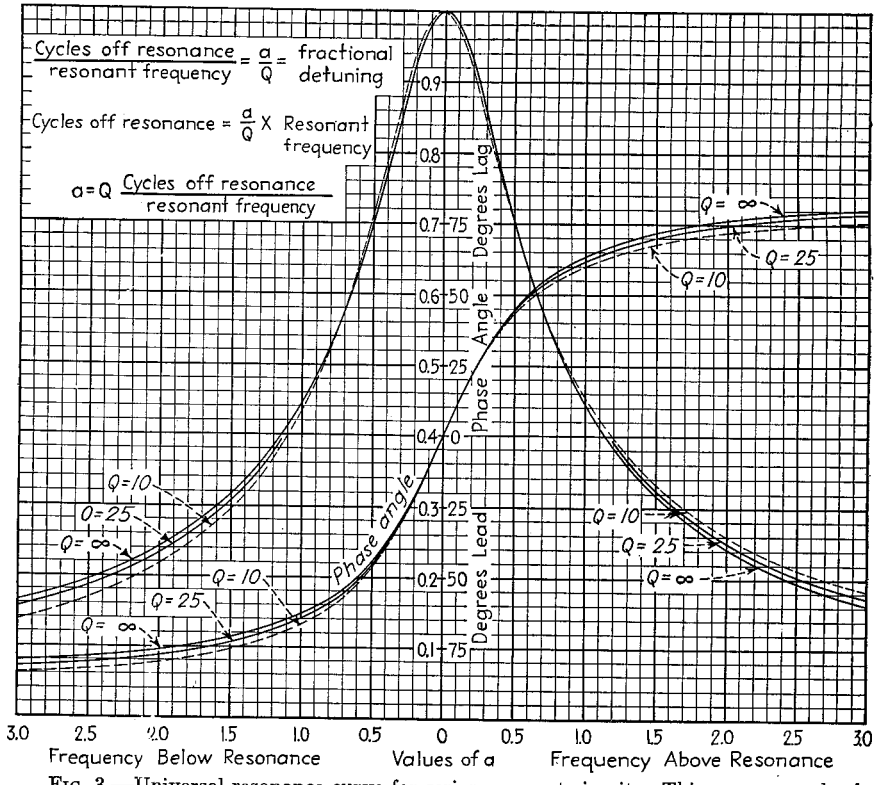


FIG. 3.—Universal resonance curve for series resonant circuit. This curve can also be applied to the parallel resonant circuit by considering the vertical scale to represent the ratio of actual parallel impedance to the parallel impedance at resonance. When applied to parallel circuits the angles shown in the figure as leading are lagging, and *vice versa*.

The only assumption involved in the universal resonance curves is that Q is the same at the frequency being considered as at the resonant frequency. *When this is true, the universal resonance curves involve no approximations whatsoever.* Over the limited range of frequencies near resonance represented in Fig. 3, the variations in Q are so small as to introduce negligible (*i.e.*, less than 1 per cent) errors from the use of the curves, provided the value of Q existing at resonance is used in determining the fractional detuning.

The universal resonance curve also gives the resonance curve of voltage developed across the inductance or capacity with an accuracy sufficient for all ordinary purposes.

This is because under ordinary conditions the frequency range covered by the curves extends only a few per cent on either side of resonance. For this limited range of frequencies, the inductive (and capacitive) reactances are so nearly constant that the product of current and reactance gives a curve having essentially the same shape as the current curve.

Working Rules for Dealing with Series Resonant Circuits.—In making calculations of series resonant circuits or in estimating their behavior, the following rules and principles will frequently be found of value:

The resonance curve can for all practical purposes be considered as symmetrical about the resonant frequency. This is apparent from an examination of the universal resonance curves.

The current at resonance is always E/R, while the voltage developed across the capacitive (or inductive) reactance at resonance is Q times the applied voltage and is 90° out of phase with this voltage.

When the frequency of the applied voltage deviates from the resonant frequency by an amount that is $1/2Q$ of the resonant frequency, the current that flows is reduced to 70.7 per cent of the resonant current, and the current is 45° out of phase with the applied voltage. Correspondingly, when the fractional detuning is $1/Q$, the current is approximately 45 per cent of the current at resonance, and is 63.5° out of phase with the voltage.[1]

At frequencies that differ from the resonant frequency by amounts that exceed $3/Q$ of the resonant frequency, one can neglect the circuit resistance in determining the magnitude of the circuit impedance. The magnitude of the current is then

$$\frac{\text{Actual current when far off resonance}}{\text{Current at resonance}} = \frac{1}{Q\gamma\left(1 - \frac{1}{\gamma^2}\right)} \tag{7}$$

where $\gamma = \dfrac{\text{actual frequency}}{\text{resonant frequency}}$.

The phase angle of the current can be calculated without error by the use of Eq. (2c).

$$\tan\theta = Q\left(1 - \frac{1}{\gamma^2}\right)$$

[1] Other convenient cases are as follows:

TABLE 1

Fractional detuning $\left(=\frac{\text{cycles off resonance}}{\text{resonant frequency}}\right)$	$\frac{\text{Actual current}}{\text{Current at resonance}}$	Phase of current, °
$\frac{1}{6Q} = \frac{1}{3} \times \frac{1}{2Q}$	0.95	18½
$\frac{1}{4Q} = \frac{1}{2} \times \frac{1}{2Q}$	0.90	26½
$\frac{1}{2Q} = 1 \times \frac{1}{2Q}$	0.707	45
$\frac{1}{Q} = 2 \times \frac{1}{2Q}$	0.447	63½
$\frac{2}{Q} = 4 \times \frac{1}{2Q}$	0.242	76
$\frac{4}{Q} = 8 \times \frac{1}{2Q}$	0.124	83

The voltage across the inductance and also the voltage across the condenser vary with frequency in the immediate vicinity of resonance in a way that is for all practical purposes the same as the current resonance curve. At frequencies differing somewhat from the resonant frequency, there is a slight but ordinarily unimportant difference in shape between the voltage and current curves. At frequencies that deviate from resonance by at least $3/Q$ times the resonant frequency, it is permissible to neglect the resistance in making calculations. This leads to the following formulas:

For voltage across the inductance:

$$\frac{\text{Actual voltage when far off resonance}}{\text{Voltage at resonance}} = \frac{1}{Q\left(1 - \frac{1}{\gamma^2}\right)} \tag{8a}$$

For voltage across the condenser:

$$\frac{\text{Actual voltage when far off resonance}}{\text{Voltage at resonance}} = \frac{1}{Q(\gamma^2 - 1)} \tag{8b}$$

Practical Calculation of Series Resonance Curves.—Series resonance curves can of course be calculated with the aid of Eqs. (1) to (3). Practically, however, this is not a very satisfactory way to obtain them, since in order to obtain reasonable accuracy in the region near resonance, it is necessary to use logarithms in forming the difference $\omega L - 1/\omega C$ in cases where the Q is at all appreciable.

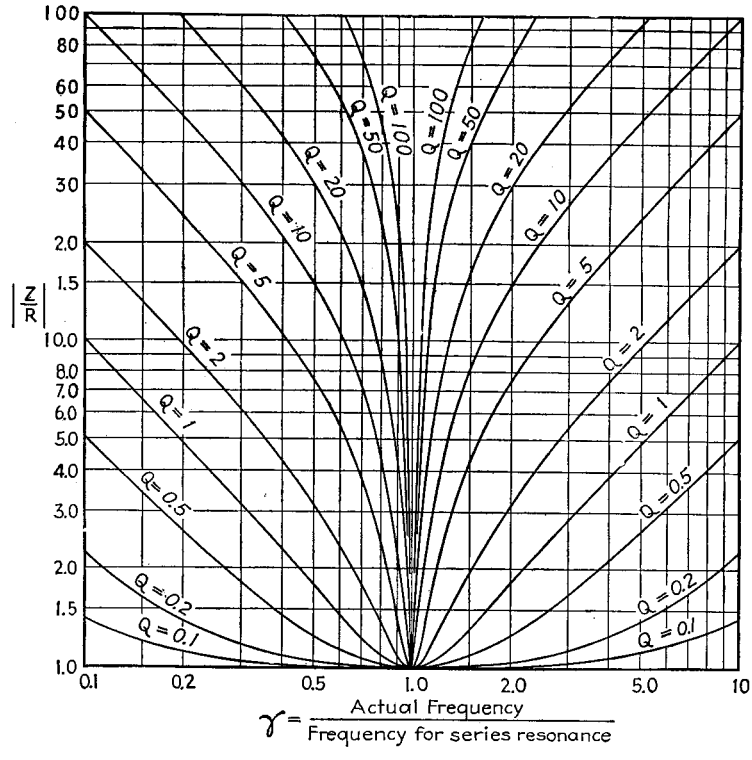
A preferable procedure is first to calculate the resonant frequency and the response at resonance. The former can be conveniently done with the aid of the reactance charts given on the inside back cover of this volume. The working rules given above can then be applied to obtain the response at frequencies $1/2Q$, $1/Q$, etc., on either side of resonance. This gives a picture of the resonance curve that is sufficient for most practical purposes. Additional points in the vicinity of resonance can be calculated when needed with the aid of the universal resonance curve of Fig. 3. Calculations for frequencies too far from resonance to be included on the universal resonance curve can be made by the use of Eq. (8*a*) (or 8*b*) and Eq. (2*c*).

Logarithmic Decrement.—The sharpness of the series resonance curve is normally expressed in terms of the Q of the circuit. In the early days of radio, when the oscillating spark discharge represented the chief source of radio-frequency energy, the sharpness of resonance was expressed in terms of the *logarithmic decrement*, defined by the equation

$$\text{Logarithmic decrement} = \frac{R}{2f_0L} \tag{9}$$

where R and L represent circuit resistance and inductance, respectively, and f_0, the resonant frequency. It will be noted that the logarithmic decrement is equal to π/Q.

Series Resonance by Varying Circuit Reactance.—The preceding discussion of series resonance has been for the case where the circuit resistance, inductance, and capacity were constant and the frequency was varied. Under many circumstances, however, the frequency is kept constant and the circuit is adjusted to series resonance by varying either the capacity or the inductance. When the circuit reactance is varied in this way, the curve of current in the circuit as a function of the capacitive reactance (or inductive reactance) has essentially the same shape as the series resonance curves already considered. In particular, maximum current and unity power factor still occur when the capacitive and inductive reactances are equal, provided the circuit resistance does not change with the change in reactance. The curve of voltage developed across the inductance (or capacity) will, however, not be quite the same when the reactance is varied as in the case of frequency variation, since the maximum will occur at a slightly different frequency.



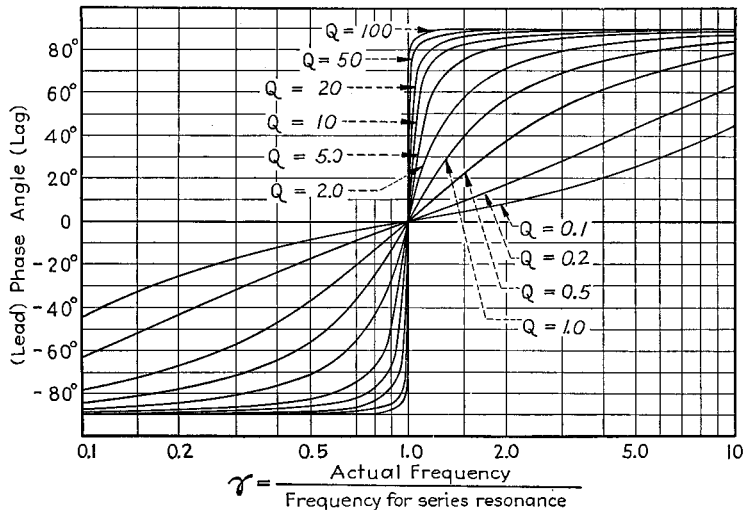


FIG. 4.—Universal curves giving the impedance as a function of frequency for a series resonant circuit having constant resistance.

Series Resonant Circuits with Constant Resistance.—Series resonant circuits having a substantially constant resistance independent of frequency, particularly circuits with a relatively low Q, are often used in equalizers and in other applications. The general expression for the impedance of a series resonant circuit as given in Eq. (2*a*) can for the case of constant circuit resistance be conveniently rewritten as follows:

$$\frac{Z}{R} = 1 + jQ_0\left(\frac{\gamma^2 - 1}{\gamma}\right) \tag{10}$$

where Q_0 = circuit Q at resonance.

$\gamma = \dfrac{\text{actual frequency}}{\text{resonant frequency}}$.

Z = series impedance.

R = circuit resistance (assumed independent of frequency).

Results given by Eq. (10) are expressed graphically in Fig. 4, which not only is useful in showing at a glance the manner in which the magnitude and phase angle of the impedance will vary under different conditions but also can be used in the design of series resonant circuits to have desired characteristics.

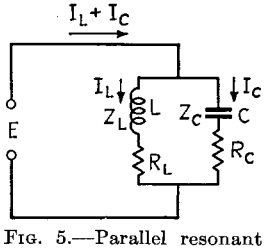


FIG. 5.—Parallel resonant circuit.

2. Parallel Resonant Circuits.—A parallel resonant circuit is obtained when the generator voltage is applied to an inductance and a capacity connected in parallel, as in Fig. 5. It will be noted that the only difference between the series and parallel circuit is in the manner of connection. In the analyses given below of the parallel-resonant circuit of Fig. 5, the following notation will be used:

E = voltage applied to circuit.

$Z_c = R_c - j/\omega C$ = impedance of capacitive branch.

$Z_L = R_L + j\omega L$ = impedance of inductive branch.

$Z_s = Z_c + Z_L$ = series impedance of circuit.

Z = parallel impedance of circuit.

$R_s = R_c + R_L$ = total series resistance of circuit.

$\omega = 2\pi$ times frequency.

$\omega_0 = 2\pi$ times frequency at which $\omega L = 1/\omega C$.

$Q = \omega L/R_s$ = circuit Q.

Q_0 = value of Q at $\omega_0/2\pi$ cycles.

The quantities L, C, R_L, and R_c refer to the inductance, capacity, and resistance components of the circuit, as indicated in Fig. 5.

The impedance offered to the generator by a parallel resonant circuit is the product of the two branch impedances divided by their sum; *i.e.*

$$\text{Parallel impedance} = Z = \frac{Z_c Z_L}{Z_c + Z_L} = \frac{Z_c Z_L}{Z_s} \tag{11}$$

$$= \frac{\left(R_c + \dfrac{1}{j\omega C}\right)(R_L + j\omega L)}{(R_c + R_L) + j\left(\omega L - \dfrac{1}{\omega C}\right)} \tag{11a}$$

The magnitude and phase of the impedance of a parallel circuit can be conveniently obtained from the universal curves of Fig. 6, which give the parallel impedance in terms of the impedance of one branch and the vector ratio of the branch impedances.[1]

[1] Somewhat similar curves are given by D. G. Fink, Parallel Impedance Chart, *Electronics*, Vol. 11, p. 31, May, 1938.

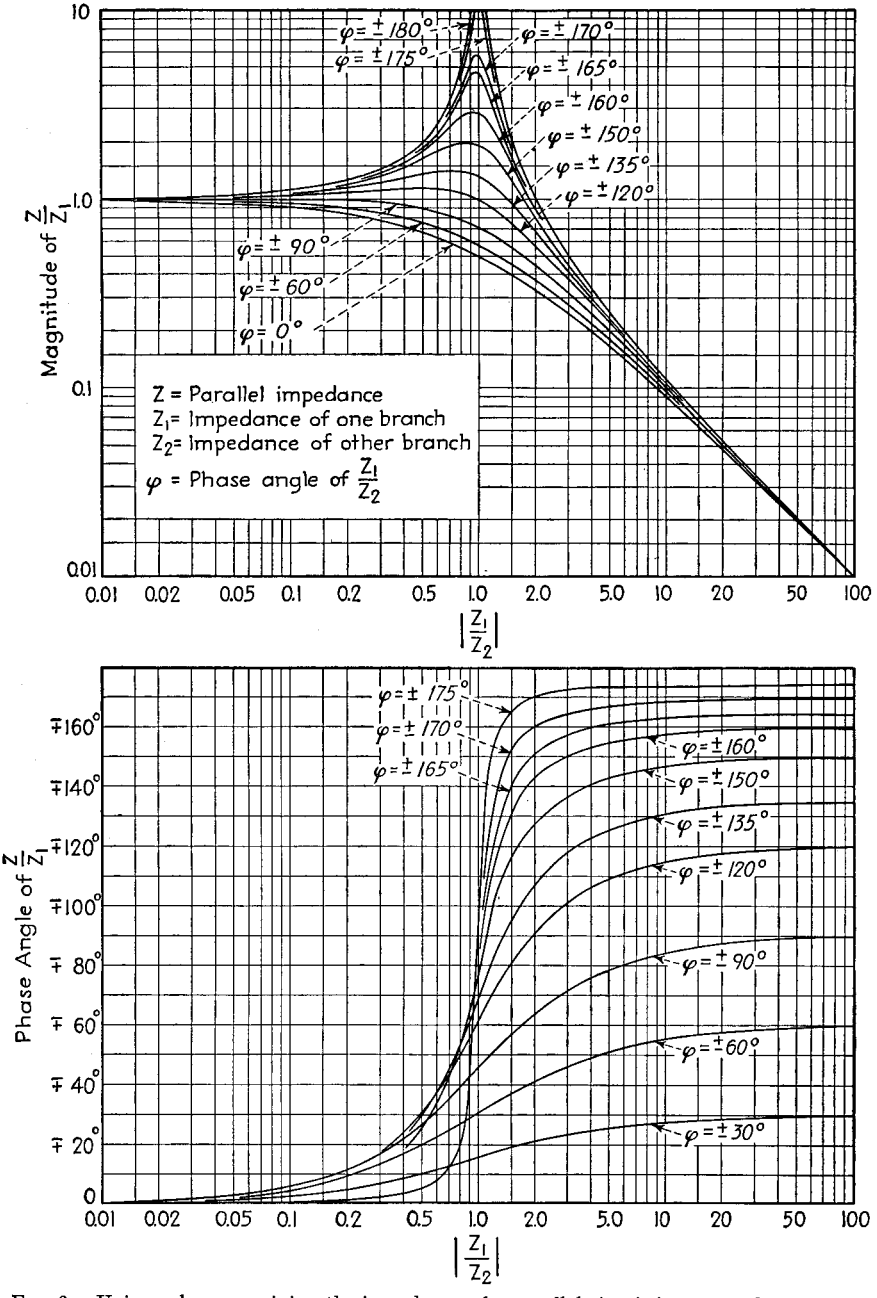


FIG. 6.—Universal curves giving the impedance of a parallel circuit in terms of the magnitudes and phases of the branch impedances.

The current that the generator delivers to the parallel circuit can be termed the *line current*, and is the applied voltage E divided by the impedance or

$$\text{Line current} = \frac{E}{Z} \tag{12}$$

The currents in the branches are equal to the applied voltage divided by the respective branch impedances, and so are given by the equations

$$\text{Current in capacitive branch} = I_c = \frac{E}{Z_c} \tag{13a}$$

$$\text{Current in inductive branch} = I_L = \frac{E}{Z_L} \tag{13b}$$

Equations (11), (12), and (13) are the fundamental equations of the parallel circuit, and apply for all conditions, irrespective of circuit Q, frequency, or division of resistance between branches.

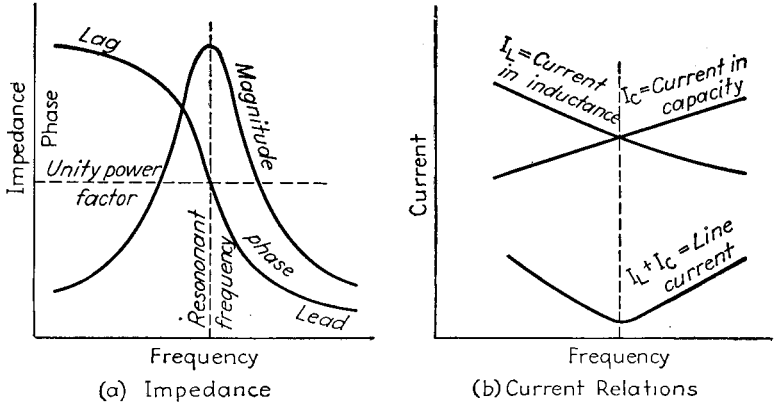


FIG. 7.—Variation of impedance and current with frequency in a parallel resonant circuit.

When the Q of a parallel circuit is not too low, the impedance that is offered to an applied voltage varies with frequency in the manner illustrated in Fig. 7*a*. At low frequencies, the inductive branch draws a large lagging current while the leading current of the capacitive branch is small, resulting in a large lagging line current and a low lagging circuit impedance. At high frequencies, the capacitive branch has a low reactance compared with the inductive branch, resulting in a large leading line current and a corresponding low circuit impedance that is leading. Between these extremes, there is a frequency, commonly termed the resonant frequency, at which the lagging current taken by the inductive branch and the leading current in the capacitive branch are equal, and since these currents are nearly 180° out of phase, they neutralize to leave only a small resultant in-phase line current. The impedance of the parallel circuit will then be a high resistance, as is apparent in Fig. 7*a*.

As a result of the fact that the parallel circuit has a high impedance at resonance and relatively low impedance at other frequencies, the line current varies with frequency as shown in Fig. 7*b* and becomes very small at resonance. The currents in the individual branches vary much more slowly with frequency, and are almost but not quite constant in the vicinity of resonance.

The resonant frequency of a parallel circuit can be defined in several ways as follows: (1) the frequency at which $\omega L = 1/\omega C$, *i.e.*, the resonant frequency of the same

circuit when operating in series resonance; (2) the frequency at which the parallel impedance of the circuit is maximum; (3) the frequency at which the parallel impedance of the circuit has unity power factor. These three definitions lead to frequencies that differ by an amount depending upon the circuit Q and the division of the resistance between the branches. A discussion of the practical importance of these various cases is given below.

Analysis of Parallel Resonant Circuits Having High Q.—The most important practical case of parallel resonance exists when the circuit Q is at least moderately high (*i.e.*, $Q \gtrsim 10$). Under these conditions, the exact way in which the resistance is divided between the two branches is of little consequence, and it is permissible to assume that the absolute magnitude of the impedance of the individual branches is not affected by the resistance in these branches.

To the accuracy allowed by these approximations, the three possible resonant frequencies are virtually identical, so that *for all practical purposes the resonant frequency of a high Q parallel resonant circuit can be taken as the resonant frequency of the same circuit when connected in series.*

The shape of the impedance curve of a high Q parallel resonant circuit is also the same as the current curve of the same circuit acting as a series resonant circuit. The only difference between series and parallel resonance under these conditions is that the phase angles are reversed, *i.e.*, the phase angle of parallel impedance has the opposite sign from the phase angle of the current in the series circuit.

In terms of the notation given above, the equations giving the behavior of parallel resonant circuits with moderate or high Q's are

$$\text{Resonant frequency} = f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{14}$$

$$\text{Parallel impedance} = \frac{(\omega_0 L)^2}{Z_s} = \frac{\left(\frac{1}{\omega_0 C}\right)^2}{Z_s} = \frac{\left(\frac{L}{C}\right)}{Z_s} \tag{15}$$

At resonance

$$\left.\begin{matrix}\text{Parallel impedance} \\ \text{at resonance}\end{matrix}\right\} = \frac{(\omega_0 L)^2}{R_s} = Q_0\omega_0 L = R_s Q_0^2 = \frac{L}{RC} \tag{16}$$

At frequencies differing from resonance by at least $3/Q$ times the resonant frequency, it is permissible to neglect the effect of the circuit resistance R in determining the magnitude of the parallel impedance. This gives

$$\frac{\text{Impedance at frequencies far from resonance}}{\text{Impedance at resonance}} = \frac{1}{Q\gamma\left(1 - \frac{1}{\gamma^2}\right)} \tag{17a}$$

where γ is the ratio of the actual to the resonant frequency. The phase angle at all frequencies is very nearly the negative of the angle of the current for series operation, and from Eq. (2*c*) is found to be

$$\tan\theta = Q\left(\frac{1}{\gamma^2} - 1\right) \tag{17b}$$

The recommended procedure for calculating parallel resonance in the case of moderate or high Q circuits is as follows: First, obtain the resonant frequency and the impedance at resonance, using Eqs. (14) and (16). Second, calculate the frequencies at which the impedance is 70.7 per cent and 44.7 per cent of the value at resonance with the aid of the same working rules as for series resonance, and then where other points are needed in the vicinity of resonance use either the universal resonance curve or Table 1. Finally, for frequencies too far off resonance to be within the range

of the universal resonance curve, the magnitude and phase of the impedance are obtained by Eqs. (17*a*) and (17*b*). This procedure gives an accuracy far better than obtainable by slide-rule calculations using Eq. (11), and also involves considerably less work.

The impedance offered by a parallel resonant circuit can be expressed in form $R_{eq} + jX_{eq}$. A universal curve giving the variation of R_{eq}/Z_0 and X_{eq}/Z_0 as a function of frequency, where Z_0 is the impedance at resonance, is shown in Fig. 8 for the case of a high Q circuit. It will be noted that the curve of resistance is much like a resonance curve, but differs in that it has steeper sides. In particular, the resistance drops to 50 per cent of the resonant impedance when the fractional detuning is $1/2Q$

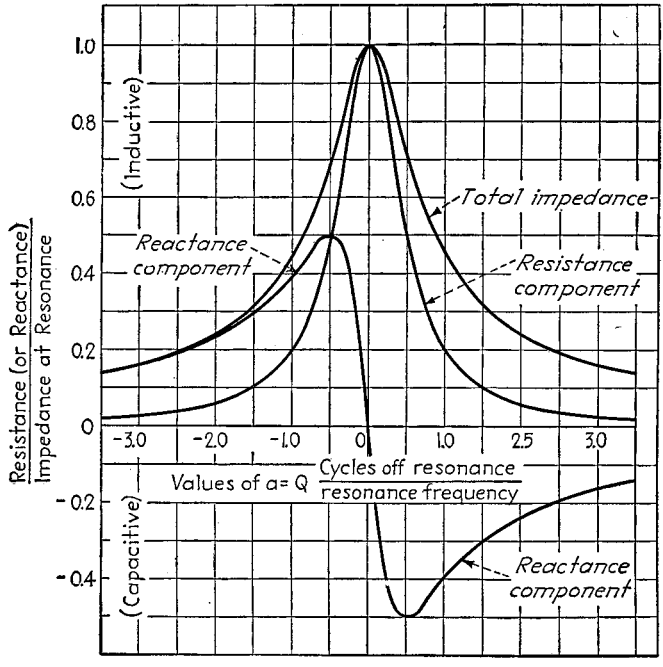


FIG. 8.—Universal curve giving the resistance and reactance components of the impedance of a high-Q parallel resonant circuit as a function of frequency.

instead of 70.7 per cent, as in the case of a resonance curve. The reactance curves of Fig. 8 are characterized by maximum inductive and capacitive reactances at frequencies that are likewise $1/2Q$ below and above resonance, respectively. These reactance maxima have an amplitude exactly 50 per cent of the impedance at resonance.

General Case of Parallel Resonance with Particular Reference to Low Q Circuits.—In the general case where the circuit Q is low, the curve of circuit impedance still tends to have the shape of a resonance curve unless the circuit Q approaches (or is less than) unity, but the peak of the curve does not necessarily occur at the frequency of series resonance, and the condition for unity power factor does not necessarily occur either at this frequency or when the impedance is a maximum. Furthermore, the details of the behavior of a low Q parallel circuit also depend upon the division of resistance between the inductive and capacitive branches, upon the way in which the resistance varies with frequency, and upon whether the adjustment to resonance is made by

varying the frequency, inductance, or capacity. Some of the effects that can be obtained under different conditions are illustrated in Fig. 9.

The behavior of the impedance for any desired condition can be calculated with the aid of Eq. (11). Special relations can also be derived for particular cases, but there are so many of these special cases, and each is of such limited application that the equations are not particularly useful.[1] The results in a number of limiting cases

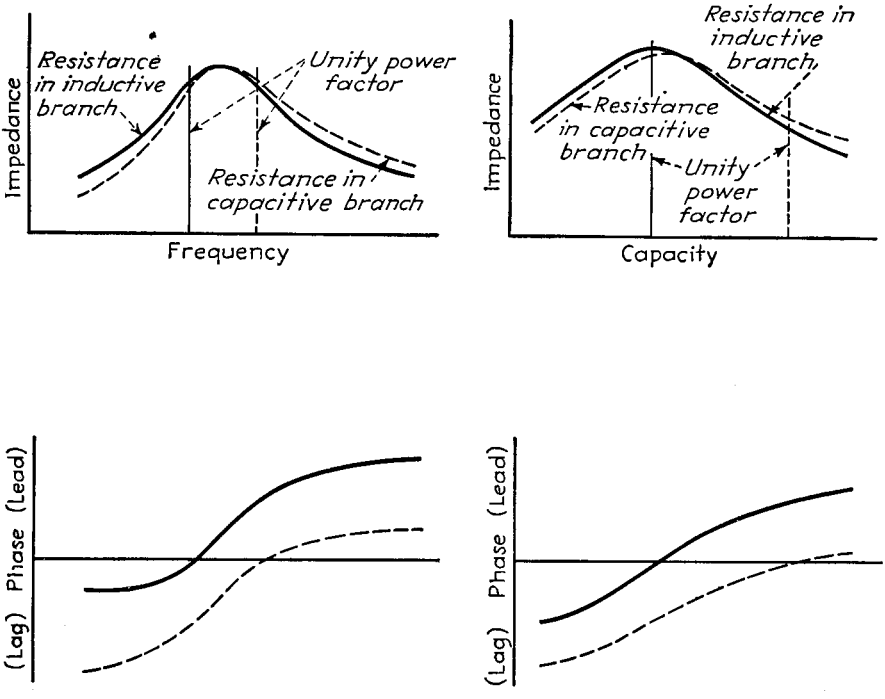


FIG. 9.—Typical characteristics of parallel resonant circuits having low Q.

are as follows: *First*, when the R, L, and C of the circuit are constant, and all of the resistance is in one branch, the frequency of unity power factor is related to the frequency of series resonance by the following equations:

For resistance in inductive branch:

$$\omega = \omega_0 \sqrt{1 - \frac{1}{Q_0^2}} \tag{18}$$

For resistance in capacitive branch:

$$\omega = \omega_0 \frac{1}{\sqrt{1 - \frac{1}{Q_0^2}}} \tag{19}$$

The notation is the same as previously used, *i.e.*, Q_0 is the circuit Q at the series resonant frequency, and $\omega_0/2\pi$ corresponds to the series resonant frequency. *Second*, when the frequency is varied, the impedance curve is not greatly affected by the

[1] A good discussion of some of these special cases is given in pp. 35–45 of R. S. Glasgow, "Principles of Radio Engineering," McGraw-Hill, New York, 1936.

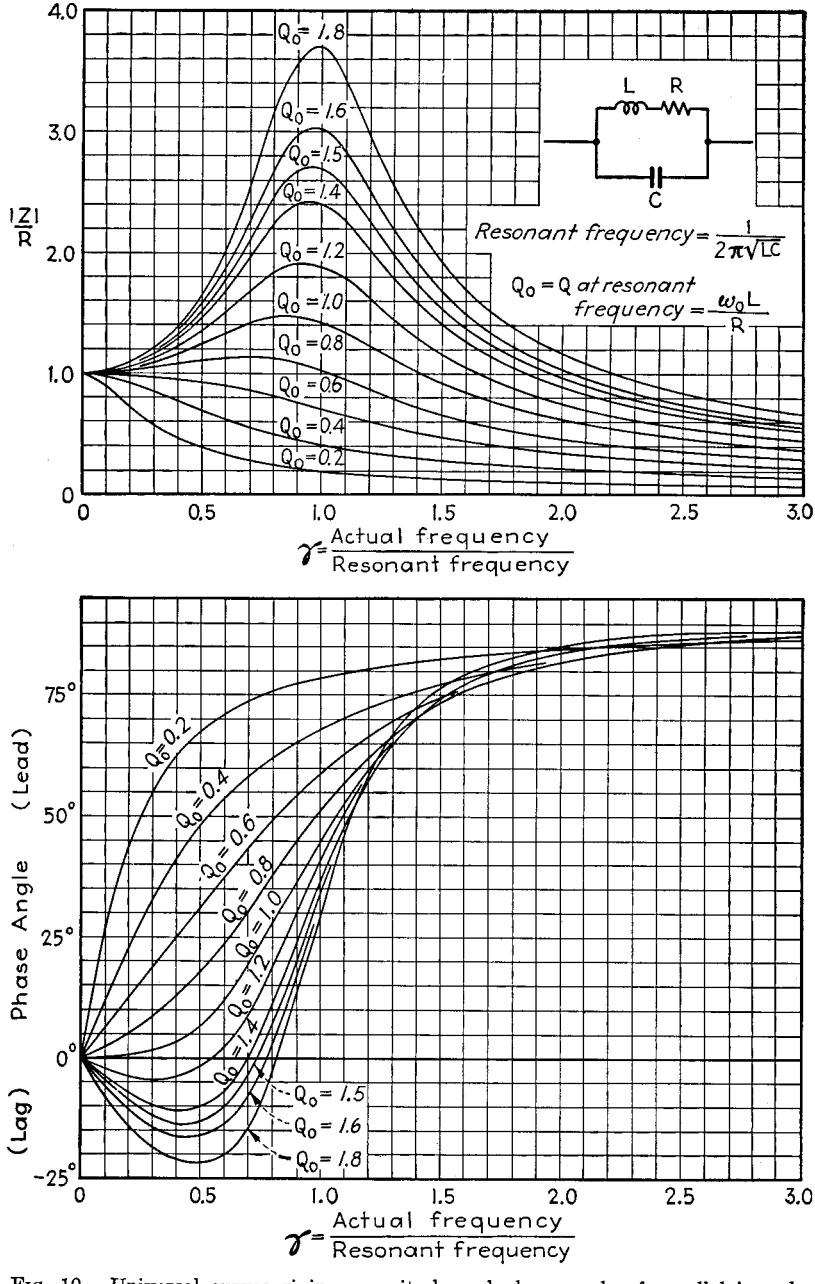


FIG. 10.—Universal curves giving magnitude and phase angle of parallel impedance as a function of frequency in parallel resonant circuits having low Q and constant circuit resistance.

division of resistance between the branches, and the maximum impedance occurs very near, but not necessarily exactly at, the frequency of series resonance. *Third*, when all the resistance is in the inductive branch and the circuit capacity is varied while keeping resistance, inductance, and frequency constant, the capacity giving maximum impedance also corresponds to unity power factor. *Fourth*, when all the resistance is in the capacitive branch and the inductance is varied while keeping resistance, capacity, and frequency constant, the inductance giving maximum impedance also corresponds to unity power factor. *Fifth*, tuning by adjusting circuit by any other manner than the preceding cases, as, for example, by varying the inductance when the resistance is in the inductance branch, causes the condition for unity power factor to differ from the condition for maximum impedance.

Special Cases of Very Low Q.—Inductance and capacity are sometimes associated with a constant resistance in such a way as to form a parallel resonant circuit in which the resistance is constant and in the inductive branch, and the Q is extremely low, even less than unity. Typical examples are a resistance unit designed for low reactance effects (see Fig. 8, Sec. 2), and the coupling impedance to give shunt peaking in a video amplifier (see Fig. 51*a*, Sec. 5). Universal curves giving the magnitude and phase angle of impedance for this particular case are given in Fig. 10.

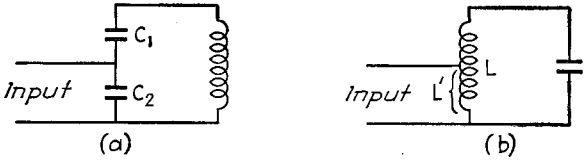


FIG. 11.—Tapped parallel resonant circuits.

Another important case of a very low Q parallel circuit is where the resistances in the two branches are the same, and are each equal to $\sqrt{L/C}$ ohms. Under this condition, the parallel impedance is a resistance at all frequencies, and has a value of $\sqrt{L/C}$ ohms.

Parallel Resonance with Tapped Circuits.—Parallel resonance circuits are frequently operated with the input connected across only a portion of the total circuit, as illustrated in Fig. 11. At frequencies near resonance, the effect of this method of operation is to make the impedance offered to the input terminals less in magnitude than the parallel impedance of the entire circuit without changing the character of the impedance curve as far as shape or equivalent Q is concerned. Tapping a parallel resonance circuit accordingly offers a means of adjusting the magnitude of impedance obtained without changing the circuit itself.

With the circuit of Fig. 11*a*, one has near the resonant frequency

$$\frac{\text{Impedance offered to input}}{\text{Parallel impedance of circuit}} = \left(\frac{C_1}{C_1 + C_2}\right)^2 \tag{20}$$

Similarly, for Fig. 11*b*

$$\frac{\text{Impedance offered to input}}{\text{Parallel impedance of circuit}} = \left(\frac{M_{eq}}{L}\right)^2 \tag{21}$$

where M_{eq} is the total equivalent mutual inductance between L' and the entire coil L, including both common and inductive coupling.

3. Inductively Coupled Circuits.—When mutual inductance exists between coils that are in separate circuits, these circuits are said to be inductively coupled. The effect of the mutual inductance is to make possible the transfer of energy from one circuit to the other by transformer action; *i.e.*, the alternating current flowing in one

circuit as a result of a voltage applied to that circuit produces magnetic flux that induces a voltage in the coupled circuit, resulting in induced currents and a transfer of energy from the first or primary circuit to the coupled or secondary circuit.

Analysis of Inductively Coupled Circuits.—The behavior of inductively coupled circuits can be determined by making use of the following principles:[1] *First*, as far as the primary circuit is concerned, the effect of the presence of the coupled secondary circuit is exactly as though an impedance $(\omega M)^2/Z_s$ had been added in series with the primary, where

M = mutual inductance.
$\omega = 2\pi f$.
Z_s = series impedance of secondary circuit when considered by itself.

Second, the voltage induced in the secondary circuit by a primary current of I_p has a

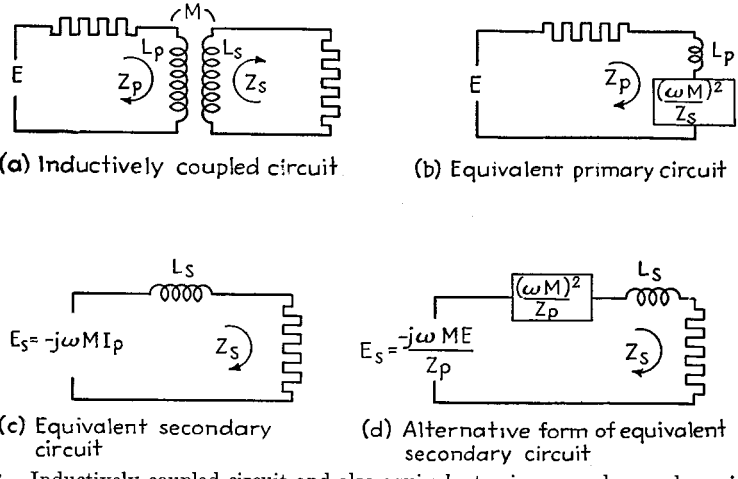


FIG. 12.—Inductively coupled circuit and also equivalent primary and secondary circuits.

magnitude of ωMI_p and lags the current that produces it by 90°. *Third*, the secondary current is exactly the same current that would flow if the induced voltage were applied in series with the secondary and if the primary were absent.

The preceding analysis shows that the actions taking place in the primary and secondary circuit are as though one had the equivalent circuits shown in Figs. 12*b* and 12*c*. Here the primary circuit consists of the actual primary impedance Z_p in series with an additional impedance $(\omega M)^2/Z_s$ that takes into account the effect the secondary produces on the primary current. This extra impedance $(\omega M)^2/Z_s$ is termed the *coupled impedance*. The equivalent secondary circuit of Fig. 12*c* consists of the voltage $-j\omega MI_p$ induced in the secondary by the primary current, in series with the secondary impedance Z_s as shown.

An alternative form of equivalent secondary circuit is given in Fig. 12*d*. This is obtained by applying Thévenin's theorem and noting that when the secondary is opened, the open circuit voltage that appears is $-j\omega MI_p$, while when the applied voltage is short-circuited, the equivalent impedance as viewed from the opened secondary terminals is $Z_s + (\omega M)^2/Z_p$. This second form of the equivalent secondary

[1] These rules for handling coupled circuits are direct consequences of the mesh equations, which for this case can be written as

$$E = I_pZ_p + j\omega MI_s \quad (22)$$

$$\text{Induced voltage} = -j\omega MI_p = I_sZ_s \quad (23)$$

circuit leads to the same results as the first form, since the extra voltage that is considered as acting in the second case is just compensated for by the voltage drop in the extra impedance considered as existing in the secondary circuit.

The resistance and reactance components of the coupled impedance are

$$\text{Resistance component} = \frac{(\omega M)^2}{R_s^2 + X_s^2} R_s \tag{24}$$

$$\text{Reactance component} = -j \frac{(\omega M)^2}{R_s^2 + X_s^2} X_s \tag{25}$$

where R_s and X_s are, respectively, the resistance and reactance components of the secondary impedance Z_s. The effect that the secondary has on the primary circuit is exactly as though the resistance and reactance given by Eqs. (24) and (25) had been inserted in series with the primary circuit. The energy and reactive volt-amperes consumed by the primary current flowing through this hypothetical resistance and reactance represent the energy and reactive volt-amperes that are transferred to the secondary circuit.

Conditions for Maximum Secondary Current.[1]—In dealing with coupled circuits, it is desirable to keep in mind the adjustments giving maximum secondary current. The behavior in the more important cases is summarized below for the case of two circuits inductively coupled, with resistance and reactance in each circuit.

Case 1. Maximum Secondary Current with Variation of Secondary Reactance.—When frequency, mutual inductance, primary impedance, and secondary resistance are constant, the secondary current is maximum when

$$X_s = \frac{(\omega M)^2}{R_p^2 + X_p^2} X_p \tag{26}$$

where X_p and X_s = reactance of primary and secondary circuits, respectively.
R_p and R_s = resistance of primary and secondary circuits, respectively.
$\omega = 2\pi$ times frequency.
M = mutual inductance.

Case 2. Maximum Secondary Current with Variation of Primary Reactance.—If the primary rather than secondary reactance is varied, maximum secondary current is obtained when

$$X_p = \frac{(\omega M)^2}{R_s^2 + X_s^2} X_s \tag{27}$$

Case 3. Maximum Possible Secondary Current.—The secondary current has its maximum possible value when the adjustments are such that the resistance and reactance coupled into the primary circuit by the secondary are equal to the resistance of the primary circuit and the negative of the primary reactance, respectively; *i.e.*, when

$$\begin{aligned} \frac{(\omega M)^2}{R_s^2 + X_s^2} R_s &= R_p \\ -\frac{(\omega M)^2}{R_s^2 + X_s^2} X_s &= -X_p \end{aligned} \tag{28}$$

It will be noted that in order to obtain maximum possible secondary current, it is necessary that two independent variables be adjusted. One of these is commonly the mutual inductance; the other, the reactance of one or both circuits. Unless the mutual inductance satisfies the relation $\omega M \geq \sqrt{R_p R_s}$, it is impossible to satisfy the condition for maximum possible secondary current.

[1] For a more detailed discussion, see G. W. Pierce, "Electric Oscillations and Electric Waves," Chap. XI, McGraw-Hill, New York, 1920.

Inductive Coupling Considered as a Transformer.—Two coils inductively coupled represent the general case of a transformer, and can be reduced to the equivalent transformer circuit of Fig. 13. Here the primary and secondary inductances are divided into leakage and closely coupled components. The leakage inductances correspond to magnetic flux lines that produce no linkages to any secondary circuit, while the coupled inductances are taken as having unity coefficient of coupling.

In a conventional transformer with iron core, the coefficient of coupling between primary and secondary is so large that the leakage inductances represent only a very small fraction of the total circuit inductances. Under these conditions it is convenient to analyze the behavior in terms of the turn-ratio, leakage inductance, etc., as is commonly done with 60-cycle power transformers. However, in the more general case where the coefficient of coupling may be small, most of the inductance is leakage, and turn ratio has relatively little significance. It is then preferable to use the general method of analysis given above.

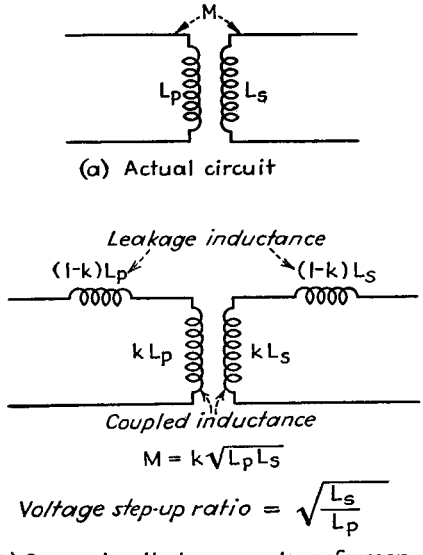


FIG. 13.—Equivalent circuit of a transformer expressed in terms of primary and secondary inductances and coefficient of coupling.

4. Analysis of Some Simple Coupled Circuits. *Untuned Secondary Consisting of a Pure Reactance.*—This case corresponds to a coil serving as the primary, with a metal mass of low resistivity serving as the secondary. In this case, which is illustrated in Fig. 14*a*, one has

$$\text{Coupled reactance} = -\frac{(\omega M)^2}{\omega L_s} = -k^2\omega L_p \tag{29}$$

where k is the coefficient of coupling between the coil and the secondary. The presence of the secondary reduces the inductance as viewed from the coil terminals to an equivalent value $(1 - k^2)L_p$.

It will be noted that the amount of reduction in inductance is determined solely by the coefficient of coupling and not by the number of primary turns, etc. It will be also observed that the effect of the secondary decreases rapidly as the coefficient of coupling becomes small.

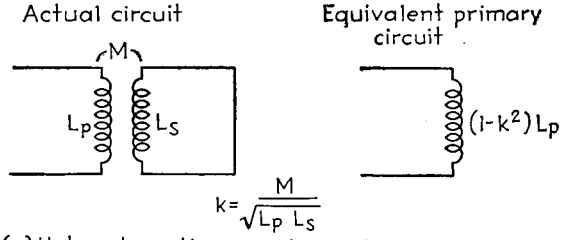
Untuned Secondary Having Both Resistance and Reactance.—This corresponds to the practical case of a coil in a shield, or near a metal panel or other mass, with the resistance of the metal secondary taken into account. For such a secondary, the coupled impedance as given by Eqs. (24) and (25) consists of a resistance and a negative reactance. The effect of the secondary upon the primary is hence to lower the equivalent inductance and increase the equivalent resistance of the primary coil as viewed from its terminals.

Under practical conditions, a secondary consisting of a shield or other metal object is usually of copper or aluminum, so that the secondary reactance is considerably greater than the secondary resistance. Under these conditions, the equivalent primary circuit has the values and form given in Fig. 14*b*.

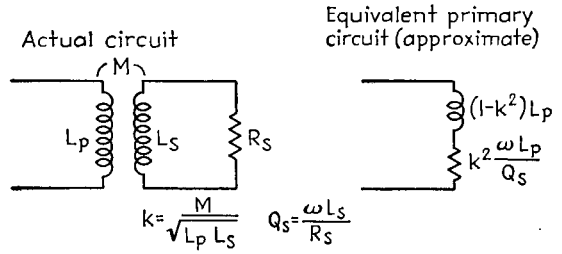
Tuned Secondary, Untuned Primary.—This circuit, illustrated in Fig. 15*a*, is of importance because it represents the equivalent circuit of the transformer-coupled tuned radio-frequency amplifier, with the primary resistance R_p representing the plate

resistance of the tube. When the secondary is a resonant circuit, the curve of coupled impedance $(\omega M)^2/Z_s$ with variation in frequency has substantially the same shape and characteristics as the parallel impedance curve of the secondary, but the absolute magnitude is determined by the mutual inductance.

The important characteristic of this circuit is the curve of voltage developed across the secondary condenser as a function of frequency under conditions where $R_p >> \omega L_p$. Such a characteristic is shown in Fig. 15*b*, and has substantially the same shape as a resonance curve with the peak at the resonant frequency of the secondary and an effective Q somewhat lower than the actual Q of the secondary circuit.



(a) Untuned reactive secondary with zero losses



(b) Untuned secondary with small losses

FIG. 14.—Inductively coupled circuits with untuned secondaries.

On the assumption that $R_p >> \omega L_p$, the response at resonance is

$$\frac{\left\{\begin{matrix}\text{Voltage across secondary} \\ \text{condenser at resonance}\end{matrix}\right\}}{\text{Voltage applied to primary}} = \frac{\omega M}{R_p R_s + (\omega M)^2} \times \omega L_s \tag{30}$$

With given primary and secondary circuits, the response at resonance is maximum when the coupled resistance equals the primary resistance, *i.e.*, when

$$\omega M = \sqrt{R_p R_s} \tag{31}$$

With greater or less coupling, the response is less.

The relationship between the effective Q of the response curve and the actual Q of the secondary circuit is, still assuming that $R_p >> \omega L_p$

$$\frac{\text{Effective } Q \text{ of amplification curve}}{\text{Actual } Q \text{ of tuned circuit}} = \frac{1}{1 + \dfrac{(\omega M)^2/R_s}{R_p}} \tag{32}$$

The effective Q of the response curve approaches the actual Q of the secondary circuit when the coupling is very small, and is exactly one-half of the actual Q when the coupling corresponds to the value giving maximum response.

When ωL_p is not negligible compared with the primary resistance, the response curve still has the shape of a resonance curve, but the peak is shifted to a slightly higher frequency. This shift in location of the peak is due to the fact, expressed by Eq. (26), that maximum secondary current occurs when secondary is sufficiently

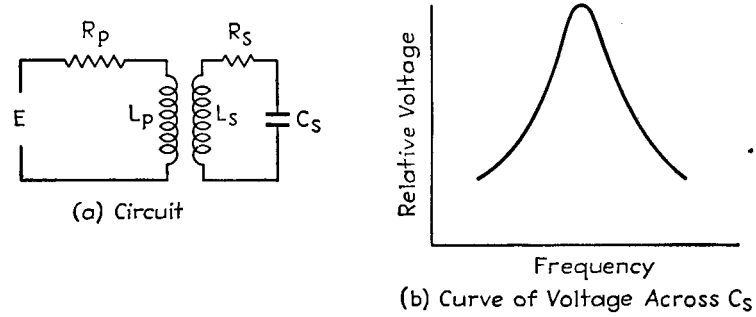


FIG. 15.—Inductively coupled circuit with tuned secondary and untuned primary.

detuned that its coupled reactance neutralizes the reactance of the primary. When the primary reactance is not negligible, this detuning is appreciable. The coupling required for maximum secondary current is also greater as the ratio $\omega L_p/R_p$ is increased.

Analysis of the output of coupled systems with untuned primary and tuned secondary can be most conveniently carried out by using the equivalent secondary circuit of Fig. 12*d*. Such equivalent circuits are shown in Fig. 16. These show how the asso-

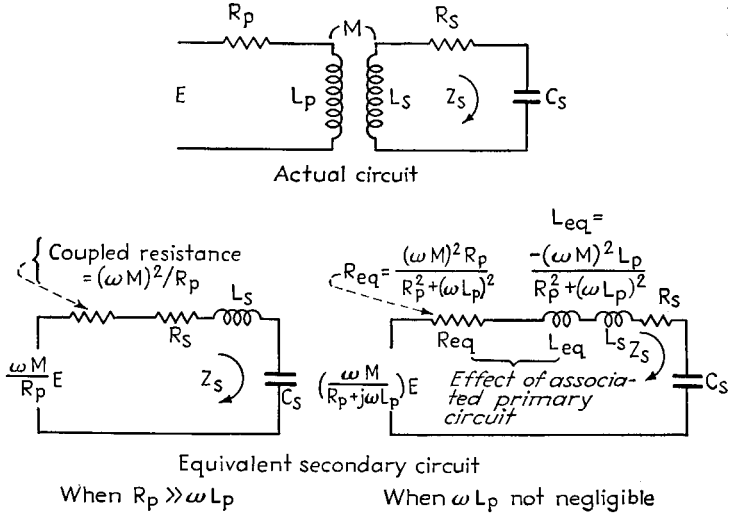


FIG. 16.—Equivalent secondary circuit for case of inductive coupling with tuned secondary and untuned primary.

ciation of the primary with the resonant secondary has the effect of increasing the effective secondary resistance and hence of lowering the effective Q, and also that an inductive primary neutralizes some of the secondary inductance and so raises the resonant frequency.

5. Two Resonant Circuits Tuned to the Same Frequency and Coupled Together.[1] When two circuits resonant at the same frequency are coupled together, the resulting behavior depends very largely upon the coupling, as shown in Fig. 17. When the coefficient of coupling is small, the curve of primary current as a function of frequency approximates closely the series resonance curve of the primary circuit considered alone. The secondary current at the same time is small and varies with frequency according to a curve having a shape approximating the product of the resonance curves of

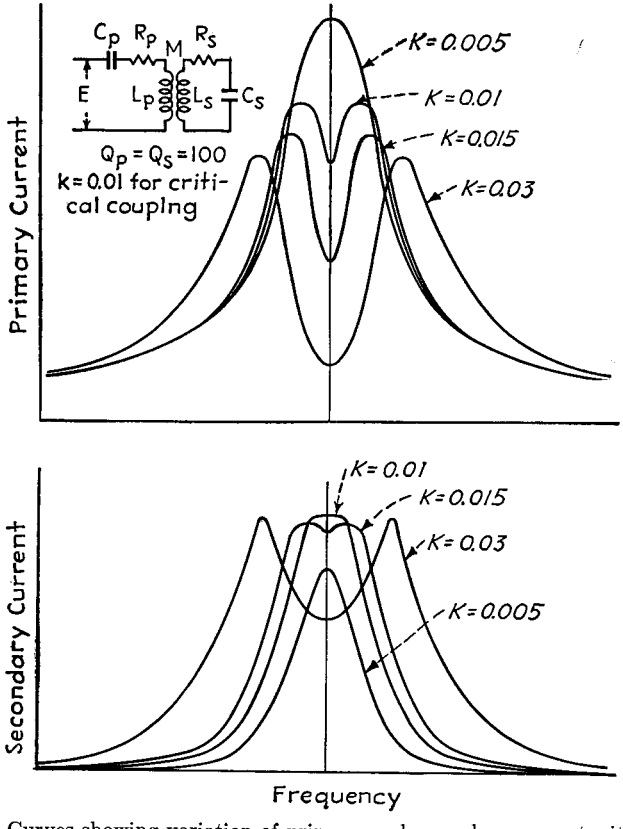


FIG. 17.—Curves showing variation of primary and secondary current with frequency for different coefficients of coupling when the primary and secondary are separately tuned to the same frequency.

primary and secondary circuits taken alone. As the coupling is increased, the curve of primary current becomes broader and the peak value of primary current is reduced. At the same time, the secondary current becomes larger and the sharpness of the secondary current curve is reduced. These trends continue until the coupling is such that the resistance that the secondary circuit couples into the primary at resonance is equal to the primary resistance. This coupling is called the *critical coupling*, and causes the secondary current to have the maximum possible value that it can attain.

[1] A very complete discussion of this subject is given by C. B. Aiken, Two-mesh Tuned Coupled Circuit Filters, *Proc. I.R.E.*, Vol. 25, p. 230, February; p. 672, June, 1937. See also E. S. Purington, Single and Coupled-circuit Systems, *Proc. I.R.E.*, Vol. 18, p. 983, June, 1930.

The curve of secondary current is then somewhat broader in the immediate vicinity of the resonant frequency than the resonance curve of the secondary circuit, and the primary curve shows double peaks. With still greater coupling, the double humps in the primary current curve become more prominent, and the peaks spread farther apart. At the same time the secondary current curve begins to show double humps, with the peaks becoming progressively more pronounced and farther apart as the coefficient of coupling is increased.[1]

Basic Equations.—In practical work, the most important property of the coupled system under consideration is the ratio of voltage developed across the condenser in the secondary circuit to the voltage applied in series with the primary circuit. A useful expression for this important ratio is obtained by rearranging Eqs. (22) and (23) to give the relation[2]

$$\frac{\text{Voltage across secondary condenser}}{\text{Voltage applied in series with primary}} = \frac{E_c}{E} = \frac{-1}{\gamma^2}\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 + \frac{1}{Q_pQ_s} - \left(1-\frac{1}{\gamma^2}\right)^2 + j\left(1-\frac{1}{\gamma^2}\right)\left(\frac{1}{Q_p}+\frac{1}{Q_s}\right)\right]} \tag{33}$$

where $\omega = 2\pi$ times frequency.

$\omega_0 = 2\pi$ times resonant frequency.

$\gamma = \dfrac{\omega}{\omega_0} = \dfrac{\text{actual frequency}}{\text{resonant frequency}}$.

E = voltage applied in series with primary.

$Q_p = \omega L_p/R_p$ for primary circuit.

$Q_s = \omega L_s/R_s$ for secondary circuit.

$k = M/\sqrt{L_pL_s}$ = coefficient of coupling.

L_p = total inductance of primary circuit.

L_s = total inductance of secondary circuit.

This equation involves no approximations, other than assuming that the circuits are resonant at the same frequency.

In the special case where the circuit losses are zero, or where the frequency differs sufficiently from the resonance frequency to allow the losses to be neglected, Eq. (33) takes the form

$$\frac{E_c}{E}\text{ for zero losses} = \frac{-1}{\gamma^2}\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 - \left(1-\frac{1}{\gamma^2}\right)^2\right]} \tag{34}$$

Effect of Coupling on Response at Resonance—Critical Coupling.—The response at resonance is obtained by substituting $\gamma = 1$ in Eq. (33), which gives

$$\frac{E_c}{E}\text{ at resonance} = -\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 + \frac{1}{Q_pQ_s}\right]} \tag{35}$$

This expression is maximum when the coefficient of coupling has the value

$$k_c = \frac{1}{\sqrt{Q_pQ_s}} \tag{36}$$

[1] When the circuit Q's are not equal, double humps do not appear in the curve of secondary current until the coupling is somewhat greater than the critical value. This is illustrated in the universal curves of Fig. 22.

[2] For derivation, see Terman, *op. cit.*, p. 81.

This coupling k_c is often termed the *critical coupling.* The corresponding mutual inductance is

$$\omega M = \sqrt{R_p R_s} \tag{37}$$

With the mutual inductance for critical coupling, the resistance coupled into the primary circuit at resonance by the presence of the secondary equals the primary resistance. This gives the maximum possible secondary current. The resulting maximum possible voltage across the secondary condenser is

$$\text{Maximum possible } \frac{E_c}{E} = \frac{1}{2k_c}\sqrt{\frac{L_s}{L_p}} = \frac{\sqrt{Q_p Q_s}}{2}\sqrt{\frac{L_s}{L_p}} \tag{38}$$

where k_c is the critical coupling as given by Eq. (36). The voltage developed across the condenser at resonance with other couplings is less than value given by Eq. (38) by an amount depending upon the ratio of actual to critical coupling, as shown in Fig. 18.

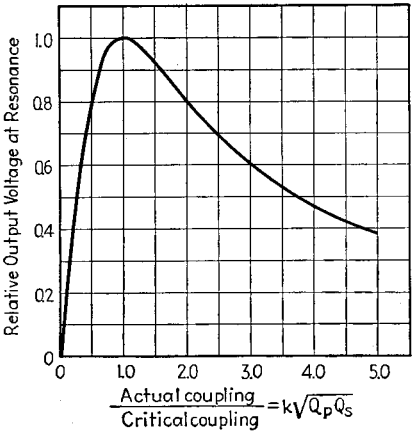


FIG. 18.—Relative output voltage obtained at resonance as a function of coefficient of coupling in the case of two circuits tuned to the same frequency and coupled together.

Conditions for Existence of Double Peak —Height, Location, and Width of Peaks.— The curve of response in the secondary circuit exhibits double peaks whenever the coefficient of coupling is sufficient to cause the right-hand side of Eq. (33) to show two maxima. This occurs when[1]

$$\frac{\text{Actual coefficient of coupling}}{\text{Critical coefficient of coupling}} \geqslant \left[\frac{1}{2}\left(\frac{Q_p}{Q_s} + \frac{Q_s}{Q_p}\right)\right]^{1/2} \tag{39}$$

It will be noted that when $Q_p = Q_s$, double peaks occur whenever the coupling exceeds the critical value, but if the Q's are different, double peaks do not occur until the coupling exceeds the critical value by an amount determined by the Q ratio.

When double humps occur, the location of the peaks in relation to the resonant frequency depends upon the actual coupling, the critical coupling, and the Q ratio. Analysis based upon Eq. (33) shows that to an accuracy sufficient for all ordinary purposes the location of the peaks is given by[2]

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = \gamma_p = \frac{1}{\sqrt{1 \pm k\left[1 - \frac{k_c^2}{2k^2}\left(\frac{Q_p}{Q_s} + \frac{Q_s}{Q_p}\right)\right]^{1/2}}} \tag{40}$$

where k_c is the critical coefficient of coupling as given by Eq. (36) and the remaining notation is the same as above. The ratio represented by the spacing between coupling peaks to the resonant frequency, as a function of the coefficient of coupling,

[1] Equation (39) assumes that the γ^2 multiplying the denominator of Eq. (33) is constant. The approximation that results is so small as to be of negligible consequence, but results in enormous simplification of the equations.

[2] Eq. (40) involves the same approximation as does Eq. (39), but again the error introduced is negligible and the simplification considerable.

is given in Fig. 19 for various conditions, in terms of a parameter g given by the lower part of Fig. 19.

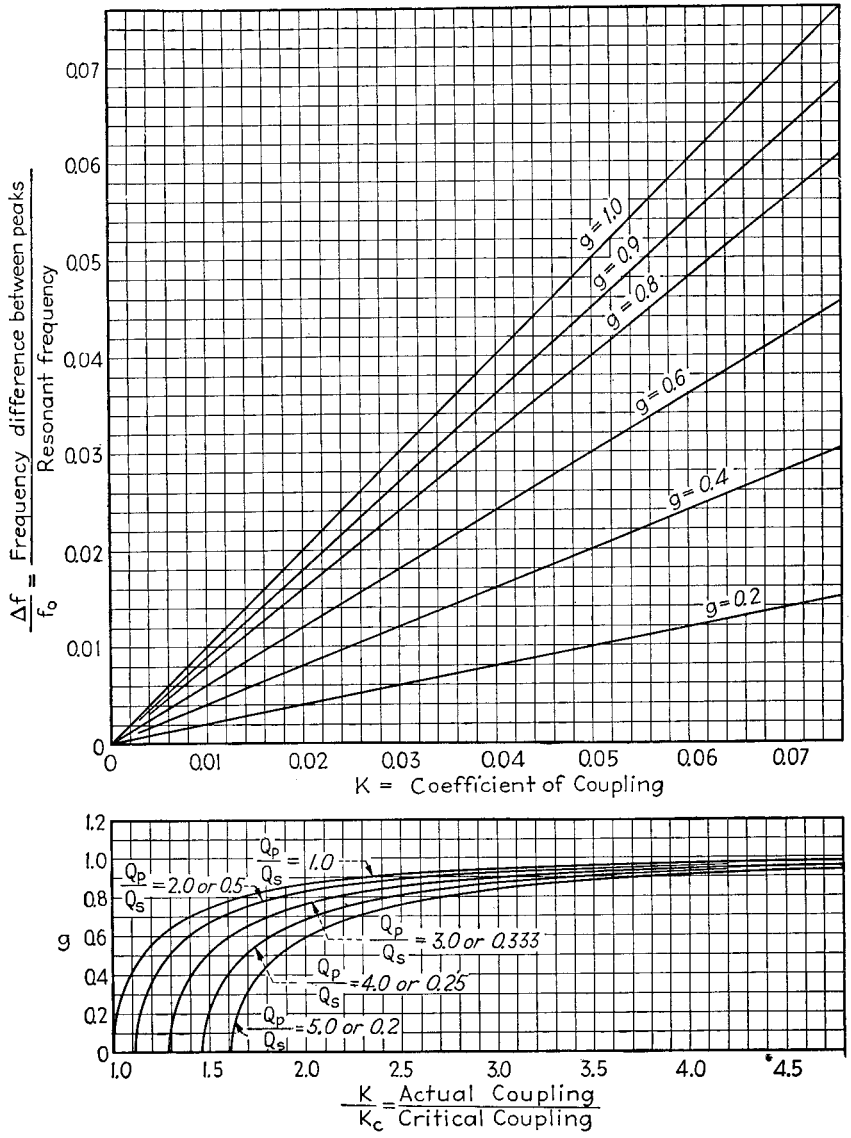


FIG. 19.—Curves from which the separation between peaks can be determined for the case of two coupled circuits resonant at the same frequency.

When the circuit Q's are not too unequal, and the coefficient of coupling greatly exceeds the critical value, Eq. (40) reduces to

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = \frac{1}{\sqrt{1 \pm k}} \qquad (41a)$$

Furthermore, in the very common case when the coefficient of coupling is small, Eq. (41*a*) becomes

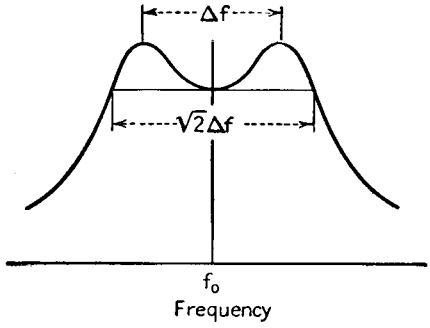


FIG. 20.—Relationship between band width, and width between secondary peaks existing when two circuits resonant at the same frequency are coupled together.

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = 1 \pm \frac{k}{2} \quad (41b)$$

In practical work, one is commonly interested in the frequency band over which the response in the secondary circuit equals or exceeds the response at resonance. This band is illustrated schematically in Fig. 20, and can be shown to be equal to $\sqrt{2}$ times the width of the frequency band between coupling peaks.[1]

The height of the peaks of response in the secondary circuit can be obtained by substituting the value of γ_p from Eq. (40) into Eq. (33). This gives

$$\text{Value of } \frac{E_c}{E} \text{ at peaks} = \frac{1}{\gamma_p^2}\sqrt{\frac{L_s}{L_p}}\,\frac{\sqrt{Q_pQ_s}}{\left[\sqrt{\frac{Q_p}{Q_s}}+\sqrt{\frac{Q_s}{Q_p}}\right]\left[1-\frac{k_c^2}{4k^2}\left(\sqrt{\frac{Q_s}{Q_p}}-\sqrt{\frac{Q_p}{Q_s}}\right)^2\right]^{1/2}} \quad (42)$$

It will be noted that the lower frequency peak is the highest peak. However, since

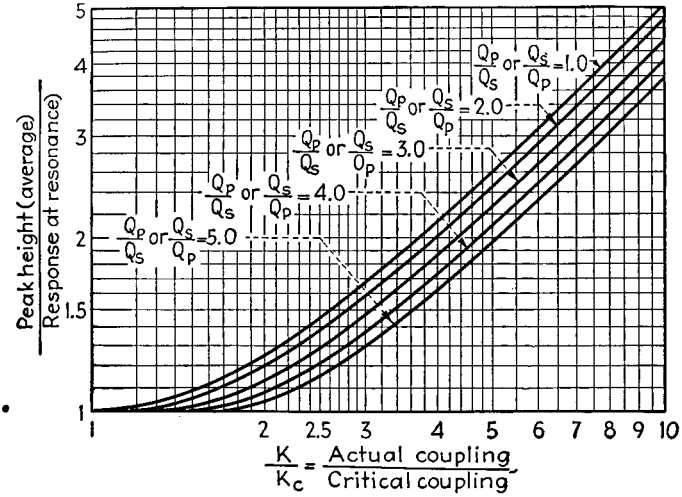


FIG. 21.—Universal curves giving the ratio of actual peak height to the response at resonance as a function of the coefficient of coupling for two coupled circuits resonant at the same frequency.

γ_p differs only slightly from unity in ordinary cases, it is customary to consider that the peaks are identical.

[1] See Aiken, *loc. cit.*

When $Q_p = Q_s$, the value of E_c/E approximates the maximum possible response that can be obtained, as given by Eq. (38). On the other hand, when $Q_p \neq Q_s$, the peaks always have a height less than the maximum possible response in the secondary.

The ratio of the mean of the responses at the two peaks to the response at resonance is plotted in Fig. 21 for a number of Q ratios.

Universal Response Curves.—Equations (33) and (38) can be combined to give

$$\frac{\left\{\begin{array}{l}\text{Actual voltage across}\\ \text{secondary condenser}\end{array}\right\}}{\left\{\begin{array}{l}\text{Maximum possible voltage across secondary condenser with optimum coupling}\end{array}\right\}} = \frac{2n}{\left(1+\frac{pk_c}{2}\right)^2\left[(n^2+1)-(mp)^2+j(mp)\left(\sqrt{\frac{Q_p}{Q_s}}+\sqrt{\frac{Q_s}{Q_p}}\right)\right]} \tag{42a}$$

where $n = \dfrac{k}{k_c} = \dfrac{\text{actual coefficient of coupling}}{\text{critical coefficient of coupling}}$.

$$p = \frac{\text{cycles off resonance}}{\left(\frac{k_c}{2}\right) \times \text{resonant frequency}}.$$

$$m = \frac{1+\frac{pk_c}{4}}{\left(1+\frac{pk_c}{2}\right)^2}.$$

Universal curves based on Eq. (42*a*) are given in Figs. 22*a* and 22*b* for $k_c = 0$, and $k_c = 0.05$. These curves, especially when considered in conjunction with Figs. 18, 19, and 21, provide a complete and accurate picture of the response that can be expected when two resonance circuits are tuned to the same frequency and coupled together.

Figure 22 shows that when both primary and secondary circuits have the same Q, the response curve separates into double peaks whenever the coefficient of coupling exceeds the critical value, and these peaks maintain a substantially constant height irrespective of their location. On the other hand, when the two circuits have different Q's, then if the coupling exceeds the critical value by only a small amount, the response curve still has a single peak, but with reduced response at resonance. Double humps do not occur until the coupling exceeds the critical value by an amount that depends upon the ratio Q_p/Q_s, and the peaks become slightly lower as the coupling increases.

Practical Calculation of Coupled Systems Involving Two Resonant Circuits Tuned to the Same Frequency.—When it is desired to obtain the response curve of a given system, the relative shape of the curve of voltage developed across the secondary condenser can be determined with an accuracy sufficient for all ordinary purposes by the use of the universal curves of Fig. 22, interpolating between these in accordance with the Q ratio involved. The absolute magnitude of the curve can then be obtained by calculating the response at the resonant frequency with the aid of Eq. (35).

An alternative procedure for the case where the curve shows two peaks is to calculate the response at resonance by Eq. (35), determine the location and height of the peaks by Figs. 19 and 21, or Eqs. (40) and (42), and then to note that the response beyond the peaks falls to the response at resonance when the frequency is $\sqrt{2}$ times as far from resonance as the coupling peaks. This information is sufficient to define the shape of the double peaked curve in the vicinity of resonance. If quantitative information is desired at frequencies differing from resonance by at least twice as

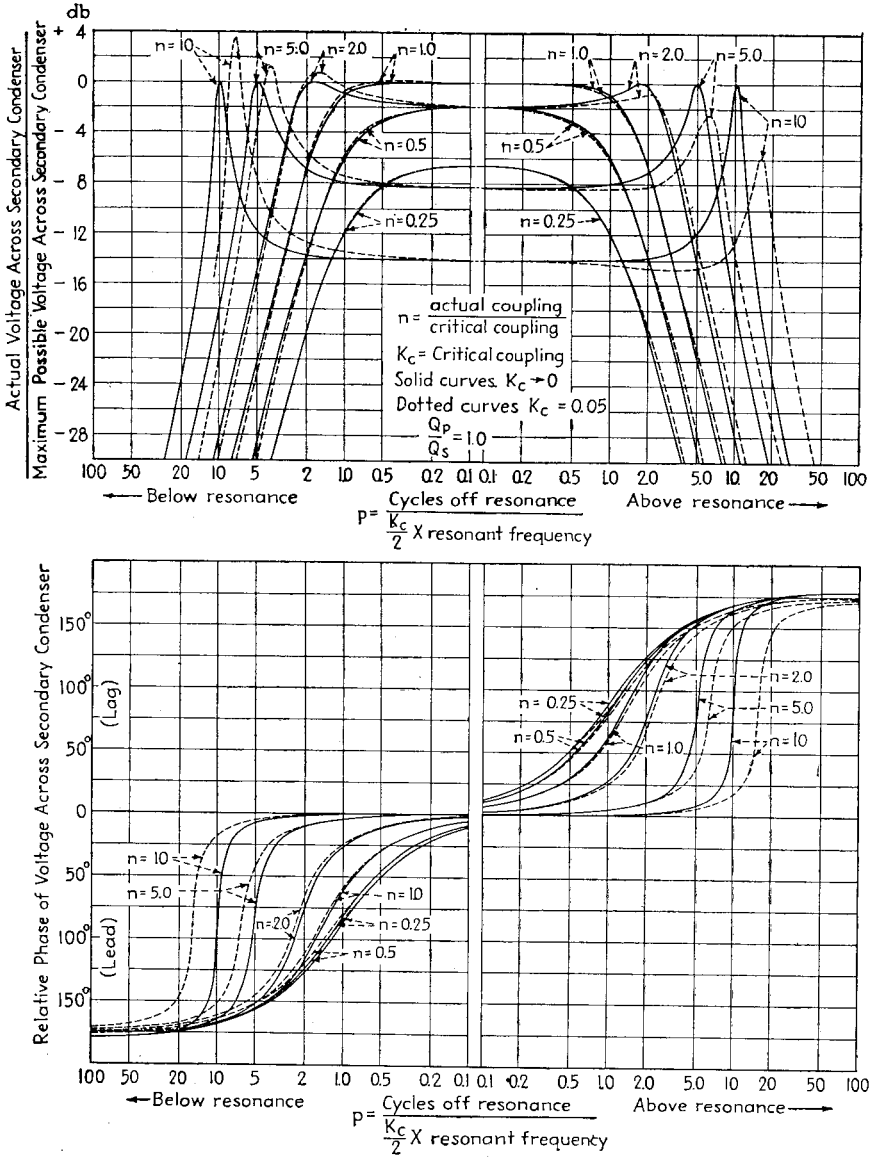


FIG. 22*a*.—Universal curves giving the phase and relative magnitude of the voltage across the secondary condenser for the case of two coupled circuits resonant at the same frequency and having a Q ratio of unity.

many cycles as do the peaks, this can be readily obtained by neglecting the circuit resistance and using Eq. (34).

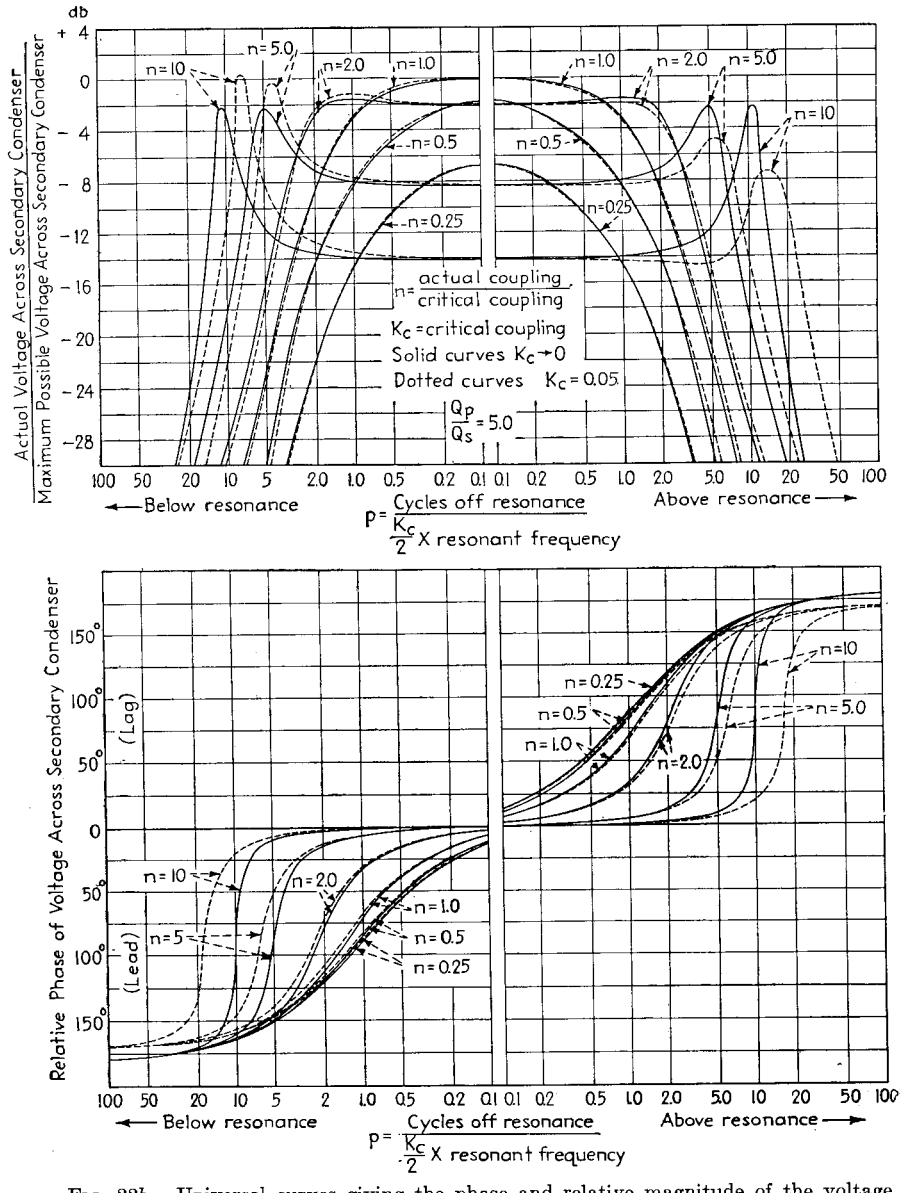


FIG. 22*b*.—Universal curves giving the phase and relative magnitude of the voltage across the secondary condenser for the case of two coupled circuits resonant at the same frequency and having a Q ratio of 5.

Parallel Feed in the Primary Circuit.—The analysis and discussion given above have all been for the case where the voltage was applied in series with the primary

circuit, as shown in Fig. 17. In many cases where coupled resonant circuits are employed, the excitation of the primary circuit is obtained by applying the voltage in parallel with the primary circuit through a resistance, as shown in Fig. 23*a*. This arrangement can be termed *parallel excitation*, and corresponds to the case when the coupled circuits are excited from a vacuum tube.

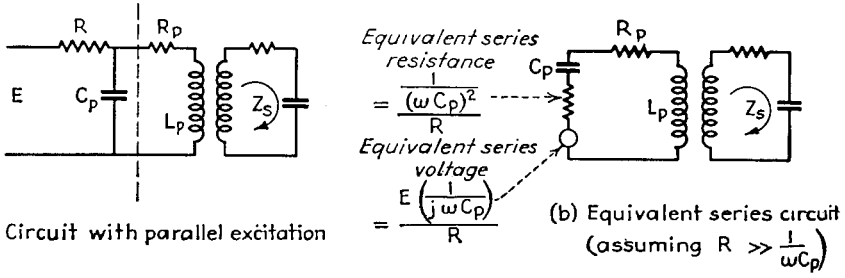
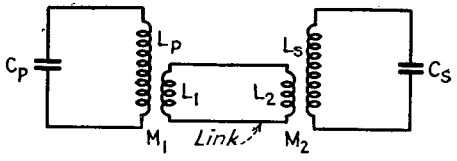


FIG. 23.—Actual and equivalent circuits for the case of parallel excitation of two coupled circuits.

The parallel excited circuit of Fig. 23*a* can be reduced by Thévenin's theorem to the equivalent arrangement shown in Fig. 23*b*. This is accomplished by considering that the portion of the circuit to the left of the dotted line is the equivalent generator circuit supplying the remainder of the circuit. The principal effect of parallel excitation is to introduce an added resistance in series with the primary circuit that becomes greater the lower the series resistance R used in the parallel feed. The equivalent



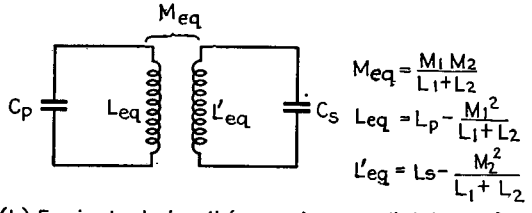


FIG. 24.—Actual and equivalent circuits for link coupling.

series voltage also varies somewhat with frequency, but in a limited range, such as the region in which the resonance phenomena occur, can be considered as being substantially constant.

Link-coupled Circuits.—Two resonant circuits are often coupled together by a link circuit, as shown in Fig. 24*a*. When this arrangement is reduced to an inductively coupled circuit by the method of Par. 8, the result for negligible losses in the link is

as shown in Fig. 24*b*. It will be noted that the use of link coupling has the effect of reducing the equivalent primary and secondary inductance, and hence of raising the resonant frequencies. In other respects, the use of link coupling gives the same behavior as obtainable from ordinary inductive coupling.

6. Coupled Circuits with Resonant Primary and Secondary Tuned to Slightly Different Frequencies.—When two circuits resonant at slightly different frequencies are coupled together, the behavior depends upon the coefficient of coupling and the relative and absolute circuit Q's. When $Q_p = Q_s$, the response curve of secondary current (or voltage) has a shape of the same type as would be obtained if the two circuits were both resonant at the same frequency and coupled with an increased coefficient of coupling. In other words, the effect of detuning the circuits for the case of equal Q's is equivalent, as far as shape is concerned, to increasing the effective coupling. The only difference is that detuning causes the absolute magnitude of the response curve to be less than when the same shape is obtained without detuning.

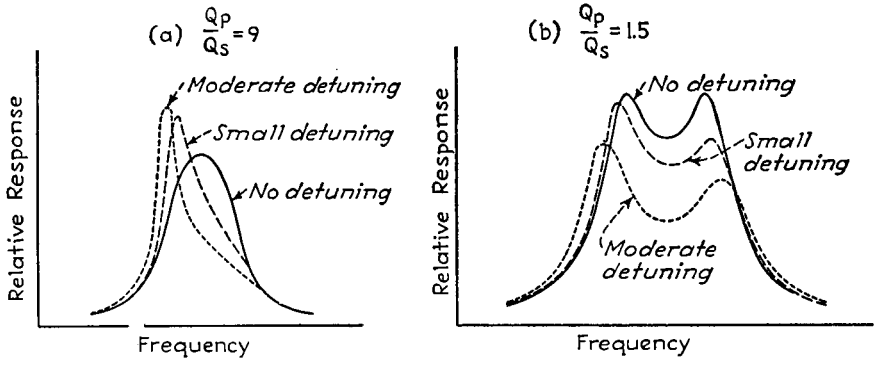


FIG. 25.—Curves illustrating the effect produced on the shape of the response curve by detuning primary and secondary circuits when the primary and secondary circuits do not have identical Q's.

Analysis shows that when $Q_p = Q_s$ the shape with detuning is as though the circuits were both tuned to the same frequency and coupled by an amount[1]

$$\left.\begin{matrix}\text{Equivalent coupling}\\ \text{corresponding to}\\ \text{detuned operation}\end{matrix}\right\} = \sqrt{k^2 + \left(\frac{\Delta}{f_0}\right)^2} \qquad (43)$$

where k = actual coefficient of coupling.

Δ = difference between resonant frequency of primary and secondary circuits.

f_0 = frequency midway between primary and secondary resonant frequencies.

In the more general case where the circuit Q's are not the same, the secondary response curve is no longer symmetrical about the mean frequency, as shown in Fig. 25. In cases where unsymmetrical peaks are obtained, the low-frequency peak will be depressed when the secondary is tuned to a higher frequency than the primary and the secondary Q is higher, or when the secondary is tuned to a lower frequency and the secondary Q is the lower. Otherwise the high frequency peak will be depressed.

Detuning produces much more effect upon the shape of the primary current curve than upon the secondary response, and if the detuning is at all large the primary

[1] See Harold A. Wheeler and J. Kelly Johnson, High Fidelity Receivers with Expanding Selectors, *Proc. I.R.E.*, Vol. 23, p. 594, June, 1935; or C. B. Aiken, *loc. cit.*

Equation (43) involves the assumption that Δ/f_0 is small compared with unity.

current curve will have only a single peak, even though the secondary response shows pronounced double peaks.

7. Capacitive and Direct Inductive Coupling.—Examples of capacitive and direct inductive coupling are given in Fig. 26. The behavior of these circuits follows the same general character as that discussed for inductive coupling. Thus the secondary circuit can be considered as producing an equivalent coupled impedance in the primary circuit, and the primary circuit can be considered as inducing in the secondary a voltage that gives rise to the secondary current. The analysis for these cases can be worked out in a manner exactly analogous to the inductively coupled case. In fact, with direct inductive coupling, the same equations apply without change if the coupled inductance L_m is substituted for the mutual inductance M appearing in the inductively coupled equations.

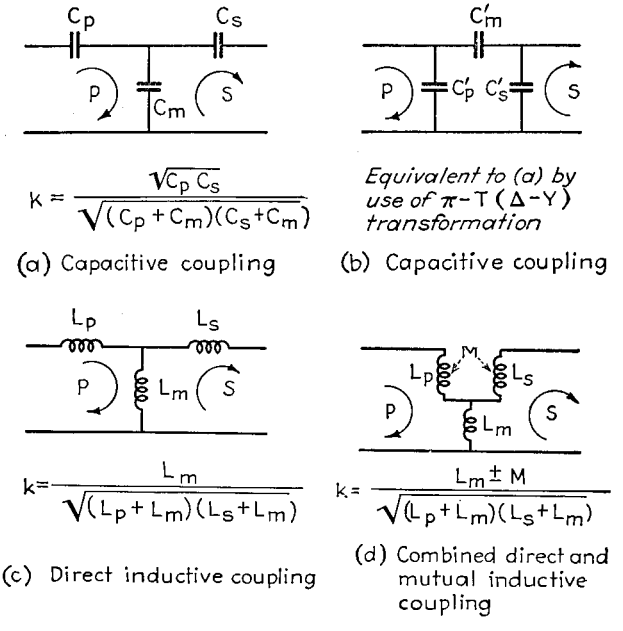


FIG. 26.—Circuits with simple capacitive and inductive coupling.

In the case of capacitive coupling, the equations already given for inductive coupling still apply, provided that one substitutes $1/j\omega C_m$ for $j\omega M$ in the equations for inductive coupling. This means that the mutual impedance varies inversely instead of directly with frequency, being now $1/\omega C_m$ instead of ωM. However, in a limited frequency range, such as that represented by the region about resonance, the difference is inconsequential. As a result, capacitively coupled circuits with a given coefficient of coupling act substantially the same as the corresponding inductively coupled circuit with the same coefficient of coupling. Hence when primary and secondary are both tuned to the same frequency, the secondary current characteristic has two humps if the coupling is large, *i.e.*, if the condenser C_m is small, whereas there is only one peak of secondary current when the coupling is small, *i.e.*, when condenser C_m is large.

8. Systems in Which Coupling Varies with Frequency.—Systems of this type are often used in arrangements involving circuits tunable over a wide-frequency range.

Examples are the tuned radio-frequency circuits and antenna-coupling circuits of radio receivers, which are normally adjustable to be resonant over a three to one frequency range. In such arrangements, it is desirable to maintain substantially constant response over the entire tuning range.

Two methods are commonly used for attacking this problem. The first combines either tuning the circuits by variable condensers and using capacitive coupling or tuning by variable inductances and using inductive or direct inductive coupling. In the first of these arrangements, the coefficient of coupling is proportional to $1/f^2$; in the latter, it is proportional to f^2.

The second method is to employ combined electrostatic and magnetic coupling.[1] A simple example of such coupling is illustrated in Fig. 27. Here the coupling is capacitive at low frequencies and inductive at high frequencies, because the coupling combination of C_m in series with L_m has capacitive and inductive reactance under these respective conditions. In between, at the resonant frequency of L_m and C_m, there is zero coupling.

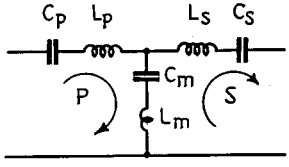


FIG. 27.—Circuit with combined capacitive and combined inductive coupling.

An almost unlimited number of arrangements for combining electrostatic and magnetic coupling are possible. A number of representative cases, and the ones most likely to be encountered in radio work, are shown in Figs. 28 to 34.[2] All these arrangements show a tuned secondary circuit LC with primary input terminals indicated by P. The part between the input terminals and the secondary can be thought of as representing the coupling system.

Analysis of Circuits with Combined Coupling.—The simplest method of analyzing these various forms of coupled circuits is to take advantage of the fact that all of them can be reduced to the simple coupled circuit of Fig. 35, provided that suitable values

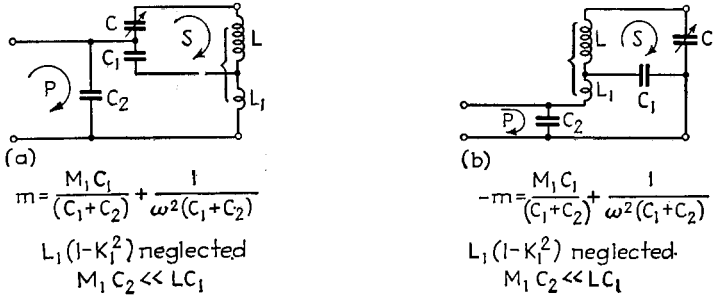


FIG. 28.—Circuits with complex coupling having coupling condensers in series with the tuning condensers.

are assigned to Z_p, Z_s, and M. The rules that determine the values of these quantities in the simple equivalent circuit are as follows:

1. The equivalent primary impedance Z_p of the equivalent circuit is the impedance that is measured across the primary terminals of the actual circuit when the secondary circuit has been opened.
2. The secondary impedance Z_s of the equivalent circuit is the impedance that is measured by opening the secondary of the actual circuit and determining the impedance between these open points when the primary is open-circuited.

[1] An exhaustive treatment of this is given by H. A. Wheeler and W. A. MacDonald, Theory and Operation of Tuned Radio-frequency Coupling Systems, *Proc. I.R.E.*, Vol. 19, p. 738, May 1931. See also Edward H Loftin and S. Young White, Combined Electromagnetic and Electrostatic Coupling and Some Uses of the Combination, *Proc. I.R.E.*, Vol. 14, p. 605, October, 1926.

[2] These are from Wheeler and MacDonald, *loc. cit.*

3. The equivalent mutual inductance M is determined by assuming a current I_0 flowing into the primary circuit. The voltage that then appears across an open circuit in the secondary is equal to $-j\omega MI_0$.

In making use of the equivalent circuit of Fig. 35, it is to be remembered that the values of Z_p, Z_s, and M may all vary with frequency, so that it is generally necessary to determine a new equivalent circuit for each frequency at which calculations are to be made.

After the actual coupled circuit has been reduced by the preceding procedure to its equivalent form, shown in Fig. 35, one can then apply the formulas that have already been given for inductively coupled circuits, using the appropriate values M, Z_s, Z_p, as determined for the equivalent circuit. This procedure has the advantage of using the same fundamental formulas to handle all types of coupling, and makes it possible to carry on the analysis in the same manner for all cases.

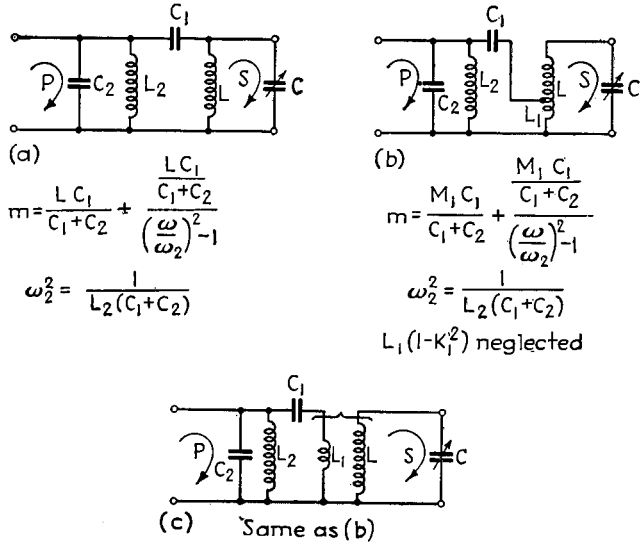


FIG. 29.—Circuits with complex coupling having high-inductance primaries not inductively coupled to the secondaries.

The quantity M that appears in the equivalent circuit represents the effective coupling that is present between the primary and secondary circuits. It is not necessarily a real mutual inductance of the inductive type, but rather a sort of mathematical fiction that gives the equivalent effect of whatever coupling is really present. If the actual coupling is capacitive, the numerical value of M will be found to be negative; if the coupling is of a complex type representing both resistive and reactive coupling, the numerical value of M will be found to have both real and imaginary parts. The proper procedure, in any case, is to take the value of M as it comes and substitute it with its appropriate sign and phase angle whenever M appears in the expressions for inductively coupled circuits.

Application of Analysis to Circuits of Figs. 28 to 34.—Equations for the equivalent mutual inductance of circuits of Figs. 28 to 34 are shown below each circuit diagram. In these equations and circuits, a uniform nomenclature has been used as follows:

L = tuned secondary coil.
C = secondary tuning condenser.
L_1 = primary coil of low inductance closely coupled to secondary coil.

L_2 = primary coil of high inductance not closely coupled to secondary.
M_1 = mutual inductance between L and L_1.
M_2 = mutual inductance between L and L_2.
M_{12} = mutual inductance between L_1 and L_2.
k_1, k_2, and k_{12} = coupling coefficients corresponding to M_1, M_2, and M_{12}, respectively.
C_1 = coupling condenser.
C_2 = capacity across input terminals (such as tube capacity).
ω_2 = resonant frequency of primary circuit as viewed from input terminals P.

Relative polarity of coils is indicated by the way in which coils are drawn: thus in Fig. 30*c*, L and L_2 have opposite polarity from L and L_1 in Fig. 30*d*.

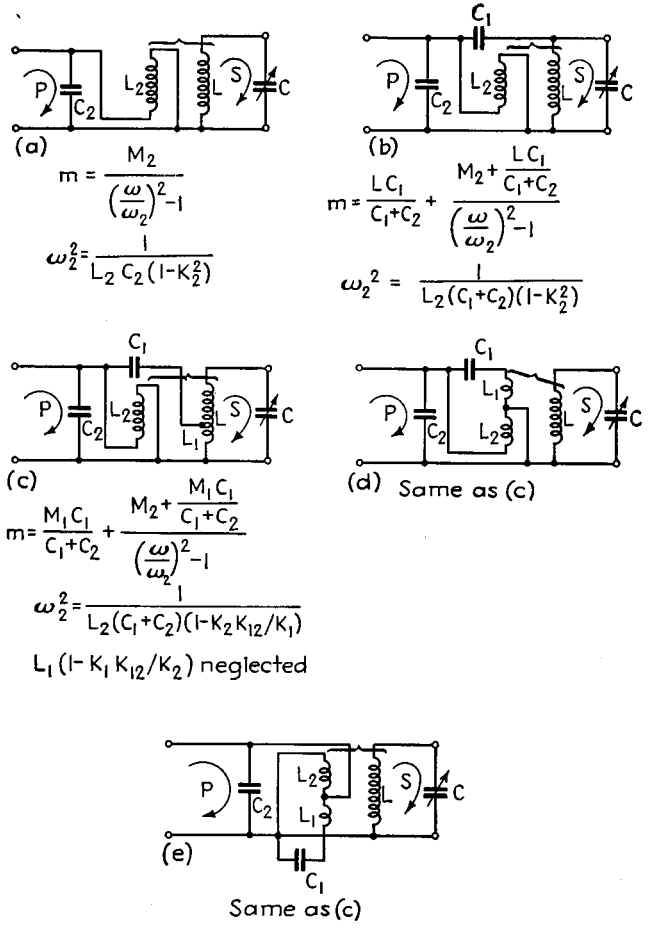
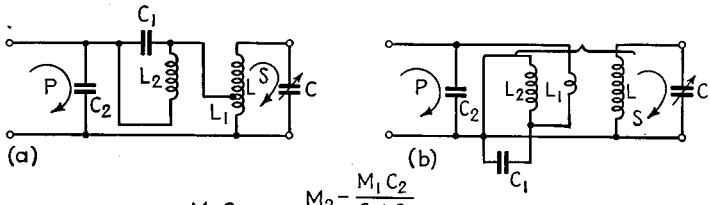


FIG. 30.—Circuits with complex coupling having high-inductance primaries inductively coupled to secondaries.

In the circuits of Figs. 28 to 34, C_1 can in some cases be supplied by the stray capacity between coils and in other cases, by the distributed capacity of the high inductance primary coil L_2. Nearly all the circuits of Figs. 28 to 33 are character-



$$m = \frac{M_1 C_1}{C_1+C_2} + \frac{M_2 - \frac{M_1 C_2}{C_1+C_2}}{\left(\frac{\omega}{\omega_2}\right)^2 - 1}$$

$$\omega_2^2 = \frac{1}{L_2(C_1+C_2)(1-K_2K_{12}/K_1)}$$

$L_1(1-K_1K_{12}/K_2)$ neglected

Fig. 31.—Circuits with complex coupling having high-inductance primaries inductively coupled to secondaries.

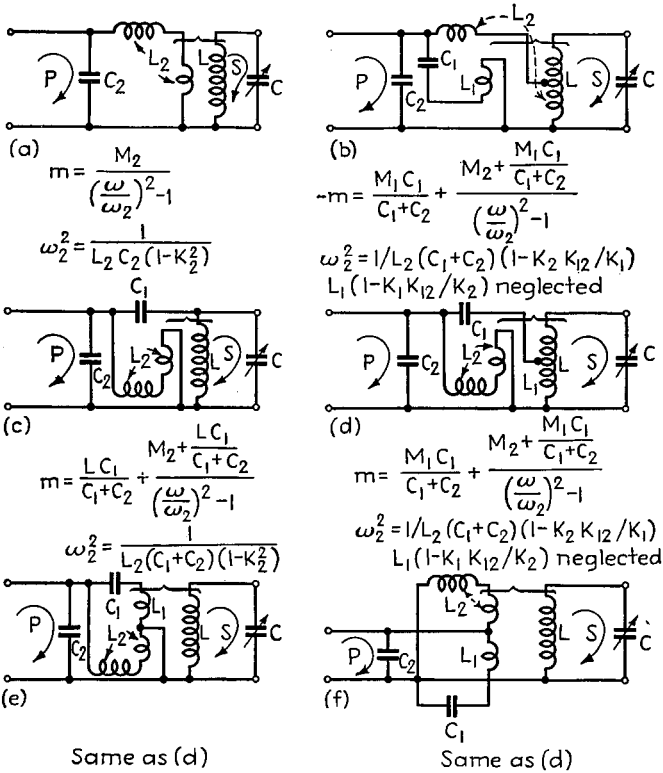


Fig. 32.—Circuits with complex coupling having high-inductance primaries coupled to secondaries by coupling coils.

ized by having an equivalent mutual inductance that is the sum of two components, one of which is independent of frequency, the other of which decreases rapidly with frequency. In all the arrangements having a high inductance primary L_2, the rate of

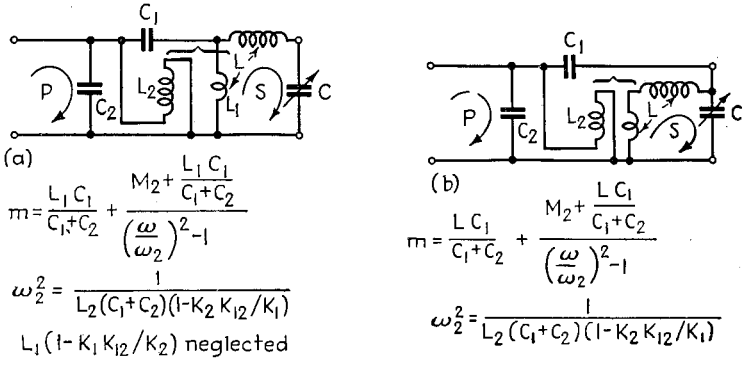


FIG. 33.—Circuits with complex coupling having high-inductance primaries coupled to coupling coils in secondary circuits.

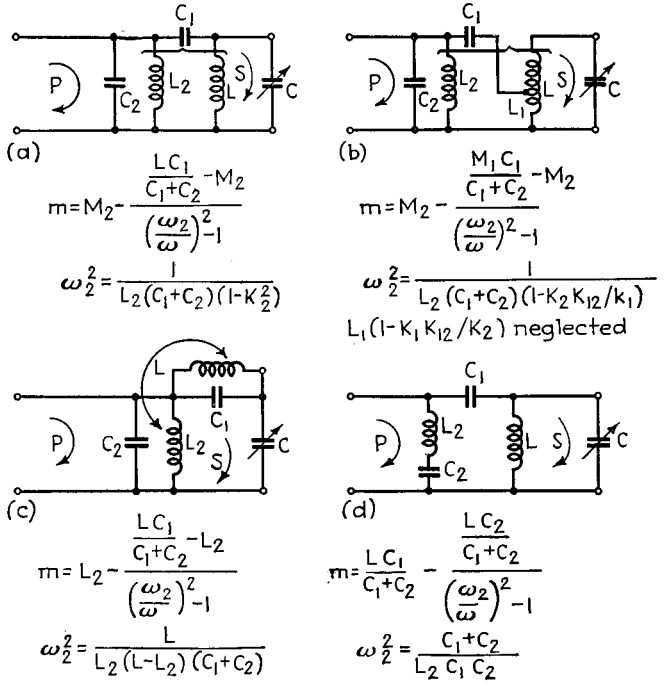


FIG. 34.—Circuits with complex coupling in which the equivalent mutual inductance varies according to a different law from that applying to Figs. 28–33.

rise of the variable component at low frequencies is determined by how close the resonant frequency $\omega_2/2\pi$ of the primary circuit is to the frequency under consideration. The circuits of Fig. 34 differ from those in Figs. 28 to 33 in that the equivalent

mutual inductance equals a fixed value minus a component that increases with frequency.

Typical curves showing ways in which the equivalent mutual inductance can be made to vary with frequency with different coupling systems are shown in Fig. 36.

9. Band-pass Filters Based on Coupled Resonant Circuits.—A response curve of secondary voltage having a relatively flat top with steep sides can be obtained by properly coupling two circuits resonant at the same frequency, as is apparent by examining Fig. 22. The resulting arrangement is commonly termed a band-pass filter, and has characteristics particularly desirable for handling modulated waves, because the response can be made practically the same to side-band frequencies as to the carrier. In contrast with this, ordinary resonance curves have a rounded top and so discriminate against the higher side-band frequencies in favor of the carrier and lower side-band frequencies.

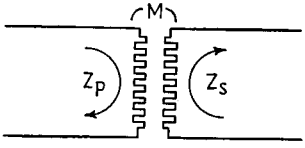


FIG. 35.—Equivalent form for any coupled circuit.

Design of Band-pass Circuits Where Primary and Secondary Are Tuned to the Same Frequency.—In order to obtain band-pass characteristics in this way, it is necessary to adjust the coefficient of coupling and the circuit Q's properly in order to obtain the desired width and flatness of the response characteristic. Whereas the width and flatness of the top of the response curve are both affected by the coefficient of coupling and the circuit Q's, the width of the top is determined primarily by the coefficient of coupling, and the flatness of the top depends mainly on the circuit Q's. Large couplings correspond to wide tops, high Q's give pronounced double peaks, and low Q's cause the top to be rounded off.

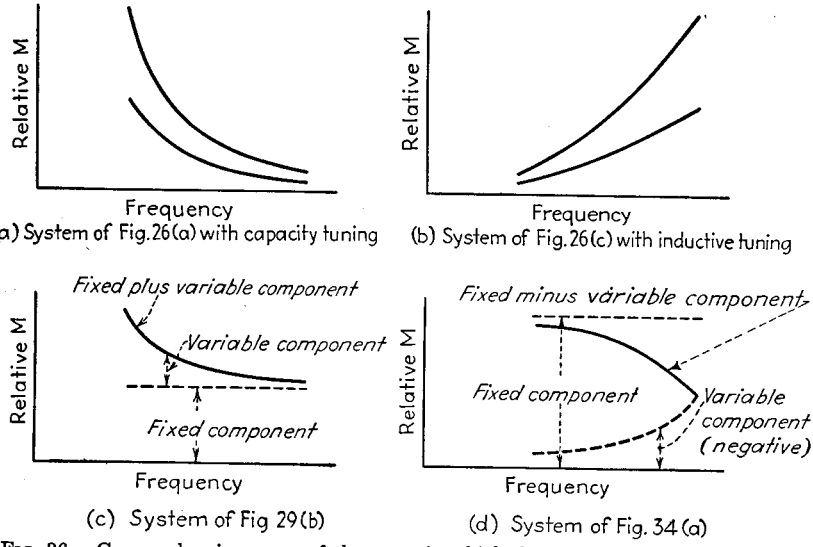


FIG. 36.—Curves showing some of the ways in which the equivalent mutual inductance can be made to vary with frequency in systems combining electrostatic and magnetic coupling.

The actual design of band-pass filters to give a desired performance can be carried out by the use of Figs. 19 and 20 and Eqs. (39) to (42) The method of doing this is illustrated by the following example:

Example.—It is desired to design a band-pass filter such that over a 10-kc band the output voltage is at least equal to the response at the middle of the band and does not exceed this response by a factor greater than 1.2. The pass band is to be centered on 162 kc, and the primary and secondary circuits are to have identical Q's.

The band width between peaks is $10/\sqrt{2} = 7.07$ kc, so that $\Delta f/f_0 = 7.07/162 = 0.0436$. Also, reference to Fig. 21 shows that for $Q_p = Q_s$ and a peak 1.2 times the response at resonance, k/k_c must be 1.86. If we turn to Fig. 19, we see that this corresponds to $g = 0.84$, and $k = 0.051$. Hence

$$Q = \frac{1}{k_c} = \frac{1.86}{0.051} = 36.5.$$

As a rough design rule for the case where Q_s and Q_p are approximately equal, a reasonably flat topped characteristic can be obtained over a frequency band approximately $1.2\ kf_0$, where f_0 is the frequency at the center of the band and k is the coefficient of coupling. The corresponding circuit Q's should then approximate the relation $\sqrt{Q_pQ_s} = 1.75/k$.[1]

Effect of Resonant Frequency on Band Width.—Where a band-pass filter is used for tuning purposes instead of a simple resonant circuit, the position of the pass band is ordinarily controlled by the use of identical primary and secondary variable condensers mounted on a common shaft. The width of the pass band in cycles is then approximately proportional to the resonant frequency of the circuits when inductive or direct coupling is used and approximately inversely proportional to the resonant frequency when capacitive coupling is employed. If the width of the pass band is to be constant irrespective of the frequency at which the band is located, it is necessary for the coefficient of coupling to be inversely proportional to the resonant frequency. Such a coefficient can be obtained by varying the coupling with the same control that adjusts the primary and secondary circuits, or can be approximated by the use of combined inductive and capacitive coupling.

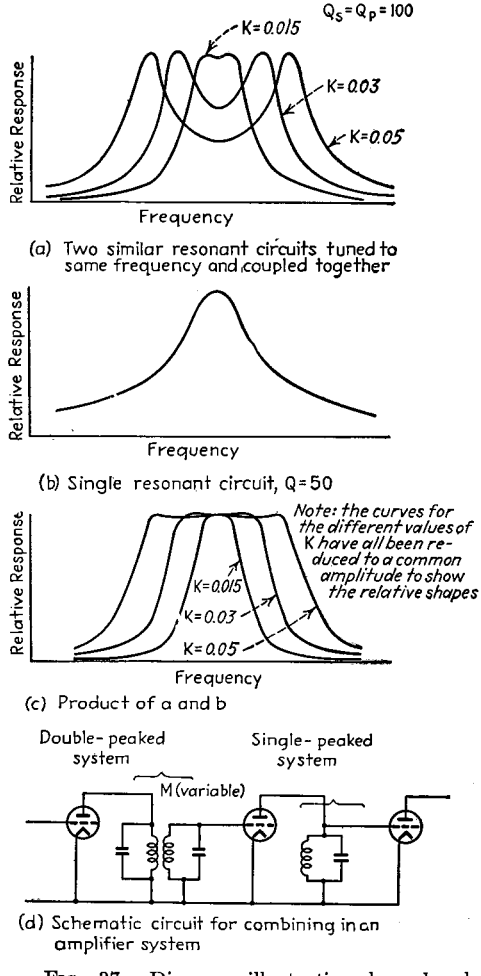


FIG. 37.—Diagram illustrating how band-pass action can be realized by combining a double-humped curve obtained from a pair of coupled circuits, with a simple resonant curve. By using proper circuit proportions, band-pass action is realized irrespective of the coefficient of coupling between the pair of coupled circuits.

Band-pass Characteristics by Detuning of Coupled Circuits.—When the primary and secondary circuits have the same Q, a slight detuning produces the same effect, as far as the shape of the secondary response curve is concerned, as increasing the coefficient

[1] Design proportions in the case where Q_s and Q_p differ appreciably are discussed in Par. 19, Sec. 5.

of coupling to a value $\sqrt{k^2 + (\Delta/f_0)^2}$, where k is the actual coefficient of coupling, Δ is the difference between the resonant frequencies, and f_0 is the average of the two resonant frequencies. The only essential difference between detuning and increasing the coupling is that with detuning the magnitude of the response is less than if the same-shaped curve is obtained by increasing coupling.

When the primary and secondary circuits have appreciably different Q's, detuning causes one side of the response curve to be depressed, as illustrated in Fig. 25, with a resulting undesirable characteristic.

Detuning effects in band-pass filters are of importance primarily because such detuning may occur accidentally through improper adjustments. This method of obtaining band-pass characteristics is not generally employed because of the necessity of maintaining equal primary and secondary circuit Q's and because of the practical difficulty of adjusting the circuits to the proper resonant frequencies.

Band-pass Action Obtained by Combining a Single- and Double-peaked Resonance Curve with a Simple Resonance Curve.[1]—Band-pass action can also be obtained by combining a double-peaked curve with the resonance curve of a simple single circuit, as shown in Fig. 37. Good band-pass characteristics result when the circuits are so proportioned that the product of the double-peaked and single-peaked curve is the same at resonance and at the two peaks. Analysis shows that this condition is obtained when the circuit Q's satisfy the relation[2]

$$Q_3 = \frac{Q_1}{1 + \frac{Q_1}{Q_2}} \tag{44}$$

where Q_1 and Q_2 apply to the two circuits that are coupled to give the double-peaked curve, and Q_3 applies to the simple resonant circuit. For $Q_1 = Q_2$, Eq. (44) requires that the Q_3 of the single resonant circuit should be half that of the two resonant circuits that are coupled together.

The condition specified in Eq. (44), is independent of the coupling. Hence if the Q's of the various circuits satisfy Eq. (44), variation of the coupling between the two coupled circuits will vary the width of the pass band without destroying the band-pass action, as shown in Fig. 37.

TRANSMISSION LINES[3]

10. Basic Equations and Concepts.—The voltage and current relations that exist on a transmission line can be expressed by the equations

$$\begin{aligned} E_s &= E_r \cosh \sqrt{ZY}l + I_r Z_0 \sinh \sqrt{ZY}l \\ I_s &= I_r \cosh \sqrt{ZY}l + \frac{E_r}{Z_0} \sinh \sqrt{ZY}l \end{aligned} \tag{45}$$

where $Z_0 = \sqrt{Z/Y}$ = characteristic impedance.

$\sqrt{ZY}$ = propagation constant (or hyperbolic angle per unit length).

[1] A more detailed discussion of band-pass filters of this type is given by Harold A. Wheeler and J. Kelly Johnson, High Fidelity Receivers with Expanding Selectors, *Proc. I.R.E.*, Vol. 23, p. 594, June, 1935, and Ho-Shou Loh, On Single and Couple Tuned Circuits Having Constant Response-band Characteristics, *Proc. I.R.E.*, Vol. 26, p. 469, April, 1938.

[2] Actually, the band-pass characteristic is slightly improved if the single peak curve corresponds to a Q slightly less than the value called for by Eq. (44).

[3] More detailed information on the theory of transmission lines can be found in many books on the subject. A good treatment for the radio engineer is to be found on pp. 94–178 of W. L. Everitt, "Communication Engineering," 2d ed., McGraw-Hill, New York. Other good books are L. F. Woodruff, "Electric Power Transmission and Distribution,' 2d ed., Wiley, New York; F. W. Norris and L. A. Bingham, "Electrical Characteristics of Power and Telephone Transmission Lines," International Textbook, Scranton, Pa.; A. E. Kennelly, "Hyperbolic Functions as Applied to Electrical Engineering," McGraw-Hill, New York.

$Z = R + j\omega L$ = series impedance per unit length.
$Y = G + j\omega C$ = shunt admittance per unit length.
l = length from receiver in same units as Z and Y.
R, L, C, and G = resistance, inductance, capacity and leakage, respectively, of line per unit length.
E and I are voltage and current, respectively.
Subscript s denotes sending or generator-end quantity.
Subscript r denotes receiving-end or load quantity.

Characteristic Impedance.—The characteristic impedance Z_0 appearing in Eq. (45) is given by the relation

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{L}{C}}\sqrt{\frac{1 + \frac{R}{j\omega L}}{1 + \frac{G}{j\omega C}}} \tag{46}$$

The characteristic impedance has the dimension of ohms, and is determined by the line proportions and the frequency.

At all radio frequencies, and also usually at all but the lowest audio frequencies, the ratios $R/\omega L$ and $G/\omega C$ are so small that the characteristic impedance can be taken as a pure resistance of $\sqrt{L/C}$ ohms. Furthermore, even when the line losses are not negligible, if $R/\omega L = G/\omega C$, the characteristic impedance is still a pure resistance equal to $\sqrt{L/C}$ ohms.

At frequencies so low that R is comparable with ωL and G comparable with ωC, the characteristic impedance varies with frequency, and approaches $\sqrt{R/G}$ as the frequency approaches zero.

Formulas and curves giving the characteristic impedance of various types of lines with air insulation are given in Figs. 38 and 39 for the case of high frequencies.[1] When dielectric insulation is employed, as with cables, the characteristic impedance is lower than the value with air insulation by the factor of $1/\sqrt{k}$, where k is the dielectric constant of the insulation.

Attenuation Constant.—The propagation constant $\sqrt{ZY}$ is commonly divided into real and imaginary parts

$$\sqrt{ZY} = \alpha + j\beta \tag{47}$$

The real part α is termed the *attenuation constant* and determines the energy dissipated in the line. The attenuation constant is determined by the resistance and leakage of the line, the frequency, and the extent to which R/L differs from G/C. At low frequencies, $\alpha = \sqrt{RG}$, which is the lowest possible value that the attenuation constant can have with a given resistance and leakage. As the frequency increases, the attenuation constant will in general become larger until at frequencies such that $\omega L >> R$ and $\omega C << G$, it reaches the limiting value

$$\left.\begin{matrix}\text{Attenuation constant} \\ \text{at high frequencies}\end{matrix}\right\} = \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{48}$$

In the special case where $R/L = G/C$, the attenuation does not increase with frequency and maintains its value $\alpha = \sqrt{RG}$ for all frequencies, even including very high frequencies, as shown by substitution in Eq. (48). A line with these particular

[1] The formulas in Fig. 38 are from J. F. Morrison, Transmission Lines, *Western Electric Pickups*, December, 1939. See also R. D. Duncan, Jr., Characteristic Impedance of Grounded and Ungrounded Open-wire Transmission Lines, *Communications*, Vol. 18, p. 10, June, 1938.

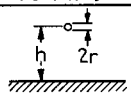
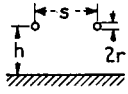
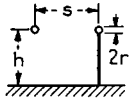
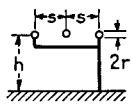
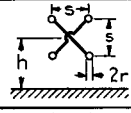
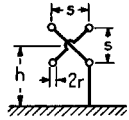
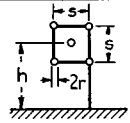
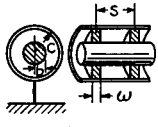
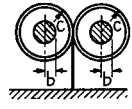
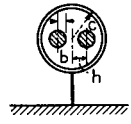
LOGARITHMS TO THE BASE 10			I_1 = GENERATOR CURRENT
LINE CONFIGURATION		CHARACTERISTIC IMPEDANCE	NET GROUND-RETURN CURRENT
Single wire		$Z_0 = 138 \text{ Log } \frac{2h}{r}$	$I_{Gnd} = I_1$
2-Wire balanced		$Z_0 = 276 \text{ Log } \frac{s}{r}$	$I_{Gnd} = 0$
2-Wire 1 wire grounded		$Z_0 \approx 276 \frac{\text{Log } \frac{s}{r} \text{ Log} \left[\rho^2 \frac{s}{r}\right]}{\text{Log} \left[\rho^2 \left(\frac{s}{r}\right)^2\right]}$ $\rho = \frac{2h}{s}$	$I_{Gnd} \approx I_1 \frac{\text{Log } \frac{s}{r}}{\text{Log } \frac{2h}{r}}$
3-Wire 2 wires grounded		$Z_0 \approx 69 \left[\text{Log } \frac{s^3}{2r^3} - \frac{\left(\text{Log } \frac{s}{2r}\right)^2}{\text{Log } \frac{2h^2}{rs}}\right]$	$I_{Gnd} \approx I_1 \frac{\text{Log } \frac{s}{2r}}{\text{Log } \frac{s\rho^2}{2r}}$ $\rho = \frac{2h}{s}$
4-Wire balanced		$Z_0 = 138 \left(\text{Log } \frac{s}{r}\right) - 21$	$I_{Gnd} = 0$
4-Wire 2-wires grounded		$Z_0 \approx 138 \left[\frac{\text{Log } \frac{s}{r\sqrt{2}} \text{ Log} \left[\rho^4 \frac{s}{r\sqrt{2}}\right]}{\text{Log} \left[\rho^4 \left(\frac{s}{r\sqrt{2}}\right)^2\right]}\right]$ $\rho = \frac{2h}{s}$	$I_{Gnd} \approx I_1 \frac{\text{Log } \frac{s}{r\sqrt{2}}}{\text{Log } \frac{\rho^2 s}{r\sqrt{2}}}$
5-Wire 4 wires grounded		$Z_0 \approx 138 \left[\text{Log } \frac{2h}{r} - \frac{\left[\text{Log } 2\rho^2\right]^2}{\text{Log} \left[\rho^{\frac{3h\sqrt{2}}{r}}\right]}\right]$ $\rho \frac{2h}{s}$	$I_{Gnd} \approx I_1 \frac{\text{Log } \frac{s}{r4\sqrt{2}}}{\text{Log } \frac{s\rho^4}{r\sqrt{2}}}$
Concentric (coaxial)		$Z_0 = 138 \frac{\text{Log } \frac{c}{b}}{\sqrt{1 + \frac{(\varepsilon - 1)\omega}{s}}}$ ε = Dielectric constant of insulating material	
Double coaxial balanced		$Z_0 = 276 \frac{\text{Log } \frac{c}{b}}{\sqrt{1 + \frac{(\varepsilon - 1)\omega}{s}}}$	
Shielded pair balanced		$Z_0 = \frac{120}{\sqrt{\varepsilon}} \left[2.303 \text{ Log} \left(2v \frac{1-\sigma^2}{1+\sigma^2}\right) - \frac{1+4v^2}{16v^4} \left(1 - 4\sigma^2\right)\right]$ ε = Dieletric constant of medium ε = Unity for gaseous medium $v = \frac{h}{b}$; $\sigma = \frac{h}{c}$	

FIG. 38.—Characteristics of various types of transmission lines erected parallel to a perfectly conducting earth.

proportions is said to be a *distortionless* line because it has the same attenuation for all frequencies.

In lines in which the insulation is primarily air, the leakage conductance G approaches zero. Also, at high radio frequencies, skin effect is so great that the

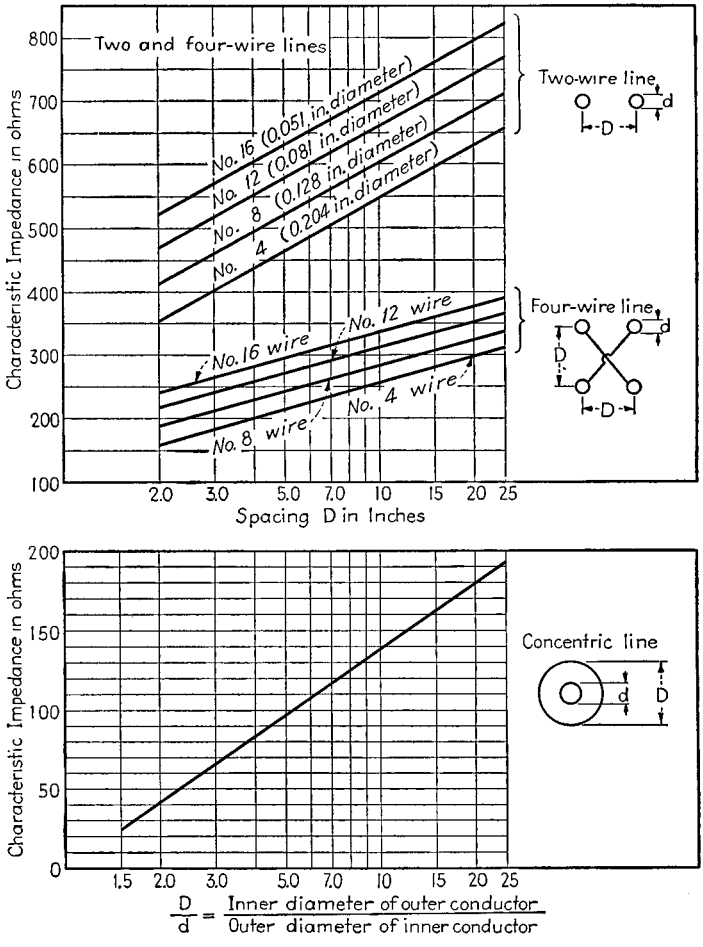


FIG. 39.—Characteristic impedance of common types of transmission lines.

resistance of the conductors is given accurately by Eq. (7), Sec. 2. Under these conditions for a copper line, one has the following:

For two-wire line:[1]

$$\alpha = 0.00362 \frac{\sqrt{f_{mc}}}{d_{in} \log_{10}\left(\frac{2D}{d}\right)} \text{ per 1,000 ft} \tag{49}$$

[1] This neglects proximity effect. For very close spacing, this value of α should be multiplied by the factor given in Fig. 6, Sec. 2.

For concentric line:

$$\alpha = 0.00362 \frac{\sqrt{f_{mc}}\left(1 + \frac{D}{d}\right)}{D_{in} \log_{10}\left(\frac{D}{d}\right)} \text{ per 1,000 ft} \tag{50a}$$

where f_{mc} is the frequency in megacycles and d and D have the meanings shown in Fig. 39, and are in inches (indicated by the subscript *in*).

In the case of a concentric line, if D is fixed, the lowest attenuation occurs when $D/d = 3.6$, corresponding to $Z_0 = 77$ ohms and giving

$$\left.\begin{array}{l}\text{Attenuation } \alpha \text{ of optimum} \\ \text{concentric line per} \\ \text{1,000 ft}\end{array}\right\} = \frac{0.0299\sqrt{f_{mc}}}{D_{in}} \tag{50b}$$

The power that can be transmitted for a given maximum voltage gradient in a concentric line having an outer conductor diameter D is a maximum when $D/d = 1.65$, corresponding to $Z_0 = 30$ ohms.

Values of attenuation constants expressed in db per 1,000 feet are plotted in Fig. 40 for two-wire and concentric copper lines having air insulation.[1] When solid insulation is employed, the attenuations will be much greater. Thus cable with rubber insulation can be expected to have an attenuation from 10 to 50 times that of a corresponding 600-ohm air-insulated two-wire line.[2]

Phase Constant, Wave Length, and Velocity of Phase Propagation.—The imaginary part β of $\sqrt{ZY}$ is termed the *phase constant*, and determines the wave length and velocity of phase propagation.

A wave length is defined as a length of line l such that $\beta l = 2\pi$. Hence

$$\left.\begin{array}{l}\text{Length corresponding} \\ \text{to one wave length}\end{array}\right\} = \lambda = \frac{2\pi}{\beta} \tag{51}$$

In the case of lines with air insulation, the distance corresponding to a wave length will be only very slightly less than the wave length of a radio wave of the same frequency traveling in space. Accordingly

$$\left.\begin{array}{l}\text{Approximate wave length for} \\ \text{lines with air insulation}\end{array}\right\} = \frac{300{,}000{,}000}{f} \text{ meter} \tag{52}$$

where f is in cycles.

With dielectric insulation, as in the case of cables, the wave length is less than with air insulation by the factor $1/\sqrt{k}$, where k is the dielectric constant of the insulation.

The product of frequency and wave length corresponds to velocity, and is termed the *velocity of phase propagation* or simply *wave velocity*. This is not a true velocity in the sense that something tangible actually travels at a certain speed, but rather means that the relative phases at different places on the line are such as to give the appearance of velocity. With air insulation, the velocity at high frequencies is practically the same as the velocity of light; while with dielectric insulation the velocity is $1/\sqrt{k}$ as great.

Hyperbolic Functions and Complex Components.—The hyperbolic functions cosh x and sinh x appearing in Eq. (45) (where $\sqrt{ZY}l = x$) represent certain combinations of exponents as follows:

[1] The attenuation in db is 8.686 αl, and gives the energy loss in db for the line operated as a nonresonant line (see Par. 14).

[2] For further information, see C. C. Harris, Losses in Twisted Pair Transmission Lines at Radio Frequencies, *Proc. I.R.E.*, Vol. 24, p. 425, March, 1936.

$$\cosh x = \frac{\epsilon^{x} + \epsilon^{-x}}{2}$$
$$\sinh x = \frac{\epsilon^{x} - \epsilon^{-x}}{2} \tag{53}$$

The more common and useful properties of hyperbolic functions are given in Par. 15, Sec. 1. Values of the function when x is real are tabulated in Pars. 6, 7, and 8, Sec. 1.

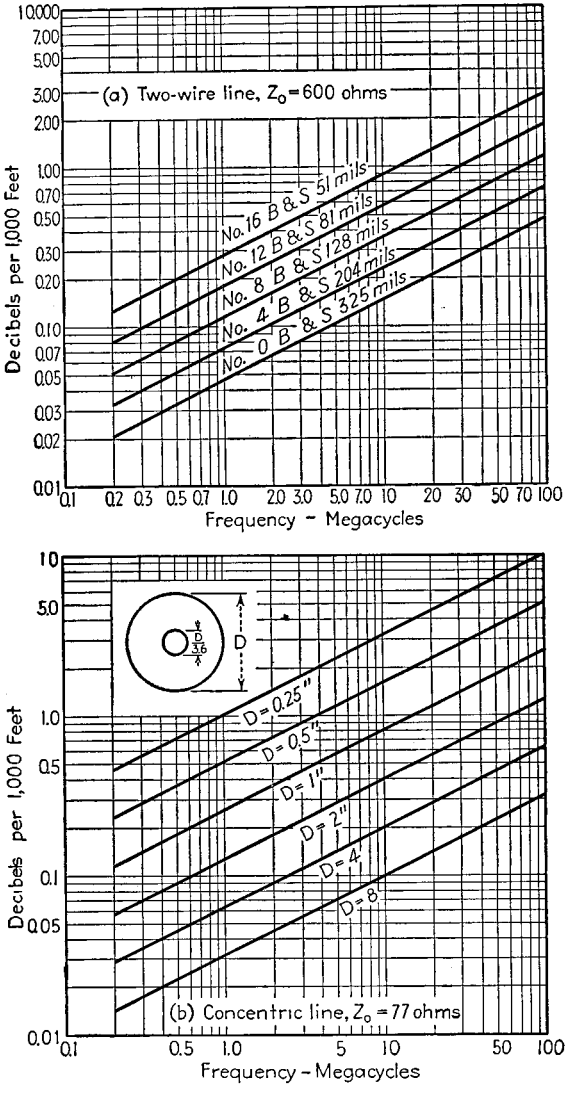


FIG. 40.—Attenuation of two-wire and concentric transmission lines.

When x is complex, as is the case in Eq. (45) where $x = \sqrt{ZY}l = \alpha l + j\beta l$, the hyperbolic function can be evaluated by any one of the following methods: (1) the

use of Kennelly's charts,[1] (2) expansion of the complex hyperbolic function into functions of real circular and real hyperbolic angles, and (3) direct evaluation of the complex exponents appearing in the definition of the function. Examples illustrating the details involved in applying the second and third methods are as follows:

Example 1.—Evaluate $\sinh 2/\underline{60°}$ by the expansion method.

$$\sinh 2/\underline{60°} = \sinh (1.0 + j1.732)$$

Referring to Par. 15, Sec. 1

$$\begin{aligned}\sinh (1.0 + j1.732) &= \sinh 1.0 \cos 1.732 + j \cosh 1.0 \sin 1.732 \\ &= 1.175 \times (-0.1599) + j1.547 \times 0.987 \\ &= -0.188 + j1.525 \\ &= 1.53/\underline{97°0'}\end{aligned}$$

Example 2.—Evaluate $\sinh 2/\underline{60°}$ by direct evaluation of the complex exponents.

$$\sinh 2/\underline{60°} = \frac{\epsilon^{2/\underline{60°}} - \epsilon^{-2/\underline{60°}}}{2}$$

But $2/\underline{60°} = 1.0 + j1.732$

$$\begin{aligned}\therefore \epsilon^{2/\underline{60°}} &= \epsilon^{(1.0+j1.732)} = \epsilon^{1.0}\ /\underline{1.732} \text{ radians} \\ &= 2.718/\underline{99°12'} = -0.434 + j2.675\end{aligned}$$

Similarly

$$\begin{aligned}\epsilon^{-2/\underline{60°}} &= \frac{1}{\epsilon^{2/\underline{60°}}} = 0.368/\underline{-99°12'} \\ &= -0.0589 - j0.363\end{aligned}$$

Hence

$$\begin{aligned}\sinh 2/\underline{60°} &= \frac{-0.434 + 0.059 + j(2.675 + 0.363)}{2} \\ &= -0.187 + j1.52 \\ &= 1.53/\underline{97°0'}\end{aligned}$$

The relationships represented by $\cosh (\alpha + j\beta)l$ and $\sinh (\alpha + j\beta)l$ are wavy curves with regular but progressively decreasing amplitude of oscillation and increasing mean amplitude as l becomes greater, as shown in Fig. 41. The difference between $\cosh x$ and $\sinh x$ is that the latter begins at zero, the former at unity. In both cases, a maximum or minimum occurs whenever βl passes through a multiple of $\pi/4$. On the other hand, the amplitude of the maxima and minima increases as αl becomes greater, whereas the range of the oscillations becomes less, as shown in Fig. 41.

Note on Handling Complex Quantities in Polar Form.—All the terms appearing in Eqs. (45) are vector quantities. In order to carry on the multiplications, divisions, and extraction of roots that are indicated, it is convenient to express Z, Y, Z_0, $\cosh \sqrt{ZY}l$, etc., in polar form. The following rules for manipulation then apply:

Products are formed by multiplying magnitudes of the vectors and adding their angles.
Quotients are formed by dividing magnitudes and subtracting angles.
Sums and differences are obtained by converting to rectangular form and adding the real and imaginary parts separately, after which conversion can be made to polar form if desired.
Square roots are obtained by taking the square root of the vector magnitude and dividing the angle by two to obtain the resulting vector.
Powers are obtained by raising the length to the appropriate power and multiplying the angle by the power.

11. Voltage and Current Distribution—Wave Trains and Their Reflection.—The expressions for the voltage and current relations of a transmission line given by Eqs. (45) can be rearranged as follows:

[1] A. E. Kennelly, Chart Atlas of Complex Hyperbolic and Circular Functions, Harvard Univ. Press, Cambridge, Mass., 1924.

$$E_s = \left(\frac{E_r + I_r Z_0}{2}\right) \epsilon^{\sqrt{ZY}l} + \left(\frac{E_r - I_r Z_0}{2}\right) \epsilon^{-\sqrt{ZY}l}$$
$$I_s = \left(\frac{I_r + \frac{E_r}{Z_0}}{2}\right) \epsilon^{\sqrt{ZY}l} + \left(\frac{I_r - \frac{E_r}{Z_0}}{2}\right) \epsilon^{-\sqrt{ZY}l} \tag{54}$$

It is to be noted that l is distance from the load. This shows that the voltages and currents existing on the line can be considered as comprising the sum of two components. The first of these consists of a voltage having a magnitude $(E_r + I_r Z_0)/2$ accompanied everywhere by a current that is equal to this voltage divided by the characteristic impedance. This first component is proportional to $\epsilon^{\sqrt{ZY}l}$, and can be termed a *wave train traveling toward the receiver.* The second component consists of a voltage $(E_r - I_r Z_0)/2$ accompanied by a current equal to the voltage divided by $-Z_0$. This component is proportional to $\epsilon^{-\sqrt{ZY}l}$, and can be termed a *wave train traveling from the receiver toward the generator.* This second wave train can be thought of as resulting from the reflection of the first wave train by the load impedance at the receiver end of the line. For this reason, the wave traveling to the receiver is often designated as the *incident wave,* while the second wave is termed the *reflected wave.*

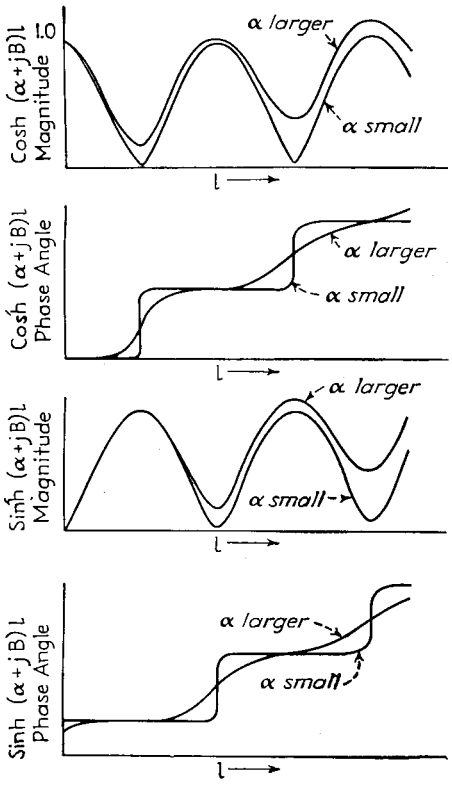


FIG. 41.—Curves showing the nature of the functions cosh $(\alpha + j\beta)$ and sinh $(\alpha + j\beta)$.

Reflection of Wave Trains.—The incident and reflected waves existing at the receiving point must be so related as to give the required receiver voltage and current, and at the same time satisfy the properties of the wave trains. The vector ratio of voltage of the reflected wave to voltage of the incident wave is termed the *coefficient of reflection,* and depends upon the vector ratio Z_L/Z_0 of load impedance to characteristic impedance according to the equation

$$\text{Coefficient of reflection} = \frac{\frac{Z_L}{Z_0} - 1}{\frac{Z_L}{Z_0} + 1} \tag{55}$$

The vector ratio of voltage of the incident wave to the receiver voltage is

$$\frac{\text{Voltage of incident wave}}{\text{Receiver voltage}} = \frac{1 + \frac{Z_0}{Z_L}}{2} \tag{56}$$

The magnitude and phase angle of the reflection coefficient can be conveniently obtained from Fig. 42 for a given Z_L/Z_0.[1]

Characteristics of Wave Trains.—Considering first the incident wave, as the distance from the receiver increases (*i.e.*, as $\sqrt{ZY}l$ increases), the size and phase of this wave varies by the factor $\epsilon^{\sqrt{ZY}l} = \epsilon^{(\alpha+j\beta)l} = \epsilon^{\alpha l}/\underline{\beta l}$. The amplitude hence increases exponentially according to the factor $\epsilon^{\alpha l}$. At the same time, the phase advances by the factor βl radians, or β radians per unit length.

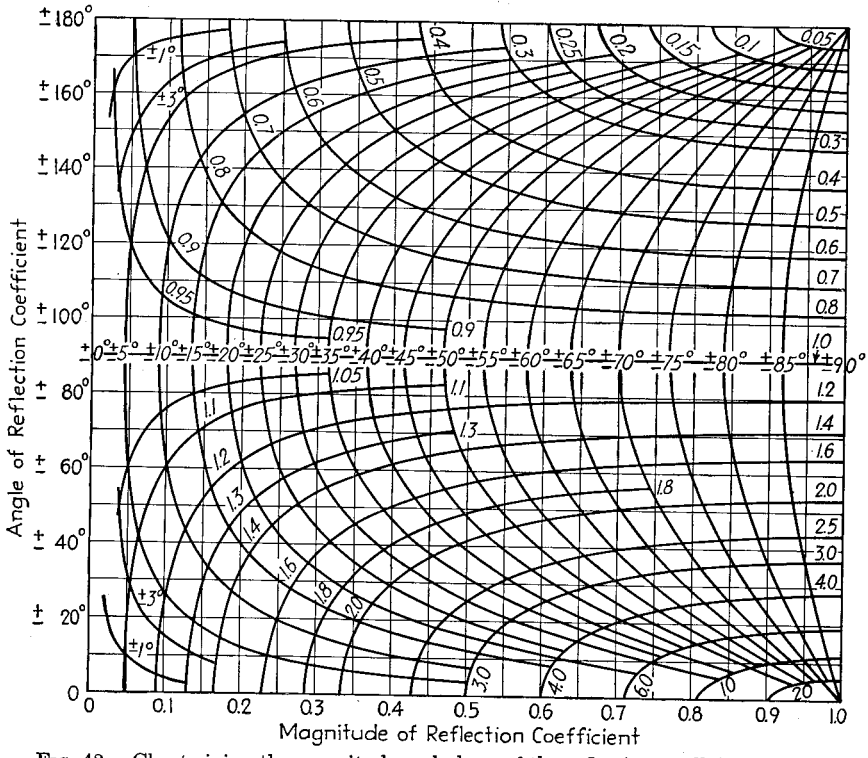


FIG. 42.—Chart giving the magnitude and phase of the reflection coefficient at the load in terms of the vector ratio Z_L/Z_0 of load to characteristic impedance. The numbers on the curvilinear system of coordinates represent the magnitude and phase of the ratio Z_L/Z_0.

If we turn next to the reflected wave, the variation of magnitude and phase with increase in l is according to the factor $\epsilon^{-\sqrt{ZY}l} = \epsilon^{-(\alpha+j\beta)l} = \epsilon^{-\alpha l}/\underline{-\beta l}$. The amplitude now varies according to $\epsilon^{-\alpha l}$, and so decreases exponentially with increasing distance from the receiver. At the same time, the phase drops back β radians per unit length, or by a total amount βl radians.

It will be noted that the phase shift of the individual wave trains is 2π radians per wave length or 90° per quarter wave length. It is also to be noted that when αl is small there is very little change in magnitude of the individual waves.

Voltage and Current Distribution with Different Receiving-end Terminations.—The actual voltage and current existing on a transmission line is always the vector sum

[1] An equivalent chart expressed in a different form is given by P. S. Carter, Charts for Transmission-line Measurements and Computations, *R.C.A. Rev.*, Vol. 3, p. 355, January, 1939.

of the two wave trains described above. The nature of the resulting distribution depends upon the relative amplitude and phase of the two wave trains at the receiving point, upon the phase shift βl, *i.e.*, upon the distance to the receiver measured in wave lengths, and upon the attenuation αl to the receiver.

With an open-circuit receiver, Eq. (56) shows that the reflection coefficient is $+1$. This means that at the reflection point, the incident and reflected waves have the

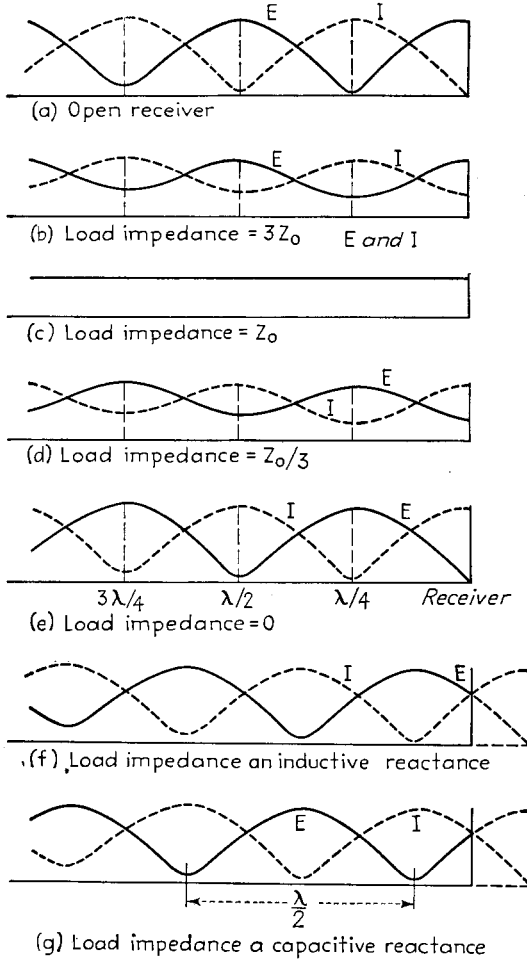
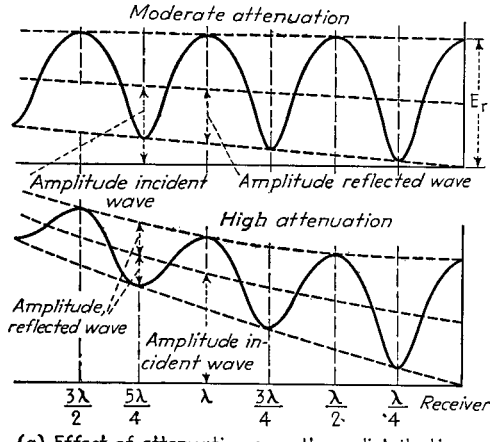


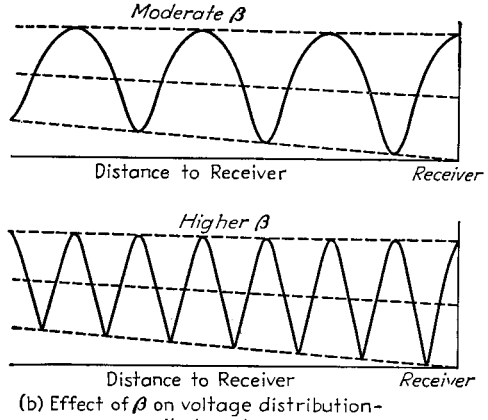
FIG. 43.—Types of voltage and current distributions produced on transmission lines by different load impedances.

same amplitude, with both voltages in the same phase and the two currents in phase opposition. The vector sum of two such wave trains is shown in Fig. 43*a*. The voltage minima occur at distances from the receiver corresponding to an odd number of quarter wave lengths. These are points at which the phase advance βl of the incident wave and the phase retardation $-\beta l$ of the reflected wave are such as to cause the two voltages to be out of phase. The current maxima and minima coincide with voltage minima and maxima, respectively.

With a short-circuited receiver, the reflection coefficient is -1. The reflection is again complete, but takes place with reversal in voltage and no change in phase of current. Accordingly, the distributions are similar to those obtained with an open-circuit receiver, except that the voltage and current curves have been interchanged, as shown in Fig. 43*e*.



(a) Effect of attenuation on voltage distribution-open circuited receiver



(b) Effect of β on voltage distribution-open circuited receiver

FIG. 44.—Curves showing the effect of the attenuation constant and phase constant on the voltage distribution in an open-circuited transmission line.

When the load impedance equals the characteristic impedance, the reflection coefficient is zero, *i.e.*, there is no reflected wave train. The current and voltage distribution then follow an exponential law, as shown in Fig. 43*c*, and no resonances are present.

With load impedances having the same phase angle as the characteristic impedance but greater magnitude, the reflection coefficient is positive but less than unity. The reflected wave then has the same phase relations as for the open-circuit receiver, but is less in magnitude. This results in a distribution of the type shown in Fig. 43*b*, which is similar to the open-circuit case, but with less pronounced resonances. The corresponding situation when the load impedance is less than the characteristic impedance resembles the short-circuit case, but again the resonances are less pronounced, as in Fig. 43*d*.

Load impedances having a phase angle differing from the characteristic impedance by 90° (corresponding to a reactive load and a resistance characteristic impedance) give a reflection coefficient that is unity, but that has a phase shift associated with it. The resonances are then just as pronounced as in the open- and short-circuit cases, but the position where the first voltage minimum occurs is displaced toward the receiver with capacitive loads and away from the receiver with inductive loads, as compared with the open-circuited line (see Figs. 43*f* and 43*g*). The amount of this displacement is $\delta/4\pi$ wave lengths, where δ is the phase angle of the reflection coefficient.

In the case where the load impedance has both a reactive and a resistive component, the reflected wave is smaller than the incident wave and also has its phase shifted. This causes the ratio of maximum to minimum amplitudes to be smaller than with complete reflection and also modifies the distance from the receiver at which the first minimum occurs.

Effect of Attenuation Constant and Phase Constant (or Frequency) on Distribution Curves.—The attenuation constant determines the rate at which the amplitude of the individual wave trains varies. When the attenuation is very small, the waves change only very slightly in amplitude when traveling large distances, and the envelopes of the voltage and current distribution curves therefore change only slightly. On the other hand, when the attenuation constant is large enough so that $\epsilon^{\alpha l}$ differs appreciably from unity, the wave trains change their amplitude relatively rapidly. As the distance from the receiver is increased, the incident wave rapidly becomes larger, and the reflected wave dies out. This results in the behavior that is shown in Fig. 44*a*.

The phase constant β is almost exactly proportional to the frequency, and determines the length required to give a quarter wave length. Increasing β (as, for example, by increasing the frequency) accordingly causes the maxima and minima of the distributions to be more closely spaced, as shown in Fig. 44*b*.

Phase Shifts Produced by Lines.—In the case of a nonresonant line (load impedance = characteristic impedance), there is only a single wave train. The voltage (or current) then shifts in phase uniformly at the rate of 2π radians for each wave length, the phase advancing as the generator is approached.

When resonances occur, the phase relations are more complicated. The phase of the voltage (or current) distribution still advances 180° from one minimum in the distribution to the next minimum toward the generator, but unlike the resonant line, most of the phase shift tends to take place in the vicinity of the minima.

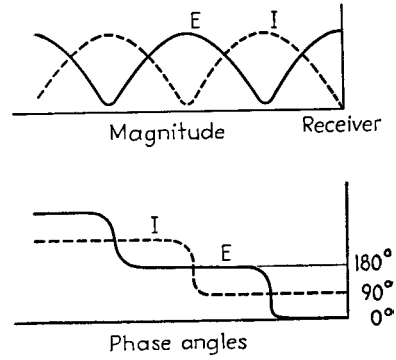


FIG. 45.—Variation of the phase of voltage and current on an open-circuited transmission line with increasing distance to the load.

Phase relations are shown in Fig. 45 for conditions corresponding to the case illustrated in Fig. 43*a*.

12. Impedance Properties of Transmission Lines.—The sending-end impedance of a transmission line is

$$\left.\begin{matrix}\text{Sending-end}\\ \text{impedance}\end{matrix}\right\} = Z_s = Z_0 \frac{\left(\frac{Z_L}{Z_0}\right)\cosh\sqrt{ZY}l + \sinh\sqrt{ZY}l}{\left(\cosh\sqrt{ZY}l + \frac{Z_L}{Z_0}\sinh\sqrt{ZY}l\right)} \tag{57}$$

In special limiting cases, this reduces to

Open-circuit receiver $\left(\frac{Z_L}{Z_0} = \infty\right)$:

$$Z_s = \frac{Z_0}{\tanh\sqrt{ZY}l} \tag{58}$$

Short-circuited receiver $\left(\frac{Z_L}{Z_0} = 0\right)$:

$$Z_s = Z_0 \tanh\sqrt{ZY}l \tag{59}$$

Nonresonant line ($Z_L = Z_0$):

$$Z_s = Z_0 = Z_L \tag{60}$$

In these equations, $\tanh \sqrt{ZY}l = \sinh \sqrt{ZY}l / \cosh \sqrt{ZY}l$, and the remaining notation is the same as used above.

The variation of line impedance with frequency (or length) is given by the same factors that control the voltage and current distribution. As the length is varied, the impedance is maximum when the voltage is maximum, and is lowest when the voltage goes through a minimum.

The sending-end impedance of a line of fixed length follows a curve such as is shown in Fig. 46*a* when the frequency is varied. Here the impedance maxima are spaced at frequency intervals such that each new maximum corresponds to an extra half wave length in the line. When the receiver is open-circuited, these impedance maxima

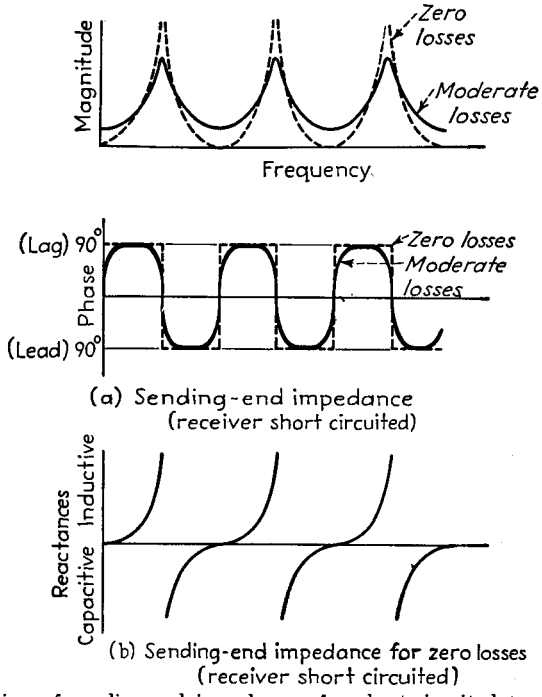


FIG. 46.—Variation of sending-end impedance of a short-circuited transmission line of fixed length as frequency is changed.

occur at frequencies such as to make the line an even number of quarter wave lengths long, while with a short-circuited receiver, the impedance maxima occur at frequencies that make the line an odd number of quarter wave lengths long.

The power factor of the line impedance is largely reactive except at the peaks and troughs, where the power factor goes through unity. The power factor alternates between leading and lagging each time the length measured in wave lengths increases a quarter of a wave length, as shown in Fig. 46.

13. Properties of Transmission Lines with Zero Losses.—In the usual line used in the transmission of radio-frequency power, the losses are so small compared with the power being transmitted that it is permissible for most purposes to assume that the line losses are zero. Also, in most radio antennas, the errors from assuming that the current distribution is as though the losses were zero introduce only a second

order effect on the principal properties of the antenna, so that again the assumption of zero losses is commonly permissible.

When the resistance and leakage of a line are zero, $\alpha = 0$, $Z_0 = \sqrt{L/C}$, and $\sqrt{ZY} = j2\pi/\lambda$. Equations (45) then become

$$E_s = E_r \cos\left(2\pi \frac{l}{\lambda}\right) + jI_r \sqrt{\frac{L}{C}} \sin\left(2\pi \frac{l}{\lambda}\right)$$
$$I_s = I_r \cos\left(2\pi \frac{l}{\lambda}\right) + j \frac{E_r}{\sqrt{\frac{L}{C}}} \sin\left(2\pi \frac{l}{\lambda}\right) \tag{61}$$

where l and λ are in the same units.

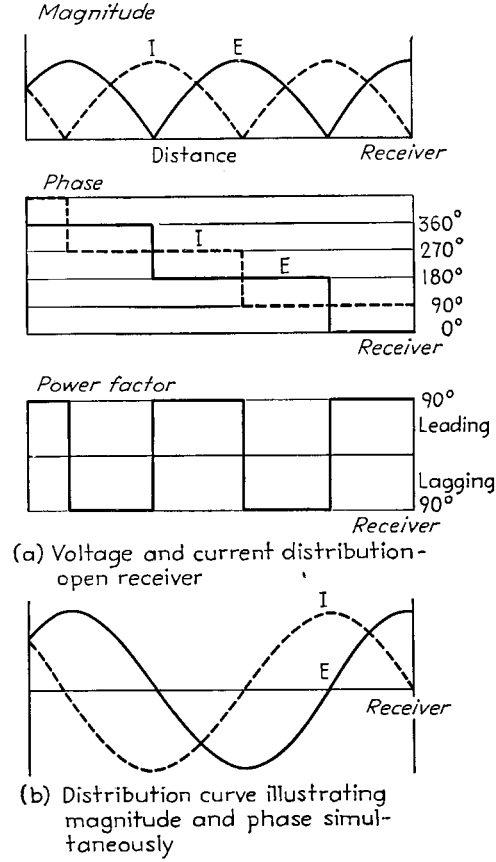


FIG. 47.—Voltage and current relations on an ideal open-circuited transmission line having zero losses.

The maxima in the distribution curves of current or voltage are similar to those of Fig. 43, except that all maxima have the same height when the attenuation is assumed to be zero, as also do the minima. The ratio of minimum to maximum is

$$\frac{(1-k)}{(1+k)},$$

where k is the magnitude of the reflection coefficient. In the special case where $k = 1$, corresponding to an open- or short-circuited receiver or a reactive load, the distribution curves are sections of half sine waves, as shown in Fig. 47*a*, the phase of the voltage (or current) jumps 180° as each minimum is passed through, and the power factor jumps from 90° leading to 90° lagging (or vice versa) at each minimum, as indicated in the figure. The distribution curves for this case can then be drawn as in Fig. 47*b*, which simultaneously shows both magnitude and phase by using negative amplitudes to indicate opposite phase from positive amplitudes.

In the case of zero losses, Eq. (57), giving the impedance of the line, reduces to

$$Z_s = Z_0 \frac{\frac{Z_L}{Z_0}\cos\left(2\pi\frac{l}{\lambda}\right) + j\sin\left(2\pi\frac{l}{\lambda}\right)}{\cos\left(2\pi\frac{l}{\lambda}\right) + j\frac{Z_L}{Z_0}\sin\left(2\pi\frac{l}{\lambda}\right)} \tag{62}$$

In the special case where *the length is exactly a whole number of half wave lengths, the sending-end impedance is exactly equal to the load impedance.* Also, when the length is exactly an odd number of quarter wave lengths, then

$$Z_s = \frac{Z_0^2}{Z_L} \tag{63}$$

When the reflection is unity, corresponding to open-circuited, short-circuited, or reactive loads, the impedance goes to infinity at the frequencies that make the impedance a maximum, and the power-factor angle is everywhere 90° except at these transition points (see dotted curves, Fig. 46*a*). The curve of reactance as a function of frequency can then be drawn as in Fig. 46*b*.

14. Transmission of Radio-frequency and Audio-frequency Power by Lines.[1]—An important use of transmission lines is in carrying radio-frequency power from a transmitter to an antenna or from an antenna to a radio receiver. Lines used for this purpose are commonly said to be *resonant* or *nonresonant*, according to the method of operation.

Nonresonant Lines.—In the nonresonant lines, the terminating impedance Z_L at the end to which the line delivers the power is made equal to the characteristic impedance. There is then only the incident wave train present, and the power is transmitted at substantially unity power factor. The power loss and peak voltage to which the line is subjected are then less than with any other mode of operation.

The losses[1] in a nonresonant line depend upon the line attenuation αl, and are usually expressed in decibels, using the relation

$$\text{Decibels} = 8.686\,\alpha l \tag{64}$$

With air insulation the attenuation constant α at radio frequencies is inversely proportional to the linear dimension of the line (assuming fixed proportions) and directly to $\sqrt{f}$ (see Par. 10).

Curves giving the attenuation for two-wire copper lines spaced so that the characteristic impedance is 600 ohms, and corresponding curves for a concentric line of optimum proportions ($D/d = 3.6$) are given in Fig. 40.

Resonant Lines.—When the load impedance does not equal the characteristic impedance, a reflected wave exists, and the line is said to be resonant because of the resulting oscillatory character of the current and voltage distributions. Resonant operation has the advantage of not requiring an impedance match between the line and the load. At the same time it has the disadvantage that the average power factor

[1] An excellent discussion of the subject is given by E. J. Sterba and C. B. Feldman, Transmission Lines for Short Wave Radio Systems, *Proc. I.R.E.*, Vol. 20, p. 1163, July, 1932.

along the line is lower than for nonresonant action. This results in increased power loss and greater voltage stress in proportion to transmitted power than for the corresponding nonresonant case. Curves giving the increase in loss and in voltage stress of a resonant line over a nonresonant line as a function of the reflection coefficient are given in Fig. 48. Nonresonant operation also leads in many cases to loss in power transferred, as a result of mismatching.

Transmission of Audio-frequency Power over Moderate Distances.—Transmission lines are frequently used for transmitting audio-frequency energy for short distances, such as from one part of a building to another part. Such lines differ from radio-frequency lines and ordinary telephone lines in that the length, at even the highest audio frequencies, is only a small fraction of a quarter wave length. The possibility of resonance, accordingly, does not need to be considered, and there is no advantage in making terminating impedance equal to the characteristic impedance. The impedance as viewed from the sending end is essentially equal to the terminating load

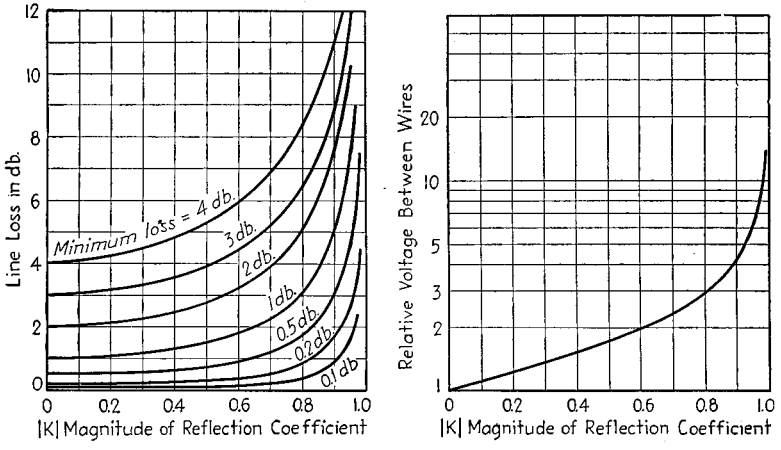


FIG. 48.—Effect of resonance on transmission line loss, and on maximum voltage between wires. The minimum loss is the loss for $k = 0$.

impedance. Such short lines are frequently referred to as 50-ohm lines, 100-ohm lines, 500-ohm lines, etc., according to the load impedance used, quite independently of the characteristic impedance.

15. Impedance Matching.—When it is desired to obtain nonresonant operation, it is necessary to make provision for coupling the load impedance to the line in such a way that the effective terminating impedance of the line equals the characteristic impedance. The methods commonly used for doing this at radio frequencies are:

Impedance Matching Network.—This arrangement makes use of a four-terminal reactive network so proportioned that the input impedance offers the desired resistance load to the line when the actual load impedance is connected to the output terminals. Impedance matching networks can also be designed to introduce any desired phase shift.

Impedance matching networks can take a variety of forms. Thus in broadcast work, unbalanced T, π, or L networks are commonly used to couple a grounded antenna to a concentric transmission line, as shown schematically in Fig. 49*a*. The design of such networks is discussed in Par. 25.

Transformers are also frequently used in impedance matching. A simple arrangement is shown in Fig. 49*b*, in which the tuning of the circuits and the mutual induct-

ance are chosen to make the load impedance offered to the line a resistance of the proper magnitude. Transformers are sometimes combined with T, π, and L networks, and are always necessary, either alone or in combination with other circuit elements, when a balanced transmission line is to be coupled to a load that is unbalanced to ground.

In any impedance matching network it is to be noted that the network consumes no power, provided it is composed of reactive elements having negligible loss. Practically all the power delivered to the input terminals of such an impedance matching network finds its way to the load.

Resonant Line Coupling.—This arrangement is shown in Fig. 49*c*, and employs an auxiliary line as an impedance matching network. The load is connected to one end

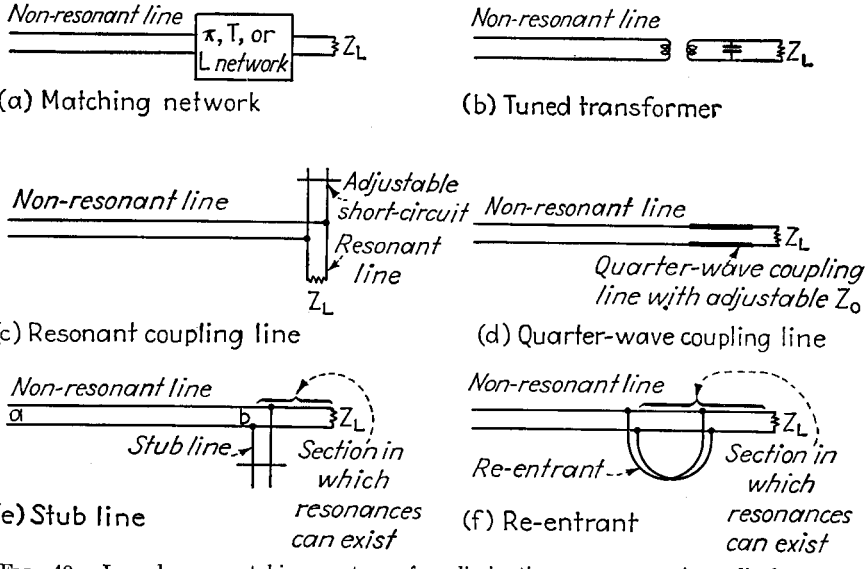


FIG. 49.—Impedance matching systems for eliminating resonances in radio-frequency transmission lines.

of this auxiliary line, and the position of an adjustable short circuit at the other end is varied until the line with the associated load is in resonance. The impedance offered to an external source of power connected between the two wires of the auxiliary line is then a resistance having a value depending upon the point at which the connection is made to the auxiliary line. The characteristic impedance of the nonresonant line can now be matched by connecting to the auxiliary line at a suitable point. The length of the auxiliary depends upon the magnitude and power factor of the actual load, and may range up to a half wave length.

Quarter-wave Coupling Line.—This arrangement, illustrated in Fig. 49*d*, makes use of the fact that a quarter-wave line of negligible losses will transform impedance according to Eq. (63). The reactance component of the load is first made zero by tuning the load to resonance, after which the characteristic impedance of the coupling line is varied by adjusting its spacing until the impedance at the transmission line end of the quarter wave coupling section has the required value.

Stub-line Impedance Matching System.—This arrangement is illustrated in Fig. 49*e*. The load is first connected directly to the transmission line, as in the case of resonant operation. Then, at a point within a wave length or less of the load, a stub

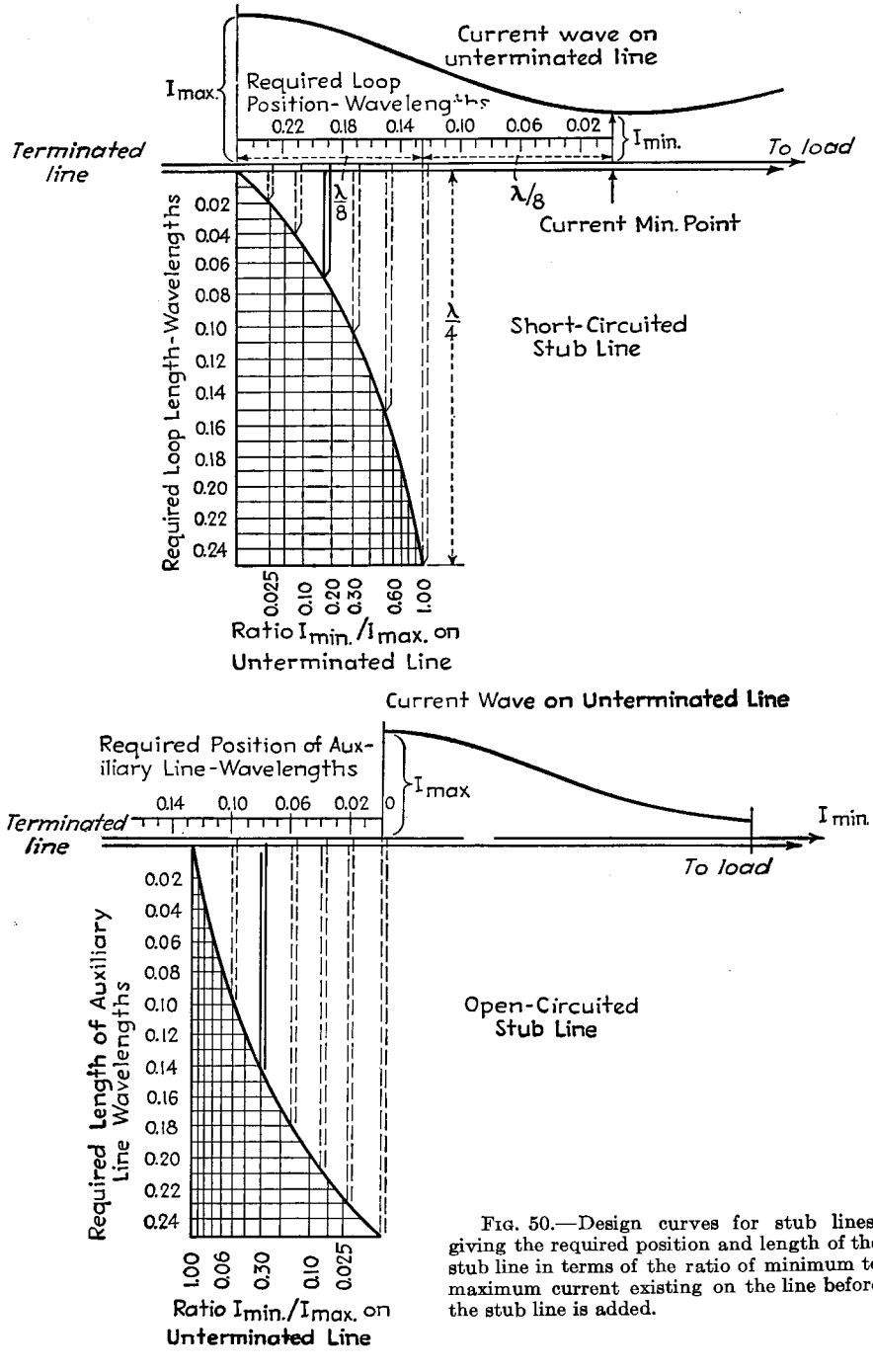


FIG. 50.—Design curves for stub lines, giving the required position and length of the stub line in terms of the ratio of minimum to maximum current existing on the line before the stub line is added.

line is connected in shunt with the line. By properly proportioning the system, the impedance offered to the line *ab* at the point *b* can be made equal to the characteristic impedance of the line. The length of the stub and its proper location with respect to a current maximum are given by Fig. 50 in terms of the ratio of maximum to minimum current (or voltage) existing on the line before the stub is added.[1]

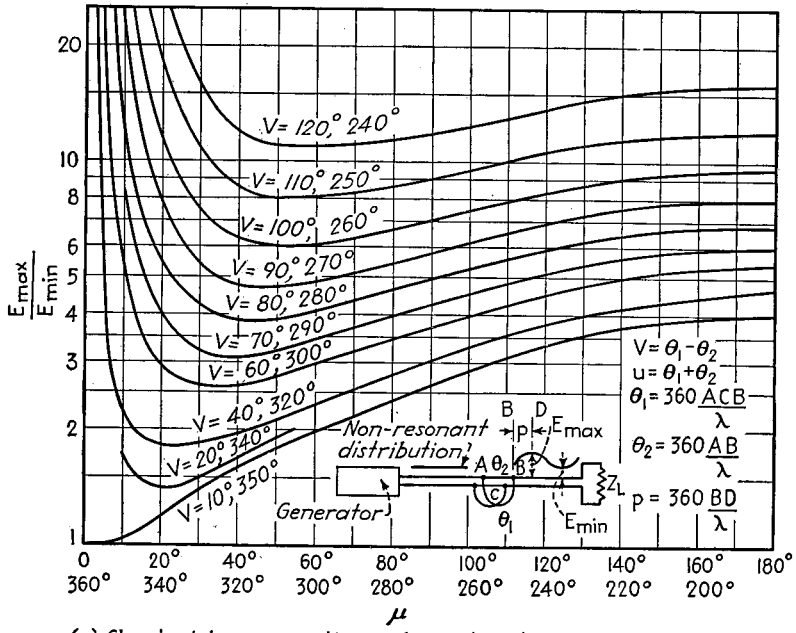


FIG. 51.—Chart for designing re-entrant sections to eliminate resonances on transmission lines. This chart gives the proportions of the re-entrant and its position with respect to a voltage maximum, in terms of the voltage maxima and minima existing on the line before the addition of the re-entrant.

Re-entrant Transmission-line Sections.[2]—This arrangement is illustrated in Fig. 49*f*, and employs an auxiliary line with both ends connected to the actual transmission line as shown. This "re-entrant" section produces an effect very similar to that of a stub line in that it can be made to introduce a compensating irregularity that causes the line on the generator side of the re-entrant to be terminated in an imped-

[1] From Sterba and Feldman, *loc. cit.*

[2] A. Alford, High-frequency Transmission Line Networks, *Elec. Comm.*, Vol. 17, p. 301, January, 1939.

ance equal to the characteristic impedance of the line. The design of re-entrant sections can be carried out with the aid of Fig. 51. It will be noted that in any particular case there are an infinite number of combinations that will give a particular result.

16. Transmission Lines as Resonant Circuits.[1]—A transmission line can be used to perform the functions of a resonant circuit when properly operated. Thus a quarter wave-length line open-circuited at the receiver has a receiving-end voltage much higher than the applied voltage, and a quarter-wave line short-circuited at the receiver offers a high impedance to an applied voltage similar to the resonant impedance of a parallel circuit. Resonant lines are frequently used in place of resonant circuits at high frequencies because of their superior performance.

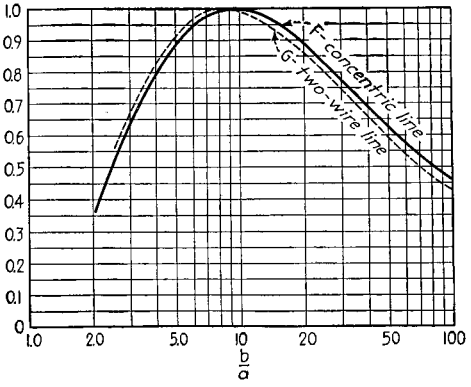


FIG. 52.—Factors F and G for use in Eqs. (66) and (67).

Lines as Parallel Resonant Circuits. A line offers a high impedance at the sending end when the receiver is short-circuited and the length is an odd multiple of a quarter wave length, or when the receiver is open-circuited and the line is an even number of quarter wave lengths long. In either case

$$Z_s = \frac{8Z_0^2 f}{Rnc} \tag{65}$$

where Z_s = sending-end impedance of line at resonance.
Z_0 = characteristic impedance of line.
f = frequency in cycles.
R = line resistance per unit length.
n = number of quarter wave lengths in line.
c = velocity of light in same units of length as R.

At high frequencies with air insulation and copper conductors, Eq. (65) becomes for the case of the concentric line

$$Z_s = \frac{11.11\sqrt{f}\,b}{n} F \tag{66}$$

where F = factor given by Fig. 52.
b = inner radius of outer conductor, cm.
a = outer radius of inner conductor, cm.

With a two-wire line (neglecting radiation resistance)

$$Z_s = \frac{23.95\sqrt{f}\,bG}{n} \tag{67}$$

where G = factor given in Fig. 52, which is a function of b/a.
b = spacing of wire centers, cm.
a = radius of wire, cm.

The curve of resonant-line impedance as a function of frequency in the region near resonance has the same shape as a resonance curve, with an equivalent Q given by the

[1] For further information see F. E. Terman, Resonant Lines in Radio Circuits, *Elec. Eng.*, Vol. 53, p. 1046, July, 1934.

formula

$$Q = \frac{2\pi Z_0 f}{Rc} \tag{68}$$

For copper lines with air insulation at radio frequencies, this equation becomes

For concentric lines:

$$Q = 0.0839\sqrt{f}\,bH \tag{69}$$

For two-wire lines (neglecting radiation resistance):

$$Q = 0.0887\sqrt{f}\,bJ \tag{70}$$

where H and J are factors given in Fig. 53 and the remaining notation is as before. Curves giving the Q of concentric lines of optimum proportions ($b/a = 3.6$) are given in Fig. 54.

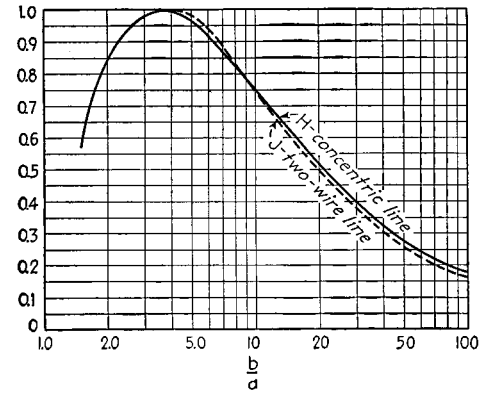


FIG. 53.—Factors H and J for use in Eqs. (69) and (70).

Resonant Lines as Series Circuits. A transmission line acts as a voltage step-up device when the receiver is open-circuited and the length is an odd number of quarter wave lengths. At frequencies near resonance, the step-up varies with frequency in exactly the same manner as a resonance curve having an equivalent Q given by Eq. (68), and the voltage step-up ratio is $Q \times 4/\pi n$ instead of Q, as in the case of the ordinary series resonant circuit.

Lines as Low-loss Reactances.—A line can serve as a low-loss inductance or capacity by employing the proper combination of length, frequency, and terminating condition. Thus a line short-circuited at the receiver will offer an inductive reactance when less than a quarter wave length long and a capacitive reactance when between a quarter and a half wave length long. With an open-circuited receiver, the conditions for inductive and capacitive reactance are interchanged. The reactances obtained are as follows:

For short-circuited receiver:

$$X_s = jZ_0 \tan\left(2\pi\frac{l}{\lambda}\right) \tag{71}$$

For open-circuited receiver:

$$X_s = -j\frac{Z_0}{\tan\left(2\pi\frac{l}{\lambda}\right)} \tag{72}$$

where l/λ is the line length measured in wave lengths and Z_0 is the characteristic impedance.

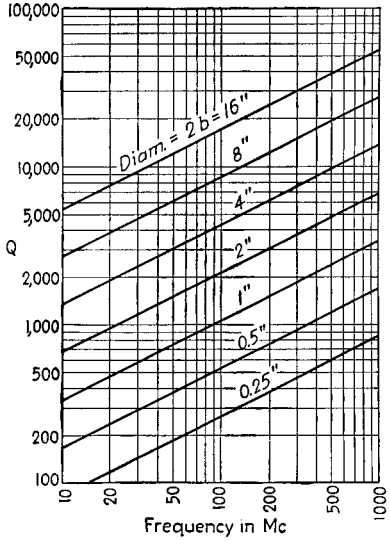


FIG. 54.—Q's of resonant concentric lines of optimum proportions ($b/a = 3.6$).

The Q of the reactance is related to the Q of the resonance curve of the corresponding quarter-wave line by the equation

$$\frac{Q \text{ of reactance}}{Q \text{ of resonant quarter-wave line}} = \frac{\sin\left(4\pi\frac{l}{\lambda}\right)}{2\pi\frac{l}{\lambda}} \tag{73}$$

It will be noted that the Q of the reactance decreases with l/λ, so that for high Q the line should be less than a quarter of a wave length long.

Reactances obtained from resonant lines differ from ordinary reactances in that their impedance varies more rapidly with frequency than is the case with an inductance or a capacity.

17. Miscellaneous Characteristics. *Irregularities.*—An irregularity in a transmission line produces a reflected wave having a magnitude determined by the extent of the irregularity. Resonances are then produced on the generator side of the irregularity, even though the line is properly terminated, as shown in Fig. 55. Irregularities should generally be avoided, for they give rise to unexpected resonances, often produce unbalance in the transmission line, and in general make the behavior different from that anticipated.

Irregularities are introduced whenever the constants of the line are modified. Typical causes are leads, bends in which the two sides of the circuit are not symmetrical, poor joints, etc.

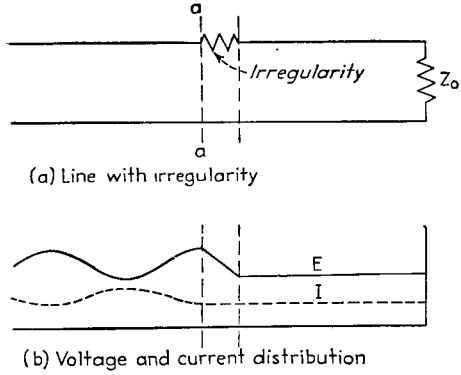


FIG. 55.—Resonances produced on a transmission line by a typical irregularity.

Balanced and Unbalanced Transmission Lines.—A line is said to be balanced when its two sides are symmetrical with respect to ground. Thus a horizontal two-wire line, and four-wire lines, are balanced structures, while a two-wire line in vertical configuration and a concentric line are unbalanced structures. Except for the concentric line, in which the outer conductor acts as a shield, unbalance causes earth currents to flow, because part of the current that should be carried by the conductor having the largest capacity to earth is diverted to the earth. This causes the current in the two sides of the transmission line to be unequal, and results in a component of current that flows out along the two transmission lines in parallel, and returns through the earth. Such unbalanced currents are to be avoided, because they give rise to extra energy loss, cause the radiation from the system to be greatly increased, and in general serve no useful purpose.

Radiation from Transmission Lines.—All transmission lines, except those of the concentric type, radiate some energy. Such radiation is often of importance, since it represents an additional energy loss, and in the case of transmission lines used in association with directional antennas, may seriously modify the directional pattern of the complete system.

The radiation from a two-wire nonresonant line is given approximately by the following formula,[1] provided that the length is at least twenty times the spacing and the spacing is not greater than one-tenth of a wave length and the line is nonresonant:

$$\frac{\text{Radiated power}}{I^2} = 160\left(\frac{\pi D}{\lambda}\right)^2 \tag{74}$$

[1] From Sterba and Feldman, *loc. cit.*

where D/λ is the spacing in wave lengths, and I is the rms line current. This radiation is twice that resulting from a doublet antenna carrying the same current as the line and having a length equal to the line spacing. In addition to the radiation given by Eq. (74), the terminating connections also produce radiation, so that the total radiation from the line with its terminations will be approximately four times the power radiated from a doublet having a length equal to the line spacing, and carrying the line current.

In the case of resonant lines the amount of energy radiated in proportion to the power transmitted to the load will be somewhat greater than is the case with the nonresonant line. However, unless the reflection coefficient of the load departs greatly from unity, the increase will not be great.

Unbalanced currents are much more effective in producing radiation than the normal balanced currents because of the considerable distance between the two sides of the circuit through which the unbalanced currents flow. The amount of radiation depends both on the height of the line above ground and on the line length. The order of magnitude of the factors involved can be estimated with the aid of Fig. 56.

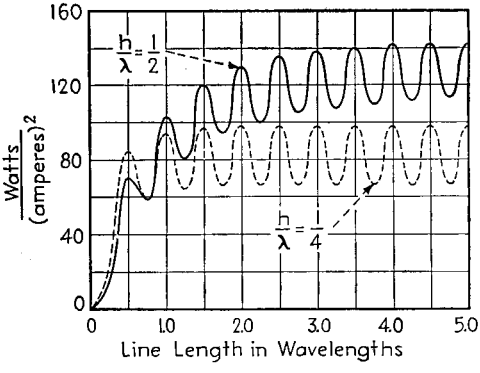


FIG. 56.—Radiation produced by unbalanced current in a transmission line.

When a two-wire line is used as a resonant circuit, radiation is quite important and may materially reduce the effective Q and parallel impedance obtained.[1]

Artificial Lines.—An artificial line is a four-terminal network composed of resistance, inductance, and capacity elements so proportioned that as far as the terminals are concerned the artificial line behaves in exactly the same way as an actual line. Artificial lines are usually in the form of symmetrical π or T networks, in which case the impedances composing the artificial line are related to the characteristic impedance Z_0 and hyperbolic angle $\sqrt{ZY}l$ of the corresponding real line, as indicated in Fig. 57*b*.[2]

Any given symmetrical T or π section is equivalent to some actual transmission line. The hyperbolic angle θ and characteristic impedance Z_0 of the real line equivalent to a given artificial line are given by the equations

For π or T line:

$$\cosh \theta = 1 + \frac{Z_1}{2Z_2} \tag{75a}$$

or

$$\sinh \frac{\theta}{2} = \sqrt{\frac{Z_1}{4Z_2}} \tag{75b}$$

For T line:

$$Z_0 = Z_T = \sqrt{Z_1 Z_2 + \tfrac{1}{4}Z_1^2} \tag{76}$$

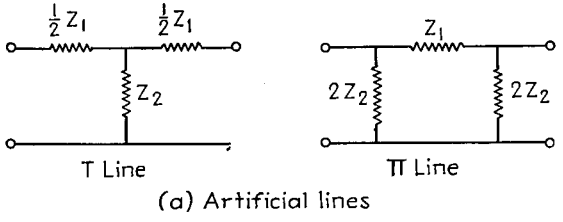
[1] Radiation is not present in concentric lines, provided that the outer conductor is in the form of a complete surface with closed ends. This is because the outer conductor acts as a nearly perfect shield for the fields existing within such a system.

[2] The relations given in this figure sometimes call for negative resistance elements in the artificial line, which is then physically unrealizable. This situation is likely to occur when the length of line being represented by the artificial section exceeds a quarter wave length, and can be handled by dividing the actual line into several shorter parts and representing each part by a separate artificial line.

For π line.

$$Z_0 = Z_\pi = \frac{Z_1 Z_2}{\sqrt{Z_1 Z_2 + \frac{1}{4} Z_1^2}} \tag{77}$$

In these equations, Z_1 and Z_2 have the significance shown in Fig. 57*a*.



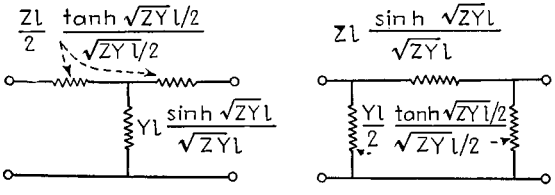


FIG. 57.—Artificial T and π lines for representing real transmission lines.

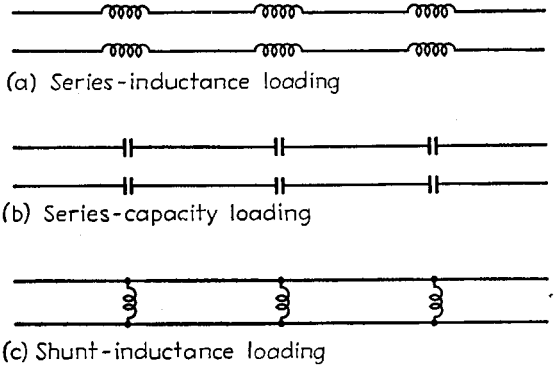


FIG. 58.—Transmission lines with various types of reactive loading.

Loaded Lines.—A loaded line is an ordinary transmission line to which lumped elements, usually capacities or inductances, are added at regular intervals. The most common use of loading is in telephone lines, particularly telephone cables, where inductance coils are commonly added at regular intervals, as in Fig. 58*a*. Such inductive loading makes the equivalent inductance per unit length of the transmission line greater than the actual inductance of the unloaded line at voice frequencies. The result is an increase in the characteristic impedance, a lowering of the phase velocity,

and also, when most of the loss is in the series resistance of the line, a reduction of the attenuation constant.

Loading by the use of series capacity, or shunt inductance, as illustrated in Fig. 58, is also sometimes employed in transmission lines used in radio work. The first of these arrangements reduces the inductance per unit length of the line, and the second reduces the equivalent shunting capacity. Both arrangements reduce the imaginary part of $\sqrt{ZY}$ and so reduce β. In this way, the phase velocity can be made to exceed the velocity of light and even approach infinity.

In order that lumped loading impedances may have the same effect as though their impedance were uniformly distributed, it is necessary that the loading impedances be spaced at distances that do not appreciably exceed a quarter wave length. If the

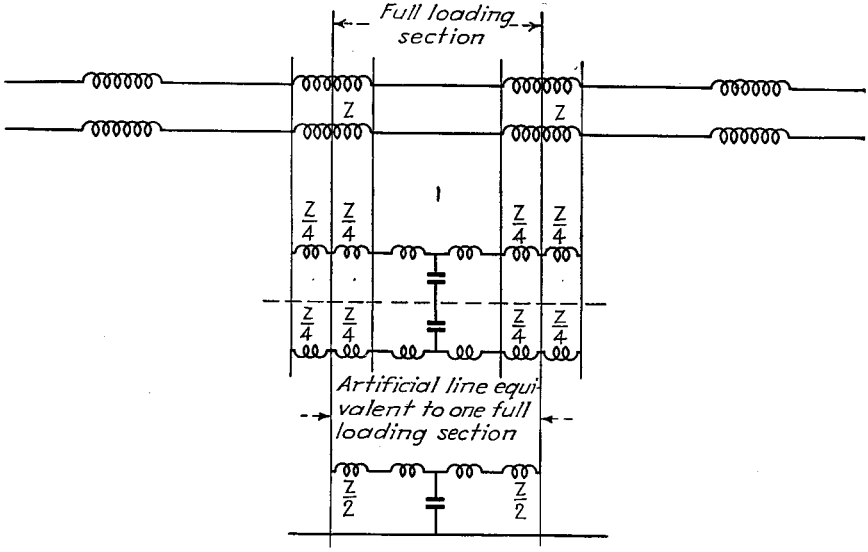


FIG. 59.—Transmission line with inductive loading, showing steps involved in determining equivalent characteristic impedance and hyperbolic angle of the loaded line.

spacing is greater, the loading impedances tend to act as irregularities or chokes and interfere with transmission.[1]

Exponentially Tapered Lines.[2]—An exponentially tapered line is one in which the spacing between conductors varies in such a way that the inductance and capacity vary exponentially with distance. Such a line acts as a combined high-pass filter and impedance-transforming device; *i.e.*, it transmits all frequencies above a certain critical or cut-off frequency with little or no attenuation (provided that the line losses are low or negligible) while attenuating lower frequencies, and at the same time has different characteristic impedances when viewed from the two ends.

[1] The effect of loading on a line can be analyzed exactly as follows: Referring to Fig. 59, the section of line between the loading points is first reduced to an equivalent artificial section. This section is then merged with the loading impedance as shown, thereby giving an artificial section that is equivalent to the actual line plus the loading corresponding to one loading section. The hyperbolic angle and characteristic impedance of this new artificial section are then the values applying to the actual loaded line per loading section.

[2] For further information see Harold A. Wheeler, Transmission Lines with Exponential Taper, *Proc. I.R.E.*, Vol. 27, p. 65, January, 1939. Charles R. Burrows, Exponential Transmission Line, *Bell System Tech. Jour.*, Vol. 17, p. 555, October, 1938. An excellent elementary summary of the latter paper is to be found in the *Bell Lab. Rec.*, Vol. 18, February, 1940.

The cut-off frequency depends upon the rate of taper, and is lower the more gradual the taper, in accordance with the relation

$$\frac{L_2}{L_1} = \frac{C_1}{C_2} = \epsilon^{\left(4\pi \frac{l}{\lambda_c}\right)} \tag{78}$$

where L_1 and C_1 are the inductance and the capacity per unit length at the closely spaced end, L_2 and C_2 are the corresponding values at the wide end, l is the length of the line, and λ_c the distance corresponding to a wave length at the cut-off frequency (in the same units of length as l), calculated on the basis of the constants at the narrow end.

The ratio of impedance transformation is approximately

$$\text{Ratio of impedance transformation} = \sqrt{\frac{Z_{02}}{Z_{01}}} = \epsilon^{2\pi \frac{l}{\lambda_c}} \tag{79}$$

where Z_{02} is the characteristic impedance on the basis of the spacing at the wide end of the line, Z_{01} is the characteristic impedance based on the spacing at the other end of the line, and l and λ_c have the same meaning as above. The ratio of impedance transformation accordingly depends upon the difference in spacing at the ends of the line, *i.e.*, upon the total taper, and can be easily made as great as two to one. The impedance match obtained at the ends of the exponential line may be improved by adding terminal reactances of complexity determined by the extent the reflection coefficient at the terminals must be minimized over a frequency range.[1]

NETWORK THEORY, FILTERS, AND EQUALIZERS

18. Network Definitions.—A network is made up of resistances, inductances, capacities, and mutual inductances connected together in some manner. The resistances, inductances, mutual inductances, and capacities involved are termed *network constants* or *parameters*. When these parameters are constant, independent of the current going through them, the network is said to be *linear*.

A typical network is illustrated in Fig. 60. The junctions *a*, *b*, *c*, etc., at which the current can divide are termed *branch points*, and the sections of the network between branch points are termed *branches*. A series of branches that form a complete loop is termed a *mesh*. Examples of meshes are shown by the arrows in Fig. 60, where five separate meshes are designated.

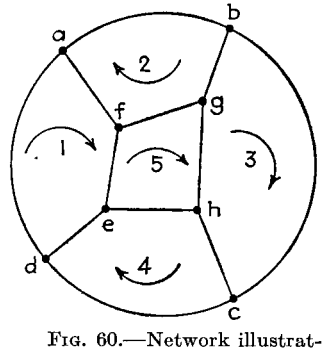


FIG. 60.—Network illustrating meshes, branches, and branch points.

A *passive network* is a network containing no source of energy, in which no energy is dissipated other than that accounted for by the resistance elements of the network. An *active network* is a network containing one or more sources of energy, or some sink of energy.

The term *two-terminal* is applied to networks operated under the conditions shown in Fig. 61*a*. Here the only applied voltage is that shown acting between terminals 1-2 obtained by opening some branch of the network. The term *four-terminal* is applied to a network operating under the conditions illustrated in Fig. 61*b*. Here a voltage E is applied in series with one branch of the network by opening up this branch to form terminals 1-2, while another branch of the network is opened up to form terminals 3-4, between which an output or load impedance is inserted. A four-terminal network is equivalent to a transmission system, and is sometimes referred to as a *transducer*.

[1] An excellent discussion of this subject is given by Wheeler, *loc. cit.*

19. Network Theorems. *Superposition Theorem.—The current that flows in a linear network, or the potential difference that exists between any two points in such a network, resulting from the simultaneous application of a number of voltages distributed in any manner whatsoever throughout the network is the sum of the component currents at the first point (or the component potential differences between the two points) that would be caused by the individual voltages acting separately.*

Reciprocity Theorem.—In any network composed of linear impedances, if an electromotive force E applied between two terminals produces a current I at some branch in the network, then the same voltage E acting at the second point in the circuit, will produce the same current I at the first point.

Thévenin's Theorem.—Any linear network containing one or more sources of voltage and having two terminals behaves, insofar as a load impedance connected across the terminals is concerned, as though the network and its generators were equivalent to a simple generator having an internal impedance Z and a generated voltage E, where E is the voltage that appears across the terminals when no load impedance is connected and Z is the impedance that is measured between the terminals when all sources of voltage in the network are short-circuited.[1]

Compensation Theorem.—If an impedance ΔZ is inserted in a branch of a network, the resulting current increment produced at any point in the network is equal to the current

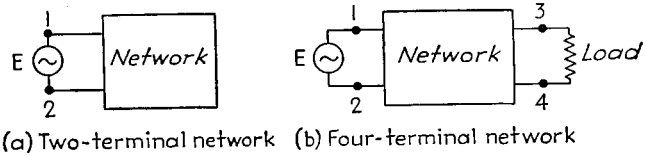


FIG. 61.—Two- and four-terminal networks.

that would be produced at that point by a compensating voltage acting in series with the modified branch, whose value is $-I\Delta Z$, where I is the original current that flowed where the impedance was inserted before the insertion was made.

20. General Mesh Equations of a Network and Their Solution.[2]—The response of a network to an applied voltage can be most readily expressed in terms of the mesh currents, *i.e.*, the currents that can be considered as circulating around the closed meshes, as indicated in Fig. 60. The branch currents can then be obtained as the vector sum of the various mesh currents that flow through the branch in question.

Mesh Equations.—The voltage and current relations in a network containing n independent meshes can be written as

$$\left.\begin{aligned} Z_{11}I_1 + Z_{12}I_2 + Z_{13}I_3 + \ldots . Z_{1n}I_n &= E_1 \\ Z_{21}I_1 + Z_{22}I_2 + Z_{23}I_3 + \ldots . Z_{2n}I_n &= E_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ Z_{n1}I_n + Z_{n2}I_n + Z_{n3}I_n + \ldots . Z_{nn}I_n &= E_n \end{aligned}\right\} \qquad (80)$$

In these equations, I_1, I_2, etc., designate the individual mesh currents; E_1, E_2, etc., represent the vector sums of the applied voltages acting around the individual meshes numbered 1, 2, etc. Z_{11}, Z_{22}, Z_{33}, etc., represent the self-impedances of individual meshes, *i.e.*, the impedance around the mesh if all other branches of the network other than those included in the mesh in question were open-circuited. The impedance Z_{12}

[1] When the sources of energy in the network are constant-current generators, instead of constant-voltage generators, the internal impedance Z is the impedance observed between the terminals when all constant-current generators are open-circuited.

[2] An excellent discussion of the general mesh equations of a network is given in Chap. IV of E. Guillemin, "Communication Networks," Vol. I, Wiley, New York, 1931.

represents the mutual impedance (coupling) between meshes 1 and 2, as a result of which, a current in mesh 2 produces a voltage drop in mesh 1, and Z_{13} represents the mutual impedance whereby current in mesh 3 produces voltage drop in mesh 1, etc. It is to be noted that reversing the order of subscripts of the Z's does not alter the value of mutual impedance. Thus $Z_{12} = Z_{21}$. Coupling may result either through mutual inductance or from impedance elements common to the two meshes.

In setting up the system of Eq. (80), care must be taken to be consistent in the matter of signs. The positive directions for the mesh currents are assigned arbitrarily. An impedance that is common to two branches is then considered to be a positive mutual impedance when the arrows representing the corresponding mesh current pass through the impedance in the same direction. If the arrows indicate that the corresponding mesh currents pass through the common impedance in opposite directions, then the mutual impedance is the negative of this common impedance. A mutual inductance is positive or negative according to whether it acts with a polarity the same as or opposite to that of a corresponding common inductance.

In setting up a system of relations such as is represented in Eq. (80), it is possible to designate the meshes in a variety of ways, subject only to the limitation that each branch of the network must be included in at least one mesh. For example, mesh 1 in Fig. 60, instead of following the configuration *afeda*, could have been defined as *abgfeda*. This would have modified the details in Eq. (80), but would have resulted in the same individual branch currents.

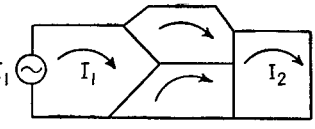


FIG. 62.—Network for illustrating meanings of driving point and transfer impedances.

In setting up mesh equations and selecting the meshes, there sometimes may be a question as to the number of independent meshes present. In such cases one can use the relation

$$\left.\begin{matrix}\text{Number of independent}\\ \text{meshes}\end{matrix}\right\} = \left.\begin{matrix}\text{number of}\\ \text{branches}\end{matrix}\right\} - \left.\begin{matrix}\text{number of}\\ \text{branch points}\end{matrix}\right\} + 1 \qquad (81)$$

A solution of the system of Eq. (80) shows that the current I_k in the kth mesh that flows as the result of the voltage E_j acting in the jth mesh is

$$I_k = E_j \frac{B_{jk}}{D} \qquad (82)$$

Here D is the determinant of the system of Eq. (80), and is given by

$$D = \begin{vmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix} \qquad (83)$$

Methods of evaluating a determinant (as well as its more important properties) are found in mathematical texts. The quantity B_{jk} in Eq. (82) is the principal minor of D and is formed by canceling the jth row and kth column and then moving the remainder together to form a new determinant with one less row and column than D. In evaluating B_{jk}, this new determinant is prefixed with the sign $(-1)^{j+k}$.

Input and Transfer Impedance.—Consider a network having a single applied voltage, with the meshes so arranged that this voltage acts in a branch that is part only of a single mesh, as in Fig. 62. The impedance that the network offers to this applied voltage, *i.e.*, the ratio E_1/I_1, is termed the *input impedance* or *driving-point impedance*

of the network. From Eq. (82) this input impedance can be expressed as

$$\text{Input impedance} = \frac{D}{B_{11}} \tag{84}$$

In an analogous manner the *transfer impedance* is defined as the ratio of the voltage E_1 applied in mesh 1 to the resulting current I_2 of mesh 2, as indicated in Fig. 62. This transfer impedance can be expressed as

$$\left.\begin{matrix}\text{Transfer impedance from} \\ \text{mesh 1 to mesh 2}\end{matrix}\right\} = \frac{D}{B_{12}} \tag{85}$$

where the meshes are so selected that I_2 is observed in a branch that is in only the second mesh of the network and the branch in which the voltage is applied is contained only in the first mesh of the system.

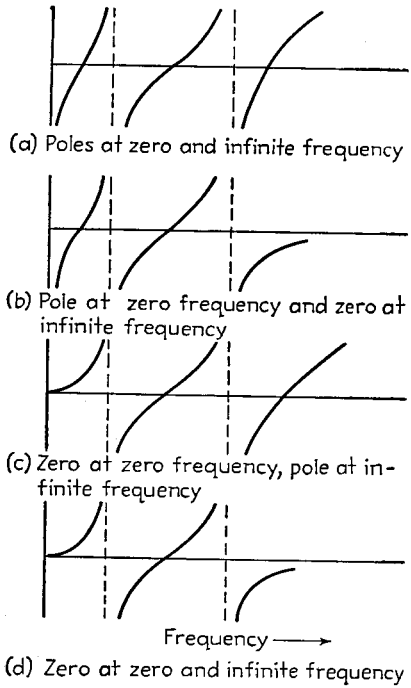


FIG. 63.—Reactance as a function of frequency for various classes of reactive networks.

21. Two-terminal Reactive Networks. The characteristics of a two-terminal network composed of ideal reactances having zero losses are important, because such networks approximate very closely the reactive arms that are used to build up filters, impedance matching networks, etc.

Foster's Reactance Theorem.[1]—The driving-point impedance of a two-terminal reactive network behaves as shown in Fig. 63. The impedance curve consists of segments going from minus infinity to plus infinity (except possibly at zero and infinite frequency where a segment may start or stop, respectively, at zero impedance). The slope of the curve is everywhere positive, and is greater than the slope of a straight line drawn to the origin. The frequencies at which the impedance is infinity are termed *poles*, and the frequencies at which the impedance is zero are termed *zeros*.

Foster has shown that the driving-point impedance of a reactive network is uniquely specified by the location of the internal zeros and poles, plus one additional piece of information.[2] Expressed analytically, the reactance function can be written as follows:

For a pole at the origin (Figs. 63*a* and 63*b*):

$$\left.\begin{matrix}\text{Driving point} \\ \text{impedance}\end{matrix}\right\} = Z = \pm j\frac{H}{\omega}\frac{(\omega^2-\omega_1^2)(\omega^2-\omega_3^2)\ \ldots\ (\omega^2-\omega_p^2)}{(\omega^2-\omega_2^2)(\omega^2-\omega_4^2)\ \ldots\ (\omega^2-\omega_q^2)} \tag{86a}$$

For a zero at the origin (Figs. 63*c* and 63*d*):

$$\left.\begin{matrix}\text{Driving point} \\ \text{impedance}\end{matrix}\right\} = Z = \pm j\omega H\frac{(\omega^2-\omega_1^2)(\omega^2-\omega_3^2)\ \ldots\ (\omega^2-\omega_p^2)}{(\omega^2-\omega_2^2)(\omega^2-\omega_4^2)\ \ldots\ (\omega^2-\omega_q^2)} \tag{86b}$$

[1] Ronald M. Foster, A Reactance Theorem, *Bell System Tech. Jour.*, Vol. 3, p. 259, April, 1924. See also Chap. IV of E. Guillemin, "Communication Networks," Vol. II, Wiley, New York, 1935.

[2] Poles or zeros at the origin, or at infinity, are referred to as external, and play no part in the specification of the reactance function.

where the angular velocities $\omega_1, \omega_3, \ldots, \omega_p$ designated by odd subscripts correspond to the internal zeros of the reactance function, and the angular velocities $\omega_2, \omega_4 \ldots, \omega_q$ designated by even subscripts, correspond to the internal poles of the reactance (see Fig. 63). The plus sign applies when there is a pole at infinite frequency, while the minus sign applies with a zero at infinite frequency. The sum of the number of poles and number of zeros is one less than the number of independent meshes of the network. The quantity H is a positive real constant that takes into account the fact that one additional piece of information is required to complete the specification of the reactance function.

Foster's reactance theorem shows that the impedance characteristics obtainable from a physically realizable reactance network are quite restricted. This is important, because it limits the characteristic obtainable from filters and other networks.

Synthesis of Reactive Two-terminal Networks.—Any driving-point reactance characteristic that can be obtained from any conceivable two-terminal reactive network can be realized by either one of the two networks shown in Fig. 64. The first of these

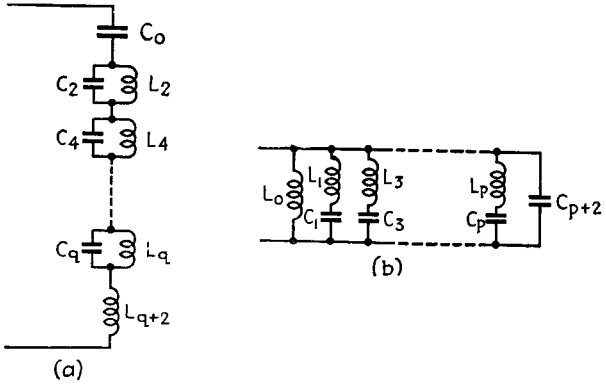


FIG. 64.—General forms of reactive networks.

consists of parallel resonant circuits connected in series, with one parallel resonant circuit corresponding to each internal pole. The series condenser C_0 is omitted in the event that the network has a zero at the origin, while the series inductance L_{q+2} is omitted if the network has a zero at infinity. The magnitudes of the circuit components required in the equivalent network of Fig. 64*a* to realize a desired impedance characteristic are given by the following relations:

$$C_k = \left|\frac{j\omega_k}{Z_k}\right| \qquad (k = 2, 4, \ldots, q) \tag{87}$$

where Z_k is the quantity obtained by omitting the term $(\omega^2 - \omega_k^2)$ from the denominator of the corresponding expression (86) for Z and evaluating the modified expression for Z with $\omega = \omega_k$. Corresponding to each C_k, one has

$$L_k = \frac{1}{\omega_k^2 C_k} \tag{88}$$

If the network has a pole at infinity, then

$$L_{q+2} = H \tag{89}$$

If the network has a pole at zero frequency

$$C_0 = \left|\frac{1}{Z_0}\right| \tag{90}$$

where Z_0 is the quantity obtained by omitting the ω in the denominator of Eq. (86*a*) under H and evaluating the modified expression for Z with $\omega = 0$.

An alternative method of synthesizing any desired impedance characteristic is to use the arrangement of Fig. 64*b*, in which the two-terminal network is built up of a number of series resonant circuits connected in parallel, with one resonant circuit for each internal zero. In the event that the poles and zeros are so arranged that the network has a pole at the origin, inductance L_0 is omitted. Similarly, if there is a pole at infinite frequency, capacity C_{p+2} is omitted. The values of the elements in the circuit of Fig. 64*b* required to give a reactance function corresponding to specified zeros, poles, and the given value of H, is obtained from the following equation which can be deduced by a partial fraction expansion of $1/Z_{11}$ as given by Eq. (86).

$$L_k = \left| j\omega_k Z_k' \right|, \qquad (k = 1, 3, \ldots, p) \tag{91}$$

where Z_k' is the quantity obtained by omitting $(\omega^2 - \omega_k^2)$ from the numerator of expression (86) for Z and evaluating the modified expression for $\omega = \omega_k$. Corresponding to each L_k, one has

$$C_k = \frac{1}{\omega^2 L_k} \tag{92}$$

If the network has a zero at infinite frequency

$$C_{p+2} = \frac{1}{H} \tag{93}$$

If the network has a zero at zero frequency

$$L_0 = Z_0' \tag{94}$$

Where Z_0' is the quantity obtained by omitting the ω that multiplies H in expression (86*b*) and then evaluating the modified expression for Z with $\omega = 0$.

The networks of Fig. 64 represent networks having the least possible number of reactive elements that can be used to realize a specified impedance characteristic. The same impedance characteristic may also be realized by many other networks other than the two shown,[1] but alternative arrangements will in most cases have additional circuit elements that are superfluous. The least number of circuit elements required is one more than the sum of the internal poles and zeros. One, or even more than one, element beyond this minimum is possible for a given number of internal zeros and poles, but such networks have exactly the same impedance characteristic as the simplest forms, such as given in Fig. 64, which are called fundamental or canonic forms because they have the least possible number of elements.

22. Inverse or Reciprocal Impedances.[2]—Two impedances Z_1 and Z_2 are said to be reciprocal with respect to an impedance Z if they are so related as to satisfy the relation

$$Z_1 Z_2 = Z^2 \tag{95}$$

Under practical conditions where reciprocal impedances are of importance, the impedance Z in Eq. (95) is always a resistance.

The process of deriving a reciprocal impedance from a given impedance is termed *reciprocation*. Reciprocation of a ladder network consisting of alternate series and shunt arms, as in Fig. 65*a*, leads to a corresponding ladder network of alternate shunt and series arms, as shown in Fig. 65*b*.

[1] Ladder networks starting with either a series impedance or a shunt impedance also can be used to develop a specified two-terminal impedance utilizing the minimum possible number of circuit elements. The circuit constants for such networks are obtained from continued fraction expansion of the reactance function. For further information, see Guillemin, *op. cit.*, pp. 198–207.

[2] See Guillemin, *op. cit.* p. 204; also, A. C. Bartlett, "Theory of Electrical Artificial Lines and Filters," pp. 53–58, Wiley, New York, 1931.

Various special cases of reciprocation are shown in Fig. 66. These assume that reciprocation is with respect to a resistance R, *i.e.*, Z in Eq. (95) is taken as R. It will

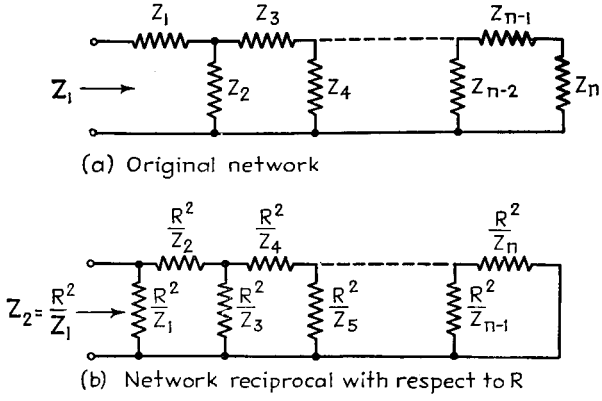


FIG. 65.—Reciprocal ladder networks.

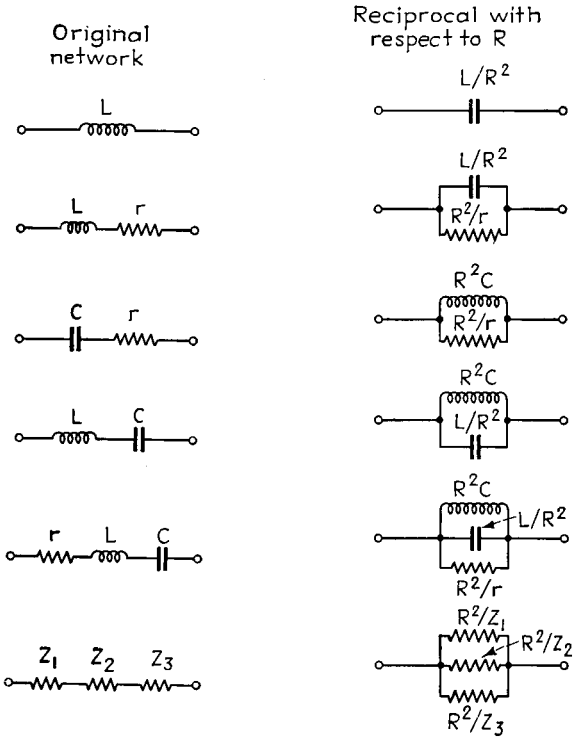


FIG. 66.—Basic examples of reciprocation.

be noted that the reciprocal of a number of impedance elements in series consists of a number of impedance elements in shunt, with each shunt element being the reciprocal of one of the series elements, and vice versa. It is also to be noted that when two

reactive networks are reciprocal, the poles of one coincide with the zeros of the other impedance, and vice versa.

23. Fundamental Relations Existing in Four Terminal Networks.[1] *Methods of Expressing Network Characteristics.*—Insofar as the four terminals are concerned, the properties of a four-terminal network at any one frequency can be expressed in terms of any three independent properties of the network, irrespective of how complicated the network is. Although there are an unlimited number of ways in which three independent constants can be defined, the ones most commonly employed in communication networks are (1) image impedances and image transfer constant; (2) open- and short-circuit impedances; and (3) iterative impedances and iterative transfer constant.

Four-terminal Network Behavior Expressed in Terms of Image Impedances.—A network is said to be operated under image-impedance conditions when the internal impedance Z_s of the source of power acting on the input terminals of the network, and the load impedance Z_L at the output terminals, are so related to the network that the impedance looking into the network from the terminals 1-2 with the load connected

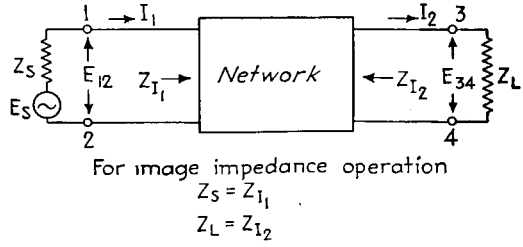


FIG. 67.—Four-terminal network operated with terminal impedances.

(see Fig. 67) is equal to the generator impedance, and, similarly, so that the impedance looking into the network at terminals 3-4 with the generator connected equals the load impedance. The generator and load impedances required to produce this condition are properties of the network, and are termed the *image impedances.* They can be designated by the symbol Z_{I_1} and Z_{I_2} for the input and output terminals, respectively (see Fig. 67).

In dealing with image impedances, the third independent property of the network required to finish the specification of the network behavior is taken as the *image transfer constant* θ, which is defined in terms of the relations

$$\frac{E_{34}}{E_{12}} = \sqrt{\frac{Z_{I_2}}{Z_{I_1}}}\epsilon^{-\theta}, \qquad \frac{I_2}{I_1} = \sqrt{\frac{Z_{I_1}}{Z_{I_2}}}\epsilon^{-\theta} \tag{96}$$

The notation is illustrated in Fig. 67. The image transfer constant θ has the same value irrespective of the direction of transmission of energy through the network.

The three network parameters Z_{I_1}, Z_{I_2}, and θ can be defined in terms of the network determinant and its minors, and also in terms of the open- and short-circuit impedance of the network, according to the equations

$$Z_{I_1} = \sqrt{\frac{DB_{22}}{B_{11}B_{1122}}} = \sqrt{Z_{oc}Z_{sc}} \tag{97}$$

$$Z_{I_2} = \sqrt{\frac{DB_{11}}{B_{22}B_{1122}}} = \sqrt{Z'_{oc}Z'_{sc}} \tag{98}$$

[1] Additional information on these subjects is given in the following books: K. S. Johnson, "Transmission Circuits for Telephonic Communication"; T. E. Shea, "Transmission Networks and Wave Filters," Van Nostrand, 1929; Guillemin, *op. cit.*

$$\tanh \theta = \sqrt{\frac{Z_{sc}}{Z_{oc}}} = \sqrt{\frac{Z'_{sc}}{Z'_{oc}}} = \sqrt{\frac{DB_{1122}}{B_{11}B_{22}}} \tag{99}$$

where Z_{oc} and Z_{sc} are the impedances at terminals 1-2 with terminals 3-4 open- and short-circuited, respectively, Z'_{oc} and Z'_{sc} are the impedances at 3-4 with 1-2 alternately open- and short-circuited, while D is the determinant for the network formed by short-circuiting terminals 1-2 and 3-4. The B's are minors of this determinant, with B_{1122} being the minor formed by striking out both first and second rows and first and second columns of the determinant.

Image-impedance operation of a network can be conveniently related to the behavior of a transmission line expressed in terms of wave trains. Although wave trains obviously cannot exist in a network having lumped constants, it is nevertheless frequently convenient to explain the behavior of a four-terminal system, insofar as the terminals are concerned, in terms of wave trains just as though these wave trains actually existed within the network instead of being hypothetical. The image impedances correspond to the characteristic impedance of the transmission line, but unless the network is symmetrical about its midpoint, there will be two image impedances because of the fact that the network can be considered as equivalent to a transmission line that is unsymmetrical and so has an impedance transforming action upon a

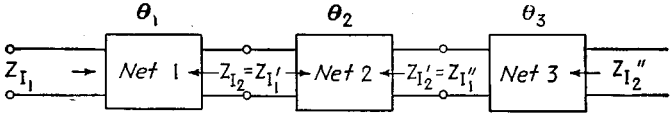


FIG. 68.—Networks connected in cascade on an image-impedance basis.

wave train. The image transfer constant θ of the network is likewise analogous to the hyperbolic angle of the transmission line. The real part of the image transfer constant is called the *attenuation constant*, and can be considered as causing the hypothetical wave train in the network to be attenuated in magnitude. The imaginary part of the image transfer constant corresponds to the phase constant of the transmission line hyperbolic angle, and causes a shift in phase of the hypothetical wave train. If the load impedance does not equal the image impedance on the output side of the network, the effect on the voltage and current relations is as though a wave train existed in the network, and was reflected by the load impedance, just as in the case of a transmission line.

When several networks are connected together in cascade on an image-impedance basis, as illustrated in Fig. 68, then the image impedances at the input and output terminals are the image impedances at the input terminals of the first network and the output terminals of the last network, respectively (Z_{I_1} and Z_{I_2}'' in Fig. 68). The image transfer constant of such a system is

$$\theta + \theta_1 + \theta_2 + \ldots + \theta_m \tag{100}$$

where θ_1, θ_2, etc., represent the image transfer constants of the first, second, etc., component networks of the system.

In a system consisting of a number of networks connected in cascade on an image-impedance basis, it is customary to refer to the image impedance existing at a particular junction point as the *impedance level* at that point.

The image-impedance method of expressing the properties of a four-terminal network is extremely important because most four-terminal networks used in communication systems are operated under conditions that approach very closely image-impedance operation. This is particularly true of filters, equalizers, and impedance matching networks.

Four-terminal Networks Operated on an Iterative-impedance Basis.—A network operated on an iterative-impedance basis requires that the load impedance be such that the input impedance of the network with the load connected is equal to the load impedance. At the same time, the impedance observed by looking into the network from the output terminals toward the generator with the internal impedance of the generator connected across the input terminals of the network must equal the generator impedance. These two impedances can be designated as Z_{k_1} and Z_{k_2}, respectively, and are properties of the network. The third property necessary to specify completely the network characteristics is then taken as the iterative transfer constant P, which is defined by the equation

$$\frac{I_2}{I_1} = \epsilon^{-P} \tag{101}$$

where I_1 and I_2 are the input and output currents of the network, respectively, when operated under iterative conditions. Iterative-impedance action becomes of importance in handling problems involving L and ladder types of attenuators.

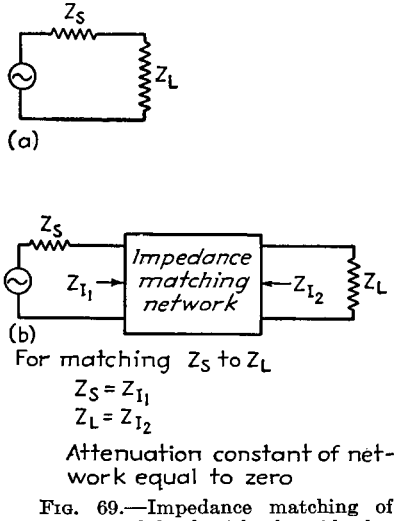


FIG. 69.—Impedance matching of generator and load with the aid of a network.

Impedance Matching.—Two impedances are said to be matched when they have the same magnitude and the same phase angle. Thus in Fig. 69*a*, the load impedance Z_L is said to be matched to the generator impedance Z_s if the load impedance equals the generator impedance. If the load and generator impedances are not matched, it is possible to obtain matching by inserting between generator and load, as shown in Fig. 69*b*, a network having $Z_{I_1} = Z_s$, $Z_{I_2} = Z_L$, and zero attenuation (*i.e.*, real part of image transfer constant θ must equal zero). The generator then sees a load impedance $Z_{I_1} = Z_s$, and the load receives its power from a source (the network output terminals) having an internal impedance $Z_{I_2} = Z_L$. Such a system has its impedances matched on an image basis at all junction points.

The ratio of load current that would be delivered by a particular generator to a particular load without matching, as in Fig. 69*a*, to the ratio of the same currents when the impedances are matched, as in Fig. 69*b*, is designated by such terms as *mismatching factor*, *reflection factor*, or *transition factor*. The absolute value of this ratio gives the loss of load current that results when no means are provided to couple the generator to the load impedance on an image-impedance basis. The value of the mismatching factor depends only upon the ratio of load to generator impedance, and is

$$\frac{\text{Load current without matching}}{\text{Load current with matching}} = k = \frac{\sqrt{\frac{4Z_s}{Z_L}}}{1 + \frac{Z_s}{Z_L}} \tag{102}$$

Expressed in decibels, the loss of load current resulting from mismatching is

$$\text{Mismatching loss in db} = 20 \log_{10} (1/k) \tag{103}$$

A chart giving loss from mismatching based on Eq. (103) is given in Fig. 70, where θ_s and θ_L are the phase angles of Z_s and Z_L, respectively.

When the generator impedance Z_s is a resistance, then image-impedance matching corresponds to the condition for which the power delivered by the generator to the load is maximum, and if the load impedance fails to match the generator impedance, a loss of load power results. However, if the generator impedance has a reactive component, then failure to match the load impedance to the generator impedance can, under certain conditions, result in an increase in the load current.[1] Under these conditions, the mismatching factor will be greater than unity.

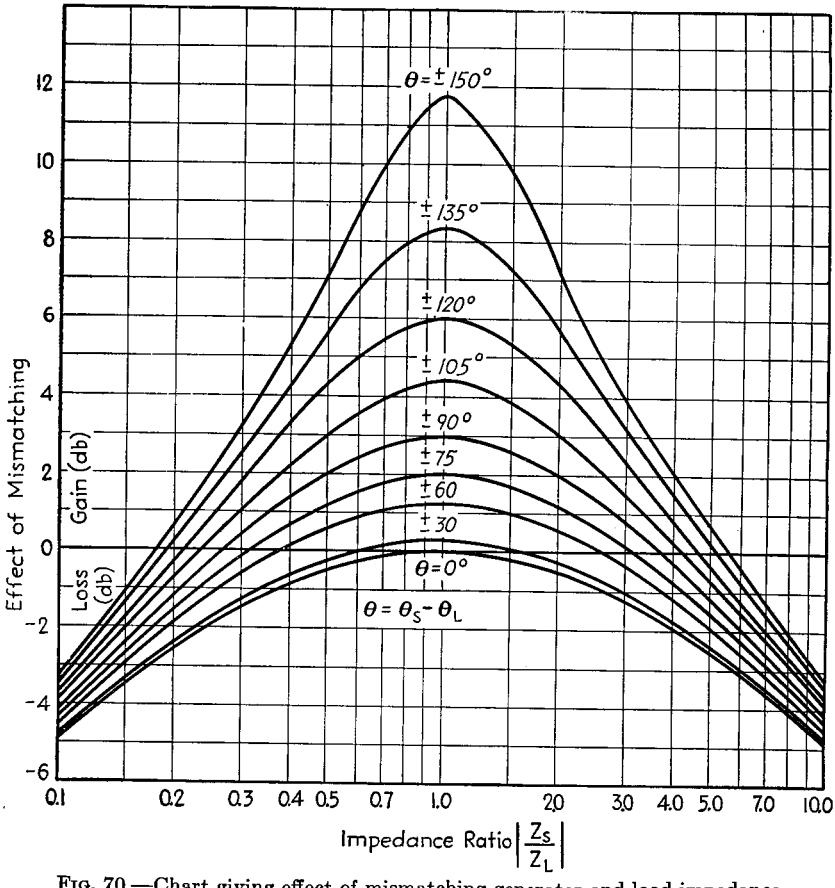


FIG. 70.—Chart giving effect of mismatching generator and load impedance.

Insertion Loss.—When a network is inserted between a sending-end impedance Z_s and a load impedance Z_L, as in Fig. 69*b*, the ratio of current in the load impedance when the network is present to the load current in the absence of the network is termed the *insertion loss*, since this is the loss in output current resulting from the insertion of the network between generator and load.

The insertion loss can be conveniently expressed by the formula

$$\text{Insertion loss} = \frac{k_1 k_2}{k} \sigma\epsilon^{-\theta} \tag{104}$$

[1] The condition for maximum possible transfer of energy to the load is realized when the resistance component of the load impedance is equal to the resistance component of the generator impedance and when at the same time the reactive component of the load impedance is equal in magnitude but opposite in sign to the reactive component of the generator impedance.

where k_1 = mismatching factor of Z_s and Z_{I_1}.
k_2 = mismatching factor of Z_{I_2} and Z_L.
k = mismatching factor of Z_s and Z_L.
θ = image transfer constant.

$$\sigma = \text{interaction factor} = \frac{1}{1 - \left(\frac{Z_{I_2} - Z_L}{Z_{I_2} + Z_L}\right)\left(\frac{Z_{I_1} - Z_s}{Z_{I_1} + Z_s}\right)\epsilon^{-2\theta}}.$$

The interaction factor σ is a second-order effect representing a modification of the insertion loss that occurs when there is mismatching at *both* the input and output terminals of the network. The interaction factor takes into account the effect of a wave that is reflected from the load back to the generator through the network, and there reflected back to the load. The interaction factor becomes unity whenever at least one end of the network is matched on an image-impedance basis, or when the network attenuation is such that a wave that has traveled through the network twice will have been reduced to negligible amplitude.

Under practical conditions, the interaction factor in networks intended to be operated on an image-impedance basis is at most only a few decibels.

24. Fundamental Types of Four-terminal Networks. *π and T Networks.*—The fact that any four-terminal network can have its properties represented, insofar as the

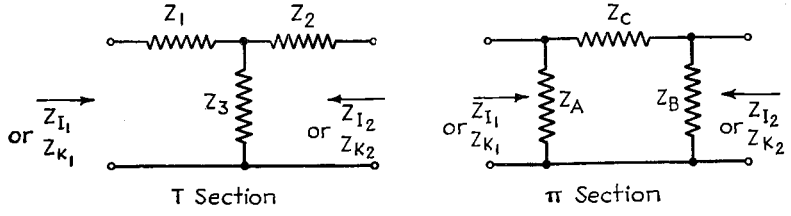


FIG. 71.—General T and π networks.

terminals are concerned, by three independent constants means that such networks can be always represented by three properly chosen independent impedances arranged in the form of a T or π, as shown in Fig. 71.

The relationship between the impedances composing such networks, and the characteristics of the system expressed on an image-impedance basis, are

For T section (see Fig. 71):

$$\left.\begin{aligned} Z_{I_1} &= \sqrt{(Z_1Z_2 + Z_1Z_3 + Z_2Z_3)\left(\frac{Z_1 + Z_3}{Z_2 + Z_3}\right)} \\ Z_{I_2} &= \sqrt{(Z_1Z_2 + Z_1Z_3 + Z_2Z_3)\left(\frac{Z_2 + Z_3}{Z_1 + Z_3}\right)} \\ \tanh\theta &= \sqrt{\frac{(Z_1Z_2 + Z_1Z_3 + Z_2Z_3)}{(Z_1 + Z_3)(Z_2 + Z_3)}} \end{aligned}\right\} \quad (105)$$

For π section (see Fig. 71):

$$\left.\begin{aligned} Z_{I_1} &= Z_A\sqrt{\frac{(Z_B + Z_C)}{(Z_A + Z_C)}\frac{Z_C}{Z_A + Z_B + Z_C}} \\ Z_{I_2} &= Z_B\sqrt{\frac{(Z_A + Z_C)}{(Z_B + Z_C)}\frac{Z_C}{Z_A + Z_B + Z_C}} \\ \tanh\theta &= \sqrt{\frac{Z_C(Z_A + Z_B + Z_C)}{(Z_A + Z_C)(Z_B + Z_C)}} \end{aligned}\right\} \quad (106)$$

In the design of T and π networks, one normally knows the desired image impedances, and wishes to realize a particular transfer constant θ. The relations are then

For T section (see Fig. 71):

$$\left.\begin{aligned} Z_3 &= \sqrt{Z_{I_1} Z_{I_2} \left(\frac{1}{\tanh^2 \theta} - 1\right)} \\ Z_2 &= \frac{Z_{I_2}}{\tanh \theta} - Z_3 \\ Z_1 &= \frac{Z_{I_1}}{\tanh \theta} - Z_3 \end{aligned}\right\} \tag{107}$$

For π section (see Fig. 71):

$$\left.\begin{aligned} Z_C &= \sqrt{Z_{I_1} Z_{I_2}} \sinh \theta \\ Z_B &= \frac{1}{\frac{1}{Z_{I_2} \tanh \theta} - \frac{1}{Z_C}} \\ Z_A &= \frac{1}{\frac{1}{Z_{I_1} \tanh \theta} - \frac{1}{Z_C}} \end{aligned}\right\} \tag{108}$$

Examination of these equations shows that in the case of networks with reactive elements, the image impedances are either pure resistances or pure reactances. Furthermore, when the image impedances are resistive, the image transfer constant is a pure imaginary, while when the image impedances are reactive, the image transfer constant has a real component, and so attenuation is introduced. These relations are particularly important in the case of filters and impedance-matching networks.

L Networks.—An L network is shown in Fig. 72, and can be considered as a special case of a T or π network in which one of the impedance arms has become either zero or infinity.

The properties of an L network can be expressed in terms of image impedances and a transfer constant, as in the case of any four-terminal network. However, since there are only two impedance arms in an L network, a relationship must exist between the image impedances and image transfer constant such that if two of these are defined the third is likewise determined.

The formulas relating the image impedance and image transfer parameters of an L network with the impedance elements of the L are

$$\left.\begin{aligned} Z_1 &= \sqrt{Z_{I_1}(Z_{I_1} - Z_{I_2})} \\ Z_3 &= Z_{I_2} \sqrt{\frac{Z_{I_1}}{Z_{I_1} - Z_{I_2}}} \\ \tanh \theta &= \sqrt{\frac{Z_{I_1} - Z_{I_2}}{Z_{I_1}}} \end{aligned}\right\} \tag{109}$$

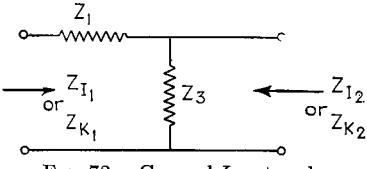


FIG. 72.—General L network.

The corresponding formulas for iterative impedance operation are

$$\left.\begin{aligned} Z_1 &= Z_{k_1} \left(\frac{\epsilon^P - 1}{\epsilon^P}\right) \\ Z_3 &= \frac{Z_{k_1}}{\epsilon^P - 1} \\ Z_{k_2} &= \frac{Z_{k_1}}{\epsilon^P} \end{aligned}\right\} \tag{110}$$

where Z_{k_1} and Z_{k_2} are the iterative impedances of the two ends of the network, as shown in Fig. 72, and P is the iterative transfer constant, as defined by Eq. (101). L networks find their chief use in impedance matching systems and in attenuators.

Lattice Sections.—A lattice is a symmetrical balanced four-terminal network composed of two pairs of impedances, arranged as shown in Fig. 73. It will be noticed that the lattice is essentially a bridge in which the input is applied across one diagonal of the bridge, and the output is taken from the other diagonal.

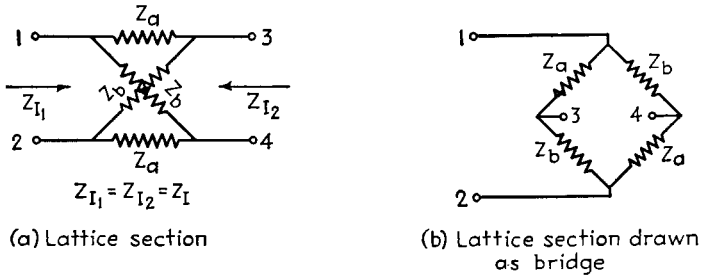


FIG. 73.—General lattice network.

The basic formulas of the lattice section in terms of image impedance and image transfer constant are

$$Z_I = Z_{I_1} = Z_{I_2} = \sqrt{Z_a Z_b}$$
$$\tanh\left(\frac{\theta}{2}\right) = \sqrt{\frac{Z_a}{Z_b}} \tag{111}$$

If one is given θ and Z_I, then

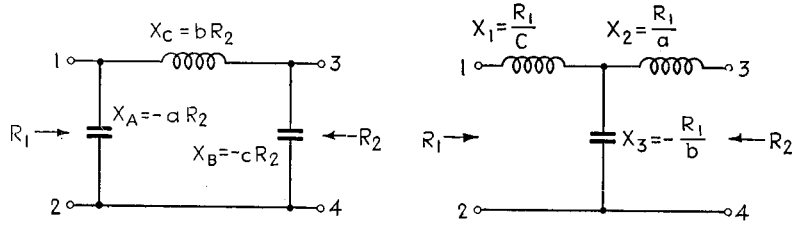
$$Z_a = Z_I \tanh\left(\frac{\theta}{2}\right)$$
$$Z_b = \frac{Z_I}{\tanh\left(\frac{\theta}{2}\right)} \tag{112}$$

It is possible to represent any symmetrical four-terminal network by a lattice having physically realizable impedance arms. In contrast, the π or T network equivalent of a complicated four-terminal network will sometimes require negative circuit elements in some of the arms, and hence be physically unrealizable. The image impedance of a lattice depends only upon the product of the two branch impedances, whereas the image transfer constant depends only upon the ratio of these impedances. It is therefore possible in the lattice to control the transmission characteristics entirely independently of the image-impedance behavior.

Lattice networks are used in filters and equalizers.

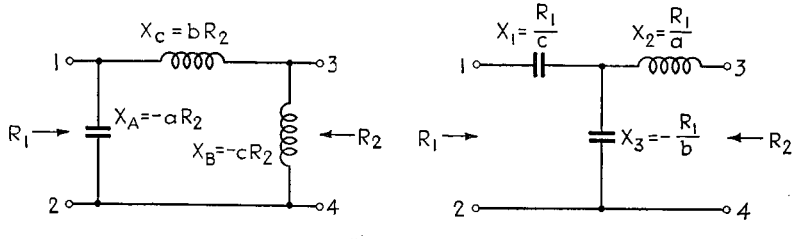
25. Reactive T, L, and π Networks for Matching Impedances.[1] *T and π Reactive Networks.*—T and π networks having impedance arms composed of reactive elements are widely used for matching an antenna to a transmission line in order to give a nonresonant characteristic impedance termination for the line. It is possible with such networks to transform any resistance load that may be offered by the antenna system to any other value of resistance that may be needed to give a characteristic impedance

[1] W. L. Everitt, Output Networks for Radio-frequency Power Amplifiers, *Proc. I.R.E.*, Vol. 19, p. 725, May, 1931; Coupling Networks, *Communications*, Vol. 18, p. 12, September, 1938; p. 12, October, 1938; Carl G. Dietsch, Terminating Concentric Lines, *Electronics*, Vol. 9, p. 16, December, 1936; Ralph P. Glover, R-f Impedance Matching Networks, *Electronics*, Vol. 9, p. 29, January, 1936.



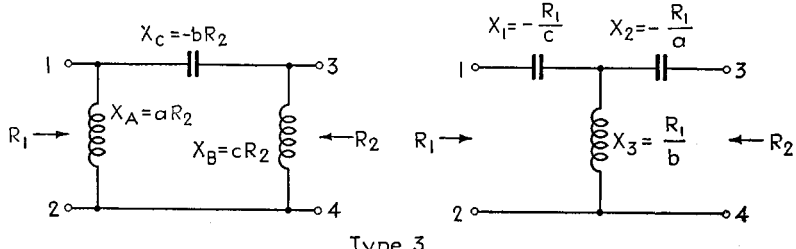
Type I

Phase retarded by large angle (C positive)



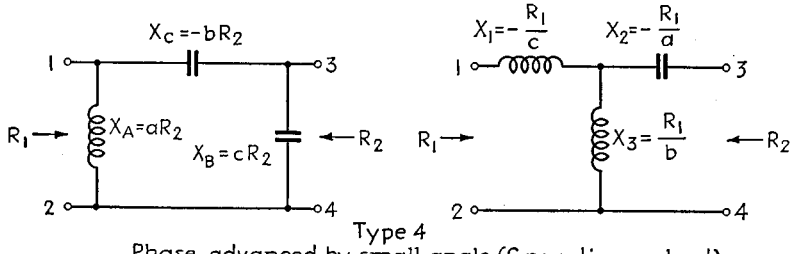
Type 2

Phase retarded by small angle (C negative)



Type 3

Phase advanced by large angle (C positive on chart)

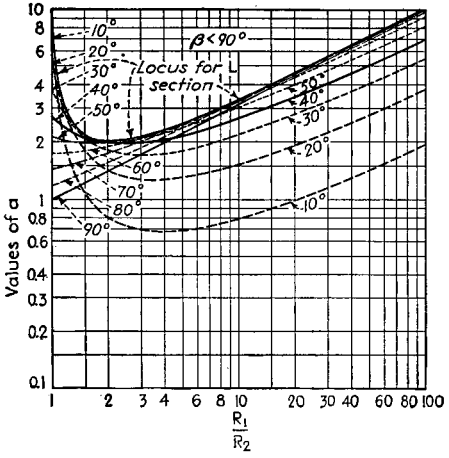


Type 4

Phase advanced by small angle (C negative on chart)

FIG. 74.—Three-element reactive networks which may be used for impedance matching.

load for the transmission line. At the same time, the phase shift introduced by the impedance matching network can have any desired value.



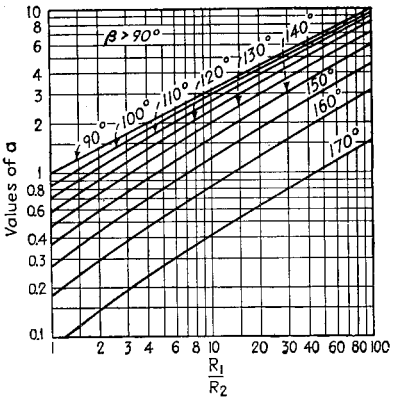


FIG. 75.—Design charts, giving values of a for different values of β for use in networks of Fig. 74.

On the assumption of an ideal network composed of reactive impedances with zero losses, the design relations represented by Eqs. (107) and (108) can be written as

For T section:

$$\left.\begin{aligned} Z_1 &= -j\frac{R_1\cos\beta - \sqrt{R_1R_2}}{\sin\beta} \\ Z_2 &= -j\frac{R_2\cos\beta - \sqrt{R_1R_2}}{\sin\beta} \\ Z_3 &= -j\frac{\sqrt{R_1R_2}}{\sin\beta} \end{aligned}\right\} \quad (113)$$

For π section:

$$\left.\begin{aligned} Z_A &= j\frac{R_1R_2\sin\beta}{R_2\cos\beta - \sqrt{R_1R_2}} \\ Z_B &= j\frac{R_1R_2\sin\beta}{R_1\cos\beta - \sqrt{R_1R_2}} \\ Z_C &= j\sqrt{R_1R_2}\sin\beta \end{aligned}\right\} \quad (114)$$

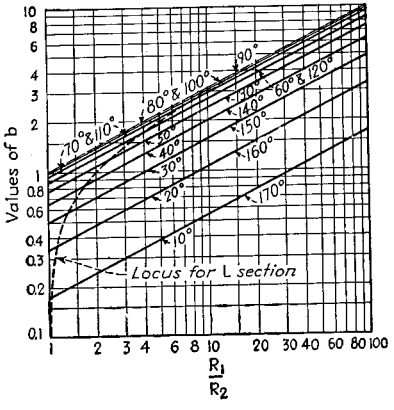


FIG. 76.—Design charts, giving values of b for different values of β for use in networks of Fig. 74.

where R_1 and R_2 are the two image impedances and β is the angle by which the phase at the output terminals lags behind the phase at the input terminals. Design charts derived from these equations are given in Figs. 75, 76, and 77, and together with Fig. 74 will give the reactances required. The charts can be used for negative as well as positive values of β. For negative values of β, the magnitudes of the constants a, b, and c, are the same as for positive, but the signs used in front of the constants are reversed, as shown in Type 3 and 4 networks of Fig. 74. In the design curves of Figs. 75 to 77, it has been assumed that R_1/R_2 is greater than unity. This is no restriction, since terminals 1-2 can be placed at either the generator or the load end of the network, according to whichever must match the higher resistance.

L Reactive Networks.—An L network composed of reactive elements is able to transform a given resistance to make it look like any other resistance by making the image impedance of the network at the terminals facing the load equal the load resistance and the image impedance at the other terminal equal the desired resistance. The phase shift introduced by an L section is determined by the ratio of image impedances, and cannot be specified independently, because the L section has only two impedance arms. The L section can be considered as a special case of Fig. 74, for which constant c

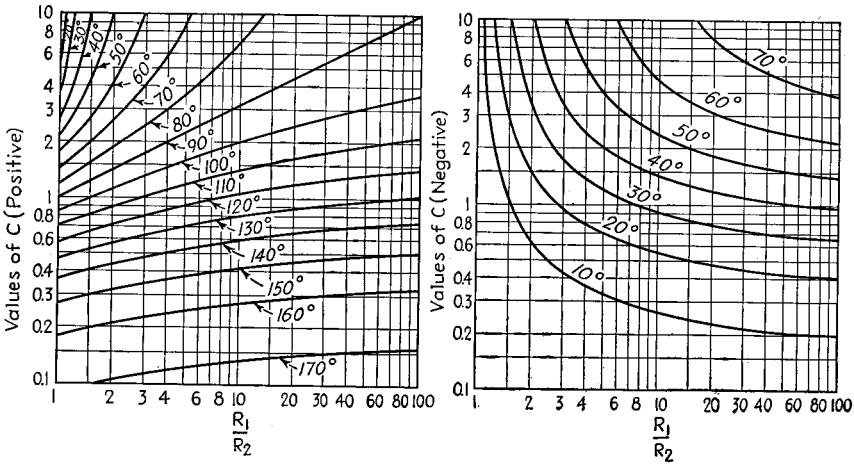


FIG. 77.—Design charts, giving values of c for different values of β for use in networks of Fig. 74.

has the value $c = \infty$, leading to sections as in Fig. 78. The loci corresponding to this condition are shown on Figs. 75 and 76. With an L section, the maximum phase shift obtainable is $\pm 90°$.

Dissipation of Power in Reactive Networks.—In practical networks, the condensers have negligible loss, but the resistance of the inductances is not entirely negligible. On the assumption that the currents in the various network branches are not appreci-

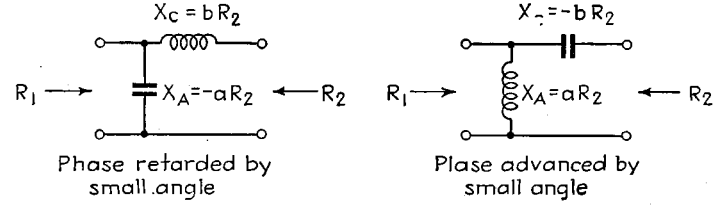


FIG. 78.—Section of L type that is obtained when c in Fig. 74 becomes infinite.

ably affected by the dissipation of the inductive elements and that the various inductive elements have a ratio of reactance to resistance (*i.e.*, Q) that is the same for all inductances, one can write[1]

$$\frac{\text{Power lost in network}}{\text{Power delivered to network}} = \frac{\delta}{Q} \tag{115}$$

where δ is a constant given by either Fig. 79 or 80, according to the type of section

[1] See Everitt, *loc. cit.*

involved. These curves apply for L as well as T and π networks, the locus for the L network being dotted on the curves.

A study of Figs. 79 and 80 shows that the efficiency of the network is implicitly determined by the impedance transformation ratio and the phase shift. There is no choice between T and π networks and between networks that advance and retard the phase, insofar as efficiency is concerned. The loss increases with increasing transformation ratio, and tends to be large when the phase shift of the network is either very

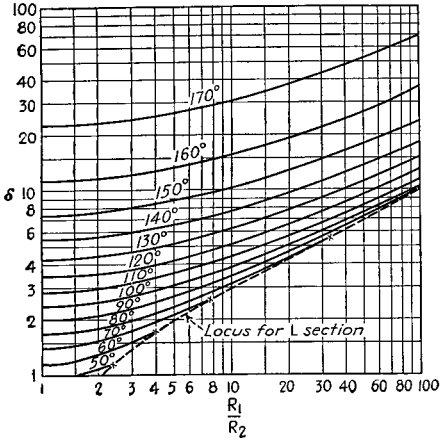


FIG. 79.—Values of δ for use in Eq. (115), applicable for Type 1 and Type 3 networks of Fig. 74. The angles are values of β.

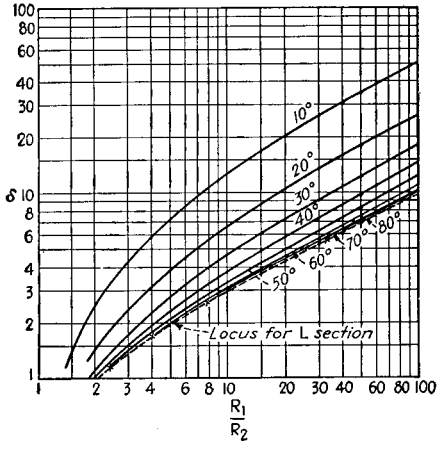


FIG. 80.—Values of δ for use in Eq. (115) applicable for Type 2 and Type 4 networks of FIG. 74. The angles are values of β.

small or very large. Finally, it will be noted that for a given transformation ratio, the L section has a lower loss than either the T or π section. In cases involving very high transformation ratios, or phase shifts that are either very small or approach 180°, an increase of efficiency can be obtained by dividing the total impedance transformation and total phase shift among two or more networks connected in tandem.

Harmonic Reduction.—Coupling networks that call for an inductance in the series arm or a capacity in a shunt arm, or both, can be readily arranged to provide dis-

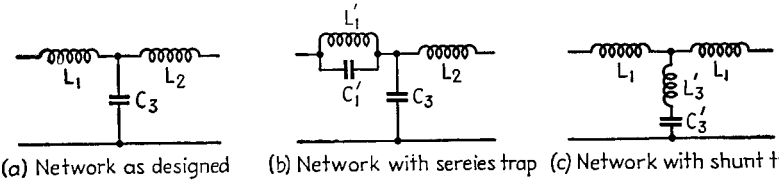


FIG. 81.—Diagram illustrating how a series inductance and a shunt capacity of an impedance matching network may be replaced by parallel and series resonant circuits respectively, to increase discrimination against an undesired harmonic.

crimination against harmonics. The discrimination against a particular harmonic can be made particularly great when the series inductive reactance is supplied by a parallel circuit, as shown in Fig. 81*b*, resonant at a frequency to be suppressed and so proportioned as to give the required inductive reactance at the frequency to be transmitted. An equivalent result is also obtainable when the shunt capacitive reactance is supplied by a suitably designed series circuit, as in Fig. 81*c*, that is resonant at the frequency to be suppressed.

The design formulas are as follows:

For Fig. 81*b*:

$$L_1' = L_1(1 - \gamma^2) \tag{116a}$$

For Fig. 81*c*:

$$C_3' = C_3(1 - \gamma^2) \tag{116b}$$

where γ is the ratio of the frequency to be transmitted to the frequency to be suppressed. C_1' in Fig. 81*b* and L_3' in Fig. 81*c* are assigned values that will make the resonance occur at the frequency to be suppressed.

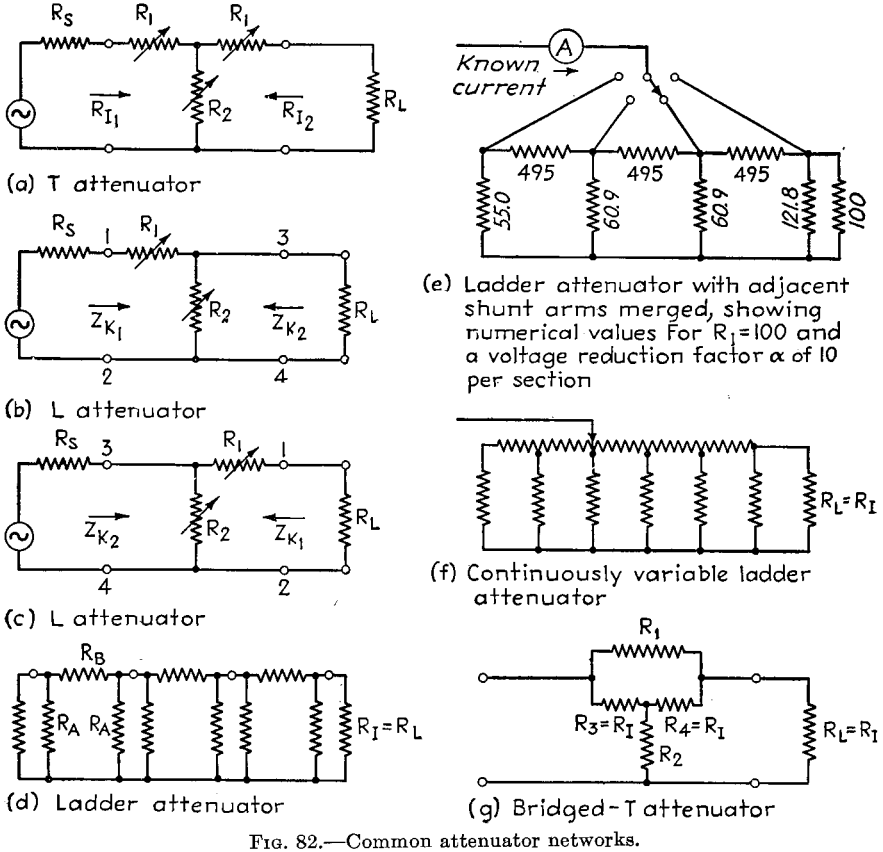


FIG. 82.—Common attenuator networks.

26. Attenuators.[1]—Attenuators are resistance networks used for the purpose of reducing voltage, current, or power in controllable and known amounts. The most commonly used types of attenuators are the *T*, *L*, ladder, and bridged-*T* types.

T Attenuators.—The *T* attenuator consists of three resistances arranged in a *T*, as shown in Fig. 82*a*. Such an attenuator is usually designed with the two series impedances of the *T* identical, thereby making the two image impedances the same.

[1] For further information on attenuators, and in particular for design information including tables giving the resistance values required in the arms for different values of attenuation, see P. K. McElroy, Designing Resistive Attenuating Networks, *Proc. I.R.E.*, Vol. 23, p. 213, March, 1935; Guy C. Omer, Jr., Lattice Attenuating Networks—Complete Design Tables, *Wireless Eng.*, Vol. 17, p. 206, May, 1940; R. E. Blakey, Network Resistances for Balanced Attenuators, *Electronics*, Vol. 8, p. 446, November, 1935.

For such a symmetrical T network composed of resistance elements, the relations expressed by Eqs. (107) can be rewritten as

$$\left.\begin{aligned} R_1 &= R_I\left(\frac{\alpha - 1}{\alpha + 1}\right) \\ R_2 &= R_I\left(\frac{2\alpha}{\alpha^2 - 1}\right) \end{aligned}\right\} \tag{117}$$

where R_1 and R_2 are the attenuator resistances, as shown in Fig. 82*a*, R_I is the image impedance of the section, and α is the number whose natural logarithm is the image transfer constant (*i.e.*, $\alpha = \epsilon^\theta$). Thus when the generator and load impedances both equal the image impedance R_I, the insertion loss of the attenuator is α, *i.e.*, with image operation

$$\alpha = \frac{\text{load current without attenuator}}{\text{load current with attenuator}} \tag{118}$$

T attenuators are used where it is important that the presence of the attenuator in the circuit, and the amount of attenuation have no effect upon the impedance relations existing in the circuit. This is achieved by making the image impedance of the T attenuator equal either the generator or load resistance.

L Attenuators.—The L attenuator consists of two resistance arms arranged as shown in Fig. 82*b* or 82*c*. The arms are so proportioned that one of the iterative impedances of the attenuator is constant irrespective of the attenuation introduced. The corresponding design formulas are

For Z_{k_1} constant (Fig. 82*b* or *c*):

$$\left.\begin{aligned} R_1 &= Z_{k_1}\left(\frac{\alpha' - 1}{\alpha'}\right) \\ R_2 &= \frac{Z_{k_1}}{\alpha' - 1} \end{aligned}\right\} \tag{119}$$

For Z_{k_2} constant (Fig. 82*b* or *c*):

$$\left.\begin{aligned} R_1 &= Z_{k_2}(\alpha' - 1) \\ R_2 &= Z_{k_2}\left(\frac{\alpha'}{\alpha' - 1}\right) \end{aligned}\right\} \tag{120}$$

Here $\alpha' = \epsilon^P$, where P is the iterative transfer constant. When the impedance connected between the pair of terminals opposite from the terminals having constant iterative impedance is equal to this design value of iterative impedance, then

$$\alpha' = \frac{\text{current in load without attenuator}}{\text{current in load with attenuator}} \tag{121}$$

L attenuators are less expensive than the T type, because only two instead of three variable resistances are required to control the attenuation. At the same time the L attenuator maintains impedance independent of attenuation at only one pair of terminals as the attenuation is varied, whereas the T attenuator maintains constant impedance at both terminals. L attenuators are commonly used where a number of loads are associated with a common generator and it is necessary to control the power delivered to each load without altering the impedance offered to the source of power.

Ladder Attenuators.—Ladder attenuators consist of a series of symmetrical π sections designed so that the required ratio of voltage loss per section is obtained with image impedance operation. A typical arrangement is shown in Fig. 82*d*, where there is shown a chain of three π sections terminated at both ends with a resistance equal to the image impedance, and designed to produce the required attenuation when operated

on an image impedance basis. The appropriate design formulas for the individual sections are

$$R_A = R_I \left(\frac{\alpha + 1}{\alpha - 1}\right)$$
$$R_B = R_I \left(\frac{\alpha^2 - 1}{2\alpha}\right) \tag{122}$$

Here R_A and R_B have the meanings shown in Fig. 82*d*, R_I is the image impedance, and α is the factor by which each section reduces the current with image-impedance conditions (*i.e.*, if $\alpha = 10$, then each section reduces the current in the load to $\frac{1}{10}$ of the load current that would be obtained with the section removed).

The impedance between any junction point in the ladder attenuator and the common side of the system is one-half the image impedance.

Ladder attenuators such as are illustrated in Fig. 82*d* are used in signal generators and in other devices requiring that voltages and currents be reduced in known ratios. A typical arrangement for producing known voltages of small magnitude is illustrated in Fig. 82*e*. Here a known current is supplied to the switch and directed so that it flows from one of the junction points to the common side of the system. This produces a known voltage across that particular branch of the shunt attenuator. The voltage appearing across the output terminals of the attenuator is then this voltage reduced in accordance with the number of attenuator sections between the input point and the load.

A continuously adjustable ladder attenuator is shown in Fig. 82*f*, and is the same as the step attenuator of Fig. 82*d*, except that the switch has been replaced by a slider that permits continuous variation of the input point. As actually constructed, the resistance shown horizontally along the top is an ordinary slide-wire potentiometer, to which suitable shunt resistances have been connected at regular intervals to form a series of π sections. Such an attenuator is inexpensive to construct, and maintains its input and output resistances constant within reasonably narrow limits for all attenuations except extreme values.

Bridged-T Attenuators.—A bridged-T attenuator consists of four resistance arms arranged as shown in Fig. 82*g*. When proportioned so that $R_3 = R_4 = R_I$, and $R_1R_2 = R_I^2$, the image impedance will be a constant value R_I, irrespective of attenuation. With image-impedance operation

$$\frac{\text{Load current with attenuator}}{\text{Load current without attenuator}} = \frac{R_I}{R_1 + R_I} \tag{123}$$

The bridged-T attenuator is equivalent to a simple T attenuator but requires only two instead of three variable resistances.

Decimal Attenuators.[1]—A decimal attenuator is a system of attenuators so arranged that a voltage or current can be reduced in decimal fractions. Such an attenuator is shown in Fig. 83, where any voltage from 0.001 to 1.0 times the input voltage can be obtained in steps of 0.001 volt. This decimal attenuator includes three similar L-type attenuators designed to operate between equal generator and load resistances and to have a constant output resistance equal to the generator resistance R_g. Each attenuator is adjustable in steps such that $1/\alpha$ can be made 0, 0.1, 0.2, 0.3, etc., up to 1.0. The second of these L attenuators feeds into a single T attenuator having an input and output resistance equal to the output resistance of the L section and having $\alpha = 10$. The third of these attenuators delivers its output to two such T sections in tandem. These three systems hence give output voltages in steps of 0.1, 0.01, and 0.001, respectively. The outputs of all three attenuating systems are connected in parallel, so that when

[1] Decimal attenuators as described here were developed by the General Radio Company.

the output terminals are open-circuited the load resistance for any one attenuator system consists of the output resistance of the other two systems in parallel. The output voltages of the three attenuators are thus superimposed upon each other, and add up directly without being affected by the fact that the load resistance for any one attenuator is supplied by the other two attenuators, which, at the same time, produce their own output voltages. The fact that there is a mismatch of resistance at the output need not be corrected for, since this changes all output voltages by the same percentage and so does not disturb the relative outputs. The output resistance to any load connected across the output terminals is the output resistances of the three attenuator systems in parallel. The addition of such a load merely increases the mismatch on the output side of the attenuator, and changes all output voltages by the same percentage without altering the relative values at the different attenuator settings.

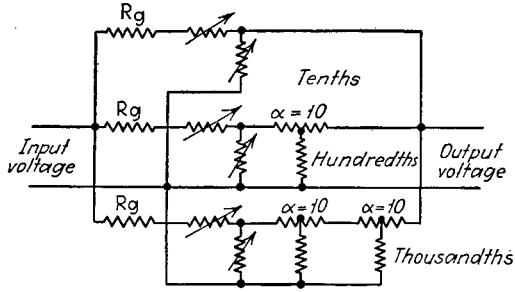


FIG. 83.—Circuit diagram of decimal attenuator.

27. Relation between Attenuation and Phase Shift in Four-terminal Networks.[1]— For any four-terminal network connected between resistive terminal impedances, there is always a minimum possible phase change that can be associated with a given transmission characteristic. In such a *minimum-phase-shift network*, the phase characteristics can be calculated if the transmission characteristic is known, and vice versa.

Any actual four-terminal network will be a minimum-phase-shift network unless (1) the network contains a transmission line or an equivalent circuit with distributed constants, or (2) the circuit includes an all-pass section, either as an individual structure or in a combination that can be replaced by an all-pass filter section plus some other physical structure. It is to be noted particularly that all ladder networks are automatically of the minimum-phase-shift type, since it is impossible to form an all-pass filter section from alternate series and shunt impedances.

Phase-area Theorem.—One of the simplest relations existing between attenuation and phase shift is

$$\int_{-\infty}^{+\infty} B d\mu = \frac{\pi}{17.37} (A_\infty - A_0) \tag{124}$$

where B = phase shift, radians.

μ = $\log_e (f/f_0)$, where f is the actual frequency and f_0 is any convenient reference frequency.

A_∞ = attenuation, db, at infinite frequency.

A_0 = attenuation, db, at zero frequency (not at frequency f_0).

The relation described by Eq. (124), expressed in words, is to the effect that the total area under the phase characteristics, when plotted on a logarithmic frequency scale, depends only upon the difference between the transmission (or attenuation) at zero and infinite frequency, and does not depend in any way upon the way in which the

[1] H. W. Bode, Relations between Attenuation and Phase in Feedback Amplifier Design, *Bell System Tech. Jour.*, Vol. 19, p. 421, July, 1940; also, U.S. Patent No. 2,123,178; Y. W. Lee, Synthesis of Electrical Networks, *Jour. of Math. and Physics*, Vol. 11, p. 83, 1931–32.

transmission varies between these limits; nor does it depend upon the physical configuration of the network, provided a minimum-phase-shift structure is employed. The equality of the phase areas under different conditions is illustrated in Fig. 84.

The relation expressed by Eq. (124) can be termed the *phase-area theorem*, and has several very important practical consequences. It means, for example, that if a network is employed to change the transmission from one fixed value to another fixed value and if, at the same time, the maximum phase shift that can be permitted at any frequency is limited to some relatively low value, then the region where the attenuation is changing must be spread out over a sufficiently wide frequency range so that the necessary area can be obtained under the phase curve without this curve having a maximum exceeding the allowable value. This fact is of considerable significance in connection with the design of feedback amplifier systems, since here the transmission around the feedback loop must have associated with it a phase shift that is always less than 180°, even when the transmission is varying.

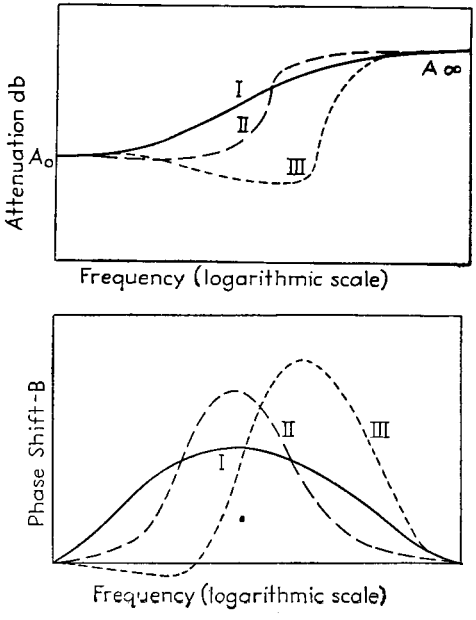


FIG. 84.—Diagram to illustrate relation between area under phase-shift curve and the change in attenuation.

Phase Shift as a Function of Attenuation Slope.—The minimum phase change associated with a given attenuation characteristic can be expressed as[1]

$$B_c = \frac{\pi}{12}\left(\frac{dA}{d\mu}\right)_c + \frac{1}{6\pi}\int_{-\infty}^{+\infty}\left[\left(\frac{dA}{d\mu}\right) - \left(\frac{dA}{d\mu}\right)_c\right]\log_\epsilon \coth\frac{|\mu|}{2}\,d\mu \qquad (125)$$

where B_c = phase shift, radians, at the frequency f_c.
$dA/d\mu$ = slope of attenuation curve, db per octave.
$\mu = \log_\epsilon (f/f_c)$, where f is frequency and f_c the frequency at which B_c is desired.
coth x = hyperbolic cotangent of x ($= 1/\tanh x$).
subscript c denotes evaluated at $f = f_c$.
log coth $\left|\frac{\mu}{2}\right|$ denotes the real part of log coth $\frac{\mu}{2}$ (which is complex when μ is negative).

Equation (125) shows that the minimum phase change can be expressed as the sum of two components. The first of these is proportional to the slope of the attenuation characteristic when plotted on a logarithmic frequency scale, and amounts to 180° when the attenuation is varying at the rate of 12 db per octave at the reference frequency. The second term is determined by an integral, the integrand of which is proportional to the weighting function $\log_\epsilon \coth (|\mu|/2)$, which is plotted in Fig. 85, and the difference between the slope of the attenuation characteristic at the desired frequency f_c and the slope elsewhere. Because of the symmetrical character of the weighting function, this second contribution to the phase shift is determined by

[1] The π in the denominator in front of the integral represents the correction of an error in Eq. (10) of Bode's U.S. Patent No. 2,123,178, upon which the equation given here is based.

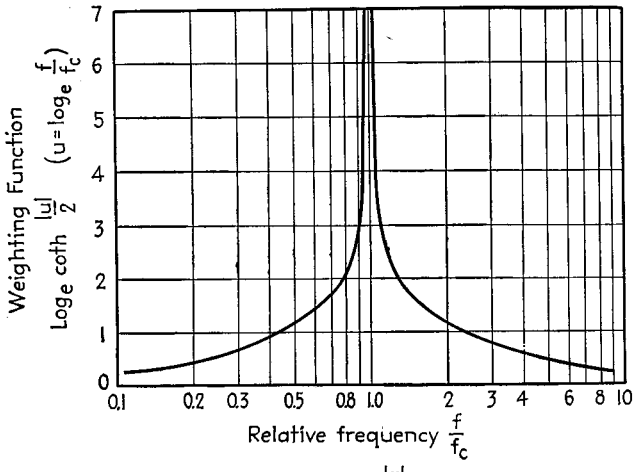


FIG. 85.—Weighting function $\log_\epsilon \coth \frac{|\mu|}{2}$ appearing in Eq. (125).

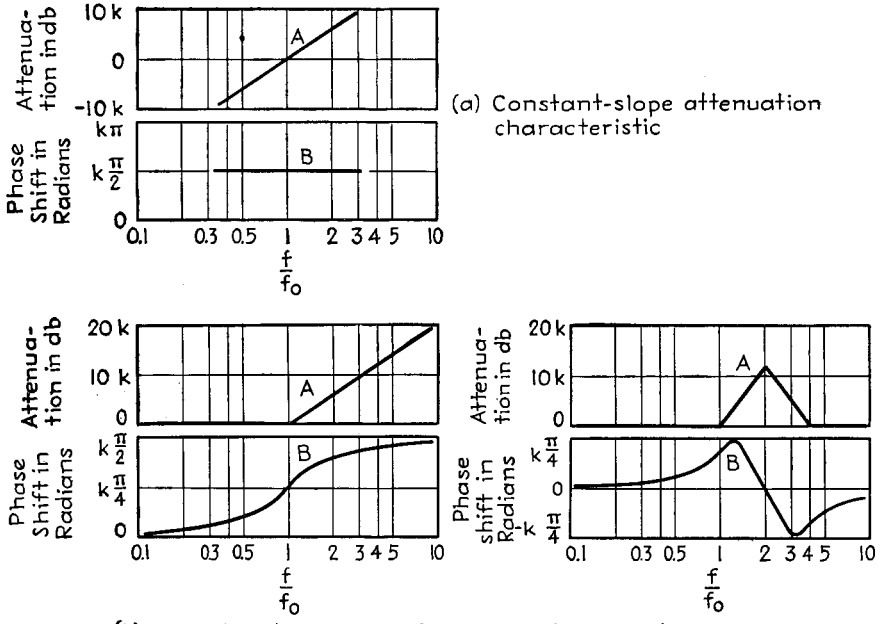


FIG. 86.—Examples of the relationship between phase and attenuation characteristic in several idealized examples.

the extent that the attenuation-slope characteristic fails to have negative symmetry (*i.e.*, odd-function symmetry) about the reference frequency. Also, the shape of the weighting function is such that changes in the slope of the attenuation characteristic at frequencies close to the reference frequency have far more effect upon phase shift than do changes in the slope of the characteristic at more remote frequencies.

Examples of the relationship between attenuation characteristics and phase shift are given in Fig. 86 for several idealized cases.

Specification of Complete Phase and Attenuation Characteristics from Attenuation and Phase Fragments.—Relations have been worked out from which the complete attenuation and phase characteristics of a minimum phase-shift network can be determined by prescribing the attenuation characteristic over a portion of the frequency range and prescribing the phase-shift characteristic for the remainder of the frequency range.

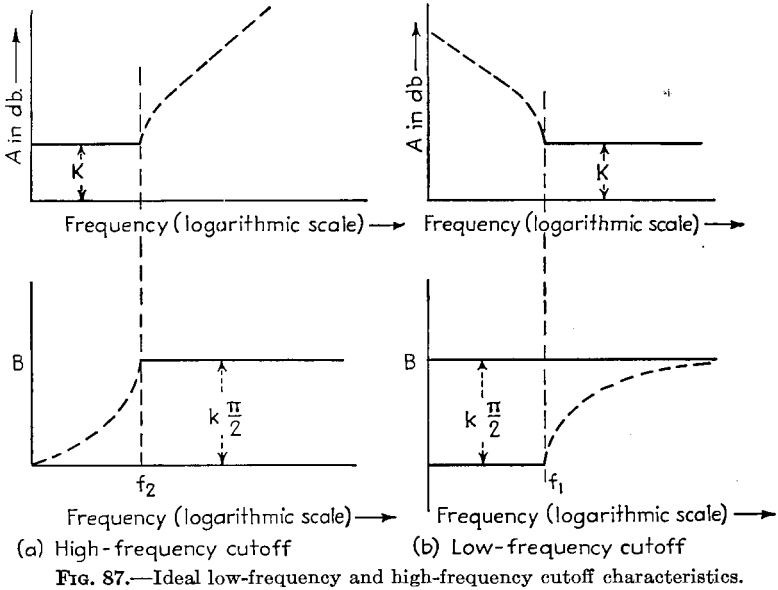


FIG. 87.—Ideal low-frequency and high-frequency cutoff characteristics.

Certain special cases of this character that are of particular importance are illustrated in Fig. 87. Here the attenuation is specified as a constant value in the useful range of the circuit, while the phase shift is specified as having a constant value (normally slightly less than 180°), outside the useful frequency range. The remainder of the characteristics are shown dotted. The essential formulas for the cases illustrated in Fig. 87 are

High-frequency cutoff (Fig. 87*a*):

$$B = k \sin^{-1} \frac{f}{f_2} \qquad f < f_2 \tag{126a}$$

$$A = K + 8.69k \log_\epsilon \left[\sqrt{\left(\frac{f}{f_2}\right)^2 - 1} + \frac{f}{f_2} \right] \qquad f > f_2 \tag{126b}$$

Low-frequency cutoff:

$$B = -k \sin^{-1} \frac{f_1}{f} \qquad f > f_1 \tag{127a}$$

$$A = K + 8.69k \log_\epsilon \left[\sqrt{\left(\frac{f_1}{f}\right)^2 - 1} + \frac{f_1}{f} \right] \qquad f < f_1 \tag{127b}$$

The notation either is as illustrated in Fig. 87 or is as previously used, with the addition that B is in radians and A in db, and K is the attenuation in the region where the attenuation is constant. These expressions can be considered as representing the fundamental formulas for the high- and low-frequency cutoff characteristics of the feedback loop of the ideal feedback amplifier.

Application of Minimum Phase-shift Principles to the Design of Feedback Amplifier Circuits.—A schematic diagram of an amplifier provided with negative feedback is shown in Fig. 88. In this diagram, the portion CDE transmits a fraction of the output of the amplifier back to the input circuit, and superimposes this upon the applied signal. In the normal frequency range, the phase relations are so adjusted that this feedback action opposes the applied signal, thereby reducing the amplification and giving negative feedback. In order that such a system will not oscillate under any conditions, it is necessary that the transmission and phase characteristics of the "feedback loop" $ABCDE$ be such that the transmission around this loop is less than unity when the phase shift reaches 180°. The transmission around this loop is obtained by breaking the circuit at x, applying a signal voltage at point A, and comparing this signal voltage with the voltage that it produces at the point E.

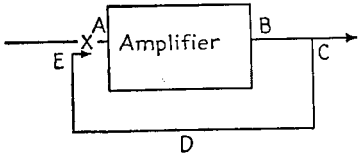


FIG. 88.—Schematic feedback amplifier diagram.

In an ideal feedback amplifier system, the transmission around the feedback loop would be constant throughout the useful band of frequencies, thereby giving constant negative feedback in the useful range. At the same time, the transmission around this loop should drop off as rapidly as possible outside the useful range of frequencies in order that the frequency range over which the transmission characteristics must be controlled accurately be no greater than absolutely necessary. Because of the phase-area theorem, this maximum rate of falling off in transmission outside the useful band will be obtained when the phase shift has a constant value that approaches as close to 180° as is practical throughout the attenuating range. The ideal characteristic for a feedback loop is accordingly that shown in Fig. 87, with k given a value that is slightly less than 2 to provide a margin of safety.

Limitations Introduced by the Asymptotic Transmission Characteristics of the Feedback Loop at Extreme Frequencies.—The ideal feedback loop characteristics cannot in many cases be realized at extremely high and extremely low frequencies, because at these extreme frequencies such factors as stray shunting capacities, grid-leak-condenser combinations, etc., take over control of the characteristics and may cause the transmission to fall off more rapidly than permissible for an ideal feedback loop. This is illustrated in Fig. 89, where ABC is the ideal characteristic, but because of the asymptotic falling off in transmission of the amplifier at high frequencies, the actual characteristic is $AB'C'$. There is then an additional phase shift at high frequencies. The excess phase shift in the region where the feedback loop transmission changes over from the ideal to the asymptotic characteristic (region C in Fig. 89), can be eliminated by the expedient shown in Fig. 90. Here a step of zero slope is introduced in the transmission characteristic just before the asymptotic characteristic is allowed to take control. The length of this step should be such that

$$\frac{f_a}{f_b} = \frac{f'_a}{f'_b} = \frac{\text{slope of ideal asymptotic characteristic in db per octave}}{\text{slope of actual asymptotic characteristic in db per octave}} \qquad (128)$$

The type of phase-shift characteristic that results is shown in Fig. 90, and the total attenuation between f'_a and f_1 (and f_2 and f_b) is very close to the largest that can possibly

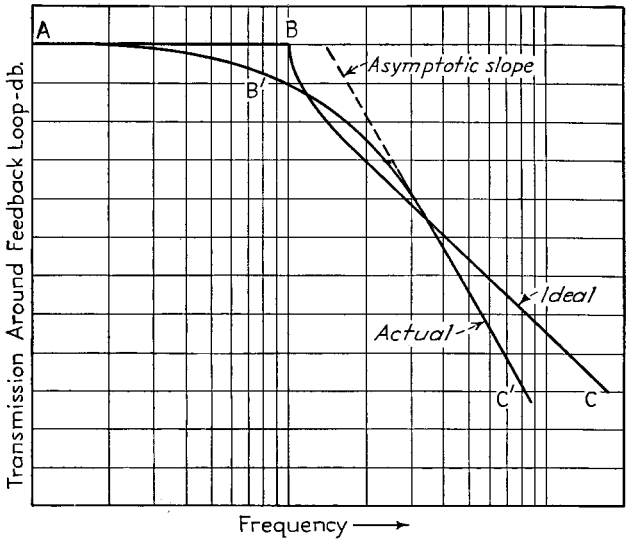


FIG. 89.—Typical amplifier cutoff characteristic, and ideal cutoff characteristic for feedback loop.

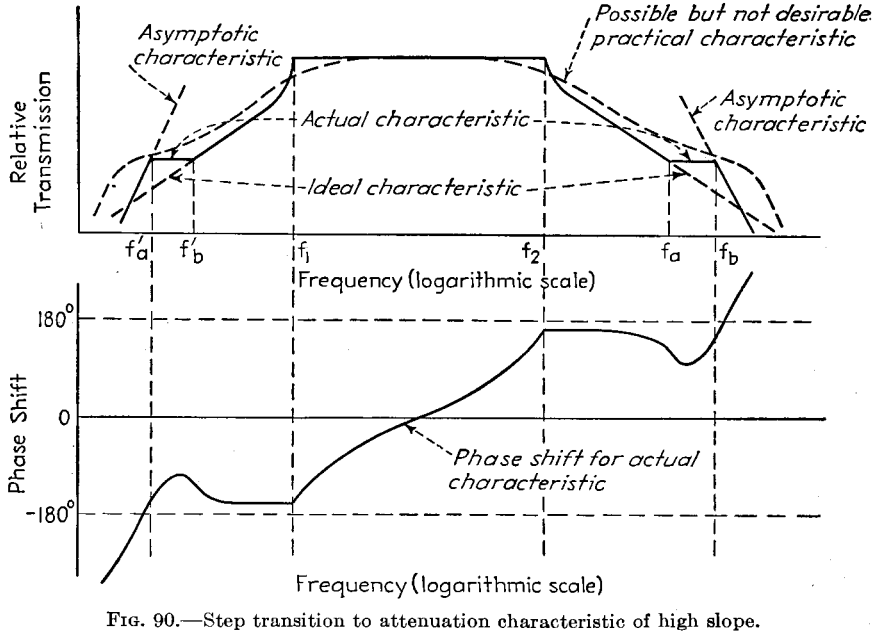


FIG. 90.—Step transition to attenuation characteristic of high slope.

be obtained in this frequency interval without the maximum phase shift exceeding a fixed maximum value.

The slope of the actual asymptotic characteristic can be estimated by inspecting the circuits. Each resistance- or impedance-coupled amplifier stage will contribute 6 db per octave to the asymptotic slope at high frequencies, as will each output transformer fed from a tube and having a resistance load. An interstage coupling transformer will add 12 db per octave to the slope unless a resistance is shunted across the secondary. Similarly, at low frequencies, each grid leak-condenser combination will cause 6 db per octave slope, as will each transformer (output, input, or interstage) excited from a resistance source as a tube or line. Transformers with resonated primaries can contribute 12 db per octave at low frequencies.

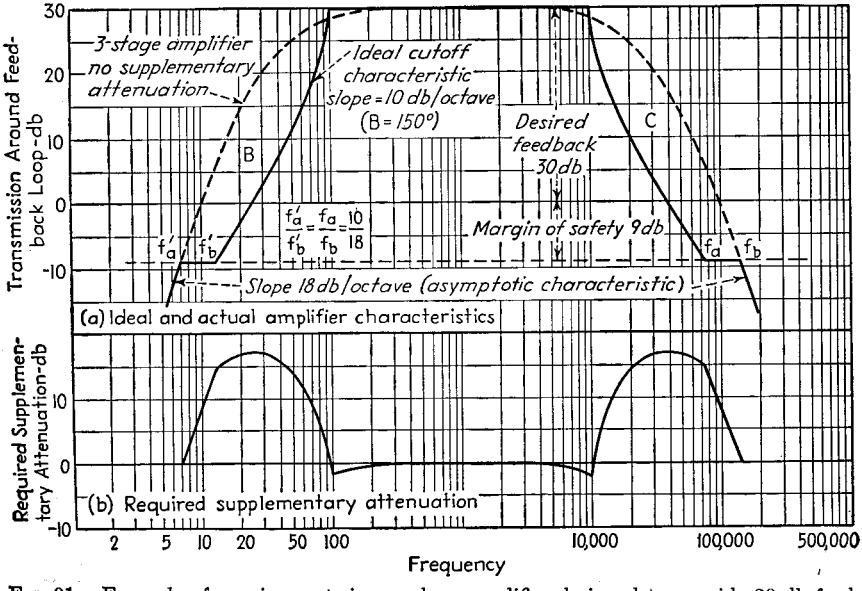


FIG. 91.—Example of requirements imposed on amplifier designed to provide 30-db feedback with 9-db margin of safety.

The asymptotic characteristic is of importance only when it has a value approaching or exceeding 12 db per octave. Thus, with many two-stage amplifiers, the asymptotic characteristic is not a design limitation, and one can theoretically hope to realize the ideal feedback loop characteristics shown in Fig. 87 instead of being forced to the step transition of Fig. 90.

Design Procedures and Considerations.—The design of a feedback amplifier under conditions where the asymptotic characteristic must be considered can be best understood by considering several specific cases.

Example 1.—Assume that a three-stage amplifier has a useful frequency range of 100 to 10,000 cycles and that 30-db feedback is desired with an amplitude margin of safety of 9 db, a phase margin of safety of 30°, and that the asymptotic slope is 18 db per octave for both high- and low-frequency cutoffs. The ideal transmission characteristic of the feedback loop can be calculated from Eqs. (126) and (127) for high and low frequencies, respectively. These characteristics are plotted in Fig. 91*a*, in which the transmission around the feedback loop in the useful frequency range is shown as being 30 db above zero level. The ideal characteristic is continued to 9 db below zero level to f_b' and f_b in order to provide the necessary amplitude margin, after which there is introduced a step having a length such that the frequencies at the two ends of the step are in the ratio 10:18 (or 18:10 as the case may be), which is the slope ratio of the ideal and asymptotic characteristics (slope of the ideal characteristic with 30° phase

margin = 12 × 15%80). Beyond the step, *i.e.*, below f_a' and above f_b, the asymptotic characteristic is permitted to take control, as shown. The actual amplifier would then be built so that in the absence of feedback the amplification would fall to 30 + 9 = 39 db below the amplification in the useful range at the frequencies f_a' and f_b. The frequencies f_a' and f_b for this case, when calculated as above on the basis of the ideal step characteristic, are 7 cycles and 150,000 cycles, respectively. In order to obtain the transmission actually required in order to avoid oscillation, *i.e.*, the characteristics indicated by the heavy line in Fig. 91*a*, then additional circuit elements must be added either to the amplifier or to the feedback loop in order to cause added loss in the regions marked *B* and *C* in Fig. 91. It will be noted that the amplifier must be made much broader than the actual amplification characteristic needed.

Example 2.—Consider a broadcast transmitter in which it is desired to employ 20 db feedback, with 10 db of safety margin in amplitude and a 30° phase margin of safety (k = 1.67), and is to maintain uniform response up to 10 kc. Also, assume that the transmission characteristic around the feedback loop, as observed experimentally, when no special circuit elements are introduced to control feedback characteristics, is as shown by *AHIJ* in Fig. 92*a*, with 24 db per octave asymptotic slope. The optimum transmission characteristic, as calculated by Eq. (126), is the curve *ABCDG*, but because of the asymp-

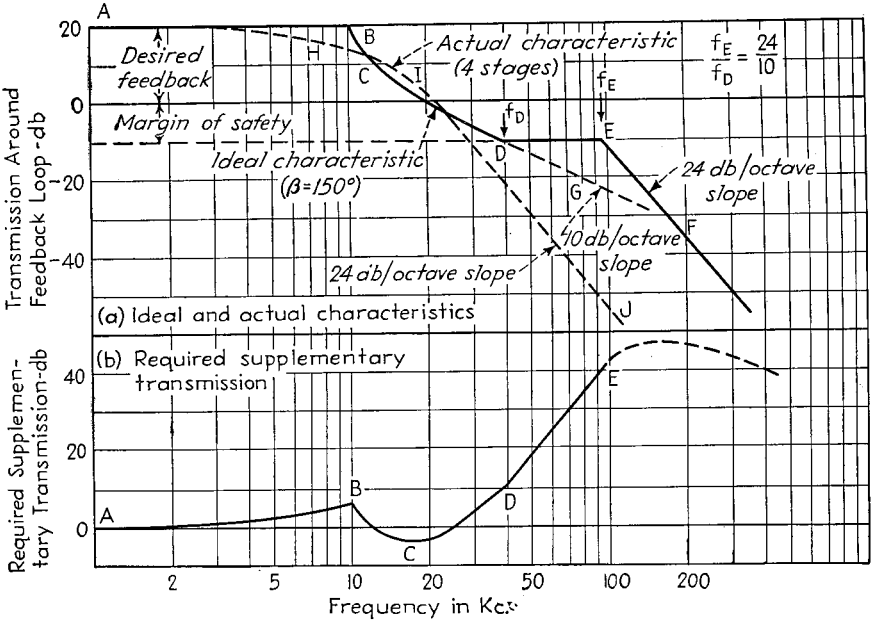


FIG. 92.—Example involving application of negative feedback to a broadcast transmitter.

totic slope near *G*, the practical curve required to give the required amplitude margin of 10 db is the step curve *ABCDEF*. When this is compared with the actual curve *AHIJ*, it is seen that in order to obtain the required characteristic, one must add circuit elements somewhere in the feedback loop that will add to the curve *AHIJ* a characteristic such as illustrated by the curve *ABCDE* of Fig. 92*b*. This can be done by suitable design of circuit elements in the feedback loop, or it may be achieved by adding further stages of amplification either in the feedback path or in the audio-frequency part of the system. These additional stages need contribute little or nothing to the gain of the system in the useful range of frequencies, but must have a rising amplification in the frequency range *DE*. Beyond the point *E*, the amplification of the extra stages can fall off in any convenient manner. If additional stages of amplification are employed to give this required supplementary rising characteristic, these additional stages must be designed so that their asymptotic characteristic does not take control until a frequency is reached that is very much higher than the frequency corresponding to *E* in Fig. 92.

Control of the way in which the transmission around the feedback loop falls off outside the useful range of frequencies can be obtained in a variety of ways. Some of the simpler and more important expedients that can be employed, together with the corresponding transmission characteristics, are illustrated in Fig. 93. The number of such networks that can be devised is of course without limit, and under some circumstances the networks involved may become very complicated.

The frequency range over which the transmission characteristics of the feedback loop must be controlled is surprisingly high, and represents the price that is paid in order to obtain the benefits of negative feedback. The cost is approximately one octave for each 10 db of useful feedback plus about one or two octaves extra as a margin of safety and to take care of failure to realize exactly the optimum characteristic. Thus, if an amplifier with a useful frequency range of 60 to 15,000 cycles is to have 30 db of feedback, the characteristics of the feedback loop must be carefully controlled for at least four octaves beyond this range, or from about 4 to 240,000 cycles (compare with Fig. 91).

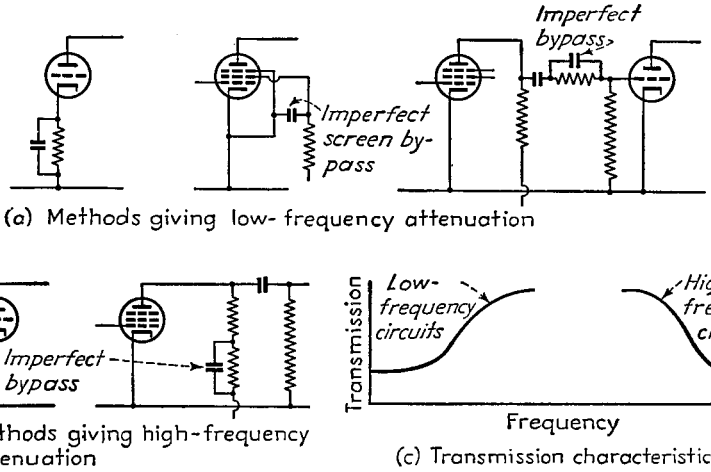


FIG. 93.—Simple methods of modifying amplifier characteristics at high and low frequencies.

28. *M*-derived (Ladder) Filters.[1]—A filter can be considered as a four-terminal network in which the image transfer constant has a value that is either small or zero in a particular range of frequencies and a value that is relatively large for other frequencies.

Fundamental Filter Equations.—Practical filters of the ladder type consist of symmetrical T or π sections composed of reactive elements. When the impedance arms are designated as in Fig. 94, then the image transfer constant θ and the image impedances Z_T and Z_π for the T and π sections are

$$\cosh \theta = 1 + \frac{Z_1}{2Z_2} \tag{129a}$$

or

$$\sinh \frac{\theta}{2} = \sqrt{\frac{Z_1}{4Z_2}} \tag{129b}$$

$$Z_T = \sqrt{Z_1 Z_2 + \frac{1}{4} Z_1^2} \tag{130}$$

$$Z_\pi = \frac{Z_1 Z_2}{\sqrt{Z_1 Z_2 + \frac{1}{4} Z_1^2}} = \frac{Z_1 Z_2}{Z_T} \tag{131}$$

Half sections have an image transfer constant exactly half of the transfer constant of a

[1] Otto J. Zobel, Theory and Design of Uniform and Composite Electric Wave-Filters, *Bell System Tech. Jour.*, Vol. 2, p. 1, January, 1923.

full section. The image impedances of the two sides of the half sections differ, being Z_T on one side and Z_π on the other (see Fig. 94).

An ideal filter in which the impedance arms are pure reactances with zero loss has zero attenuation (*i.e.*, the real part of θ is 0) for all frequencies that make $Z_1/4Z_2$ lie between 0 and -1. Such frequencies are termed *pass frequencies*, and the range in which they lie is termed the *pass band*. The phase constant β of the image transfer constant in the pass band is

$$\cos \beta = 1 + \frac{Z_1}{2Z_2} \tag{132}$$

The image impedance in the pass band is a pure resistance.

All frequencies other than those which make $Z_1/4Z_2$ lie between 0 and -1 suffer attenuation (*i.e.*, the real part α of θ is not zero). Such frequencies are termed *stop frequencies*, and are said to lie in the *stop band* of the filter. The phase shift in the stop

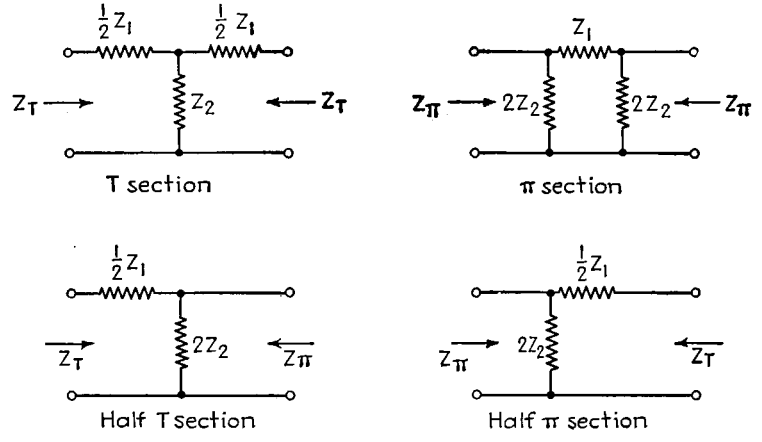


FIG. 94.—Filter sections, showing nomenclature.

bands is either zero or $\pm 180°$, and the image impedance is a pure reactance. The magnitude of the attenuation constant α for the stop frequencies is

$$\cosh \alpha = \left| 1 + \frac{Z_1}{2Z_2} \right| \tag{133}$$

Filter Design.—Practical ladder filters are built up in the manner illustrated in Fig. 95. The middle portion of the filter is composed of a series of symmetrical T or π sections connected in cascade, and at each end of the filter there is a half section as shown. All sections in the filter are matched together at each junction on an image-impedance basis.

The intermediate sections used in constructing a practical filter are of a class in which one can vary the attenuation characteristics of the section by suitably designing the series and shunt arms, without at the same time affecting in any way the image impedance. This makes it possible to arrange matters so that the frequencies for which one section gives only small attenuation are then strongly attenuated by some other section. Sections of this class, however, have an image impedance that is far from constant in the pass band of the filter. This means that a filter consisting only of such sections as might be used in the intermediate portion of a filter would not operate satisfactorily in association with a load resistance having a value inde-

TABLE 2.—DESIGN OF LOW-PASS SECTIONS

Fundamental Relations

R = load resistance f_2 = cut-off frequency (highest frequency transmitted) f_∞ = a frequency of very high attenuation

$$L_k = \frac{R}{\pi f_2} \qquad C_k = \frac{1}{\pi f_2 R} \qquad m = \sqrt{1 - \left(\frac{f_2}{f_\infty}\right)^2}$$

Design of Sections

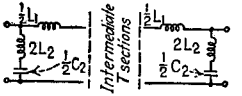
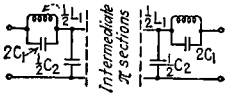
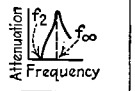
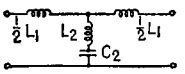
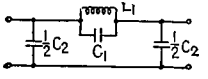
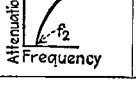
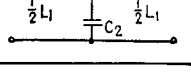
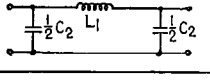
Type	Attenuation characteristic	*A.* Filters having T intermediate sections		*B.* Filters having π intermediate sections	
		Configuration	Formulas	Configuration	Formulas
End (m of approximately 0.6)			$L_1 = mL_k$ $L_2 = \frac{1 - m^2}{4m} L_k$ $C_2 = mC_k$		$L_1 = mL_k$ $C_1 = \frac{1 - m^2}{4m} C_k$ $C_2 = mC_k$
I			$L_1 = mL_k$ $L_2 = \frac{1 - m^2}{4m} L_k$ $C_2 = mC_k$		$L_1 = mL_k$ $C_1 = \frac{1 - m^2}{4m} C_k$ $C_2 = mC_k$
II ($f_\infty = \infty$)			$L_1 = L_k$ $C_2 = C_k$		$L_1 = L_k$ $C_2 = C_k$

TABLE 3.—DESIGN OF HIGH-PASS SECTIONS

Fundamental Relations

R_1 = load resistance f_1 = cut-off frequency f_∞ = a frequency of very high attenuation (lowest frequency transmitted)

$$L_k = \frac{R}{4\pi f_1} \qquad C_k = \frac{1}{4\pi f_1 R} \qquad m = \sqrt{1 - \left(\frac{f_\infty}{f_1}\right)^2}$$

Design of Sections

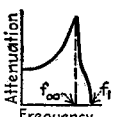
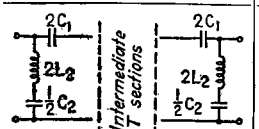
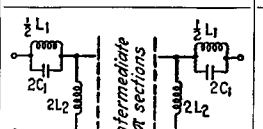
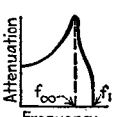
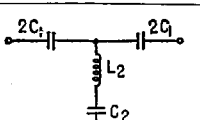
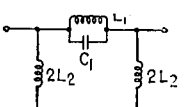
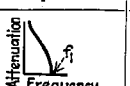
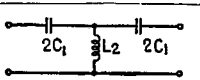
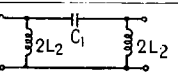
Type	Attenuation characteristic	*A.* Filters having T intermediate sections		*B.* Filters having π intermediate sections	
		Configuration	Formulas	Configuration	Formulas
End (m of approximately 0.6)			$C_1 = \frac{C_k}{m}$ $C_2 = \frac{4m}{1 - m^2} C_k$ $L_2 = \frac{L_k}{m}$		$L_1 = \frac{4m}{1 - m^2} L_k$ $C_1 = \frac{C_k}{m}$ $L_2 = \frac{L_k}{m}$
I			$C_1 = \frac{C_k}{m}$ $C_2 = \frac{4m}{1 - m^2} C_k$ $L_2 = \frac{L_k}{m}$		$L_1 = \frac{4m}{1 - m^2} L_k$ $C_1 = \frac{C_k}{m}$ $L_2 = \frac{L_k}{m}$
II $f_\infty = 0$			$C_1 = C_k$ $L_2 = L_k$		$C_1 = C_k$ $L_2 = L_k$

TABLE 4.—DESIGN OF BAND-PASS SECTIONS

Fundamental Relations

R = load resistance

f_1 = lower frequency limit of pass band

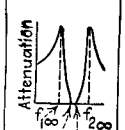
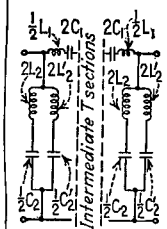
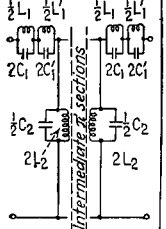
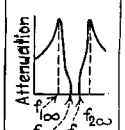
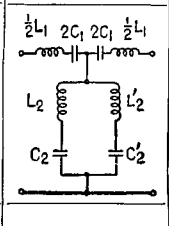
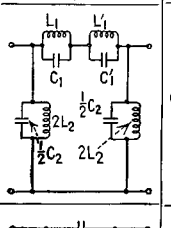
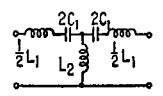
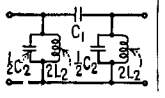
f_2 = higher frequency limit of pass band

$f_{1\infty}$ = a frequency of very high attenuation in low-frequency attenuating band

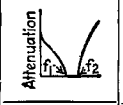
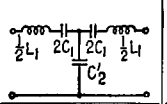
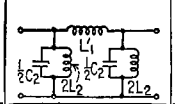
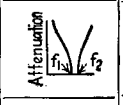
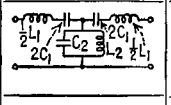
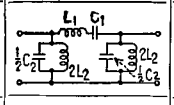
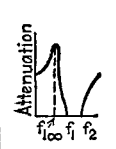
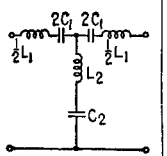
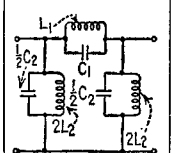
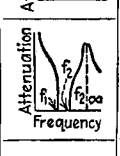
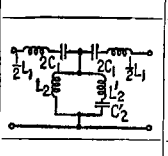
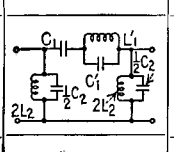
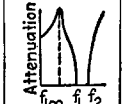
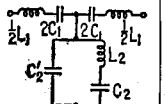
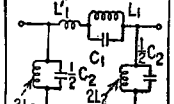
$f_{2\infty}$ = a frequency of very high attenuation in high-frequency attenuating band

$$L_{1k} = \frac{R}{\pi(f_2 - f_1)} \qquad L_{2k} = \frac{(f_2 - f_1)R}{4\pi f_1 f_2} \qquad C_{1k} = \frac{f_2 - f_1}{4\pi f_1 f_2 R} \qquad C_{2k} = \frac{1}{\pi(f_2 - f_1)R}$$

Design of Sections

Type	Attenuation characteristic	A. Filters having T intermediate sections — Configuration	A. Filters having T intermediate sections — Formulas	B. Filters having π intermediate sections — Configuration	B. Filters having π intermediate sections — Formulas	Notation for both T and π sections
End ($m_1 = m_2$ = approximately 0.6)			$L_1 = m_1 L_{1k}$ $L_2 = a L_{1k}$ $L_2' = c L_{1k}$ $C_1 = \frac{C_{1k}}{m_2}$ $C_2 = \frac{C_{1k}}{b}$ $C_2' = \frac{C_{1k}}{d}$		$L_1 = \frac{L_{2k}}{b}$ $L_1' = \frac{L_{2k}}{d}$ $L_2 = \frac{L_{2k}}{m_2}$ $C_1 = aC_{2k}$ $C_1' = cC_{2k}$ $C_2 = m_1 C_{2k}$	$g = \sqrt{\left(1 - \frac{f_{1\infty}^2}{f_1^2}\right)\left(1 - \frac{f_{1\infty}^2}{f_2^2}\right)}$ $h = \sqrt{\left(1 - \frac{f_1^2}{f_{2\infty}^2}\right)\left(1 - \frac{f_2^2}{f_{2\infty}^2}\right)}$ $m_1 = \frac{\frac{f_1 f_2}{f_{2\infty}^2} g + h}{1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}}$ $m_2 = \frac{g + \frac{f_{1\infty}^2}{f_1 f_2} h}{1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}}$ $a = \frac{(1-m_1^2) f_{2\infty}^2}{4g f_1 f_2}\left(1 - \frac{f_1^2}{f_{2\infty}^2}\right) = \frac{(1-m_2^2) f_1 f_2}{4g f_{1\infty}^2}\left(1 - \frac{f_{1\infty}^2}{f_2^2}\right)$ $b = \frac{(1-m_2^2)}{4g}\left(1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}\right)$ $c = \frac{(1-m_1^2)}{4h}\left(1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}\right)$ $d = \frac{(1-m_1^2) f_{2\infty}^2}{4h f_1 f_2}\left(1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}\right) = \frac{(1-m_2^2) f_1 f_2}{4h f_{1\infty}^2}\left(1 - \frac{f_{1\infty}^2}{f_{2\infty}^2}\right)$ when ($m_1 = m_2$), $g = h,\ a = d,\ b = c,\ f_{1\infty} = \frac{f_1 f_2}{f_{2\infty}},\ m_1 = m_2 = \frac{h}{1 - \frac{f_1 f_2}{f_{2\infty}^2}}$ and $f_{2\infty}^2 = \frac{f_1^2 + f_2^2 - 2m^2 f_1 f_2}{2(1-m^2)} + \left[\left(\frac{f_1^2 + f_2^2 - 2m^2 f_1 f_2}{2(1-m^2)}\right)^2 - f_1^2 f_2^2\right]^{1/2}$
I			$L_1 = m_1 L_{1k}$ $L_2 = a L_{1k}$ $L_2' = c L_{1k}$ $C_1 = \frac{C_{1k}}{m_2}$ $C_2 = \frac{C_{1k}}{b}$ $C_2' = \frac{C_{1k}}{d}$		$L_1 = \frac{L_{2k}}{b}$ $L_1' = \frac{L_{2k}}{d}$ $L_2 = \frac{L_{2k}}{m_2}$ $C_1 = aC_{2k}$ $C_1' = cC_{2k}$ $C_2 = m_1 C_{2k}$	
II $f_{1\infty} = 0$ $f_{2\infty} = f_2$			$L_1 = \frac{f_1 R}{\pi f_2 (f_2 - f_1)}$ $L_2 = \frac{(f_1 + f_2)R}{4\pi f_1 f_2}$ $C_1 = C_{1k}$		$C_1 = \frac{f_1 + f_2}{4\pi f_1 f_2 R}$ $C_2 = \frac{f_1}{\pi f_2 (f_2 - f_1) R}$ $L_2 = L_{2k}$	

These formulas apply for both end and Type I sections

III $f_{1\infty} = f_1$ $f_{2\infty} = \infty$			$L_1 = L_{1k}$ $C_2' = \dfrac{1}{\pi(f_1+f_2)R}$ $C_1 = \dfrac{f_2 - f_1}{4\pi f_1^2 R}$		$L_1' = \dfrac{R}{\pi(f_1+f_2)}$ $L_2 = \dfrac{(f_2-f_1)R}{4\pi f_1^2}$ $C_2 = C_{2k}$	
IV $f_{1\infty} = 0$ $f_{2\infty} = \infty$			$L_1 = L_{1k}$ $L_2 = L_{2k}$ $C_1 = C_{1k}$ $C_2 = C_{2k}$		$L_1 = L_{1k}$ $L_2 = L_{2k}$ $C_1 = C_{1k}$ $C_2 = C_{2k}$	
V $f_{2\infty} = f_2$			$L_1 = m_1 L_{1k}$ $L_2 = \dfrac{(1-m_1^2)}{4m_1} L_{1k}$ $C_1 = \dfrac{C_{1k}}{m_1}$ $C_2 = \dfrac{4m_2}{1-m_2^2} C_{1k}$ See notation for m_1 and m_2		$L_1 = \dfrac{4m_2}{1-m_2^2} L_{2k}$ $L_2 = \dfrac{L_{2k}}{m_2}$ $C_1 = \dfrac{(1-m_1^2)}{4m_1} C_{2k}$ $C_2 = m_1 C_{2k}$ See notation for m_1 and m_2	$m_1 = \dfrac{f_1}{f_2} m_2 \qquad m_2 = \sqrt{\dfrac{1-\frac{f_1^2}{f_{2\infty}^2}}{1-\frac{f_1^2}{f_2^2}}}$
VI $f_{1\infty} = f_1$		Same circuit as above for Type V	Same formulas as above for Type V. See notation for m_1 and m_2	Same circuit as above for Type V	Same formulas as above for Type V See notation for m_1 and m_2	$m_1 = \sqrt{\dfrac{1-\frac{f_2^2}{f_{2\infty}^2}}{1-\frac{f_1^2}{f_{2\infty}^2}}} \qquad m_2 = \dfrac{f_1}{f_2} m_1$
VII $f_{1\infty} = 0$			$L_1 = m_1 L_{1k}$ $L_2 = aL_{1k}$ $L_2' = \dfrac{(1-m_1^2)}{4h} L_{1k}$ $C_1 = C_{1k}\ C_2' = \dfrac{h}{a} C_{1k}$		$L_1' = \dfrac{h}{a} L_{2k}\ \ L_2 = L_{2k}$ $C_1 = aC_{2k}$ $C_2 = m_1 C_{2k}$ $C_1' = \dfrac{(1-m_1^2)}{4h} C_{2k}$	$h = \sqrt{\left(1-\frac{f_1^2}{f_{2\infty}^2}\right)\left(1-\frac{f_2^2}{f_{2\infty}^2}\right)} \qquad m_1 = \dfrac{f_1 f_2}{f_{2\infty}^2} + h$ $a = \dfrac{(1-m_1^2) f_{2\infty}^2}{4 f_1 f_2}$
VIII $f_{2\infty} = \infty$			$L_1 = L_{1k}\ \ L_2 = \dfrac{d}{g} L_{1k}$ $C_1 = \dfrac{C_{1k}}{m_2}$ $C_2 = \dfrac{4g}{1-m_2^2} C_{1k}$ $C_2' = \dfrac{C_{1k}}{d}$		$L_1 = \dfrac{4g}{1-m_2^2} L_{2k}$ $L_2 = \dfrac{L_{2k}}{m_2}\ \ L_1' = \dfrac{L_{2k}}{d}$ $C_1 = \dfrac{d}{g} C_{2k}\ \ C_2 = C_{2k}$	$g = \sqrt{\left(1-\frac{f_{1\infty}^2}{f_1^2}\right)\left(1-\frac{f_{1\infty}^2}{f_2^2}\right)} \qquad m_2 = g + \dfrac{f_{1\infty}^2}{f_1 f_2}$ $d = \dfrac{(1-m_2^2) f_1 f_2}{4 f_{1\infty}^2}$

pendent of frequency. This difficulty is overcome by employing suitably designed terminal half sections, as shown in Fig. 95, that match the intermediate portion of the filter on an image-impedance basis on one side, but offer a more desirable type of image impedance at their external terminals.

Sections suitable for use in low-pass, high-pass, and band-pass filters, together with the necessary design formulas, are given in Tables 2, 3, and 4. All the sections illustrated in these tables are suitable for use in the intermediate portion of the filter except those sections specifically designated as end sections. All intermediate sections designed from one of these tables for given cutoff frequencies, and for the same load resistance, will have identical image-impedance characteristics, and so can be connected in cascades on an image-impedance basis. The image transfer constant, *i.e.*, the attenuation and phase-shift characteristics, of the individual sections will, however, depend upon the type of section involved and upon the value assigned to the design parameter m (or m_1 and m_2).

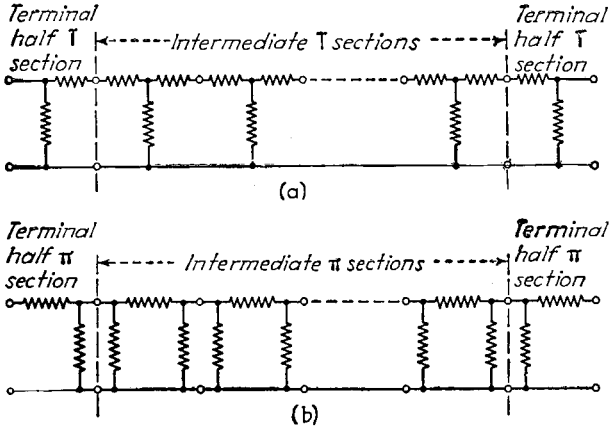


FIG. 95.—Method of building up an m-derived filter.

The terminal half sections are designed according to the formulas given in the respective tables, with the design parameter m assigned a value of approximately 0.6. When the terminal half sections are designed by the tables for the same cutoff frequencies and same load resistance as the intermediate sections, their image impedances will match the intermediate sections at one pair of terminals. Also, when the appropriate value of m is used in the design, the image impedance at the other terminals will be substantially constant over almost the entire pass band of the filter. This is illustrated in Fig. 96, where the effect of using various values of m is shown. Examples of actual filters designed according to Tables 2, 3, and 4 are given in Figs. 97, 98, and 99.

The procedure involved in the design of a filter to meet a given set of conditions falls into several well-defined steps, as follows:

1. Determine the cutoff frequencies, *i.e.*, the frequencies that mark the edge of the pass band, and the load resistance that will meet the design requirements, and decide whether T or π intermediate sections are to be used.

2. Design the terminating sections according to the proper table, using the load resistance and cutoff frequencies selected in step 1.

3. Decide on the number of intermediate sections to be used. The more sections selected the greater will be the attenuation that can be obtained in the stop band.

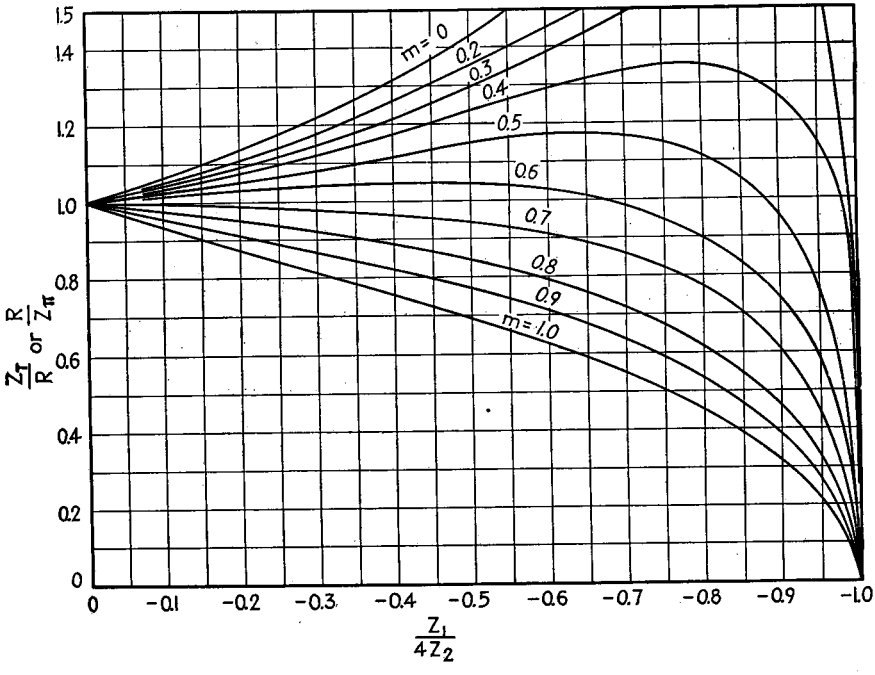
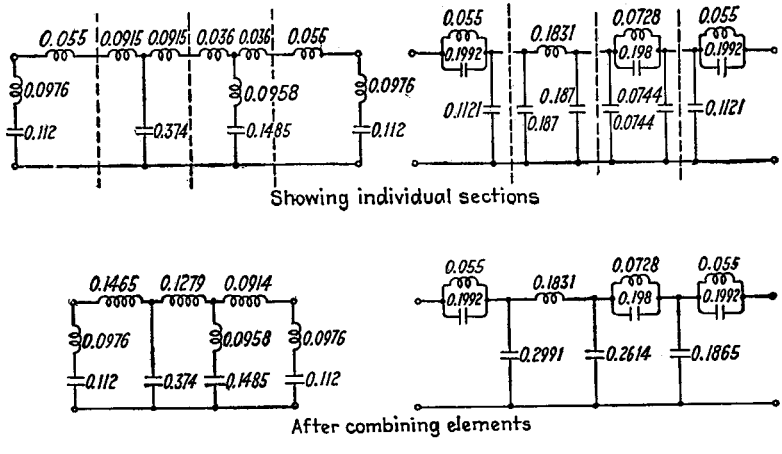


FIG. 96.—Effect of design parameter m on image-impedance characteristics in the pass band.



(a) Intermediate sections of "T" type (b) Intermediate section of "π" type

FIG. 97.—Example of a low-pass filter having a cut-off frequency of 1213 cycles, two intermediate sections, and proportioned to operate with a load resistance of 700 ohms.

One intermediate section will give a moderately good filter, and will be sufficient for many requirements. More than two intermediate sections are seldom required.

4. Select the frequencies at which the different intermediate sections will have their maximum attenuation, and then design these sections for the chosen load resistance and cutoff frequencies, using formulas in the appropriate tables.

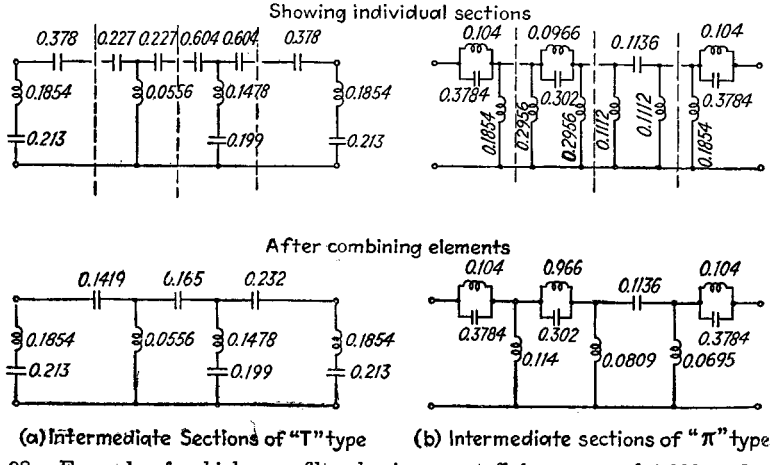


FIG. 98.—Example of a high-pass filter having a cutoff frequency of 1,000 cycles, two intermediate sections, and proportioned to operate with a load resistance of 700 ohms.

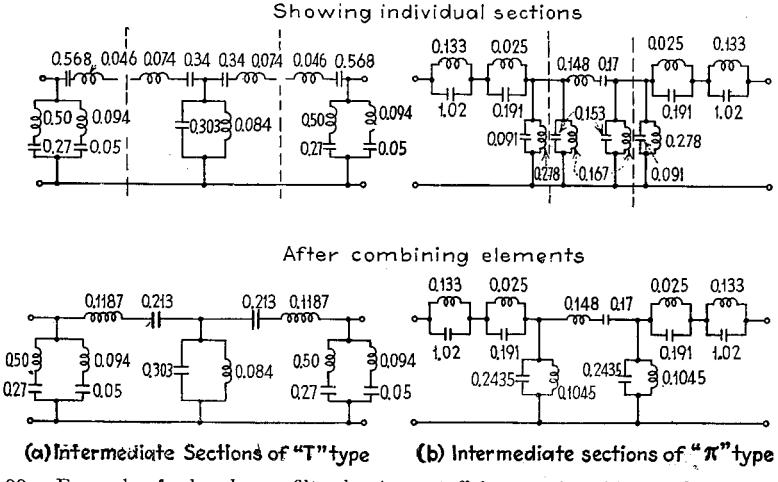


FIG. 99.—Example of a band-pass filter having cutoff frequencies of 500 and 2,000 cycles, one intermediate section, and proportioned to operate with a load resistance of 700 ohms.

The choice between T and π sections for the middle portion of the filter is primarily based upon considerations of convenience. Electrically, the performance of the two types will be identical.

The location of the frequencies for which the intermediate sections have high attenuation must be carefully selected, for upon this choice rest the attenuating properties of the filter. These frequencies of high attenuation should in general be

different for the different sections, and should be so staggered throughout the attenuating bands that every frequency suffers considerable attenuation by at least one section. Inasmuch as the terminal half sections determine one frequency of high attenuation in each attenuating band, these sections should be designed before the intermediate sections in order that the attenuation of the central portion of the filter may best supplement that of the end sections.

It is sometimes desirable to obtain a very rapid rise of attenuation at the edge of the pass band of the filter. Although it is theoretically possible in the ideal case of perfect reactances to obtain a frequency of high attenuation as close to the cutoff frequency as desired, practical considerations require that such frequencies of high attenuation be not too close to the cutoff frequencies. There are two reasons for this. In the first place, as the frequency of high attenuation is moved closer and closer to the cutoff frequency, the attenuating characteristics become narrower, as shown in Fig. 100. The result is that a section designed to attenuate frequencies extremely close to cutoff will be of little use in attenuating frequencies appreciably different

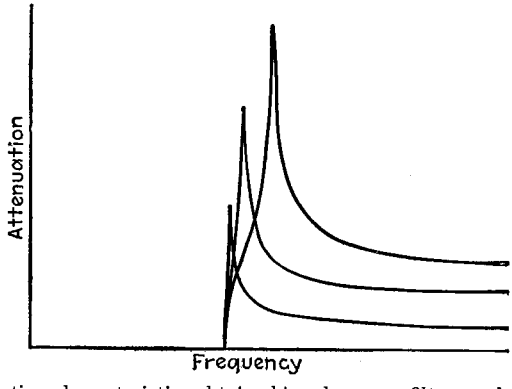


FIG. 100.—Attenuation characteristics obtained in a low-pass filter section as the frequency of high attenuation is placed closer and closer to cutoff.

from cutoff. In the second place, the reactance arms required in a filter having a frequency of high attenuation very close to cutoff assume impractical proportions, and if these impractical circuit elements are built, it will be found that their losses prevent the high attenuation that is theoretically possible from being actually realized. This is illustrated in Fig. 100. In the case of ordinary coils with reasonably high Q, the frequencies of infinite attenuation should differ from cutoff by not less than two to five per cent. Where quartz crystals are used as filter elements, the higher Q of such resonators permits this difference to be reduced greatly.

Filters with MM′ and Fractional Terminations, Filters in Series and Parallel.—Filters such as are described above, employing terminating half sections designed for $m \cong 0.6$, provide an image-impedance characteristic for the external terminals of the filter that is sufficiently constant over the pass band to meet most requirements. Where a closer approximation to a constant resistance image impedance is required, more elaborate terminating sections can be employed. The next steps in complexity beyond the simple sections already discussed, are given in Fig. 101 and are termed MM' terminations.[1] These arrangements give a decided improvement in image impedance at the external terminals. Thus, in the case of a low-pass filter, the usual $m = 0.6$ termination gives an image impedance constant to within 4 per cent over

[1] Otto J. Zobel, Extensions to the Theory and Design of Electric Wave-filters, *Bell System Tech. Jour.*, Vol. 10, p. 284, April, 1931.

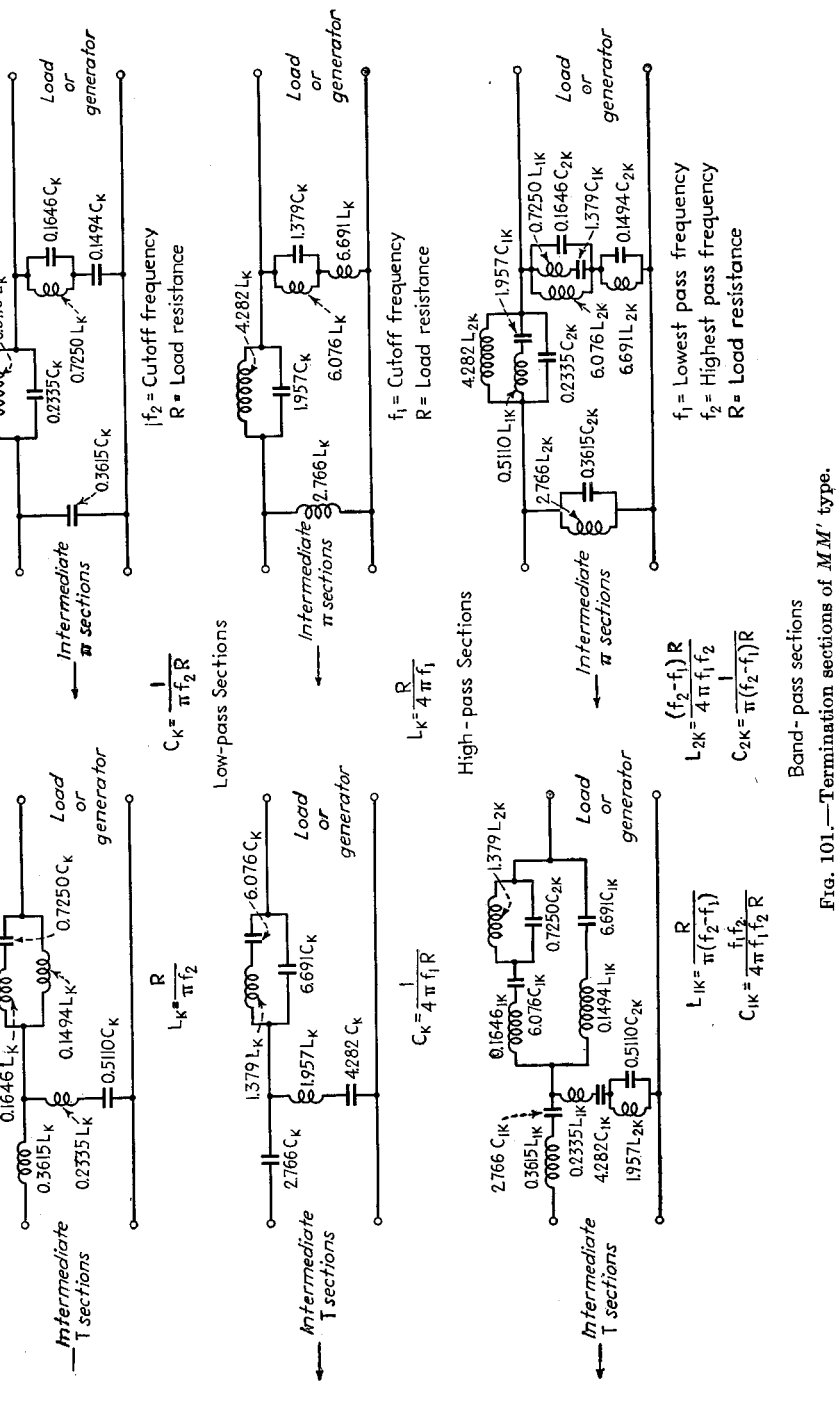


FIG. 101.—Termination sections of *M'* type.

90 per cent of the pass band, and the MM' termination gives an image impedance constant to within 2 per cent over 96 per cent of the pass band.

The MM' terminating section is characterized by having two frequencies of infinite attenuation in each stop band, as shown in Fig. 102.

It is sometimes necessary to parallel the input or the output terminals of a number of filters. In such an arrangement, each filter shunts its own image impedance across the common terminals. If one then considers the situation at a particular frequency, it is apparent that the filters that have a pass band for this frequency offer a resistance impedance, whereas the remaining filters, *i.e.*, those for which this frequency is in the stop band, provide a reactive shunting impedance that varies with frequency. This modifies the impedance presented by the combination of filters in parallel, and affects the insertion loss in an unfavorable manner.

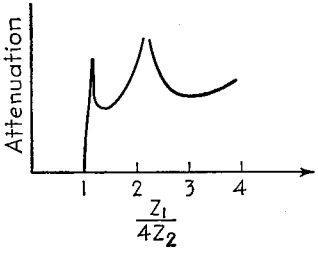


FIG. 102.—Attenuation characteristic of MM' terminating section.

This difficulty can be handled practically by employing filters in which the external terminals are shunted by a reactive arm, *i.e.*, by using a filter having T intermediate sections. The shunting reactances of all the filters in parallel then can be thought of as forming a single shunting reactance, which can then be modified as required to make the equivalent impedance seen when one is looking toward the paralleled terminals approximate a resistance in the pass band of all filters. This requires, in some instances, merely that the magnitude of the combined shunting impedance be modified in size; in other cases, an additional shunting reactive network must be placed across the common terminals.[1]

A particularly important case of filters in parallel is furnished by complementary filters, *i.e.*, where those frequencies that lie in the stop band of one filter are the pass

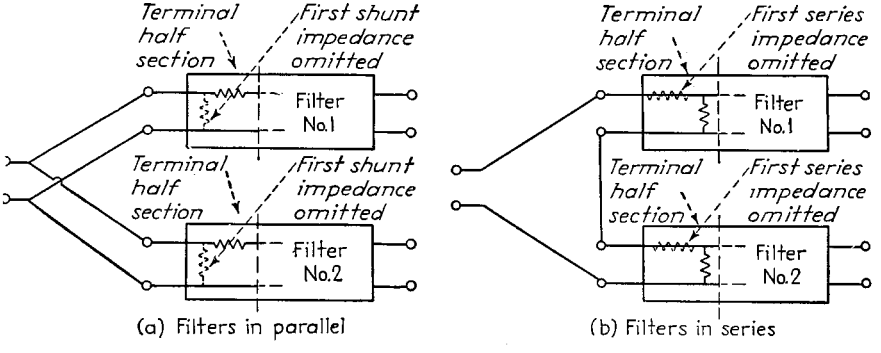


FIG. 103.—Schematic illustration of complementary filters in parallel and series (x terminations).

frequencies of the other filter. In such an arrangement, it is found that if the terminating half sections are originally both designed on the basis $m = 0.6$ and the same load resistance, then the reactive input impedance of one filter in its stop band is almost exactly equal to the reactive impedance that should be provided by the first shunting impedance of the complementary filter in its pass band, and vice versa. The result is that, if the shunting impedances at the inputs of both filters are simply

[1] Further discussion is given by Guillemin, *op. cit.*, pp. 356–366. See also O. Zobel, U.S. Patents Nos. 1,557,229 and 1,557,230.

omitted, then the normal impedance relations will be maintained at all frequencies. Such an arrangement is termed a *fractional* or *x termination*, and is illustrated in Fig. 103*a*.

Filters that are connected in series at one side, such as the input, present a problem analogous to that of filters in parallel. Here each filter in its stop band places a reactance in series with the other filter or filters. The proper method of attack here consists in employing filters that have reactances in series with the external terminals *i.e.*, filters with π intermediate sections. These series reactances are then lumped together to form a single series impedance, which is then modified as required to take into account the reactances contributed by the filters in the stop bands. In the case of complementary filters in series, one designs the terminating half sections for $m = 0.6$. The series elements of the sections are then omitted, as shown in Fig. 103*b*, and the input reactance of the filter in the stop band is used to supply the series reactance required by the complementary filter in its pass band, and vice versa.

Losses.—The ideal filter with pure reactive elements discussed above can never be fully realized in practice, for although condensers can be made with negligible losses, this is never true with inductances. Experience shows, however, that when the Q of the coils is of the order of 15 or higher, the effect of the losses is only secondary and the essential conclusions and design procedures based on an ideal filter with zero losses are not seriously invalidated.

The principal effects of a moderate amount of energy loss in a filter are:

1. A small attenuation is introduced in the pass band.

2. Frequencies for which infinite attenuation would otherwise be obtained are found to have only finite attenuation.[1]

3. The abrupt transitions occurring at the cut-off frequencies in an ideal filter are rounded off by the losses.

29. Lattice Filters.[2]—*Fundamental Relations.*—A lattice will act as a filter when the impedance arms are suitably designed reactances. The lattice structure provides the most general type of symmetrical filter section that can be devised, and includes T and π sections (including m-derived types) as special cases.

The image impedance Z_1 and image transfer constant θ of a lattice are

$$Z_I = \sqrt{Z_a Z_b} \tag{134}$$

$$\tanh\left(\frac{\theta}{2}\right) = \sqrt{\frac{Z_a}{Z_b}} \tag{135}$$

Here Z_a and Z_b are the lattice impedances, as shown in Fig. 73. Since the image impedance of a lattice depends upon the product of the two lattice impedances, whereas the image transfer constant depends upon the quotient of these impedances, *it is possible to specify the image impedance and the image transfer constant of a lattice filter independently of each other.*

A pass band is obtained in a lattice network whenever the two reactances Z_a and Z_b have opposite signs. Under these conditions the attenuation constant α is zero, the image impedance is a resistance, and the phase shift β is

$$\tan\frac{\beta}{2} = -j\sqrt{\frac{Z_a}{Z_b}} \tag{136}$$

[1] It is, however, possible in many cases to obtain infinite attenuation by addition of a resistance to the network in such a manner as to cancel out the residual transmission at this particular frequency. See Vernon D. Landon, M-derived Band-pass Filters with Resistance Cancellation, *R.C.A. Rev.*, Vol. 1, p. 93, October, 1936.

[2] H. W. Bode, A General Theory of Electric Wave Filters, *Jour. Math. & Phys.*, Vol. 13, p. 275, November, 1934; *Monograph* B-843 of Bell Telephone System; H. W. Bode and R. L. Dietzold, Ideal Wave Filters, *Bell System Tech. Jour.*, Vol. 14, p. 215, April, 1935.

The lattice filter has a stop band whenever the two reactances Z_a and Z_b have the same sign. Under these conditions, the image impedance is a pure reactance, the phase shift is either zero or π radians, and the attenuation constant α is

For $Z_b > Z_a$ ($\beta = 0$):

$$\tanh\left(\frac{\alpha}{2}\right) = \sqrt{\frac{Z_a}{Z_b}} \tag{137}$$

For $Z_b < Z_a$ ($\beta = \pi$):

$$\tanh\left(\frac{\alpha}{2}\right) = \sqrt{\frac{Z_b}{Z_a}}$$

The pass bands, cutoff frequencies, and stop bands of a lattice filter can be given the desired locations by properly arranging the zeros and poles of the lattice impedances Z_a and Z_b. The zeros and poles possessed by the impedance Z_a that lie within the pass band must coincide with poles and zeros, respectively, of impedance Z_b. At each cutoff frequency, one of the lattice impedances must contain either a pole or a zero that is not matched by a corresponding critical frequency in the other impedance. In the stop or attenuating band, zeros and poles of one of the lattice impedances must coincide, respectively, with the zeros and poles of the other lattice impedance. When these requirements are met, the impedances Z_a and Z_b will be of opposite sign throughout the pass band, but will have the same sign in the desired stop band. Arrangements of poles and zeros for several typical cases are shown in Fig. 104.

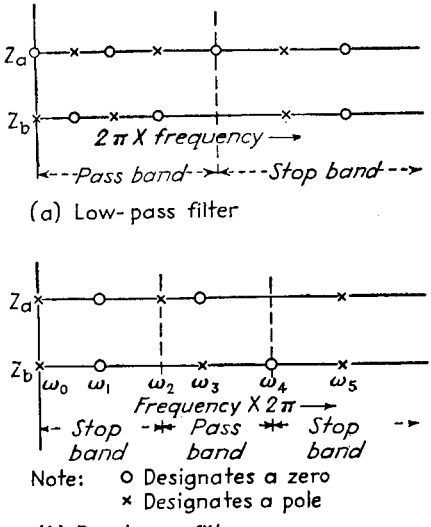


FIG. 104.—Examples of arrangements of zeros and poles in lattice filters.

It is apparent that there is a great variety of possibilities for any particular type of filter. Thus the number of critical frequencies in the pass and stop bands may be varied, as well as the position of these critical frequencies in relationship to cutoff. Likewise, the cutoff frequencies may be delineated by zeros or poles in impedance Z_a or by corresponding critical frequencies in Z_b.

Formulas for the image impedance and image transfer constant of a lattice filter can be readily derived in any particular case by substituting for the lattice impedances Z_a and Z_b that appear in Eqs. (134) and (135) the corresponding expressions for these impedances in terms of zeros and poles given by Eqs. (86*a*) and (86*b*). When such expressions are derived, it will be found that the image impedance will depend *only upon the number and location of the critical frequencies that are present in the stop band of the filter* and upon the cutoff frequencies, while the image transfer constant, *i.e.*, the phase shift in the pass band and the attenuation in the stop band, will depend *only upon the number and location of the critical frequencies that appear in the lattice impedances in the pass band* and upon the cutoff frequencies.

The procedure for setting up formulas for image transfer constant and image impedance in a particular case can be illustrated by considering Fig. 104*b*. Imped-

ances of the lattice, by Eqs. (86*a*) and (86*b*), are

$$Z_a = -j\frac{H_a}{\omega}\frac{(\omega^2-\omega_1^2)(\omega^2-\omega_3^2)}{(\omega^2-\omega_2^2)(\omega^2-\omega_5^2)} \tag{138}$$

$$Z_b = -j\frac{H_b}{\omega}\frac{(\omega^2-\omega_1^2)(\omega^2-\omega_4^2)}{(\omega^2-\omega_3^2)(\omega^2-\omega_5^2)} \tag{139}$$

Substituting these values for Z_a and Z_b into Eqs. (134) and (135) gives

$$Z_I = -\sqrt{H_aH_b}\,\frac{(\omega^2-\omega_1^2)}{\omega(\omega^2-\omega_5^2)}\sqrt{-\frac{(\omega^2-\omega_4^2)}{(\omega^2-\omega_2^2)}} \tag{140}$$

$$\tanh\frac{\theta}{2} = \sqrt{\frac{H_a}{H_b}}\,(\omega^2-\omega_3^2)\,\frac{1}{\sqrt{(\omega^2-\omega_2^2)(\omega^2-\omega_4^2)}}$$

In these equations, H_a and H_b represent the values of the constant H in Eq. (86) for the impedances Z_a and Z_b, respectively.

If the lattice impedances Z_a and Z_b are interchanged, the only effect upon the behavior of the lattice network is to reverse the polarity of the output voltage, *i.e.*, to introduce a phase shift of π radians. If *either* one of the lattice impedances (but not both) is replaced by its reciprocal impedance with respect to a constant value R, the effect is to interchange the pass and stop bands, *i.e.*, the new filter is complementary to the original filter.

Design of Lattice Filters.—The transmission characteristics of a lattice filter, which are the phase shift in the pass band and the attenuation in the stop band, are determined by the number and location of the critical frequencies within the pass band and by the ratio H_a/H_b. By employing a sufficient number of critical frequencies and properly disposing them within the pass band, it is possible to realize almost any desired phase and attenuation characteristics. Similarly, the design factors controlling the image impedance of the filter are the critical frequencies of the lattice impedances that lie in the stop band and the value of the product $\sqrt{H_aH_b}$. As before, by employing a sufficient number of such critical frequencies and properly distributing them, it is possible to realize almost any desired image-impedance characteristic.

An ideal filter would have constant image impedance throughout the pass band, a phase shift proportional to frequency within the pass band, and a very high attenuation in the stop band. It is accordingly customary to distribute the critical frequencies within the stop band in such a manner as to maintain the image impedance as nearly constant as possible over most of the pass band. In distributing the critical frequencies within the pass band, one may either give maximum weight to linearity of the phase characteristic or place the greatest emphasis upon obtaining the highest possible attenuation within the stop band. Designs that give the most linear phase-shift characteristics with a given number of critical frequencies within the pass band do not have such desirable attenuation characteristics as can be obtained with the same number of critical frequencies and some sacrifice in linearity of the phase curve.

The attenuation of a lattice filter will be infinite whenever $\sqrt{Z_a/Z_b} = 1$, and will be smaller the farther the departure of $\sqrt{Z_a/Z_b}$ from unity. Accordingly, the attenuation characteristics for a given number of critical frequencies will approach most closely the ideal when the critical frequencies within the pass band are so distributed that the value of $\sqrt{Z_a/Z_b}$ oscillates with equal amplitudes above and below the value unity, as shown in Fig. 105*a*. A systematic procedure for selecting critical frequencies to accomplish this result has been developed by Cauer and is described in some detail by Guillemin.[1] When the design parameters are selected in the Cauer

[1] See Guillemin, *op. cit.*, pp. 394–412.

manner to provide the best attenuation characteristic, the phase characteristic is far from linear.

The image impedance of a lattice filter will be constant to the extent that the quantity $\sqrt{Z_a Z_b}/R_L$ approximates unity, where R_L is the load resistance. This ideal is most closely realized for a given number of critical frequencies when $\sqrt{Z_a Z_b}/R_L$ oscillates about the value unity with equal positive and negative deviations, as shown

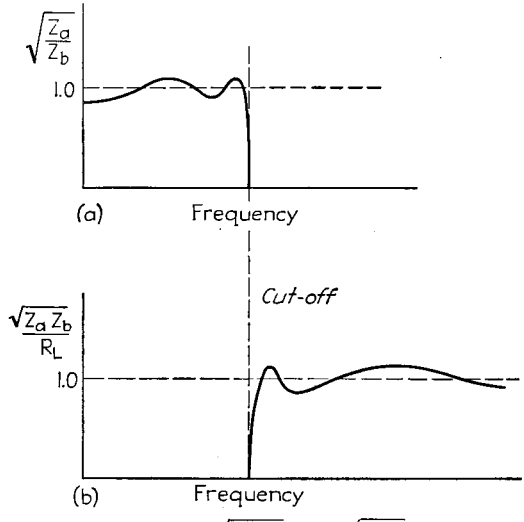


FIG. 105.—Variation of design factors $\sqrt{Z_a/Z_b}$ and $\sqrt{Z_a Z_b}/R_L$ corresponding to optimum attenuation, and optimum characteristic impedance characteristics.

in Fig. 105*b*. The procedure for locating the critical frequencies within the stop band in order to accomplish this is similar in every respect to the procedure for locating the critical frequencies in the pass band to provide an oscillatory approximation to the ideal attenuation characteristic. The detailed steps are given by Cauer and Guillemin.

A phase-shift characteristic that is substantially linear over most of the pass band, combined with an attenuation characteristic that rises rapidly beyond cutoff, can be obtained by locating the critical frequencies within the pass band according to the

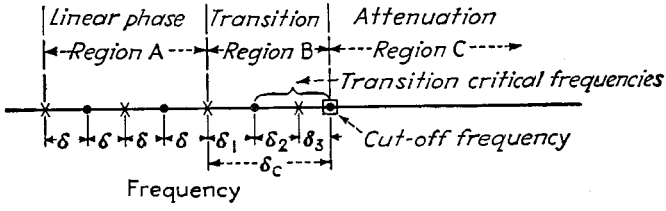


FIG. 106.—Distribution of critical frequencies in a Bode-type lattice filter.

manner devised by Bode,[1] and illustrated in Fig. 106. Here the pass band is divided into two parts, region A and region B. Region A, which comprises most of the pass band, is to have a substantially linear phase-shift characteristic; region B is a transition region in which the phase shift is not linear. In region A, critical frequencies are located at a uniform frequency interval δ as shown. Region B is provided with at

[1] Bode and Dietzold, *loc. cit.*

TABLE 5.—SPACING INTERVALS*

Number of transition frequencies	δ_1/δ	δ_2/δ	δ_3/δ	δ_4/δ	δ_5/δ	δ_c/δ
1	0.500	.....	.....	.....	.....	0.500
2	0.853	0.354	.....	.....	.....	1.207
3	0.950	0.727	0.278	.....	.....	1.955
4	0.981	0.880	0.630	0.231	.....	2.723
5	0.992	0.946	0.810	0.555	0.197	3.501

* See Fig. 106.

least one critical frequency, corresponding to cutoff, and often two or more critical frequencies. In order to obtain the most favorable phase characteristic in region A, these critical frequencies in the transition region B must be located in accordance with Table 5. When this is done, the phase characteristic will be almost perfectly linear in region A, provided that there are not too few critical frequencies in region A. The associated attenuation characteristic is as shown in Fig. 107, and rises with a rapidity that increases as more transition factors are employed in region B. The attenuation in a Bode design never goes to infinity, although it quickly becomes quite large.

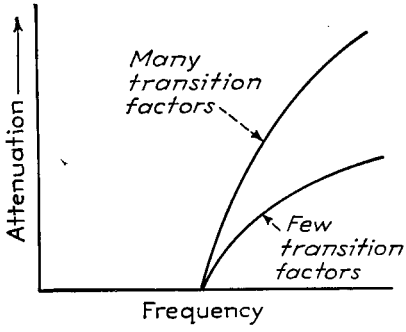


FIG. 107.—Attenuation characteristics of Bode-type lattice filter.

Relation between Ladder and Lattice Structures.—Any symmetrical T or π section can be converted to an equivalent lattice section by so selecting the lattice impedances that the open- and short-circuit impedances of the lattice are the same as for the T (or π) arrangement. The transformations necessary to accomplish this are illustrated in Fig. 108.

The reverse transformation, *i.e.*, the representation of a lattice network by a symmetrical T (or a symmetrical π) will not necessarily lead to a physically realizable T or π. Examination of Fig. 108 shows that the conversion to a symmetrical T will result in a realizable structure only when it is possible to subtract the series arm Z_a of the lattice from the lattice diagonal and have a physically realizable remainder. Similarly, the conversion from a lattice to a symmetrical π will result in a physical network only when it is physically possible to subtract the admittance $1/Z_b$ of the diagonal lattice impedance from the admittance of the series impedance arm and have a physically realizable remainder.

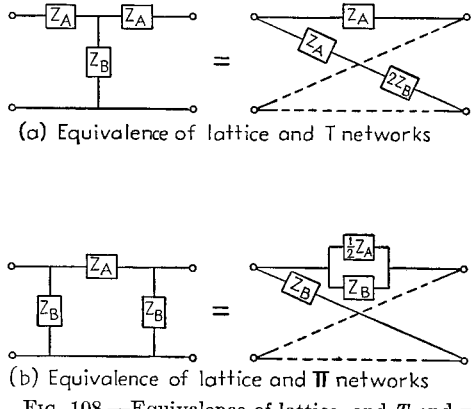
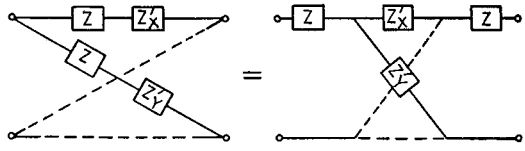
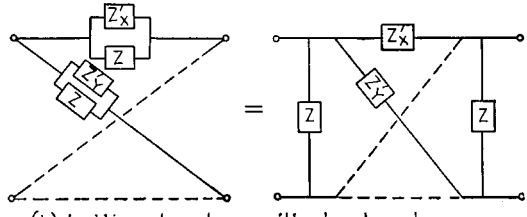


FIG. 108.—Equivalence of lattice, and T and π networks.

A series impedance that is common to both arms of a lattice can be removed from the lattice and placed in series with the external terminals, as shown in Fig. 109*a*. The resulting combination of series impedance and modified lattice has exactly the same properties as the original lattice. Similarly, an impedance that is in shunt with both arms of a lattice may be removed from the lattice and placed in shunt across the external terminals, as shown in Fig. 109*b*. The validity of these transformations is



(a) Lattice structure with developed series impedances



(b) Lattice structure with developed shunt impedances

FIG. 109.—Development of lattice network into lattice plus series (or shunt) impedances.

based on the fact that the open- and short-circuit impedances possessed by the modified lattice and its external series (or shunt) impedances are exactly the same as the corresponding open- and short-circuit impedances of the original lattices.

Successive applications of these transformations can be used to transform a lattice having complicated impedance arms into a ladder network of alternate series and shunt impedances in association with a simpler lattice. In some cases, the

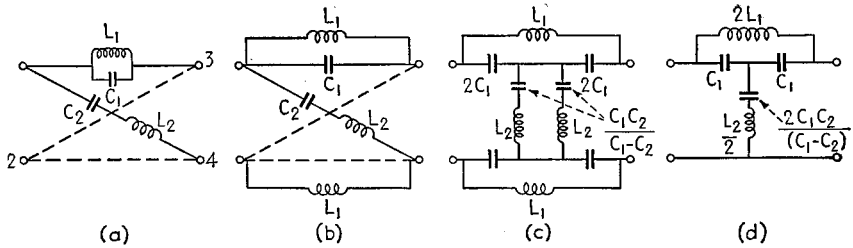


FIG. 110.—Development of a lattice which can be transformed into a bridged-T network.

residual lattice degenerates into a pair of shunt arms in parallel, or into two series arms, in which case the conversion from the original lattice to a ladder network is complete.

A lattice that cannot be converted to a ladder network can sometimes still be transformed into a bridged-*T* section. This is illustrated in Fig. 110, where by considering inductances L_1 as bridging the terminals 1-3 and 2-4, the lattice reduces to

arms C_1 and C_2L_2, and when $C_1 > C_2$ can be developed into a ladder network, shown at Fig. 110*d*, that, when bridged by the inductance $2L_1$, is equivalent to the original lattice.

Comparison of Lattice and Ladder Filters.—The lattice structure provides the most general form of symmetrical section that it is possible to realize, and so provides the greatest possible choice of characteristics that can be obtained in a filter. At the same time, the lattice generally requires more coils and condensers than does the corresponding *m*-derived type of filter, and does not provide an input and output terminal at a common ground potential unless the lattice can be fully developed into a ladder or bridged-*T* network. It is found that for most communications purposes calling for filters, the *m*-derived structure is quite adequate.

30. Equalizers.—An equalizer is a network placed between the generator and load such that the current that the generator produces in the load will vary in some desired manner with frequency. An *attenuation* equalizer is an equalizer that is inserted in order to control the magnitude of the load current as a function of frequency, without any particular regard to phase relations. A *phase* equalizer is an all-pass filter designed to introduce a desired phase shift as a function of frequency in the load current.

For most purposes it is sufficient to equalize only for attenuation. If, however, circumstances require phase as well as attenuation equalization, the procedure followed is first to equalize for attenuation and then afterward to add a phase equalizer to the system to correct for any undesired features in the phase characteristic of the original system plus the attenuation equalizer.

Attenuation Equalizers.[1]—The networks commonly used as attenuation equalizers are shown in Fig. 111, which also gives the insertion loss of these networks when the load impedance R_L has the design value $R_L = R_0$.

These networks are conveniently divided into classes as indicated. The simple series and shunt equalizing networks designated as Type I find some use because of their simplicity, but have the very serious disadvantage of causing the impedance seen by the generator in looking toward the load, and also the impedance seen by the load in looking toward the generator, to depend upon the amount of attenuation introduced by the equalizer.

The *L* equalizers, designated as Type II, partially overcome this disadvantage of the Type I equalizer, since when $Z_1Z_2 = R_0^2 = R_L^2$, *i.e.*, by making Z_1 and Z_2 reciprocal networks with respect to $R_0 = R_L$, the impedance on the input side of the equalizer will be a constant resistance equal to R_L, irrespective of the amount of attenuation introduced. The resistance as seen by the load when one looks toward the generator will, however, depend upon the attenuation.

The Type III equalizer networks are the types most widely used in practical work. They are characterized by the fact that if the impedances Z_1 and Z_2 are reciprocal networks with respect to the resistance R_0, *i.e.*, if $Z_1Z_2 = R_0^2$, then the equalizer has both image impedances equal to R_0 at all frequencies. Such a network is said to be of the *constant-resistance* type. Insertion of such an equalizer in a system therefore does not disturb the impedance relations, provided that the load impedance matches the image impedance of the equalizer.

The equalizers indicated as Type IV represent more general forms of bridged-*T* networks of the constant-resistance type, and Type V represents a general lattice network of the constant-resistance type. The networks IIIc and III*d* are special cases of the more general Types IV*a* and V, respectively.

[1] Much of the material given here is based upon Chap. XVI of "Motion Picture Sound Engineering," Van Nostrand, New York, 1938.

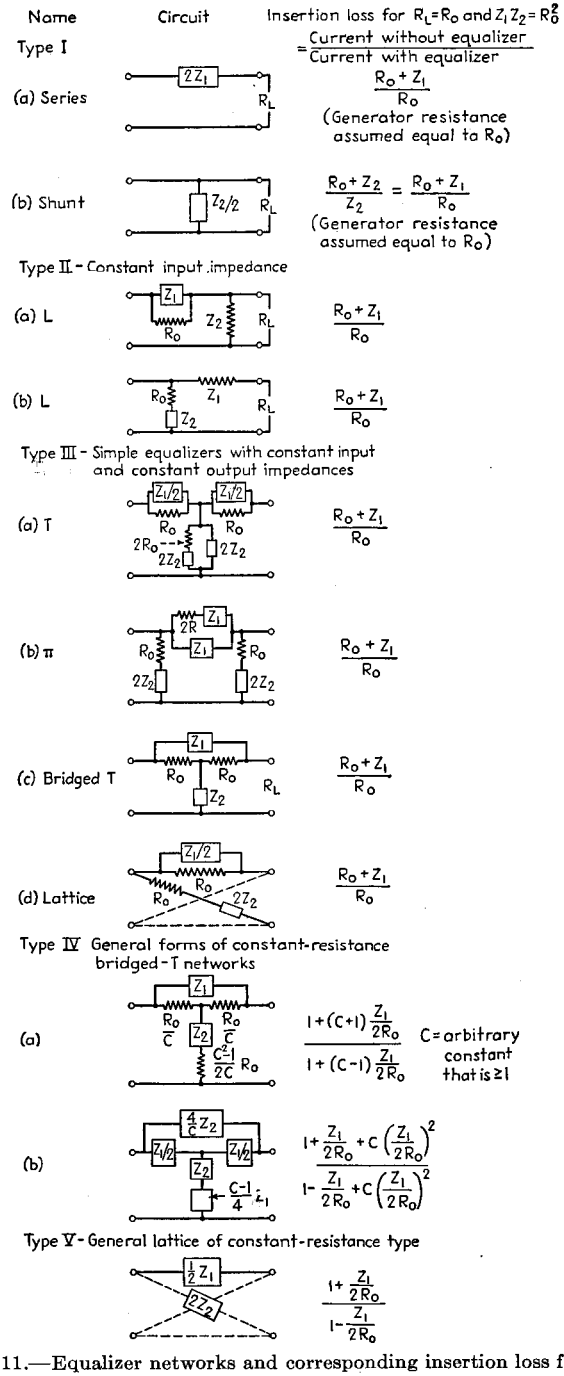


FIG. 111.—Equalizer networks and corresponding insertion loss formula.

Design of Attenuation Equalizing Networks.[1]—The most important equalizers from a practical point of view are the Type III networks on Fig. 111, with the Type II sections also finding some application. The attenuation formulas are the same for

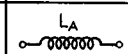
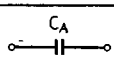
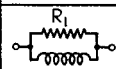
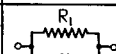
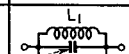
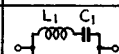
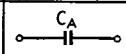
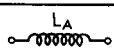
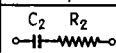
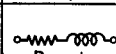
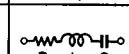
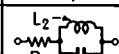
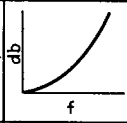
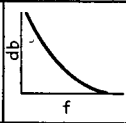
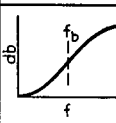
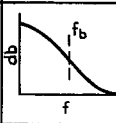
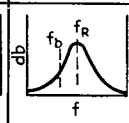
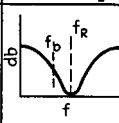
	a	b	c	d	e	f
Z_1 network						
Z_2 network						
Insertion loss characteristic in db						
Insertion loss in db	$10\log_{10}\left[1+\left(\frac{f}{f_a}\right)^2\right]$	$10\log\left[1+\left(\frac{f_a}{f}\right)^2\right]$	$10\log_{10}\left[1+\frac{K^2-1}{1+K\left(\frac{f_b}{f}\right)^2}\right]$	$10\log_{10}\left[1+\frac{K^2-1}{1+K\left(\frac{f}{f_b}\right)^2}\right]$	$10\log_{10}\left[1+\frac{K^2-1}{1+K\left(\frac{f}{f_R}\frac{f_R}{f}\right)^2}\right]$	$10\log_{10}\left[1+\frac{K^2-1}{1+K\left(\frac{b-\frac{1}{b}}{\frac{f}{f_R}-\frac{f_R}{f}}\right)^2}\right]$
Design formulas	$L_A=\frac{R_0}{2\pi f_a}=\frac{R_0}{\omega_a}$ $C_A=\frac{1}{2\pi f_a R_0}=\frac{1}{\omega_a R_0}$ $f_a=\frac{1}{2\pi\sqrt{L_A C_A}}$ $R_0=\sqrt{\frac{L_A}{C_A}}$		$L_1=L_B\frac{K-1}{\sqrt{K}}$ $L_2=L_B\frac{\sqrt{K}}{K-1}$ $C_1=C_B\frac{\sqrt{K}}{K-1}$ $C_2=C_B\frac{K-1}{\sqrt{K}}$		$L_1=L_B\frac{K-1}{\sqrt{K}}\frac{b^2-1}{b^2}$ $L_2=L_B\frac{\sqrt{K}}{K-1}\frac{1}{b^2-1}$ $C_1=C_B\frac{\sqrt{K}}{K-1}\frac{1}{b^2-1}$ $C_2=C_B\frac{K-1}{\sqrt{K}}\frac{b^2-1}{b^2}$	$L_1=L_B\frac{K-1}{\sqrt{K}}\frac{1}{b^2-1}$ $L_2=L_B\frac{\sqrt{K}}{K-1}\frac{b^2-1}{b^2}$ $C_1=C_B\frac{\sqrt{K}}{K-1}\frac{b^2-1}{b^2}$ $C_2=C_B\frac{K-1}{\sqrt{K}}\frac{1}{b^2-1}$
			$R_0=\sqrt{\frac{L_B}{C_B}}$ $\quad R_1=R_0(K-1)$ $\quad R_2=R_0\frac{1}{K-1}$ $L_B=\frac{R_0}{2\pi f_b}=\frac{R_0}{\omega_b}$ $\quad C_B=\frac{1}{2\pi f_b R_0}=\frac{1}{\omega_0 R_0}$ $\quad f_b=\frac{1}{2\pi\sqrt{L_B C_B}}$			
Notes	f_R = Resonant frequency of Z_1 and Z_2 arms f_a = Frequency of 3db insertion loss f_b = Frequency where loss is one half maximum value (loss measured in db) f = Any frequency $b=\frac{f_R}{f_b}$ = defined as greater than unity		R_0 = Equalizer resistance Pad loss = maximum loss = $20\log_{10}K$ L = Inductance in henries C = Capacity in farads			

FIG. 112.—Design information for fundamental equalizer types.

Types I, II, and III, so that the insertion loss obtained with any one type can be duplicated in the other types by using the same R_0 and Z_1 (or Z_2). The choice between types is accordingly based on convenience in construction, cost, and the importance of maintaining the impedance relations.

[1] An excellent set of design curves is given in Chaps. XVI and XVII of "Motion Picture Sound Engineering," Van Nostrand, New York, 1938.

The bridged-T structure designated as Type IIIc in Fig. 111 is the most widely used unbalanced structure; while if symmetry with respect to ground is required, Types IIIc and IIId networks are best. The bridged T is superior to the Types IIIa and IIIb sections, since it requires fewer reactive elements. Compared with the Type II L sections, the bridged T requires only one additional fixed resistance, and, in return, gives a constant resistance irrespective of attenuation from both sides, instead of from only one side.

The types of attenuation characteristics obtainable with Type II and Type III sections when Z_1 has simple configurations are shown in Fig. 112. When Z_1 is supplied by an inductance, the attenuation rises at high frequencies, while if it is a

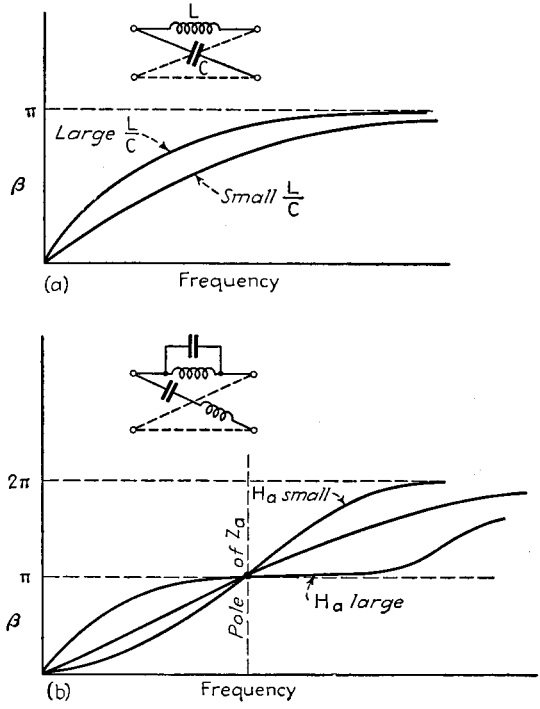


FIG. 113.—Examples of phase characteristics of simple all-pass sections (phase equalizers).

capacity, as in Fig. 112b, the attenuation rises at low frequencies. If the inductance is shunted with a resistance, as in Fig. 112c, the attenuation at high frequencies first rises and then levels off to a limiting value determined by the shunting resistance. Similarly, a resistance in shunt with a capacity as in Fig. 112d causes a corresponding leveling off of the rise in attenuation that would otherwise occur at low frequencies. Constructing the impedance Z_1 as a parallel resonant circuit shunted by a resistance, as in Fig. 112e, results in an attenuation peak at the parallel resonant frequency, with the magnitude of the peak determined by the shunting resistance. Such an arrangement is sometimes referred to as a "dip pad." When the impedance Z_1 is supplied by a series resonant circuit shunted by a resistance, as in Fig. 112f, the result is an attenuation at low and high frequencies determined by the shunting resistance, with negligible attenuation at the series resonant frequency.

Formulas arranged in convenient form for designing equalizers having simple impedance configurations for Z_1 are given in Fig. 112 for Types I*a*, II, and III sections.

The simple Type I, II and III equalizer sections are capable of meeting most practical requirements. This is true even when relatively complicated equalization characteristics are desired, since one may break up the required attenuation characteristic into the sum of several simpler characteristics, and then obtain each simpler attenuation characteristic from a Type I, II, or III section with a relatively simple Z_1.

The general lattice network (Type V) is the most general form of symmetrical equalizer that can be devised. All other symmetrical equalizers, such as Types I*a*, III, and IV, are equivalent to special cases of the general lattice. It is accordingly possible to base the design of all equalizers upon the general lattice even though the

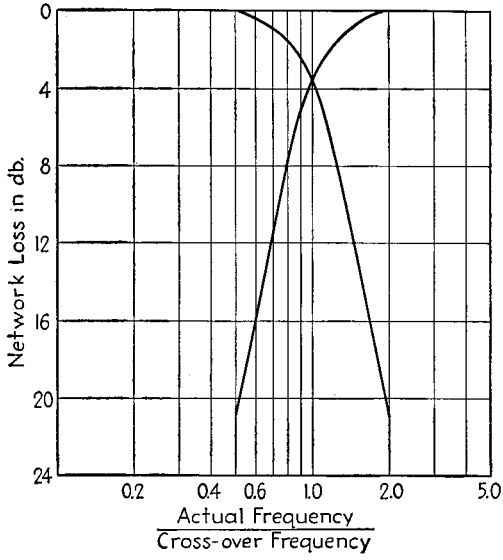


FIG. 114.—Characteristics of a typical dividing network.

network may be built in the form of a bridged T where this conversion leads to a physically possible structure.[1]

Phase Equalizers. All-pass Filters.—An all-pass filter is a filter having zero attenuation for all frequencies from zero to infinity. Such filter sections introduce phase shift without affecting attenuation, and so are employed as phase equalizers to correct phase distortion introduced by other parts of a system, and can also be used to introduce a time delay.

All-pass action can be obtained by making use of a lattice network in which the lattice impedances Z_a and Z_b are reactances that are reciprocal with respect to the desired image impedance R. This leads to an image impedance that is constant at the desired value for all frequencies.

The phase shift β introduced by an all-pass section under image-impedance operation is

[1] A discussion of this approach to equalizer design, together with design information on a considerable variety of configurations for the lattice impedances, is given by Otto J. Zobel, Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks, *Bell System Tech. Jour.*, Vol. 7, p. 438, April, 1928. A very clear discussion of the theoretical basis of the design method developed by Zobel for the general lattice equalizer is given by Everitt, *op. cit.*, pp. 287–293.

$$\tan \frac{\beta}{2} = \pm j \sqrt{\frac{Z_a^2}{R^2}} \tag{141}$$

where R is the desired image impedance.

The variation of the phase shift β with frequency can be controlled by the number and location of the internal zeros and poles in the impedance Z_a, and by the value assigned to the quantity H in the expression for Z_a given by Eq. (86). Typical phase-shift characteristics for several simple cases are illustrated in Fig. 113.

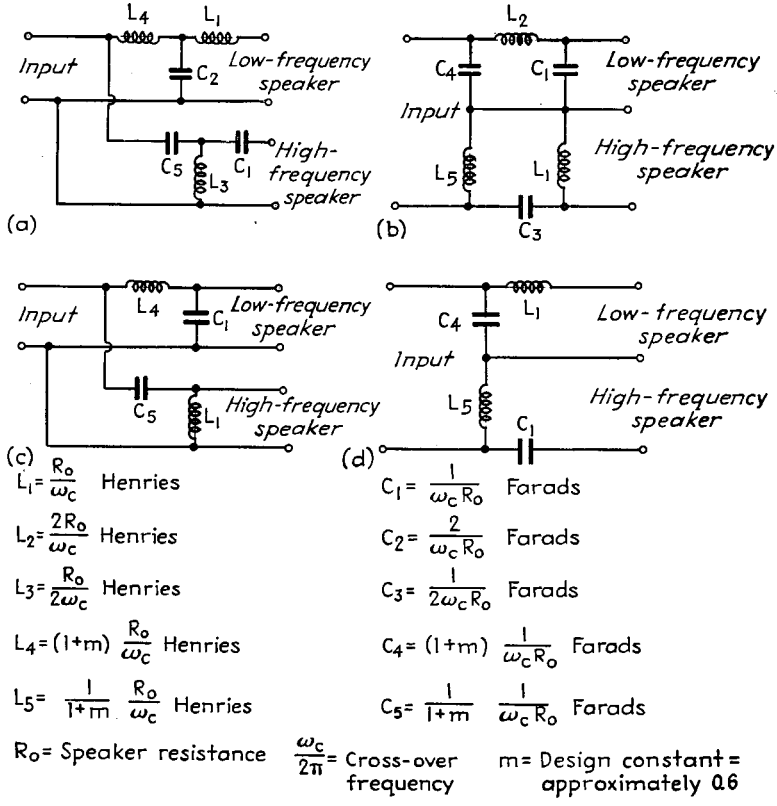


FIG. 115.—Dividing networks of the filter type.

31. Dividing Networks.[1]—The term *dividing network* is applied to a coupling system so arranged that at low frequencies power is delivered to a low-frequency loud-speaker, while at high frequencies it is delivered to a high-frequency speaker. The transmission characteristics of a typical dividing network are shown in Fig. 114. The frequency at which the power delivered to the two outputs is equal is termed the *cross-over* frequency. Experience indicates that the dividing network should provide at least twelve db attenuation one octave away from the cross-over frequency, as compared with the cross-over attenuation, whereas attenuations of more than eighteen db per octave are not necessary or desirable.

[1] See Chap. XX, "Motion Picture Sound Engineering," Van Nostrand, New York, 1938, John K. Hilliard, Loud Speaker Dividing Networks, *Electronics*, Vol. 14, p. 26, January, 1941.

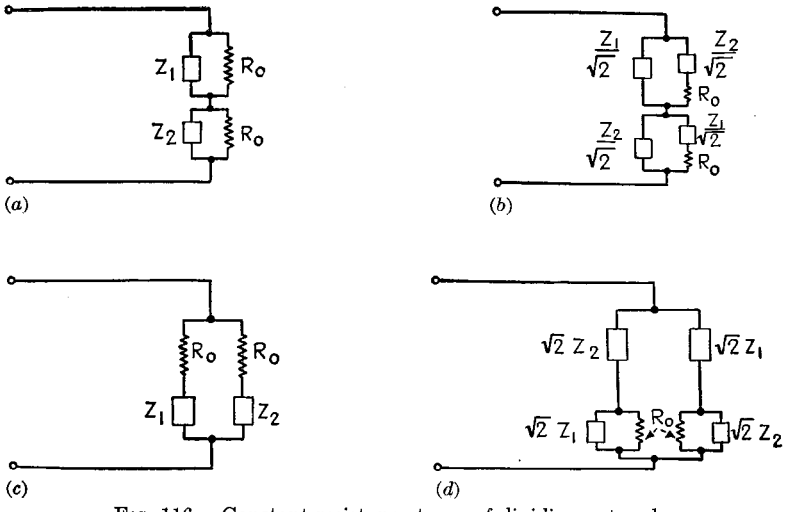


FIG. 116.—Constant-resistance types of dividing network.

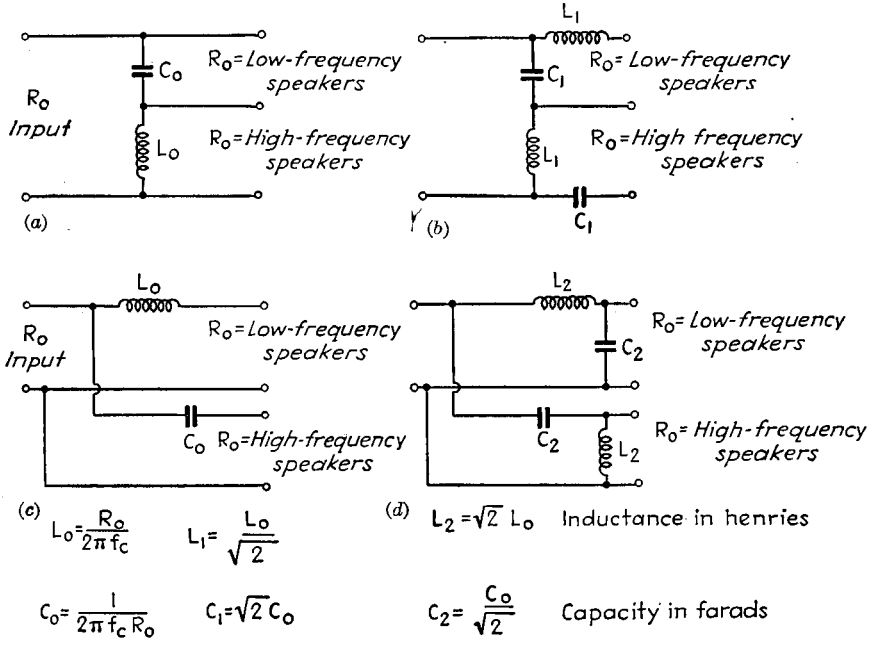


FIG. 117.—Practical constant-resistance dividing networks.

There are two basic types of dividing networks. The first consists of complementary low- and high-pass filters connected with inputs either in parallel or in series. Such networks are shown in Fig. 115, which also gives the necessary design formulas. The different parts are designed as complementary low- and high-pass filters with fractional terminations. The networks shown at (*a*) and (*b*) in Fig. 115 consist of one full section of the $m = 1$ network type, with an input half section designed for $m \cong 0.6$. The input shunting impedance called for in the design of (*a*) and the series input impedance called for in the design of (*b*) are then omitted to provide the fractional termination. Such an arrangement provides an attenuation of approximately 18 db for the first octave beyond cross-over. Sections shown at Fig. 115*c* and 115*d* differ from those at (*a*) and (*b*) in that the final half section of the $m = 1$ network type is omitted. This reduces the attenuation to approximately 12 db for the first octave beyond the cross-over frequency. It would be possible to design the dividing network so that the output terminals were also provided with an $m = 0.6$ termination, but this is never done, because satisfactory performance can be obtained with the simpler arrangement.

Dividing networks of another kind, referred to as the *constant-resistance* type, are shown in Fig. 116. These are characterized by possessing an input impedance that is a constant resistance equal to R_0 when the two impedances Z_1 and Z_2 are so related that $Z_1Z_2 = R_0^2$. Practical dividing networks making use of these circuit arrangements are shown in Fig. 117, together with the necessary design formulas. The simple circuits (*a*) and (*c*) have an attenuation of approximately 6 db for the first octave beyond cutoff, and since this is smaller than desirable, such arrangements are seldom used. The two-element dividing networks of (*b*) and (*d*) are similar to those of Figs. 115*d* and 115*c*, respectively, but because of slightly different circuit proportions provide an exactly constant input resistance, whereas the circuits of Fig. 115 only approximate a constant resistance. The attenuation provided by the networks of Figs. 116*b* and *d* is about 10 db for the first octave beyond cutoff.

Dividing networks carry the full output power that the amplifier operating the loud-speaker system is capable of developing, and so must be designed to have a low transmission loss. When care is taken to employ coils of the lowest practical resistance, this loss is of the order of 0.5 db in systems providing from 12 to 18 db per octave of attenuation.

WAVES GUIDES AND RESONATORS[1,2]

32. Fundamental Concepts of Waves Guides.—The term wave guide refers to a hollow conducting tube used for the transmission of electromagnetic waves. At very high frequencies wave guides provide an alternative to the ordinary transmission line, and are comparable with or superior to a concentric line.

[1] This presentation of wave guides and cavity resonators is based on material prepared by the author's associate, Karl Spangenberg.

[2] The practical possibilities of wave guides as transmission systems for very high-frequency waves was discovered independently and almost simultaneously by W. L. Barrow and G. C. Southworth. The fundamental papers on the subject are as follows: W. L. Barrow, Transmission of Electromagnetic Waves in Hollow Tubes of Metal, *Proc. I.R.E.*, Vol. 24, p. 1298, October, 1936; G. C. Southworth, Hyper-frequency Wave Guides—General Considerations and Experimental Results, *Bell System Tech. Jour.*, Vol. 15, p. 284, April, 1936; Some Fundamental Experiments with Wave Guides, *Proc. I.R.E.*, Vol. 25, p. 807, July, 1937; John R. Carson, Sallie P. Mead, and S. A. Schelkunoff, Hyper-frequency Wave Guides—Mathematical Theory, *Bell System Tech. Jour.*, Vol. 15, p. 310, April, 1936; S. A. Schelkunoff, Transmission Theory of Plane Electromagnetic Waves, *Proc. I.R.E.*, Vol. 25, p. 1457, November, 1937; L. J. Chu and W. L. Barrow, Electromagnetic Waves of Hollow Metal Tubes of Rectangular Cross Section, *Proc. I.R.E.*, Vol. 26, p. 1520, December, 1938; L. J. Chu, Electromagnetic Waves in Elliptic Hollow Pipes of Metal, *Jour. Applied Phys.*, Vol. 9, p. 583, September, 1938.

Physical Picture of Wave-guide Propagation.—A wave in order to exist in a wave guide must satisfy Maxwell's equations throughout the wave guide. It must also satisfy the boundary conditions, which are that to the extent that the walls of the wave guide are perfect conductors, there can be no tangential component of electric field at the walls of the wave guide. An exact solution for the field existing within a wave guide is therefore a relatively complicated mathematical expression.

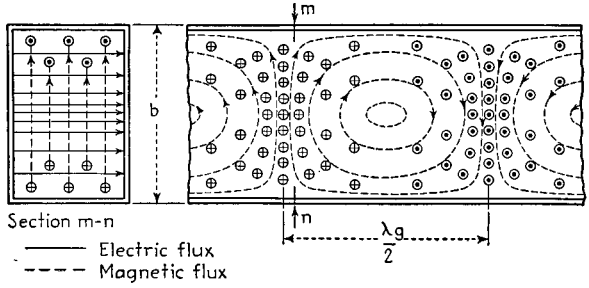


FIG. 118.—Typical field arrangement existing in a rectangular wave guide ($TE_{0,1}$ wave).

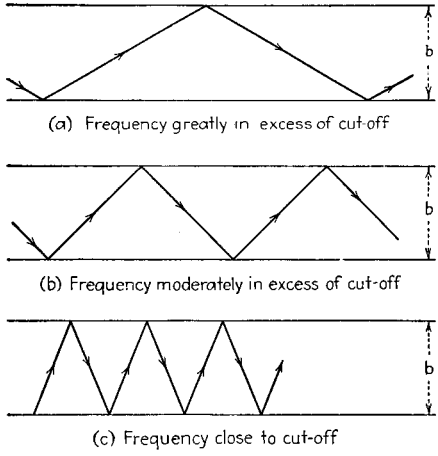


FIG. 119.—Paths followed by waves traveling back and forth between the walls of a wave guide for frequencies exceeding the cutoff frequency by varying amounts.

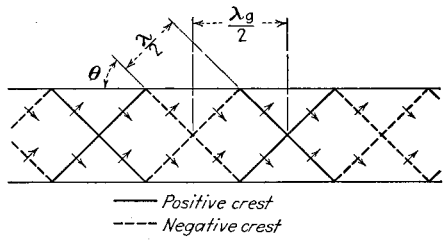
It is possible, however, to obtain an understanding of many of the properties of wave-guide propagation from a simple physical picture of the mechanisms involved.[1] Consider the fields shown in Fig. 118, which correspond to a typical kind of wave propagation in a rectangular guide. These can be considered as the resultant fields produced by an ordinary plane electromagnetic wave that travels back

[1] This is due to L. Page and N. I. Adams, Electromagnetic Waves in Conducting Tubes, *Phys. Rev.* Vol. 52, p. 647, September, 1937; Léon Brillouin, Propagation of Electromagnetic Waves in a Tube, *Rev. gén élec.*, Vol. 40, p. 227, August, 1936. See also H. H. Skilling, "Fundamentals of Electric Waves," Wiley, 1942

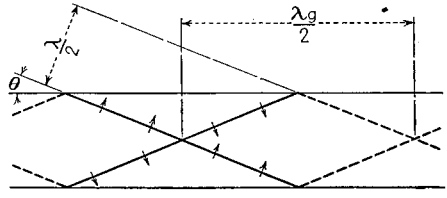
and forth between the sides of the guide, as illustrated in Fig. 119. The electric and magnetic component fields of this plane wave are in time phase, but are geometrically at right angles to each other and to the direction of propagation. Such a wave travels with the velocity of light, and upon encountering the conducting walls of the guide is reflected with a reversal of the electric field and with an angle of reflection equal to the angle of incidence.

A picture of the wave fronts involved in such propagation in a rectangular wave guide is shown in Fig. 120. When the angle θ is such that the successive positive and negative crests traveling in the same direction just fail to overlap inside the guide, as is the case in Fig. 120*a*, then it can be shown that the summation of the various waves and their reflections leads to the field distribution illustrated in Fig. 118, which travels down the wave guide and represents propagation of energy. The angle θ that the component waves must have with respect to the wave guide in order to satisfy the conditions for wave-guide propagation in a rectangular guide is given by the relation

$$\cos \theta = \frac{\lambda}{2b} \qquad (142)$$



(a) Wave fronts corresponding to Fig. 119 b



(b) Wave fronts corresponding to Fig. 119 c

FIG. 120.—Wave front corresponding to the situation illustrated in Figs. 119*b* and 119*c*.

where b is the width of the wave guide (see Fig. 118) and λ is the wave length of the wave on the basis of the velocity of light.

Because of the fact that the component waves that can be considered as building up the actual field in the wave guide all travel at an angle with respect to the axis of the guide, the rate at which energy propagates down the guide is less than the velocity of light. This velocity with which energy propagates is termed the *group velocity* and in the case of Fig. 120 is given by the relation

$$\frac{\text{Group velocity}}{\text{Velocity of light}} = \sin \theta = \sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} \qquad (143)$$

At the same time, the fields of the component waves combine in such a manner that the distance λ_g, as shown in Fig. 120, is greater than the wave length λ of the waves in free space. The result is that the fields in the wave guide possess an *apparent* or *phase velocity* that is greater than the velocity of light. Thus one has

$$\frac{\text{Wave length in guide}}{\text{Wave length in free space}} = \frac{\lambda_g}{\lambda} = \frac{1}{\sin \theta} \qquad (144)$$

$$\frac{\text{Phase velocity in guide}}{\text{Velocity of light}} = \frac{\lambda_g}{\lambda} = \frac{1}{\sqrt{1 - \left(\frac{\lambda}{2b}\right)^2}} \qquad (145)$$

The notation is the same as used above. The phase velocity is of particular importance as it gives the phase shift β per unit length that a wave suffers when traveling

down the guide. The relationship is

$$\beta = \frac{2\pi}{\lambda_g} \tag{146}$$

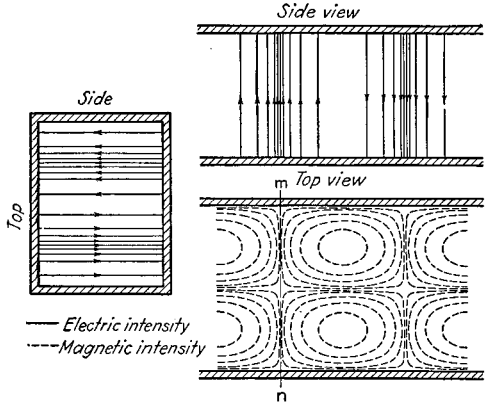
The phase and group velocities v_p and v_g, respectively, are related by the fact that

$$\sqrt{v_p v_g} = \text{velocity of light} \tag{147}$$

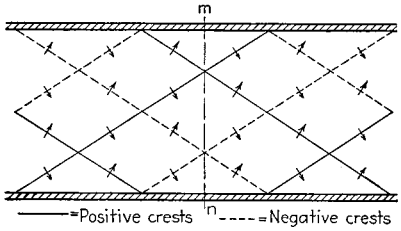
As the wave length is increased, the component waves must travel more nearly at right angles to the axis of the wave guide, as shown in Fig. 120*b*. This causes the group velocity to be lowered, and the phase velocity to be still greater than the velocity of light, until finally one has $\theta = 90°$. The component waves then bounce back and forth across the wave guide at right angles to its axis, and do not travel down the tube at all. Under these conditions the group velocity is zero, the phase velocity becomes infinite, and propagation of energy ceases. This frequency is termed the *cutoff frequency*. The wave length corresponding to cutoff is given in the case of Fig. 118 by the relation

$$\text{Wave length for cutoff} = 2b \tag{148}$$

where b is the width of the guide (see Fig. 118). It will be noted that the wave guide acts as a high-pass filter with the cutoff frequency determined by the dimensions of the guide. The fact that in order to obtain propagation, the wave guide must have a dimension comparable to a wave length limits the practical use of wave guides to extremely high frequencies.



(a) Second-order $TE_{0,2}$ wave in rectangular guide



(b) Wave fronts corresponding to (a)

FIG. 121.—Second-order mode in a rectangular wave guide.

When the frequency is made very high, the mode of propagation shown in Fig. 118 corresponds to a component wave traveling almost parallel to the axis of the guide (see Fig. 119*a*), so that the group and phase velocities both approach the velocity of light.

At frequencies very much greater than the cutoff frequency, it is possible for higher order modes of transmission to exist in a wave guide. Thus if the frequency is high enough, propagation of energy can take place down the guide when the system of component waves that are reflected back and forth between the walls has the character shown in Fig. 121*a*. This leads to fields within the guides illustrated in Fig. 121*b*, which is a second-order mode of propagation. This second mode corresponds to a field distribution equivalent to two distributions of the fundamental mode placed side by side, but with reversed polarity. Higher order modes of propagation can exist whenever the frequency is high with respect to the cutoff frequency, or conversely, higher order modes can exist for a given frequency if the dimensions of the wave guide are unnecessarily large.

This concept that has been presented of wave-guide propagation as involving a wave suffering successive reflections between the sides of the guide can be applied to all types of waves and to other than rectangular guides. The way in which the concept works out in these other cases is normally not as simple, however, as in the wave of Fig. 118.

Transmission Modes and Their Classification.—Waves of a variety of types may propagate down a wave guide. These include several types of fundamental modes, such as illustrated in Fig. 122 and Fig. 124, together with higher orders of each of these modes shown in Fig. 123.

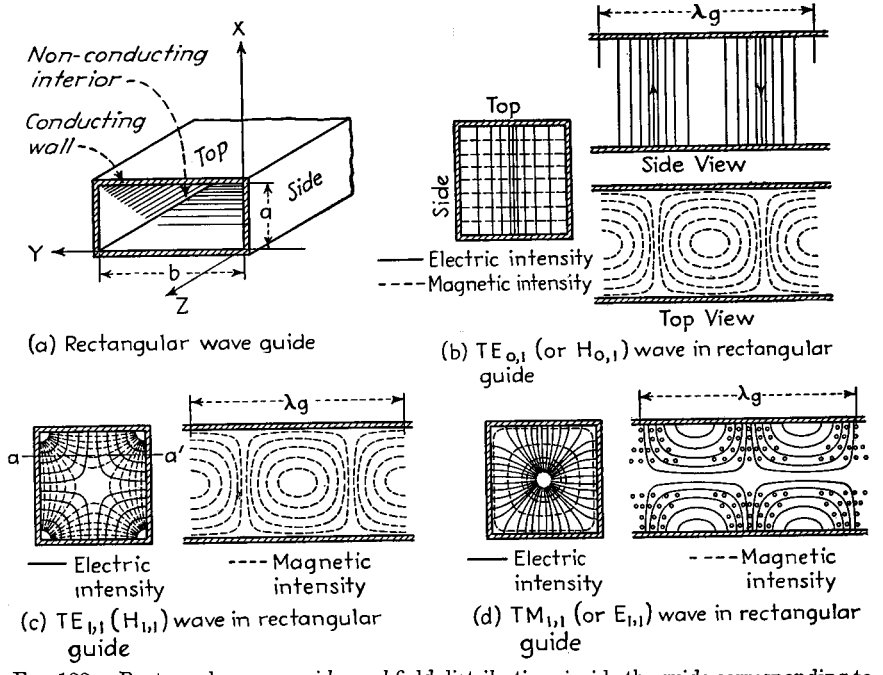


FIG. 122.—Rectangular wave guide, and field distributions inside the guide corresponding to fundamental forms of wave-guide transmission.

A study of the fields that can exist inside a closed tube shows that if the electric field has a component in the direction of propagation then the magnetic field does not. The converse of this statement is also true. Hence all modes of transmission may be put into two groups according to the axial component of field that exists. Modes of transmission having a component of electric field in the direction of propagation but no component of magnetic field in that direction are known as *E* waves or *TM* (transverse magnetic) waves. Modes of transmission having a component of magnetic field in the direction of propagation but no component of electric field in that direction are known as *H* waves or *TE* (transverse electric) waves. All possible modes that can exist in any type of hollow guide can be placed in one of these two groups. The designation in terms of the transverse field is preferred.[1]

[1] Memorandum on the Terminology of Guided Waves, Committee on Radio Wave Propagation of the Institute of Radio Engineers, Conference of Jan. 23, 1940.

33. Properties of Guides.[1] *Rectangular Wave Guides.*[2]—The three simplest modes of wave-guide propagation that can take place in rectangular guides are shown in Fig. 122. Examples of higher order modes of these are given in Fig. 123.

A double-subscript notation is used to designate the various TE and TM waves, leading to $TE_{0,1}$, $TE_{1,1}$, etc., waves, as shown in Figs. 122 and 123. In the case of rectangular guides these subscripts denote the number of half-period variations of the fields along the x and y axes, respectively.

The properties of the three principal modes that can exist in rectangular guides are given in Table 6.[3]

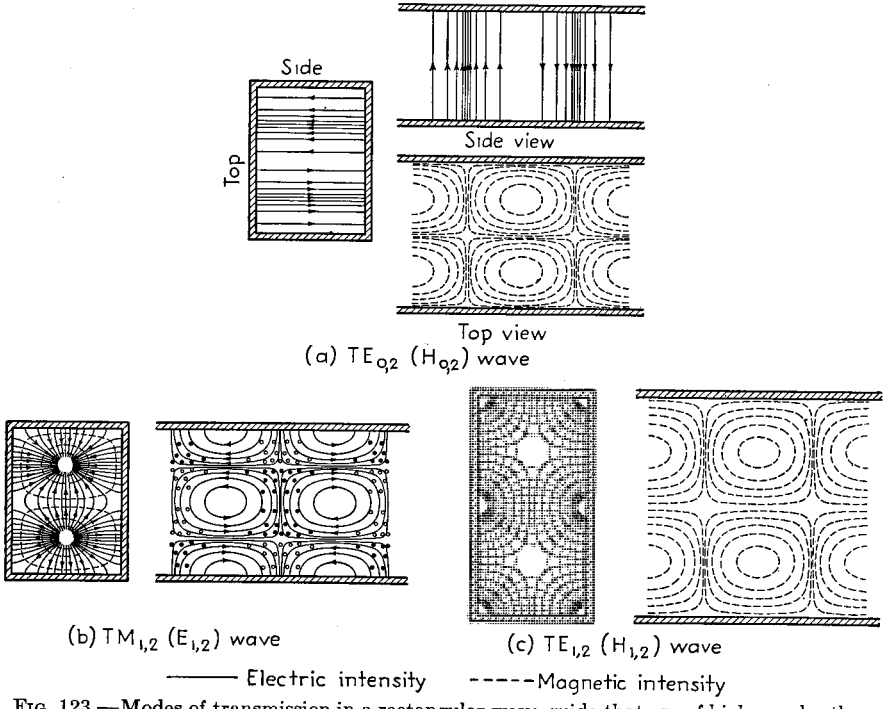


FIG. 123.—Modes of transmission in a rectangular wave guide that are of higher order than those shown in Fig. 122.

Cylindrical Wave Guides.[4]—The general relations for cylindrical guides are the same as those for the rectangular guides, though the specific form of the field equations differ. The commonest cylindrical modes are shown in Fig. 124. It will be noted that the waves of Figs. 124*a* and 124*b* are closely related to the waves of Figs. 122*b* and 124*d* of a rectangular guide.

A double-subscript notation is used to designate the various modes in a circular guide (see Fig. 124). As applied to circular guides, the first subscript denotes the number of full cycles in the radial field pattern that is traversed when the angle ϕ (see Fig. 125) passes through 360°, while the second subscript denotes the number of

[1] Information on other than rectangular and circular guides is given by Chu, *loc. cit.*, and Schelkunoff, *loc. cit.*

[2] L. J. Chu and W. L. Barrow, Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section, *Proc. I.R.E.*, Vol. 26, p. 1520, December, 1938.

[3] From Chu and Barrow, *loc. cit.*

[4] For further information see Southworth, Carson, *et al.*, and Chu, *loc. cit.*

TABLE 6.—PROPERTIES OF RECTANGULAR WAVE GUIDES

Notation

a = transverse dimension of guide, x dimension (see Fig. 122).
b = transverse dimension of guide, y dimension (see Fig. 122).
f = frequency, cycles per sec.
$\omega = 2\pi f$, angular velocity, radians per sec.
β = phase constant, radians per cm, based on phase velocity.
λ = wave length, cm (in free space).
c = velocity of light, 3×10^{10} cm per sec.

Operating Characteristics Common to All Rectangular Modes

Wave length in guide: $\lambda_g = 2\pi/\beta$ Wave length in free space: $\lambda = c/f$

Phase velocity: $v_p = \omega/\beta$ Group velocity: $v_g = d\omega/d\beta = c^2/v_p$

Mode	Cutoff frequency, f_0	Cutoff wave length, λ_0	Phase constant, β
$TE_{0,1}(H_{0,1})$*	$c/2b$	$2b$	$\frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\lambda}{2b}\right)^2}$
$TE_{1,1}(H_{1,1})$†	$\frac{c}{2}\sqrt{\left(\frac{1}{a}\right)^2+\left(\frac{1}{b}\right)^2}$	$\frac{2}{\sqrt{\left(\frac{1}{a}\right)^2+\left(\frac{1}{b}\right)^2}}$	$\frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\lambda}{2a}\right)^2-\left(\frac{\lambda}{2b}\right)^2}$
$TM_{1,1}(E_{1,1})$‡	$\frac{c}{2}\sqrt{\left(\frac{1}{a}\right)^2+\left(\frac{1}{b}\right)^2}$	$\frac{2}{\sqrt{\left(\frac{1}{a}\right)^2+\left(\frac{1}{b}\right)^2}}$	$\frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\lambda}{2a}\right)^2-\left(\frac{\lambda}{2b}\right)^2}$

* See Fig. 122*b*.
† See Fig. 122*c*.
‡ See Fig. 122*d*.

TABLE 7.—PROPERTIES OF CYLINDRICAL GUIDE MODES

Notation

a = guide radius, cm.
f = frequency, cycles per sec.
$\omega = 2\pi f$, angular velocity, radians per sec.
β = phase constant, radians per cm, based on phase velocity.
λ = wave length, cm (in free space).
c = velocity of light, 3×10^{10} cm per sec.

Operating Characteristics Common to All Cylindrical Modes

Wave length in guide: $\lambda_g = 2\pi/\beta$ Wave length in free space: $\lambda = c/f$

Phase velocity: $v_p = \omega/\beta$ Group velocity: $v_g = d\omega/d\beta = c^2/v_p$

Mode	Cutoff frequency, f_0	Cutoff wave length, λ_0	Phase constant, β
$TM_{0,1}(E_0)$*	$0.383\,\frac{c}{a}$	$2.61a$	$\sqrt{\left(\frac{2\pi}{\lambda}\right)^2-\left(\frac{2.405}{a}\right)^2}$
$TE_{0,1}(H_0)$†	$0.610\,\frac{c}{a}$	$1.64a$	$\sqrt{\left(\frac{2\pi}{\lambda}\right)^2-\left(\frac{3.83}{a}\right)^2}$
$TE_{1,1}(H_1)$‡	$0.289\,\frac{c}{a}$	$3.46a$	$\sqrt{\left(\frac{2\pi}{\lambda}\right)^2-\left(\frac{1.84}{a}\right)^2}$

* See Fig. 124*b*.
† See Fig. 124*c*.
‡ See Fig. 124*a*.

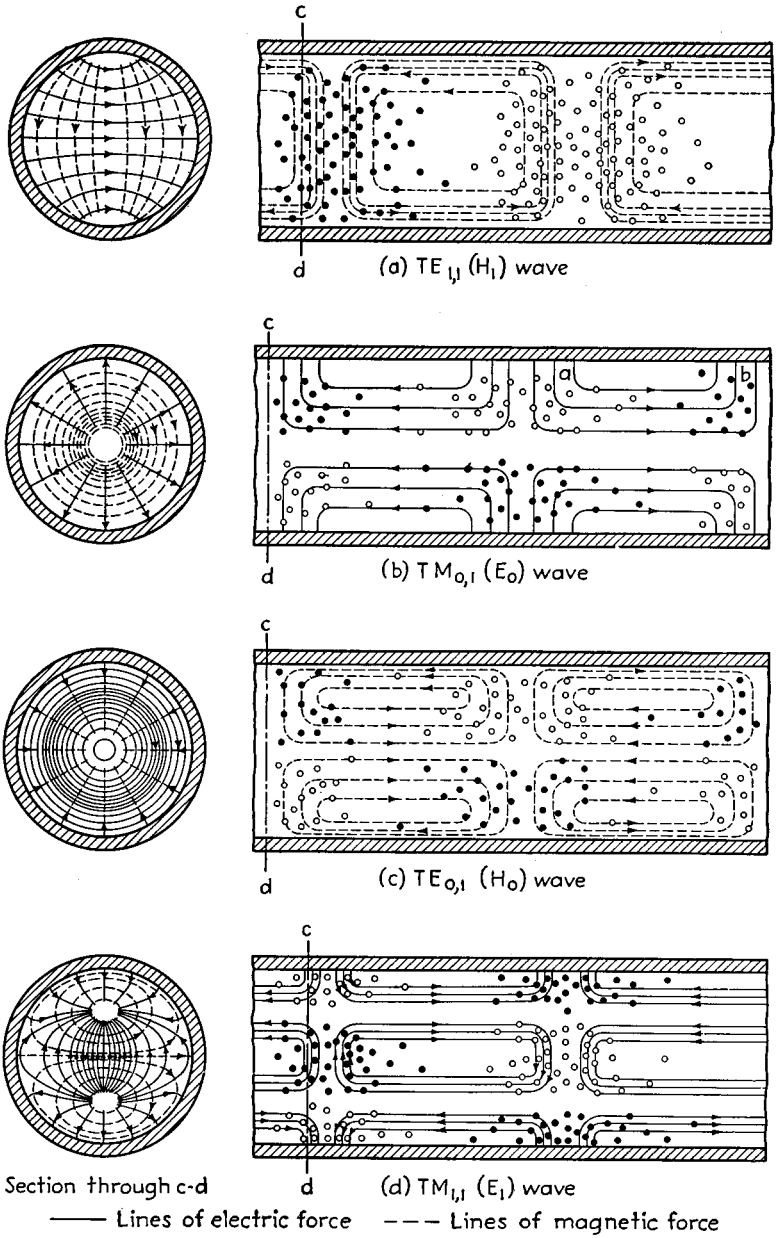


FIG. 124.—Field distribution corresponding to the fundamental modes of transmission possible in a circular wave guide.

half cycles of field variation along a radial component between the center and the walls. Waves in circular guides are also sometimes denoted by single subscripts, as E_0, H, etc., types, as shown in Fig. 124.

The properties of the first three modes of Fig. 124 are given in Table 7.[1]

Cutoff Frequency and Properties Related Thereto.—With any mode of wave-guide propagation, as the frequency is decreased, the longitudinal loops whether magnetic or electric become longer until at the cutoff frequency they become infinite in extent, and the phase velocity is infinite. As the frequency is increased the loops become smaller and their axial length approaches a half wave length as measured in free space. The distance between corresponding points on successive waves inside the guide is always greater than the distance between corresponding points of a plane wave in free space because the phase velocity is always greater than the velocity of light.

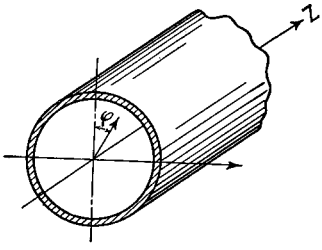


FIG. 125.—Diagram illustrating notation in circular wave guides.

The relation between the various factors, such as frequency, wave length, and velocity, is shown in Fig. 126. From this it is seen that as the frequency decreases, *i.e.*, as the wave length increases, both the phase velocity and the wave length in the guide increase to an infinite value at cutoff. This is a universal relation that holds for all modes and all shapes of guides.

All wave guides act like high-pass filters, with the cutoff frequency determined by the shape and size of the guide and by the mode of transmission. In Fig. 127 are shown the longest wave lengths passed by the simplest modes of a square and a circular wave guide. In general, the critical wave length is of the order of the cross-sectional dimension of the guide.

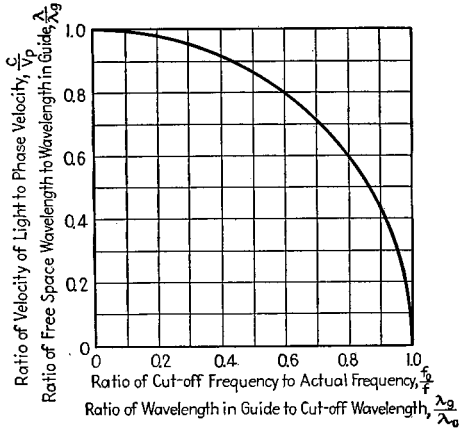


FIG. 126.—Variation of phase velocity and wave length in guide as the ratio of cutoff frequency to actual frequency is varied.

For the square guide, the lowest frequency mode has a critical wave length that is twice the width of the guide. This particular mode is unique in that its critical wave length depends only upon the width of the guide but not upon the guide height, so that a rectangular guide of a given width has the same critical wave length for any height. For a square guide, the $TE_{1,1}$ and $TM_{1,1}$ modes each have a critical wave length equal to the diagonal of the square. The critical wave length for these modes depends upon both of the dimensions of the guide if the guide shape is rectangular. It will be noted that if the dimensions of a rectangular guide are suitably chosen in relation to wave length, only the $TE_{0,1}$ mode of propagation is possible, and one is assured that the fields in the guide will correspond to those of a single pure mode.

[1] From Chu, *loc. cit.*

Similar relations hold for the circular guide. Here the critical wave length depends upon the diameter and the mode of transmission. As frequency is increased from below the lowest cutoff frequency for a given circular guide, the guide will first pass only the cylindrical $TE_{1,1}$ mode, then also the circular magnetic mode $TM_{0,1}$, and finally the circular electric mode $TE_{0,1}$. After that, higher order modes will be passed.

Excitation of Wave-guide Modes.—Each wave-guide mode must be excited in a different way. The guide can be thought of as a region of space into which energy is radiated from the coupling or exciting circuit. Just as waves can be set up in space by straight wires and loops, so can the wave-guide modes be excited by electric probes and loops.

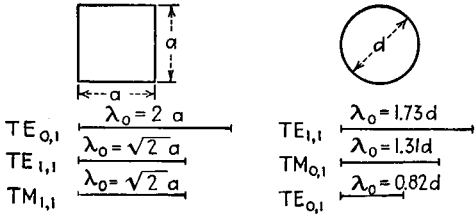


FIG. 127.—Diagram illustrating relative length of waves corresponding to cutoff frequency for the three simplest modes of propagation in square and circular wave guides.

The location of an exciting probe or loop must bear a definite relation to the fields in the guide corresponding to the desired transmission mode. For instance, if the transmission mode has an axial component of electric field and if its magnetic flux lines are transverse closed loops, then it can be excited either by an antenna probe that produces an axial component electrical field or by a loop placed so that it produces a magnetic field in the proper direction. In general, any mode can by excited either by a suitable antenna probe (or dipole) or by a suitable current loop.

Excitation of various transmission modes in a rectangular wave guide by means of antenna probes is shown in Fig. 128.[1] Corresponding arrangements for circular guides are shown in Fig. 129.[2] In all cases the probe creates an electric or magnetic field of the proper orientation.

Unless the dimensions of the guide are such that only one transmission mode can exist, any excitation device will set up several modes simultaneously, though the desired mode is usually strongest. When this is the case the wave may be purified by inserting grids or wires that will short out the electric field of the undesired mode.

The pickup device at the end of a wave guide is of the same form as the exciting device. It is rather difficult to terminate a wave guide so that there will be no reflections from the load end, unless the termination has both a reactive and a resistive adjustment.

Attenuation in Wave Guides.—Hollow-pipe guides have attenuation characteristics that are similar to those of conventional high-pass filters in which the electrical elements are lumped rather than distributed. The chief features of the attenuation characteristics of high-pass filters are that there is a high attenuation in the attenuation band (*i.e.*, for all frequencies below cutoff) and a low but finite attenuation in the pass band (*i.e.*, for frequencies above cutoff). The nature of the attenuation in the attenuation and pass bands is different. At frequencies below the cutoff frequency, power is rejected at the source because the guide dimensions and frequency do not permit the existence of the type of waves that transmit power. In this region the attenuation is so high that no effective transmission of power is possible over any appreciable distances.

[1] From Chu and Barrow, *loc. cit.*
[2] From Southworth, *loc. cit.*, and Chu, *loc. cit.*

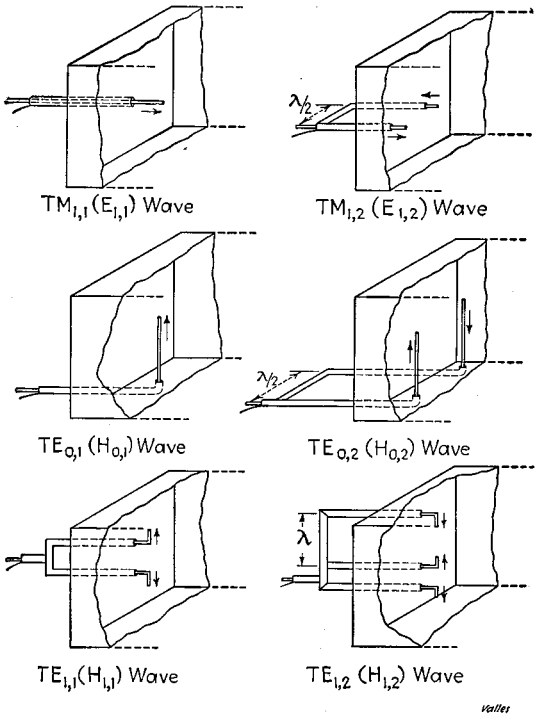


FIG. 128.—Typical probe arrangements for exciting different types of waves in rectangular guides.

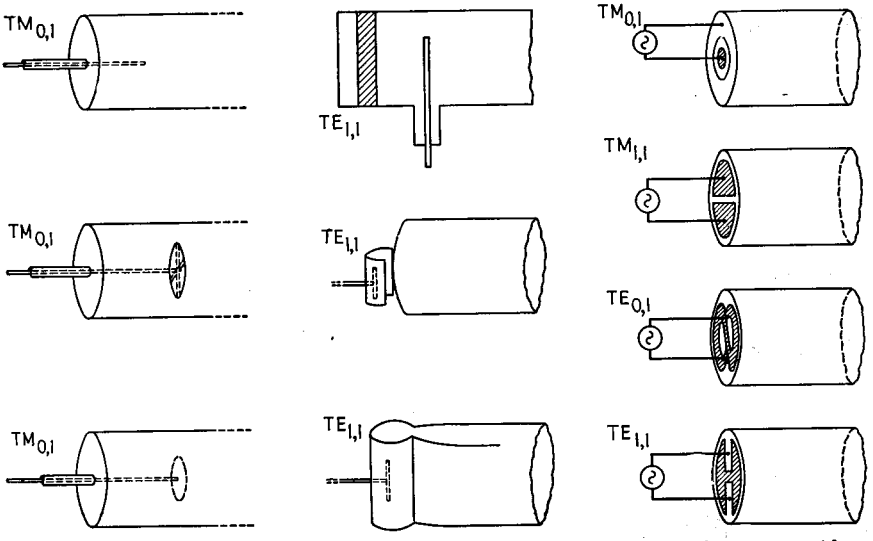


FIG. 129.—Typical methods of exciting different types of waves in circular wave guides.

In the pass band, as frequency increases, the attenuation of a wave guide first decreases from a very large value (theoretically infinite at cutoff), then levels off at some low but finite value, and then increases as shown in Fig. 130.[1] This is the result of two opposing factors that take place as the frequency increases. First, the field configurations become more favorable for keeping the losses low, and, second, the skin effect becomes greater as the frequency is raised. This variation of attenuation with frequency is found in all shapes of guides and all modes with the single exception of the cylindrical $TE_{0,1}$ mode, for which the attenuation decreases without limit as the frequency increases.[2]

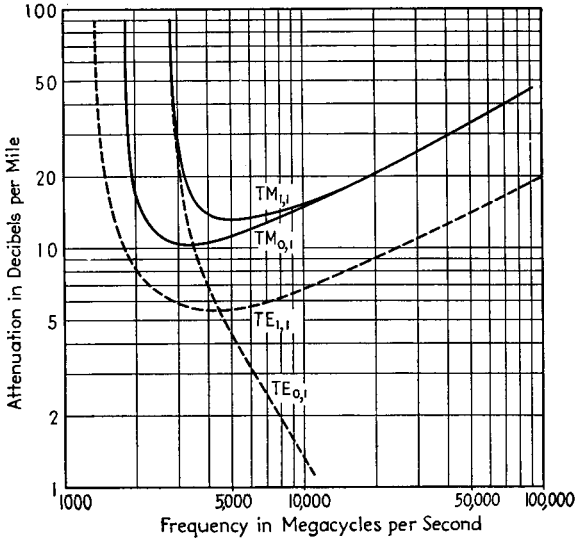


FIG. 130.—Attenuation of different types of waves in a 5-inch copper cylindrical wave guide for various frequencies.

Formulas for the attenuation of the various basic modes in cylindrical and square guides are given in Table 8.[3] Results calculated for a circular copper guide 5 inches in diameter are given in Fig. 130. Examination of Tables 8 and 9 and Fig. 130 shows that in all cases except that of the $TE_{0,1}$ mode, the attenuation decreases from a very high value at cutoff, reaches a minimum at from about 2 to 3 times the cutoff frequency, and then increases with frequency, finally increasing with the square root of frequency at very high frequencies. The minimum is quite broad, and the minimum attenuation low. Thus in Fig. 130 there is a band of frequencies several thousand megacycles wide in which the attenuation varies only 1 db, though it takes a very high frequency or a very large guide to realize this property.

[1] From Southworth, *loc. cit.*

[2] No experimental confirmation of this property, or even of the existence of the $TE_{0,2}$ mode, has been published. In fact, Chu, *loc. cit.*, and Chu and Barrow, *loc. cit.*, question the stability of this mode.

[3] These results are from Chu and Chu and Barrow, *loc. cit.*, and Carson, Mead, and Schelkunoff, *loc. cit.*

An alternative method of calculating the attenuation in a rectangular guide, based upon waves being reflected back and forth between the walls of the guide, is given by John Kemp, Electromagnetic Waves in Metal Tubes of Rectangular Cross Section, *Jour. I.E.E.*, Vol. 88, Pt. III, p. 213, September, 1941. See also a discussion of this by G. W. O. Howe, *Wireless Eng.*, Vol. 19, p. 93, March, 1942.

TABLE 8.—WAVE-GUIDE ATTENUATION FORMULAS FOR AIR DIELECTRIC INSIDE GUIDE

Notation

r = guide radius (of circular guide), cm.
a = length of side of square guide, cm.
f = actual frequency, cycles.
f_0 = cutoff frequency, cycles.
K_1 = a constant depending on guide walls, and given at end of table.

Circular Guide

$TM_{0,1}$ (E_0) mode:

$$\alpha = \frac{0.4378K_1}{r^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

$TM_{1,1}$ (E_1) mode:

$$\alpha = \frac{0.5520K_1}{r^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

$TE_{0,2}$ (H_0) mode:

$$\alpha = \frac{0.5520K_1}{r^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{1/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

$TE_{1,1}$ (H_1) mode:

$$\alpha = \frac{0.3824K_1}{r^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{-1/2} + \frac{1}{2.38}\left(\frac{f}{f_0}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

Square Guide

$TE_{0,1}$ mode:

$$\alpha = \frac{K_1}{a^{3/2}} \frac{\frac{1}{2}\left(\frac{f}{f_0}\right)^{3/2} + \left(\frac{f}{f_0}\right)^{-1/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

$TE_{1,1}$

$$\alpha = \frac{1.189K_1}{a^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{3/2} + \left(\frac{f}{f_0}\right)^{-1/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

$TM_{1,1}$

$$\alpha = \frac{1.189K_1}{a^{3/2}} \frac{\left(\frac{f}{f_0}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_0}\right)^2 - 1}} \text{ db per mile}$$

Value of K_1

Copper	236.5
Aluminum	313.6
Lead	821.3
Iron (μ = 100)	5,696

The minimum attenuation associated with a particular mode is given by the relation

For circular guide:

$$\text{Minimum attenuation} = \frac{K_2}{r^{3/2}} \text{ db per mile} \qquad (149a)$$

For square guide:

$$\text{Minimum attenuation} = \frac{K_2}{a^{3/2}} \text{ db per mile} \qquad (149b)$$

where K_2 is a constant depending on the mode of propagation as given in Table 9 and r and a have the same meanings as in Table 8. The ratio of the frequency corresponding to minimum attenuation to the cutoff frequency is given in Table 9.

TABLE 9.—ATTENUATION PROPERTIES OF SQUARE AND CIRCULAR GUIDES

Mode	$\left(\frac{\text{Frequency of minimum attenuation}}{\text{Cutoff frequency}}\right)$	Minimum attenuation, relative value	K_2 (copper)
Circular Guide			
$TE_{1,1}$	3.150	1.00	170
$TM_{0,1}$	1.732	2.00	340
$TM_{1,1}$	1.732	2.53	430
Square Guide			
$TE_{0,1}$	2.960	1.867	265.5
$TE_{1,1}$	2.438	3.967	562.6
$TM_{1,1}$	1.732	3.20	453.3

Examination of the foregoing shows that *for a given periphery*, the $TE_{1,1}$ mode in the circular guide has the lowest possible attenuation. In all cases the attenuation for a cylindrical guide is less than that for the corresponding mode in a square guide of the same guide periphery. The lowest attenuation that can be obtained from a rectangular mode for a given guide periphery occurs for the $TE_{0,1}$ mode of Fig. 122 when the ratio of width to height is 1.18, though for this condition the attenuation is only a few per cent lower than for a guide of square cross section of the same periphery excited in the same mode.

34. Cavity Resonators.[1]—Any closed surface having conducting walls can support oscillating electromagnetic fields within it, and possesses certain resonant frequencies when excited by electrical oscillations. Resonators of this type are commonly termed *cavity resonators* or *rhumbatrons*, and find extensive use as resonant circuits at extremely high frequencies. For such use cavity resonators have the advantages of simplicity, relatively large physical size, high Q, and very high shunt impedances. At wave lengths well below one meter cavity resonators become vastly superior to the corresponding resonators with lumped circuit constants, and are fully as desirable—if not more so—as resonant concentric lines for such frequencies.

Types of Cavity Resonators and Modes of Oscillation.—Practical cavity resonators can be conveniently divided into two types. The first of these is the nonreentrant

[1] The basic paper on cavity resonators is that of W. W. Hansen, A Type of Electrical Resonator, *Jour. Applied Phys.*, Vol. 9, p. 654, October, 1938. See also E. U. Condon, Forced Oscillations in Cavity Resonators, *Jour. Applied Phys.*, Vol. 12, p. 129, February, 1940.

type, which in idealized form is typified by the cylinder, prism, sphere, ellipsoid, etc., as shown in Figs. 131*a* to 131*d*. The second type is illustrated in Figs. 131*e* to 131*j* in various forms, some idealized and some practical, and is characterized by having a reentrant portion or portions.[1]

The fields corresponding to the normal or fundamental mode of operation of typical cavity resonators are shown in Fig. 131. In the case of a cylindrical resonator, the oscillations can be described crudely as involving electric charges or currents that flow up and down the sides of the cylinder and charge the electrostatic capacity

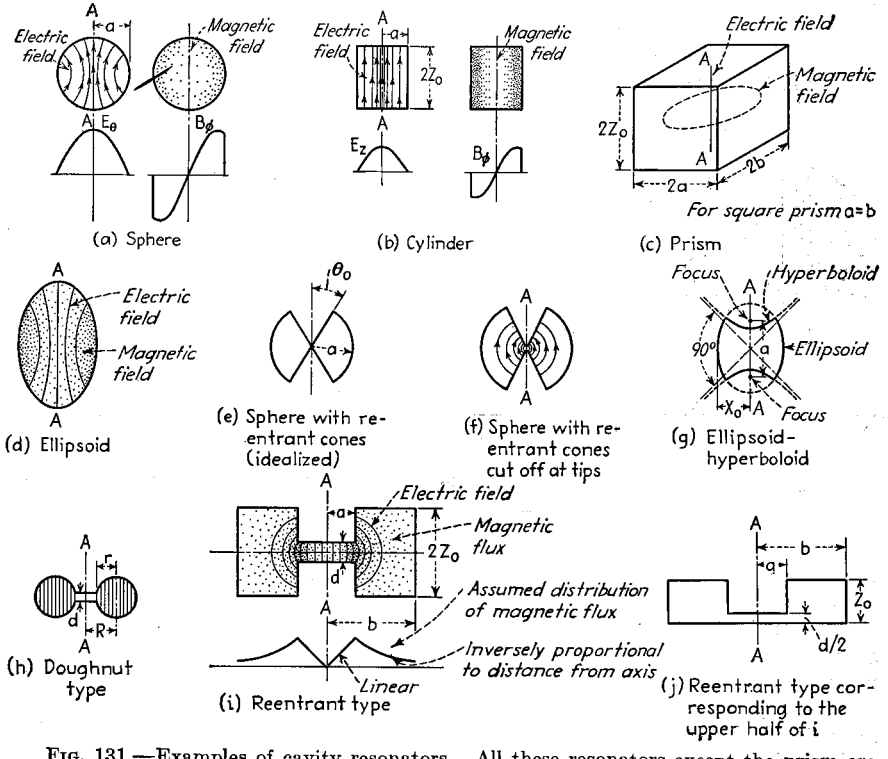


FIG. 131.—Examples of cavity resonators. All these resonators except the prism are shown as cross sections of a figure of revolution. The flux distributions shown for certain of the resonators correspond to the distributions with the normal or fundamental mode of operation.

formed by the top and bottom ends of the cylinder. The sphere is analogous. With resonators of the reentrant type operating as in Figs. 131*e* to 131*i*, the electric field is concentrated largely in the region of the reentrant portions, with most of the magnetic flux existing in the remaining regions and encircling the reentrant portions as shown. Because the electric and magnetic fields tend to concentrate in different parts of the reentrant-type cavity resonator, such a resonator is much more closely related to a circuit with lumped constants than the other forms of cavity resonators.

Higher order modes of many varieties can exist in cavity resonators. These correspond in the case of cylindrical and prism types of cavity resonators to the

[1] Characteristics of resonators of the reentrant type are discussed by W. W. Hansen and R. D. Richtmyer, On Resonators Suitable for Klystron Oscillators, *Jour. Applied Phys.*, Vol. 10, p. 189, March, 1939.

higher order modes that can exist in circular and rectangular wave guides, while in the case of reentrant resonators they correspond to the higher order modes of propagation that can exist in concentric transmission lines when the conductor spacing is of the order of a half wave length or more.[1] Typical examples are shown in Fig. 132. Higher order modes correspond to a higher resonant frequency than the fundamental mode. Other properties of the resonator, such as the Q and shunt impedance, also depend upon the mode, and may be either greater or less than for the fundamental mode, according to circumstances. In particular, certain types of higher order modes have very low losses and give unusually high Q and shunt impedance. The examples of Fig. 132 are of this type. Here the fact that the magnetic flux density adjacent to the walls of the cylinder is less than for fundamental operation causes the current flowing in the walls to be small, with correspondingly lower losses and higher Q. In spite of such possible advantages obtained from higher mode operation, cavity

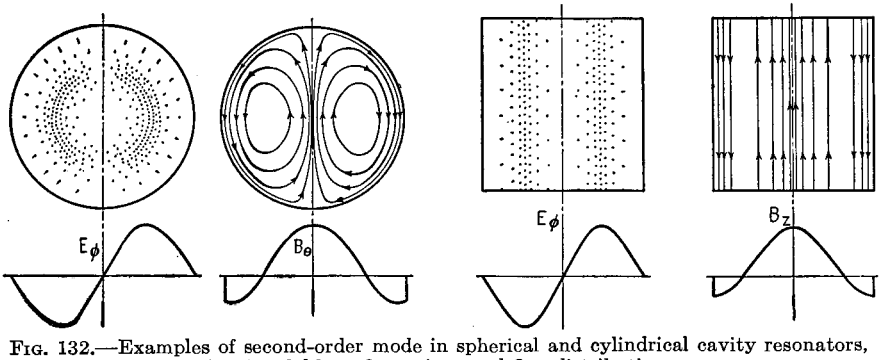


FIG. 132.—Examples of second-order mode in spherical and cylindrical cavity resonators, showing field configurations and flux distributions.

resonators are ordinarily operated in the normal or fundamental mode. This is because in the region where higher modes are possible, the frequency separation between different modes is relatively small, and there are even certain conditions under which the cavity resonator can be simultaneously resonant to more than one mode. Under these circumstances there is always the possibility that the resonator will be excited in undesired ways.

Resonant Frequency of Cavity Resonators.—The resonant frequency of a cavity resonator corresponds to a possible solution of Maxwell's equations for the electric and magnetic fields within the resonator, which will also satisfy the boundary conditions. The calculation of resonant frequency (or wave length) can be carried out exactly for geometrical shapes such as spheres, cylinders, prisms, and for certain idealized cases involving reentrant sections. Approximate solutions have been derived for certain other cases.

Formulas giving the wave length at resonance for the fundamental mode, for various simple geometrical shapes, are given in Table 10, Fig. 133, and Eqs. (150) and (151). This resonant wave length is proportional to the size of the resonator; *i.e.*, if all dimensions are doubled, the wave length corresponding to resonance will likewise be doubled. This fact simplifies the construction of resonators of shapes that cannot be calculated. To obtain a resonator operating exactly at a desired frequency, one first constructs a resonator of convenient size and of the desired

[1] See Hansen, *loc. cit.* An extensive discussion of certain types of higher modes is given by W. L. Barrow and W. W. Mieher, Natural Oscillations of Electrical Cavity Resonators, *Proc. I.R.E.*, Vol. 28, p. 184, April, 1940.

TABLE 10.—PROPERTIES OF CAVITY RESONATORS

	Figure showing notation	Wave length, cm	Q	Shunt resistance, ohms	Impedance coupled into loop of area Σ sq cm, ohms
Sphere	131*a*	$2.28a$	$1.024\frac{a}{\delta}$	$81.6\frac{a}{\delta}$	$354\frac{\Sigma^2}{a^3\delta}$
Cylinder	131*b*	$2.61a$	$1.414\frac{a}{\delta}\frac{1}{1+\frac{a}{2z_0}}$	$204\frac{z_0}{\delta}\frac{1}{1+\frac{a}{2z_0}}$	$204\left(\frac{\Sigma}{az_0}\right)^2\frac{z_0}{\delta}\frac{1}{1+\frac{a}{2z_0}}$
Prism (square)	131*c*	$2.83a$	$1.41\frac{a}{\delta}\frac{1}{1+\frac{a}{2z_0}}$	$182.1\frac{z_0}{\delta}\frac{1}{1+\frac{a}{2z_0}}$	$296\left(\frac{\Sigma}{az_0}\right)^2\frac{z_0}{\delta}\frac{1}{1+\frac{a}{2z_0}}$
Spheres with reentrant cones	131*e*	$4a$	Fig. 134	Fig. 136	
Ellipsoid-hyperboloid	131*g*	See Fig. 133	Fig. 135	Fig. 137	
Toroidal type	131*h*	See Eq. (151)			
Toroidal type	131*i*	See Eq. (150)	$\frac{2.83\frac{z_0}{\lambda}\log_e\frac{b}{a}}{\frac{z_0}{a}\left(1+\frac{a}{b}\right)+\log_e\frac{b}{a}}\frac{\lambda}{\delta}$	$2{,}130\left(\frac{z_0}{\lambda}\right)^2\frac{\log_{\epsilon^2}\frac{b}{a}}{\frac{z_0}{a}\left(1+\frac{a}{b}\right)+\log_e\frac{b}{a}}\frac{\lambda}{\delta}$	
Toroidal type	131*j*	See Eq. (150)	Same as Fig. 131*i*	$1{,}065\left(\frac{z_0}{\lambda}\right)^2\frac{\log_{\epsilon^2}\frac{b}{a}}{\frac{z_0}{a}\left(1+\frac{a}{b}\right)+\log_e\frac{b}{a}}\frac{\lambda}{\alpha}$	

All dimensions in centimeters.
δ defined as in Eq. (153); for copper $\delta = 54 \times 10^{-6}\sqrt{\lambda_{cm}}$.
Formulas for Q and shunt resistance for type of Figs. 131*i* and 131*j* assume flux distribution of Fig. 131*i*.
Notation as in Fig. 131.

proportions. The resulting resonant frequency is then measured. The ratio of this wave length to the desired resonant wave length gives a scale factor that is applied to every dimension of the final resonator to obtain the dimensions of the resonator giving the desired wave length.

With cylinders the resonant wave length is determined only by the diameter, and is not affected by the height. Similarly, the height has no effect in the case of the prism. A sphere with reentrant cones (Fig. 131*e*) has a wave length independent of the cone semiangle θ, and only moderately greater than for the same sphere without reentrant cones. These relations apply to fundamental modes.

The resonant frequency of resonators such as illustrated in Figs. 131*h* to 131*j*, which have a large lumped capacity, can be calculated crudely on the assumption of lumped constants; *i.e.*, one determines the capacity between the flat surfaces AA,

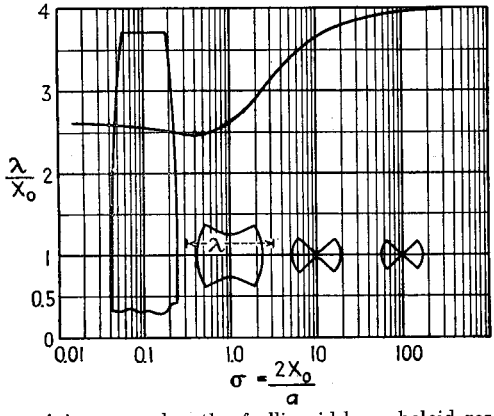


FIG. 133.—Curve giving wave length of ellipsoid-hyperboloid resonators of the type shown in Fig. 131*g* for different values of the shape factor σ_0. The little drawings show resonator cross sections corresponding to different values of σ_0, and are so scaled that all give the same wave length. Only the upper half of the drawing corresponding to $\sigma_0 = 0.1$ is shown.

and assumes that this tunes a lumped inductance corresponding to that of a one-turn toroidal coil, having a cross section such as indicated by the shading in Fig. 131*h*. For the reentrant resonator of Figs. 131*i* and 131*j* this leads to the relation[1]

$$\left.\begin{matrix}\text{Wave length in cm}\\ \text{for resonance}\end{matrix}\right\} = 2\pi\left(\frac{Z_0 a^2}{d}\log_\epsilon\frac{b}{a}\right)^{1/2} \tag{150}$$

Similarly, for Fig. 131*h*

$$\left.\begin{matrix}\text{Wave length in cm}\\ \text{for resonance}\end{matrix}\right\} = 2\pi\left\{\frac{\pi r^2 R}{d}\left(1-\sqrt{1-\left(\frac{r}{R}\right)^2}\right)\right\}^{1/2} \tag{151}$$

The notation is as shown in Fig. 131 (dimensions in cm). Results calculated by Eqs. (150) and (151) always give a wave length less than the correct value. The error depends upon the extent to which the electric field is concentrated in the narrow gap, *i.e.*, upon the smallness of the dimension d. The factor by which Eq. (150) must be multiplied to obtain the correct wave length is commonly of the order of 1.25 to 1.75.

[1] A more accurate determination can be made for this case by methods described by W. W. Hansen, On the Resonant Frequency of Closed Concentric Lines, *Jour. Applied Phys.*, Vol. 10, No. 1, January, 1939.

The resonant frequency of a cavity resonator can be changed by altering the mechanical dimensions, by coupling reactance into the resonator, or by means of a copper paddle. Small changes in mechanical dimensions can be achieved by flexing walls, while large changes require some type of sliding member. The resonant frequency of the reentrant type of resonator is particularly sensitive to the gap between the reentrant portions, especially if this gap is small. Reactance can be coupled into the resonator by coupling loops or probes in the manner discussed below, and affects the resonant frequency. A copper paddle placed inside the resonator in such a manner as to intercept some of the flux lines corresponding to the normal distribution will raise the resonant frequency. Rotation of the paddle will vary this effect and so can be used to adjust the resonant frequency, exactly as a copper paddle can be used to produce small variations in the inductance of a coil.

Q of Cavity Resonators.—The Q of a cavity resonator has the same significance as for an ordinary resonant circuit, and can be defined on the basis that Q is a quantity such that when the response has dropped to 70.7 per cent of the response at resonance, the cycles off resonance is the resonant frequency divided by $2Q$. In the case of cavity resonators, it is also convenient to think of Q as expressing the ratio of the energy stored in the fields of the resonator to the energy lost per cycle, according to the relation

$$Q = 2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}} \tag{152}$$

In terms of the magnetic field H within the resonator, this is

$$Q = \frac{2\sqrt{2}}{\delta} \frac{\int H^2\, dv}{\int H^2 |da|} \tag{153}$$

where H = magnetic intensity.

$\delta = \sqrt{\rho/\pi\omega}$.

ρ = resistivity in abohms (emu) per cm cube (for copper $\rho = 1{,}724$ abohms per cm cube).

$\omega = 2\pi$ times frequency.

The integral of the numerator is evaluated throughout the volume v of the resonator, and the integral of the denominator is taken over the area a of the inner surface of the resonator.

The Q of a resonator of any shape can be evaluated with the aid of Eq. (153) for any mode of oscillation provided that the distribution of magnetic flux is known. In simple geometrical shapes, such as spheres and cylinders, the flux distribution may be expressed analytically. The indicated integrations can then be carried out without particular difficulty, and exact formulas for Q are obtained. In other cases it is necessary to assume an approximate distribution of flux, and then to carry out the integrations either analytically or graphically as the occasion permits. The value of Q obtained in this way will not be exactly correct, but the error is small even if the assumed flux distribution differs appreciably from that actually present. This is because the flux appears both in the numerator and denominator of Eq. (153).

An important implication of Eq. (152) is that to obtain high Q, the resonator should have a large ratio of volume to surface area. The physical reason for this is apparent when it is realized that energy is stored in the volume of the resonator, whereas all the energy lost is lost at the walls. As a consequence, resonators such as spheres, cylinders, and prisms can in general be expected to have higher Q than the corresponding resonators with reentrant sections. The loss in Q resulting from reentrant sections is not excessive, however, unless the reentrant portions become very sharp spines. With resonators of the same proportions but different size, Eq. (152) shows that the Q's will be proportional to the square root of the wave length.

TABLE 11.—TYPICAL CAVITY RESONATOR PROPERTIES
Copper walls — $\lambda = 40$ cm*

Type of resonator	Physical size	Q	Shunt resistance, megohms
Sphere	$a = 17.5$ cm	53,000	4.2
Cylinder (diameter = height; $z = a$)	$a = 15.3$ cm	48,800	6.1
Sphere with reentrant cones ($\theta_0 = 34°$)	$a = 10$ cm	18,200	3.5
Reentrant type (Fig. 131*i*): $\frac{b}{a} = 3.33$, $\frac{Z_0}{a} = 1.20$	$Z_0 = 2.8$ cm, $a = 2.8$ cm, $b = 3.9$ cm } approx.	10,200	0.66

* For other values of λ, the dimensions are proportional to λ, while Q and shunt resistance vary as $\lambda^{1/2}$

Values of Q for a number of types of resonators can be obtained from Table 10, Figs. 134 and 135. In the case of a sphere with reentrant cones (Fig. 131*c*) the Q passes through a broad maximum when the half angle of the cone is 34°. The Q of the ellipsoid-hyperboloid cavity resonator of Fig. 131*f* is substantially the same as that of a sphere having the diameter $2x_0$, provided that the distance a between the focuses of the ellipse is somewhat greater than the diameter. When the reverse is the case, the Q drops to about one-third that of the corresponding sphere.

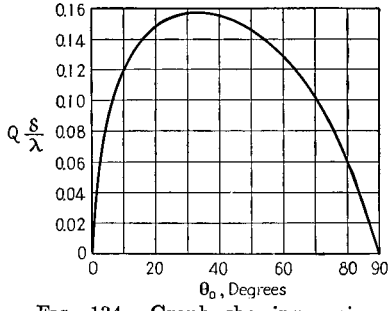


FIG. 134.—Graph showing variation in Q for the resonator of Fig. 131*e* with the angle θ_0. The notation is the same as used in Table 10 and Fig. 131*e*.

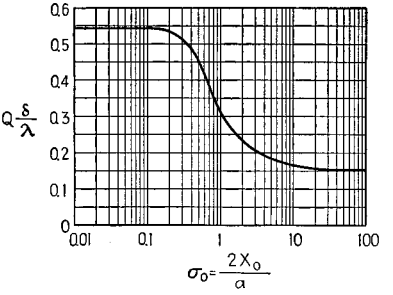


FIG. 135.—Curve giving Q of ellipsoid-hyperboloid type of cavity resonator shown in Figs. 131*g* and 133. The notation is as in Fig. 133 and Table 10.

Typical values of Q obtainable in practical cavity resonators are given in Table 11, and are extremely high compared with those encountered with ordinary resonant circuits.

Shunt Impedance of Cavity Resonators.—The shunt impedance of a cavity resonator such as those of Fig. 131 can be defined as the square of the line integral of voltage along a path such as AA, divided by the power loss in the resonator. This impedance corresponds to the parallel resonant impedance of a tuned circuit, and at resonance becomes a resistance termed the shunt resistance of the resonator.

The shunt impedance of a number of types of cavity resonators can be calculated with the aid of Table 10 or Figs. 136 and 137. The values obtainable are very large, as seen from typical results given in Table 11.

The shunt resistance obtainable with cavity resonators of the reentrant type is less than without reentrants. However, this impedance is developed across a much shorter distance in the reentrant case; so the reentrant type of cavity resonator gives a much greater shunt impedance *in proportion to distance* across which the impedance

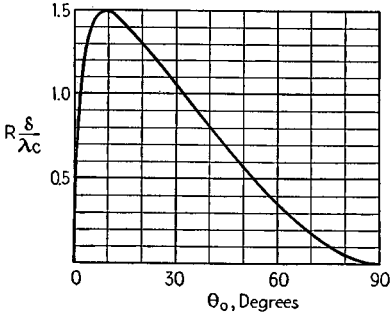


FIG. 136.—Variation of shunt impedance R of spherical resonator with reentrant cone (Fig. 131*e*) as a function of the semiangle θ of the cones (for notation see Fig. 131*e* and Table 10).

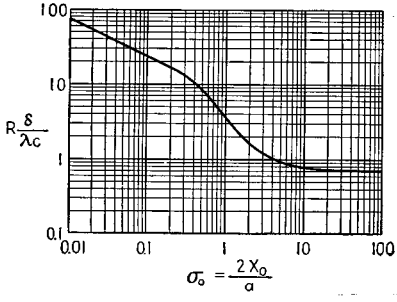


FIG. 137.—Curve giving shunt resistance for ellipsoid-hyperboloid resonators of the type shown in Fig. 131*g* for the various shapes shown in Fig. 133 (for notation see Fig. 131*g* and Table 10).

is developed. This is of importance when the resonator is associated with an electron beam, since it gives a high impedance across a distance small enough to correspond to a very short transit time for electrons.

Coupling to Cavity Resonators.—Coupling to a cavity resonator may be achieved by means of electron beams, by coupling loops, or coupling electrodes. Coupling by means of an electron beam can be accomplished by passing the beam through the resonator, as indicated in Fig. 138. In such an arrangement, it is necessary that the transit time required by the electrons to pass through the resonator be small compared with the time of a cycle. This in general requires the use of resonators of the reentrant type, in order to keep down the distance that the electrons must travel within the resonator. If the electrons in such a beam pass through the resonator in the form of bunches, one to each cycle, then the resonator will be excited by the beam, and if the resonator proportions are such that the resonant frequency approximates the frequency of the bunches, then oscillations of large amplitude will be built up in the cavity. On the other hand, if the resonator is excited by some external source of energy, and a continuous stream of electrons is passed through the resonator, then the emerging electrons will be alternately accelerated and slowed down by the oscillations within the resonator.

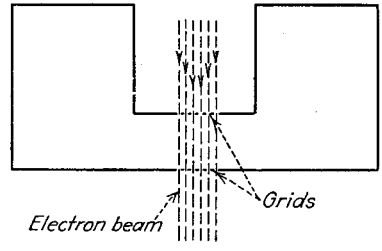
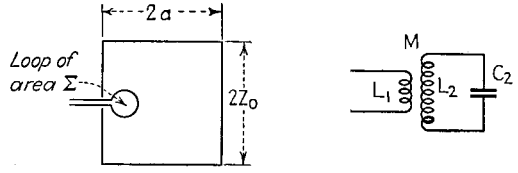


FIG. 138.—Cavity resonator of the type shown in Fig. 131*j* coupled to an electron beam that passes through the resonator.

Magnetic coupling to a cavity resonator can be obtained by means of a small coupling loop so oriented as to inclose magnetic flux lines corresponding to the desired mode of operation. A current passed through such a loop will then excite oscillations

of this mode, while conversely oscillation existing in the resonator will induce a voltage in such a coupling loop.

The combination of coupling loop and cavity resonator is equivalent to the ordinary coupled circuit of Fig. 139. The impedance that the cavity resonator couples into the loop (*i.e.*, the equivalent impedance that can be considered as in



(a) Cylindrical resonator with coupling loop (b) Equivalent coupled circuit

FIG. 139.—Cylindrical resonator with coupling loop, together with equivalent coupled circuit.

series with L_1) is given in Table 10 for the case of a cylinder, square prism, and sphere, on the assumption that the coupling loop is placed close to the outer edge where the magnetic flux density is maximum. The ratio between the impedance coupled into the loop and the shunt resistance of a cavity resonator is roughly equal to the square of the ratio (loop area)/(half the cross-sectional area of the resonator). This rule may be used to estimate the size of coupling loop that will be required in resonators of types other than those given in Table 10. These relations assume that the loop is oriented so that its area is in the most favorable position for inclosing magnetic flux. The magnitude of coupling can be readily changed by simply rotating the loop, the coupling reducing to zero when the plane of the loop is parallel to the magnetic flux.

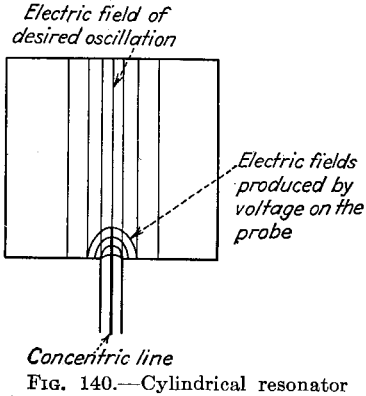


FIG. 140.—Cylindrical resonator with coupling probe, showing fields of desired oscillation and fields produced by the probe. It will be noted that the probe fields have a component in the direction of the desired field, and therefore are capable of exciting them.

Cavity resonator excitation by means of an electrode is illustrated in Fig. 140. Here there is a component of electric field produced by the presence of the electrode (or probe) that is in the same direction as the electric field of the desired oscillation in the resonator. Under these conditions a voltage applied to the electrode will excite oscillations in the resonator, and conversely oscillations in the resonator will deliver energy to the concentric line associated with the electrode.

It is also possible to couple cavity resonators to wave guides. This is accomplished by arranging so that the waves entering the cavity from the guide produce a field within the cavity that corresponds to the desired mode of cavity-resonator operation.

Miscellaneous Considerations.—A hole in a cavity resonator will cause energy to be radiated if oscillations exist inside the resonator, or, conversely, will cause oscillations to be produced within the resonator if oscillating fields external to the resonator are present. The loss of energy caused by radiation from holes is of particular importance under conditions where grids must be used in the sides of the resonator to allow for the entrance or exit of electron beams. The magnitude of energy loss through a

hole is proportional to the cube of the hole area (sixth power of the hole radius). In the case of a square prism one has[1]

$$Q' = 2\pi \frac{\text{energy stored inside resonator}}{\text{energy lost through hole per cycle}}$$
$$= 37 \left(\frac{a^2}{\Sigma}\right)^3 \frac{z_0}{a}$$

where Σ is the area of the hole and a and z_0 are the same as in Fig. 131c and Table 10. Note that Q' is the value of Q the resonator would have if there were no losses in the resonator other than the energy leaking out of the hole. It is easy to see that the losses from a hole will be quite negligible unless this hole is of relatively large size.

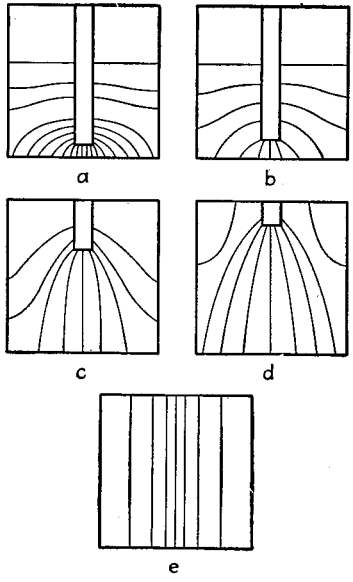


FIG. 141.—Transition of a reentrant cavity resonator of the concentric-line type to a cavity resonator that is a cylinder.

In designing a cavity resonator, it is desirable that all joints be parallel with the lines of current flow. Joints that must be traversed by currents can introduce large losses, because the currents that must be carried are extremely high and even a very low resistance in the joint will seriously reduce the Q and shunt resistance of the resonator.

The dividing line between resonant transmission lines and cavity resonators is not at all distinct. For example, one can consider a quarter-wave resonant concentric transmission line, as illustrated in Fig. 141,[2] as being a cavity resonator with a reentrant section. Furthermore, by progressively shortening this reentrant section as shown in the figure, one finally ends up with a cylindrical cavity resonator. In a like manner, it is possible to show that a cavity resonator may be also viewed as a section of wave guide in which standing waves exist.

[1] See Hansen, *loc. cit.*
[2] From Barrow and Mieher, *loc. cit.*

SECTION 4

VACUUM TUBES AND ELECTRONICS[1]

LAWS OF ELECTRON BEHAVIOR

1. Physical Properties of Electrons and Ions.—For purposes of radio engineering, the most fundamental particle is the electron. The electron is the smallest particle whose mass and charge have been determined. It is the particle that constitutes most of the current flow encountered in vacuum tubes. It is the negative planetary particle of the hydrogen atom.

Another fundamental particle is the proton. The proton is the positive nuclear particle of the hydrogen atom. It has the same magnitude of harge as the electron, but its charge is positive, whereas the charge of the electron is negative. The mass of the proton is roughly 1,800 times that of the electron.

Other charged particles called ions may exist. These are usually created from a normal atom or molecule by the addition or removal of one or more electrons or protons from its structure. The mass of such ions is always much greater than that of the electron, and their charge is an integral multiple of the charge of the electron and is not restricted to the negative sign.

The principal properties of electrons and the commonest ions are given in Table 1.[2]

TABLE 1.—PHYSICAL PROPERTIES OF ELECTRONS AND IONS

Charge	
Electronic charge (e)	4.767×10^{-10} statcoulomb
	1.590×10^{-20} abcoulomb
	1.590×10^{-19} coulomb
Mass	
Mass of an electron*	9.038×10^{-28} gram
Mass of a proton	1.6608×10^{-24} gram
Mass of a hydrogen atom	1.6617×10^{-24} gram
Mass of hypothetical atom of unit atomic weight (m_0)	1.649×10^{-24} gram
Ratios	
Charge to mass of an electron*	1.769×10^{7} abcoulombs per gram
	5.303×10^{17} statcoulombs per gram
	1.769×10^{8} coulombs per gram
Mass of hydrogen atom to that of electron*	1,848.0
Atomic Weights	
Argon	39.94
Helium	4.002
Hydrogen	1.0078
Mercury	200.61
Nitrogen	14.008
Neon	20.183
Oxygen	16.000
Sodium	22.997

* Values from deflection measurements. Spectroscopic values differ by at most 0.1 per cent.

[1] This entire section was written by the author's associate, Dr. Karl Spangenberg, associate professor of electrical engineering at Stanford University.

[2] R. T. Birge, Values of the General Physical Constants, *Rev. Modern Phys.*, Vol. 1, No. 1, p. 1, July, 1929.

2. Motion of Electrons. *Uniform Electrostatic Field.*—The laws of motion of an electron in a uniform electrostatic field are the same as those of a body falling freely under the influence of gravity. In the case of the electron the gravitational field is replaced by the factor eF/m, when the relations are expressed in electrostatic units, where e is the charge of the electron, F is the constant gradient of potential, and m is the mass of the electron.

The principal relations for the electron in a uniform electrostatic field are given below:

Acceleration:

$$a = \frac{eF}{m} \text{ cm per sec per sec} \qquad \text{(esu)} \tag{1}$$

Velocity (starting from rest):

$$v = \frac{eFt}{m} \text{ cm per sec} \qquad \text{(esu)} \tag{2}$$

$$v = \sqrt{\frac{2eF}{m}\, s} \text{ cm per sec} \qquad \text{(esu)} \tag{3}$$

Distance traveled (starting from rest):

$$s = \frac{eFt^2}{2m} \text{ cm} \qquad \text{(esu)} \tag{4}$$

where a = electron acceleration, cm per sec per sec.
v = electron velocity, cm per sec.
e = electron charge, statcoulombs.
m = electron mass, grams.
F = gradient of potential, statvolts per cm.
E = potential, statvolts.
s = distance, cm.
t = time, sec.

In practical units the velocity acquired by an electron in falling through a given potential is

$$v = 5.97 \times 10^7 \sqrt{E} \text{ cm per sec} \tag{5a}$$

where the potential E is given in volts.

Velocity is often expressed in terms of volts, the number of volts corresponding to a given velocity being given by Eq. (5*a*). Thus by a velocity of 1,000 volts is meant the velocity that an electron would acquire upon being accelerated through 1,000 volts. This is not quite correct dimensionally, since voltage to the first power is proportional to energy, but no confusion should result if the term velocity is associated with the voltage.

Other ions follow the same law as do electrons, except that the scale factor is different because of the increased mass and in some cases also because of a different charge. The relation between velocity and potential for any ion is given by

$$v = 5.97 \times 10^7 \sqrt{\frac{Enm_e}{m}} \text{ cm per sec} \tag{5b}$$

where m_e is the mass of the electron, m the mass of the ion, and n the ratio of ion charge to electron charge.

Relativity Correction for Velocity.—Equation (5) is valid until the velocity of the charged particle is no longer negligible compared with the velocity of light. Then a correction must be made for the relativistic change of mass. This change, and the resulting error in the velocity, is less than one per cent as long as the velocity of the particle is less than one-tenth of the velocity of light. The potential at which an

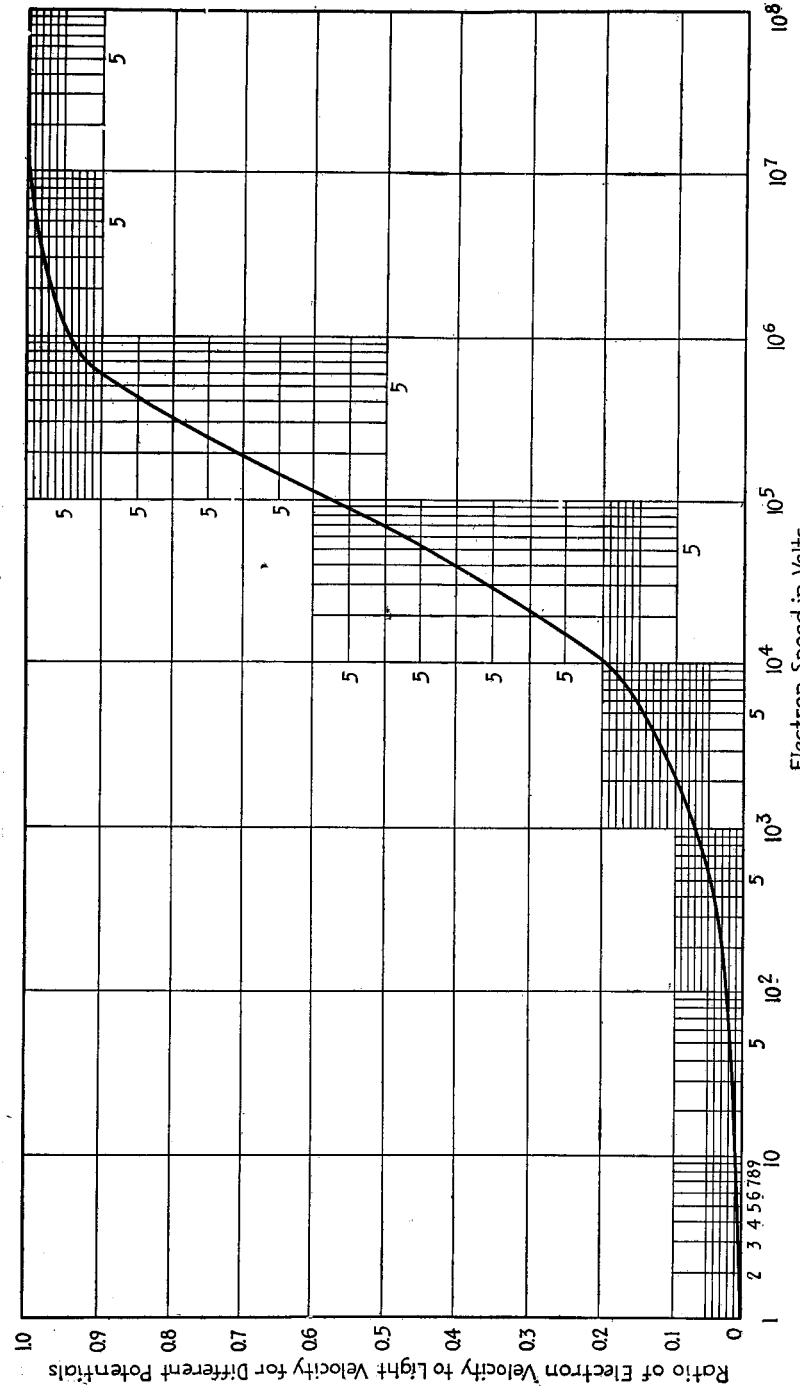


FIG. 1.—Relation between electron velocity and potential.

electron reaches this velocity is about 2,500 volts. If there were no change of mass, an electron would reach the velocity of light at about 250,000 volts.

The change in mass of an electron with velocity is given by

$$m = \frac{m_e}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \tag{6}$$

which for an electron, in terms of potential is

$$m = m_e(1 + 1.94 \times 10^{-6}E) \tag{7}$$

where m is the mass in grams of the electron at velocity v or potential E, c is the velocity of light (3×10^{10} cm per sec), and E is the potential in volts

The velocity corresponding to any potential is given by

$$v = c\sqrt{1 - \frac{1}{(1 + 1.94 \times 10^{-6}E)^2}} \tag{8}$$

A plot of this equation is given in Fig. 1.

Initial Velocity at an Angle to Gradient of Uniform Field.—If an electron traveling with an initial velocity is injected into a region of uniform field, the path of the electron in the field will be parabolic and will be like that of a projectile fired from a gun in a gravitational field.

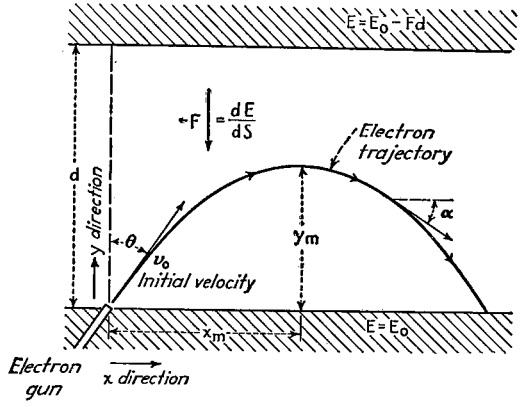


FIG. 2.—Parabolic electron trajectory in a uniform electric field.

Consider Fig. 2, which shows the path of an electron that is injected into the constant field between two electrodes. Here there is an initial component of velocity in the horizontal or x direction that is unaffected by the field, and that remains constant. The component of velocity in the vertical or y direction varies like that of a ball thrown vertically into the air.

The equation of the parabolic trajectory followed is

$$y = \frac{-Fx^2}{4E_0 \sin^2 \theta} + \frac{x}{\tan \theta} \tag{9}$$

the origin being taken at the point where the electron enters the field. The components of velocity at any point are

$$v_x = v_0 \sin \theta \text{ cm per sec} \tag{10}$$

$$v_y = \sqrt{v_0^2 \cos^2 \theta - \frac{2eFy}{m}} \text{ cm per sec} \tag{11}$$

The angle that the instantaneous velocity makes with the horizontal at any point is given by

$$\tan \alpha = \frac{-xF}{2E_0 \sin^2 \theta} + \frac{1}{\tan \theta} \tag{12}$$

The maximum height to which the electron will rise, and the x component corresponding to this position, are

$$y_m = \frac{E \cos^2 \theta}{F} \tag{13}$$

$$x_m = \frac{2E \sin \theta \cos \theta}{F} \tag{14}$$

The initial components of velocity are

$$v_x = \sqrt{Ee \sin^2 \theta} \tag{15}$$

$$v_y = \sqrt{Ee \cos^2 \theta} \tag{16}$$

In the relations given above the symbols have the following significance:

x = horizontal displacement of electron, cm.
y = vertical displacement of electron, cm.
F = gradient of potential, statvolts per cm.
θ = angle that direction of original electron velocity makes with gradient (see Fig. 2).
v_x = horizontal velocity component of electron, cm per sec.
v_y = vertical velocity component of electron, cm per sec.
v_0 = original velocity of electron, cm per sec.
E_0 = potential at point where electron enters field, statvolts.
α = slope angle of parabolic path with horizontal at any point along electron trajectory.
e = electron charge, statcoulombs.

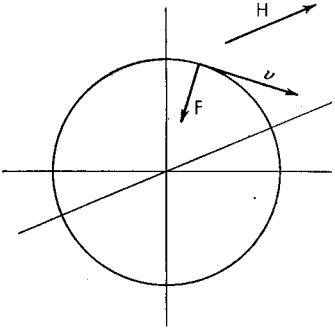


FIG. 3.—Direction of force on an electron relative to the magnetic field and the direction of velocity.

Electron in a Uniform Magnetic Field.—An electron moving at right angles to a uniform magnetic field experiences an instantaneous force whose direction is mutually perpendicular to the field and the electron velocity. This is because the electron in motion constitutes an electric current and as such is subject to the sidewise force that is experienced by a conductor in a magnetic field. The sidewise force changes the direction but not the magnitude of the velocity, with the result that the electron moves in a circular path such that the magnetic force balances the centrifugal force of the circular motion. The geometrical relation between velocity, field, and force is shown in Fig. 3. The force developed is always at right angles to both the velocity and the field.

The expression for the radius of the circular path obtained by equating the magnetic and centrifugal force is

$$r = \frac{m_e v}{eB} \tag{17}$$

where r = radius of circular path, cm.
e = electron charge, abcoulombs.
m_e = electron mass, grams.
v = electron velocity, cm per sec.
B = magnetic field, lines per sq cm.

Expressing this in practical units and introducing the expression for velocity in terms of the square root of potential:

$$r = \frac{3.37\sqrt{E}}{B} \text{ cm} \tag{18}$$

where E is the electron velocity expressed in volts. The radius of the circular path is seen to be proportional to the square root of potential and inversely proportional to

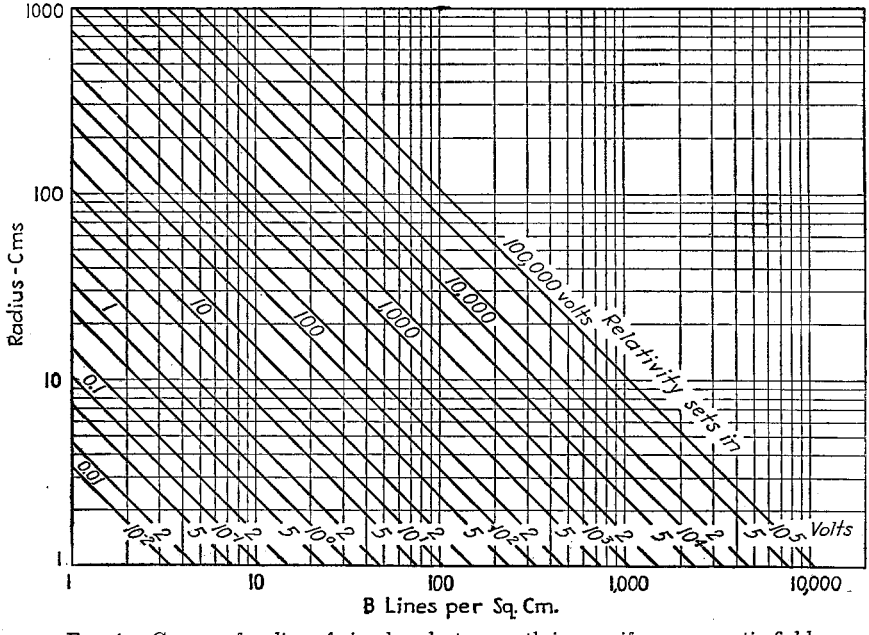


FIG. 4.—Curves of radius of circular electron path in a uniform magnetic field.

the magnetic field strength. The relation between the factors is given by the curve of Fig. 4.

If the particle is not an electron but an ion of mass m and with n times the charge of an electron, then the result of Eq. (18) must be multiplied by the factor $(m/m_e n)^{1/2}$.

The time required for the electron to make a complete circle, *i.e.*, the period, does not depend upon the voltage at which the electron is injected into the magnetic field. This is because the radius, and hence the circumference, of the circle is proportional to the velocity, thus making the period independent of either. The period T is given by

$$T = \frac{2\pi m_e}{e}\frac{1}{B} \text{ seconds} \qquad \text{(emu)} \tag{19}$$

In practical units this becomes

$$T = \frac{0.355 \times 10^{-6}}{B} \text{ seconds} \tag{20}$$

The value of the period can be obtained from the curves of Fig. 4 by observing that the period in microseconds is the same as the radius in cm when the potential is 0.01122 volts.

For particles of mass m, having n times the charge of the electron, the value of Eq. (20) must be multiplied by the factor $(m/m_e n)$.

Electron at an Angle with a Uniform Magnetic Field.—If an electron is injected into a magnetic field at an angle with the field, the initial velocity can always be resolved into components parallel and perpendicular to the field. The component parallel to the magnetic field experiences no force due to the field and so is unchanged in either magnitude or direction. The component normal to the field is subject to a constant sidewise force and so undergoes the same change in direction as that described above, giving rise to a circular component of motion. The combination of the circular and linear motion gives rise to a helical path. The equations given above are directly applicable to the circular component of motion if for the velocity there be substituted the normal component of velocity.

The radius of the helix is given by

$$r = \frac{3.37 \sqrt{E} \sin \theta}{B} \text{ cm} \tag{21}$$

where θ is the angle that the initial velocity makes with the field, E is the velocity in volts, and B is the strength of the magnetic field in lines per sq cm.

The pitch p of the helix is given by

$$p = \frac{21.2 \sqrt{E} \cos \theta}{B} \text{ cm} \tag{22}$$

All the preceding remarks have assumed a uniform magnetic field. When the field is strong enough and the electron velocity low enough then the electron will move in a tightly wound helix that will follow magnetic flux lines even though these be slightly curved. Use is made of this property in the Orthicon television pickup tube.

3. Emission.[1]—The emission of electrons from metals can occur in several ways. If a metal is heated to a sufficiently high temperature emission will occur. This is known as primary or thermionic emission. Secondary emission is that which occurs when high-velocity electrons or ions strike a metal and knock out other electrons. Emission may also occur when light rays strike a material and give up energy that liberates electrons; this is known as photoemission. Thermionic and secondary emission are of most importance in ordinary vacuum tubes, and these types of emission will be the subject of this section.

Theory of Emission.—In any conductor there are a large number of free electrons moving around with various velocities. As the temperature of the conductor increases the average velocity of the free electrons increases. At the surface of the conductor there exists a restraint upon the movement of the electrons in the form of an electrostatic gravitational force. Were it not for this all the electrons in the metal would soon escape. As it is, only those having enough energy to overcome the surface forces manage to get away. The energy in electron volts that an electron must have

[1] A good introductory survey to the subject of thermionic emission is given by S. Dushman, Electron Emission, *Elec. Eng.*, Vol. 53, p. 1054, July, 1934.

before it can leave a metal is known as the *work function* of the metal. It would be expected that the emission of electrons would depend upon the temperature of the metal and the work function of the particular metal involved. This is found to be the case.

A law giving the relation between emission current and temperature was first proposed by Richardson.[1] Later another emission equation was proposed by Dushman.[2] Dushman's equation is only slightly different from Richardson's, but is

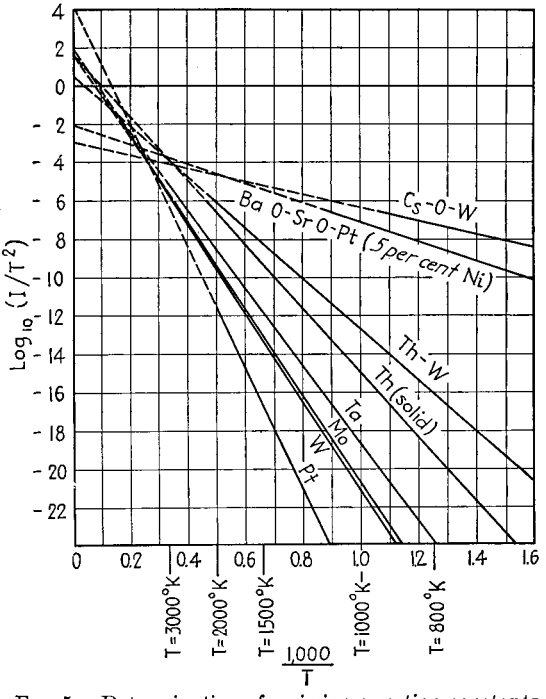


FIG. 5.—Determination of emission equation constants.

based upon better theoretical considerations and is now deemed to be the best expression available.

Dushman's equation is

$$I = AT^2\epsilon^{\frac{-b_0}{T}} \tag{23}$$

where I = current, amperes per cm^2.

A = 120.4 amperes per cm^2 deg^2, a universal constant.

T = absolute temperature, °K.

b_0 = temperature equivalent of the work function.

= 11,600 E_w, °K.

E_w = work function of the metal, volts.

The form of the emission equation is such that if the logarithm of I/T^2 be plotted against $1/T$, a straight line results whose slope gives the quantity b_0, and whose intercept on the I/T^2 axis gives the constant A. Such curves are shown in Fig. 5

[1] O. W. Richardson, "Emission of Electricity from Hot Bodies," Longmans, New York, 1921.

[2] S. Dushman, Thermionic Emission, *Rev. Modern Phys.*, Vol. 2, No. 4, p. 382, October, 1930. This paper gives a review of the subject up to its date, including extensive references.

and represent the most accurate known method of determining the emission constants. Examination of the lines in Fig. 5 shows considerable variation in the intercept, with the pure metals tending to have an intercept of about 1.78, corresponding to a value of A of 60, instead of an intercept of 2.08, corresponding to a value of A of 120.4. Since the actually observed values of the constant A range from higher to lower than the theoretical value, an explanation of the effect must be found elsewhere than in incorrectness of the theory. It has been suggested that the observed value of the constant A of 60, which is about half of the theoretically predicted value of 120.4, is due to a temperature variation of the work function. A decrease in the work function of 6 parts per 100,000 per degree would effect the observed change.

Emission of Pure Metals.—Pure metals follow the form of the Dushman equation, as indicated by the straight-line nature of the curves of Fig. 5. The constants of the commonest pure metals as determined from emission measurements are given in Table 2. The next to last column of this table gives the current per square centimeter emission at the absolute temperature given in the last column.

TABLE 2.—EMISSION OF PURE METALS

Metal	A, $\frac{\text{amperes}}{\text{cm}^2\ \text{deg}^2}$	$b_0 \times 10^{-4}$, °K	E_w, volts	I_T, $\frac{\text{amperes}}{\text{cm}^2}$	T, °K
Cs	162	2.10	1.81	2.5×10^{-11}	500
Mo	60.2	5.09	4.38	2.34×10^{-2}	2,000
Pt	1.7×10^4	7.25	6.27	9.2×10^{-10}	1,600
Ta	60.2	4.72	4.07	1.38×10^{-2}	2,000
W	60.2	5.24	4.52	1.0×10^{-3}	2,000
Th	60.2	3.89	3.35	4.3×10^{-3}	1,600
Zr	330	4.79	4.13	8.5×10^{-5}	1,600

Because of their relatively low emission, pure metals are not used much, tungsten being the only pure-metal emitter of practical importance. Extensive data on the emission properties of pure tungsten have been collected, largely by Langmuir.[1] The emission constant of platinum, as reported by various investigators, shows great variation. Most recently reported values give a smaller value of A than indicated in Table 2. The variation is probably due to the difficulty of obtaining a clean surface free from occluded oxygen.

The only metal other than tungsten that has possibilities as a practical emitter is tantalum. Although tantalum cannot be made as hot as tungsten because its melting temperature is 3300°K, compared with 3655°K for tungsten, its work function is sufficiently lower, so that at any temperature less than 2500°K, its emission is at least 10 times that from tungsten. A disadvantage of tantalum is that it is easily contaminated by residual gases, which form oxides that greatly reduce the emission.

Emission of Thoriated Tungsten.—There are many metals that have a higher emission than tungsten, but it is not possible to take advantage of this by using the metals in the pure form, because most of them melt at a relatively low temperature. It is, however, possible to coat tungsten with a thin layer of other metals and thus

[1] Well summarized by E. L. Chaffee, "Theory of Thermionic Vacuum Tubes," pp. 96–102, McGraw-Hill, New York, 1933. See also I. Langmuir, The Characteristics of Tungsten Filaments as Functions of Temperature, *Phys. Rev.*, Vol. 7, p. 302; C. Davisson and L. H. Germer, The Thermionic Work Function of Tungsten, *Phys. Rev.*, Vol. 20, p. 300, April, 1922; S. Dushman, H. N. Rowe, J. W. Ewald, and C. A. Kidner, Electron Emission from Tungsten, Molybdenum, and Tantalum, *Phys. Rev.*, Vol. 25, p. 338, March, 1925; W. E. Forsythe and A. G. Worthing, The Properties of Tungsten and the Characteristics of Tungsten Lamps, *Astrophys. Jour.*, Vol. 61, p. 146, 1925.

take advantage of the high melting temperature of tungsten and the high emission of the other metals. It has been discovered that thorium on tungsten gives the highest emission of any practical combination of this character.[1] The layers of thorium resulting from the formation procedure are molecular in thickness.

Thoriated tungsten emitters are formed by using tungsten wire that has from 1 to 2 per cent of thorium oxide mixed with it. The filament is placed in tubes in the usual way, and the tubes are degassed and baked out. It is then necessary to activate the filament to obtain the high emission. The activation procedure consists of several steps. *First*, the filament is heated to a high temperature just short of the melting temperature, such as 2800°K. This does two things: it cleans the tungsten and reduces the thorium oxide to metallic thorium. Such thorium as finds its way to the surface is immediately vaporized at this temperature. *Second*, the filament is maintained at 2100°K for a considerable time. At this temperature there is a combination action of diffusion of metallic thorium to the surface and an evaporation of some of this. The purpose of this step in the procedure is to accumulate as thick a layer of thorium as possible at the surface. The amount of thorium that gathers on the surface depends upon the relative rates of diffusion and evaporation. At 2100°K there is a pronounced maximum in the emission of thoriated filaments, indicating that at this temperature there exists the best combination of diffusion and evaporation possible for accumulating thorium on the surface. *Third*, the filament temperature is reduced to 1600°K, and some napthalene, acetylene, or other hydrocarbon gas is admitted. This causes a layer of tungsten carbide to form on the surface of the filament. Although this "carbonizing" reduces the emission somewhat, it is found that the presence of the layer reduces the subsequent evaporation of thorium by a factor of six. *Fourth*, the filament temperature is set to 1800 to 2000°K for normal operation.

Several interesting and remarkable properties of thoriated tungsten filaments should be pointed out. The thorium will adhere to the tungsten at temperatures considerably higher than that at which metallic thorium will evaporate. The work function of thoriated tungsten is lower than that of either tungsten or thorium alone. The approximate values of the work functions are

Tungsten	4.5 volts
Thorium	3.4 volts
Thoriated tungsten	2.6 volts

The reduction of the work function is apparently due to the lowering of the electrostatic surface gravitational forces by virtue of the existence of a double layer of charge caused by partially ionized electropositive metal on the surface.

Emission of Oxide-coated Cathodes.[2]—The most efficient type of emitter is the oxide-coated cathode. It consists of a metallic base of some metal or alloy with a coating of the oxides of some of the alkaline-earth metals.

The base metals, known as cores, upon which the oxide coating is formed may be of pure metal such as platinum. Better emission is, however, obtained with certain alloys such as platinum-iridium, nickel-platinum, and nickel-silicate. The best emission results from the use of an alloy of nickel, iron, cobalt, and titanium known as Konel metal.

[1] I. Langmuir, Electron Emission from Thoriated Tungsten Filaments, *Phys. Rev.*, Vol. 22, p. 357, October, 1923; S. Dushman and J. W. Ewald, Electronic Emission from Thoriated Tungsten, *Phys. Rev.*, Vol. 29, p. 857, June, 1927; W. H. Brattain and J. A. Becker, Thermionic and Absorption Characteristics of Thorium on Tungsten, *Phys. Rev.*, Vol. 43, p. 428, March, 1933.

[2] See J. A. Becker, Thermionic Electron Emission, *Bell System Tech. Jour.*, Vol. 14, p. 413, July 1935; also, *Rev. Modern Phys.*, April, 1935; J. A. Becker and R. W. Sears, Phenomena in Oxide-coated Filaments, Pt. II, Origin of Enhanced Emission, *Phys. Rev.*, Vol. 38, p. 2193, December, 1931.

The best alkaline earths are barium and strontium. The coating is applied to the core in the form of barium and strontium carbonates.

It is necessary to employ an activation procedure to the carbonate coatings. This procedure consists of heating the coating to about 1500°K for a short time and applying a voltage of about 150 volts through a protective resistor. This does two things: it reduces the carbonates to oxides thermally, and bombards the oxide coating with positive ions of considerable energy. The emission will tend to rise rapidly at these temperatures, and the cathode temperature is then reduced to about 1000°K for normal operation. The exact details of this activation, or "forming" process, vary greatly in actual commercial production.

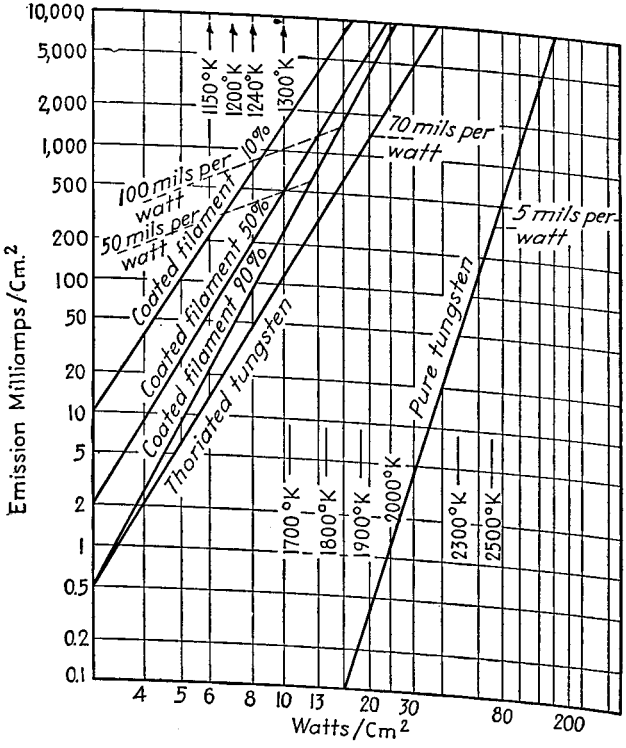


FIG. 6.—Emission efficiency of various emitters.

The mechanism of the emission of the oxide coatings is not exactly understood. Evidence exists to show that the nature of the core has a considerable influence. It is believed that there is emission from particles of the pure alkaline-earth metal that have been reduced from the oxides. The emission constants of oxide-coated cathodes are subject to some considerable variation. It is not possible to obtain the constants from curves such as those of Fig. 5 with any accuracy, because there is no sharply defined saturation current, as with tungsten and thoriated tungsten. However, the order of the work function is about 1 volt, and the emission constant A of Eq. (23) is about 0.01 ampere per cm² per degree².

Comparison of Emitters.[1]—The relative emission properties of the commonest types of emitters are shown in Fig. 6. Here are shown curves of emission against cathode heating power on a specially devised paper, known as power-emission paper, on which

[1] A good discussion is given by O. W. Pike, Cathode Design, *Communications*, Vol. 21, p. 4, October, 1941.

these curves appear as straight lines. The curves for the oxides were obtained from tests on a large number of tubes, and the percentages indicate the relative portion of tubes having higher emission than that indicated by the particular curve. The 50 per cent curve thus corresponds to an average emission for these particular oxide-coated filaments.

The curves of Fig. 6 also indicate approximately the emission efficiency of the various types of emitters. The emission efficiency is determined by the ratio of the number of milliamperes that can be obtained per square centimeter of surface area to the heating power in watts. The emission efficiency depends upon the cathode and filament structure, since the same power into different structures will raise the emitter coating to different temperatures. Some typical emission efficiencies that can be obtained at normal operating temperature are

Pure tungsten filaments.....	2–10 ma per cm^2 per watt
Thoriated tungsten filaments.....	5–100 ma per cm^2 per watt
Oxide-coated cathodes.....	10–200 ma per cm^2 per watt
Oxide-coated filaments.....	200–1,000 ma per cm^2 per watt

Practical Emitting Structures.—The physical form that the various emitters take is influenced by practical considerations. In the case of tungsten, thoriated-tungsten

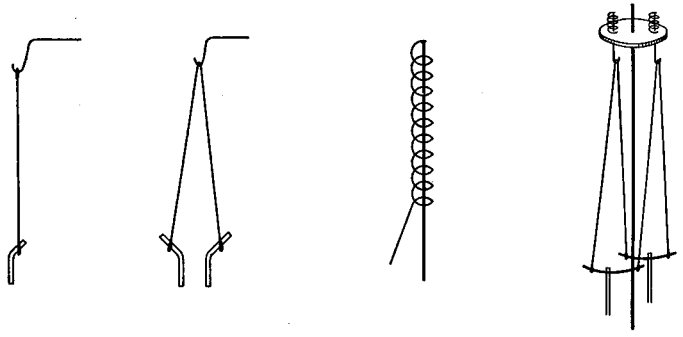


FIG. 7.—Typical filament structures.

and oxide-coated filaments, it is desired to have the emission efficiency as high as possible. This dictates a structure in which the thermal radiation is a minimum for a given wire length. It is also necessary to allow for the expansion of the wire filament. The various types of filamentary structures shown in Fig. 7 are typical. All these are arranged so that the temperature is as high as possible for a given input power. All have spring suspensions that allow for the expansion of the filament wire.

The indirectly heated cathodes consist of cylinders of thin sheet nickel on which the emitting coating is placed and in which there is contained the heater wire. The heater wire is invariably tungsten, covered with some insulating coating such as aluminum oxide. Typical arrangements are shown in Fig. 8.

Secondary Emission.—Almost all metals and some insulators will emit secondary electrons; *i.e.*, they will emit electrons when bombarded by a stream of electrons or ions.[1]

The number of secondary electrons emitted per primary electron depends upon the velocity of the primary bombarding electrons and upon the nature of the material and the condition of its surface. The number of secondaries liberated per primary electron may be greater than one. This is not a violation of the law of conservation

[1] H. E. Farnsworth, *Phys. Rev.*, Vol. 31, p. 405, 1928; Vol. 31, p. 419, 1928; Vol. 25, p. 41, 1925; Vol. 20, p. 358, 1922; L. R. G. Treolar, Measurement of Secondary Emission in Valves, *Wireless Eng.*, Vol. 15, p. 535, October, 1938; D. E. Wooldridge, Theory of Secondary Emission, *Phys. Rev.*, Ser. 2, Vol. 56, p. 562, Sept. 15, 1939.

of energy, since the velocity of the secondary electrons is lower than that of the primary electrons. Most of the secondary electrons have velocities of about 10 volts, though a small fraction have velocities approaching that of the primary electrons. The number of secondary electrons per primary electron varies typically as shown in Fig. 9. No secondary electrons are produced for primary velocities below about

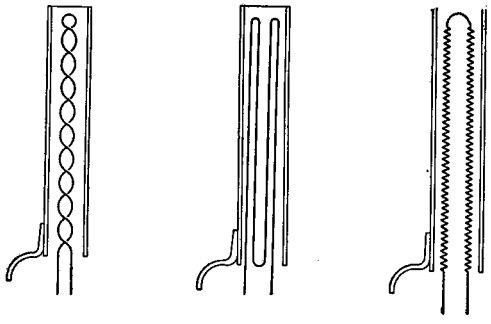


FIG. 8.—Typical heater cathode structures.

9 volts. With increasing potential of the bombarding primary electrons, the ratio of secondary to primary electrons increases, reaching a maximum in the vicinity of about 400 volts for all metals, and then decreasing. Most of the metals reach a maximum ratio of secondary to primary electrons of between one and three. Some of the complex alkali metals reach ratios of secondary to primary electrons of as high as five to eight. Maximum ratios depend greatly upon processing and surface condition.

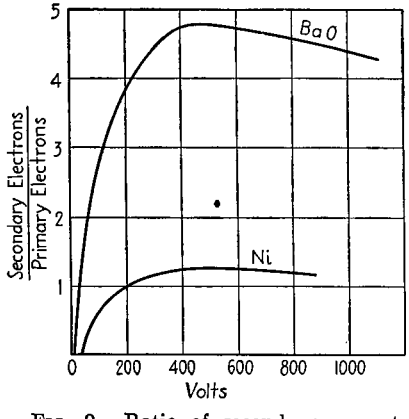


FIG. 9.—Ratio of secondary current to primary current as a function of primary potential for typical surfaces.

The number of liberated secondary electrons that succeed in escaping from a surface depends upon the potential conditions surrounding the bombarded surface. If there is an adjacent electrode at a potential higher than that of the bombarded surface, it will collect all the secondary electrons. If the adjacent electrode has a potential lower than that of the bombarded surface, the former will receive only the highest velocity secondaries, and the low-velocity electrons will be returned to the surface from which they came.

4. Space-charge Effects.—Most vacuum tubes are designed so that the emitting surface gives off a surplus of electrons, with the result that the actual current that flows is limited by the mutual repulsion between the electrons and not by the emission capabilities of the cathode. Under these conditions the current is said to be space-charge limited.

The space-charge limitation of current is brought about by the presence of electrons in the space between electrodes. These electrons introduce a negative charge that reduces the potential in the region. The reduction of potential slows down the electrons and thus increases the negative charge density, which further reduces the

potential. The potential is finally reduced to the point at which the potential gradient at the emitting surface is zero or even slightly negative. This is a limiting equilibrium condition that sets a limit to the current that can flow for a given potential difference between electrodes.

Plane Electrode Diode.—The theory relating current density, potential, and electrode spacing for a plane electrode diode was developed by Child and Langmuir[1] and has been verified experimentally.

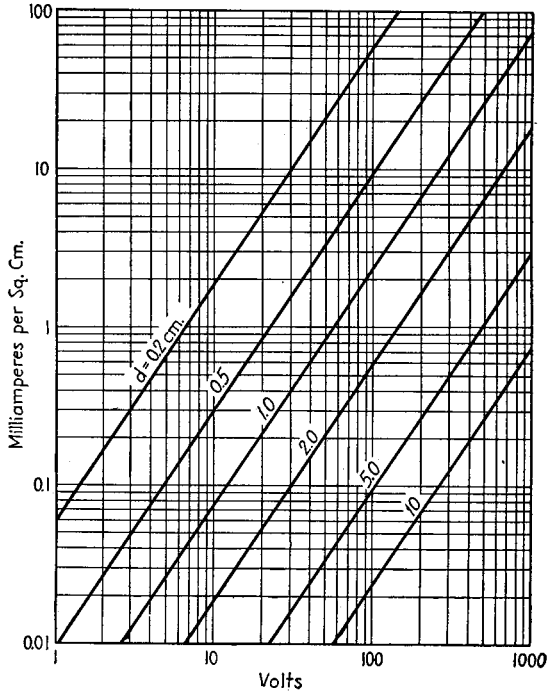


FIG. 10.—Space-charge limited current in a plane electrode diode (d = electrode spacing).

The Child-Langmuir law for parallel plane electrodes is

$$I = \frac{\sqrt{2}}{9\pi}\left(\frac{e}{m}\right)^{1/2}\frac{E^{3/2}}{d^2} \text{ statamperes per sq cm} \qquad (24)$$

where e = electron charge, statcoulombs.
E = potential difference between cathode and plate, statvolts.
m = electron mass, grams.
d = cathode-plate distance, cm.

In practical units this becomes

$$I = \frac{2.331 \times 10^{-6} E^{3/2}}{d^2} \text{ amperes per sq cm} \qquad (25)$$

A plot of this equation, from which numerical values are readily obtained, is given in Fig. 10.

[1] D. C. Child, Discharge from Hot CaO, *Phys. Rev.*, Vol. 32, p. 492, May, 1911; I. Langmuir, The Effect of Space Charge and Residual Gas on Thermionic Currents in High Vacuum, *Phys. Rev.*, Ser. 2, Vol. 2, p. 450, December, 1913.

Equation (25) shows that the current density varies as the three-halves power of potential for a given electrode spacing. The manner in which various quantities vary in the interelectrode space may be determined by treating the current as a constant in Eq. (24). When this is done it is seen that in a given tube the potential difference between electrodes required to obtain a given current I varies as the $\frac{4}{3}$ power of the distance between electrodes. Likewise, the potential gradient in the interelectrode space varies as the one-third power of the distance between anode and cathode, and is zero at the cathode. The velocity of the electrons in the interelectrode space for a given current I varies as the two-thirds power of the distance d, assuming the velocity is zero at the cathode. The space charge density varies as the negative two-thirds power of distance from the cathode, and is theoretically infinite at the cathode because of the assumption of zero initial velocity.

Cylindrical Electrode Diode.—Relations similar to those which hold for the plane electrode diode have been worked out for the cylindrical electrode diode.[1] The current-voltage law here has the form

$$\frac{I}{l} = \frac{2}{9}\sqrt{\frac{2e}{m}}\frac{E^{3/2}}{r_p\beta^2} \qquad \text{(esu)} \tag{26}$$

where I/l is the current in statamperes per unit length of axial emitter, E is the potential difference between concentric cylindrical cathode and plate in statvolts, r_p is the plate radius, and e and m are the charge and mass of the electron. The quantity β is a dimensionless factor depending upon the ratio of the plate radius r_p to the cathode radius r_c, and is given by the series

$$\beta = u - \frac{2}{5}u^2 + \frac{11}{120}u^3 - \frac{47}{3{,}300}u^4 + \cdots \tag{27}$$

in which $u = \log_e r_p/r_c$. Values of β^2 for various values of r_p/r_c are given in Table 3.

TABLE 3

r_p/r_c	β^2	r_p/r_c	β^2	r_p/r_c	β^2
1.00	0.000	6.0	0.838	30	1.091
1.50	0.116	7.0	0.887	45	1.095
2.00	0.275	8.0	0.925	67	1.089
2.50	0.405	9.0	0.955	122	1.072
3.00	0.512	10.0	0.978	221	1.053
4.00	0.665	12.0	1.012	735	1.023
5.00	0.775	16.0	1.051	2,440	1.006
		20.0	1.072	22,026	0.999

In practical units the current-voltage relation becomes

$$\frac{I}{l} = 14.68 \times 10^{-6}\frac{E^{3/2}}{r_p\beta^2} \text{ amperes per unit length} \tag{28}$$

A plot of this equation from which numerical values are readily obtained is given in Fig. 11 for a value of $\beta^2 = 1$. If β^2 does not equal one, the value from the curves of Fig. 11 must be multiplied by the value of β^2.

[1] I. Langmuir and K. B Blodgett, Currents Limited by Space Charge between Coaxial Cylinders, *Phys. Rev.*, Ser. 2, Vol. 22, p. 347, October, 1923.

A solution taking into account the velocity of emission of the emitted electrons and giving the location and magnitude of the potential minimum near the anode is obtained by E. L. E. Wheatcroft, *Jour. I.E.E.*, Vol. 86, p. 473, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 94, June, 1940.

The current-voltage law for spherical electrodes has a form similar to that for cylindrical electrodes. The current again varies as the three-halves power of potential.[1] In fact, the three-halves power relation between current and potential for a space-charge-limited condition is independent of the electrode geometry and holds for any shapes.[2]

Space Charge in the Grid-anode Region.—In screen-grid tubes it is found that under certain conditions the space charge of the electrons produces a negative gradient of

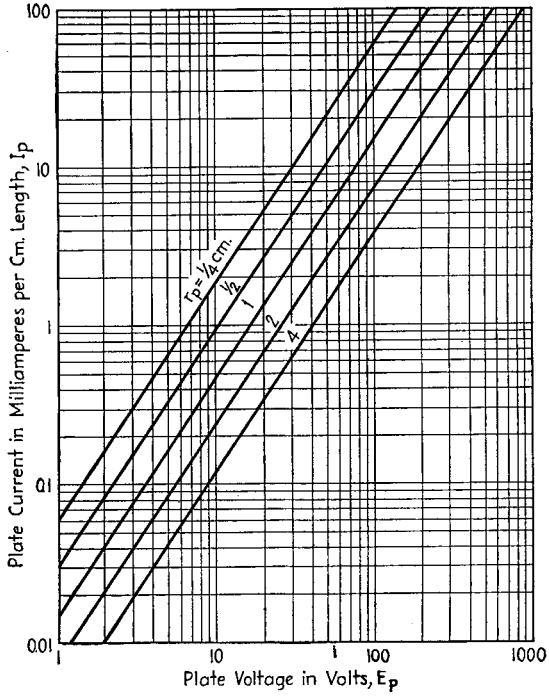


FIG. 11.—Space-charge limited current in a cylindrical electrode diode, $\beta^2 = 1$.

potential at the plate that eliminates the necessity of a suppressor grid in reducing secondary emission effects.[3] Screen-grid tubes especially designed to utilize these screen-anode space-charge effects are the *beam power tubes*. Beam tubes have aligned control and screen grids, which cause the electrons to move in sheets. They also have a screen-plate distance considerably greater than that in the conventional screen-grid tube.

A study of the effects that may occur in beam power tubes shows that the excellent plate current-voltage characteristics obtained are the result of the space-charge action

[1] I. Langmuir and K. T. Compton, Electrical Discharges in Gases, Pt. II, *Rev. Modern Phys.*, Vol. 3, p. 191, April, 1931.

[2] Langmuir and Compton, *op. cit.*

[3] J. H. O. Harries, The Anode to Accelerating Electrode Space in Thermionic Values, *Wireless Eng.*, Vol. 13, p. 190, April, 1936, "Critical Distance" Tubes, *Electronics*, Vol. 9, p. 33, May, 1936; O. H. Schade, Beam Power Tubes, *Proc. I.R.E.*, Vol. 26, p. 137, February, 1938; B. Salzberg and A. V. Haeff, Effects of Space Charge in Grid-anode Region of Vacuum Tubes, *R.C.A. Rev.*, Vol. 2, p. 336, January, 1938; C. E. Fay, A. L. Samuel, and W. Shockley, *Bell System Tech. Jour.*, Vol. 17, p. 49, January, 1938

in the screen-anode region. In this region, in addition to a possible negative gradient of potential at the anode, it is also possible to have a virtual cathode, *i.e.*, a point of zero potential and zero-potential gradient, from which electrons are partially returned to the screen grid and partially transmitted to the plate.

The different types of space-charge action that occur in the screen-anode region of a beam power tube are best classified in conjunction with the potential distributions in the interelectrode space with which they are associated. All possible potential distributions fall into one of four types. These four types are shown in Fig. 12, and have the characteristics listed below:

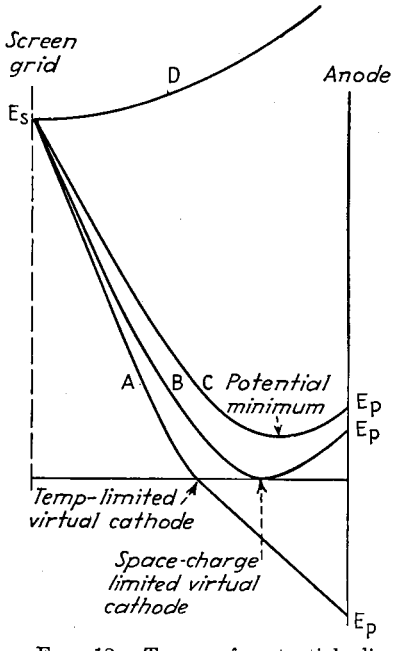


FIG. 12.—Types of potential distributions encountered in the screen-plate region of beam power tubes.

Type A.—Plate is at a negative potential. All the current injected into the region is reflected from a point at zero potential but having a negative gradient of potential. The potential distribution on the anode side of the point of zero potential is a straight line. On the screen side of the zero-potential point the potential distribution is the same as for a temperature-limited case in which the emission current is assumed to be twice the current injected into the screen-anode space through the screen.

Type B.—Both screen grid and plate are positive, with a single point at both zero potential and zero-potential gradient between them, resulting in a space-charge-limited virtual cathode. Part of the current injected through the screen is then transmitted to the plate, and the remainder is returned to the screen. The potential distribution is then proportional to the four-thirds power of distance on both sides of the virtual cathode, and the anode current is proportional to the three-halves power of anode voltage.

Type C.—Both screen grid and plate are positive, with a potential minimum at a positive potential between them. All the injected current is then transmitted to the plate.

Type D.—Both screen grid and plate are positive, with no potential minimum between them. All the injected current is transmitted to the plate.

The most important potential distribution formulas and current-voltage relations associated with the potential distributions are summarized below. The notation used is as follows:

x = distance from screen grid to a point between screen grid and plate where the potential is E.

$x_0 = 1.527 \times 10^{-3} \frac{E_s^{3/4}}{I^{1/2}}$ cm, cathode-plate spacing of an equivalent plane electrode diode, which would give a current I with a plate potential E_s.

E_s = screen-grid potential, volts.

I = current injected into the screen-grid-plate region from the screen side, amperes.

b = constant appearing in the Type A distributions and related to the straight-line portion of the potential distribution curve beyond the virtual cathode by

$$\frac{dE}{dx} = \frac{-4\sqrt{2}}{3}\frac{E_0}{x_0}b^{-1/4}.$$

E = potential at point x.

E_0 = potential at point x_0.

x_1 = distance from the virtual cathode of a point at potential E in the Type B distributions, and from the potential minimum in the Type C distributions when this point lies between the screen grid and the minimum.

x_2 = distance from the virtual cathode of a point at potential E in the Type B distribution, and from the potential minimum in the Type C distribution, when this point lies between the potential minimum and the plate.

T = fraction of the current I injected through the screen grid that reaches the plate in the Type B distributions.

The potential distributions of the Type A are like those of a temperature-limited diode, and are given by

$$x = \frac{x_0}{\sqrt{2}}\left\{(1 - 2b^{1/2})(1 + b^{1/2})^{1/2} - \left[\left(\frac{E}{E_s}\right)^{1/2} - 2b^{1/2}\right]\left[\left(\frac{E}{E_s}\right)^{1/2} + b^{1/2}\right]^{1/2}\right\} \quad (29)$$

The potential distributions of the Type B are like those of a space-charge-limited diode. The potential rises from a zero value at the virtual cathode located between screen grid and plate, and follows a four-thirds power variation with distance on either side of the virtual cathode, with the scale depending upon the fraction of the current transmitted past the virtual cathode. The potential distribution between the virtual cathode and the screen grid is given by

$$x_1 = x_0\left[\frac{1}{(2 - T)^{1/2}} - \frac{\left(\frac{E}{E_s}\right)^{3/4}}{(2 - T)^{1/2}}\right] \quad (30)$$

The potential distribution between the virtual cathode and plate is given by

$$x_2 = x_0\left[\frac{1}{(2 - T)^{1/2}} + \frac{\left(\frac{E}{E_s}\right)^{3/4}}{T^{1/2}}\right] \quad (31)$$

The potential distributions of the Type C are characterized by a potential minimum between the screen grid and plate occurring at a positive potential. The distribution on the screen side of the potential minimum is given by

$$x_1 = -x_0\left\{\left[\left(\frac{E}{E_s}\right)^{1/2} + 2\left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]\left[\left(\frac{E}{E_s}\right)^{1/2} - \left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]^{1/2} + \left[1 + 2\left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]\left[1 - \left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]^{1/2}\right\} \quad (32)$$

The distribution to the right of the potential minimum, *i.e.*, between the potential minimum and the plate, is given by

$$x_2 = x_0\left\{\left[\left(\frac{E}{E_s}\right)^{1/2} + 2\left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]\left[\left(\frac{E}{E_s}\right)^{1/2} - \left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]^{1/2} + \left[1 + 2\left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]\left[1 - \left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]^{1/2}\right\} \quad (33)$$

The potential minimum occurs at a distance

$$x_{\min} = x_0 \left[1 + 2\left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]\left[1 - \left(\frac{E_{\min}}{E_s}\right)^{1/2}\right]^{1/2} \tag{34}$$

from the screen.

Potential distributions of the Type D are not of great importance, since for this distribution all the injected current is transmitted to the plate.

It is found that there is a maximum current that can be passed through the screen-grid–plate region for any combination of potentials, current, and electrode spacing. This maximum plate current $I_{p_{\max}}$ that can be passed is

$$I_{p_{\max}} = \frac{2.33 \times 10^{-6} \times (E_s^{1/2} + E_p^{1/2})^3}{a^2} \text{ amperes per cm}^2 \tag{35}$$

where a is the grid-anode distance in cm and E_p is the plate voltage.

The most important consequences of the preceding potential distributions are that tubes of the beam power type may exhibit some unusual dynamic characteristics and that the family of plate current-voltage curves has a larger region in which the curves are approximately parallel and equally spaced than is the case with the usual pentode. The plate current-voltage curves are similar to those of a pentode except that the transition between the region where the plate current increases with plate voltage and the region in which the plate current is substantially constant is smaller and occurs at a lower plate potential than for the pentode, with the result that the "shoulder" formed by this region of decreasing slope is sharper for the beam power tube than for the pentode (see Fig. 20). The reason that the plate current-voltage curves of the beam power tube are steeper at low plate voltages than are the corresponding portions of the pentode characteristics is that in the beam power tube the action of the virtual cathode between the screen grid and the plate is more uniform in the absence of the suppressor grid of the pentode, which tends to give a kind of variable-mu action.

5. Tube Noise.[1]—Vacuum tubes act as sources of noise because of the inherent electronic nature of their operation. Noise arises from the fact that the current consists of the movement of discrete particles rather than the flow of a continuous fluid, giving rise to what is known as the *shot effect*. If the emitting material is irregular in its nature, there may also be large low-frequency variations resulting from chunks of the emitting material varying in their emission, causing what is known as the *flicker effect*. Noise will also arise from variations in the secondary emission, from ionization within the tube, and from a random variation of the division of current between elements in multielectrode tubes. Of all the above-mentioned effects, the largest and most important source of noise is the shot effect.

Shot effect is noise due to the fact that electrons are discrete particles emitted from the cathode in a random way, so that any current resulting from such emission has a random or statistical variation that is termed "noise." The energy correspond-

[1] For further information on noise in tubes the reader is referred to the following series of papers published in the *R.C.A. Rev.* under the general heading Fluctuation Noise in Space-charge-limited Currents at Moderately High Frequencies: Pt. I, B. J. Thompson, General Survey, Vol. 4, p. 269, January, 1940; Pt. II, D. O. North, Diodes and Negative Grid Triodes, Vol. 4, p. 441, April, 1940; Vol. 5, p. 106; July, 1940; Pt. III, D. O. North, Multicollectors, Vol. 5, p. 244, October, 1940; Pt. IV, B. J. Thompson and D. O. North, Fluctuations Caused by Collision Ionization, Vol. 5, p. 371, January, 1941; Pt. V, W. A. Harris, Fluctuations in Vacuum Tube Amplifiers and Input Systems, Vol. 5, p. 505, April, 1941; Vol. 6, p. 114, July, 1941.

Other articles of especial importance are D. A. Bell, A Theory of Fluctuation Noise, *Jour. I.E.E.*, Vol. 82, p. 522, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 97, June, 1938; A. J. Rack, Effect of Space Charge and Transit Time on the Shot Noise in Diodes, *Bell System Tech. Jour.*, Vol. 17, p. 592, October, 1938.

ing to these variations (*i.e.*, the energy of the resulting a-c components of the current) is found to be proportional to the magnitude of the average or d-c current. The noise energy is distributed evenly over the frequency spectrum. The shot noise in a tube, as well as depending upon the magnitude of the current, also depends upon the conditions under which this current flows. The noise for a given current is maximum when the plate is absorbing all the electrons that are liberated by the cathode, *i.e.*, when the emission is temperature limited. If the plate does not receive all the electrons emitted by the cathode, as happens when there is a very copious emission causing the current to be limited by the space-charge repulsion between electrons, then the noise is much less because of a cushioning effect upon the variations in the rate of emission produced by the great number of electrons emitted.

The rms fluctuation current from a temperature-limited diode, measured in a frequency band Δf, is given by[1]

$$i^2 = 3.18 \times 10^{-19} I \, \Delta f \tag{36}$$

where i = rms fluctuation current, amperes.
I = diode current, amperes.
Δf = frequency band in which the noise is measured in cycles per sec.

It is convenient to compare the noise from tubes with that which occurs in resistors because of thermal agitation of the electrons (see Par. 26, Sec. 5). The fluctuation current in a short-circuited resistor of resistance R at 290°K (63°F) is

$$i^2 = \frac{1.59 \times 10^{-20} \, \Delta f}{R} \tag{37}$$

where i = rms value of noise current, amperes.
Δf = width of the frequency band in which the noise is measured, cycles per sec.
R = resistance of the short-circuited resistor producing the noise, ohms.[2]

When Eq. (36) is compared with Eq. (37), it is seen that the resistance R_{eq}, which would produce the same noise as a temperature-limited diode with a current I, is given by

$$R_{eq} = \frac{0.05}{I} \text{ ohms} \tag{38}$$

In a tube in which the current to the plate is space-charge-limited, it is found that the noise is less than given by Eq. (36). Over a wide range of conditions in which only a small fraction of the emitted current reaches the plate (as is the case for all tubes operated at their rated conditions), the square of the rms fluctuation current is approximately 0.04 as great as given by Eq. (36) for the anode current in question, *i.e.*, the noise energy of a diode operating at a given current is roughly 0.04 times as great when the current flow is space-charge-limited as it is when it is temperature limited. For the diode operating under a condition of space-charge-limited current flow, the value of resistance that would produce the same noise energy as does the diode is given by[3]

$$R_{eq} = \frac{0.002}{I} \text{ ohms} \tag{39}$$

[1] W. Schottky, Spontaneous Current Fluctuations in Various Conductors, *Ann. Physik*, Vol. 57, p. 541, 1918.

[2] H. Nyquist, Thermal Agitation of Electric Charge in Conductors, *Phys. Rev.*, Vol. 32, p. 110, July, 1938.

[3] The case where part of the flow is space-charge-limited and part is temperature limited is covered by D. A. Bell, Fluctuation Noise in Partially Saturated Diodes, *Jour. I.E.E.*, Vol. 84, p. 723, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 177, June, 1939.

where R_{eq} is an equivalent short-circuited noise-producing resistance in ohms and I is the plate current of the diode in amperes.

The equivalent noise-producing resistance of a space-charge-limited triode is[1]

$$R_{eq} = \frac{\theta}{\sigma}\frac{T}{T_0}\frac{1}{g_m} \tag{40}$$

where R_{eq} = that resistance which, if inserted in the grid circuit of the given tube, would produce as much noise energy as does the tube itself.

θ = approximately 0.667, the ratio of noise energy of tube to noise energy of a resistance equal to dynamic tube resistance and at cathode temperature.

σ = approximately 0.75, the ratio of the transconductance of the triode to the conductance of an equivalent diode (*i.e.*, the diode having a cathode-plate spacing equal to the cathode-grid spacing of the triode and having a potential of $(E_g + E_p/\mu)$ on the plate).

T = cathode temperature, °K = 273 + C°.

T_0 = room temperature, °K.

g_m = transconductance of the triode, mhos.

In a typical triode for which T = 1000°K, T_0 = 300°K, with the use of the approximate value of θ and σ indicated above, one has approximately

$$R_{eq} = \frac{3}{g_m} \text{ ohms} \tag{41}$$

where g_m is in mhos.

The noise energy in pentode, beam, and screen-grid tubes is higher than in triodes with similar characteristics, because there is an added component of noise due to the random division of current between the screen and anode. The value of the resistance that, if inserted in the grid circuit of a pentode and similar tubes, would produce as much noise energy as does the tube itself is approximately

$$R_{eq} = \frac{I_b}{I_b + I_{c2}}\left(\frac{2.5}{g_m} + \frac{20I_{c2}}{g_{m2}}\right) \text{ ohms} \tag{42}$$

where I_b = plate current, amperes.

I_{c2} = screen current, amperes.

g_m = transconductance of the pentode, mhos.

The noise energy from a pentode will be about three to five times as great as that from a triode producing an equivalent amplification.[2]

As previously indicated, there are other sources of noise in tubes, such as ionization, secondary emission, emission of positive ions, and reflection of electrons from a virtual cathode, that are not included in the foregoing discussion. These sources produce noise that is usually low compared to that from the shot effect alone. In particular, noise from ionization may generally be ignored when the grid gas current is less than a few hundredths of a microampere.[3]

VACUUM TUBES

6. Basic Tube Types. *Diodes.*—A diode is a two-element vacuum tube containing an emitter of electrons, known as the cathode, and a collector electrode, termed the anode or plate. The emitter may be either an indirectly heated cathode or a filament. The chief uses of diodes are as detectors, and as rectifiers for either control or power

[1] North, *loc. cit.*
[2] North, *loc. cit.*, and Harris, *loc. cit.*
[3] Thompson and North, *loc. cit.*

purposes. The detector and rectifier tubes differ in size and current capacity but have the same general type of characteristics.

The plate current-voltage characteristics of a typical diode are shown in Fig. 13. For the lower portion of the curve the current is approximately proportional to the three-halves power of the anode voltage, because the current is limited by the space charge. For higher values of voltage, the current ceases to increase at the three-halves power rate and finally begins to flatten off with increasing voltage at a point determined by the emission of the cathode, which, in turn, depends primarily upon the cathode temperature. The higher the cathode temperature the higher the emission and the higher the voltage at which the current curve begins to flatten or "saturate."

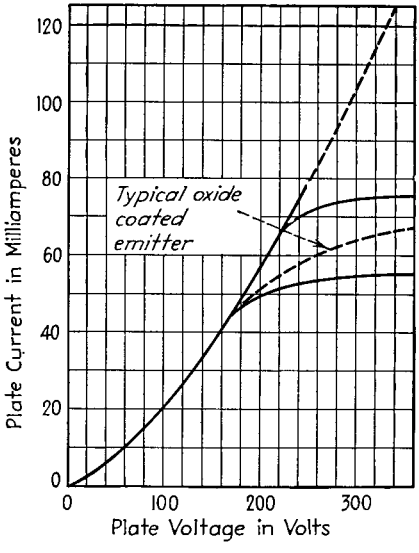


FIG. 13.—Current-voltage characteristics of a diode.

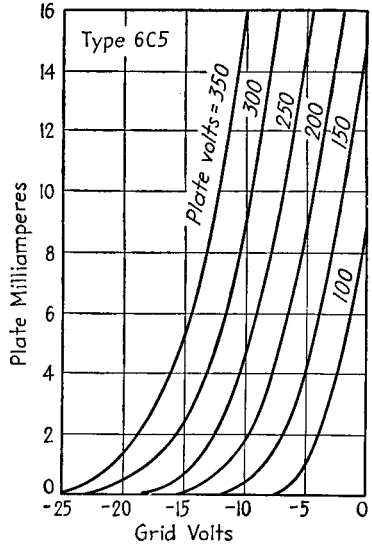


FIG. 14.—Plate-current–grid-voltage characteristics of a triode.

All current-voltage characteristics of diodes are similar at low voltages in the region in which the emission is space-charge limited. However, the portion of the plate current-voltage curves in which the current is limited by the cathode emission does not have the same shape for all types of tubes, but rather depends upon the nature of the electron-emitting surface. Thus oxide-emitting surface has less tendency to saturate sharply than a tungsten or thoriated-tungsten surface, as shown in Fig. 13.

Triodes.—A triode is a three-electrode vacuum tube containing an emitting electrode called the cathode, a control electrode called the grid, and a current-collecting electrode called the anode or plate.

The emitting electrode may be either an indirectly heated cathode, an oxide-coated filament, or a filament of tungsten or thoriated tungsten.

The control electrode, usually in the form of a grid of fine wire, surrounds the emitter and is, in turn, surrounded by the plate. The grid is usually operated at a slight negative potential and thus does not attract electrons. By virtue of its proximity to the cathode, it is able to influence the electrostatic field at the cathode to a greater extent than can the plate, and thus the grid is able to control the flow of current from the cathode without itself taking any of this current.

Some typical characteristics of a triode, showing the way in which the plate current varies with grid voltage for various fixed values of plate voltage, are given in Fig. 14. The plate current is seen to increase if either the plate voltage or grid voltage is increased. The increase in plate current for a given increase in grid voltage is always much larger than the increase in plate current for the same increase in plate voltage.

The relative effectiveness of the plate and grid potentials in controlling the plate current is known as the amplification factor of the tube, symbol μ. Specifically, the *amplification factor* of a triode or any multi-electrode tube is the ratio of a small change in plate voltage to a small change in control grid voltage under the conditions that the plate current remains unchanged and that all other electrode voltages are maintained constant. The sense is usually taken as positive when the voltages are changed in opposite directions. In mathematical notation

$$\mu = -\left(\frac{dE_p}{dE_g}\right)_{I_p \text{constant}} \tag{43}$$

Observation of the curves of Fig. 14 shows that they are all similar but are only displaced from one another. This is because the plate current depends upon an equivalent voltage of value $(E_g + E_p/\mu)$ in the same way that a diode depends upon its plate voltage E_p. The current to the plate of a triode may be written as

$$I_p = k\left(E_g + \frac{E_p}{\mu}\right)^{3/2} \tag{44}$$

where I_p = plate current.
k = a constant, sometimes known as the perveance.
E_g = grid voltage.
E_p = plate voltage.
μ = amplification factor.

In work with vacuum-tube circuits it is desirable to know the resistance of the plate circuit of the tube to a small superimposed a-c voltage. This resistance is known as the *dynamic plate resistance* of the tube, and is generally designated by the symbol R_p. The mathematical definition of the dynamic plate resistance in terms of increments of plate voltage and current is

$$R_p = \left(\frac{dE_p}{dI_p}\right)_{E_g \text{constant}} \tag{45}$$

Another factor of interest in triodes and multi-electrode tubes is the relation between the a-c components of grid voltage and plate current. The ratio between these two is known as the *transconductance* or *mutual conductance* and is defined by

$$g_m = \left(\frac{dI_p}{dE_g}\right)_{E_p \text{constant}} \tag{46}$$

It will be observed from the defining relations of the tube constants mentioned above that the product of the transconductance and plate resistance is numerically equal to the amplification factor

$$g_m R_p = \mu \tag{47}$$

Other representations of the triode than the plate-current–grid-voltage characteristics of Fig. 14 are possible. In Fig. 15 is shown a typical plate-current–plate-voltage characteristic of a triode. Another possible representation of the triode

characteristics is the constant-current characteristics of Fig. 16. These curves show loci of constant current and indicate the combinations of plate and grid voltage necessary to keep the plate current constant.

For each of the three representations of the triode characteristics given, the slope of the characteristics is related to one of the tube constants. For the plate-current–

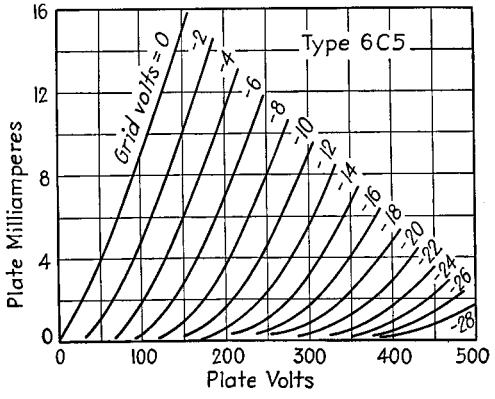


FIG. 15.—Plate-current-voltage characteristics of a triode.

grid-voltage characteristics of Fig. 14, the slope of the characteristic at any point is the transconductance of the tube at the potentials involved. In the plate-current–plate-voltage characteristics of Fig. 15, the slope of the characteristic at any point gives the reciprocal of the dynamic plate resistance. In the constant-current representation of Fig. 16 the slope of the characteristic gives the negative reciprocal of the amplification factor of the tube.

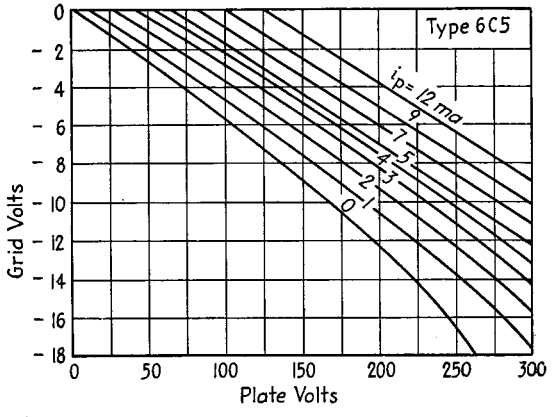


FIG. 16.—Constant-current curves of a triode.

Examination of the triode characteristics shows that the so-called tube "constants" are really not constants but vary considerably with plate current. The nature of this variation is shown in Fig. 17. It is seen that the amplification constant is most constant but tends to drop with low current. This drop occurs because of a variable-mu action that comes into play near cutoff. Some parts of the tube struc-

ture have a lower mu than do others, and so tend to cut off at a lower voltage.[1] The transconductance varies nearly as the one-third power of plate current, and the plate resistance as the reciprocal of this.

Pentodes.—A pentode is a five-electrode high-vacuum tube. The five electrodes, in the order in which they occur in the tube, are cathode, control grid, screen grid, suppressor grid, and plate. The control grid is operated at a slight negative potential relative to the cathode. The screen grid is operated at a relatively large positive potential relative to the cathode. The suppressor grid is operated at cathode potential. The plate is operated at a large positive potential relative to cathode, and is generally but not always larger than the screen-grid potential. The control grid is usually of very fine mesh in order to obtain a high transconductance. The screen-

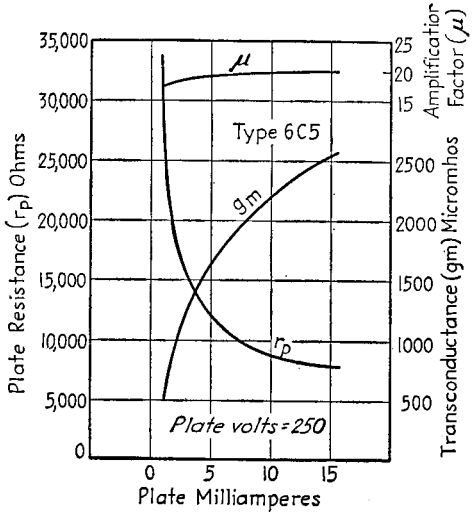


FIG. 17.—Variation of triode constants with plate current.

grid mesh is less fine than that of the control grid. The suppressor-grid mesh is very coarse so as to obstruct the flow of electrons as little as possible.

The pentode resulted from efforts to improve the characteristics of the triode and screen-grid tube. The screen grid eliminates substantially all direct capacity between control grid and plate. The suppressor grid prevents the exchange of secondary electrons between screen and plate, as occurs in the screen-grid tube. This the suppressor grid is able to do because it is operated at zero potential and hence produces a negative gradient of potential at the plate that causes secondary electrons liberated from the plate to be returned to the plate. The suppressor grid, though operated at zero potential, is able to pass electrons freely through the spaces between the wires. The suppressor grid also aids the shielding action of the screen grid by introducing some additional electrostatic shielding between the control grid and the plate.

The total current passed by the control grid to the screen grid is determined mainly by the control-grid potential, to a lesser degree by the screen-grid potential, and hardly at all by the plate potential. Thus

$$I_s = k\left(E_g + \frac{E_s}{\mu_{sg}}\right)^{3/2} \tag{48}$$

[1] See F. E. Terman and A. L. Cook, Note on the Variations in the Amplification Factor of Triodes, *Proc. I.R.E.*, Vol. 18, p. 1044, June, 1930.

where I_s = space current, the sum of screen-grid and plate current.
k = a constant determined by the tube electrode geometry.
E_g = control-grid potential.
E_s = screen-grid potential.

$\mu_{sg} = \left(\frac{dE_s}{dE_g}\right)_{I_s \text{constant}}$, analogous to the amplification factor of a triode.

The fraction of the space current that is delivered to the plate depends upon the relative screen-grid and plate potentials. When the plate potential is larger than

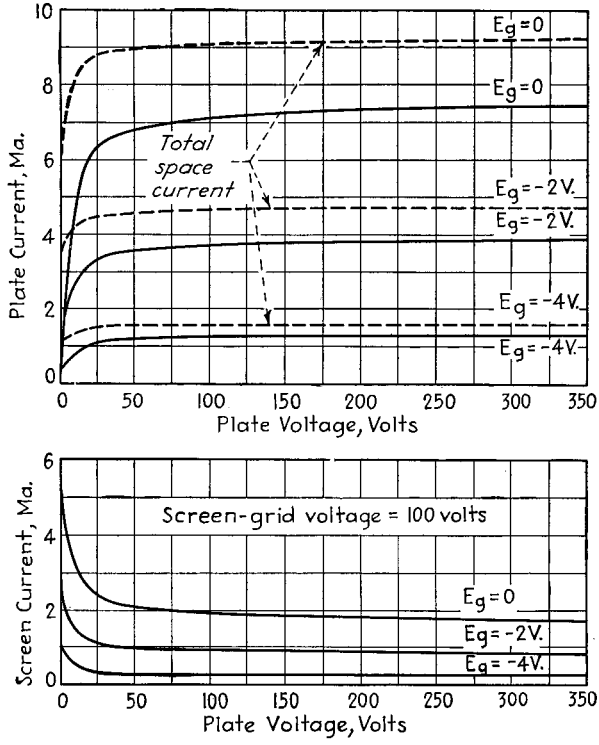


FIG. 18.—Plate-current–plate-voltage characteristics of a pentode; space-current–plate-voltage characteristics of a pentode; screen-current–plate-voltage characteristics of a pentode.

the screen potential, the screen grid intercepts a fraction of the space current that is a little larger than the ratio of projected area of the screen-grid wires to the area of the cylindrical surface in which the screen grid lies. When the plate potential is much less than the screen-grid potential, some of the electrons passed by the screen grid fail to reach the plate and are turned back to the grid, thus decreasing the fraction of the total current that is delivered to the plate.

As a result of the preceding action, both space current and plate current increase only very slightly with plate voltage when the plate voltage exceeds a certain minimum, and the plate current is then nearly a constant fraction of the space current. When the plate voltage is very low compared with the screen-grid voltage, the plate current drops as the plate voltage is decreased, and the space current drops too. The

plate current drops because a smaller fraction of the space current is delivered to the plate, and because the space current drops somewhat. The space current drops because the current returned from the screen-grid–plate region increases the space charge around the cathode. Some typical curves showing how space current and plate current vary with plate voltage are shown in Fig. 18.

The screen-grid current at low plate voltages decreases as plate potential increases because a smaller fraction of the current originally transmitted by the screen grid is returned from the screen-grid–plate region, and because the screen intercepts a smaller fraction of the passing electrons. Current is returned from the screen-grid–plate

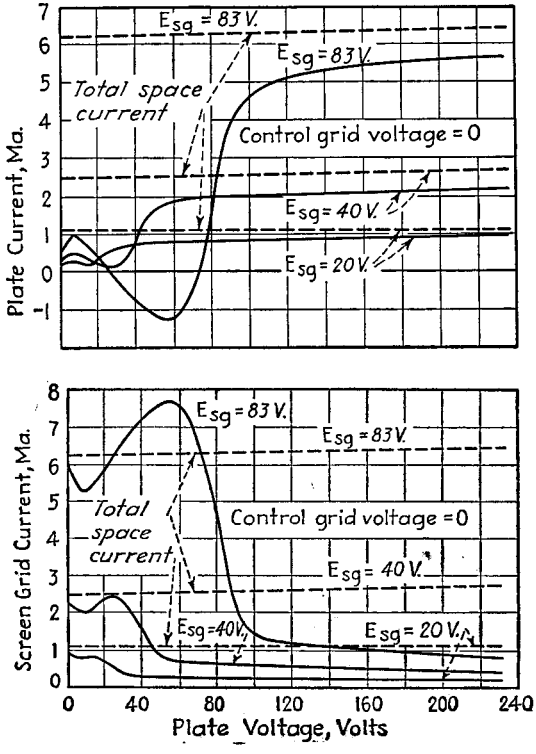


FIG. 19.—Plate and screen-grid current characteristics of a screen-grid tube.

region when the plate potential is low, because then the electrons are slowed down in the vicinity of the suppressor grid so much that some of them are actually stopped, creating a region of zero potential known as a virtual cathode. Electrons with velocity higher than average will get past this region, while those with velocity less than average will be reflected from this region back toward the screen grid. Curves of screen current as a function of plate voltage are also shown in Fig. 18.

Tetrodes.—A tetrode is a four-element vacuum tube. The four elements are cathode, control grid, screen grid, and plate. The tetrode is also known as a screen-grid tube.

The tetrode is the historical predecessor of the pentode. It was designed to have a low electrostatic capacity between control grid and plate, this low capacity being achieved by inserting an extra grid, known as the screen grid, between the control grid and plate of a triode. The tube was successful in providing the low capacity

desired but suffered distortions in the plate-current–plate-voltage characteristics because of secondary emission effects within the tube. These detrimental effects are overcome in the pentode by the addition of the suppressor grid between screen grid and plate, and the pentode has superseded the screen-grid tube in all but a few applications.

If there were no such thing as secondary emission, the current-voltage characteristics of a screen-grid tube would look like those of a pentode. As it is, they have the form shown in Fig. 19. This is because the electrode currents are made up not only of the primary electrons, *i.e.*, the electrons emitted by the cathode, but also of secondary electrons, which are electrons knocked out of other electrodes by the impact of the primary electrons.

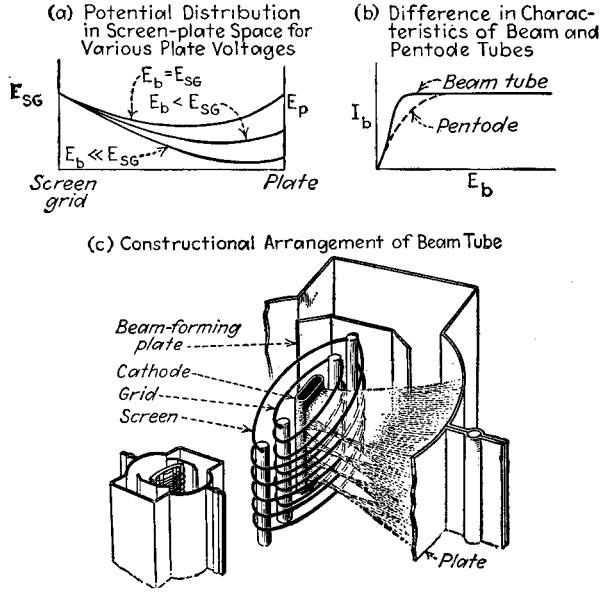


FIG. 20.—Characteristics and constructional features of the beam power tube.

Secondary electrons are always emitted when primary electrons strike an electrode with an energy of more than about 10 electron volts. The reason that the secondary-electron effects have not appeared in the triode and pentode characteristics is that the potentials within these tubes are such that under ordinary operating conditions the secondary electrons are drawn back to the electrode from which they originated.

In screen-grid tubes the secondary electrons from the screen are attracted to the plate when the plate potential is higher than the screen-grid potential, and the secondary electrons from the plate are attracted to the screen grid when the screen-grid potential is more positive than the plate potential. The latter effect is more pronounced because the plate receives more electrons and so gives rise to more secondary electrons than does the screen grid. The effect of the action indicated above results in the characteristics shown in Fig. 19. The dip in the plate-current–plate-voltage curves occurs in the region in which the plate potential is lower than the screen potential and is caused by the fact that the plate current is decreased by the amount of the secondary-electron current that flows from plate to screen grid. In this same region the screen-grid current is increased by the amount of the secondary-electron

current that flows from plate to screen grid. The amount of the reduction of plate current in this region may be sufficient to cause the plate current to be negative.

Beam Power Tubes.—A beam power tube is a tetrode in which the effect of a suppressor grid is achieved by a special design that causes the mutual repulsion between the electrons in the screen-grid–plate region to depress the potential. This effect is achieved by using a long screen-grid–plate spacing and by causing the electrons to move in sheetlike beams, thus increasing the electron density.

The construction of a beam power tube is shown in Fig. 20. The control grid and screen grid are of the same pitch and are aligned so that the screen-grid wires lie in the shadow of the control-grid wires. End plates at cathode potential are used to keep the electrons flowing in a rectangular section of the tube. The dimensions of the grids are chosen so that the electrons are focused in the interspaces of the screen grid. This causes the screen-grid current to be much lower than the value calculated on the basis of the intercepted area.

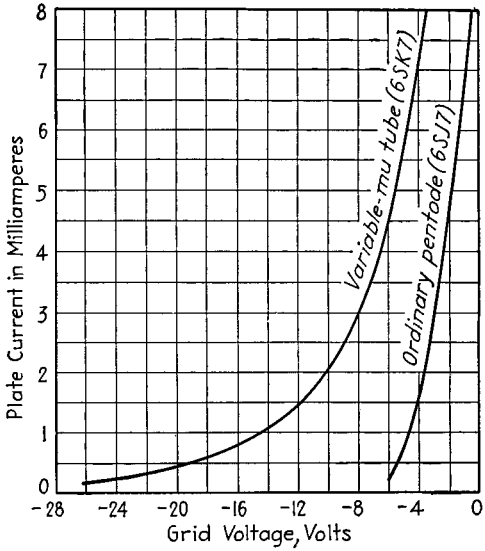


FIG. 21.—Characteristic curve of a variable-mu pentode compared with corresponding characteristic of an ordinary pentode.

The distinguishing features of the beam-power-tube characteristics are that the plate current rises very rapidly with plate voltage at low plate voltages, and then flattens off very sharply (see Fig. 20*b*). The curves show no secondary emission action except at very low currents. Here secondary-emission effects evidence themselves because the current is not high enough to depress the interelectrode potential to the point where a minimum will appear and present a negative gradient of potential to secondary electrons leaving the plate. The screen current decreases with increasing plate voltage as expected, and is relatively small in magnitude if the plate voltage is not too small.

In the screen-grid–plate region the electrons move in essentially parallel paths, and the space charge causes the potential here to be reduced. If the plate voltage is extremely low, the potential in the screen-grid–plate region is reduced to the point where a virtual cathode is formed before the plate. There is then a plane of zero potential before the plate, from which the potential rises on both sides according to

the same law as when one has the cathode of a diode whose emission is space-charge-limited (see curve *B* of Fig. 12). It is the presence of this virtual cathode that causes the plate current to rise so rapidly with plate voltage with small plate voltages. The rapid rise is due to a combination of increasing transmission of current past the virtual cathode, and movement of the virtual cathode toward the plate, causing further increase of the current. The rise in plate current with plate voltage is much more rapid in the beam power tube than in the usual pentode, and results in the characteristics shown in Fig. 20*b*.

Variable-mu Tubes.—A variable-mu tube is a tube, usually a pentode, that has a gradual cutoff as the control grid is made negative, as shown in Fig. 21. The variable-mu characteristic is obtained by employing a control-grid structure having a varying pitch between grid wires. Such a tube acts like a tube made up of a number of tubes with successively decreasing amplification factors in parallel. As the bias on such a tube is increased, the currents in the component parts are cut off one by one, until only those parts corresponding to the lowest mu portions of the grid structure are conducting. Variable-mu tubes are also known as remote cutoff tubes, and "supercontrol" tubes.

Coefficients of Screen-grid, Beam, and Pentode Tubes.—The number of possible coefficients of tubes with four or more electrodes becomes very large. These naturally divide into three types of coefficients, the dynamic resistances of the various electrodes, particularly the screen and plate, the mu-factors, and the transconductances.

The dynamic electrode resistance is the resistance that the electrode offers to a small increment in applied voltage. In the case of beam, pentode, and screen-grid tubes the resistances of the plate and screen circuits are of importance. They are defined as follows:

$$\text{Plate resistance} = R_p = \left.\frac{dE_p}{dI_p}\right|_{E_g \text{ and } E_{sg} \text{ constant}} \tag{49}$$

$$\left.\begin{matrix}\text{Screen-grid} \\ \text{resistance}\end{matrix}\right\} = R_{sg} = \left.\frac{dE_{sg}}{dI_{sg}}\right|_{E_g \text{ and } E_p \text{ constant}} \tag{50}$$

The plate resistance of these types of tubes is very high when the plate voltage is great enough to make the plate current substantially independent of plate voltage. This is particularly true with pentode tubes, and values of plate resistance of megohms are typical for small pentodes under operating conditions. On the other hand, the screen resistances are moderate, being usually about five times the plate resistance of the same tube operated as a triode by connecting the screen and plate electrodes together.

A mu factor of a tube is defined as the relative effect that some two electrodes have on some current in the tube. In the case of pentode and similar tubes one could accordingly define mu factors for the relative effectiveness of control grid and screen on the plate current, and on the total space current, for the relative effectiveness of control grid and plate on each of these currents, etc. In general, the mu factor of electrode 1 relative to electrode 2 with respect to current 3 in a tube is given by the relation

$$\text{mu factor} = \left.\frac{dE_2}{dE_1}\right|_{I_3 \text{ constant}} \tag{51}$$

With pentode, beam, and screen-grid tubes the mu factors of importance are two. *First* is the relative effectiveness of plate and control-grid voltage on the plate current, which corresponds to the amplification factor of a triode and is usually called simply the amplification factor. Under conditions where the plate resistance of the tube is

high, this amplification factor is likewise high, and for the same reason. Values of amplification factor in excess of 1,000 are typical for small pentodes. The *second* mu factor of importance in pentode, screen-grid, and beam tubes is μ_{sg} of Eq. (51), the relative effectiveness of control grid and screen on the *total* space current. This is analogous to the amplification factor of a triode, and can be called the "cutoff" amplification factor, since it determines the relative screen and control-grid potentials giving plate-current cutoff. Values for μ_{sg} of the order of 5 to 25 are typical.

The transconductances are defined in the general case as the change of current to electrode 2 as a result of a voltage increment applied to electrode 1; *i.e.*

$$\text{Transconductance} = \frac{\partial I_2}{\partial E_1} \tag{52}$$

In the case of pentodes and similar tubes one is primarily interested in the transconductance from control grid to plate, and occasionally in the transconductance from control grid voltage to screen current. The former corresponds to the transconductance of a triode and has about the same numerical value as in the corresponding triode. The transconductance to the screen current is usually only about one-fourth that to the plate, because the screen current is only about one-fourth of the plate current. The transconductance from the control-grid voltage to the plate current is the most important single coefficient of a pentode tube when operated under normal conditions.

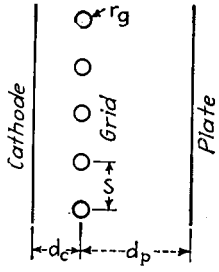


FIG. 22.—Parallel-plane triode.

7. Amplification-factor Formulas.—Formulas for the amplification factor of a triode tube may be obtained by calculating the relative effectiveness of the plate and control grid in controlling the gradient of potential at the cathode.

For a *plane electrode triode* whose electrodes comprise a plane cathode, a plane plate, and a grid of parallel, evenly spaced, equal diameter wires lying in a plane as shown in Fig. 22, the amplification factor is[1]

$$\mu = \frac{\left(\frac{2\pi d_p}{s}\right) - \log_\epsilon \cosh\left(\frac{2\pi r_g}{s}\right)}{\log_\epsilon \coth\left(\frac{2\pi r_g}{s}\right)} \tag{53}$$

where d_p = grid-plate spacing.
s = grid-wire spacing.
r_g = grid-wire radius.

For tubes in which the *screening fraction*, which is the percentage projected area that the grid occupies in its plane, is less than one-sixth, and the ratio of the grid-wire spacing to cathode-grid distance is less than one, the foregoing formula is accurate[2] to within two per cent and becomes more accurate as the two above-mentioned ratios decrease, being within one-half per cent in ordinary tubes. The amplification factor increases as the grid-plate distance is made larger, as the grid-wire radius is increased, or as the grid-wire spacing is decreased. The amplification factor is independent of the cathode-grid distance.

[1] F. B. Vodges and F. R. Elder, Formulas for the Amplification Constant for Three Element Tubes, *Phys. Rev.*, Vol. 24, p. 683, December, 1924.

[2] Formulas for cases where these limitations do not hold are given by Bernard Salzberg, Formulas for the Amplification Factor for Triodes, *Proc. I.R.E.*, Vol. 30, p. 134, March, 1942.

For a *cylindrical triode* having a cathode and plate that are concentric cylinders, and a squirrel-cage grid of evenly spaced wires parallel to the axis and arranged in a circle as shown in Fig. 23, the amplification factor is[1]

$$\mu = \frac{N \log_{10}\left(\frac{s_p}{s_g}\right) - \log_{10} \cosh\left(\frac{N r_g}{s_g}\right)}{\log_{10} \coth\left(\frac{N r_g}{s_g}\right)} \tag{54}$$

where s_p = plate radius.
s_g = grid-wire-circle radius.
r_g = grid-wire radius.
N = number of grid wires.

The foregoing expression becomes inaccurate if the screening fraction, in this case $N r_g/\pi s_g$, is greater than one-sixth, and if the distance between grid wires, approximately $2\pi s_g/N$, is greater than the cathode-grid distance. For the cylindrical triode, the amplification factor increases as the number of grid wires is increased, as the gridwire radius is increased, and as the radius of the grid wire circle is decreased.

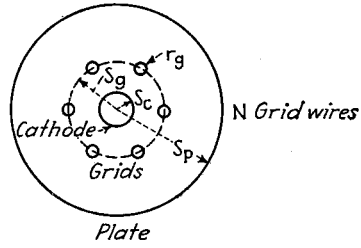


FIG. 23.—Cylindrical triode, end view.

In most tubes the idealized structures shown in Figs. 22 and 23 are only partially realized. Equation (53*a*) may, however, be generalized to include all plane electrode triodes having a grid lying in a plane, regardless of the form of the grid. This is done by recognizing that the length L_g of grid wire per unit area of grid plane is given by the reciprocal of the grid-wire spacing s

$$L_g = \frac{1}{s} \tag{55}$$

and that the screening fraction S is

$$S = \frac{2r_g}{s} \tag{56}$$

Making these substitutions in Eq. (53*a*), the *general amplification factor formula* for a *plane electrode triode* is[2]

$$\mu = \frac{(2\pi d_p L_g) - \log_\epsilon \cosh(\pi S)}{\log_\epsilon \coth(\pi S)} \tag{57}$$

where d_p = grid-plate spacing.
L_g = length of grid wire per unit area of grid plane.
S = screening fraction = ratio of projected area of grid wires to total area of grid plane.

The corresponding formula for the *amplification factor* of a *cylindrical electrode triode* is

$$\mu = \frac{L_g \log_{10}\left(\frac{s_p}{s_g}\right) - \log_{10} \cosh(\pi S)}{\log_{10} \coth(\pi S)} \tag{58}$$

[1] Vodges and Elder, *loc. cit.*
[2] Y. Kusonose, Calculation of the Characteristics and the Design of Triodes, *Proc. I.R.E.*, Vol. 17, No. 10, pp. 1706–1749, October, 1929.

where L_g = length of grid wire per unit area of grid cylinder.
S = screening fraction of grid structure.
s_p = radius of plate cylinder.
s_g = radius of grid cylinder.

The values of L_g for the most commonly encountered grid structures will now be given. In the *cylindrical triode*, if the grid is a cylindrical *screen* of *square mesh*, then

$$L_g = \frac{4\pi s_g}{d} \tag{59}$$

where s_g = radius of grid cylinder.
d = the distance between wires, as shown in Fig. 24.

In the preceding it is presumed that the wires in the grid mesh are small in diameter compared to their spacing. If the diameter of the grid wires is not small compared to the spacing, then

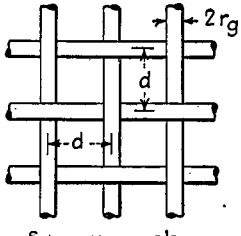


FIG. 24.—Square mesh grid.

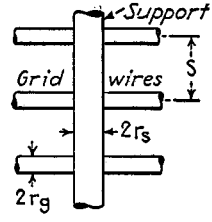


FIG. 25.—Supported parallel grid rings.

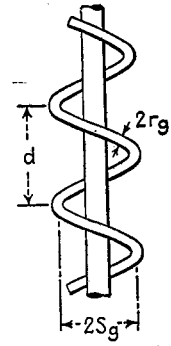


FIG. 26.—Helical grid.

$$L_g = \frac{4\pi s_g}{d}\left(1 - \frac{r_g}{d}\right) \tag{60}$$

For a cylindrical triode having a grid of *parallel rings* with supports parallel to the axis, as shown in Fig. 25,

$$L_g = \frac{2\pi s_g}{s} + n_s - \frac{2n_s r_s}{s} \tag{61}$$

where n_s = number of supports.
s = spacing of the grid rings.
r_s = radius of the support wires.
s_g = radius of the grid rings.

For a cylindrical triode with a *helical grid*, as in Fig. 26

$$L_g = \sqrt{1 + \frac{2\pi s_g^2}{d}} \tag{62}$$

where s_g = the radius of the helix.
d = distance between turns.

The *screening fractions* for the cases listed above are readily evaluated. For the cases in which the grid is a *mesh* of *wires*, a set of *parallel wires*, or a *helix*

$$S = \frac{L_g r_g}{\pi s_g} \tag{63}$$

where r_g = radius of the grid wires.
s_g = radius of the grid wire circle.
For *parallel grid rings with supports*

$$S = \frac{2r_g(L_g - n_s) + 2n_s r_s}{2\pi s_g} \tag{64}$$

Many actual tubes have electrodes that are neither strictly plane nor cylindrical in shape. For such tubes it has been found empirically that a combination of the values of the plane and cylindrical formulas will give the correct amplification factor.[1] The combination formula that gives the actual amplification factor is

$$\mu = \mu_p - K(\mu_p - \mu_c) \tag{65}$$

where μ_p = amplification factor assuming plane electrodes, Eq. (53).
μ_c = amplification factor assuming cylindrical electrodes, Eq. (54).

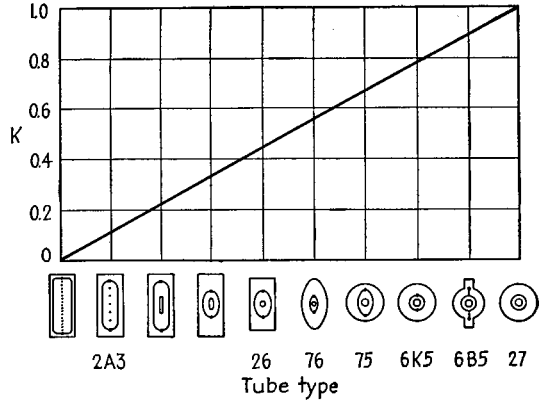


FIG. 27.—Factor K for use in Eq. (65) and associated tube types.

The factor K is a constant that depends upon the tube structure. It is an empirical constant and takes into account the effect of the grid supports. Values of K are given in Table 4 and Fig. 27.

TABLE 4.—FACTOR K FOR USE IN EQ. (65)

Tube Type	K
2A3	0.11
...	0.22
...	0.33
26	0.44
76	0.55
75	0.66
6K5	0.77
6B5	0.88
27	1.00

8. Transconductance Formulas.—The transconductance formulas of triodes and pentodes are obtained from the space-charge relations given in Par. 4 and the equivalent voltage concept discussed in Par. 6.

Triode Transconductance.—The transconductance of a triode with a plane electrode structure is

[1] E. R. Jervis, Amplification Factor Chart, *Electronics*, Vol. 12, p. 45, June, 1939.

$$g_m = \frac{3.5\sqrt{E_g + \frac{E_p}{\mu}}}{\left[d_c^{4/3} + \frac{(d_c + d_p)^{4/3}}{\mu}\right]^{3/2}} \text{ microamperes per volt per unit area} \qquad (66)$$

where E_g = grid voltage.
E_p = plate voltage.
μ = amplification factor.
d_c = cathode-grid distance.
d_p = grid-plate distance.

and the electrode structure is that of Fig. 22. The transconductance is seen to depend upon the equivalent voltage $(E_g + E_p/\mu)$ and the electrode dimensions. Unlike the amplification factor, the transconductance does depend upon the cathode-grid distance. The transconductance of a plane electrode triode may also be expressed in terms of the plate current:

$$g_m = \frac{2.64 \times 10^{-4} i^{1/3}}{\left[d_c^{4/3} + \frac{(d_c + d_p)^{4/3}}{\mu}\right]^{3/2}} \text{ amperes per volt per unit area} \qquad (67)$$

where i is the current density in amperes per unit area and the other symbols have the same significance as in Eq. (66).[1]

The transconductance of a *cylindrical electrode triode* with the configuration of Fig. 23 is

$$g_m = \frac{2.2\left(E_g + \frac{E_p}{\mu}\right)^{1/2}}{\left[(s_g\beta_{cg}^2)^{2/3} + \frac{1}{\mu}(s_p\beta_{cp}^2)^{2/3}\right]^{3/2}} \text{ microamperes per volt per unit length of structure} \qquad (68)$$

where E_g = grid voltage.
E_p = plate voltage.
μ = amplification factor.
s_g = radius of grid-wire circle.
s_p = plate radius.
β_{cg}^2 = value of β^2 from Table 3 for argument s_g/s_c.
β_{cp}^2 = value of β^2 from Table 3 for argument s_p/s_c.
s_c = cathode radius.

9. Ultra-high-frequency Effects and Tubes.[2]—In ordinary applications of vacuum tubes the transit time of the electrons in the tube is short compared with the period of the applied voltages. As frequency is increased this becomes less and less true until finally the electron will require an appreciable fraction of a cycle to pass from one electrode to another. When this happens the behavior of the tube changes markedly.

Ultra-high-frequency Behavior of Diodes.—Even the simple diode tube experiences a difference in its behavior at ultra-high frequencies. The dynamic plate resistance of the diode drops, and at certain frequencies even becomes negative, and there are also modifications produced in the plate-cathode capacity.

When transit-time effects are of importance it is convenient for the purpose of analysis to consider that the impedance between the plate and cathode of the tube,

[1] J. H. Fremlin, Calculation of Triode Constants, *Elec. Comm.*, Vol. 18, p. 39, July, 1939.

[2] A comprehensive discussion and summary of transit-time effects is given in the book by F. B. Llewellyn, "Electron Inertia Effects," Cambridge (London), 1941. Much of this material is also given by F. B. Llewellyn, Operation of Ultra-high-frequency Tubes, *Bell System Tech. Jour.*, Vol. 15, p. 575 October, 1936.

including the active part of the interelectrode capacity[1] as well as the resistance effects due to electron current flow, be represented by a resistance R in series with a capacitive reactance X (*i.e.*, $Z = R - jX$). In terms of this representation one has for complete space charge[2]

$$R = R_p \frac{12}{\theta^4} [2(1 - \cos \theta) - \theta \sin \theta] \tag{69}$$

$$X = -\left\{\frac{1}{\omega C} + R_p \frac{12}{\theta^4} [\theta(1 + \cos \theta) - 2 \sin \theta]\right\} \tag{70}$$

where R_p = dynamic plate-cathode resistance of diode at low frequencies.
θ = transit-time angle, radians ($\theta = \omega$ times transit time, sec.).
$\omega = 2\pi$ times frequency.
C = active part of the plate-cathode capacity of the tube measured with the cathode cold.

At frequencies where the transit-time effects are small, Eqs. (69) and (70), after suitable transformations, become

$$R = R_p \left[1 - \frac{\theta^2}{15}\right] \tag{71}$$

$$X = -\frac{3\theta}{10} R_p \left[1 - \frac{\theta^2}{25.2}\right] \tag{72}$$

It will be noted that θ is the transit time in radians, on the basis that the time represented by one cycle is 2π radians. A curve showing the variation of R and X with

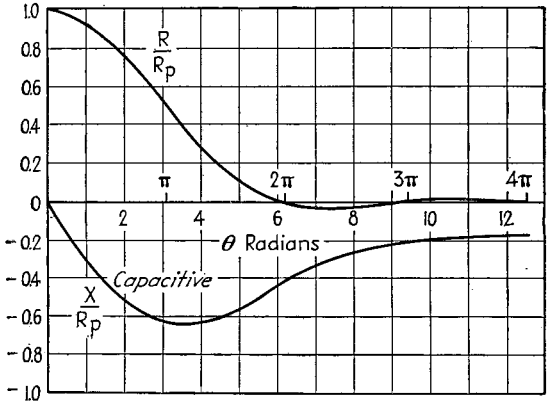


FIG. 28.—Resistance and reactance components of diode plate-cathode impedance as a function of transit angle θ (series representation).

transit angle is given in Fig. 28. It is seen that as the frequency is increased the resistance R first drops increasingly from its low-frequency value R_p until it becomes zero when the transit time is a full cycle, and with still greater transit times then

[1] The active part C of the interelectrode capacity is the capacity between parts of the tube that are effective in causing a current flow. The capacity between leads, or inactive parts of the plate or cathode structure, is not included, and represents an additional capacity that shunts the "active" impedance $R + jX$.

[2] See Llewellyn, *op. cit.*, p. 48; W. E. Benham, Electron Transit Time, *Wireless Eng.*, Vol. 16, p. 598, December, 1939. Graphical means of analyzing transit-time effects in diodes are described by R. W. Sloane and E. G. James, Transit-time Effects in Diodes in Pictorial Form, *Jour. I.E.E.*, Vol. 79, p. 291, 1936; also, Wireless Section, *I.E.E.*, Vol. 11, p. 247, September, 1936; and Rudolf Kompfner, Transit-time Phenomena in Electronic Tubes, *Wireless Eng.*, Vol. 19, p 3, January, 1942.

oscillates about the zero value, being slightly negative at times. The reactive component of the impedance is always capacitive, but varies greatly. For very large transit times the tube impedance approaches that offered by the "cold" interelectrode capacity C, and the resistance component becomes very small.

In the preceding, the effect of the active part of the interelectrode capacity has been represented by a reactance in series with the equivalent plate resistance R. At frequencies where the transit-time effects are small, it is, however, more convenient to think of the tube plate-cathode impedance as representing a parallel combination of a capacity in parallel with an equivalent plate resistance. This equivalent resistance at low frequencies R is given by Eq. (69) or (71), while the equivalent low-frequency shunting capacity is $0.6C$. This low-frequency representation holds up to transit times of about $\theta = 1$. It will be noted that the presence of the electron stream in the interelectrode space causes the "hot" capacity of the tube to be only a fraction of the "cold" capacity C.

Input Resistance of Control Grids at Ultra-high Frequencies.[1]—At very high frequencies the transit time contributes a conductance component to the grid input admittance. A current will be induced in an electrode by the approach or passage of an electron even if the electron does not hit the electrode.[2] In the case of a triode or a tube with a screen grid, the currents induced in the control grid by approaching and receding electrons are different because the differences in velocities and the finite transit time impart an in-phase component to the induced current. The result of this action is that the conductance increases with frequency according to the formula[3]

$$G_g = Kg_m f^2 T^2 \tag{73}$$

where G_g = conductive component of grid input admittance.
g_m = transconductance of tube.
f = frequency.
T = transit time to a reference point near the grid plane.
K = a constant for any particular tube, a function of the cathode-grid and grid-plate transit times.

Examination of the preceding formula shows that if all the dimensions of a tube are reduced by a factor M, then the input conductance will be reduced by a factor M^2. If all the voltages are increased by a factor N, the input conductance will be reduced by a factor $N^{1/2}$.

An idea of the magnitude of the transit-time effect upon grid input conductance can be gained from the fact that the input resistance of a 57 tube is of the order of megohms at 5 megacycles, while at 30 megacycles it has dropped to 20,000 ohms and at 100 megacycles it has dropped to about 1,500 ohms. The acorn pentode 954, which makes use of a very small cathode-grid spacing, has an input resistance that is of the order of 10 times as great as that of ordinary triodes, being about 20,000 ohms at 100 megacycles. The theoretically predicted variation of conductance with the square of the frequency is borne out by experimental observations.

Ultra-high-frequency Behavior of Triodes.—The behavior of triodes at ultra-high frequencies has been worked out to include both the effects of transit time and space charge.[4] Analysis shows that it is possible to set up a circuit that is the equivalent of

[1] The discussion here assumes that the electrode on the side of the control grid away from the cathode (screen or plate, as the case may be) is at ground potential to very high frequencies.

[2] B. J. Thompson, Review of Ultra-high Frequency Vacuum Tube Problems, *R.C.A. Rev.*, Vol. 3, p. 146, October, 1938.

[3] W. R. Ferris, Input Resistance of Vacuum Tubes as Ultra-high Frequency Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 82, January, 1936; D. O. North, An Analysis of the Effects of Space Charge on Grid Impedance, *Proc. I.R.E.*, Vol. 24, p. 108, January, 1936.

[4] F. B. Llewellyn, Operation of Ultra-high Frequency Tubes, *Bell System Tech. Jour.*, Vol. 14, p. 112, October, 1935.

the vacuum tube and its transit-time effects, in terms of parameters that are independent of frequency and that are not complex. This is done by using a delta of impedances with the individual arms represented by the proper series or parallel combination of elements as shown in Fig. 29.[1] In this circuit the need for complex tube constants is avoided by the proper arrangements of circuit elements. The actual values applicable to the various parts of the circuit of Fig. 29 depend on the transit times, the dynamic plate resistance at low frequencies, and the capacities between the active parts of the various electrodes. The relations involved are too complicated and specialized to be given here, but are to be found in the references cited. The equivalent circuit of Fig. 29 is valid up to frequencies of 100 mc or more with ordinary tubes.

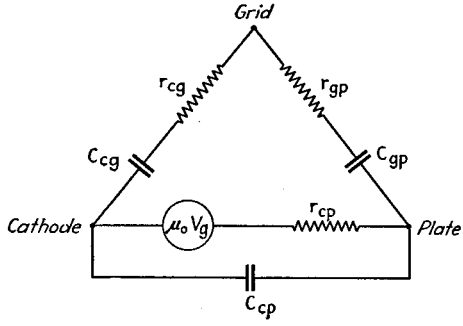


FIG. 29.—Equivalent delta of a triode at ultra-high frequencies.

The circuit of Fig. 29 is valid only for the active part of the tube, *i.e.*, the portion of the tube elements that intercept electrons. The effect of the lead impedances is not included, an important limitation to keep in mind, since the effects of the lead impedances may be as great as the effects of transit time of the electrons. Neither are the capacities between inactive parts of the electrodes and leads included. These capacities act in shunt with the equivalent circuit of Fig. 29, between the appropriate electrodes.

As far as the plate circuit of a triode is concerned, the effect of transit time can be taken into account by assuming that the ordinary equivalent plate circuit consisting of a generator voltage μe_s acting in series with the plate resistance R_p still holds (see Par. 3, Sec. 5), but that the value of amplification factor and plate resistance (and hence transconductance) is now modified in magnitude and a phase angle is associated with them. In particular, the transconductance tends to decrease in magnitude and lag by an increasingly large amount as the transit time increases,[2] while the amplification factor likewise tends to decrease and have an increasingly large phase angle.

Ultra-high-frequency Negative Grid Tubes.—Triode, beam, and other similar tubes intended for negative grid operation at very high frequencies must be especially designed for such service. The desired characteristics are those outlined in Par. 11, Sec. 6, for oscillator tubes, namely: (1) close electrode spacing to minimize transit time and give a high transconductance in proportion to electrode capacities; (2) short leads of relatively large diameter to minimize lead inductance and power losses in the leads (achieved in many cases by the use of multiple leads from the same electrode); (3) electrode and lead arrangements that facilitate operation with resonant transmission lines, the ideal being when the leads and electrodes of the tube represent extensions of the transmission line tank circuits.[3]

The tube should also be designed so that there can be no stray electrons that can shoot out of the open ends of the tube structure, and take a circuitous route in

[1] F. B. Llewellyn, Equivalent Networks of Negative-grid Vacuum Tubes at Ultra-high Frequencies, *Bell System Tech. Jour.*, Vol. 15, p. 575, October, 1936.

[2] Thus see F. B. Llewellyn, Phase Angle of Vacuum-tube Transconductance at Very High Frequencies, *Proc. I.R.E.*, Vol. 22, p. 947, August, 1934; Thompson, *loc. cit.*

[3] A tube in which this idea is carried to the limit is described by I. E. Mouromtseff and H. V. Noble, A New Type of Ultra-short-wave Oscillator, *Proc. I.R.E.*, Vol. 20, p. 1328, August, 1932.

traveling from the cathode to the plate.[1] Such electrons are likely to bombard the glass envelope and thereby release secondary electrons. These indirect paths, either with or without involving secondary electrons, result in some electrons arriving at the anode with excessively long transit times, with undesirable consequences. In order to avoid difficulties of this sort, it is desirable to close the ends of the tube structure so that it will not be possible for even a few electrons to stray along indirect routes.

When high voltages are involved, as is the case with transmitter tubes, the structural arrangement must be such as to provide an adequate amount of insulation. This is particularly important as the frequency increases, because the dielectric loss in the glass is proportional to the square of the frequency, and can reach sufficiently large values at extremely large frequencies to start a cumulative heating that results in puncture of the glass envelope.

It is possible to obtain satisfactory ultra-high-frequency operation of receiving tubes up to higher frequencies than can be achieved with power tubes. This is because the close spacings that favor ultra-high-frequency operation inevitably carry with them the necessity of making the tube dimension small, and this limits the power-dissipating ability of the tube. The amount of power that can be developed in the output of an ultra-high-frequency tube is accordingly less as the frequency limit is pushed to higher values. Ingenious arrangements have been devised for increasing the ability to dissipate heat, as, for example, the use of cooling fins attached to the plate, and in some cases also to the grid, and the use of water or forced-air cooling, even in tubes of moderate power rating. Such arrangements at best, however, merely improve an otherwise unfavorable situation, but in no case do they eliminate it.

A variety of ultra-high-frequency tubes are available. These include receiving tubes of the triode, pentode, and beam types, and also triode, beam, and pentode power tubes. Most of these are air-cooled, but a few of the larger types designed for such applications as television transmitters are arranged for water or forced-air cooling. In a few instances tubes are arranged in duplex (two tubes in a single envelope) for push-pull operation.[2]

10. Special Considerations Involved in Power Tubes.[3]—Tubes intended for use as power amplifiers must be capable of dissipating plate and grid losses in proportion to the desired power. Likewise, the electron emission from the cathode and the amount of voltage that can be applied to the anode with safety must be in proportion to the power output that the tube is to develop.

[1] W. G. Wagener, The Developmental Problems and Operating Characteristics of Two New Ultra-high-frequency Triodes, *Proc. I.R.E.*, Vol. 26, p. 401, April, 1938.

[2] Tubes especially designed for ultra-high-frequency use represent the accumulation of a great deal of specialized technical experience, much of which is of interest only to the designer of tubes. Because of this, and the fact that the subject is so new that the techniques are being steadily improved and modified, detailed descriptions of the various ultra-high-frequency tubes now available on the open market are not included here. The reader wishing to obtain a background in the subject is referred to the following articles: B. J. Thompson and G. M. Rose, Jr., Vacuum Tubes of Small Dimensions for Use at Extremely High Frequencies, *Proc. I.R.E.*, Vol. 21, p. 1707, December, 1933; M. J. Kelly and A. L. Samuel, Vacuum Tubes as High Frequency Oscillators, *Elec. Eng.*, Vol. 53, p. 1504, November, 1934; A. L. Samuel and N. E. Sowers, A Power Amplifier for Ultra-high Frequencies, *Proc. I.R.E.*, Vol. 24, p. 1464, November, 1936; A. L. Samuel, A Negative Grid Triode Oscillator and Amplifier for Ultra-high Frequencies, *Proc. I.R.E.*, Vol. 25, p. 1243, October, 1937; A. K. Wing, A Push Pull Ultra-high-frequency Beam Tetrode, *R.C.A. Rev.*, Vol. 4, p. 62, July, 1939; A Transmitter for Frequency-modulated Broadcast Service Using a New Ultra-high-frequency Tetrode, *R.C.A. Rev.*, Vol. 5, p. 327, January, 1941; A. K. Wing, Jr., and J. E. Young, A New Ultra-high-frequency Tetrode and Its Use in a 1-kilowatt Television Sound Transmitter, *Proc. I.R.E.*, Vol. 29, p. 5, January, 1941; K. C. DeWalt, Three New Ultra-high-frequency Triodes, *Proc. I.R.E.*, Vol. 29, p. 475, September, 1941; Cecil B. Haller, The Design and Development of Three New Ultra-high-frequency Transmitting Tubes, *Proc. I.R.E.*, Vol. 30, p. 20, January, 1942.

[3] An outstandingly fine discussion on the construction of large tubes is given by J. Bell, J. W. Davies, and B. S. Gossling, High-power Valves: Construction, Testing, and Operation, *Jour. I.E.E.*, Vol. 83, p. 176, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 177, September, 1938.

Small power tubes are often merely large receiving tubes constructed according to the standard receiving-tube techniques, involving oxide-coated cathodes, "getters" to obtain and maintain vacuum, etc. Such tubes are not capable of standing much abuse, however, and the maximum power that can be developed with this type of construction is limited.

Larger tubes, up to dissipations of several hundred watts, are normally still inclosed in a glass envelope but are constructed along lines entirely different from those of receiving tubes. The plates are commonly of molybdenum or tantalum, carbon sometimes being used. Grids are usually molybdenum, with tungsten and tantalum as alternates. Cathodes are sometimes oxide-coated, but more frequently are thoriated tungsten carburized for long life. The exhaust procedure is so thorough that a high degree of vacuum can be maintained without a getter. The various electrodes inside the tube are heated to high temperatures during exhaust, and the glass envelope is simultaneously baked at a temperature just below the softening point, and this process is continued until all the occluded gases have been removed. It is possible in this way to obtain a degree of vacuum such that a satisfactory life can be obtained with thoriated tungsten filaments even at high anode voltages, particularly if tantalum is used for the plate.

In tubes where the anode dissipation is of the order of kilowatts, the anodes of the tube are in the form of copper cylinders, which are part of the envelope and are cooled by means of circulating water or by being soldered to a radiating structure with many fins, which are cooled by means of an air blast. Analysis and experience indicate that water and forced-air cooling are about equally effective when used with vacuum tubes having external anodes.[1] Decision as to which type of cooling is preferable depends upon the individual factors involved in the particular case. Thus air-cooled tubes behave better in operation than water-cooled tubes with respect to "flashback." With forced-air cooling the anodes tend to run hotter, but the glass ends run cooler than with water cooling, which has some advantages and some disadvantages. With forced-air cooling the anode capacity to ground is increased by the large cooling structure, something to be avoided at very high frequencies.

The most effective way to employ water cooling is to arrange the water jacket so that the water passes over the anode in a very thin stream of high velocity. In this way any steam bubbles that may be formed on the anode tend to be scraped away by the velocity of the cooling water. The tendency for such bubbles to form limits the amount of heat that can be dissipated by a water-cooled anode, since, if the tendency of the bubble to stick to the copper is not overcome by the flow of cooling water, then that particular spot on the anode ceases to be cooled by direct contact with the water, and may readily become hot enough to cause failure of the tube.

In tubes with external anodes that are water-cooled, there is a tendency for the grid to produce a focusing effect on the electrons, which causes hot spots to be produced on the anode at the points where the electrons concentrate.[2] This is much more troublesome with water cooling than with forced-air cooling. With water cooling these hot spots tend to cause steam bubbles to form on the surface of the anode.

The cathodes of tubes employing water or forced-air cooling are always tungsten. This is because such high-power tubes are operated at very high anode voltages, and it is not practical to exhaust copper-anode tubes so thoroughly as glass-envelope tubes, where the anodes can be brought to incandescent heat during the exhaust process.

[1] Further discussion of both air and water cooling is given by I. E. Mouromtseff, Water and Forced-air Cooling of Vacuum Tubes, *Proc. I.R.E.*, Vol. 30, p. 190, April, 1942. See also E. M. Ostlund, Air Cooling Applied to External-anode Tubes, *Electronics*, Vol. 13, p. 36, June, 1940.

[2] I. E. Mouromtseff, The Influence of Grid Focusing Effect on Plate Dissipation Limit of a Vacuum Tube, *Communications*, Vol. 18, p. 9, December, 1938.

Thoriated tungsten filaments accordingly do not behave satisfactorily, and pure tungsten must be employed.[1]

The thoriated-tungsten and oxide-coated cathodes should be operated at the rated voltage to obtain maximum life. With tungsten filaments, however, decreasing the filament voltage will increase the useful life at the expense of reduced electron emission (and hence lowered peak anode current). Varying the voltage applied to a tungsten filament by 5 per cent from normal will double or halve the life as the case may be, while a 10 per cent variation affects the life by a factor of 4. A study of the optimum diameter of a tungsten filament, if the costs of power and tube replacement are taken into account, indicates that the optimum diameter is one that would correspond to a life of the order of 10,000 hours or more.[2]

In water-cooled or forced-air cooled tubes operating at high anode voltages (such as 15 to 20 kv), trouble is frequently encountered from arc-backs or "flash arcs" in the tube.[3] These are ordinary high-current arcs that suddenly form between cathode and plate for no apparent reason, and short-circuit the plate supply. After such an arc has been broken by opening of the circuit breakers, and the anode voltage has been reapplied, the tube may continue to operate for many hours or days without any further trouble, while in other cases another flash arc may occur almost immediately. A tube that has a tendency to arc-back will generally improve in this respect if it is put through a conditioning process[4] that involves applying an alternating potential of about 25,000 volts between plate and filament of the tube with the filament unlighted, with suitable relays and current-limiting devices in the circuit to prevent damage from arcs. This results in the production of many flash arcs in the tube, which gradually clean up the gas that is causing the trouble, and bring the tube to a satisfactory operating condition.

Power tubes are commonly operated so that the grid goes positive during a part of the cycle. This results in grid current, and causes power dissipation at the grid of the tube, which is sometimes the limiting factor in tube operation.[5] As a consequence, the grids of power tubes often operate at relatively high temperatures, and such materials as molybdenum, tungsten, or tantalum are accordingly generally used. The fraction of the primary electrons intercepted by the grid depends upon the grid potential relative to the anode potential, and upon the grid structure. In ordinary triodes with equal grid and plate voltages the effective grid area, insofar as intercepting the flow of primary electrons is concerned, is between 120 and 180 per cent of the actual grid area.[6] The grid heating that takes place is determined by the number of primary electrons intercepted by the grid, and by the grid voltage. The actual d-c grid current as measured by a meter may differ from the number of primary electrons received by the grid as a result of secondary emission causing the grid to lose secondary electrons at the same time that it receives primary electrons. The amount of current thus lost through secondary emission will be affected by the electrode potentials, by the grid temperature, and by the character of the grid surface.

[1] It is possible by keeping a water-cooled tube continuously on a vacuum pump, to maintain a degree of vacuum about one decimal point greater than feasible when the tube is sealed off. Under these conditions thoriated-tungsten filaments are entirely practical even at very high anode voltages. Such continuously evacuated tubes have been developed by various individuals and organizations for very high-power service, but as yet have not had any extensive continuing commercial use.

[2] Thus see J. J. Vormer, Filament Design for High Power Transmitting Valves, *Proc. I.R.E.*, Vol. 26, p. 1399, November, 1938.

[3] See Bell, Davies, and Gossling, *loc. cit.;* B. S. Gossling, The Flash-arc in High-power Valves, *Jour. I.E.E.*, Vol. 71, p. 460, 1932; also, Wireless Section, *I.E.E.*, Vol. 7, p. 192, September, 1932.

[4] Further details are given in Vacuum Tube Reconditioning, *Electronics*, Vol. 14, p. 84, January, 1941.

[5] I. E. Mouromtseff and H. N. Kozanowski, Grid Temperature as a Limiting Factor in Vacuum Tube Operation, *Proc. I.R.E.*, Vol. 24, p. 447, March, 1936.

[6] Karl Spangenberg, Current Division in Plane-electrode Triodes, *Proc. I.R.E.*, Vol. 28, p. 226, May, 1940.

In the case of thoriated tungsten and oxide-coated cathodes the secondary emissions may, under some conditions, become quite large as a result of cathode material that has been deposited upon the grid.

11. Receiving Tubes.—Receiving tubes can be divided into two main classes: the glass-envelope tubes, and the so-called "metal" tubes. The glass-envelope tubes, as the name implies, have a glass envelope, and the electrodes are mounted on leads running through either a stem or a glass cup, to which the bulb is sealed after the electrodes are mounted and adjusted. The metal-envelope tube employs a metal shell in place of the glass bulb, with all the leads, except perhaps a control-grid lead, brought through a glass button or glass beads in the base of the tube.

The cathodes of receiving tubes, either filaments or heaters, are practically always of the oxide-coated type. The grid and plate electrodes and side rods are usually constructed from nickel, and a "getter" is used to obtain the desired degree of vacuum. The exhaust, and also many of the other assembly operations, are performed on automatic machines.

The actual technical details of receiving tubes represent primarily a production problem, and accordingly are beyond the scope of this work. The literature available on the subject is relatively limited, and always lags behind the current practice in industry.[1]

12. "Getters."[2]—A high vacuum is obtained in receiving tubes by means of a "getter," which is volatilized inside the tube for the purpose of removing residual gas by either chemical or mechanical action. Magnesium is widely used as a getter, but other materials, such as barium, barium berylliate,[3] zirconium, phosphorus, etc., can also be employed, as well as various mixtures. A getter should be initially inert, but should be of such character that it can be made highly active by some process such as a chemical change, or vaporization of material inert at room temperature.

Some getters also act as "keepers" in that they not only remove what gas is present in the tube at the time of flashing but also combine with any gas that may subsequently be liberated within the tube.

Getters cannot be used in tubes where the anode dissipation is large. This is because the large amount of heat liberated in such tubes would vaporize the getter and destroy the vacuum.

Certain metals have the property of absorbing gases when raised to a high temperature. For example, tantalum will absorb gases when very hot, and this is one of the things that make tantalum a desirable metal to use in power tubes for plates and grids.

Zirconium is also a useful metal in this regard.[4] Zirconium at 1400° will absorb copious quantities of such gases as oxygen, nitrogen, carbon monoxide, and carbon dioxide. The optimum temperature for taking up hydrogen is, however, approximately 350°C, and at higher temperatures hydrogen begins to be given off by the metal. As a result of these properties it is possible to maintain an extremely high

[1] Useful articles on receiving tubes are M. Benjamin, C. W. Cosgrove, and G. W. Warren, Modern Receiving Valves: Design and Manufacture, *I.E.E. Wireless Proc.*, Vol. 12, p. 65, June, 1937 E. R. Wagner, Raw Materials in Vacuum Tube Manufacture, *Electronics*, Vol. 7, p. 104, April, 1934; Processes in Vacuum Tube Manufacture, *Electronics*, Vol. 7, p. 213, July, 1934; G. E. Moore, Improved Repeater Tubes, *Bell Lab. Rec.*, Vol. 18, p. 219, March, 1940; Newell R. Smith and Allen H. Schooley, Development and Production of the New Miniature Battery Tubes, *R.C.A. Rev.*, Vol. 4, p. 496, April. 1940; S. R. Mullard, The Development of the Receiving Valve, *Jour. I.E.E.*, Vol. 76, p. 10, 1935, also Wireless Section, *I.E.E.*, Vol. 10, p. 1, March, 1935.

[2] A good discussion of getters is given by E. A. Lederer and D. H. Wamsley, "Batalum," a Barium Getter for Metal Tubes, *R.C.A. Rev.*, Vol. 2, p. 117, July, 1937.

[3] See E. A. Lederer, Recent Advances in Barium Getter Technique, *R.C.A. Rev.*, Vol. 4, p. 310, January, 1940.

[4] Further information is given by J. D. Fast, Zirconium and Its Compounds with a High Melting Point, *Phillips Tech. Jour.*, World's Fair Issue, 1939.

vacuum, provided that there is zirconium in the tube at these two temperatures. This can be accomplished either by employing two zirconium filaments with different filament currents or by using a single piece of zirconium in which different parts of the metal are raised to different temperatures, so that the temperature range extends from 300 to 1400°C. This expedient is frequently used in experimental tubes to clean up gases that may be evolved during use, and can also be employed in special tubes that are particularly difficult to evacuate.

13. Effect of Gas upon Tube Characteristics—Maximum Allowable Resistance in Grid Circuits.—Very small traces of gas in vacuum tubes affect the characteristics adversely in a number of ways as a result of the positive ions produced in the tube by collision between the gas molecules and the electrons flowing to the anode. The positive ions travel in the opposite direction from the electrons, and normally end their existence by falling into the cathode or the negative control grid. Electrons that bombard the cathode tend to destroy the emission of thoriated-tungsten and oxide-coated cathodes. Positive ions collected by the negative grid result in grid current, which limits the resistance that may be inserted in series with a negative grid, and which also introduces noise.

Positive-ion currents to the grid limit the d-c resistance that may safely be placed in series with the control grid and the cathode, because the voltage drop that such a grid current produces across the resistance has a polarity that makes the grid less negative than would otherwise be the case. Thus if the tube begins to liberate gas with resulting positive-ion grid current, the grid becomes less negative, thereby increasing the space current. This increases the number of positive ions produced, and will cause additional grid current, and still greater reduction in the negative grid potential. If the resistance in the grid circuit is high enough, this process can become cumulative, and in some types of tubes can easily result in the destruction of the tube as a result of excessive plate current caused from loss of grid bias. The maximum resistance that it is permissible to place in series with the grid electrode depends upon the tube characteristics and the method of obtaining bias. It is of the order of several megohms in small tubes used for voltage amplification at audio and radio frequencies. With small power tubes, such as the output tubes of radio receivers and public-address systems, the allowable grid resistance is much less, particularly if a fixed bias is employed. The use of self-bias permits an increased resistance to be employed in the grid-cathode circuit, since self-bias provides an automatic protection against excessive increase in plate current.

Positive-ion current to a control grid has a "shot" effect associated with it, since the current flow is composed of individually charged particles, and is not a uniform fluid. The noise that is produced by this current flowing through an impedance Z_g between grid and cathode can be expressed by the formula

$$R_n = 19.3\, I_g |Z_g|^2 \tag{74}$$

where $|Z_g|$ is the absolute value of the external circuit impedance between grid and cathode, I_g is the grid current in amperes, and R_n the resistance that, when connected between grid and cathode, would produce the same thermal agitation voltage as is actually produced by the grid current flowing through the impedance Z_g.

In addition to the noise effect represented by Eq. (74), the presence of gas introduces an additional noise factor through the fact that positive ions not captured by the grid travel toward the cathode and produce irregularities in the space charge around the cathode. Little is known about this effect, however, beyond the fact that it is probably small under ordinary circumstances compared with the effect represented by Eq. (74).

14. Miscellaneous Types of Tubes. *Space-charge-grid Tubes.*—In the space-charge-grid tube there is an auxiliary grid, called the space-charge grid, located between the cathode and the control grid, and operated at a low positive potential. This grid increases the number of electrons drawn out of the space charge near the cathode. Although some of these electrons are immediately attracted to the space charge grid, many pass through its meshes into the region between the space-charge grid and the control grid. Here they are slowed down by the retarding field and form a space charge or virtual cathode, as shown in Fig. 30. The virtual

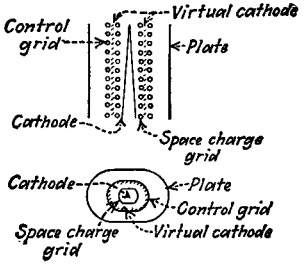


FIG. 30.—Details of a space-charge grid tube.

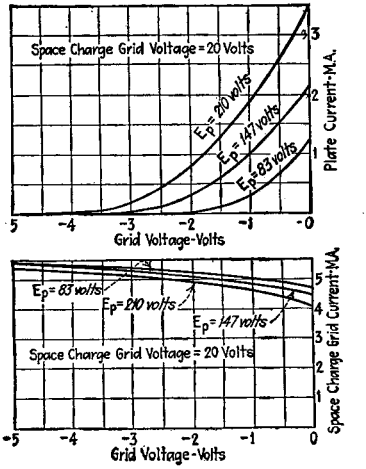


FIG. 31.—Characteristic curves of space-charge grid triode.

cathode serves as an actual cathode as far as the remainder of the tube is concerned. The characteristics that result are similar to those for conventional tubes, as is apparent from Fig. 31.

The space-charge-grid arrangement provides a virtual cathode with a large area located very close to the control grid. This gives a high transconductance in proportion to the plate potential. At the same time the characteristic curves of the tube tend to have excessive curvature when considered over an appreciable range of grid

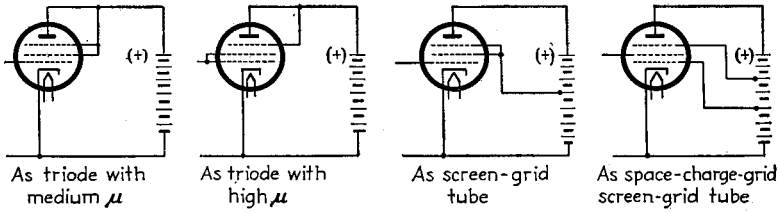


FIG. 32.—Pentode tube arranged in various ways.

voltage. Thus space-charge-grid tubes have small power-handling capacity. Furthermore, the space-charge grid draws a very heavy current, usually more than half of the total space current.

Special Connections for Conventional Tubes.—It is possible to operate conventional tubes to give special characteristics, either by employing special connections or by the proper combination of electrode voltages. Thus a pentode tube can be connected to operate as a triode with medium or high amplification factor, as a screen-grid tube, or as a space-charge-grid tube, as shown in Fig. 32.

Another rearrangement of a conventional tube is to interchange the functions of the grid and plate by making the grid the anode electrode, and by using the plate as

the negative control electrode, as shown in Fig. 33. The operation of such an *inverted tube*[1] rests on the fundamental fact that the space current flowing to the anode, which in this case is the positive grid, depends almost solely upon the electrostatic field in the vicinity of the cathode and is substantially independent of how this field is produced. Since both plate and grid potentials affect the intensity of this electrostatic field, it is possible to use a negative plate as a control electrode to serve the same purpose as the negative grid in the usual tube. The result is an amplification factor that is quite low, the value being approximately $1/\mu$ where μ is the amplification factor of the tube operated in the normal manner. The dynamic anode resistance is also much lower than in the corresponding tube operated in the normal manner. The inverted vacuum tube is useful when it is necessary to control a large current by a very high voltage, without at the same time time consuming any energy from the high potential source.

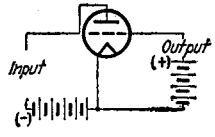


FIG. 33.—Circuit of inverted tube.

The Dynatron and Other Negative-resistance Arrangements.—In a screen-grid tube, if the screen electrode is operated at a higher voltage than the plate, and the secondary emission at the plate is sufficient, there is a range of plate voltages over which the net resultant plate current (primary electrons arriving at the plate minus secondary electrons lost to the screen) will decrease with increasing plate voltage. This effect is shown by the curve for $E_{sg} = 83$ volts in Fig. 19. The decreasing plate current with increasing plate voltage represents a negative resistance, and when a tube is used in this manner to obtain a negative resistance, it is termed a *dynatron*.[2] The negative resistance results from the fact that while the number of primary electrons that the plate receives is independent of the plate voltage, the number of secondary electrons produced at the plate increases with increasing plate voltage. With the screen more positive than the plate, all these secondary electrons flow to the screen; so if the number of secondary electrons that can be produced is considerable, the plate current not only will decrease with increasing plate voltage but will reverse in polarity at a plate potential where, on the average, each primary electron produces one secondary electron. Dynatron action is capable of giving a negative resistance as low as about 20,000 ohms, using ordinary receiving tubes in which the plates have not been specially treated to reduce secondary emission.[3] The magnitude of the negative resistance can be controlled by varying the control-grid potential. The magnitude of the negative resistance is not particularly stable with time, because secondary emission is very sensitive to the surface conditions, and these change with use.

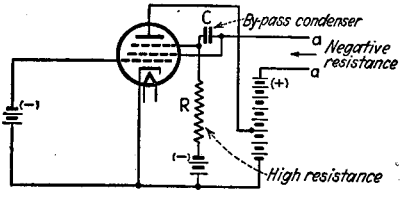


FIG. 34.—Retarding field or transitron method of producing a negative resistance.

An alternative method of obtaining negative resistance from a vacuum tube, which is generally more stable than the negative resistance obtained by dynatron action, is illustrated in Fig. 34.[4] Here a pentode tube is connected as shown and the poten-

[1] F. E. Terman, The Inverted Vacuum Tube—A Voltage Reducing Power Amplifier, *Proc. I.R.E.*, Vol. 16, p. 447, April, 1928.

[2] For further information on the dynatron, and some of its uses, see Albert W. Hull, The Dynatron—A Vacuum Tube Possessing Negative Electric Resistance, *Proc. I.R.E.*, Vol. 6, p. 5, 1918.

[3] It is generally found that some of the very early screen-grid tubes are the best for dynatron purposes. Tubes manufactured more recently commonly have the plate so treated as to reduce secondary emission. This causes the dynatron effect to be greatly reduced, or even eliminated.

[4] E. W. Herold, Negative Resistance and Devices for Obtaining It, *Proc. I.R.E.*, Vol. 23, p. 1201, October, 1935.

tials so adjusted that there is a virtual cathode between the screen and the suppressor. Under these conditions a fraction of the space current drawn from the cathode is returned back toward the cathode, to be ultimately collected by the screen. By arranging the circuit as shown, so that the screen and suppressor are at the same alternating potential as a result of by-pass condenser C, an increase in the screen potential likewise increases the suppressor voltage. This reduces the screen current because more of the electrons entering the virtual cathode pass on to the plate, and fewer return to the screen. The result is a negative-resistance effect between the terminals aa, the magnitude of which can be controlled by the control grid of the tube. This method of obtaining a negative resistance is referred to as a retarding-field negative-transconductance tube, and has also been called the transitron method. As compared with the dynatron, the transitron arrangement has the advantage that the negative resistance obtained is more stable with time, and also a much lower negative resistance can be obtained.

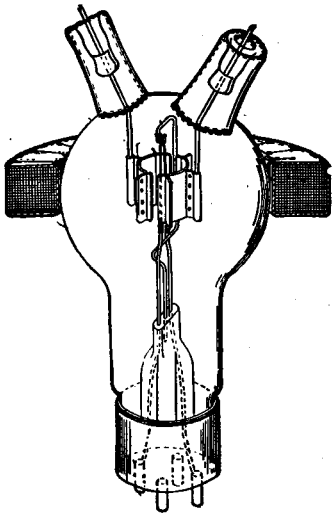


FIG. 35.—The split-anode magnetron.

It is possible to obtain negative resistance from vacuum tubes in still other ways. Some of these involve special connections of conventional vacuum tubes, such as feedback circuits and other arrangements in which the effective impedance across a pair of terminals is negative.[1] It is also possible to employ special tubes of various sorts.[2] Finally, under certain space-charge conditions in a space-charge-grid tube, it is possible for the input conductance to the control grid to be negative.[3] It is also found that associated with these negative-resistance effects obtainable through space-charge action, it is also possible to obtain a negative-capacity effect by purely electronic means involving the behavior of the virtual cathodes.

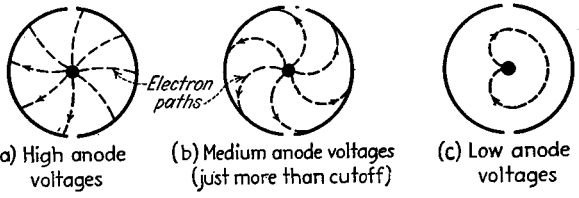


FIG. 36.—Electrode paths in split-anode magnetron under various conditions.

Magnetrons.[4]—A magnetron is a vacuum tube in which the flow of electrons from the cathode to the plate is affected by a magnetic field. The split-anode magnetron illustrated in Fig. 35 consists of a filamentary cathode and two semicylindrical plates

[1] For example, see F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 27, p. 647, October, 1939.

[2] Some possibilities are described by Harry C. Thompson, Electron Beams and Their Applications in Low Voltage Devices, *Proc. I.R.E.*, Vol. 24, p. 1276, October, 1936.

[3] This latter effect is occasionally encountered in multigrid tubes in which there is a virtual cathode in the tube, as, for example, in mixer tubes, and has been known to excite undesired oscillations. For further information on the impedance characteristics of space-charge-grid tubes, see Liss C. Peterson, Impedance Properties of Electron Streams, *Bell System Tech. Jour.*, Vol. 18, p. 465, July, 1939.

[4] Albert W. Hull, The Magnetron, *A.I.E.E. Jour.*, Vol. 40, p. 715, 1921; Léon Brillouin, Theory of the Magnetron, *Elec. Comm.*, Vol. 20, No. 2, p. 112, 1941.

in an axial magnetic field, as illustrated. When the plates are positive, the electrons that are attracted to them follow curved paths, as in Fig. 36. If the plates are very positive, the curvature of these paths is small, but at lower voltages the curvature becomes greater until at a critical potential determined by the strength of the magnetic field, the electrons follow a curved path (cardioid) back to the cathode, as shown in Fig. 36*c*, and never get to the plate, in spite of the positive plate potential. The magnetron accordingly has a critical field strength at which plate current is cut off. This field B in lines per sq cm is

$$B = \frac{6.72}{r}\sqrt{E} \tag{75}$$

where r is the anode radius in centimeters and E is the anode voltage.

If perfect symmetry were obtained and the velocity of emission of the electrons were neglected, the cutoff would be infinitely abrupt; *i.e.*, an infinitesimal change in magnetic field would be capable of turning the entire anode current off or on. Actually, because of unavoidable imperfections, and because of the initial velocity of electrons, the cutoff is rounded off slightly.[1]

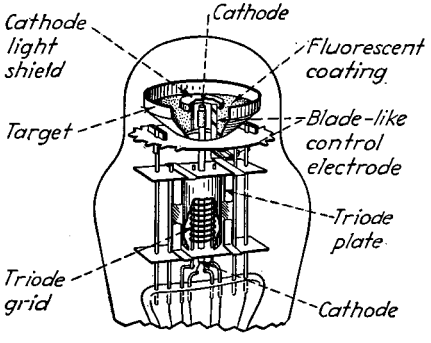


FIG. 37.—Cutaway view of visual indicator tube.

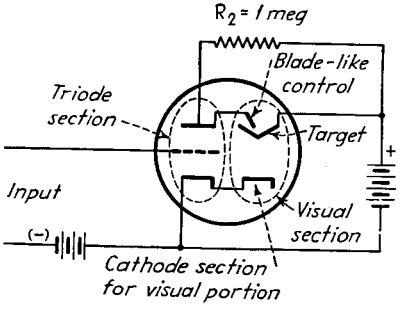


FIG. 38.—Circuit illustrating operation of visual indicator tube.

The chief value of the magnetron is in certain types of very high-frequency oscillators, such as discussed in Par. 14, Sec. 6.

Visual Indicator (Magic-eye) Tube.[2]—A typical form of visual-indicator (so-called magic-eye) tube is illustrated in Fig. 37. This consists of a high-mu triode, above which is located a cone-shaped fluorescent target connected directly to the plate-supply voltage and having its own cathode. There is also a bladelike control electrode associated with the target, which is connected directly to the triode plate. Both plate and blade are connected to the power supply through a high resistance R_2, as shown in Fig. 38. With zero potential on the grid of the triode, electrons from the cathode strike the entire cone-shaped fluorescent target except for a sector in the region of the blade electrode, which, being at a lower potential than the target, because of the voltage dropping resistance R_2, acts as a shield. However, if a negative voltage is applied to the grid of the triode the plate current becomes less. This reduces the voltage drop in resistance R_2, increasing the potential of the control electrode and reducing the sector of darkness on the fluorescent target. Thus a visual indication is obtained of the magnitude of the voltage acting on the triode tube.

[1] E. G. Linder, Effects of High Energy Electron Random Motion upon the Shape of the Magnetron Cutoff Curve, *Jour. Applied Phys.*, Vol. 9, p. 331, May, 1938.

[2] A good description of such a tube, together with some applications, is given by L. C. Waller, Applications of Visual-indicator Type Tubes, *R.C.A. Rev.*, Vol. 1, p. 111, January, 1937.

Visual-indicator tubes can be used for a wide variety of applications. Among these are tuning indicators, vacuum-tube voltmeters, overmodulation indicators, etc.

Orbital Beam Tube.[1]—An understanding of the orbital beam tube can be gained from Fig. 39. This tube is provided with a cathode of relatively flat cross section surround by a control grid G_1 and a screen grid G_2. Focusing electrodes J_1 and J_2 are employed with the outer electrode at zero potential. The inner focusing electrode J_1 is at a relatively high positive potential, so chosen that the electrons emitted from the cathode follow curved paths such as indicated by AA. The radius of these paths may be changed by varying the potential of the outer cylinder. Thus if this cylinder J_2 is made negative, the electrons will follow paths such as BB, and some of the electrons are diverted to the inner cylinder J_1. This is a feature that can be used as a method of volume control. The electrons following the path AA ultimately strike the secondary emitter electrode K_2, which is at a high positive potential, and has a surface so treated as to be a good emitter of secondary electrons. The secondary electrons

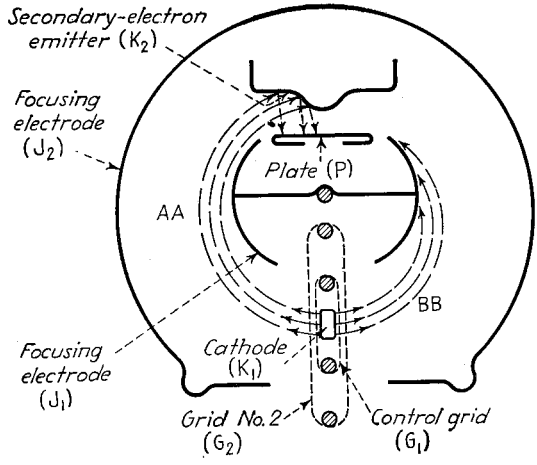


FIG. 39.—Schematic cross section of orbital-beam tube.

produced at electrode K_2 are collected by the plate P, which is at a high positive potential, and serves as the output electrode of the tube.

The orbital beam tube is able to obtain a high transconductance because of the electron multiplier action that takes place at the electrode K_2. It is possible at this electrode to obtain a number of secondary electrons for each arriving primary electron, thereby giving a large plate current in proportion to the current from the cathode. This results in a correspondingly large transconductance. Furthermore, the structural factors involved are such that this output transconductance is very large compared with the input and output capacities of the tube, so that the tube is an excellent wide-band amplifier, being superior to any of the ordinary pentode tubes in this respect.

The rather complicated arrangement of electrodes in the orbital beam tube is for the purpose of allowing the primary electrons emitted from the cathode to strike the secondary emitting surface K_2, while at the same time preventing material evaporated from the cathode from being deposited on K_2. Otherwise the secondary emitting surface would vary in its characteristics with time, and the tube behavior would be unstable.

[1] H. M. Wagner and W. R. Ferris, The Orbital-beam Secondary-electron Multiplier for Ultra-high-frequency Amplification, *Proc. I.R.E.*, Vol. 29, p. 598, November, 1941.

ELECTRON OPTICS AND CATHODE-RAY TUBES

15. Electron Optics.—There is an analogy between the behavior of an electron in an electrostatic field and the behavior of a light ray in a medium of variable index of refraction. Electrons can be reflected, refracted, and focused just as can light rays. The term electron optics implies that the behavior of electrons can be studied in terms of the known laws of physical optics.

A study of electron and physical optics reveals that there actually is a correspondence between the quantities and laws in the two fields. The quantity in electron optics that corresponds to index of refraction is electron velocity (if it is assumed that the electron starts from rest at a point of zero potential). Corresponding to the Fermat principle of least time for physical optics, there is the Hamiltonian principle of least action that applies to electron optics.

In physical optics a ray of light is refracted upon passing a boundary between a medium of one index of refraction and a medium of another. In Fig. 40, a ray passing from a region of refractive n_1 to a region of higher index of refraction n_2 is bent toward the line normal to the surface. Similarly, in the corresponding electron optical case,

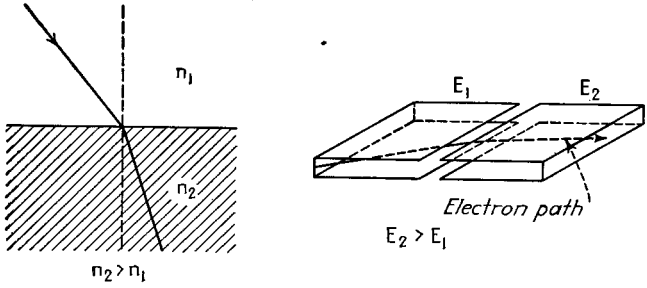


FIG. 40.—Refraction of an electron and of a light beam.

an electron passing from a region in which it has a velocity v_1, proportional to the square root of potential E_1, into a region in which it has a velocity v_2 greater than v_1 is bent toward the line normal to the boundary between the two regions. In most electron optical cases the potential changes gradually rather than abruptly, but any changing potential field is capable of causing an electron to follow a path characterized by a bend.

Any electrostatic field that has a rotational symmetry about some straight-line axis has the properties of a lens. Such a symmetrical electrostatic field is produced by any set of electrodes that themselves have rotational symmetry about the axis. The simplest type of electron lens is a circular hole in a plate, as shown in Fig. 41. If such a circular aperture is placed between planes so that there is a difference in the potential gradient on the two sides, then there is a penetration of field as shown, and electrons entering parallel to the dotted line will be uniformly deflected either toward or away from this axis. For the potentials shown in the figure the lens action is divergent. The physical lens, which is roughly the equivalent of the electron lens, is also shown in the figure.

A better electron lens is shown in Fig. 42, with its approximate physical equivalent. This lens consists of the field between two coaxial cylinders placed end to end, and at different potentials. The variation of the potential along the axis is shown. The equivalent physical lens consists of a combination of convex and concave surfaces between regions of successively higher indexes of refraction. In the equal-diameter

cylinder lens, as this is called, the equipotentials in the first cylinder are concave in the direction of increasing potential, while those in the second cylinder are convex. As a result, an electron entering from the low potential side at a small angle with the axis first experiences a force toward the axis and then experiences a force away from the axis. Thus the first part of the lens has a convergent action, while the second part of

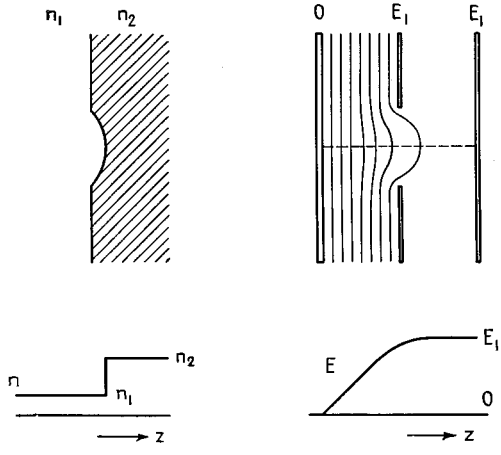


FIG. 41.—Simple electron-optical lens and physical equivalent.

the lens has a divergent action. The convergent action is stronger because the electron velocity is lower in the convergent portion of the lens and hence the electron here suffers a greater deflection for the same force.

The analogy between the electron and physical lenses indicated in Figs. 40 and 41 involves the choice of the proper segments in the physical lens. It will be observed that the front surface of the physical-lens components are curved in the same way as

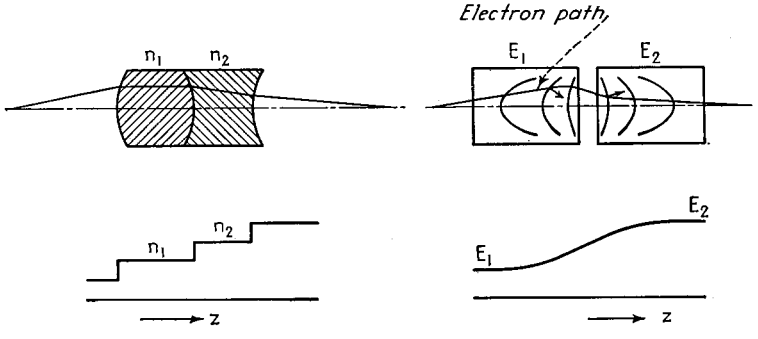


FIG. 42.—Equal-diameter cylinder lens and physical equivalent.

are the equipotentials in the electron lenses. This observation gives a means of telling by inspection how an electron lens will behave, since the property is a general one. There is also a correlation between the nature of the equipotential surfaces and the curvature of the curve of potential along the axis against distance in the axial direction. In Fig. 42 it is seen that when the curvature of the potential distribution along the axis is concave upward, then the equipotentials are concave in the direction of increasing

potential and the lens has a convergent action. In this region the second derivative with distance of the axial potential is positive. In the second cylinder the second derivative of the axial potential is negative, and the lens action is divergent. This relation between the curvature of the curve of potential along the axis and the nature of the lens action always holds for lenses with rotational symmetry.

In fields with rotational symmetry the potential at any point in the field can be determined if the variation of potential with distance along the axis is known. By an application of Laplace's equation to fields of this sort, it can be shown that

$$E(r,z) = E_0(z) - \frac{r^2}{2^2} E_0''(z) + \frac{r^4}{2^2 \cdot 4^2} E_0''''(z) + \cdots + \frac{(-1)^n}{(n!)^2} \left(\frac{r}{2}\right)^{2n} E_0^{(2n)}(z) + \cdots \tag{76}$$

where $E(r,z)$ = potential at a radial distance r from the axis and at an axial distance z from some reference point on the axis.

$E_0(z)$ = potential on the axis at the point $(0,z)$.

$E_0''(z)$ = second derivative of axial potential with respect to axial distance z.

$E_0^{(2n)}$ = 2 nth derivative of axial potential with respect to axial distance z.

The axial potential can be calculated in some cases[1] but in most cases it is simpler to obtain it by a computational method or by the use of a current flow model.[2]

When the axial potential is known its derivatives are readily obtained, and thus the potential at any point in the field can be calculated from Eq. (76).

The differential equation of an electron path through a field of rotational symmetry is

$$\frac{d^2r}{dz^2} - \frac{E_0'(z)dr}{2E_0(z)dz} + E_0'' \frac{r}{2} = 0 \tag{77}$$

This equation is good only for the so-called "paraxial" electrons, which are those electrons that are within a few per cent of the electrode radius from the axis, and that make an angle of a few degrees or less with the axis. The preceding equation is practically incapable of solution exactly. It is, however, possible to obtain good numerical approximations of its solution by various approximate methods.[3]

Although it is not possible to solve Eq. (77), it is still possible to make some deductions about the nature of the electron paths from it. Since the differential equation is of the second order, its complete solution may be expressed as the sum of two linearly independent solutions; *i.e.*, any electron path through the lens may be described as the sum of any two different and distinct paths. It is most convenient to use as a pair of independent paths those two which represent (1) an electron entering the lens parallel to the axis and (2) an electron leaving the lens parallel to the axis. These particular rays are known as the *principal rays* of the lens, and are indicated in Fig. 43. The ray that leaves the lens parallel to the axis is known as the *first principal ray.* It is sometimes shown as entering the lens parallel to the axis from the right as in Fig. 43, which is permissible because such a reversal of direction traces an identical path. The ray that enters the lens parallel to the axis from the

[1] P. Kirkpatrick and J. G. Beckerley, Ion Optics of Equal Coaxial Cylinders, *Rev. Sci. Instruments*, Vol. 7, p. 24, January, 1937; S. Bertram, Determination of Axial Potential Distribution in Axially Symmetric Fields, *Proc. I.R.E.*, Vol. 28, p. 418, September, 1940; S. Bertram, Calculations of Axially Symmetric Fields, *Jour. Applied Phys.*, Vol. 13, p. 496, August, 1942.

[2] I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938; L. M. Myers, Electron Optics, Van Nostrand, New York, 1939; O. Klemperer, "Electron Optics," Cambridge (London), 1939; M. Bowman-Manifold and F. H. Nicoll, Electrolytic Field-plotting Trough for Circularly Symmetric Systems, *Nature*, Vol. 142, p. 39, July 2, 1938.

[3] Maloff and Epstein, *loc. cit.*; R. Gans, Course of the Ray in Electron Optical Systems, *Zt. Physik*, Vol. 18, p. 41, February, 1937; K. Spangenberg and L. M. Field, Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses, *Proc. I.R.E.*, Vol. 30, p. 138, March, 1942.

left is known as the *second principal ray.* The *focal points* of a lens are those points at which the principal rays cross the axis. The intersection of the initial and final straight-line portions of the principal rays when extended toward the lens between the principal planes and the focal points of the lens are known as the focal distances or *focal lengths.* These quantities are shown in Fig. 43.

A lens such as that of Fig. 43, in which the region in which the lens action takes place is comparable in length to the focal length, is known as a thick lens. The

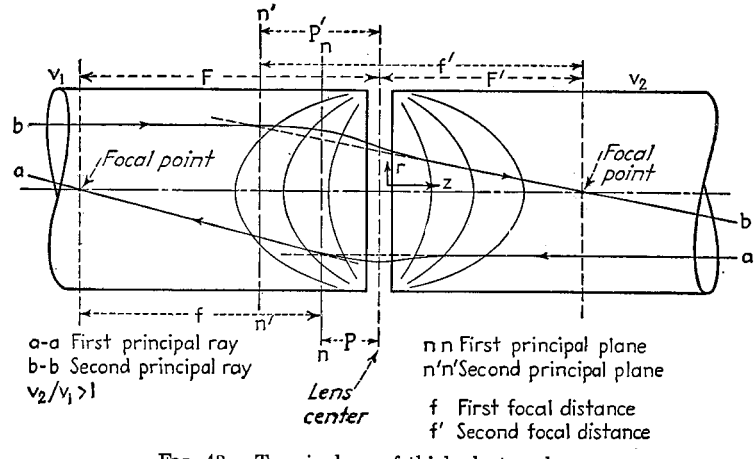


FIG. 43.—Terminology of thick electron lens.

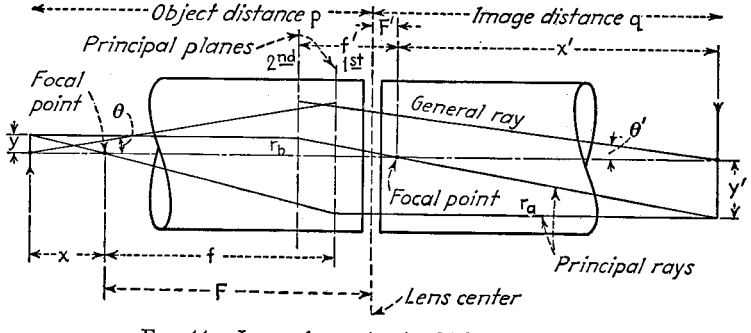


FIG. 44.—Image formation in thick electron lens.

relation between the lens characteristics and the points at which an image of a given object forms as in Fig. 44 is

$$\frac{f}{P+P} + \frac{f'}{Q+P'} = 1 \tag{78}$$

where f = first focal length.
f' = second focal length.
p = distance between object and lens center.
q = distance between image and lens center.
P = distance between first principal plane and lens center.
P' = distance between second principal plane and lens center.

This equation tells what the location of the image will be for any location of the object in terms of four fundamental quantities of the lens, f, f', P, and P'. The object and image distances as measured from their corresponding principal planes are related by Newton's formula

$$pq = ff' \tag{79}$$

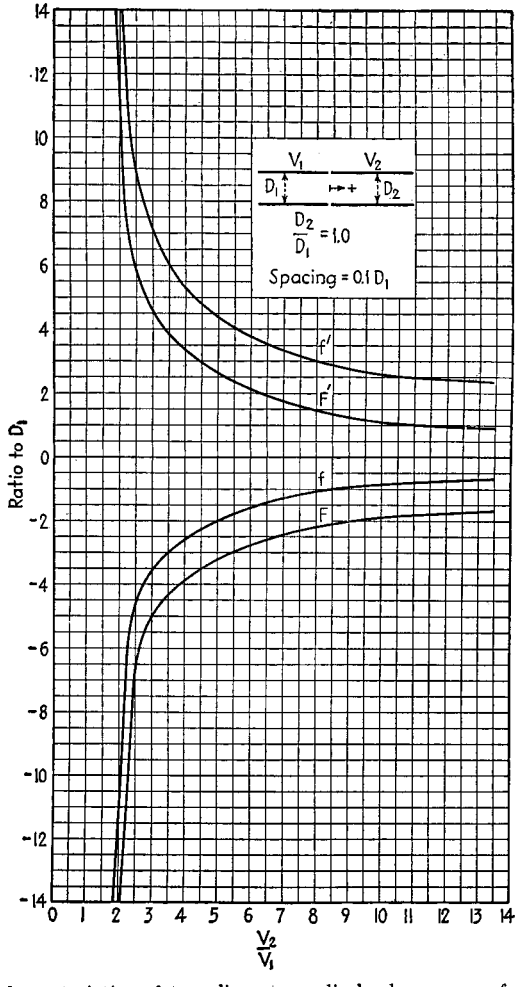


FIG. 45.—Focal characteristics of two-diameter cylinder lenses as a function of electrode voltage ratio. Figure continued on pages 327–330.

In electron lenses, the ratio of the two focal distances is related to the ratio of the electrode potentials by

$$\frac{f'}{f} = \left(\frac{V_2}{V_1}\right)^{1/2} \tag{80}$$

where V_1 = initial potential, that of first electrode.
V_2 = final potential, that of second electrode.

There is also a relationship known as Lagrange's law between the lateral magnification, the angular magnification, and the electrode voltage ratio.

$$m_1 m_a \left(\frac{E_2}{E_1}\right)^{1/2} = 1 \tag{81}$$

where $m_1 = y'/y$ in Fig. 44, lateral magnification.
$m_a = \theta'/\theta$ in Fig. 44, angular magnification.
E_2 = second electrode potential.
E_1 = first electrode potential.

Focusing Characteristics of Lenses.—The lenses of electron optics are completely described if four quantities are known. The four quantities most convenient to use

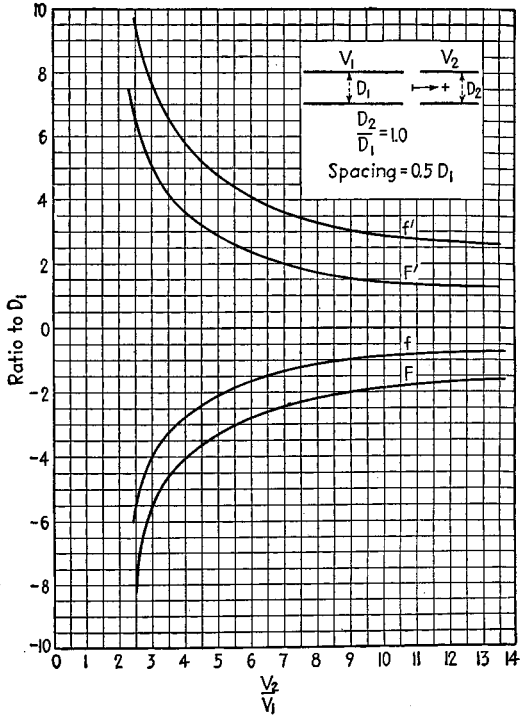


FIG. 45.—*Continued.*

are the two focal lengths and the locations of the two principal planes. Since these change with voltage ratio, it is necessary to draw curves of the four quantities listed above as a function of voltage ratio, as is done in Figs. 45 and 46.[1]

To utilize the curves of Figs. 45 and 46 use must be made of the lens relations previously given. For any voltage ratio and location of object, they enable one to calculate the location and size of the image. Since the quantities of Figs. 45 and 46 do not appear in the answers, it is desirable to present the relations between object distance, image distance, voltage ratio, and magnification in a form that does not

[1] The data presented in Figs. 45 to 49 are experimental results obtained by K. Spangenberg and L. M. Field in carrying out an investigation at Stanford University with the support of the International Standard Electric Corporation. The results are given here through the courtesy of that organization.

involve the focal distances. This has been done in Figs. 47 and 48, which present what may be called "object-image–distance curves." On these curves, which are essentially a graphical presentation of all the answers to all the problems associated with one electrode configuration, there are shown the relationships between the object and image distances and the lateral magnification m as determined by the voltage ratio. Examination of these curves shows that the magnification is approximately

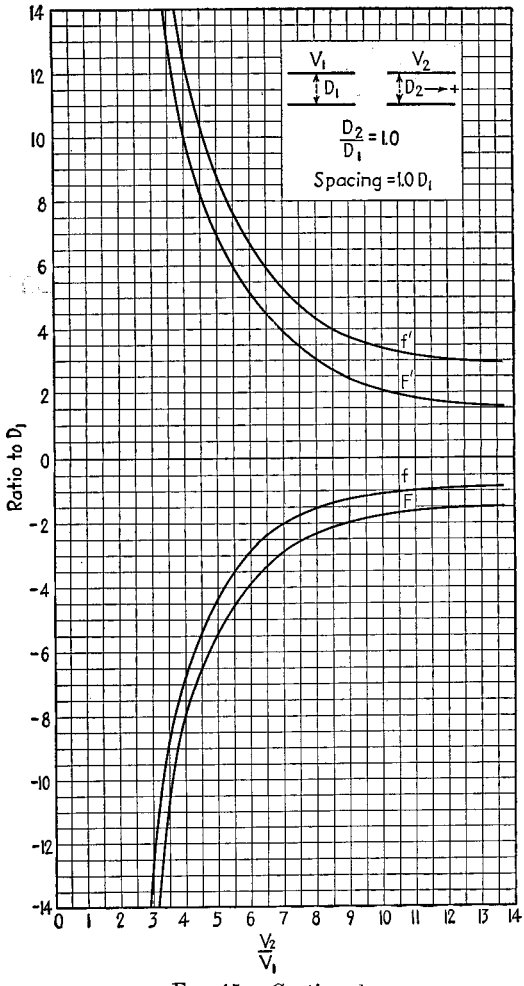


FIG. 45.—*Continued.*

0.8 of the ratio of the image to the object distance. It also shows that there is always some particular electrode configuration that makes it possible to achieve any ratio of object to image distance or magnification, provided that the proper voltage ratio is chosen.

The effect of varying the proportions of various types of lenses is illustrated in Fig. 49.

A critical examination of data such as given in Figs. 45 to 49 shows that the following properties are common to all lenses:[1] (1) Focal lengths are always uniformly decreasing functions of voltage ratio. (2) Principal planes always lie on the low-voltage side of the lens center. (3) Principal planes are crossed, with exception of

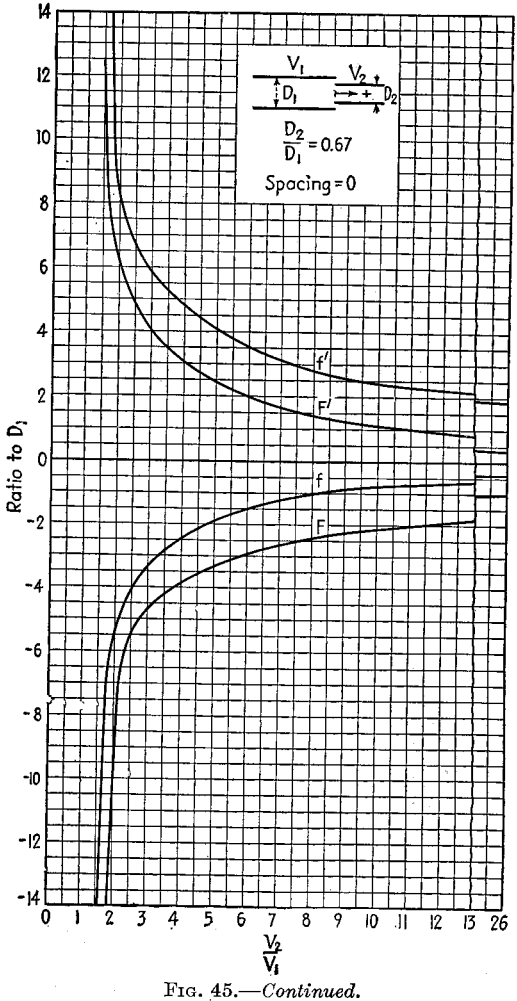


FIG. 45.—*Continued.*

large-diameter aperture lens; *i.e.*, the first principal plane lies between the second principal plane, and the lens center on the low-voltage side of the lens. (4) Focal length in the direction of increasing potential is always greater than the focal length in the other direction. (5) Position of principal planes does not change much with voltage ratio.

[1] Maloff and Epstein, *loc. cit.;* Klemperer, *loc. cit.;* Spangenberg and Field, *loc. cit.;* Frank Gray, Electrostatic Electron Optics, *Bell System Tech. Jour.*, Vol. 18, p. 1, January, 1939.

The following summarizes properties of focal lengths for the particular cases indicated: (1) The focal length of two-diameter cylinder lenses increases—*i.e.*, the lens grows weaker—for all but the highest voltage ratios, as the ratio of second to first cylinder diameter increases. (2) The focal length of equal-diameter cylinder

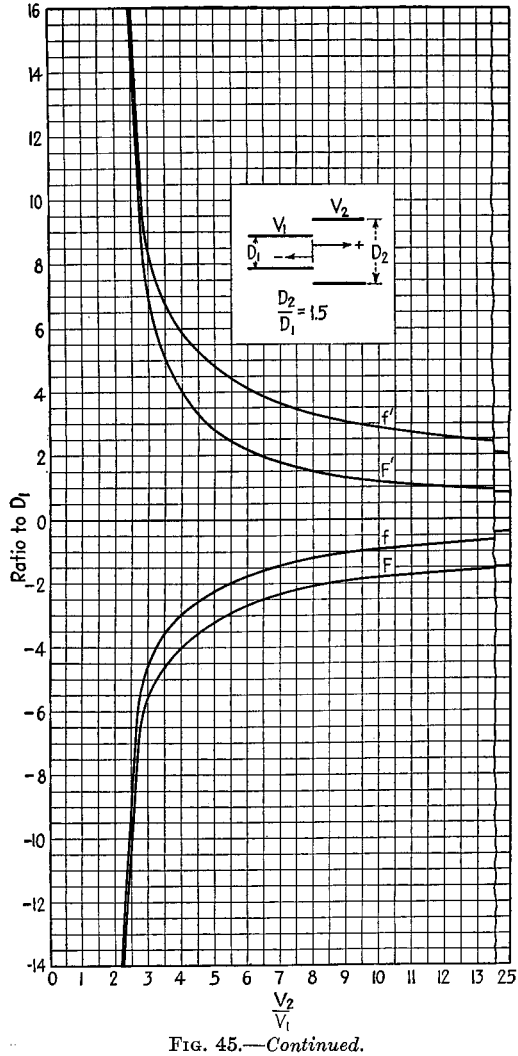


FIG. 45.—*Continued.*

lenses increases, that is, the lens grows weaker, as the axial spacing of the cylinders increases. The change is small for small spacings but increases rapidly as the spacing is increased. (3) The focal length of aperture lenses increases—*i.e.*, the lens grows weaker—as the aperture diameter increases. The change is small for small diameters but increases rapidly as the diameter increases. (4) Aperture lenses have for the most part shorter focal lengths than cylinder lenses (when the unit of length is the

diameter of the first cylinder) if the aperture spacing be taken equal to first cylinder diameter. (5) Cylinder-aperture lenses have shortest focal length of all the lenses of Figs. 45 to 49. (6) Equal-diameter lenses with axial spacing of one diameter have the longest focal length of all the lenses of Figs. 45 to 49.

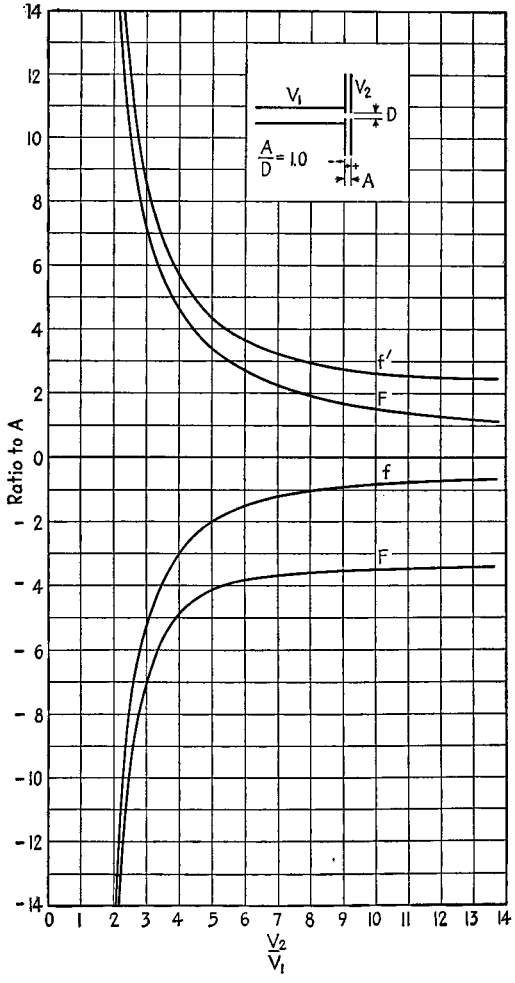


FIG. 46.—Focal characteristics of various aperture lenses as a function of electrode voltage ratio. Figure continued on pages 332–334.

The following comments apply to the magnification properties of the lenses of Figs. 45 to 49: (1) Contours of constant magnification in the object-image distance curves are approximately straight lines with a slope of one. (2) An approximate universal magnification formula that fits all lenses tested is

$$M = k\frac{q}{p} \tag{82}$$

TABLE 5.—CONSTANT k IN EQ. (82)

Type of Lens	k
Cylinder lens: $D_2/D_1 = 0.667$	0.82
Cylinder lens: $D_2/D_1 = 1$, $S = 0.1$	0.78
Cylinder lens: $D_2/D_1 = 1.5$	0.76
Cylinder lens: $D_2/D_1 = 1$, $S = 0.5$	0.80
Cylinder lens: $D_2/D_1 = 1$, $S = 1$	0.60
Aperture lens: $A/D = 5$	0.95
Aperture lens: $A/D = 3$	0.80
Aperture lens: $A/D = 1$	0.78
Cylinder aperture lens	0.82

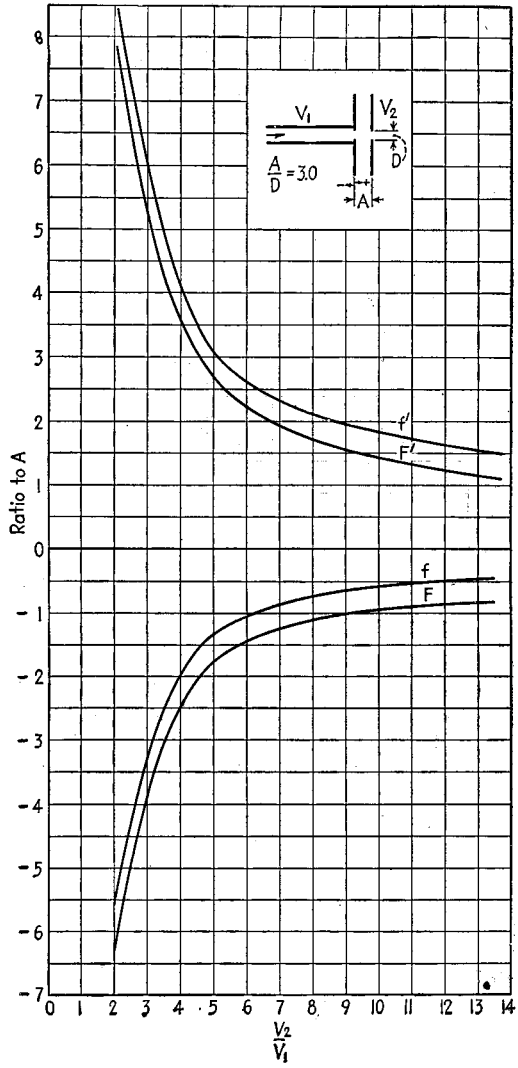


FIG. 46.—*Continued.*

where p and q are object and image distances, respectively. Values of the constant k for the lenses tested are given in Table 5, and with only two exceptions are within a few per cent of 0.8. (3) Magnification in two-diameter lenses decreases as the diameter ratio increases. The change is appreciable only for ratios less than one, however.

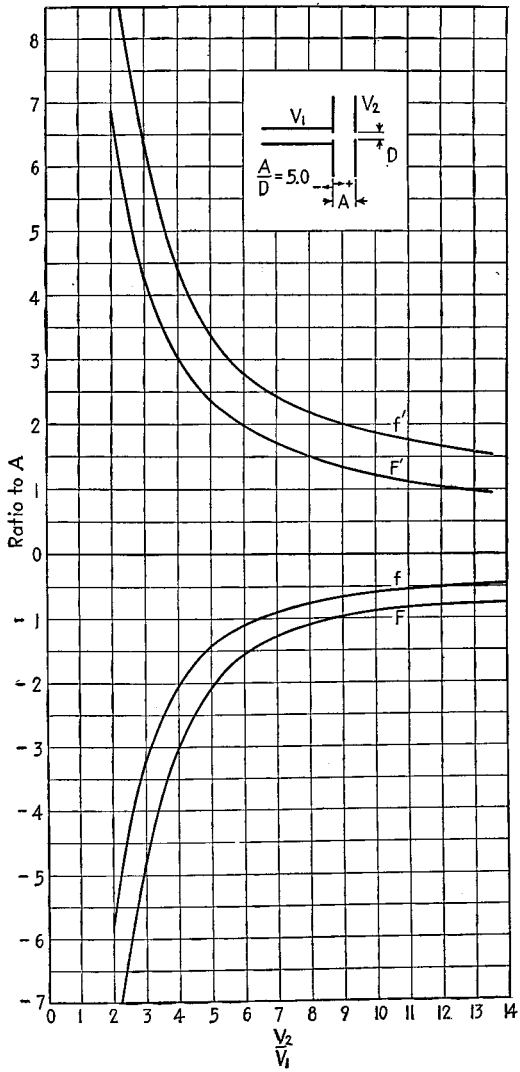


FIG. 46.—*Continued.*

(4) Magnification of equal-diameter lenses decreases as axial spacing increases. The change is inappreciable except between spacings of 0.5 and 1.0, however. (5) Magnification of aperture lenses with small apertures decreases as aperture diameter increases.

Considerations important in design that are common to all lenses of Figs. 45 to 49 are: (1) As object distance is increased at a given voltage ratio, the corresponding

image distance decreases, as does also the magnification. (2) For a given object distance the image distance and magnification decrease as the voltage ratio is increased. (3) In any lens there is a minimum object distance that can be used at any given voltage ratio. This minimum object distance is determined by the vertical asymptote to the contour of constant voltage ratio.

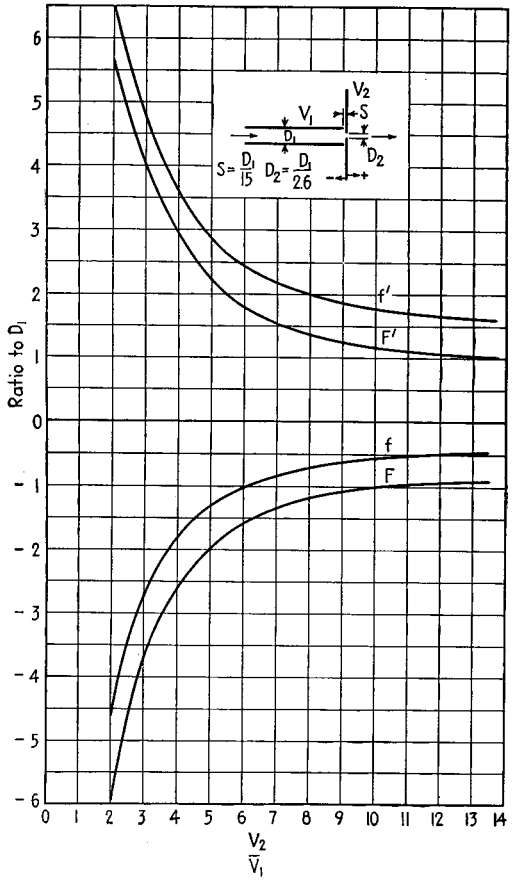
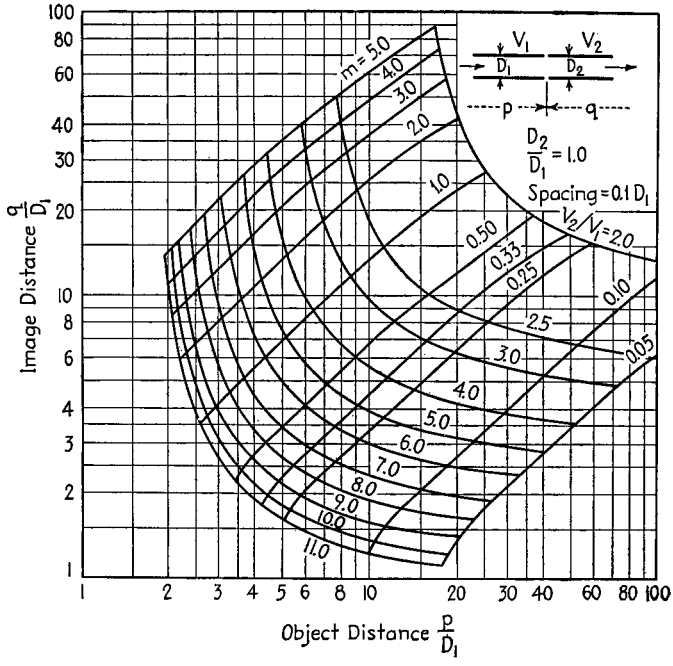


FIG. 46.—*Continued.*

The effects of object distance and voltage ratio on the relative strength of the lenses of Figs. 45 to 49 can be summarized as follows: (1) Two-diameter cylindrical lenses become stronger as diameter ratio decreases for voltage ratios between 1 and 4. This holds for all object and image distances. (2) For larger voltage ratios the lens strength is nearly independent of diameter ratio. (3) Equal-diameter lenses become stronger as the axial separation of the cylinders increases for all voltage ratios and normal values of object and image distance. (4) Aperture lenses become slightly stronger as the aperture diameter is enlarged from a very small size to a practical value, and then become greatly weaker.



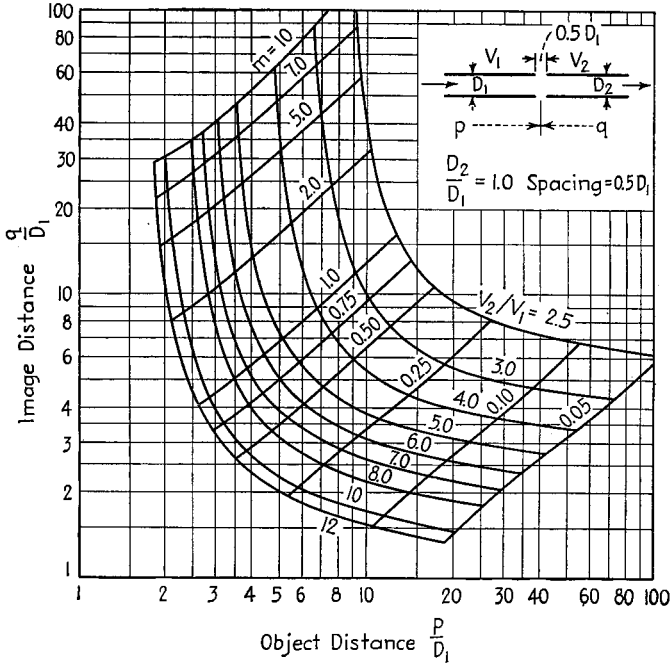
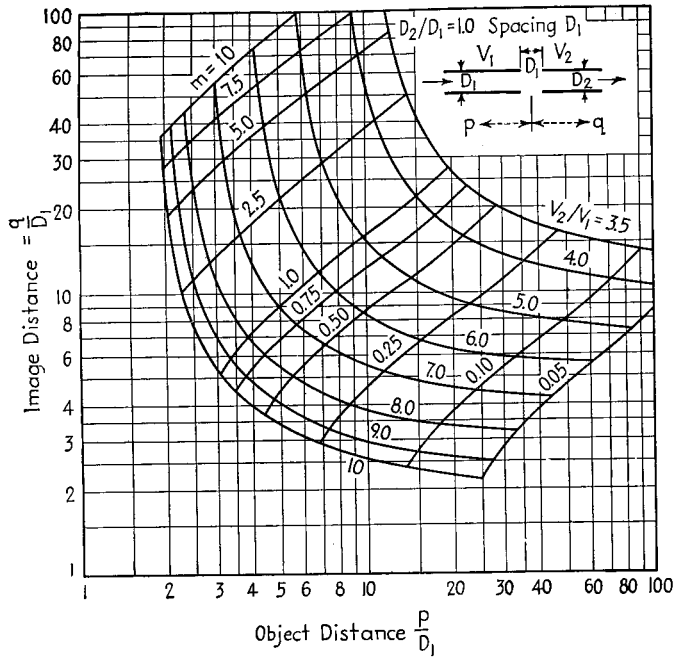


FIG. 47.—Image-object distance curves of two-diameter cylinder lenses. Figure continued on pages 336 and 337.



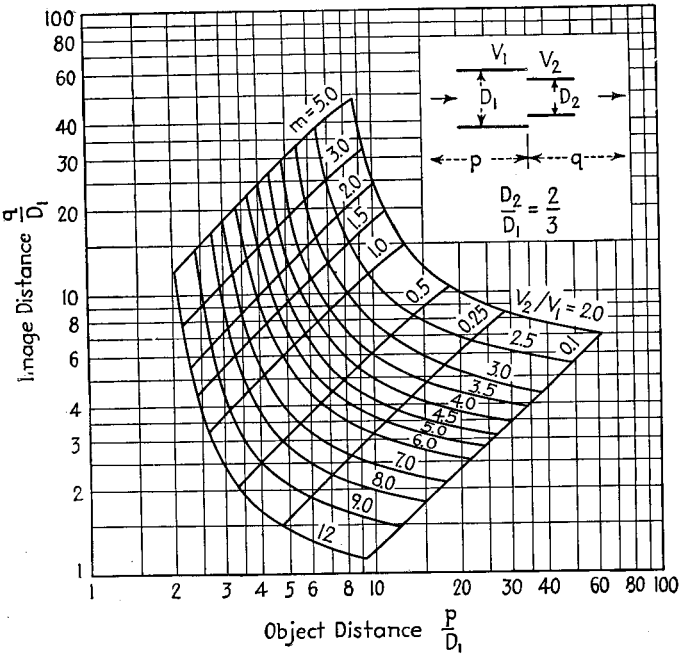


FIG. 47.—*Continued.*

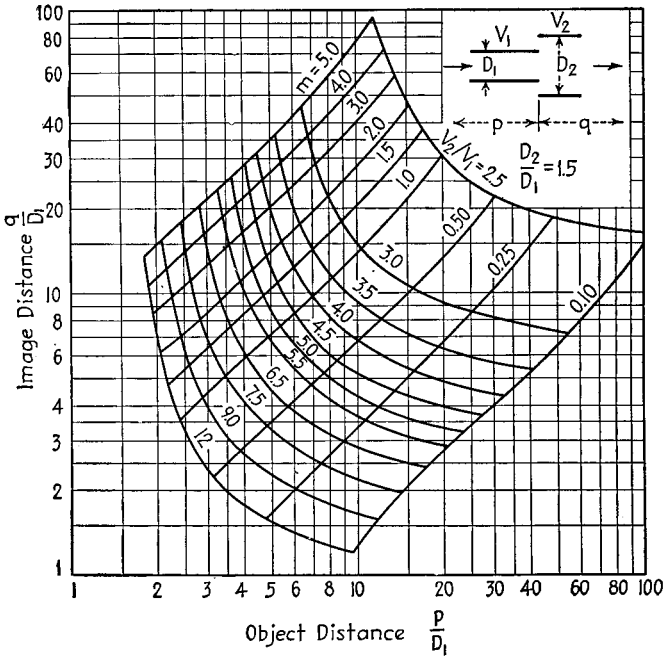


FIG. 47.—*Continued.*

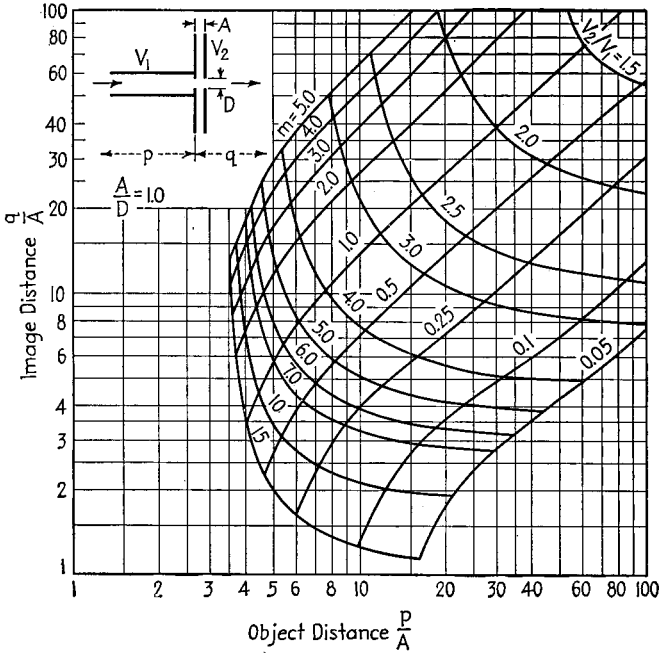
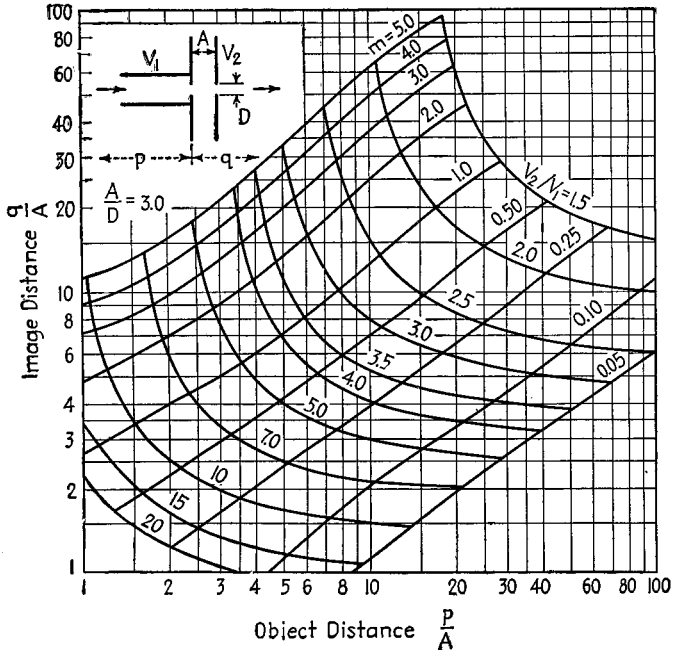


FIG. 48.—Image-object distance curves of various aperture lenses. Figure continued on pages 338 and 339.



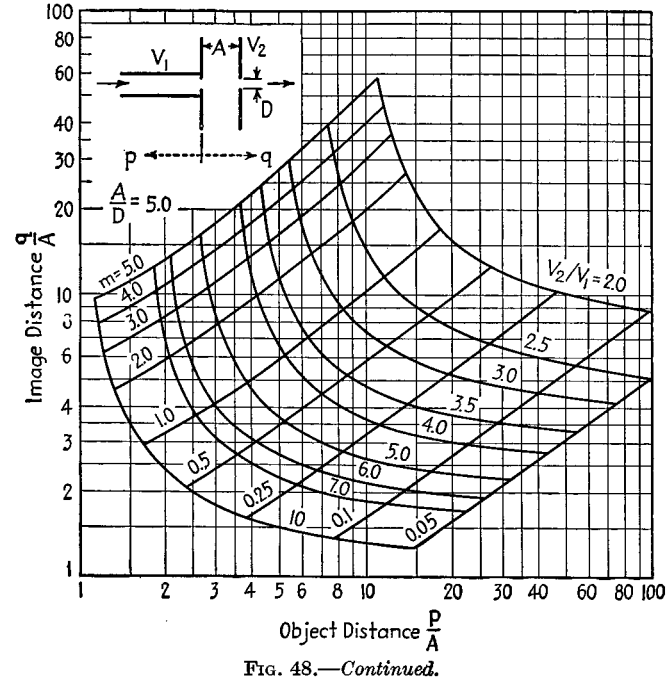


FIG. 48.—*Continued.*

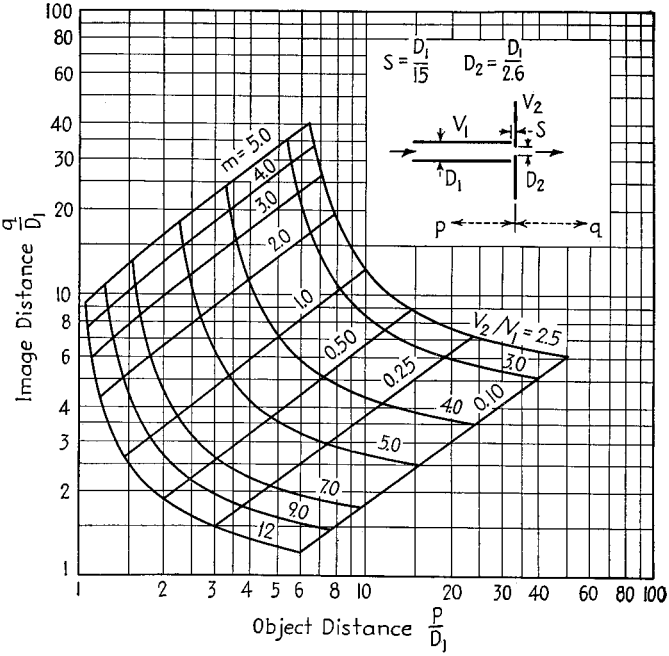
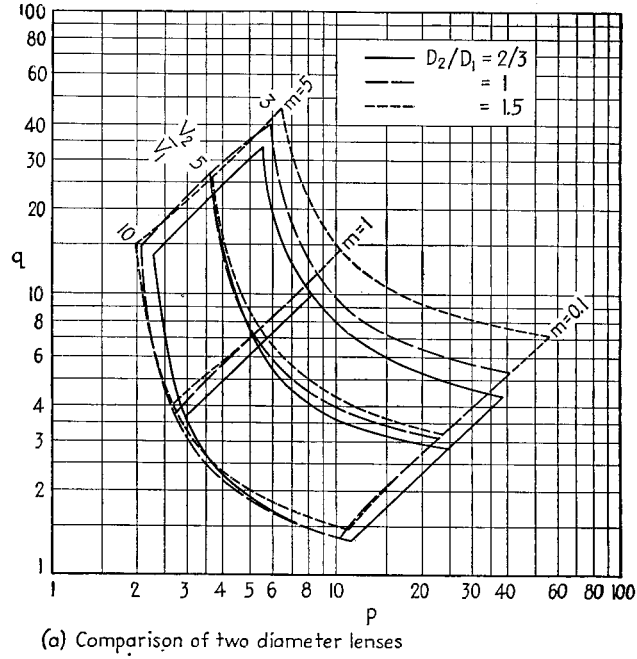
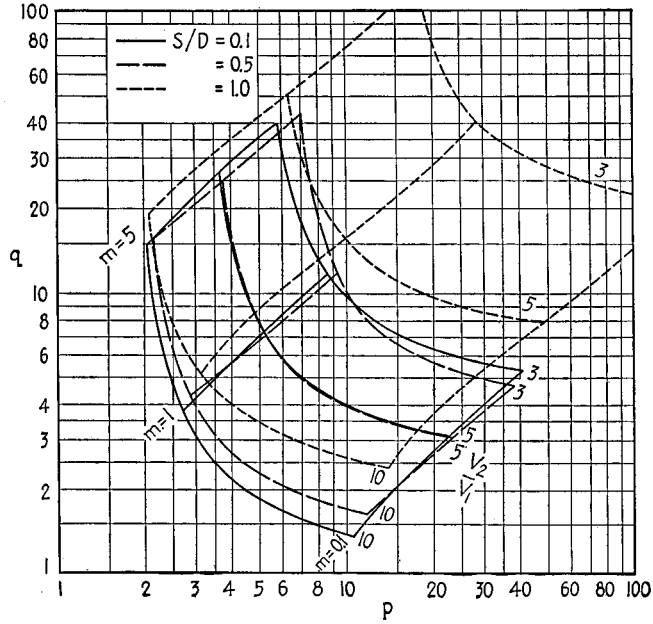


FIG. 48.—*Continued.*

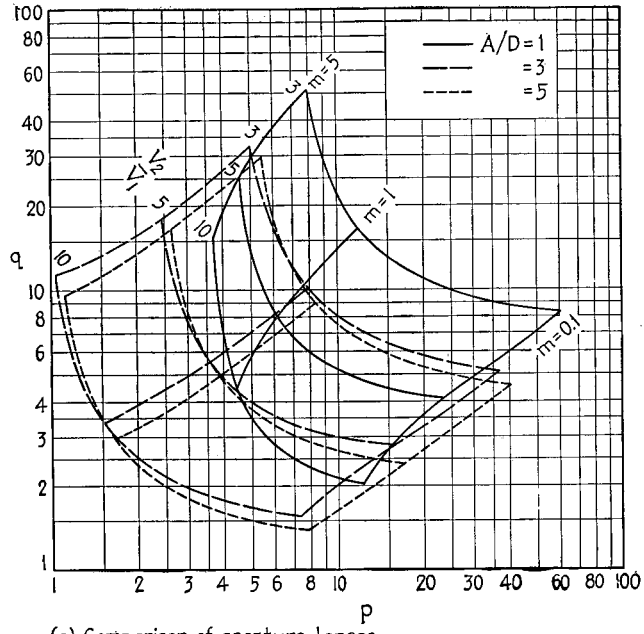


(a) Comparison of two diameter lenses

FIG. 49.—Comparison of lenses of different proportions. Figure continued on page 340.



(b) Comparison of equal diameter lenses



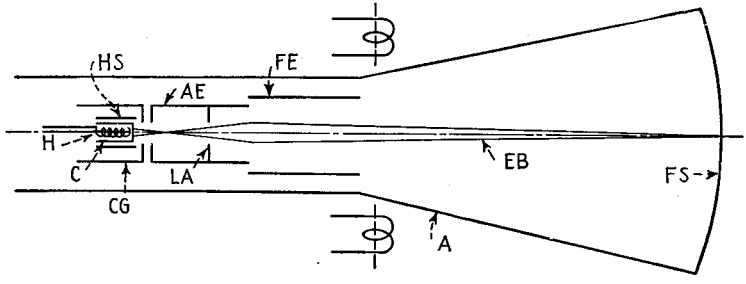
(c) Comparison of aperture lenses

FIG. 49.—*Continued.*

16. Cathode-ray Tubes.—The cathode-ray tube is an outgrowth of the modern science of electron optics. It is used extensively in engineering and scientific laboratories for such purposes as the observation of wave forms, viewing of television images, and many other applications too numerous to mention.

Every cathode-ray tube contains a source of electrons in the form of a cathode, a means for forming the emitted electrons into a pencillike beam; a control electrode for varying the beam current; a set of focusing electrodes for focusing the beam; some sort of beam-deflecting device for moving the beam about; and a fluorescent screen, which gives off light at the point where the beam strikes.

The Electron Gun.—A typical electron-gun structure is shown in Fig. 50. The cathode consists of a cylinder of sheet nickel with an oxide-coated cap on one end. The cathode is heated by a tungsten filament coated with some insulating material, such as aluminum oxide, which is inserted in the cylinder. Outside the cathode and spaced a few thousandths of an inch from it is a heat shield in the form of a nickel



H Heater
C Cathode (0.0 volts)
HS Heat shield (0.0 V)
Cg Control grid (-10 to +10 V)
AE Accelerating electrode (+200 V)
FE Focusing electrode (+800 V)
EB Electron beam
FS Fluorescent screen
LA Limiting aperture
A Aquadag coating
Potentials above are relative to cathode
Actually the focusing electrode and aquadag coating are grounded

FIG. 50.—Conventional electron gun structure with magnetic deflecting coils.

cylinder, which projects slightly beyond the end of the cathode and helps to concentrate the electrons in a beam.

The control electrode, which performs the function of a control grid in an ordinary vacuum tube, is in the form of an enveloping cylinder with aperture, properly spaced relative to the cathode.[1] The control electrode is operated at some low potential relative to the cathode, which may be either positive or negative, depending upon the electrode dimensions.

The control electrode is followed by a so-called accelerating electrode, or first anode. This usually has an aperture in the end next to the control electrode and one or two more apertures within the cylinder. The apertures within the cylinder serve to remove from the beam strongly divergent electrons that would produce a fuzzy spot.

The last electrode is the second anode, or focusing electrode, which is another cylinder following the accelerating electrode. The principal focusing action occurs in the electrostatic field between the accelerating electrode and the second anode.

[1] Details of the design of the cathode and associated electrodes are given by I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill, New York, 1938.

The beam is focused by changing the second anode potential, just as the beam intensity is controlled by changing the control-electrode potential. These two control actions are almost but not completely independent; *i.e.*, changing the beam intensity affects the focus somewhat and vice versa. The funnel-shaped part of the glass envelope beyond the second anode is frequently coated with aquadag, which is a conducting layer of carbon particles, and placed at the potential of this electrode. In this arrangement the deflecting devices are placed just beyond the second anode.

Another electron gun structure that can be used is shown in Fig. 51. This arrangement makes use of what is known as postdeflection acceleration.[1] Here the cathode is followed by a control electrode and then by an accelerating electrode, or anode, A_1. This accelerating electrode is the first of two at the same potential, and is followed by an electrode at a lower potential, which in this structure is called the focusing electrode.

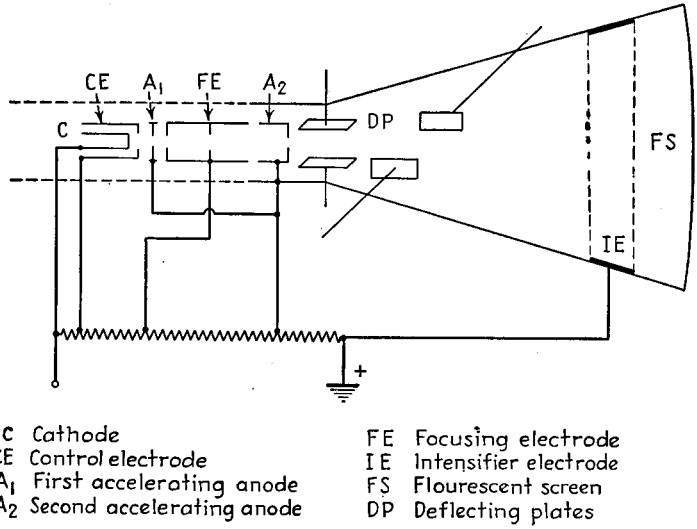


FIG. 51.—Electron gun with postelectrostatic deflection acceleration.

This focusing electrode *FE* is followed by a second accelerating anode A_2 at the same potential as the first. The second accelerating electrode is followed by the deflecting devices, placed at the beginning of the funnel-shaped part of the glass envelope, that is coated with aquadag. This conducting coating extends only part way to the fluorescent screen. It is followed by a ring of conducting material just before the screen, known as the intensifier electrode, which is maintained at about twice the potential of the accelerating electrode.

The convergent focusing action on the beam occurs between the focusing electrode and the second accelerating electrode. The action between the first accelerating electrode and the focusing electrode is divergent, as is also the action between the first accelerating electrode and the control electrode.

In the postdeflection arrangement the beam is deflected at low velocity, thus giving a good deflection sensitivity, and then accelerated further to give a good photographic sensitivity. Part of the increase in deflection sensitivity gained by this arrangement is lost because the final accelerating field is somewhat convergent, thus reducing the deflection. However, a gain of 3 to 5 over the ordinary tube sensitivity is achieved.

[1] J. de Gier, A Cathode Ray Tube with Post Acceleration, *Phillips Tech. Rev.*, Vol. 5, p. 245, September, 1940; J. R. Pierce, After Acceleration and Deflection, *Proc. I.R.E.*, Vol. 29, p. 28, January, 1941.

Deflection of Electron Beams.—The electron beam in a cathode-ray tube may be deflected either by an electrostatic or by an electromagnetic field. In the majority of applications, deflection by electrostatic means is preferred.

If a conducting plate is placed on either side of the beam at the end of the electron gun, as shown in Fig. 51, then if the mean potential of these plates is that of the last gun electrode, and yet a small potential difference is caused to appear between the plates, the electron beam will be deflected toward the more positive plate upon passing between them. This deflection occurs because there is a constant crosswise force applied to the beam as it moves through the plates with a nearly constant velocity. The result of this action is that the electrons move in a parabolic path while between the plates, and emerge at an angle with the axis that is proportional to the potential difference existing between the deflecting plates. The amount that a beam is deflected by this action is

$$d = \frac{La}{2b}\frac{E_d}{E_a} \tag{83}$$

where d = deflection at fluorescent screen.
L = length of beam from the center of deflecting plates to screen.
a = length of deflection plates.
b = spacing of deflecting plates.
E_d = deflecting-plate voltage.
E_a = beam voltage.

Ordinarily, two pairs of plates are used, one for horizontal and one for vertical deflection. To allow for the deflection of the beam, the plates are often at a small angle with the axis, so that the spacing at the end toward the fluorescent screen is greater than that at the end toward the electron gun. The second pair of plates is usually spaced farther apart than the first pair because of the deflection introduced by the first pair.

The electron beam may also be deflected by magnetic means. This is done by using the arrangement of coils shown in Fig. 50. In the region of approximately uniform magnetic field between the coils, the electrons move in portions of a circular path in a plane perpendicular to the axis of the coils. This action causes the beam to be deflected. The amount of the deflection is

$$d = \frac{La'B}{3.37\sqrt{E_a}} \tag{84}$$

where d = beam deflection at fluorescent screen, cm.
L = beam length between coil center and fluorescent screen, cm.
a' = axial length of region of uniform magnetic field, cm.
B = magnetic field strength in lines per sq cm.
E_a = beam potential, volts.

For both types of deflection, the amount of the deflection decreases as the beam potential increases. This means that if the attempt is made to increase the spot brightness, *i.e.*, the photographic sensitivity, by increasing the beam potential, then the amount of the deflection or the deflection sensitivity will be decreased. In general, any attempt to increase the photographic sensitivity decreases the deflection sensitivity, and vice versa.

The Fluorescent Screen.—Various materials will give off light upon bombardment by electrons in a vacuum. Such materials are called fluorescent in that the light they give off stops almost immediately when the excitation is removed.

Common fluorescent materials include phosphors of zinc, cadmium, and calcium. The characteristics of the principal phosphors are listed in Table 6.[1] The most com-

[1] Reprinted, by permission, from "Television," by V. K. Zworykin and G. A. Morton, John Wiley & Sons, Inc., New York, 1940.

monly used material is willemite, which is zinc orthosilicate (Zn_2SiO_4). The synthetic compound used with a manganese activator gives off a bright yellow-green light having its maximum spectral energy at 5200 angstroms and having a luminous efficiency of 10 to 20 lumens per watt. This material works best when the coating contains 7 milligrams per square centimeter.

Zinc sulphide has a long-persistence characteristic that makes it useful in the study of transients. A white screen is obtained by combining a green-blue zinc sulphide with a yellow zinc-cadmium sulphide, or by a similar combination of different colors.

Coatings are made by settling out the material from a water suspension. The water must be removed very slowly after the material has settled out. This is sometimes done by means of a clock mechanism. A mild electrolyte such as ammonium carbonate is used to prevent the particles from settling nonuniformly. Screens may also be made by spraying, using a volatile organic liquid such as acetone with a small amount of binder.

TABLE 6.—PROPERTIES OF PRINCIPAL PHOSPHORS

Phosphor	Chemical composition	Color	Spectral maximum, Å	Approximate candlepower per watt*
Zinc oxide	ZnO	Violet	Ultraviolet	<0.1
Zinc sulphide	ZnS:Ag	Blue-violet	4700–4500	5_T
Calcium tungstate	$CaWO_4$	Blue	4300	<1
Zinc silicate	$ZnO + SiO_2$	Blue	4200	<1
Zinc sulphide	ZnS	Light blue	4700	$1–5_T$
Zinc aluminate	$(ZnO + Al_2O_3)$:Mn	Green-blue	5130	~1
Zinc silicate (willemite)	$(ZnO + SiO_2)$:Mn	Blue-green	5230	3_T
Zinc sulphide	ZnS:Cu	Green	4700–5250	$>4_T$
Zinc germanate	$(ZnO + GeO_2)$:Mn	Yellow-green	5370	1.5
Beta zinc silicate	$(ZnO + SiO_2)$:Mn	Green-yellow	5600–5700	3_T
Zinc beryllium silicate	$(ZnO + BeO + SiO_2)$:Mn	Green to orange	5230–6500	$1–2_T$
Zinc cadmium sulphide	(ZnS + CdS):Ag	Blue to red	4700–> 7000	5_T
Calcium silicate	$(CaO + SiO_2)$:Mn	Green to orange	5500–6500	<1
Cadmium silicate	$(CdO + SiO_2)$:Mn	Orange-yellow	5850	~1
Magnesium silicate	$(MgO + SiO_2)$:Mn	Orange-red	6400–6700	<1
Zinc aluminate	$(ZnO + Al_2O_3)$:Cr	Red	> 7000	<1
Zinc beryllium zirconium silicate.	$[ZnO + BeO + (Ti - Zr - Th - O_2) + SiO_2]$:Mn	White	4200 + 5500–6000	$\sim 1_T$
Magnesium tungstate	$MgO + WO_3$	Very light blue	4800	<1
Zinc borate	$(ZnO + B_2O_3)$:Mn	Yellow-orange	5400–6000	$\sim 1_T$
Cadmium borate	$(CdO + B_2O_3)$:Mn	Green-orange	5300–6300	<1
Cadmium tungstate	$CdO + WO_3$	Light blue	4900	$<1_T$

* T = used in television.

GAS TUBES

17. Hot-cathode Gaseous Rectifier Tubes.[1]—A hot-cathode gaseous rectifier tube is an ordinary diode tube containing gas such as mercury vapor in equilibrium with liquid mercury, or argon, neon, etc. The gas pressure is ordinarily quite low, of the order of 10^{-3} mm of mercury. When a voltage is applied between emitting electrode and collector, the current-voltage relations are at first like those in a vacuum diode; *i.e.*, the current flow consists only of electrons emitted from the cathode and follows a three-halves power law. However, as the anode potential is increased, the electrons finally acquire enough energy to ionize the particles of gas by collision. The current that was formerly limited by the negative space charge of the electrons now has this space charge annulled by these positive ions, and the current rises extremely rapidly with voltage, quickly reaching the full electron emission. This current-voltage rela-

[1] For further information see A. W Hull, Gas-filled Thermionic Tubes, *Trans. A.I.E.E.*, Vol. 47, p. 753, 1928. See also additional discussion of rectifier tubes of this type given in Par. 1, Sec. 8.

tion is shown in Fig. 52 in comparison with the current-voltage relations of the same tube completely evacuated.

When ionization occurs in the tube there is formed a region between cathode and plate known as the "plasma," in which positive ions collect and are approximately in equilibrium with the negative electrons that are drifting through the plasma to the plate. The plasma takes up most of the room between the cathode and plate, and is at a potential slightly higher than the plate potential because of a slight surplus of positive ions in this region. The electrons liberated from the cathode strike the edge of the plasma with sufficient energy to produce ionization and then drift toward the plate. Practically all the potential drop within the tube takes place in the region immediately adjacent to the cathode. The potential distribution in the tube for a condition corresponding to the steep portion of Fig. 52 is shown in Fig. 53.

The voltage drop corresponding to the steep portion of the characteristic of Fig. 52 is approximately the ionization potential of the gas involved. In the case of mercury this is 10 to 15 volts. This voltage drop between anode and cathode is substantially constant for currents from very small values up to the full emission current of the

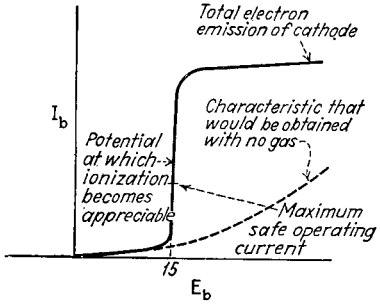


FIG. 52.—Current-voltage characteristics of a gas diode.

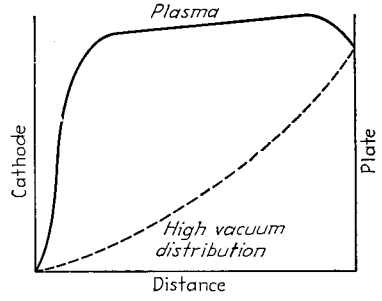


FIG. 53.—Potential distribution in a gas diode.

cathode. If there is an attempt, however, to pass more current through the tube than is represented by the cathode emission, the voltage drop will increase above the ionization potential of the gas involved. This is to be avoided because it causes cathode disintegration, and because the voltage drop in the tube, and hence the tube dissipation, is then unnecessarily large.

Study has shown that with oxide-coated and thoriated-tungsten emitters the positive ions striking the cathode in a gas-filled tube will not produce serious consequences, provided that the velocity of these positive ions does not exceed about 22 volts. However, if the velocity is greater than this, progressive disintegration of the emitting surface occurs. Damage will occur even when the voltage drop exceeds the permissible value for only a very small time. As a consequence, it is important that care be taken to arrange the circuit external to a gas tube so that the anode current in the tube is limited by the external circuit to a value safely below the cathode emission.

In the practical use of hot-cathode gas rectifier tubes, the factors that must be considered in selecting a particular tube are: (1) the allowable peak instantaneous space current; (2) the maximum average current that the tube is capable of handling; (3) the allowable peak inverse voltage. The allowable peak instantaneous current must be safely below the emitting possibilities of the cathode to ensure that at no time during life will the tube be subjected to operating conditions where the instantaneous voltage drop in the tube exceeds the ionization voltage. Every tube has a rated maximum average current that is appreciably less than the peak instantaneous current. This

rated average current should not be exceeded, since to do so will in general cause overheating of the tube and greatly shortened life. The maximum inverse voltage applied to the anode is important, because even though when the anode is negative no space current flows, there is the possibility that an arc-back will occur through the gas if the inverse voltage is too high.

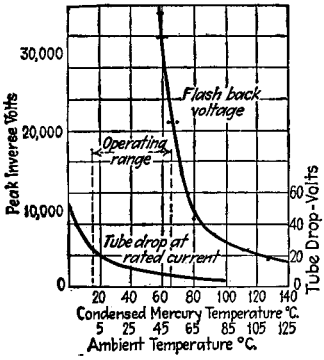


FIG. 54.—Inverse breakdown voltage and voltage-drop variation with temperature of a hot-cathode mercury-vapor diode.

An increase in the bulb temperature of a hot-cathode mercury-vapor tube causes the inverse voltage at which arc-back occurs to decrease, as indicated in Fig. 54. This same figure shows that the voltage drop in the tube increases as the temperature decreases. As a result of these characteristics, hot-cathode mercury-vapor rectifier tubes must be operated so that the temperature of the condensed mercury falls between certain limits. The highest temperature is determined by the minimum allowable inverse breakdown voltage, while the lowest allowable operating temperature is determined by the fact that the cathode disintegration potential must not be approached too closely, and the fact that the ionization in the tube is inadequate if the gas pressure (*i.e.*, temperature) is too low.

Because the current that a gas rectifier can pass is limited only by the cathode emission, and not by space charge, as in an ordinary tube, every effort is made to obtain as high an emission as possible per watt of cathode heating power. The nearly complete annulment of space charge resulting from the action of positive ions makes it possible to use cathode structures that cannot be used in vacuum tubes. Thus strongly curved surfaces, and even enclosures coated with emitting material,

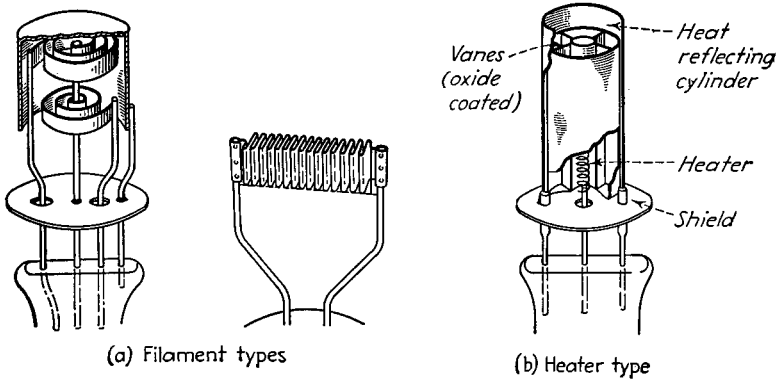


FIG. 55.—Hot-cathode emitting structures.

are permissible because the space-charge annulling positive ions are able to penetrate into these. This makes possible a higher thermal efficiency than is obtainable in vacuum tubes. Cathode structures may be either of the filamentary type or of the indirectly heated type. The filamentary type frequently makes use of a wide ribbon that is crimped and folded into a spiral to give a large emitting surface in a small volume. The indirectly heated type makes use of a heated cylinder placed within another cylinder, supported by vanes parallel to the axis. The entire interior of the

structure is coated with emitting material, and the emitted electrons escape through the ends of the structure. All types of structures sometimes have surrounding heat shields of thin sheet metal to reduce the radiation by reflecting the heat back into the structure. Some typical emitting structures are shown in Fig. 55.

It is necessary that the peak potential difference between the ends of the filament be not only less than the ionization potential of the gas, but also be less than the difference between the ionization potential and the disintegration potential. Otherwise ionization may be caused by the filament voltage, or ions may strike part of the emitter with energy in excess of that required to cause disintegration.

A compromise must be effected between heater efficiency and heating time, since the more efficient the cathode is thermally the longer it takes to reach its operating temperature. In all cases it is necessary to bring the cathode to normal operating temperature before applying the plate potential in order to avoid damaging the emitting surface. Heating times range from 30 seconds in small tubes to many minutes for the larger structures.

The nature of the current-voltage characteristics of gas tubes makes it virtually impossible to connect such tubes directly in parallel. This is because the operating part of the characteristic is so steep that a slight difference in the characteristics of two tubes would make a very large difference in the currents taken by the tubes, and the tube with the smaller voltage drop would take practically all the current. This difficulty can be circumvented by connecting resistors in series with the tubes before connecting them in parallel.

18. Grid-controlled Gas Rectifiers (Thyratrons).[1]—Grid-controlled gas rectifier tubes, also known as thyratrons and as gas triodes, are used extensively in industrial control circuit applications. The grid-controlled gas rectifier tube can be considered as a hot-cathode gas rectifier tube in which there has been inserted a control electrode between cathode and plate. This control electrode serves a function analogous to that of the control grid of an ordinary vacuum tube, although structurally it may differ.

The grid electrode is able only to initiate the flow of current, but once current flows the magnitude of the current is not influenced by the control electrode, and neither is this electrode able to stop the flow of current. The current can be stopped only by making the plate potential zero or negative for a short period of time.

The characteristics associated with the control grid are briefly as follows. If the grid is considerably more negative than the cutoff value (*i.e.*, the grid potential at which no electrons are attracted toward the anode), and then is gradually reduced, it is found that at the point where plate current would just start to flow if the tube contained no gas, the plate current suddenly jumps from zero to a high value. This is a result of the fact that as soon as plate current begins to flow, positive ions are produced as a result of ionization by collision. Some of these are attracted toward the negative grid and surround it with a sheath of positive ions that neutralize the electrostatic effect of the grid, and so destroy its normal control action. At the same time, other positive ions are attracted toward the cathode and neutralize the space charge. Hence once ionization has started, there is no space charge to limit the current flow, and the control action of the grid has been lost. The anode current then is determined by the same factors that affect anode current in the hot-cathode gas rectifier tube.

The relations between the grid and plate potential at the moment of the initiation of the arc are similar to those which apply at cutoff in a vacuum triode. As the plate potential is made less positive, the grid potential must be made relatively more positive to start the arc. The starting voltage characteristic of a typical thyratron is

[1] A. W. Hull, Gas-filled Thermionic Tubes, *Trans. A.I.E.E.*, Vol. 47, p. 753, 1928.

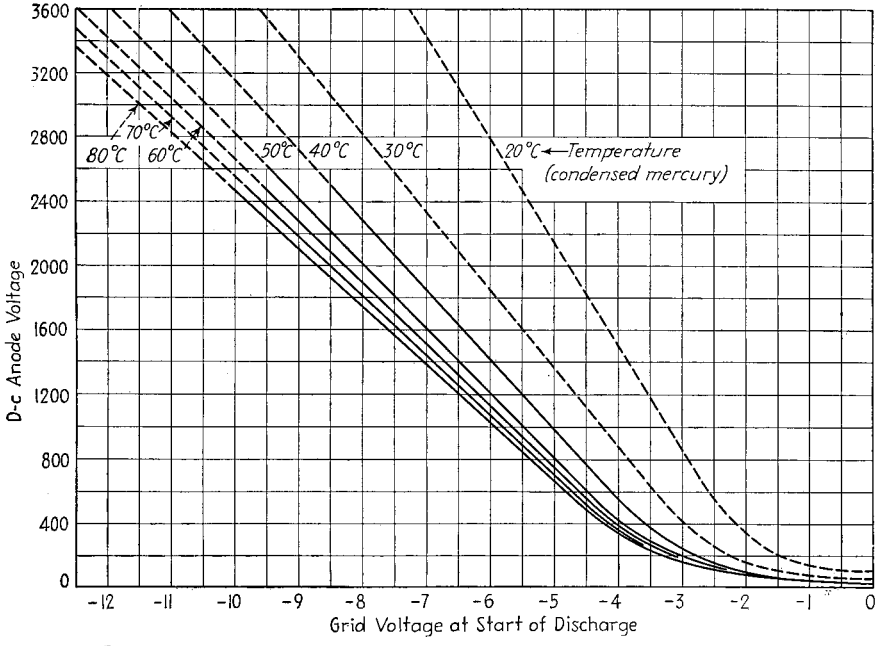


FIG. 56.—Starting characteristics of a typical thyratron using mercury.

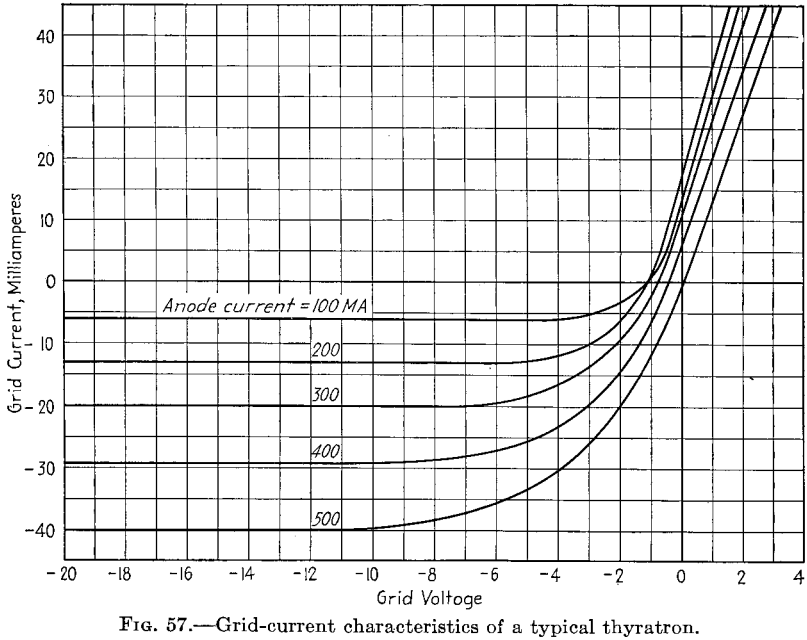


FIG. 57.—Grid-current characteristics of a typical thyratron.

shown in Fig. 56. Over a large part of the characteristic it is seen that the ratio of positive plate potential to negative grid potential is nearly constant. This ratio is known as the "control ratio."

Although the grid loses control of the current when the arc has started, it may draw an appreciable current. This current tends to be a large negative current in the form of collected electrons when the grid is positive, the magnitude of the current rising rapidly with potential. When the grid is negative the current corresponds to a positive-ion current and tends to be small and nearly constant. Because of the possibility of large grid currents the grid must be protected by a series resistor in circuit applications. The resistor cannot, however, be too large or the voltage drop in it due to the current that does flow will interfere with the normal grid action. Some typical grid-current characteristics are shown in Fig. 57.

The arc of a thyratron tube can only be extinguished by making the plate potential zero or negative for a sufficiently long period of time. This reduction of the plate potential stops the arc because it prevents electrons from reaching the plate and causes them to be drawn back to the cathode while the positive ions are drawn back to the plate. The plate must be negative long enough for most of the positive ions to

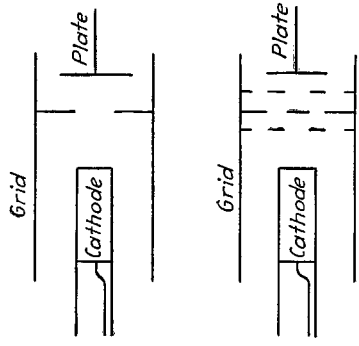


FIG. 58.—Typical thyratron electrode structures.

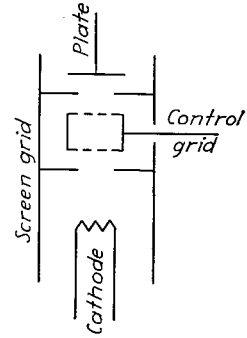


FIG. 59.—Screen-grid thyratron.

be removed from the space between electrodes, for otherwise the grid is unable to regain control of the arc. The time it takes for the grid to regain control of the arc after the plate potential has been made negative is known as the deionization time. Although the time required to initiate the arc, the ionization time, is of the order of a few microseconds, the deionization time ranges from ten to 1,000 microseconds. Hence thyratrons cannot be used in applications calling for a repetition rate of more than about fifty kilocycles.

The mode of operation of a thyratron tube makes it desirable to employ electrode structures and arrangements differing greatly from those of a vacuum triode. Cathodes are made similar to those used in gas rectifiers. The grid is frequently a nearly solid electrode surrounding the cathode, rather than a meshlike structure. Thus the usual "grid" is a cylinder having a diaphragm that separates the cathode from the plate, and that contains one or more large holes. The holes must be large enough so that they are not closed by the positive-ion sheath that forms about the grid after the initiation of the arc. The plate of the thyratron is usually just a simple disk that collects the electrons. It can be small because the low voltage drop in the tube results in a low plate heat dissipation. Some typical thyratron electrode structures are shown in Fig. 58.

Although the grid current in a typical gas triode may be only 50 microamperes or less, there are many applications in which even this much current prevents the use of a high-impedance device in the grid circuit. Grid currents may be reduced one hundredfold by the use of a screen-grid structure as shown in Fig. 59. In this the screen grid is placed next to the cathode, and the control grid is placed between the screen grid and plate. In this way the control-grid current is reduced to the point where the tube may be operated directly from a photocell. The other characteristics of the screen-grid thyratron are essentially the same as those of a three-electrode thyratron.

The limitations of a thyratron are similar to those of a gas rectifier. Thus there is a certain peak current that should not be exceeded even instantaneously lest the cathode be bombarded with ions of sufficient energy to disintegrate it. Also, since the heating of the tube depends upon the average current because of the constancy of the voltage drop in the tube, there is a maximum allowable value of average current.

The allowable inverse peak voltage depends upon the temperature of the tube and upon its internal construction. Since the pressure within the tube doubles with a rise of approximately 10°C, great care must be taken to see that the tube is not overheated.

The designer must observe a number of precautions in order to keep the rated peak inverse voltage high. The electrode geometry must be such that high potential gradients within the tube are avoided. Residual gases in the tube other than the desired gas must be kept low. The electrode shapes should be such that the ionized region is confined as much as possible. This is necessary so that the positive ions removed on the first part of the negative half of the plate-voltage cycle will only have a short distance to go, for otherwise they may strike the plate with sufficient energy to produce emission, which will cause a flashback.

19. Cold-cathode Tubes. *Diodes.*—The simplest cold-cathode tube is a diode containing neon or some other inert gas at a pressure of about 0.1 mm of mercury. There is no provision for control of the arc in a diode, and since the cathode is cold, the voltage drop is high, and the current density in the discharge is low. The voltage drop is determined by the gas pressure and the cathode material. The stability of the discharge is aided by sputtering the cathode with metals of low work function, such as cerium, lathanium, and didymium, or coating with partially reduced oxides of barium and strontium. No combination of gas and cathode material that gives a stable discharge at less than 60 volts has been found.

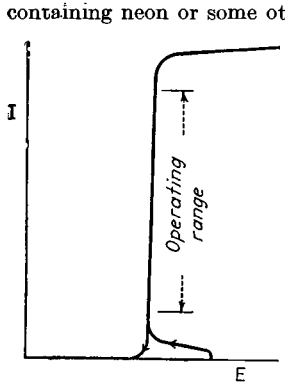


FIG. 60.—Current-voltage characteristics of a cold-cathode diode.

A small starting probe can be used as part of one electrode of the diode to give a high gradient region in which the discharge will start. As potential and current are raised the glow will spread from the starting probe until it covers the whole plate. Typical cold-cathode current-voltage characteristics are shown in Fig. 60. The striking voltage of the glow is a little higher than the normal tube drop. This is because an accumulative ionization will maintain the glow at a lower potential than that which is needed to initiate it. Cold-cathode diodes are operated in the part of the characteristic in which the voltage drop is substantially constant. If this voltage drop is exceeded the tube may be damaged because of excessive sputtering action at the cathode.

Characteristics of some typical cold-cathode diodes are given in Table 7. These tubes have all been designed to be used as voltage regulators.

TABLE 7.—CHARACTERISTICS OF VOLTAGE-REGULATOR GAS-DISCHARGE TUBES

Tube type	Maximum current, ma	Operating voltage, volts	Igniting or striking voltage
RCA 991	2	48–87	87
RCA 874	50	90	125
VR75-30	30	75	105
VR90-30	30	90	125
VR105-30	30	105	137
VR150-30	30	150	180
WE 313C	..	60	70
¼-watt neon	5	50–60	85
3-watt neon	60	50–60	85

In the *VR* series of tubes there is a correct polarity for the applied d-c voltage. Some neon lamps have current-limiting resistors in the base that must be removed for voltage-regulator application. The WE 313C has the greatest uniformity of characteristics from tube to tube; the neon lamps and the RCA 991 have the least uniform characteristics from tube to tube.

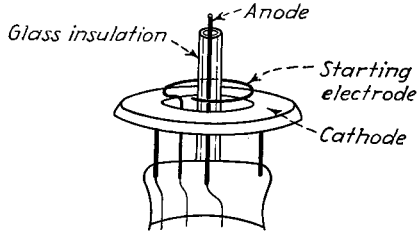


FIG. 61.—A cold-cathode discharge tube with starter electrode.

The principal uses of cold-cathode diodes are as voltage regulators, stroboscopic light sources, relaxation oscillators, polarity indicators, and surge-protection devices.

Cold-cathode Triodes.—A control type of cold-cathode tube has been developed for use as a trigger tube, especially in carrier actuated circuits.[1] This tube makes use of two fundamental principles. First, the breakdown potential between two electrodes depends upon their spacing, and second, the breakdown potential is lowered when ionization is present. Ionization is provided between cathode and a starter electrode placed close to it. The starter electrode is operated with a high resistance in series with it so that it carries little current, but the arc to it starts the main discharge to the anode, which is operated just below the starting potential corresponding to the cathode-anode spacing.

The construction of such a tube is shown in Fig. 61. The cathode is a large flat disk electrode of nickel. The control or starter electrode is a wire ring close to the cathode. The anode is a short length of wire projecting from a glass tube, which, in turn, projects through the cathode and starter electrode. An alternative construction that is also used in practice is similar, except that the cathode and starter electrodes are identical halves of a split disk. In either case, the cathode is coated with

[1] W. E. Bahls and C. H. Thomas, A New Gas-filled Triode, *Electronics*, Vol. 11, p. 14, May, 1938; S. B. Ingrahm, Cold-cathode Gas-filled Tubes as Circuit Elements, *Trans. A.I.E.E.*, Vol. 58, p. 342, July, 1939; C. N. Kimball, A New System of Remote Control, *R.C.A. Rev.*, Vol. 2, p. 303, January, 1938.

oxides of barium and strontium that are partially reduced to the metals in the evacuation process.

Since a sputtering action is present while the tube is in operation, the metal is gradually used up and replaced by a further reduction of the oxide.[1] The tube accordingly has a definite life, which is determined by the rate at which the metal is used up, and so upon the time integral of the anode current. Cold-cathode triodes

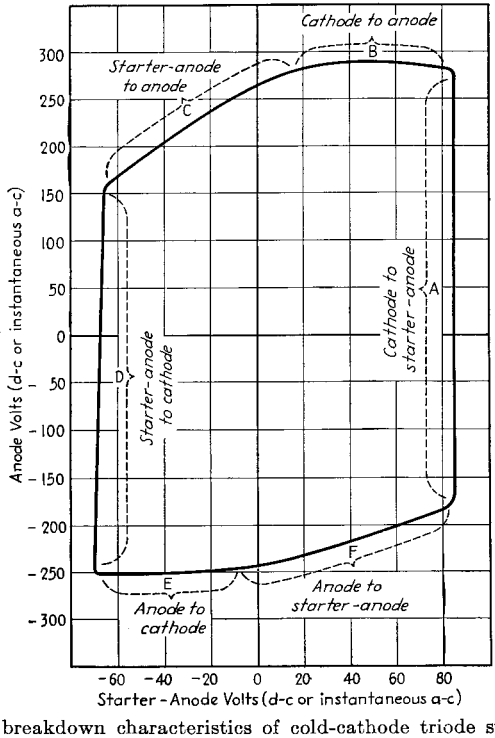


FIG. 62.—Typical breakdown characteristics of cold-cathode triode such as is shown in Fig. 61.

are designed to operate at low currents, and when possible the circuit arrangements should be such that the anode current flows as little as possible.

Since there are three electrodes in a tube such as that of Fig. 61, any one electrode may act as the control electrode to control the discharge between the other two. As a result there are six possible conditions of current flow, the breakdown characteristics of which are indicated by the closed hexagonal curve of Fig. 62. Ordinarily only the section *A* of the curve is used.

The chief uses of the cold-cathode triode are as a relay device, or trigger tube.

[1] The factors controlling life in cold-cathode tubes are well treated by G. H. Rockwood, Current Rating and Life of Cold-cathode Tubes, *Trans. A.I.E.E.*, Vol. 60, p. 901, September, 1941.

SECTION 5

VACUUM-TUBE AMPLIFIERS

GENERAL CONCEPTS

1. Classification of Amplifiers.—Amplifiers are classified in a variety of ways descriptive of their character and properties. One basis is according to the frequency range covered and leads to audio-frequency, video-frequency, radio-frequency, and direct-current amplifiers. Amplifiers are also classified as voltage or power amplifiers, according to whether the primary purpose is to develop voltage or power, respectively, in the output.

Amplifiers are also often referred to as Class A, Class AB, Class B, linear, or Class C amplifiers, according to the method of adjusting the tube. A Class A amplifier is an amplifier in which the grid-bias and alternating grid voltages are such that plate current in the tube flows at all times. Again a Class AB amplifier is an amplifier in which the grid-bias and alternating grid voltages are such that plate current in the tube flows for appreciably more than half but less than the entire electrical cycle. A Class B amplifier is an amplifier in which the grid bias is approximately equal to the cutoff value so that the plate current is approximately zero when no exciting grid voltage is applied and so that plate current flows in the tube for approximately one-half of each cycle when an alternating grid voltage is applied. Class B amplifiers employing tuned load circuits are often referred to as *linear amplifiers*. A Class C amplifier is an amplifier in which the grid bias is appreciably greater than the cutoff value, so that the plate current in each tube is zero when no alternating grid voltage is applied and so that plate current in the tube flows for appreciably less than one-half of each cycle when an alternating grid voltage is applied.[1] The suffix 1 is sometimes added to indicate that grid current does not flow during any part of the cycle of input voltage, while the suffix 2 indicates that grid current flows during some part of the cycle; thus Class A_2 means a Class A amplifier with sufficient exciting voltage to drive the grid positive.

2. Distortion in Amplifiers.—The output of an amplifier may fail to reproduce the input as a result of frequency, amplitude (or nonlinear), and phase distortion. Frequency distortion occurs when different frequency components are not amplified equally. Amplitude (or nonlinear) distortion is the result of nonlinearity existing in either the grid-cathode or plate-cathode circuits of the tube, and causes frequencies to appear in the amplifier output that are not present in the voltage applied to the amplifier input.

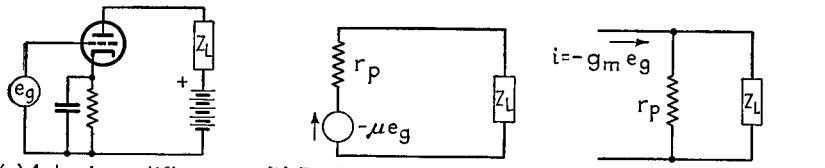
Phase Distortion and Delay Time.—Phase distortion occurs when the relative phases of the various components being amplified are not the same in the output as in the input. Phase distortion causes the output wave shape to differ from the wave shape of the applied signal, even though both may contain exactly the same frequency components in the same relative magnitudes. The difference in phase between the output and input voltage for an individual frequency component being amplified can be expressed in terms of a time of transmission through the amplifier according to the relation

$$\text{Phase shift in radians} = \tau\omega + n\pi \tag{1}$$

[1] "Standards on Electronics," p. 7, Institute of Radio Engineers, 1938.

where ω is 2π times frequency, n is an integer, and τ a quantity having the dimension of time and termed the *delay time*. In order to avoid phase distortion, τ must be the same for all frequencies; *i.e.*, the curve of phase shift as a function of frequency must be a straight line passing through an integral multiple of π at zero frequency. Any departure from such a constant delay time represents phase distortion.

3. Equivalent Amplifier Circuits.—The fundamental circuit of a vacuum-tube amplifier is shown in Fig. 1*a*. To a first-order approximation, the amplified current that flows in the plate circuit of such an amplifier tube having an amplification factor μ and transconductance g_m can be calculated by replacing the plate-cathode circuit of the tube by an equivalent generator voltage μe_g acting in series with the plate resistance of the tube. This arrangement can be referred to as the *constant-voltage generator* form by the equivalent plate circuit, and is shown in Fig. 1*b*. An alternative equivalent circuit is shown in Fig. 1*c*, in which the plate-cathode circuit of the tube is replaced by a constant-current generator supplying a current $e_g g_m$ to a circuit consist-



(a) Actual amplifier circuit (b) Equivalent amplifier circuit (constant voltage generator form (c) Equivalent amplifier circuit (constant current generator form)

FIG. 1.—Actual and equivalent plate circuits of a vacuum-tube amplifier. In the constant-voltage form the current that is produced in the plate circuit by the signal e_g acting on the grid is taken into account by postulating that the plate circuit can be replaced by an equivalent generator $-\mu e_g$ having an internal resistance equal to the plate resistance R_p. In the constant-current form the tube is considered as generating a current $-g_m e_g$ that flows through the impedance formed by the plate resistance of the tube in parallel with the load resistance. These circuits apply to all types of tubes, including pentode, beam, and screen-grid tubes as well as triodes.

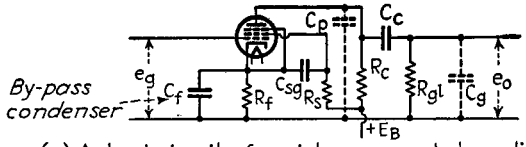
ing of the plate resistance of the tube in parallel with the load resistance. This can be referred to as the *constant-current generator* form of the equivalent amplifier circuit. It expresses the same relation as the constant-voltage form, and leads to the same result as far as the load is concerned.

The constant-current circuit is the most convenient to use in practical calculations when plate resistance of the tube is much higher than the load resistance, as is the case with pentode and beam tubes. The constant-voltage form is most convenient when the plate resistance is of the same order of magnitude or less than the load resistance, as is commonly the case with amplifiers using triode tubes.

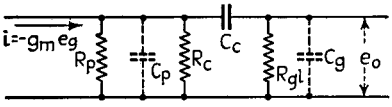
VOLTAGE AMPLIFIERS FOR AUDIO FREQUENCIES

4. Resistance-coupled Audio-frequency Amplifiers.—The ordinary voltage amplifier of audio frequencies employs a pentode tube, in the plate circuit of which is a load impedance consisting of a resistance-capacity network, as shown in Fig. 2.

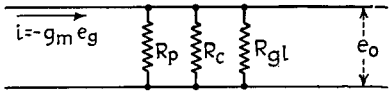
Amplification as a Function of Frequency When the Screen and Cathode By-pass Condensers Are Very Large.—When the impedance in the screen and cathode circuits of a resistance-coupled amplifier can be neglected the amplification varies with frequency as shown in Fig. 3. The amplification at high frequencies falls off because of the shunting capacity represented by the plate-cathode capacity of the tube plus the input capacity of the tube to which the amplified voltage is applied, plus incidental



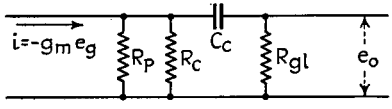
(a) Actual circuit of resistance-coupled amplifier



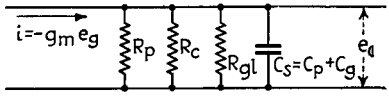
(b) Exact equivalent circuit of amplifier



(c) Simplified circuit accurate for intermediate frequencies



(d) Simplified circuit accurate for low frequencies



(e) Simplified circuit accurate for high-frequencies

e_g = signal voltage
e_o = amplified voltage
R_p = tube plate resistance
R_c = coupling resistance
R_{gl} = grid-leak resistance
R_f = bias resistance
R_s = voltage dropping resistor for screen
g_m = mutual conductance of tube
C_f = bias by-pass condenser
C_{sg} = screen by-pass condenser

C_c = coupling or blocking condenser
C_p = plate-cathode tube capacity plus stray wiring capacity to left of coupling condenser
C_g = stray wiring capacity to right of coupling condenser plus input capacity of tube or other load to which voltage is delivered
$C_s = C_p + C_g$ = total shunting capacity

FIG. 2.—Circuit of a resistance-coupled amplifier employing a pentode tube, together with equivalent plate circuits useful in making amplifier calculations, assuming large screen and bias by-pass condensers.

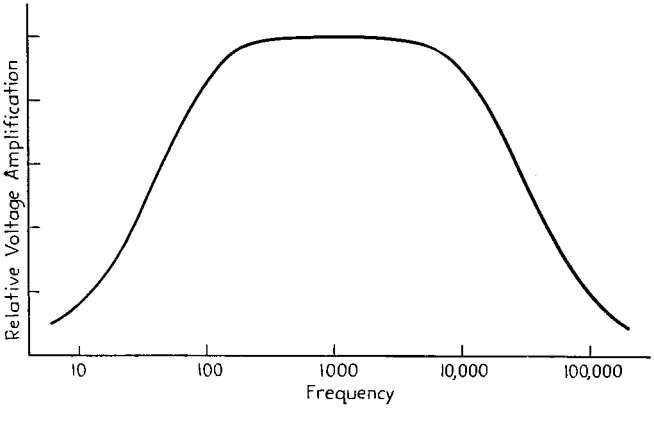


FIG. 3.—Variation of amplification with frequency in a typical resistance-coupled amplifier with large screen and bias by-pass condensers.

capacity from the wiring. The falling off at low frequencies is caused by the loss of voltage in the coupling condenser C_c.

The amplification characteristic can be calculated with the aid of the equivalent plate circuit of Fig. 2*b*. This can be simplified in ordinary amplifiers by taking advantage of the fact that the coupling condenser C_c has negligible effect except at quite low frequencies, while the shunting capacities have negligible effect except at high frequencies.[1] This results in the equivalent circuits of Figs. 2*c* to 2*e*, each of which is applicable to a limited part of the frequency range.

In the middle range of frequencies, the amplification is maximum and depends only upon the transconductance of the tube and the equivalent resistance R_{eq} formed by plate, coupling, and grid-leak resistances, all in parallel, according to the equation

$$\text{Mid-frequency amplification} = g_m R_{eq} \tag{2}$$

where $R_{eq} = \dfrac{R_c}{1 + \dfrac{R_c}{R_{gl}} + \dfrac{R_c}{R_p}}$.

g_m = transconductance of amplifier tube, mhos.

With pentode tubes the plate resistance is very much higher than the remaining resistances involved in R_{eq}, so that the amplification is given with good accuracy by the simplified equation

$$\left.\begin{matrix}\text{Mid-frequency amplification (approximate)}\end{matrix}\right\} = g_m R_c \frac{1}{1 + \dfrac{R_c}{R_{gl}}} \tag{3}$$

In this middle-frequency range there is negligible phase shift other than a phase reversal produced as a result of the way in which the tube operates.

At low frequencies the amplification is

$$\frac{\left.\begin{matrix}\text{Amplification at low frequencies}\end{matrix}\right\}}{\left.\begin{matrix}\text{Amplification in middle range}\end{matrix}\right\}} = \frac{1}{1 - j\dfrac{f_1}{f}} \tag{4}$$

where f = actual frequency.

$$f_1 = 1/2\pi C_c R \tag{5}$$

= frequency at which the reactance of C_c equals the resistance $R\left(= R_{gl} + \dfrac{R_c}{1 + \dfrac{R_c}{R_p}}\right)$ formed by grid leak in series with the combination of plate and coupling resistances in parallel.

Under all ordinary conditions with pentodes, $R_p >> R_c$, and it is permissible to take $R = R_c + R_{gl}$. *The extent to which the amplification falls off at low frequencies is determined by the ratio of the coupling condenser reactance to the equivalent resistance R formed by combining the grid leak in series with the parallel combination of coupling resistance and plate resistance. The output falls to* 70.7 *per cent* (−3 *db*) *of the mid-frequency amplification at the frequency* f_1 *for which this reactance equals the equivalent resistance R.* The falling off in amplification and the phase shift produced at low frequencies relative to the mid-frequency conditions are given in Fig. 4.

[1] Analysis of the rather infrequently encountered case where this does not hold true is given by David C. G. Luck, A Simplified General Method for Resistance-capacity Coupled Amplifier Design, *Proc. I.R.E.*, Vol. 20, p. 1401, August, 1932; Westley F. Curtis, The Limitations of Resistance-coupled Amplification, *Proc. I.R.E.*, Vol. 24, p. 1230, September, 1936.

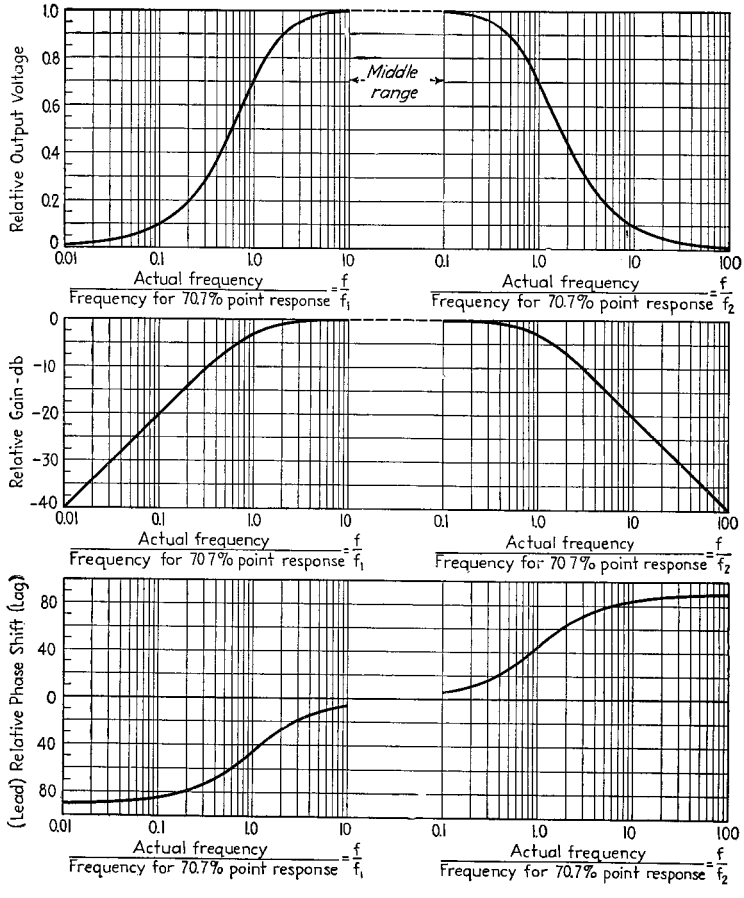


FIG. 4.—Universal amplification curve showing the falling off in amplification and the phase shift that occurs in a resistance-coupled amplifier at high and low frequencies when the screen and bias by-pass condensers have negligible effect.

The relative amplification at high frequencies is

$$\frac{\left.\begin{matrix}\text{Amplification at}\\ \text{high frequencies}\end{matrix}\right\}}{\left.\begin{matrix}\text{Amplification in}\\ \text{middle range}\end{matrix}\right\}} = \frac{1}{1 + j\frac{f}{f_2}} \tag{6}$$

where f = actual frequency.

$f_2 = 1/2\pi C_s R_{eq}$ (7)

= frequency at which the reactance of the shunting condenser C_s equals the equivalent resistance R_{eq} formed by plate, grid-leak, and coupling resistances, all in parallel.

The extent to which the amplification falls off at high frequencies is determined by the ratio that the reactance of the shunting capacity C_s bears to the equivalent resistance obtained by combining coupling resistance, grid-leak resistance, and plate resistance all in parallel. The loss in amplification at high frequencies can be conveniently estimated by the fact

that the magnitude is **70.7** *per cent* (−3 *db*) *of the mid-frequency amplification at the frequency* f_2 *for which the reactance of the shunting capacity equals the equivalent resistance* R_{eq}. The relative amplification and phase shift are given in Fig. 4.

Effect of Impedance in the Screen-grid Circuit.—Resistance-coupled amplifiers employing pentode tubes normally obtain the screen voltage from the plate-supply voltage by means of a voltage dropping resistance R_s by-passed to the cathode by condenser C_{sg}, as shown in Fig. 2*a*. Such an arrangement introduces an impedance in the screen-grid circuit that modifies the amplification, particularly at low frequencies, where the impedance is large. The effect is to reduce the output voltage, and shift its phase by the factor[1,2]

$$\left.\begin{array}{l}\text{Actual output voltage}\\ \hline \text{Output voltage with zero impedance in screen circuit}\end{array}\right\} = \beta = \frac{\dfrac{K}{1+K} + j\dfrac{f}{f_3}}{1 + j\dfrac{f}{f_3}} \tag{8}$$

where $K = \dfrac{R_{sg}}{R_s} = \dfrac{\text{dynamic resistance of screen circuit of tube}}{\text{screen dropping resistance}}$.

f = actual frequency.

$f_3 = \dfrac{1+K}{2\pi C_{sg} R_{sg}}$

= frequency at which reactance of C_{sg} equals resistance formed by R_s and R_{sg} in parallel. (9)

The magnitude and phase of the factor given by Eq. (8) are plotted in Fig. 5.

Effect of Bias Impedance.—The usual voltage amplifier obtains the grid bias by means of a resistance-condenser combination between cathode and ground, as shown by $C_f R_f$ in Fig. 2*a*. This arrangement introduces an appreciable impedance in the cathode circuit at low frequencies, and since the amplified plate current must flow through this impedance, a voltage drop is produced that causes an additional voltage to be applied between grid and cathode. This reduces the amplification and also introduces an additional phase shift. The magnitude of the effect is given by[3]

$$\left.\begin{array}{l}\text{Actual output voltage}\\ \hline \text{Output voltage with zero bias impedance}\end{array}\right\} = \gamma = \frac{1}{1 + \dfrac{g_m R_f \beta}{1 + j\dfrac{f}{f_4}}} = \frac{1}{1 + g_m R_f \eta \beta} \tag{10}$$

where β = factor given by Eq. (8) and Fig. 5.

$$\eta = \frac{1}{1 + \dfrac{f}{f_4}}$$

g_m = transconductance of the tube.

R_f = bias resistance.

f = actual frequency.

$f_4 = 1/2\pi C_f R_f$ = frequency at which reactance of C_f equals bias resistance R_f.

[1] This equation assumes that the pentode tube is operated under conditions such that the division of the total space current between the plate and screen is not affected by the plate voltage, a condition realized under ordinary conditions. For further details, see F. E. Terman, W. R. Hewlett, C. W. Palmer, W. Y. Pan, Calculation and Design of Resistance-coupled Amplifiers Using Pentode Tubes, *Trans. A.I.E.E.*, Vol. 59, p. 879, 1940.

[2] Values of R_{sg} are not yet given by the tube manufacturers, but can be approximated with sufficient accuracy for most occasions by assuming the ratio of screen resistance to plate resistance obtained when the tube is converted to a triode by connecting plate and screen together at the screen potential, is the ratio of total space current to screen current (usually about five).

[3] See Terman, Hewlett, Palmer, and Pan, *loc. cit.*

Equation (10) assumes that the a-c current flowing in the screen circuit all flows to the cathode through condenser C_{sg}, a condition closely realized in practice.

It is to be noted that the effect of the bias impedance in reducing amplification is affected not only by the bias impedance itself but also by the impedance in the screen circuit. This is in addition to the effect of the screen circuit impedance as given by Eq. (8) and Fig. 5.

In the case where the screen circuit impedance has negligible effect (*i.e.*, when $\beta \cong 1/\underline{0}°$), the effect of the bias impedance is given by Fig. 6. Results in the more general case can be readily obtained by the use of Figs. 5, 7, and 8. The first of these figures gives the factor β, and the second gives the factor η. From these one can

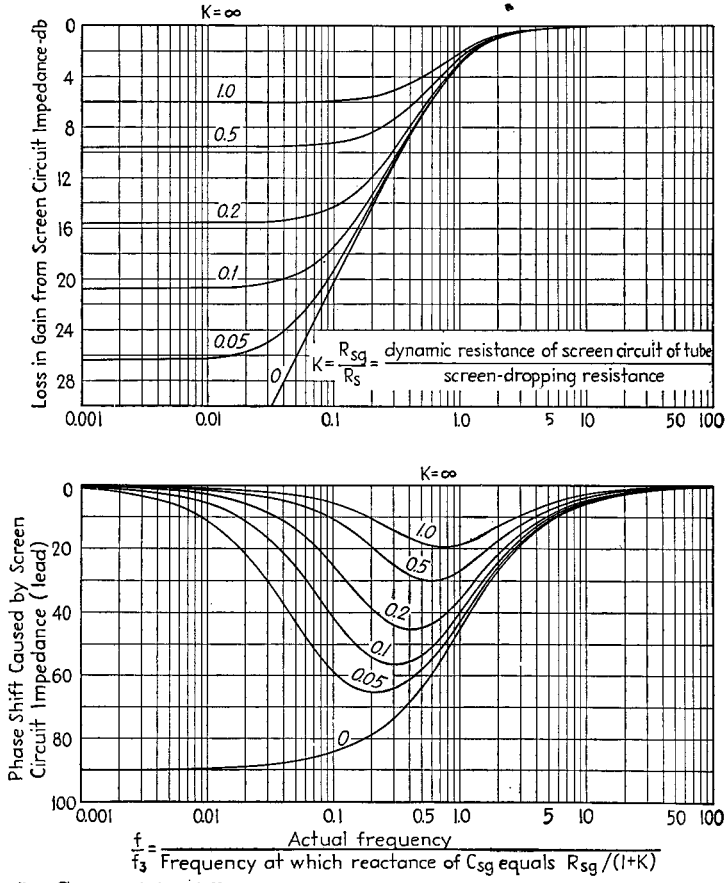


FIG. 5.—Curves giving falling off and phase shift in output voltage at low frequencies in a resistance-coupled amplifier as a result of a finite screen-cathode by-pass condenser, for various values of K.

form $R_f g_m \eta\beta$, then use Fig. 8 to give $\gamma = 1/(1 + R_f g_m \eta\beta)$. Note that the magnitudes in all these figures are given in db below unity voltage ratio, so that in forming $R_f g_m \eta\beta$, one simply adds the corresponding db values from Figs. 5 and 7 and the db value of $R_f g_m$ above unity (20 log $(R_f g_m)$ = db) to get the db magnitude $(R_f g_m)\eta\beta$, and adds angles of β and η to get the angle of $(R_f g_m)\eta\beta$.

Calculation of Actual Amplification Curve.—The actual amplification curve of an ordinary resistance-coupled amplifier such as is shown in Fig. 9 can be readily calculated by using Eq. (2) and Figs. 4, 5, 7, and 8, as illustrated by the following example:

Calculation of the Response and Phase Shift of the Amplifier Having the Constants of Fig. 9, as a Function of Frequency.—The amplification in the mid-frequency range is, according to Eq. (2),

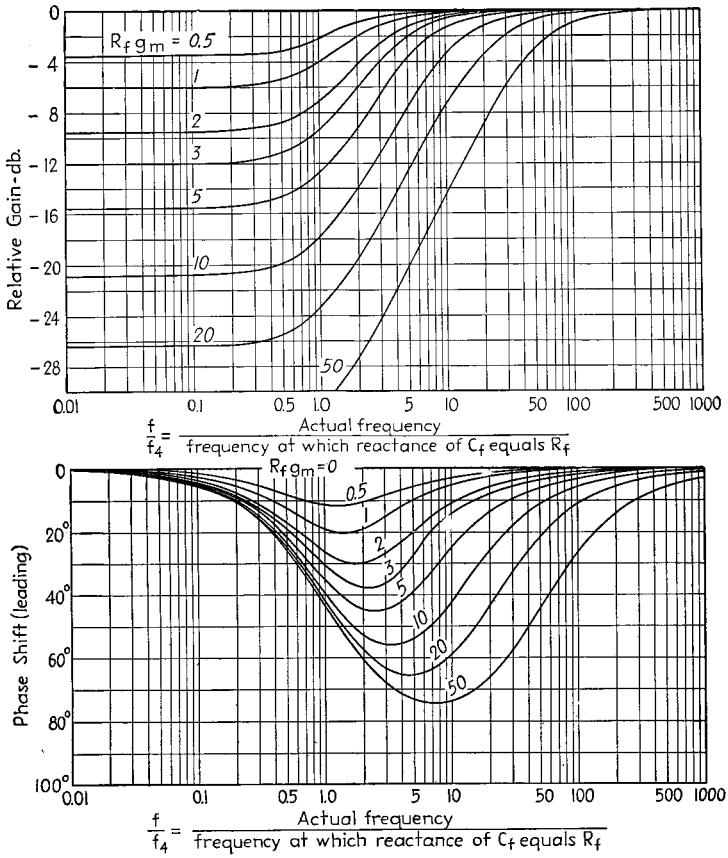


FIG. 6.—Curves giving loss in amplification and phase shift of output voltage occurring at low frequencies in a resistance-coupled amplifier as a result of a resistance-condenser bias impedance $R_f C_f$ for the case where the screen-circuit impedance has negligible effect upon the amplification, for various values of $R_f g_m$.

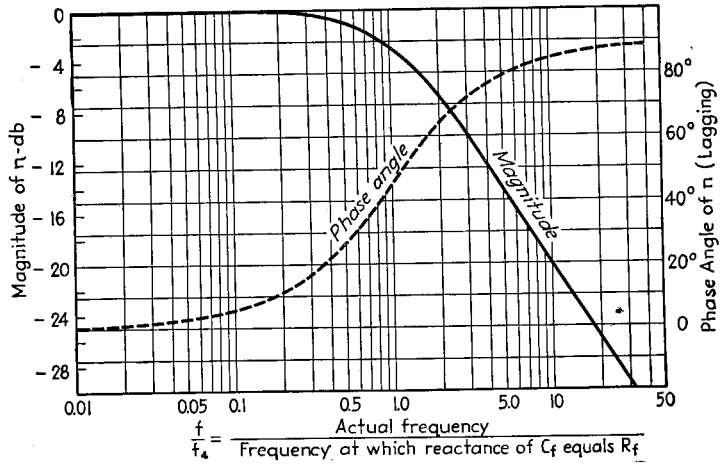


FIG. 7.—Magnitude and phase of factor η for use in Fig. 8.

$644 \times 10^{-6} \times 250{,}000 = 161$. The -3 db point on the high-frequency response occurs at a frequency $f_2 = 1/(2\pi \times 20 \times 10^{-12} \times 250{,}000) = 32{,}000$ cycles; the corresponding point f_1 at low frequencies, assuming perfect by-passing, is $f_1 = 1/(2\pi \times 0.0025 \times 10^{-6} \times 10^6) = 64$ cycles. The response and phase-shift characteristic for the case of perfect by-passing is then readily obtained with the aid of Fig. 4, and is indicated by curve *a* of Fig. 9.

Considering next the modifying effect of the impedance in the screen-grid circuit, one has $K = 125{,}000/2{,}400{,}000 = 0.052$, $f_3 = 1.052/(2\pi \times 0.04 \times 10^{-6} \times 125{,}000) = 34$ cycles. The loss

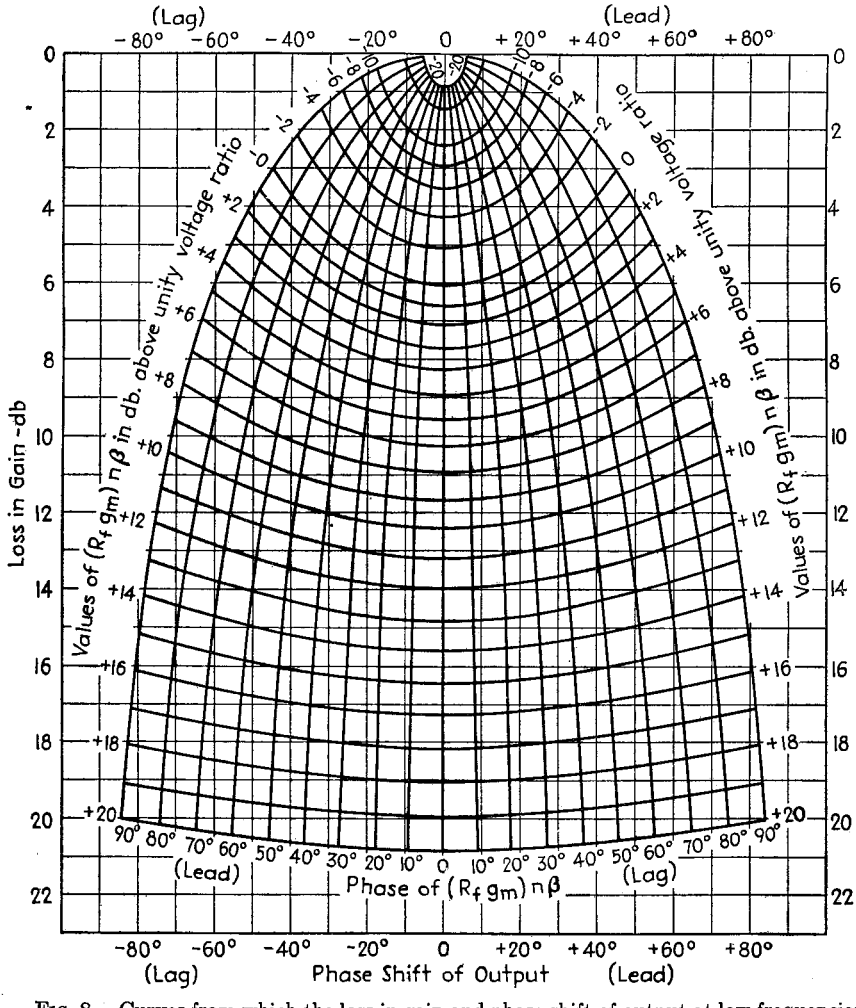


FIG. 8.—Curves from which the loss in gain and phase shift of output at low frequencies resulting from resistance-condenser bias impedance $R_f C_f$ can be calculated in a resistance-coupled amplifier for the general case where both screen and bias impedances are of importance.

of gain and phase shift that is caused by the impedance in the screen circuit is now readily obtained from Fig. 5 and gives the results shown by curve *b* of Fig. 9. Finally considering the bias impedance, $f_4 = 1/(2\pi \times 4.2 \times 10^{-6} \times 1{,}700) = 22$ cycles, and $R_f g_m = 1{,}700 \times 644 \times 10^{-6} = 1{,}095$, or $20 \log_{10} 1.095 = 0.8$ db above unity. The corresponding effects are now obtained by using Figs. 5, 7, and 8, and are shown by curve *c* in Fig. 9.

The total falling off and phase shift in the amplification at low frequencies are then the sum *d* of curves *a*, *b*, and *c* of Fig. 9.

Nonlinear Distortion.—The output voltage obtainable from a resistance-coupled amplifier is limited by the fact that amplitude distortion is proportionately greater as the applied signal is increased. In particular, the distortion increases very rapidly when the output exceeds a critical value determined by the conditions of operation. Avoidance of excessive amplitude distortion requires that the instantaneous plate current never reach zero during the negative half cycle of the applied voltage and that, similarly, on the positive half cycle the instantaneous potential at the plate of the

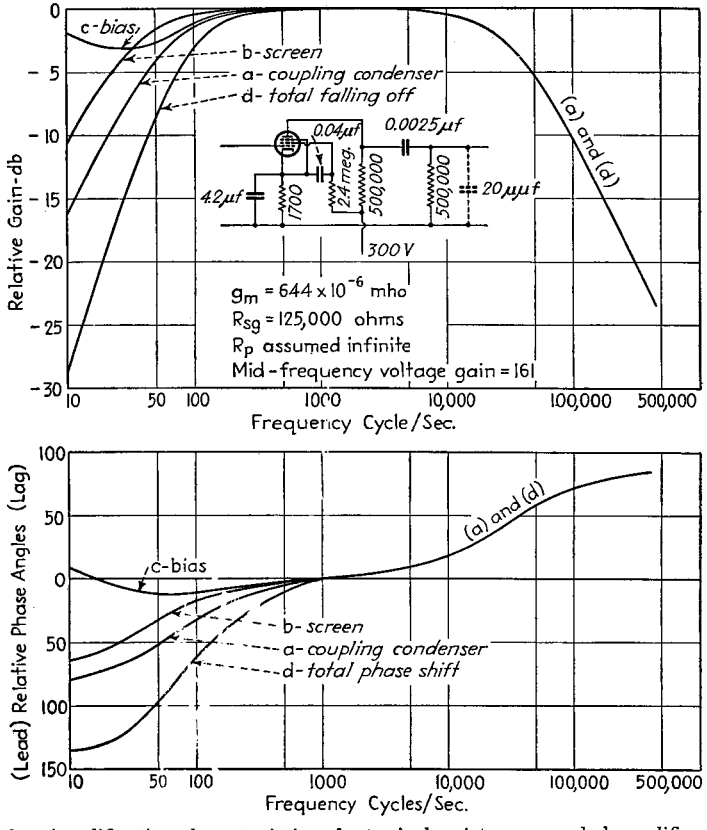


FIG. 9.—Amplification characteristics of a typical resistance-coupled amplifier, showing how the coupling condenser and the bias and screen impedances contribute to the falling off and phase shift of the output voltage at low frequencies.

tube never drops low enough to permit the formation of a virtual cathode in the vicinity of the suppressor grid. It is also necessary that the grid bias be sufficiently negative so that grid current will not be drawn at the positive peak of the applied signal. In terms of operating conditions, these requirements mean that (1) the peak alternating current I_{ac} that is available to develop voltage across the coupling resistance R_c and grid resistance R_{gl} is slightly less than the direct plate current I_{dc}; (2) the minimum potential $E_b - I_{dc}R_c - I_{ac}R_L$ reached by the plate during the cycle must exceed the plate potential at which a virtual cathode forms with a space current of $I_{dc} + I_{ac}$, where E_b is the plate-supply voltage and R_L is the load resistance formed by R_c and R_{gl} in parallel; (3) the bias voltage must exceed the peak signal amplitude by at least

1 volt. If any of these restrictions are violated, flattening of one peak or both peaks of the output wave will occur, with consequent excessive distortion.

Methods of analyzing the distortion obtained in a particular case are given in Par. 24.

Design Procedure for Obtaining Maximum Gain in Ordinary Resistance-coupled Amplifiers.[1]—In the design of an ordinary resistance-coupled amplifier, it is necessary to compromise between the desirability, on the one hand, of high gain and, on the other hand, of being able to obtain a large output voltage without excessive distortion with a reasonably high gain. It is accordingly necessary to distinguish between designs in which the objective is maximum amplification over the required frequency range without consideration being given to the amount of undistorted output voltage available and designs in which the primary object is to obtain a large output voltage without excessive distortion.

The relations governing the amplification of a resistance-coupled amplifier using a pentode tube are incorporated in the equation

$$\text{Mid-range amplification} = nK^{\frac{1}{n}}E_d^{\frac{n-1}{n}}R_{gl}^{\frac{1}{n}}\frac{\left(\frac{R_c}{R_{gl}}\right)^{\frac{1}{n}}}{1+\frac{R_c}{R_{gl}}} \tag{12}$$

where n and K are constants in the equation $i_p = K\left(E_g + \frac{E_{sg}}{\mu_{sg}}\right)^n$, E_d is the voltage drop of d-c plate current in coupling resistance R_c, and R_{gl} and R_c are the grid-leak and coupling resistances, respectively. Equation (12) assumes that the plate resistance can be taken as infinity and that the plate voltage is great enough to prevent the formation of a virtual cathode near the suppressor. These conditions are satisfactorily realized with normal operating conditions.

From Eq. (12) it is apparent that with a given plate-supply voltage and coupling resistance, the amplification will be maximum when the largest possible fraction of the supply voltage is used up as drop in the coupling resistance. Also it can be deduced from Eq. (12) that with a given E_d and grid-leak resistance R_{gl}, the voltage gain is maximum when

$$R_p = \frac{R_{gl}}{n-1} \tag{12a}$$

where n is the exponent of the space current equation $i_p = K\left(E_g + \frac{E_{sg}}{\mu_{sg}}\right)^n$. Under practical conditions this exponent ranges from about 1.5 for operation with rated plate current to something over 2 when the plate current is quite small. The coupling resistance for maximum gain is accordingly approximately equal to the grid-leak resistance, with the gain not being very critical with respect to the ratio R_c/R_{gl}.

Design for maximum amplification in an ordinary resistance-coupled amplifier is carried out as follows: The grid-leak resistance is first chosen as the highest resistance that can be safely placed in the grid circuit of the following tube (see Sec. 4, Par. 13). The coupling resistance R_c (see Fig. 2) is then made equal to the grid-leak resistance, and the high frequency f_2 giving 70.7 per cent response in Eq. (6) calculated on the basis of an estimated shunting capacity. If this frequency is lower than desired, the coupling resistance R_c should be reduced to the value that will give the desired characteristic, and the resulting loss in gain accepted as the price required to obtain the desired high-frequency characteristic. The coupling condenser C_c is now chosen to give the desired low-frequency response [*i.e.*, a suitable value of f_1 in Eq. (4)]. The

[1] See Terman, Hewlett, Palmer, and Pan, *loc. cit.*

low-frequency response should never be made any better than necessary, since otherwise trouble from motorboating may occur. A convenient grid bias, normally a few volts,[1] is now chosen, and a screen potential is then selected such that with the preceding coupling resistance, the potential at the plate of the tube will be approximately 20 per cent of the plate-supply voltage. These electrode voltages determine the screen and space current, which in turn fix the bias resistor R_f and screen dropping resistor R_s required. The screen by-pass condenser C_{sg} is next selected so that f_3 in Eq. (8) coincides with f_1, provided a sharp low-frequency cutoff is desired. If, on the other hand, the low-frequency cutoff should not be sharp, or if the phase shift at low frequencies is to be minimized, then C_{sg} should be large so that $f_3 << f_1$. Finally, the cathode by-pass condenser C_f is selected so that f_4 in Eq. (9) is not greater than f_3, provided sharp cutoff is desired, or so that $f_4 << f_1$ if minimum phase shift at low frequencies is important.

The foregoing procedure gives the maximum possible gain combined with the required high- and low-frequency response characteristics, either with or without sharp low-frequency cutoff as required. The output voltage is, however, limited to relatively small values because of the low d-c potential at the plate of the tube.

Design Procedure for Obtaining Maximum Output Voltage in Ordinary Resistance-coupled Amplifiers.—When it is important to obtain a large output voltage without excessive distortion, the design for maximum gain should be modified so that the coupling resistance does not exceed the grid-leak resistance, and the bias and screen potentials should be readjusted to reduce the plate current so that the direct-current voltage drop E_d in the coupling resistance R_c is approximately

$$\left.\begin{matrix}\text{Voltage drop in} \\ \text{coupling resistance}\end{matrix}\right\} = E_d = (1 - r)\left(\frac{1 + \frac{R_c}{R_{gl}}}{2 + \frac{R_c}{R_{gl}}}\right) E_B \tag{13}$$

$$\text{where } r = \frac{\left\{\begin{matrix}\text{lowest plate potential at which plate current} \\ \text{is independent of plate voltage when plate} \\ \text{current is twice the d-c value}\end{matrix}\right\}}{E_B}$$

The corresponding peak output voltage obtainable without excessive distortion is approximately

$$\left.\begin{matrix}\text{Peak a-c output} \\ \text{voltage}\end{matrix}\right\} = \frac{(1 - r)}{2 + \frac{R_c}{R_{gl}}} E_B \tag{14}$$

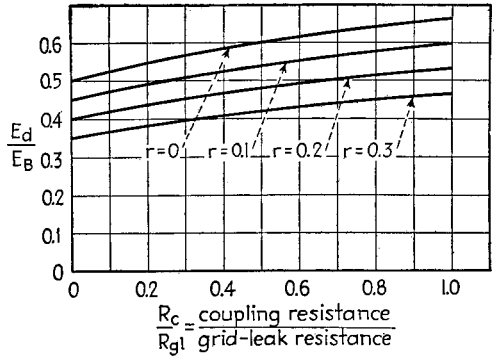
The output voltage obtainable is seen to be proportional to E_B and to decrease with R_c/R_{gl}.

The practical procedure for obtaining maximum output voltage is to start with the highest permissible value of R_{gl} and select a coupling resistance R_c that makes R_c/R_{gl} a reasonable compromise between maximum gain and maximum power output. Usual values are in the range 0.2 to 0.5, corresponding to 91 and 80 per cent, respectively, of the maximum possible output voltage. A value of r of the order of 0.1 to 0.2 is then assumed and the plate current adjusted so that the voltage drop in this coupling resistance will satisfy Eq. (13). The curves of Fig. 10 give the results of Eqs. (13) and (14) graphically. Under practical conditions it is apparent that the peak output voltage obtainable is roughly $0.35E_B$ and that the proper voltage drop in the coupling resistance will approximate $0.5E_B$.

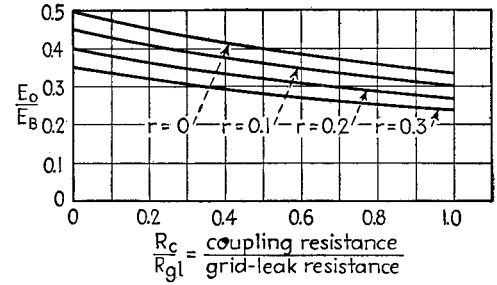
[1] In battery-operated tubes, the tube is often operated with zero bias and a very high grid-leak resistance (5 to 10 megohms).

A comparison of the amplification obtained when the amplifier is designed for maximum gain as contrasted with the design for large voltage output shows that in the former case the voltage amplification is about one-third greater.

Effect of Plate-supply Voltage and Grid-leak Resistance on Amplification.—Examination of Eq. (12) shows that for a fixed value of R_c/R_{gl}, the gain is proportional to $E_d^{(n-1)/n}$ (and hence to $E_B^{(n-1)/n}$), and also to $R_{gl}^{1/n}$. In practice, this means that the amplification obtainable from a resistance-coupled amplifier, irrespective of whether designed for maximum gain or large output voltage, can be expected to be



(a) Proper d-c voltage drop E_d in coupling resistance



(b) Maximum peak a-c voltage E_o assuming proper E_d

FIG. 10.—Curves giving the maximum output voltage obtainable from resistance-coupled amplifiers, and also the value of E_d required for maximum output.

proportional to something between the square root and cube root of the plate-supply voltage and between the square root and ⅔ power of the grid-leak resistance.

Miscellaneous Design Considerations.—Examples of designs recommended by tube manufacturers as suitable for audio-frequency work are given in Table 1.[1] These are proportioned with the main emphasis on large output voltage and sharp low-frequency cutoff.

The amplification and the output voltage obtainable both increase with plate-supply potential up to the point where with normal rated d-c plate current and the proper coupling resistance the d-c voltage at the plate of the tube equals the rated plate voltage.

The falling off in gain at high frequencies is determined by the equivalent resistance formed by the coupling and grid-leak resistances in parallel in relation to the shunting capacities. Since these shunting capacities are fixed, they fix the maximum

[1] These designs are from the RCA Radiotron tube manual.

coupling resistance that can be used to give a desired frequency range. When extremely wide frequency bands are required, the coupling resistance must be lower than the optimum value for either maximum output voltage or maximum amplification. The loss resulting from somewhat less than optimum coupling resistance is not particularly great, however, since the plate current can be increased as the resistance is reduced, thereby obtaining partial compensation for the lowered impedance.

The character of the response at low frequencies can be controlled over wide limits by the amplifier circuit proportions. A very sharp cutoff at low frequencies accompanied by a large phase shift results when the critical frequencies f_1 and f_3 coincide, and $f_4 \leq f_1$ with $R_f g_m$ approximating unity. On the other hand, a gradual low-frequency cutoff with less pronounced phase shifts is obtained when f_3 and f_4 are less than f_1 (*i.e.*, when the screen and bias by-pass condensers are larger). Examples of some of the characteristics that can be obtained are shown in Fig. 11.

In ordinary amplifiers it is usually desirable to have the low-frequency response no better than absolutely necessary, with the sharpest possible cutoff. In this way motorboating and regeneration troubles are minimized. The main exception to this is in feedback amplifiers, where the sharp cutoff is to be avoided because it introduces additional phase shifts.

Amplifiers for special purposes are sometimes required to amplify down to only a few cycles. Such requirements can be most easily met by using some other biasing system than self-bias and by deriving the screen voltage from a voltage divider of low impedance. The low-frequency response is then determined by the coupling condenser C_c in accordance with Eq. (4), and is better in proportion to C_c. The largest capacity that can be used and hence the best low-frequency response obtainable is limited by the leakage resistance of the condenser. If the condenser is very large, or the dielectric poor, the leakage current will be sufficient to produce an appreciable voltage across the grid-leak resistance and bias the grid of the next stage of amplification positively. Further discussion on the amplification of very low frequencies is given below in Par. 6.

The coupling resistors used in amplifier stages operating at low signal levels must be of a type that generates relatively little "noise" voltage when carrying the d-c plate current. Otherwise the signal-to-noise ratio of the amplifier output will be poor. Wire-wound resistors and certain special composition and film types are suitable, while ordinary carbon resistors have rather poor noise characteristics (Par. 26).

The tubes used in resistance-coupled amplifiers are normally sharp cutoff general-purpose pentodes, having a rated d-c plate current of several milliamperes.

Resistance-coupled Amplifiers Using Triode Tubes.—Triode tubes are occasionally used in resistance-coupled amplifiers. Such tubes have the disadvantage of less amplification than pentodes and also offer higher input capacity to the preceding stage (see Par. 25 below). The chief advantage of using triode tubes is that the amplitude distortion is less when large voltages are being developed.

In designing triode resistance-coupled amplifiers, the coupling resistance is normally chosen so that the d-c drop in it is about one-third to one-half of the plate-supply voltage. The coupling condenser and grid-leak resistance are determined exactly as in the case of pentode tubes. A tube having a high amplification factor must be used if appreciable gain is desired, since the gain can never exceed the amplification factor.

Typical designs for resistance-coupled amplifiers employing triode tubes are shown in Table 1.

5. Transformer-coupled Audio-frequency Amplifiers.—In the transformer-coupled amplifier the load impedance in the plate circuit of the tube is supplied by a step-up transformer that delivers its secondary voltage to the grid of another amplifier tube,

TABLE 1[1]

Pentode types: 6C6, 6J7, 57

E_b	90									180										300									E_b
R_c	0.1			0.25			0.5			0.1			0.25			0.5				0.1			0.25			0.5			R_c
R_{gl}	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	R_{gl}
R_s	0.37	0.44	0.44	1.1	1.18	1.4	2.18	2.6	2.7	0.44	0.5	0.5	1.1	1.18	1.4	2.45	2.9	2.7	0.44	0.5	0.53	1.18	1.18	1.45	2.45	2.9	2.95	R_s
R_f	1,200	1,100	1,300	2,400	2,600	3,600	4,700	5,500	5,500	1,000	750	800	1,200	1,600	2,000	2,600	3,100	3,500	500	450	600	1,100	1,200	1,300	1,700	2,200	2,300	R_f
C_{sg}	0.05	0.05	0.05	0.03	0.03	0.025	0.02	0.05	0.02	0.05	0.05	0.05	0.04	0.04	0.04	0.03	0.025	0.02	0.07	0.07	0.06	0.04	0.04	0.05	0.04	0.04	0.04	C_{sg}
C_f	5.2	5.3	4.8	3.7	3.2	2.5	2.3	2	2	6.5	6.7	6.7	5.2	4.3	3.8	3.2	2.5	2.8	8.5	8.3	8	5.5	5.4	5.8	4.2	4.1	4	C_f
C_c	0.02	0.01	0.006	0.008	0.005	0.003	0.005	0.0025	0.0015	0.02	0.01	0.006	0.008	0.005	0.0035	0.005	0.0025	0.0015	0.02	0.01	0.006	0.008	0.005	0.005	0.005	0.003	0.0025	C_c
E_0*	17	22	33	23	32	33	28	29	27	42	52	59	41	60	60	45	56	60	55	81	96	81	104	110	75	97	100	E_0*
G†	41	55	66	70	85	92	93	120	140	51	69	83	93	118	140	135	165	165	61	82	94	104	140	185	161	250	240	G†

Duplex-diode triode types: 2A6, 75

E_b	90									180									300									E_b
R_c	0.1			0.25			0.5			0.1			0.25			0.5			0.1			0.25			0.5			R_c
R_{gl}	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	0.1	0.25	0.5	0.25	0.5	1	0.5	1	2	R_{gl}
R_f	6,300	6,600	6,700	10,000	11,000	11,500	16,200	16,600	17,400	2,600	2,900	3,000	4,300	4,800	5,300	7,000	8,000	8,800	1,900	2,200	2,300	3,300	3,900	4,200	5,300	6,100	7,000	R_f
C_f	2.2	1.7	1.7	1.24	1.07	0.9	0.75	0.7	0.65	3.3	2.9	2.7	2.1	1.8	1.5	1.3	1.1	0.9	4	3.5	3	2.7	2	1.8	1.6	1.3	1.2	C_f
C_c	0.02	0.01	0.006	0.01	0.006	0.003	0.005	0.003	0.0015	0.025	0.015	0.007	0.015	0.007	0.004	0.007	0.004	0.002	0.03	0.015	0.007	0.015	0.007	0.004	0.007	0.004	0.002	C_c
E_0*	3	5	6	5	7	10	7	10	13	16	22	23	21	28	33	25	33	38	31	41	45	42	51	60	47	62	67	E_0*
G†	23	29	31	34	40	40	39	44	48	29	36	37	43	50	53	52	57	58	31	39	42	48	53	56	58	60	63	G†

Triode types: 56, 76

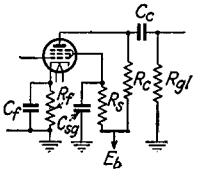
E_b	90									180									300									E_b
R_c	0.05			0.1			0.25			0.05			0.1			0.25			0.05			0.1			0.25			R_c
R_{gl}	0.05	0.1	0.25	0.1	0.25	0.5	0.25	0.5	1	0.05	0.1	0.25	0.1	0.25	0.5	0.25	0.5	1	0.05	0.1	0.25	0.1	0.25	0.5	0.25	0.5	1	R_{gl}
R_f	2,500	3,200	3,800	4,500	6,500	7,500	11,100	15,100	18,300	2,400	3,000	3,700	4,500	6,500	7,600	10,700	14,700	17,700	2,400	3,100	3,800	4,500	6,400	7,500	11,100	15,200	18,300	R_f
C_f	2	1.6	1.25	1.05	0.82	0.68	0.48	0.36	0.32	2.5	1.9	1.65	1.45	0.97	0.8	0.6	0.45	0.4	2.8	2.2	1.8	1.6	1.2	0.98	0.69	0.5	0.4	C_f
C_c	0.06	0.03	0.015	0.03	0.015	0.007	0.015	0.007	0.0035	0.06	0.035	0.015	0.035	0.015	0.008	0.015	0.007	0.0045	0.08	0.045	0.02	0.04	0.02	0.009	0.02	0.009	0.005	C_c
E_0*	16	21	23	19	23	25	21	24	28	36	48	55	45	55	57	49	59	64	65	80	95	74	95	104	82	96	108	E_0*
G†	7	7.7	8.1	8.1	8.9	9.3	9.4	9.7	9.8	7.7	8.2	9	9.3	9.5	9.8	9.7	10	10	8.3	8.9	9.4	9.5	10	10	10	10	10	G†

[1] Capacities in microfarads, resistances in megohms.

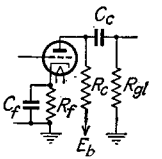
* E_0 = maximum output voltage.

† G = voltage gain.

Pentode diagram



Triode diagram



as shown in Fig. 12*a*. The amplifier tube exciting the transformer is ordinarily a triode having a plate resistance of the order of 10,000 ohms and a d-c plate current of a few milliamperes.

Equivalent Circuit and Frequency Response Characteristics of Transformer-coupled Amplifier.—The exact equivalent plate circuit of a transformer-coupled amplifier is shown in Fig. 12*b* and is quite complicated. It can, however, be simplified, as shown in Fig. 12*c*, without introducing appreciable error, by neglecting primary capacity and hysteresis loss, replacing the interwinding capacity by suitable capacity across

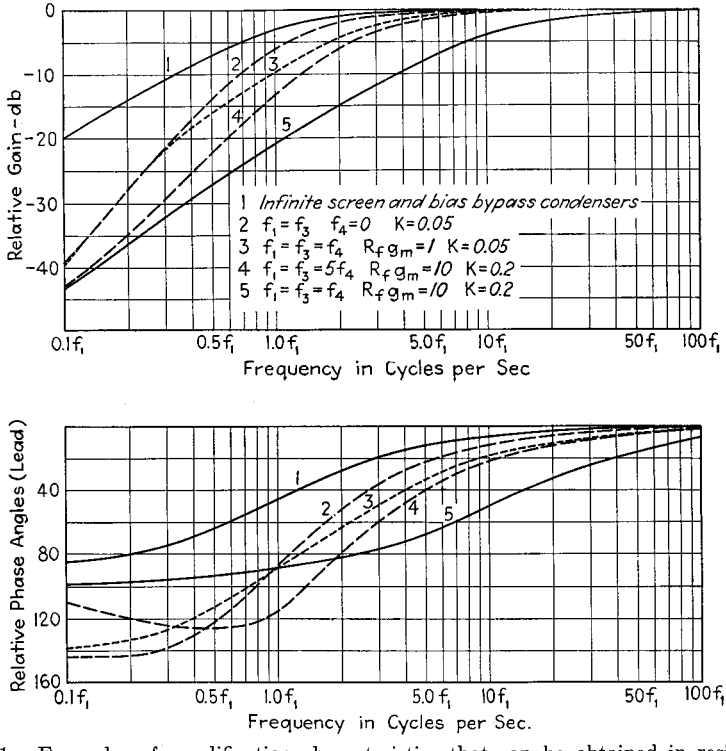


FIG. 11.—Examples of amplification characteristics that can be obtained in resistance-coupled amplifiers at low frequencies by various circuit proportions.

the secondary, lumping the primary resistance with the plate resistance of the tube, neglecting the eddy-current resistance,[1] and reducing to unity turn ratio. Still further simplifications are possible by considering only a limited range of frequencies at a time, as shown at (*d*), (*e*), and (*f*).

[1] The effect of eddy-current resistance R_e can be taken into account by using for R_p and μ values slightly lower than those actually given by the tube, according to the relations

$$\left.\begin{array}{l}\text{Effective } \mu \text{ when } R_e \text{ is} \\ \text{taken into account}\end{array}\right\} = \frac{\text{actual } \mu \text{ of tube}}{1 + \dfrac{R_p}{R_e}} \tag{15}$$

$$\left.\begin{array}{l}\text{Effective } R_p \text{ when } R_e \text{ is} \\ \text{taken into account}\end{array}\right\} = \frac{\text{actual value of } R_p}{1 + \dfrac{R_p}{R_e}} \tag{16}$$

The output voltage of a transformer-coupled amplifier can be calculated by the equivalent circuits of Fig. 12, with the use of the following equations:

$$\left.\begin{matrix}\text{Amplification in middle}\\ \text{range of frequencies}\end{matrix}\right\} = \mu n \tag{17}$$

$$\frac{\text{Amplification at low frequencies}}{\text{Amplification in middle range}} = \frac{1}{1 - j\frac{f_1}{f}} \tag{18}$$

$$\frac{\text{Amplification at high frequencies}}{\text{Amplification in middle range}} = \frac{1}{(1 - \gamma^2) + j\frac{\gamma}{Q_0}} \tag{19}$$

where $f_1 = R'_p/2\pi L_p$ = frequency at which resistance of primary incremental inductance L_p equals effective plate resistance R'_p.

$$\gamma = \frac{\text{actual frequency}}{\text{series resonant frequency of } C_s \text{ and } L_s}.$$

$$Q_0 = \frac{w_0 L}{R'_p + R_p} = \text{circuit } Q \text{ at frequency for which } C_s \text{ and } L_s \text{ are in series resonance.}$$

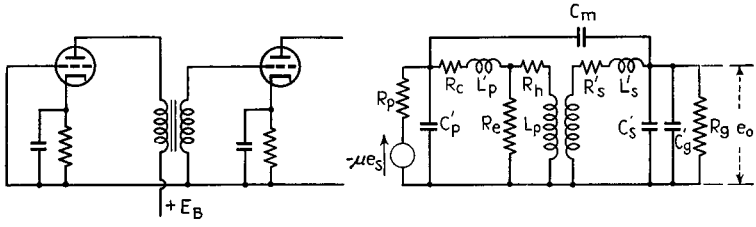
The variation in relative amplification at high and low frequencies, as given by Eqs. (18) and (19), is shown graphically in Fig. 13. At low frequencies there is a falling off in amplification because of insufficient primary inductance, *with the gain being* 70.7 *per cent* (−3 *db*) *of the mid-frequency gain at the frequency* f_1, *for which the reactance of the primary incremental inductance* $2\pi f_1 L_p$ *equals the effective plate resistance* R'_p. At high frequencies the behavior is controlled by the action of the equivalent secondary capacity C_s resonating with the leakage inductance of the transformer. The frequency at which the equivalent shunt capacity (*i.e.*, actual distributed secondary capacity plus capacity of tube connected to secondary) is in series resonance with the leakage inductance of the transformer fixes the high-frequency limit of the transformer, as the gain falls well below the mid-frequency value at frequencies from 25 to 50 per cent greater than this resonant frequency. The Q of the resonant circuit formed by the shunting capacity, leakage inductance, and plate resistance (designated by Q_0 in Fig. 13) determines the character of the high-frequency response. The amplification is most nearly uniform for $Q_0 = 0.8$, with higher values resulting in a high-frequency peak, and lower values in a drooping curve.

Transformer Characteristics.—The characteristics desired in an audio-frequency transformer are high primary incremental inductance to give good low-frequency response, high step-up ratio to give large amplification, and low leakage inductance and low distributed capacity to extend the response to high frequencies.

The common audio transformer employs a silicon steel core with a small air gap such as to give maximum incremental primary inductance with the normal d-c plate current of the tube passing through the primary. The primary and secondary are layer-wound, with the secondary on the outside. The turn ratio is commonly two to four, and the 70.7 per cent (−3 db) point for low frequencies is of the order of 40 to 100 cycles. The high-frequency response will depend upon whether or not the primary and secondary are interleaved and upon the way the interleaving is done. Relatively uniform response can be readily obtained up to 10,000 to 20,000 cycles. In general, the greater the high-frequency range the lower will be the step-up ratio and the less the gain.

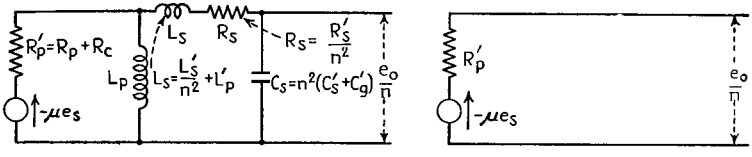
The use of core material of high permeability, such as permalloy, increases the frequency range obtainable in a given physical size, or permits the use of a physically smaller transformer for a given range. Alloy cores are desirable where weight is a factor or where the frequency range is great.

The characteristics of audio-frequency transformers can be calculated with reasonable accuracy by the method described in Sec. 2, Par. 25. The characteristics can also be measured experimentally as described in Par. 34, Sec. 13.



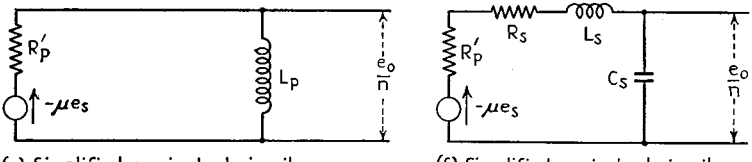
(a) Circuit of transformer-coupled amplifier

(b) Exact equivalent circuit of transformer-coupled amplifier



(c) Practical equivalent circuit of transformer-coupled amplifier reduced to unity turn ratio

(d) Simplified equivalent circuit accurate for middle range of frequencies



(e) Simplified equivalent circuit accurate at low frequencies

(f) Simplified equivalent circuit accurate at high frequencies

R_p = Plate resistance of tube
R_c = D-c resistance of primary
L'_p = Primary leakage inductance
R_e = Resistance representing eddy current loss
R_h = Resistance representing hysteresis loss
L_p = Incremental primary inductance
R'_s = Resistance of secondary winding
L'_s = Secondary leakage inductance
$L_s = L'_p + L'_s/n^2$ = total leakage inductance referred to primary side

C'_s = Secondary distributed capacity
C'_p = Capacity in shunt with primary (tube plus transformer capacity)
C_m = Capacity between primary and secondary windings of the transformer
C'_g = Input capacity of tube to which voltage e_o is delivered
R_g = Input resistance of tube to which voltage e_o is delivered
n = Ratio of secondary to primary turns
μ = Amplification factor of tube

FIG. 12.—Circuit of a transformer-coupled amplifier, together with exact and approximate equivalent plate circuits.

Miscellaneous Considerations.—For any given transformer there is a particular value of plate resistance that gives the most uniform amplification characteristic.

If the plate resistance is higher, the response falls off excessively at both high and low frequencies, while if the resistance is too low, the response is excellent at low frequencies but peaked at high frequencies. The proper plate resistance can be obtained by adjusting the bias of the amplifier tube.

Pentode tubes can be used with transformer-coupled amplifiers provided a resistance is shunted across the primary of the transformer. By making this shunting resistance equal to the plate resistance of the triode tube with which the transformer was designed to operate, the frequency response characteristic is exactly the same as with the corresponding triode, and the amplifications in the two cases will be the same if the transconductance of the pentode is the same as that of the triode.

Transformer-coupled amplifiers are sometimes operated with a resistance shunted across the transformer secondary. This improves the relative response at low frequencies, reduces the relative response at high frequencies, and lowers the absolute magnitude of the amplification. The chief use of such a resistance is to eliminate the high-frequency peak that results when a transformer is associated with a plate resistance that is too low.[1]

Uses of Transformer-coupled Amplifiers.—Transformer coupling is more expensive, gives less gain, and has poorer frequency response than obtainable with resistance coupling. Its use is accordingly limited to special circumstances where resistance coupling is not entirely suitable. The most important applications of transformer coupling are in the excitation of a push-pull amplifier from a single-ended amplifier stage, and in cases where the d-c resistance in the grid circuit of the tube being excited must be low because of grid current, or because of danger of blocking (Par. 13, Sec. 4).

6. Miscellaneous Coupling Systems for Audio-frequency Voltage Amplifiers.—Although most audio-frequency voltage amplifiers use either resistance or transformer coupling, other coupling methods can be employed, and some have advantages. The most important of these miscellaneous systems are described below.

Impedance Coupling.—In this type of coupling the coupling resistance of ordinary resistance coupling is replaced by an inductance, as illustrated in Fig. 14*a*. This arrangement gives the same shape of curve of response as does resistance coupling in the usual case when the coupling condenser C_c is large enough that its reactance is not an important factor in the useful frequency range. The response at low frequencies then falls to 70.7 per cent of the mid-frequency value at the frequency for which the reactance of the coupling inductance L_c equals the resistance formed by grid leak and plate resistance in parallel. The corresponding high-frequency point is at the frequency for which the reactance of the shunting capacities equals the same equivalent resistance. The mid-frequency gain in the case of triodes approximates the amplification factor of the tube, while with pentode tubes, it is equal to the transconductance times the equivalent resistance of grid leak and plate resistance in parallel.

Transformer Coupling with Shunt (Parallel) Feed.—This arrangement, shown in Fig. 14*b*, employs an inductance-capacity combination L_cC_c that prevents direct current from flowing through the primary of the coupling transformer.[2] The response obtained from such an arrangement at high frequencies is exactly the same as with ordinary transformer coupling using the same transformer and tube. At low frequencies, however, the response may be either peaked, flat, or drooping, according to the circuit proportions, as shown in the figure. A substantially flat response can be obtained down to the frequency f_0, at which $2\pi f_0 L_p$ equals the plate resistance of the tube. This is done by selecting a value of condenser C_c that will resonate with the

[1] Methods of analyzing and designing transformer-coupled amplifiers with resistance shunted across the secondary are discussed by Paul W. Klipsch, Design of Audio Frequency Amplifier Circuits Using Transformers, *Proc. I.R.E.*, Vol. 24, p. 219, February, 1936.

[2] The inductance L_c carrying the direct current in such a combination is sometimes referred to as a *retard* coil.

primary inductance L_p of the transformer at this frequency and at the same time by using a coupling inductance L_c that is at least 50 per cent greater than the primary

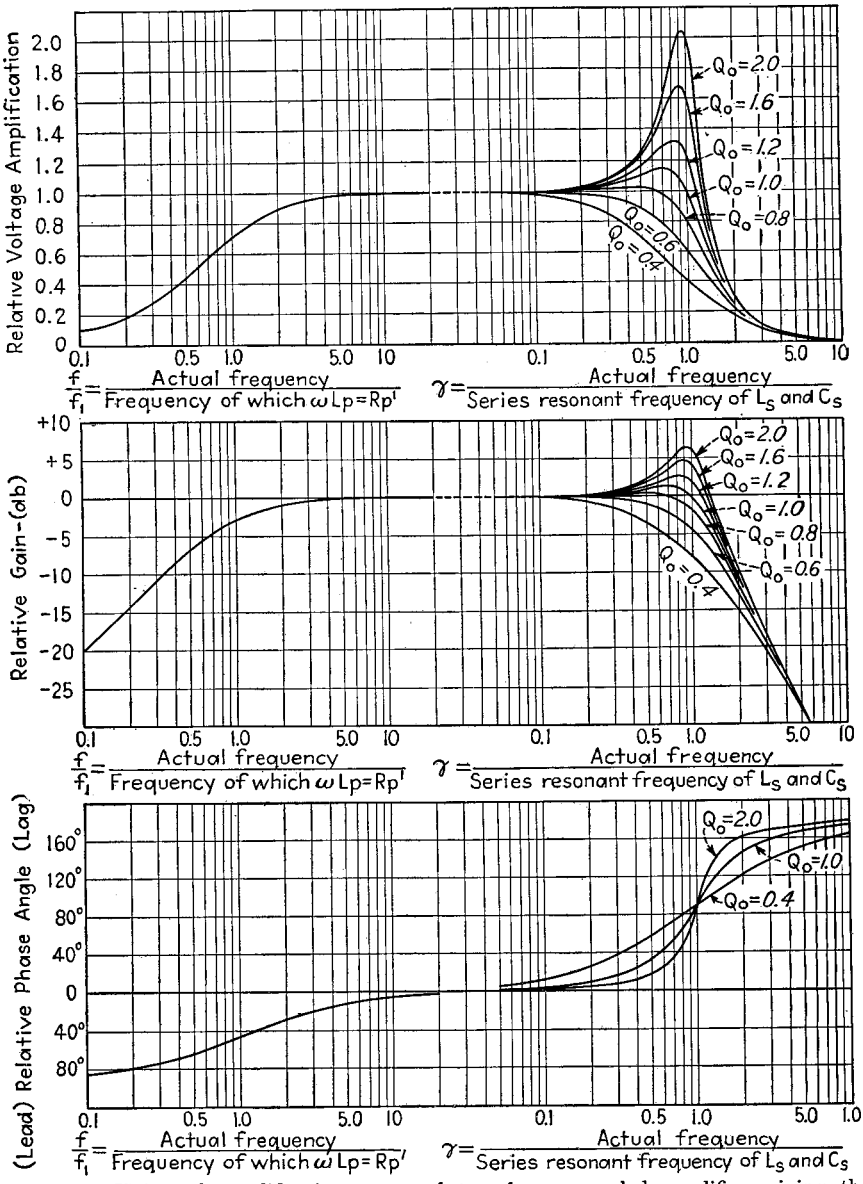


FIG. 13.—Universal amplification curve of transformer-coupled amplifier, giving the relative amplification and phase-shift characteristics.

inductance of the transformer. A smaller coupling capacity C_c gives a peaked response at low frequencies, while a very large capacity gives a falling off at low frequencies corresponding to that obtained with simple transformer coupling when the

primary inductance of the transformer is the equivalent inductance formed by L_p and L_c in parallel.

Transformer coupling with shunt feed is capable of giving a better frequency response than simple transformer coupling, provided that the transformer is designed especially for this type of operation. By eliminating the direct current from the transformer primary, it is possible to assemble the transformer core with the minimum possible air gap and also, if desired, to use alloy core material having extremely high permeability. The result is sufficient increase in primary inductance to more than make up for the shunting inductance of the retard coil, with the added advantage of being able to control the character of the low-frequency response through the size of the condenser C_c. At the same time, the presence of the retard coil L_c and the increased primary inductance do not affect either the leakage reactance or the distributed capacity of the transformer. The improved low-frequency response is hence obtained without affecting the behavior at high frequencies.

Resistance Coupling with Grid Choke.—In this arrangement the grid-leak resistance is replaced by a high inductance coil illustrated in Fig. 14*c*. Such an arrangement is occasionally used where it is necessary that the direct-current resistance be low in the grid circuit of the tube to which the amplified voltage is applied. The response of such an arrangement at high frequencies is similar to that obtained with resistance coupling, with the response falling to 70.7 per cent of the mid-frequency value at the frequency for which the reactance of the shunting capacity equals the resistance formed by coupling resistance and plate resistance in parallel. The response at low frequencies depends upon the resonant frequencies of C_c and L_g and the value of $\omega L_g/R$ at this frequency, where R is the resistance formed by the plate and coupling resistances in parallel. Substantially flat response can be obtained down to a frequency f_0 by making $R = 2\pi f_0 L_g$ and selecting a capacity C_c that resonates with L_g at f_0. A smaller capacity or lower resistance gives a peak at low frequencies, while a larger capacity or higher resistance results in a drooping low-frequency response.

Input Transformers.—A transformer used to couple a transmission line, microphone, or other low impedance source of energy to the grid of a tube is termed an *input transformer*. Such an arrangement corresponds to ordinary transformer coupling, with the source impedance associated with the primary of the transformer being equivalent to the plate resistance of the tube in a transformer-coupled stage of amplification. The only difference is that with the input transformer, the source impedance is usually much lower than the plate resistance of a tube so that the primary inductance can be proportionately reduced. This permits fewer primary turns and a higher step-up ratio, which will, in general, be inversely proportional to the square root of the source impedance.

Special Coupling Systems for Modifying Low-frequency and High-frequency Response. The addition of circuit elements to the coupling network increases the variety of characteristics obtainable. A few of the many possibilities of this sort are described below.

A sharp high-frequency cutoff in resistance-coupled amplifiers, comparable with that obtainable at low frequencies, can be realized by adding inductance in series with the screen-grid and bias by-pass condensers, as shown in Fig. 14*d*. These inductances act to increase the impedance in the screen and bias circuits at high frequencies in exactly the same way as the by-pass condensers increase the impedance in these circuits at low frequencies. In fact, Eq. (8) and Fig. 5, giving the low-frequency falling off caused by the reactance of the screen condenser C_s, also give the high-frequency falling off from a screen inductance L_s if one substitutes f_{sh}/f for f/f_s, where f_{sh} is the frequency at which the reactance of the screen inductance L_s equals the resistance formed by R_s and R_{sg} in parallel. Similarly, Eq. (10) and Fig. 8, giving the

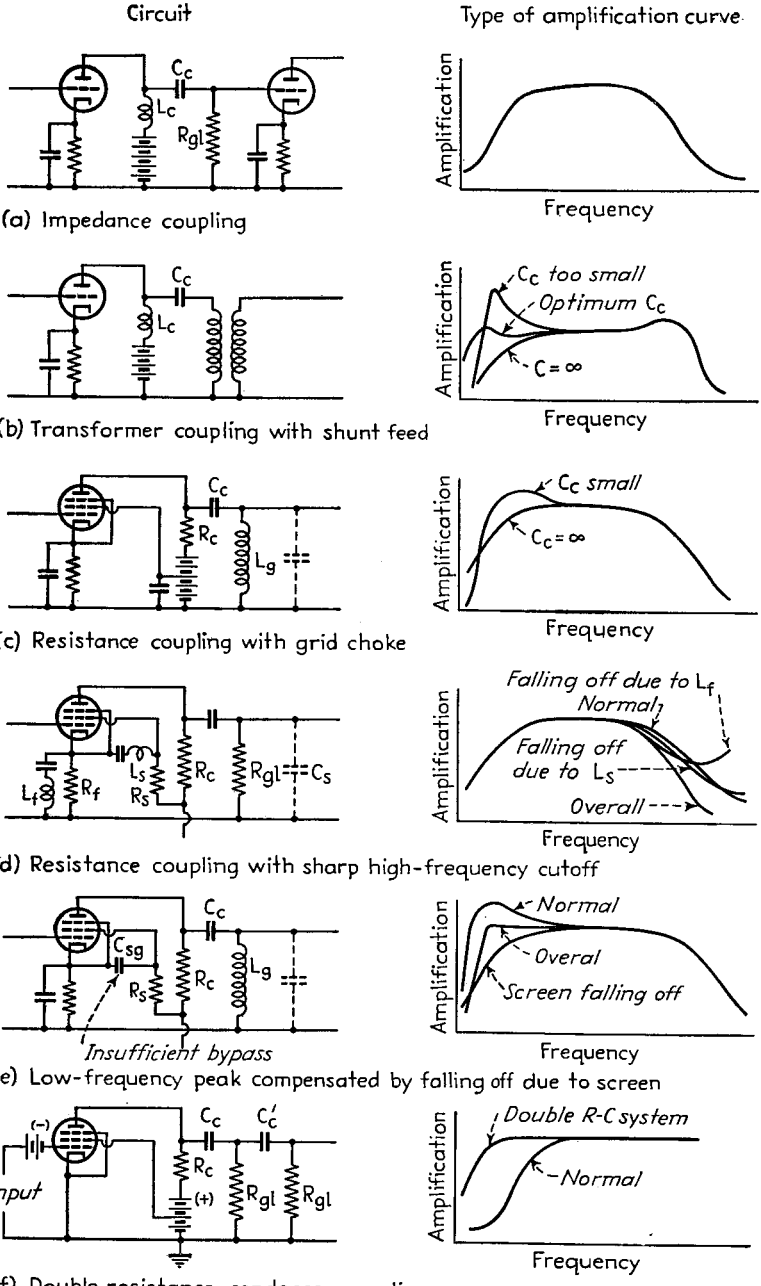


FIG. 14.—Miscellaneous coupling methods for audio-frequency amplifiers together with typical amplification curves that are obtained.

loss in amplification and the phase shift resulting from the reactance of the bias by-pass condenser C_f, also gives the effect of the inductance L_f at high frequencies, provided that f_{4h}/f is substituted for f/f_4, where f_{4h} is the frequency at which the reactance of the inductance L_f equals the bias resistance R_f.

The high-frequency response can be extended to very high frequencies while substantially constant gain and negligible phase distortion are maintained by use of the coupling systems used in video amplifiers (see Par. 16).

The low-frequency response can be modified in a number of ways. Thus in coupling systems having a low-frequency peak, as, for example, with systems using a grid choke (Fig. 14*c*), it is possible to select a screen (or bias) by-pass condenser that will give a compensating falling off in gain at low frequencies. In this way a substantially uniform gain can be obtained down to a certain point, followed by a relatively sharp cutoff at lower frequencies, as illustrated in Fig. 14*e*.

The limitation to the low-frequency response in ordinary resistance-coupled systems arising from the leakage current that flows through the coupling condenser when very large coupling condensers are used can be largely eliminated by the double resistance-condenser coupling system shown in Fig. 14*f*. This arrangement makes it possible to obtain a substantially uniform response to about one-tenth of the frequency otherwise possible. Where this is not sufficient, it is possible to use direct-coupled systems such as are discussed below.

Coupling systems giving negligible phase shift and substantially constant gain down to quite low frequencies are an essential part of video amplifiers, and are discussed in detail in Par. **7**.

Direct-coupled (Direct-current) Amplifiers.—A direct-coupled amplifier is one in which the grid of the tube receiving the output voltage is directly coupled to the plate of the amplifier tube. Such an arrangement is capable of amplifying direct as well as alternating voltages, and if there are no condensers in the screen circuit or in the arrangement for producing bias, the amplification at low frequencies will be uniform and without phase shift down to zero frequency. Examples of direct-coupled amplifiers are shown in Fig. 15. The arrangement at (*a*) is a simple amplifier in which the voltage developed across the coupling resistance R_c is applied directly to the grid of the output tube through a bias battery E_2, which compensates for the positive voltage from plate to ground of the amplifier tube. The circuit at (*b*) differs somewhat from that at (*a*) in that the coupling resistance R_c is the plate resistance of a pentode tube.[1] This provides an impedance that will pass the d-c plate current of the amplifier tube with only a moderate voltage drop yet that offers extremely high resistance to any change in this current. In this way a voltage amplification of several thousand is readily obtainable.

The circuits of both Figs. 15*a* and 15*b* require ungrounded batteries, which result in high shunting capacities to ground and resultant poor high-frequency response. This disadvantage can be avoided by the modification of Fig. 15(*a*) shown at (*e*) or by connecting the various stages of the amplifier across a single voltage divider,[2] as shown in Fig. 15*c*. Separate sources of filament power are required for each tube, but this is no disadvantage if heater tubes are employed. In order to prevent changes in plate current in one stage of such an arrangement from reacting unduly on the other stages, it is important that the current carried by the voltage dividing resistance be considerably greater than the plate currents of the individual tubes.

Direct-coupled amplifiers of all types tend to be troubled by slow drifts in the d-c plate current resulting from changes in battery potentials. Such variations are

[1] J. W. Horton, The Use of a Vacuum Tube as a Plate-feed Impedance, *Jour. Franklin Inst.*, Vol. 216, p. 749, December, 1933.

[2] Edward H. Loftin and S. Young White, Cascaded Direct-coupled Tube Systems Operated from Alternating Current, *Proc. I.R.E.*, Vol. 18, p. 669, April, 1930.

amplified just as an applied signal, since the tubes have no way of distinguishing between the desired and undesired voltages. The remedy for this is to employ some form of balancing system, a simple example of which is shown in Fig. 15*d*. Here the tubes are arranged in push-pull pairs, with the signal applied in opposite phase to each set of paired tubes, while the direct-current potentials are applied in the same phase to these tubes. As a result, changes in electrode voltages balance each other out as far as the output terminals are concerned, provided that the tubes are absolutely

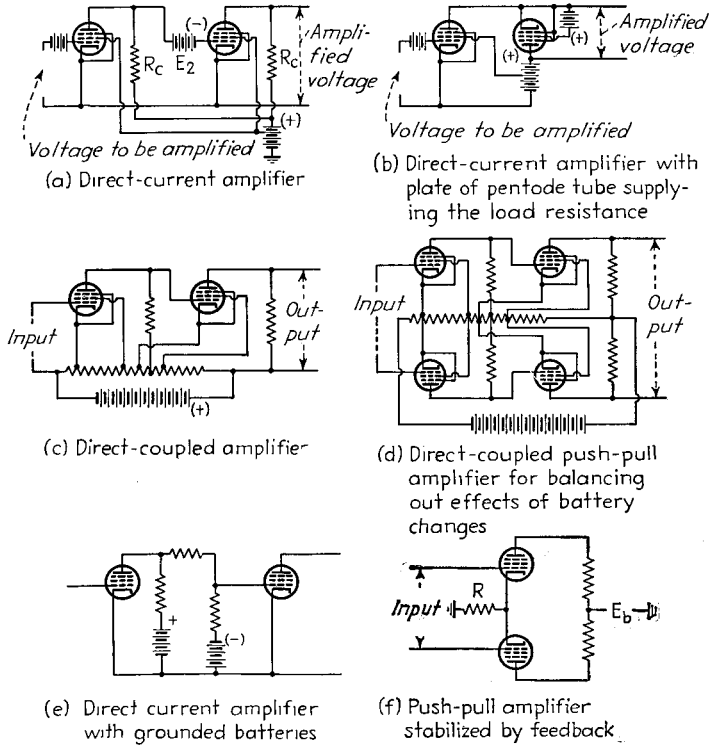


FIG. 15.—Direct-coupled amplifiers for amplifying low-frequency alternating currents without phase shift or falling off in gain, and also for amplifying direct currents.

identical. Very small d-c voltages can be successfully amplified with such balanced arrangements by employing power supplies in which the output voltage is regulated, combined with auxiliary circuits that compensate for slight differences in the characteristics of the two tubes that are paired.[1] The immunity of push-pull arrangements from drift troubles can also be greatly increased by use of negative feedback in the form of a large bias resistor R in the common cathode lead, as shown in Fig. 15*f*.[2]

[1] Discussion of a number of such balancing arrangements is given by D. B. Penick, Direct-current Amplifier Circuits for Use with the Electrometer Tube, *Rev. of Sci. Instruments*, Vol. 6, p. 115, April, 1935; G. P. Harnwell and S. N. Van Voorhis, A Balanced Electrometer Tube and Amplifying Circuit for Small Directory Currents, *Rev. Sci. Instrumenis*, Vol. 5, p. 244, July, 1934; J. M. Eglin, A Direct Current Amplifier for Measuring Small Currents, *Jour. O.S.A.*, Vol. 18, p. 393, May, 1929.

[2] The full design of a practical d-c amplifier giving a voltage gain of 6×10^6 and employing a push-pull system with negative feedback is given by Harold Goldberg, *Trans. A.I.E.E.*, Vol. 59, p 60, January, 1940.

AUDIO-FREQUENCY POWER AMPLIFIERS

7. Class A Power Amplifiers.—In a power amplifier, the object is to obtain as much power with only moderate distortion as is possible from a given tube, irrespective of the signal voltage required or the voltage amplification involved. In Class A power amplifiers, this output is developed under conditions such that the plate currents of the individual tubes flow continuously throughout the cycle.

Practical circuits of power amplifiers are shown in Fig. 16. In these, the load offers an impedance to the alternating components of plate current, but not to the d-c plate current.

Dynamic Characteristics and Load Lines.[1]—The proper operation of a Class A amplifier requires a careful balance between the load impedance, operating point, and

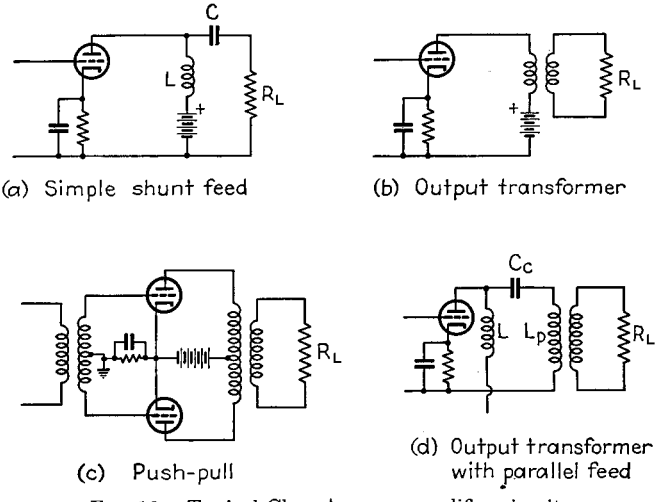


FIG. 16.—Typical Class A power-amplifier circuits.

signal voltage. The essential factors involved with resistance loads are normally determined with the aid of *dynamic characteristics* and *load lines*.

A dynamic characteristic gives the relationship between instantaneous plate current and instantaneous potential at the plate of the tube as the voltage applied to the grid of the tube is varied. A typical dynamic characteristic of a triode tube is shown in Fig. 17. Points on the dynamic characteristic can be calculated from the relation

$$I'_p = \frac{E_b - E'_b}{R_L} + I_b \tag{20}$$

where E_b = potential at plate of tube with no applied signal.
I_b = plate current with no applied signal.
E'_b = instantaneous plate potential at some point P on the dynamic characteristic.
I'_p = instantaneous current corresponding to point P.
R_L = load resistance.

[1] E. W. Kellogg, Design of Non-distorting Power Amplifiers, *Trans. A.I.E.E.*, Vol. 44, p. 302, 1925; J. C. Warner and A. V. Loughren, The Output Characteristics of Amplifier Tubes, *Proc. I.R.E.*, Vol. 14, p. 735, December, 1926.

The dynamic characteristic for a given operating point and load resistance gives the relationship between the applied signal and the output voltage, as shown in Fig. 17.

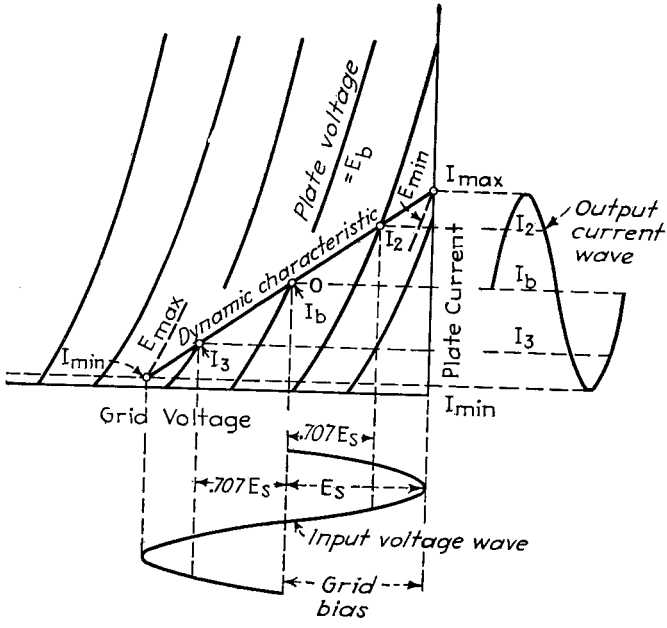


FIG. 17.—Typical dynamic characteristic for Class A power amplifier with grid driven to zero voltage, showing input and output waves and critical points.

From the dynamic characteristic one can determine the conditions for optimum operation, the wave shape of the output voltage, the output power, etc.

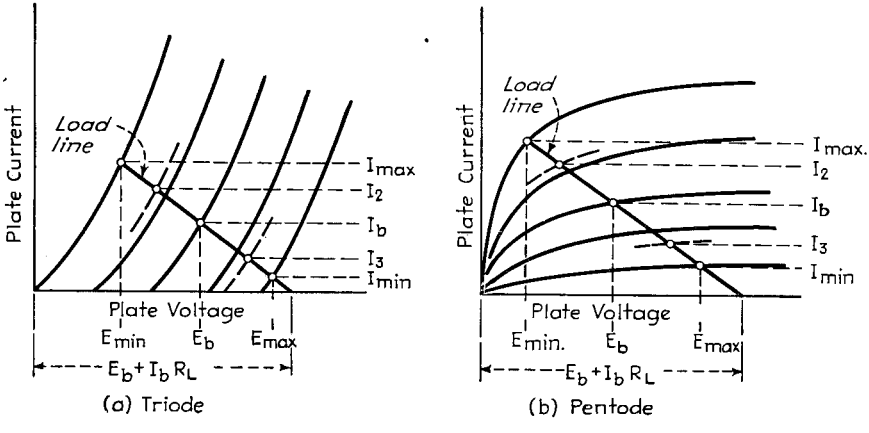


FIG. 18.—Typical load lines for triode and pentode Class A power amplifiers, showing critical points.

In the practical determination of the performance of Class A amplifiers only certain critical points on the dynamic characteristic are needed. These particular points

are most readily obtained with the aid of a *load line* drawn on the plate-voltage–plate-current characteristic curves of the tube as shown in Fig. 18. The load line is a straight line passing through the operating point (E_b, I_b) and intersecting the zero plate current axis at a plate potential of $E_b + I_bR_L$. Such a line is a dynamic characteristic plotted on the $E_b - I_b$ coordinate system.

Class A Amplifier Calculations for Resistance Loads.—The power output from a Class A amplifier with a resistance load is[1]

$$\text{Power output} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8} \tag{21a}$$

$$= \frac{E_bI_b}{2}\left(1 - \frac{E_{min}}{E_b}\right)\left(1 - \frac{I_{min}}{I_b}\right) \tag{21b}$$

where E_{max} and E_{min} = maximum and minimum instantaneous values of plate voltage, respectively, reached during the cycle.

I_{max} and I_{min} = maximum and minimum instantaneous values of plate current, respectively, reached during the cycle.

E_b and I_b = plate voltage and current, respectively, at the operating point.

Values of E_{max}, E_{min}, etc., are readily found from the load line for any specific operating point, load resistance, and signal voltage. In order to be able to obtain a large power output, it is obviously necessary that the d-c power input E_bI_b be large and also that the load resistance and grid bias be so selected that a signal voltage that makes I_{min}/I_b small on the negative peaks of the cycle likewise makes E_{min}/E_b small on the positive peaks.

The efficiency of a power amplifier is defined as the ratio of power output that can be developed with only moderate distortion, to the d-c power supplied to the plate. The efficiency of a Class A amplifier accordingly is

$$\text{Plate efficiency} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8E_bI_b} \tag{22a}$$

$$= \frac{1}{2}\left(1 - \frac{E_{min}}{E_b}\right)\left(1 - \frac{I_{min}}{I_b}\right) \tag{22b}$$

The efficiency has a theoretical maximum value of 50 per cent, and approaches this more closely as the ratios I_{min}/I_b and E_{min}/E_b are small. In practical Class A amplifiers, I_{min}/I_b is of the order of 0.1 to 0.2; while E_{min}/E_b ranges from this value for pentodes and beam tubes, and also triodes with large plate voltages or operated with the grid driven appreciably positive, to values of the order of 0.6 for triode tubes operating at low plate potentials such as 250 volts and with the grid never driven positive. The corresponding practical efficiencies are from 35 per cent down to about 20 per cent.

The load resistance R_L corresponding to specified values of E_{max}, E_{min}, I_{max}, and I_{min} is

$$R_L = \frac{(E_{max} - E_{min})}{I_{max} - I_{min}} = \frac{E_b}{I_b}\left(1 - \frac{E_{min}}{E_b}\right) \tag{23}$$

In practice, this gives load resistances of approximately E_b/I_b under conditions where the plate efficiency is reasonably high (*i.e.*, when E_{min}/E_b and I_{min}/I_b are small), and resistances about half as great for triodes operated with E_{min}/E_b approaching 0.6, as mentioned above. In triodes where the maximum instantaneous grid potential

[1] Equation (21*a*) gives the exact power output on the fundamental frequency for the case of negligible distortion, or when all the distortion is in the form of even harmonics. Equation (21*b*), however, is exact only in the absence of distortion.

reached during the cycle is exactly zero, the condition for maximum power output with only moderate distortion corresponds very closely to a load resistance twice the plate resistance at the operating point.[1]

Distortion Calculation.—Nonlinear distortion is always present to some extent in a power amplifier because the dynamic characteristic is never a straight line. The character of the distortion obtained in practice depends upon the operating conditions. In triodes, when the distortion is not excessive and the grid is not driven positive, practically all of the distortion produced with a sine wave signal is second harmonic, and can be calculated from the relations[2]

$$\text{Direct-current component} = A_0 = \frac{I_{max} + I_{min} + 2I_b}{4} \tag{24a}$$

$$\text{Second harmonic} = \frac{I_{max} + I_{min} - 2I_b}{4} \tag{24b}$$

$$\frac{\text{Second harmonic}}{\text{Fundamental}} = \frac{I_{max} + I_{min} - 2I_b}{2(I_{max} - I_{min})} \tag{24c}$$

With pentodes and beam tubes, and when the distortion is large, or when the grid is driven somewhat positive, appreciable third and fourth harmonics can also be expected. In this case, the various components of the output voltage are[3]

$$\text{Direct-current component} = A_0 = \frac{\frac{1}{2}(I_{max} + I_{min}) + I_2 + I_3 + I_b}{4} \tag{25a}$$

$$\text{Fundamental} = A_1 = \frac{\sqrt{2}(I_2 - I_3) + I_{max} - I_{min}}{4} \tag{25b}$$

$$\text{Second harmonic} = A_2 = \frac{I_{max} + I_{min} - 2I_b}{4} \tag{25c}$$

$$\text{Third harmonic} = A_3 = \frac{I_{max} - I_{min} - 2A_1}{2} \tag{25d}$$

$$\text{Fourth harmonic} = A_4 = \frac{2A_0 - I_2 - I_3}{2} \tag{25e}$$

$$\text{Power output} = \frac{A_1^2 R_L}{2} \tag{25f}$$

where I_2 and I_3 are the instantaneous plate currents when the instantaneous voltage applied to the grid is the bias voltage respectively plus and minus 0.707 times the peak signal voltage acting on the grid as indicated in Figs. 17 and 18, and the remaining notation is the same as above.

[1] Maximum power sensitivity, *i.e.*, maximum output power in proportion to the applied signal is always obtained in a Class A amplifier when the load resistance equals the plate resistance. Operation under this condition does not permit as much power to be developed without excessive distortion as does operation with the load resistance called for by Eq. (23). The difference is particularly great in the case of pentodes and beam tubes.

[2] Equations (24) and (25) are approximate because they assume that any change in the average value of plate current produced as a result of distortion encounters the same load resistance as the alternating components of the plate current. With the amplifier circuits shown in Fig. 16, this is not the case, although the resulting error is small if the distortion is not excessive. The exact solution can be obtained either by modifying the load line, as described by C. E. Kilgour, Graphical Analysis of Output Tube Performance, *Proc. I.R.E.*, Vol. 19, p. 42, January, 1931, or by using the complete equivalent circuit based upon the power-series analysis of the vacuum-tube amplifier as described in Par. 24.

[3] G. S. C. Lucas, Distortion in Valve Characteristics, *Exp. Wireless and Wireless Eng.*, Vol. 8, p. 595, November, 1931.

A slightly different procedure giving the first four harmonics is described by A. Bloch, Distortion in Valves with Resistive Loads, Wireless *Engr.*, Vol. 15, p. 592, December, 1939.

An alternative method of analysis slightly more complicated and going up to the sixth harmonic is given by J. A. Hutcheson, Graphical Harmonic Analysis, *Electronics*, Vol. 9, p. 16, January, 1936.

The "direct-current component" A_0 of Eqs. (24*a*) and (25*a*) is the d-c component of the plate current when a signal is present. When there is distortion, this will usually differ from the value I_b of d-c plate current in the absence of signal. The increase of plate current (*i.e.*, $A_0 - I_b$) caused by the application of a signal is termed the "rectified" plate current, and is usually positive. The magnitude of the rectified current is usually a fair indication of the amount of distortion present.

Equations (24) and (25) give the distortion in the output of a power amplifier when the voltage applied to the grid is sinusoidal. When the grid is driven positive, the positive peaks of the exciting waves are flattened, thereby distorting the applied voltage, as shown in Fig. 19. The amount of distortion produced in this way depends upon the internal impedance of the exciter of the power amplifier tube, the grid current, etc. The harmonics produced in the wave applied to the grid as the result of grid current are very approximately

$$\begin{aligned} \frac{\text{Second harmonic}}{\text{Fundamental}} &= \frac{Z_s I_{gm}}{4E_s} \\ \frac{\text{Third harmonic}}{\text{Fundamental}} &= \frac{Z_s i_{gm}}{4E_s} \\ \frac{\text{Fourth harmonic}}{\text{Fundamental}} &= \frac{Z_s i_{gm}}{8E_s} \end{aligned} \qquad (26)$$

where Z_s = impedance seen by grid of the tube in looking back toward source of exciting voltage.
i_{gm} = peak value of grid current.
E_s = crest value of exciting voltage in the absence of grid current.

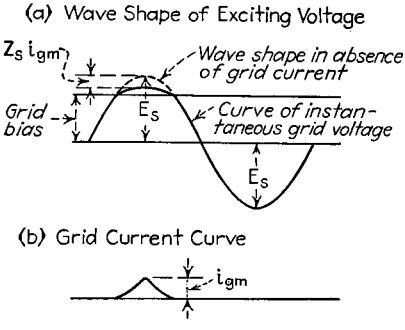


FIG. 19.—Wave shape of exciting voltage when grid is driven positive, showing flattening of the positive peaks resulting from grid current.

Selection of Operating Conditions for Class A Power Amplifiers.—Proper operating conditions for Class A operation of the commonly used power tubes are given in tube manuals, together with the amount of power that can be obtained and the accompanying distortion. It is accordingly necessary to determine operating conditions only in the case of new tubes, or tubes for which complete data are not available, or for tubes operating under special conditions.

Optimum operating conditions for triodes are selected as follows: Assume that the plate dissipation is the limiting factor, and choose a grid bias such that the d-c plate current has the maximum permissible value at the rated plate voltage. A reasonable value for $I_{\min}$ is then selected, such as one-tenth to one-fifth of the plate current at the operating point. Then to a good approximation, $I_{\max} = 2I_b - I_{\min}$. The minimum instantaneous plate potential $E_{\min}$ is then whatever value of plate voltage is required to give this value of $I_{\max}$ at the maximum potential to which it is desired to drive the grid. The corresponding maximum instantaneous plate potential is approximately $E_{\max} = 2E_b - E_{\min}$. With this information the proper load resistance and the power output can be calculated. A load line can then be drawn for this load resistance, giving the exact values of $I_{\min}$, $I_{\max}$, etc., and permitting calculation of the distortion. If the final results are not satisfactory, the operating conditions can then be modified as required. In particular, if the load resistance that this procedure leads to is less than twice the plate resistance of the tube, it will be found desirable to increase the negative bias on the grid until the load resistance called for is at least twice the plate resistance. This situation will sometimes be encountered with tubes operating at low plate voltages, *i.e.*, of the order of 250 volts.

With pentode and beam tubes the first step in selecting proper operating conditions is to decide upon a suitable plate current at the operating plate voltage, taking into account the rated plate dissipation. Values for I_{min} and I_{max} are then determined as in triodes, and the screen potential is given a value that causes the plate current to be I_{max} at the maximum grid potential desired. This maximum grid potential is nearly always either zero or only moderately positive, since there is enough distortion in pentodes and beam tubes even without introducing further distortion from grid current. The proper grid bias at the operating point is the bias that gives the operating plate current $(I_{max} + I_{min})/2$ at this screen voltage. A load line can then be drawn by selecting a load resistance such that I_{max} corresponds to a value of E_{min} sufficient to keep just above the knee of the $E_p - I_p$ curve. The exact behavior and distortion can be checked from this load line, and small readjustments of conditions made as required.

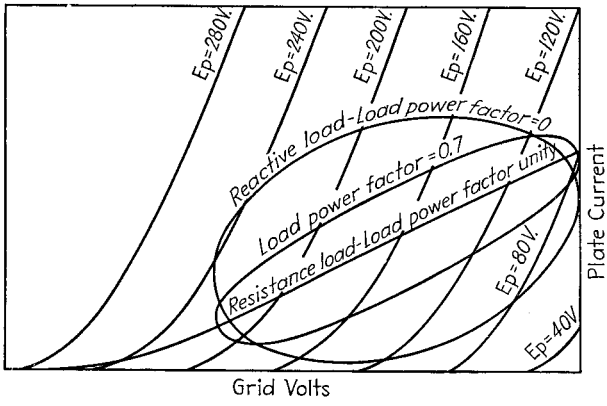


FIG. 20.—Dynamic characteristic for load impedances of varying power factors but constant magnitude, showing how reactive loads cause the dynamic characteristic to open up into an ellipse.

Reactive Loads.—When the load impedance has a reactive component, the dynamic characteristic (and also the load line) becomes a closed elliptical curve, as shown in Fig. 20.[1] Analysis is then most conveniently carried out by the power series method of Par. 24 rather than by the use of dynamic characteristics. With a reactive load, the power output obtainable is always less than with resistance load. When the load power factor is reasonably high, satisfactory results are obtained by using the same operating point as with resistance loads and making the magnitude of load impedance equal the proper resistance load. If, however, the power factor is quite low, the load impedance, bias, and applied signal may have to be modified for optimum performance.

8. Push-pull Amplifiers.—In the push-pull amplifier two tubes are arranged as shown in Fig. 16*c*, with the grids excited by equal voltages 180° out of phase and the outputs combined by means of a transformer having a center-tapped primary. With this arrangement even harmonics and even order combination frequencies cancel in the output, thereby permitting operation of triode and beam tubes in push-pull under conditions of high output per tube that would otherwise give excessive distortion. In addition, the push-pull arrangement avoids direct-current saturation in the core of the output transformer, because the d-c current in the two halves of the primary magnetize the core in opposite directions. There is also no a-c current of

[1] Manfred von Ardenne, On the Theory of Power Amplification, *Proc. I.R.E.*, Vol. 16, p. 193, February, 1928.

fundamental frequency flowing through the source of plate potential; thus minimizing regeneration in multistage amplifiers and making it unnecessary to employ a by-pass condenser across cathode biasing resistors. Hum caused by a-c filament current or ripple in the plate-supply voltage is also balanced out by the push-pull transformer connection.

These advantages of the push-pull arrangement are such that in using triodes or beam tubes, two small tubes are commonly used in a push-pull connection in preference to a single larger tube.

Class AB Push-pull Amplifiers.—The elimination of even harmonics by the push-pull connection makes it possible in triodes and beam tubes to extend operation until the instantaneous plate current is reduced to zero for a small portion of each cycle

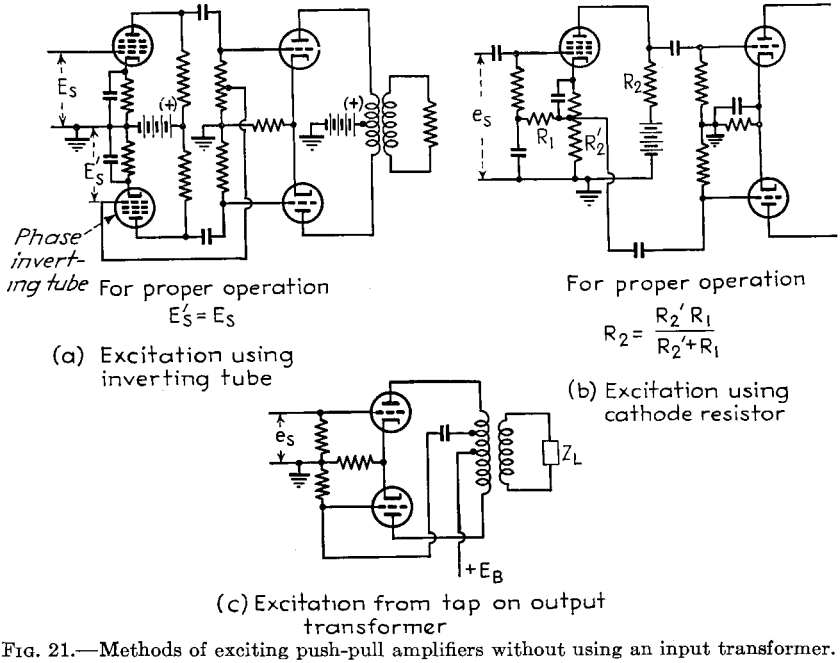


Fig. 21.—Methods of exciting push-pull amplifiers without using an input transformer.

without causing excessive distortion in the output. This is termed Class AB operation, and gives plate efficiencies of the order of 40 to 50 per cent.

Class AB operation is accompanied by a large increase in d-c plate current when the full signal voltage is applied. With fixed bias operation this is an advantage, since it permits the direct-current power input to the plate to be increased in the presence of an applied signal to a value exceeding the allowable plate dissipation. On the other hand, when self-bias is used, the full possibilities of Class AB operation cannot be realized, because if the bias keeps the plate current to an allowable value for no applied signal, the bias will be excessive when the d-c current increases in the presence of a large signal.

Excitation of Push-pull Amplifiers.—In a push-pull amplifier, the exciting voltages acting on the grids of the two tubes should be equal in magnitude and 180° out of phase. Such a voltage is usually obtained from an ordinary voltage amplifier by means of a center-tapped transformer, as shown in Fig. 16c, or by the use of an auxili-

ary *phase inverting tube*, as shown in Fig. 21*a*. Other possible means of excitation[1] are from a cathode resistor (Fig. 21*b*), by the use of a tap on the output transformer (Fig. 21*c*), and from the a-c voltage developed across the bias resistor of the push-pull stage (see Fig. 41*c*) or across a common resistor to ground in the plate circuit (R_3 in Fig. 41*d*).

Analysis of Push-pull Amplifiers.—The behavior of a push-pull amplifier can be determined with the aid of a load line plotted on composite curves, as shown in Fig. 22.[2] To obtain these composite curves, the plate voltage-current curves of the individual tubes are placed back to back as shown, with the common operating voltage superimposed. The composite characteristic curves are then derived by averaging the plate current for grid potential curves corresponding to the same applied signal. Thus the curves for the operating grid potential $E_c = 60$ in Fig. 22

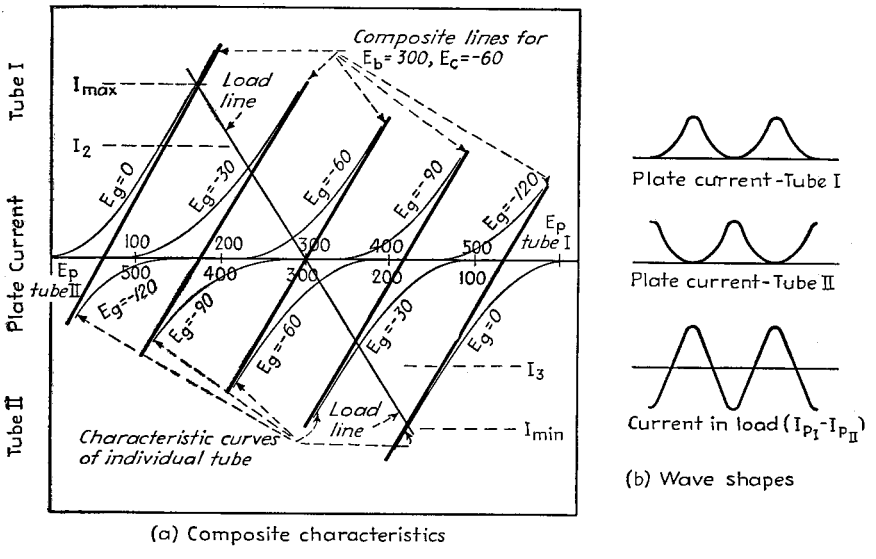


FIG. 22.—Composite characteristic curves of push-pull power amplifier, showing a load line together with the current waves for individual tubes.

are averaged because they correspond to no signal being applied, but when the signal is 30 volts on one tube it is −30 volts on the other; so one averages the grid curve for −30 volts on one tube with the curve for −90 volts on the other tube, etc. The resulting derived characteristics represent the relation between the plate-cathode voltage (which is one-half the plate-to-plate voltage) and the difference between the currents to the two anodes (which is twice the a-c current in a plate-to-plate load resistance); *i.e.*, the load line gives the voltages and currents in the output when the output transformer has a step-down ratio of 2.

The power output and distortion can be obtained in the usual way from load lines drawn on this diagram. These load lines must pass through the point on the composite characteristic corresponding to the operating plate and grid voltages, and they have a slope corresponding to one-fourth of the actual plate-to-plate load resistance. In making use of Eqs. (24) and (25) in connection with data derived from the dynamic

[1] See also Leonard Tulanskas, Bridge-type Push-pull Amplifiers, *Electronics*, Vol. 6, p. 134, May, 1933.

[2] B. J. Thompson, Graphical Determination of Performance of Push-pull Audio Amplifiers, *Proc. I.R.E.*, Vol. 21, p. 591, April, 1933.

characteristic, it is to be noted that the value of I_b used in the equations has the value zero, and that since I_{min} is in the lower half of the composite diagram, it is negative when substituted into the equations. In order to calculate the plate efficiency, it is necessary to know the d-c plate current existing in the presence of the signal. This depends upon the shape of the plate-current waves of the individual tubes, which can be derived by obtaining the instantaneous plate voltage for any given instantaneous grid voltage using the load line on the composite characteristic, and then, with the aid of the characteristic curve of the tube, finding the resulting plate-current wave. Waves for a typical case of Class AB operation are shown in Fig. 22.

When the instantaneous plate current of the individual tube goes to zero, as in Class AB operation and, as is often the case, in Class A push-pull amplifiers as well, it is possible to draw an approximate load line without the necessity of first obtaining a composite characteristic.[1] This is done by noting that the load line crosses the zero plate-current axis at a potential corresponding to the operating plate voltage and also by noting that since at the peak of the exciting voltage only one tube carries current, the composite characteristic at the extreme end of the load line coincides with the actual characteristic curve of the individual tube (see Fig. 22). This determines two points on the load line, the upper half of which can then be drawn on the $E_p - I_p$ curves of the individual tube, as in Fig. 23. In this way, it is possible to determine the value of

$$I_{max},\ I_{min}\ (= -I_{max}),\ E_{min},$$

and

$$E_{max}\ (= 2E_b - E_{min})$$

corresponding to a given load resistance and instantaneous peak voltage on the grid, or one can determine the values of E_{min}, E_{max}, and load resistance corresponding to a desired I_{max} with a given instantaneous peak grid voltage.

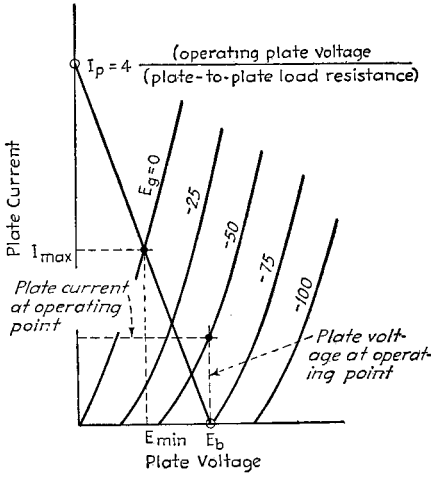


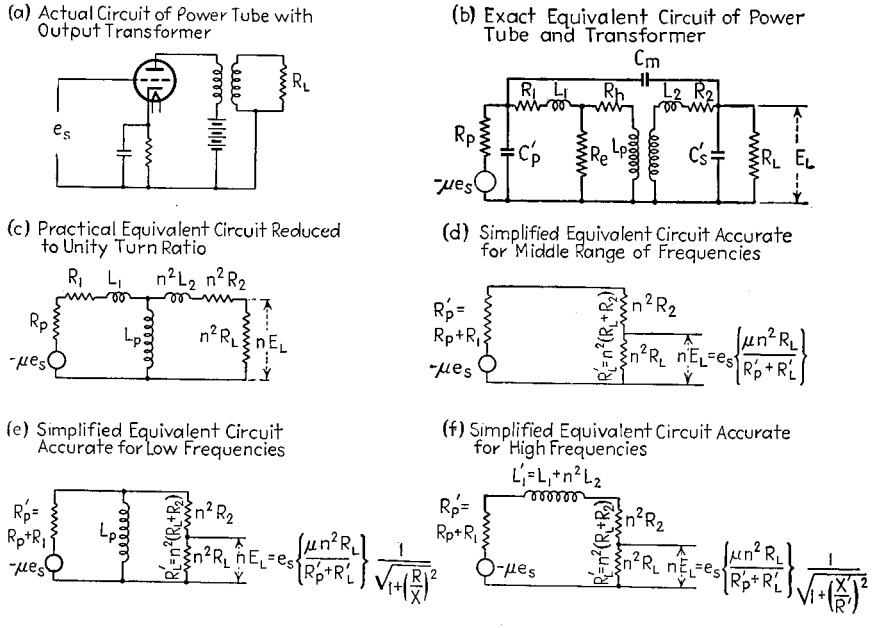
FIG. 23.—Upper half of load line for push-pull amplifier where the minimum plate current of the individual tube reaches zero. This corresponds to the part of the load line above the horizontal axis in Fig. 22.

9. Output Transformers and Other Coupling Arrangements for Class A Power Amplifiers.—The load impedance of a Class A power amplifier is normally coupled to the plate circuits of the tube by means of a transformer, as shown in Figs. 16*b* and 16*c*. This makes it possible, by the use of the proper turn ratio, to make any given load offer the desired impedance to the plate circuit of the amplifier.

Frequency Response Characteristics of Power Amplifiers Using Output Transformers.[2] The variation of output voltage with frequency that results from the use of an output transformer can be analyzed with the aid of the equivalent circuits of Fig. 24. The exact equivalent circuit shown at (*b*) can for most practical purposes be simplified by neglecting the capacities associated with the transformer and also the eddy-current and hysteresis losses, giving the result at (*c*) when reduced to unity turn ratio. Further simplifications, as shown at (*d*), (*e*), and (*f*), are possible when only limited

[1] E. W. Houghton, Class AB Push-pull Calculations, *Electronics*, Vol. 10, p. 18, June, 1937.
[2] F. E. Terman and R. R. Ingebretsen, Output Transformer Response, *Electronics*, Vol. 9, p. 30, January, 1936.

frequency ranges are considered. These equivalent circuits also apply to push-pull Class A and Class AB amplifiers, provided that the plate resistance and amplification factor used in the circuits and in the equations based on the circuits are taken as twice the values for a single tube.



μ = amplification factor of tube
R_p = plate resistance of tube (in a push-pull amplifier R_p is twice the plate resistance of one tube)
R_1 = d-c resistance of primary winding
$R'_p = R_p + R_1$ = effective plate resistance
R_2 = d-c resistance of secondary winding
R_L = load resistance
$R'_L = n^2(R_L + R_2)$ = effective load resistance reduced to unity turn ratio
$R = R_L R'_p / (R_L + R'_p)$ = resistance formed by R_p and R'_L in parallel
$R' = R_1 + R_p$ = sum of effective load and effective plate resistances
n = step down voltage ratio = ratio of primary to secondary turns
L_p = primary inductance with appropriate d-c saturation
L_1 = leakage inductance of primary winding
L_2 = leakage inductance of secondary winding
$L'_1 = L_1 + n^2 L_2$ = total leakage inductance reduced to unity turn ratio
$X = \omega L_p$ = reactance of transformer primary inductance
$X' = \omega L'_1$ = reactance of transformer leakage inductance
e_s = input voltage
E_L = output voltage

FIG. 24.—Equivalent circuits of power amplifier using output transformer.

In the middle range of frequency the output voltage of the transformer-coupled amplifier is

$$E_L = E_g\left(\frac{\mu n R_L}{R'_p + R'_L}\right) \tag{27}$$

The notation is shown in Fig. 24. At low frequencies the output falls off because of the shunting reactance of the primary inductance according to the relation

$$\text{Output voltage at low frequencies} = \frac{\text{Output voltage in middle-frequency range}}{1 - j\frac{f_1}{f}} \tag{28}$$

where f is the actual frequency, and f_1 the frequency at $R = X$. *The output voltage drops to 70.7 per cent of the mid-frequency value (output down 3 db) at the frequency* f_1

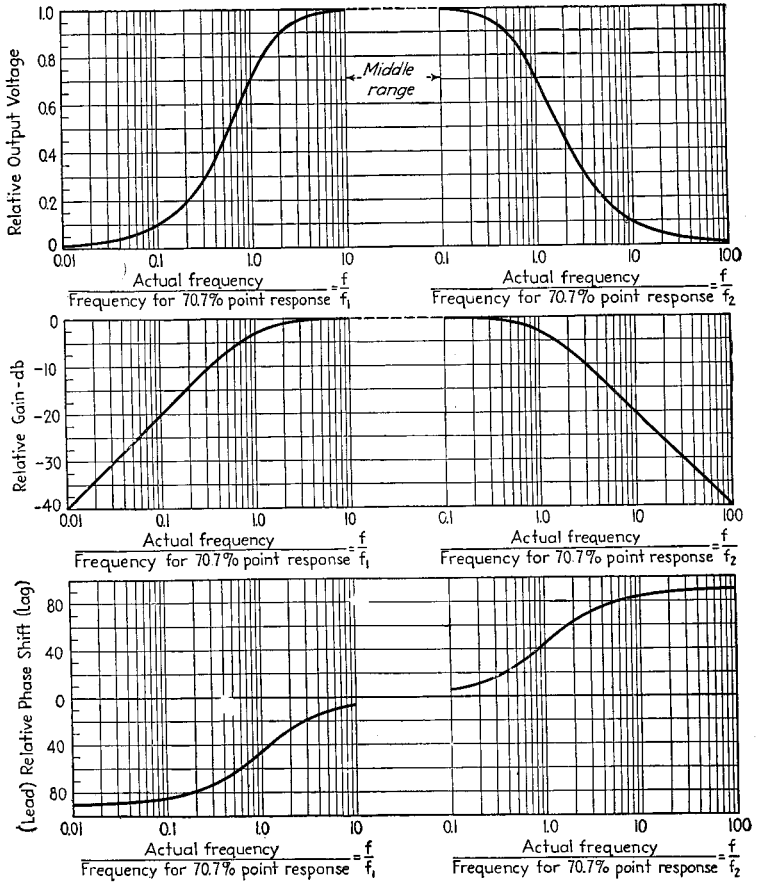


FIG. 25.—Universal amplification curve of transformer-coupled power amplifier.

for which the primary reactance X equals the resistance R formed by the effective load and effective plate resistances in parallel. At high frequencies, the output falls off because of the leakage reactance of the transformer, according to the equation

$$\text{Output voltage at high frequencies} = \frac{\text{Output voltage in middle-frequency range}}{1 + j\frac{f}{f_2}} \tag{29}$$

where f_2 is the frequency at which $X' = R'$. *The 70.7 per cent response point (output*

down 3 db) occurs at a frequency f_2 such that the leakage reactance equals the sum of the effective load and effective plate resistances.

The relative magnitude and phase of the output voltage of a power amplifier using an output transformer can be readily obtained with the aid of the universal amplification curve given in Fig. 25.

In pentode and beam tubes the plate resistance is so high that the falling off in amplification calculated from Eq. (29) does not become appreciable until the frequency is very high. Under these conditions the capacities shown in the exact equivalent circuit of Fig. 24*b* may be of importance.

Transformer Characteristics.—For uniform amplification over a wide frequency range, the output transformer should have a high primary and low leakage induct-

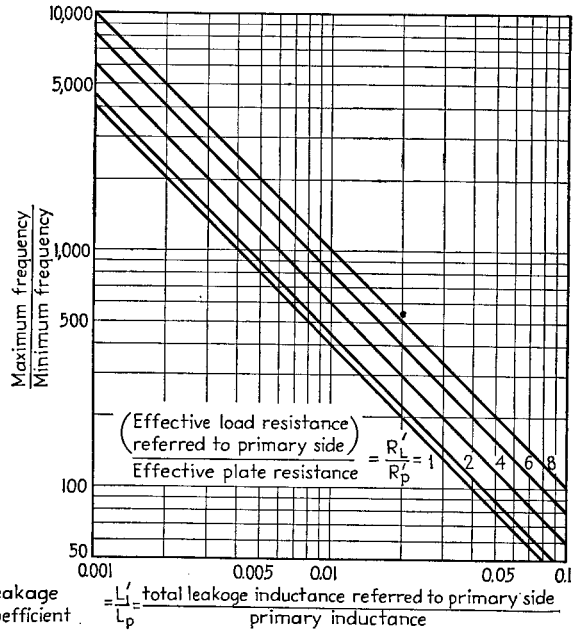


FIG. 26.—Curves giving frequency range of transformer-coupled power amplifier as a function of the leakage coefficient of the transformer.

ance. The ratio of the highest to lowest frequency over which the amplification is within 3 db of the mid-range value is given in Fig. 26 as a function of the leakage coefficient L_1'/L_p. The factors contributing to a small leakage coefficient are (1) core assembly, with a minimum possible air gap consistent with the d-c magnetization present, (2) the use of alloy cores, and (3) winding arrangements that minimize leakage inductance. In the case of transformers for push-pull operation it is also desirable that the leakage inductance between the two halves of the primary be small, as should also the leakage inductance from each half of the primary to the secondary.

The step-down ratio n for an output transformer should approximate a value such that n^2 times the load resistance equals the resistance to be presented to the tube. Actually, the turn ratio should be slightly less because of the resistances of the transformer windings. When these resistances are taken into account, the resistance offered by the transformer primary to the amplifier tube in the middle-frequency range is

$$\left.\begin{matrix}\text{Impedance offered}\\ \text{to tube}\end{matrix}\right\} = n^2R_L + (n^2R_2 + R_1) \tag{30}$$

The notation is illustrated in Fig. 24.

An output transformer consumes a fraction of the power that the tube delivers to it as a result of the winding resistances and the core losses. Practically, the primary and secondary resistances account for virtually all the power loss except at quite low frequencies. To the extent that the core losses can be neglected, one has

$$\frac{\text{Power delivered to load}}{\text{Power delivered to transformer}} = \frac{1}{1 + \dfrac{R_1 + n^2R_2}{n^2R_L}} \tag{31}$$

The maximum current that an output transformer can carry is determined by the heating of the windings, while the maximum voltage that can be applied is limited by the permissible flux density in the core. These two factors operating together determine the power rating of the output transformer.

The relationship between the applied voltage and flux density in the core is given by the equation

$$\left.\begin{matrix}\text{Effective value of}\\ \text{applied voltage}\end{matrix}\right\} = 4.44fNBA \times 10^{-8} \tag{32}$$

where N = number of turns.
f = frequency.
B = flux density in the core in lines per unit area.
A = net area of core.

The voltage producing a given flux density is proportional to frequency, so that the voltage limit occurs at low frequencies. The allowable flux density is limited by the fact that with high densities the nonlinearity of the magnetization curve introduces distortion, and depends upon the amount of distortion that can be tolerated, the magnetic characteristics of the core material, the length of the air gap, the circuits associated with the transformer, and whether or not d-c magnetization is present. In general, the higher the flux density, the greater the d-c magnetization, and the smaller the air gap in the core, the greater will be the distortion. In ordinary output transformers operated within their power rating it is commonly found that distortion due to nonlinearity of the magnetization curve often becomes serious at frequencies for which the loss of gain is less than 3 db.

Coupling Systems Employing Parallel Feed.—The term *parallel feed* (or shunt feed) is used to denote systems in which a separate choke is provided for carrying the d-c plate current. Examples of such arrangements are shown in Figs. 16*a* and 16*d*. In such circuits the shunt-feed choke is sometimes referred to as a *retard coil.*

The direct-coupled shunt-feed circuit of Fig. 16*a* is characterized by unusually good high-frequency response, with the point at which the response drops to 70.7 per cent of the mid-range value occurring when the reactance of the shunting capacity is equal to the equivalent resistance formed by the load resistance in parallel with the plate resistance of the tube. At low frequencies the amplification characteristic depends upon the inductance L and capacity C of the coupling network, and upon the ratio of plate to load resistance. Results in critical cases are shown in Fig. 27*a*. When the coupling condenser C is extremely large, the amplification falls off at low frequencies in the same way as with simple transformer coupling, with the 70.7 per cent point occurring at the frequency for which the reactance of the inductance L equals the resistance formed by the load and plate resistances in parallel. When the ratio of plate to load resistance is somewhat less than unity, as is the case with triode

tubes, a suitably chosen finite blocking-condenser capacity gives approximately the same falling off as does the infinite condenser. On the other hand, when the plate resistance of the tube is very high compared with the load resistance (corresponding to pentode tubes), Fig. 27*a* shows that substantially constant amplification can be

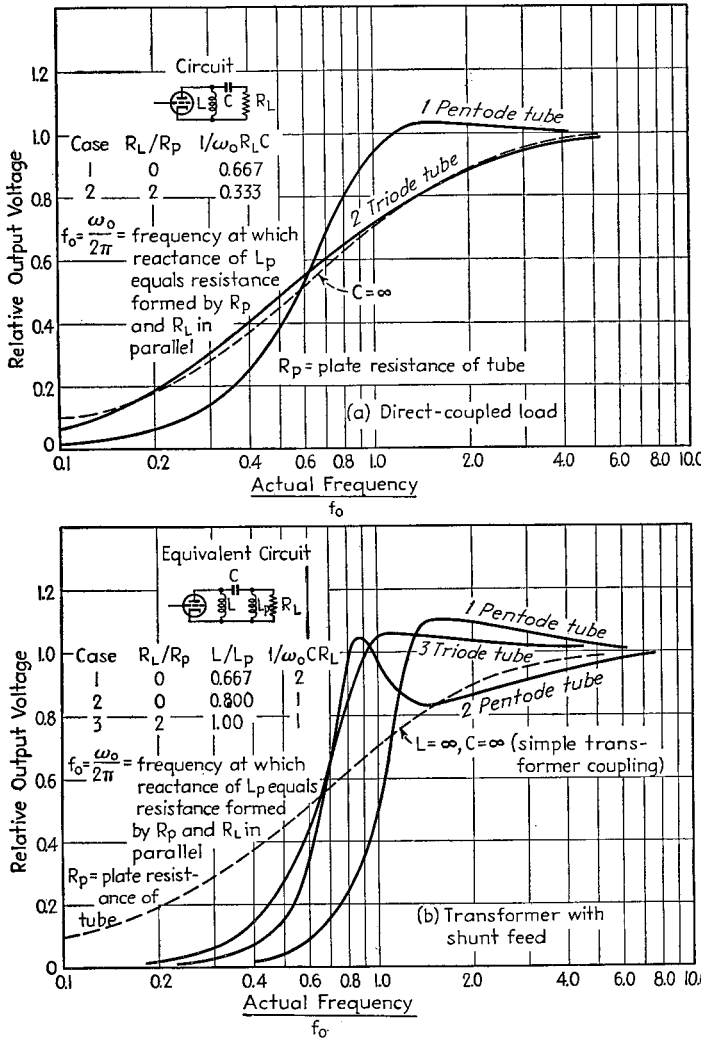


FIG. 27.—Characteristics obtainable at low frequencies by properly designed shunt-feed coupling systems.

maintained down to the frequency at which the reactance of the choke equals the load resistance, provided that the blocking capacity is such that its reactance at this same frequency is two-thirds the load resistance. The cutoff at lower frequencies is then quite sharp, however.

Transformer coupling with shunt feed gives the possibility of a better frequency response than is obtainable with a simple transformer. This is because the elimina-

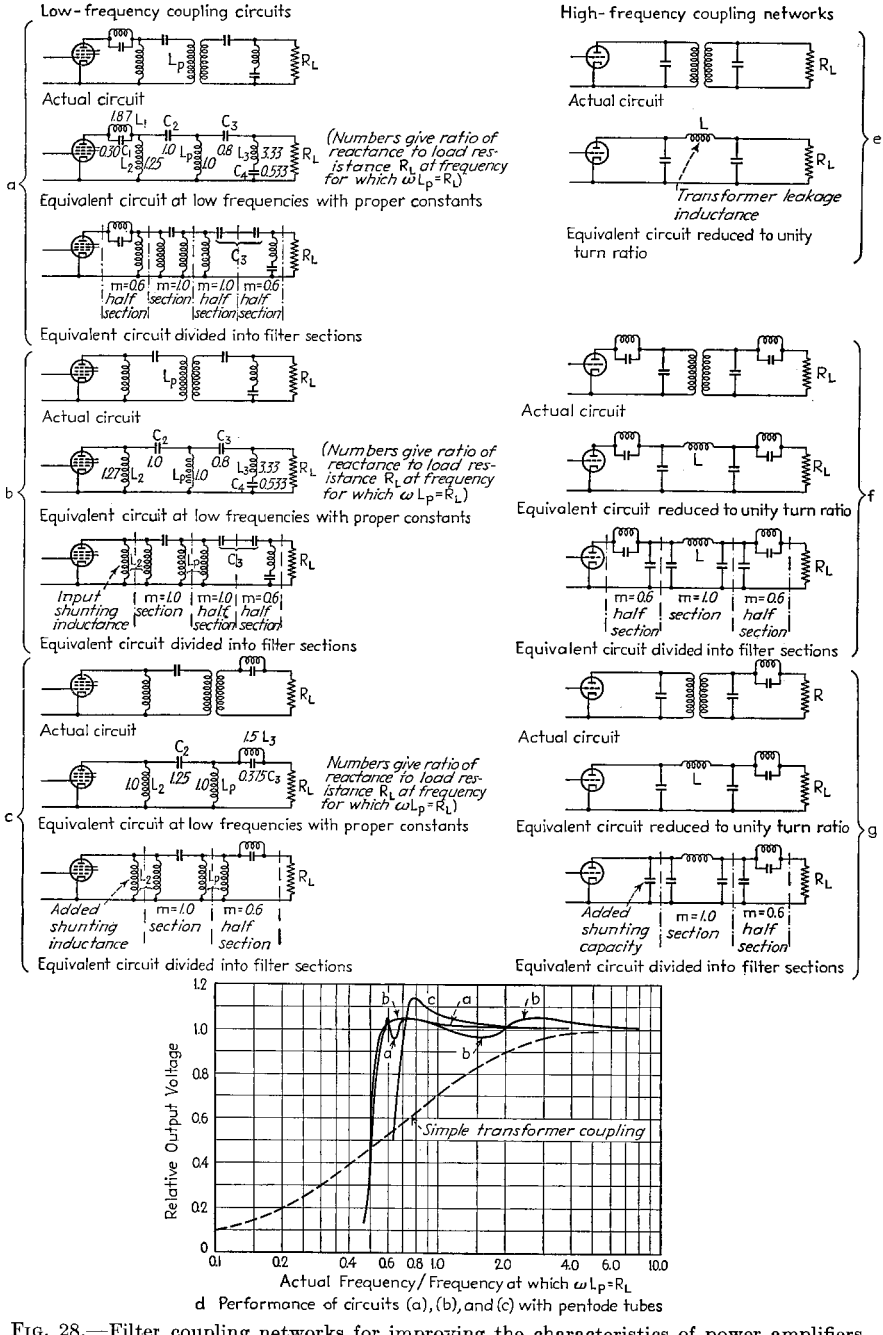


FIG. 28.—Filter coupling networks for improving the characteristics of power amplifiers, together with behavior of various low-frequency coupling systems.

tion of the direct-current magnetization from the transformer core permits the core to be assembled with a smaller air gap, and also makes it permissible to use high permeability core material. This reduces the ratio of leakage to primary inductance of the transformer, and so increases the frequency range that can be covered. The low-frequency characteristics obtained with an output transformer having shunt feed depend upon the ratio of plate to load resistance and the inductance and capacity of the shunt-feed inductance and blocking condenser, respectively. The behavior in the critical cases is shown in Fig. 27*b*. With pentode tubes, a properly selected finite shunt-feed choke and blocking condenser will give a substantially better performance than the same transformer operated with an extremely large condenser and infinite shunt-feed choke inductance (*i.e.*, simple transformer coupling having the same transformer primary inductance L_p).

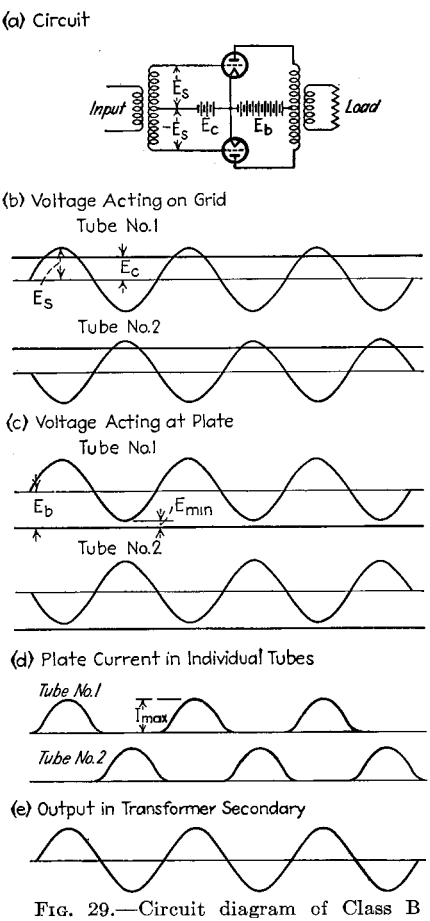


FIG. 29.—Circuit diagram of Class B audio amplifier together with oscillograms showing voltage and current relations and showing how the pulses of plate current in the individual tubes combine to produce a sinusoidal output.

Shunt-feed systems such as shown in Fig. 27 give appreciably sharper low-frequency cutoff than is obtained with simple transformer coupling.

Filter Coupling Systems.[1]—The characteristic of a coupling system involving an output transformer can be improved by incorporating the primary and leakage inductance of the transformer into a coupling network designed according to filter theory. The benefits obtained in this way are primarily an extension of the low-frequency response when pentode tubes are employed and an extension of the high-frequency response when triode tubes are used. There is also an improvement of the impedance characteristic at extreme frequencies, which increases the power-handling capacity of the amplifier under these conditions.

Filter-type coupling systems for low frequencies, together with circuit constants, are shown in Fig. 28. The circuit of Fig. 28*a* consists of ordinary filter sections arranged as shown, with the primary inductance of the transformer functioning as the full shunt inductance of an inverse-network (or constant k) section of a filter that is provided with terminating half sections. With pentode tubes, such an arrangement gives a substantially constant response down to half the frequency at which the response would be 70.7 per cent in the case of a simple transformer with the same primary inductance. With triode tubes there is little if any improvement in the

[1] Much of this material is from the thesis by C. A. Moreno, Circuit Design to Improve the Frequency Response of Output Transformers, Stanford University, 1940.

low-frequency response, but a better impedance characteristic is realized at low frequencies. The network of Fig. 28*a* can be simplified as shown in Fig. 28*b*, with little loss in frequency characteristic, although at the expense of a less satisfactory impedance at low frequencies. The simplifications involved consist primarily in omitting the input terminating half section and then modifying the shunting inductance on the input side. The resulting frequency characteristic for the case of pentode tubes is shown in Fig. 28*d*. Still further simplification of the network is possible as in Fig. 28*c*, although at the expense of a low-frequency cutoff about 25 per cent higher than for the circuit of Fig. 28*b*.

Filter type of coupling networks for high frequencies are shown in Figs. 28*e* to 28*g*. These are the high-frequency analogues of the low-frequency coupling systems discussed above. With triode tubes these arrangements offer the possibility of an appreciable improvement in the high-frequency response obtainable with a given leakage inductance.

Filter coupling systems are useful under circumstances where it is otherwise difficult to obtain the required frequency response or where a very sharp cutoff is wanted or where it is important to maintain normal impedance relations at extremes in frequency. They are not otherwise justified because of the expense of the coupling network and the necessity of adjusting its constants accurately to specified values.

10. Class B Audio-frequency Power Amplifiers.[1]—The Class B audio-frequency amplifier is a push-pull amplifier in which the tubes are biased approximately to cutoff. Operated in this way, one of the tubes amplifies the positive half cycles of the signal voltage while the other amplifies the negative half cycles, with the output transformer combining these in the output, as shown in Fig. 29. Such an amplifier is characterized by high plate efficiency and relatively high output in proportion to the average dissipation in the tubes.

The output power, plate efficiency, and proper load resistance in a Class B amplifier can be determined with an accuracy sufficient for ordinary purposes by assuming that the characteristic curves of the tube are straight lines. This simplifying assumption gives

$$\left.\begin{matrix}\text{Proper load resistance}\\ \text{from plate to plate}\end{matrix}\right\} = R_L = 4\,\frac{E_b - E_{\min}}{I_{\max}} \tag{33}$$

$$\left.\begin{matrix}\text{Power output}\\ \text{from two tubes}\end{matrix}\right\} = \frac{I_{\max}(E_b - E_{\min})}{2} \tag{34}$$

$$\text{Plate efficiency} = \frac{\pi}{4}\left(1 - \frac{E_{\min}}{E_b}\right) \tag{35}$$

Here $I_{\max}$ is the peak plate current and $E_{\min}$ the minimum plate potential reached in an individual tube during the cycle, and E_b is the plate supply voltage. Exact determination of power output, efficiency, and distortion can be carried out with the aid of a load line drawn on composite characteristic curves of the push-pull combination exactly as described in connection with Fig. 22.

The maximum possible efficiency as given by Eq. (35) is 78.5 per cent. Under practical conditions, efficiencies of the order of 50 to 60 per cent can be realized.

The distortion in Class B audio amplifiers is minimized by employing a bias corresponding to the cutoff bias that would be obtained if the main part of the $I_p - E_g$ curve for the operating plate potential were projected as a straight line, as shown in Fig. 30. Such a bias is termed *projected cutoff*, and gives an output voltage that is a

[1] L. E. Barton, High Audio Output from Relatively Small Tubes, *Proc. I.R.E.*, Vol. 19, p. 1131, July, 1931; Recent Developments of the Class B Audio- and Radio-frequency Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 985, July, 1936; True McLean, An Analysis of Distortion in Class B Audio Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 487, March, 1936.

substantially distortionless reproduction of the voltage acting on the grid, provided that operating conditions are such that E_{min} is not too low. Since projected cutoff is less than real cutoff, such operation causes the efficiency to be slightly lower than for true Class B operation, although much higher than for ordinary Class AB operation.

Class B audio-frequency amplifiers are ordinarily operated so that the grid is driven positive when the applied signal is sufficient to develop full output. This tends to distort the exciting voltage acting on the grids of the tubes. When operation is such that the grid is driven positive only at the peak of the cycle, the situation is the same as discussed in connection with Eq. (26). In some types of Class B amplifiers the tubes are operated with zero bias. Under these conditions the exciter must have sufficient power capacity to supply the energy dissipated at the grid of the Class B tubes, and must also have sufficiently low internal impedance so that the variations in instantaneous impedance offered by the grids during the cycle do not produce excessive distortion of the exciting voltage. In order to achieve this it is commonly necessary to couple the exciter to the grids of the push-pull Class B amplifier by means of a step-down transformer.

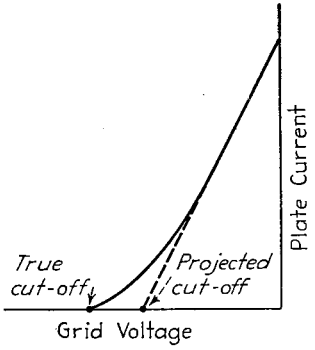


FIG. 30.—Curve illustrating meaning of projected cutoff.

The plate- and bias-supply voltages must have good regulation if best results are to be obtained in Class B operation. This is because the grid and plate currents vary with the average amplitude of the exciting voltage, and if the voltage regulation is poor the operating conditions will vary according to the amplitude of the applied signal. Regulation is much more important when the amplifier operates with a bias than when zero bias operation is used. In high-power amplifiers special compensating systems to improve the voltage regulation are often desirable.[1]

The frequency-response characteristics of Class B amplifiers depend upon the output transformer in much the same way as in Class A amplifiers. Thus the falling off in amplification at low frequencies is determined by the primary inductance of the output transformer, while the leakage inductance controls the falling off in gain at high frequencies. The exact relationships existing in Class B operation are much more complicated than in Class A amplifiers, however, because the plate currents of the individual tubes contain numerous harmonics. In particular, the intermittent character of the plate currents of the individual tubes in a Class B amplifier makes it important that the leakage inductance from one-half of the primary winding to the other half be extremely small. If this is not the case, transients will be produced in the primaries that will produce serious distortion.[2]

Class B amplifiers practically always employ triode tubes. The same tubes used in Class A and Class AB audio amplifiers and in Class C radio-frequency amplifiers are also suitable for Class B operation.[3] In addition there are special tubes designed for Class B operation at plate voltages of the order of 400, and requiring zero grid bias. These tubes have such a high amplification factor that "projected cutoff" is approximated by zero bias, and are characterized by unusually small grid current

[1] R. J. Rockwell and G. F. Platts, Automatic Compensation for Class B Bias and Plate Voltage Regulation, *Proc. I.R.E.*, Vol. 24, p. 553, April, 1936.

[2] A. Pen-Tung Sah, Quasi Transients in Class B Audio-frequency Push-pull Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 1522, November, 1936.

[3] I. E. Mouromtseff and H. N. Kozanowski, Comparative Analysis of Water-cooled Tubes as Class B Audio Amplifiers, *Proc. I.R.E.*, Vol. 23, p. 1224, October, 1935.

with positive grid voltage, and a grid-current characteristic that is reasonably linear.[1]

AUDIO-FREQUENCY AMPLIFIERS—MISCELLANEOUS

11. Feedback Amplifiers.[2]—In the feedback amplifier, a voltage derived from the amplifier output is superimposed upon the amplifier input in such a way as to oppose the applied signal in the normal frequency range (see Fig. 31). When this is properly done, the properties of the amplifier are modified in ways that are desirable for many purposes.

Effect of Feedback on Gain, Distortion, and Noise.—The amplification in the presence of feedback is given by the relation

$$\left.\begin{array}{l}\text{Voltage amplification}\\ \text{taking into account feedback}\end{array}\right\} = \frac{A}{1 - A\beta} = -\frac{1}{\beta}\frac{1}{1 - \frac{1}{A\beta}} \qquad (36)$$

where A = amplification in the absence of feedback.[3]

β = fraction of the output voltage that is superimposed upon the amplifier input.

In this equation, the signs are so chosen that when the signal voltage is opposed by the voltage fed back, β is negative.

The quantity $A\beta$ can be termed the *feedback factor*. It represents the ratio of the voltage fed back and superimposed on the applied signal to the net voltage acting on the input of the amplifier (net input = applied signal plus feedback voltage). When the feedback factor $A\beta$ is large, Eq. (36) shows that the amplification becomes the reciprocal of the fraction β of the output voltage that is superimposed upon the amplifier input, *and is substantially independent of the characteristics of the amplifier itself*. When the feedback voltage is developed by a resistance network, β has zero phase angle and is independent of frequency. Under these conditions, the amplification will be substantially independent of frequency and will have negligible phase shift for all frequencies for which $A\beta$ is large. On the other hand, when the amplification is to vary with frequency in some particular way, this can be accomplished by making the feedback or β circuit have the same transmission *loss* characteristic as the desired *gain* characteristic.

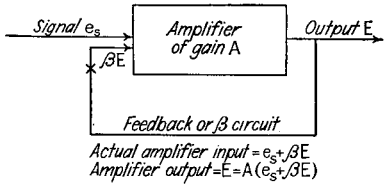


FIG. 31.—Schematic diagram of feedback amplifier.

Feedback reduces the nonlinear distortion produced in the amplifier for a given output according to the relation[4]

$$\left.\begin{array}{l}\text{Distortion}\\ \text{with feedback}\end{array}\right\} = \frac{\text{Distortion in absence of feedback}}{1 - A\beta} \qquad (37)$$

[1] See L. E. Barton, Application of Class B Audio Amplifier to A-c Operated Receivers, *Proc. I.R.E.*, Vol. 20, p. 1085, July, 1931.

[2] H. S. Black, Stabilized Feed-back Amplifiers, *Elec. Eng.*, Vol. 53, p. 114, January, 1934; also, U.S. Patent 2,102,671, Dec. 21, 1937, to H. S. Black.

[3] The symbol μ is frequently used in the literature of feedback to represent amplification, but has here been replaced by A in order to avoid confusion with amplification factor.

[4] This assumes that the amplifier produces no distortion when reamplifying the distortion voltages fed back by the β circuit, and is quite accurate if the distortion in the absence of feedback is not large. For an exact analysis without this limitation, see Robert W. Sloane, Distortion in Negative Feedback Amplifiers, *Wireless Eng.*, Vol. 14, p. 259, May, 1937; also, p. 369, July, 1937; A. C. Bartlett, A Note on Negative Feedback, *Wireless Eng.*, Vol. 15, p. 90, February, 1938.

If $A\beta$ is made large by employing a large amount of feedback, the result is a great percentage reduction in distortion and cross-modulation.

The signal-to-noise ratio is modified by feedback according to the relation

$$\frac{\text{Signal-to-noise ratio with feedback}}{\text{Signal-to-noise ratio without feedback}} = \frac{a_f}{a_0(1 - A\beta)} \tag{38}$$

where a_f and a_0 are the amplification between the place at which the noise is introduced and the output, with and without feedback, respectively. This equation assumes that the amount of noise introduced is the same both with and without feedback and that the output voltages are likewise the same in both cases. Feedback will greatly improve the signal-to-noise ratio when the noise is introduced in high-level parts of the amplifier, such as noise from a poorly filtered power supply in the plate circuit of the power tube.

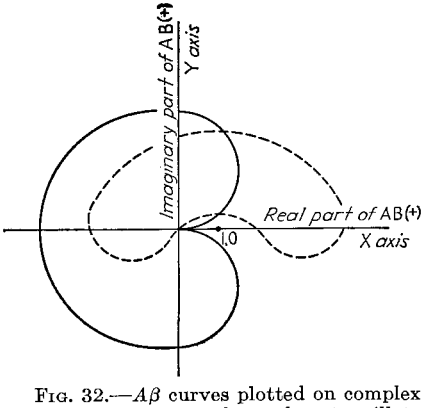


FIG. 32.—$A\beta$ curves plotted on complex plane. The two cases shown do not oscillate because the curves do not inclose the point 1, 0.

Feedback without Oscillations.[1]—The criterion for avoiding oscillations in feedback amplifiers is that when the real and imaginary parts of the feedback factor $A\beta$ and its conjugate are plotted on rectangular coordinates for frequencies from zero to infinity, with the real part along the X axis and the imaginary part along the Y axis, as shown in Fig. 32, the resulting curve must not inclose the point $X = 1$, $Y = 0$. With most amplifier circuits, this means that the feedback factor $A\beta$ should be negative in the mid-frequency range and that at high and low frequencies such that the phase of $A\beta$ is shifted $\pm 180°$ from the mid-frequency value, the magnitude of $A\beta$ must be less than unity.[2]

The fundamental factors required in the design of the feedback loop to achieve stability are discussed at length in Par. 27, Sec. 3. Briefly, the phase shift of the feedback loop (*i.e.*, phase of $A\beta$) depends upon the rate at which the magnitude of $A\beta$ varies with frequency. The more rapidly the transmission (*i.e.*, the magnitude of $A\beta$) varies with frequency, the greater tends to be the phase shift. Hence, when the phase shift is to be kept below 180°, the magnitude of $A\beta$ must not be allowed to vary too rapidly. Thus it is necessary in feedback amplifier systems to pay particular attention to the way in which the system behaves at frequencies outside the useful range, since it is here that $A\beta$ drops off in magnitude and hence it is here that large phase shifts tend to be encountered.

In the case of resistance coupling and output transformer coupling when no additional phase shift is introduced from the bias and screen impedances, the maximum phase shift that a single-stage amplifier can possibly produce will not exceed 90°.

[1] H. Nyquist, Regeneration Theory, *Bell System Tech. Jour.*, Vol. 11, p. 126, January, 1932; E. Peterson, J. G. Kreer, and L. A. Ware, Regeneration Theory and Experiment, *Proc. I.R.E.*, Vol. 22, p. 1191, October, 1934; D. G. Reid, The Necessary Conditions for Instability (or Self-oscillation) of Electrical Circuits, *Wireless Eng.*, Vol. 14, p. 588, November, 1937.

[2] The rather special case illustrated by the dotted curve in Fig. 32 violates this rule but still does not give oscillations, because the curve does not inclose the point 1, 0. Such an arrangement will, however, oscillate when the gain is reduced. Accordingly, when such an amplifier is first turned on, it will always break into oscillation as the filaments warm up and the gain is less than normal. If, however, these oscillations are momentarily stopped, as by short-circuiting one of the stages, they will not restart after the short circuit is removed.

Even with additional phase shift from bias and screen impedances, or from resonances in the coupling system (as, for example, at low frequencies in the case of a resonated primary transformer or at high frequencies in the case of an interstage transformer having leakage inductance and distributed capacity), it is still rather unusual for the phase shift of a single stage amplifier to reach 180° before the magnitude of $A\beta$ has become less than unity.

With two-stage amplifiers it is still a relatively simple matter to avoid oscillation, but more care is needed in the design of the circuits. When the individual stages are such that the phase shift that each introduces does not exceed 90°, then any desired amount of feedback may be introduced by a resistance network without oscillations being produced. However, if additional phase shifts are present from the

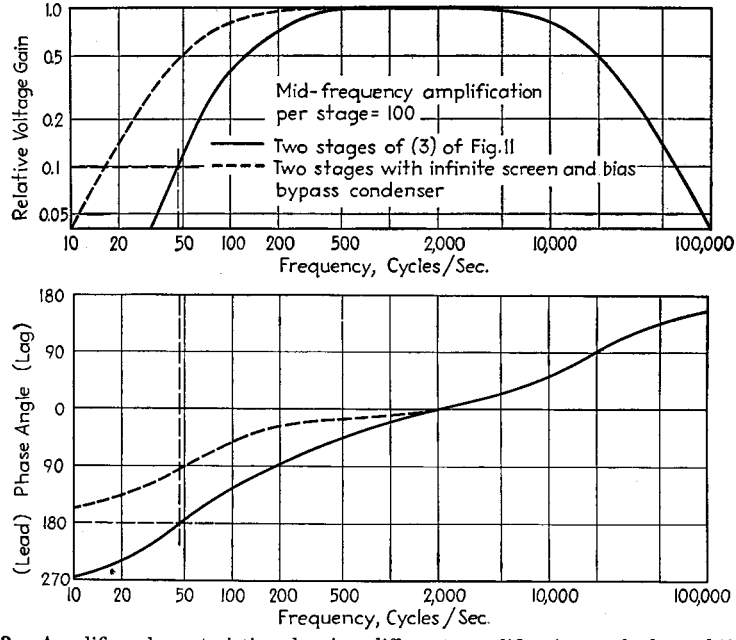


FIG. 33.—Amplifier characteristics showing different amplification and phase-shift characteristics at low frequencies.

coupling networks, from the screen and bias impedances, etc., then oscillation will occur in two-stage amplifiers when the feedback is large. This is illustrated by the solid curve in Fig. 33, where, with a two-stage amplifier of the type shown by curve 3 in Fig. 11, a 180° phase shift occurs at a frequency of 47.5 cycles, corresponding to an amplification that has dropped to 0.095 of the mid-range value. With a resistance network to develop feedback (*i.e.*, phase shift of β is zero at all frequencies), this requires that $A\beta < 10.5$ in the mid-frequency range if one is to have $A\beta < 1$ at the frequency of 47.5 cycles for which the phase shift is 180°. The corresponding value of β is 0.00105. This behavior is to be contrasted with the case of infinite bias and screen bypass condenser (shown dotted in Fig. 33) for which the phase shift does not reach 180° until the amplification has dropped to zero, so that oscillations are not possible under any conditions. If for some reason it is not possible to realize conditions corresponding to infinite bias and screen bypass condensers, and at the same time more feedback than that corresponding to $A = 10.5$ in Fig. 33 is desired, then the amplifier must be redesigned so that the amplification characteristic of one of the stages at low

frequencies is constant to a substantially lower frequency than the other stage. Doing so reduces the rate at which the amplification falls off, and so reduces the phase shift.

In amplifiers involving three or more stages the entire design revolves around controlling the amplification characteristics at high and low frequencies so that the rate at which the transmission $A\beta$ around the feedback loop falls off will not be so great as to lead to excessive phase shift and hence oscillation. When the three stages are identical, then even when the phase shift per stage has the minimum possible value

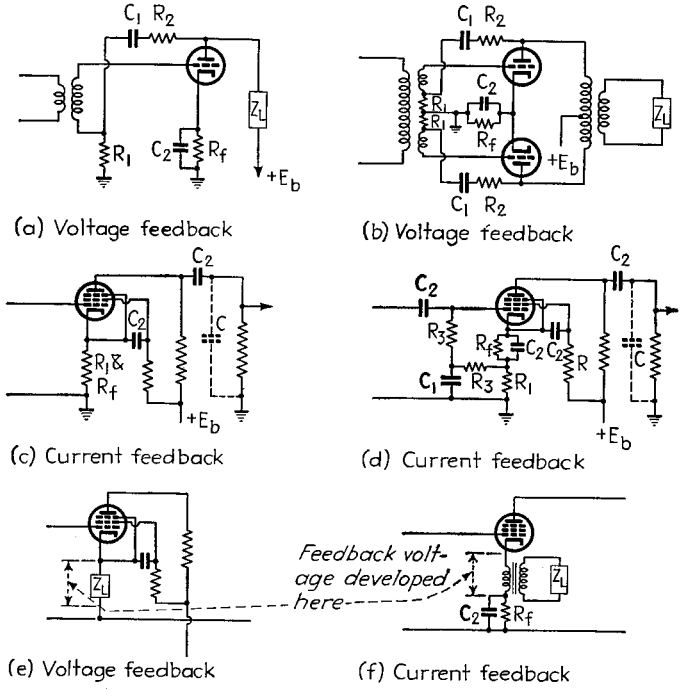


FIG. 34.—Typical feedback circuits involving a single stage of amplification.

of 90 degrees, it is necessary that $A\beta < 8$. In order to obtain a greater value of feedback it is necessary to reduce the rate at which the amplifier cuts off at high and low frequencies. A simple way of accomplishing this is to give one of the stages a flatter frequency response characteristic than the other two. Under these conditions the maximum value the feedback factor $A\beta$ can have, assuming a maximum possible phase shift of 90° per stage, is[1]

$$\left.\begin{matrix}\text{Maximum value}\\ \text{of feedback}\\ \text{factor } A\beta\end{matrix}\right\} = \frac{2(n+1)^2}{n} \tag{39}$$

$$\text{where for high frequencies } n = \frac{\left\{\begin{matrix}\text{frequency for 70.7 per cent response}\\ \text{of flat stage}\end{matrix}\right\}}{\left\{\begin{matrix}\text{frequency for 70.7 per cent response}\\ \text{of a less flat stage}\end{matrix}\right\}},$$

[1] Equation (39) was derived by Dr. R. R. Buss, former graduate student at Stanford.

and for low frequencies $n = \dfrac{\left\{\begin{matrix}\text{frequency for 70.7 per cent response}\\ \text{of a less flat stage}\end{matrix}\right\}}{\left\{\begin{matrix}\text{frequency for 70.7 per cent response}\\ \text{of flat stage}\end{matrix}\right\}}$.

When n is different at low and high frequencies, the smallest value is the one to use in Eq. (39). If the phase shift per stage exceeds 90°, or if β introduces a phase shift, the allowable mid-frequency feedback factor $A\beta$ will be less than given by Eq. (39).

An alternative procedure for designing an amplifier with three stages, and the procedure that must be followed if more than three stages are involved, is described in Par. 27, Sec. 3. This consists in designing the amplifier initially to give a wider frequency response band than is actually required and then in modifying this characteristic at high and low frequencies by placing expedients such as are illustrated in Fig. 93, Sec. 3, in the amplifier network or in the feedback connection.

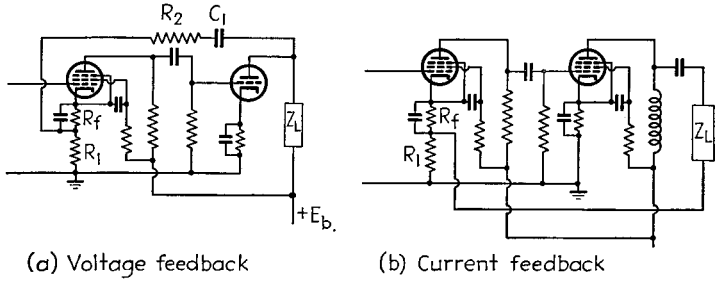


FIG. 35.—Typical feedback circuits involving two stages of amplification.

Practical Feedback Amplifiers.[1]—Typical feedback amplifier circuits are shown in Figs. 34 to 36. In these, a uniform notation is employed as follows:

R_1 = resistance across which feedback voltage is developed.

R_2 = resistance that, in conjunction with R_1, forms a voltage divider for making the feedback voltage the desired fraction of the output voltage.

R_3 = a high resistance for grid-leak or isolating purposes.

R_f = resistance across which bias voltage is developed.

C_1 = blocking condenser of such capacity as to introduce negligible reactance or phase shift at frequencies that are amplified appreciably.

C_2 = by-pass condenser.

When the feedback factor $A\beta$ is large, the feedback voltage developed across R_1 has a value almost the same in both magnitude and phase as the signal voltage applied to the amplifier input. This fact determines the way in which the feedback will affect the output voltage, since the output voltage must be whatever value is required to make the feedback voltage developed across R_1 approximate the applied signal.

The circuits of Figs. 34 to 36 can be divided into two general types, namely, those in which the feedback voltage is derived directly from the output voltage of the amplifier (*voltage feedback*), and those in which the feedback voltage is derived from the current flowing through the output of the amplifier (*current feedback*). The type of feedback involved is indicated on each of the circuits in Figs. 34 to 36. The effect of feedback upon the amplifier behavior is somewhat different with the two types of feedback. Thus voltage feedback in connection with an output transformer, as shown in Fig. 37*a*, makes the voltage across the primary reproduce accurately the signal voltage, and so improves the frequency response at low frequencies and

[1] F. E. Terman, Feedback Amplifier Design, *Electronics*, Vol. 10, p. 12, January, 1937; J. R. Day and J. B. Russell, Practical Feedback Amplifiers, *Electronics*, Vol. 10, p. 16, April, 1937.

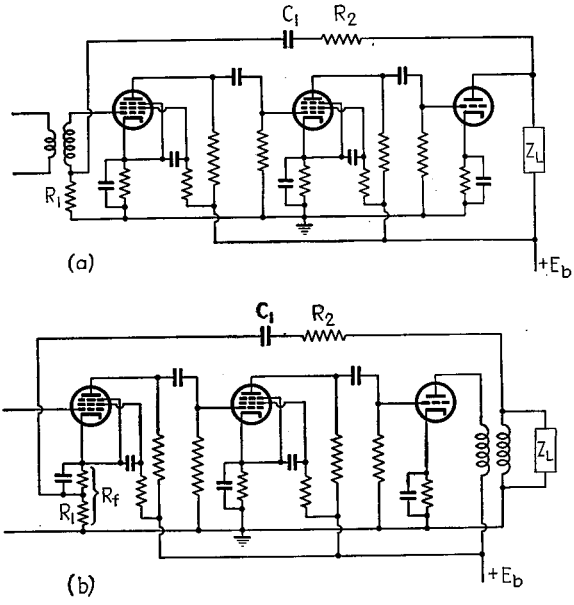


FIG. 36.—Typical feedback circuits involving three stages of amplification. In practice these circuits may be modified by expedients such as shown in Fig. 93, Sec. 3 in order to control $A\beta$ at high and low frequencies.

reduces amplitude distortion. The arrangement does not, however, improve the response at high frequencies. On the other hand, when current feedback is used with the same output transformer, as in Fig. 37*b*, the action of feedback is to tend to make the current through the transformer primary reproduce the wave shape of the applied signal and be independent of the applied frequency. The output then tends to be constant at high frequencies but falls off more at low frequencies than in the absence of feedback. Also, at low frequencies, amplitude distortion due to saturation in the core of the output transformer will be increased.

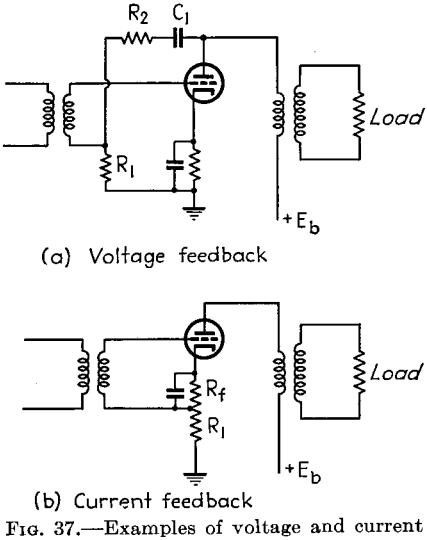


FIG. 37.—Examples of voltage and current feedback.

All the circuits shown in Figs. 34 to 36 use a resistance network to obtain feedback. It is possible to introduce reactance elements in the feedback network, and this is sometimes done, although it may lead to additional phase shifts.

Effect of Feedback on Frequency Response Characteristics.[1]—It was noted in connection with Eq. (36) that when the feedback factor $A\beta$ is large, the voltage amplifi-

[1] F. E. Terman and W. Y. Pan, Frequency Response Characteristic of Amplifiers Employing Negative Feedback, *Communications*, Vol. 19, p. 5, March, 1939; George H. Fritzinger, Frequency

cation approximates $-1/\beta$. The voltage gain is hence the inverse of the loss characteristic of the β circuit. In particular, if the β network is a resistance, then substantially constant amplification with zero phase shift independent of frequency will be obtained when $A\beta$ is large. However, at high and low frequencies the amplification A tends to fall off, so that even when $A\beta$ is large in the mid-frequency range, it will become small and ordinarily have a large phase shift at extreme frequencies.

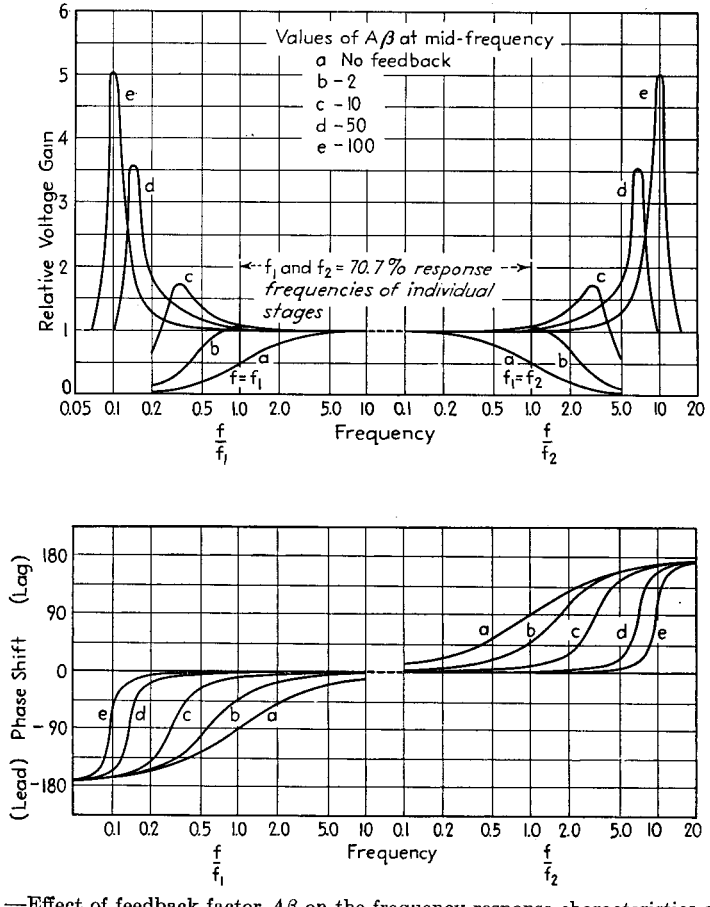


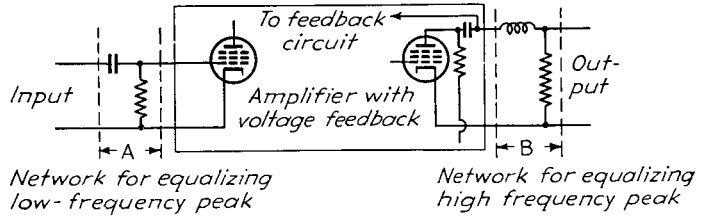
FIG. 38.—Effect of feedback factor $A\beta$ on the frequency response characteristics of a two-stage resistance-coupled amplifier having no screen grid or bias impedance.

This results in peaks of amplification at high and low frequencies if the feedback factor $A\beta$ is large in the mid-frequency range and the phase shift of $A\beta$ can approach or exceed 180° at extreme frequencies. Typical behavior of a feedback amplifier under such conditions is shown in Fig. 38. Although the use of a large amount of feedback improves the amplification and phase-shift characteristics in the middle range of frequencies, it is apparent that feedback may introduce undesirable characteristics at very low and high frequencies.

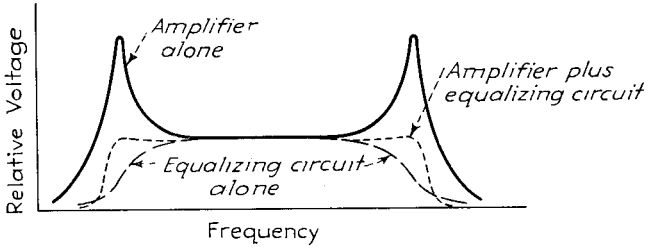
Discrimination by Inverse Feedback, *Proc. I.R.E.*, Vol. 26, p. 207, February, 1938; F. A. Everest and H. R. Johnston, The Application of Feedback to Wide-band Output Amplifiers, *Proc. I.R.E.*, Vol. 28, p. 11, February, 1940.

When peaks such as shown in Fig. 38 are not too pronounced, they can be equalized at low frequencies by a resistance-condenser combination, and at high frequencies by a resistance-inductance network, as shown in Fig. 39. These circuit elements are placed outside the amplifier proper (*i.e.*, outside the feedback loop) and are so designed as to reduce the peak of amplification to the mid-frequency gain. In this way it is possible to obtain relatively flat responses up to the frequency at which the peak occurs, followed by a sharp drop in gain, as shown.

Control of Effective Internal Impedance of Amplifiers by Feedback.[1]—The internal impedance of an amplifier can be defined as the impedance that the load observes when one looks back toward the plate circuit of the amplifier tube. The effective value of this impedance can be controlled over wide limits by the use of feedback.



(a) Amplifier with equalization for high and low frequency peaks



(b) Frequency response characteristics

FIG. 39.—Circuit for equalizing high- and low-frequency peaks produced by feedback, together with resulting characteristics.

Thus negative voltage feedback tends to make the output voltage constant, irrespective of load impedance. This effect is equivalent to decreasing the output impedance of the amplifier. The effect of feedback on the output impedance of an amplifier is given by the equation

$$\left.\begin{array}{l}\text{Effective output}\\ \text{impedance of amplifier}\end{array}\right\} = \frac{R_p - \mu A_1 \alpha}{1 - \mu A_1 \beta} \tag{40}$$

where R_p = plate resistance of output tube (equals output impedance in the absence of all feedback).

μ = amplification factor of output tube.

A_1 = amplification between grid of amplifier input and grid of last tube, in the absence of feedback.

$$\alpha = \text{current feedback factor} = \frac{\left\{\begin{array}{l}\text{feedback voltage developed}\\ \text{by current in load}\end{array}\right\}}{\text{current in load}}.$$

[1] H. F. Mayer, Control of the Effective Internal Impedance of Amplifiers by Means of Feedback, *Proc. I.R.E.*, Vol. 27, p. 213, March, 1939.

$$\beta = \text{voltage feedback factor} = \frac{\left\{\begin{matrix}\text{feedback voltage derived} \\ \text{from output voltage}\end{matrix}\right\}}{\text{output voltage}}.$$

The voltage and current feedback factors may be either positive or negative, according to whether the feedback aids or opposes the signal voltage, respectively.

The various effects that can be produced on the output impedance by feedback are revealed by Eq. (40). When $\alpha = \beta = 0$, there are no feedback effects, and the output impedance equals the plate resistance of the tube. When there is no current feedback ($\alpha = 0$), then the output impedance is $R_p/(1 - \mu A_1\beta)$, which will be greater or less than the plate resistance of the tube according to whether the feedback is positive or negative, respectively. On the other hand, when there is no voltage feedback ($\beta = 0$), the effective output impedance is $(R_p - \mu A_1\alpha)$. This will be less or more than the plate resistance of the tube, according to whether the feedback is positive or negative, respectively. In the special case where $\mu A_1\alpha = R_p$, it will be noted that the effective output resistance is zero, while when $\mu A_1\alpha > R_p$, the effective output impedance will be negative.

The gain of the amplifier with combined voltage and current feedback is given by the equation

$$\left.\begin{matrix}\text{Voltage amplification} \\ \text{taking into account} \\ \text{feedback}\end{matrix}\right\} = \frac{A_0}{1 - A_0\left(\beta + \frac{\alpha}{Z_L}\right)} \tag{41}$$

where Z_L is the load impedance and A_0 the voltage amplification without feedback. This is identical with Eq. (36) except that $A\beta$ has been replaced by $A_0\left(\beta + \frac{\alpha}{Z_L}\right)$. When the two types of feedback are of opposite sign and are so related that

$$\beta = -\frac{\alpha}{Z_L} \tag{42}$$

it will be noted that the positive and negative feedbacks balance each other as far as gain is concerned, although they will still influence the output impedance.

The criterion for avoiding oscillation with combined current and voltage feedback is the same as discussed in connection with Fig. 32, except that now the effective feedback factor is $A_0\left(\beta + \frac{\alpha}{Z_L}\right)$ instead of being $A\beta$.

Modified Forms of Feedback.—A result equivalent to that given by negative feedback can be obtained with other circuit arrangements. In the arrangement shown in Fig. 40*a*, one has an ordinary negative feedback amplifier in which the first stage is designed to have as nearly ideal characteristics as possible.[1] Positive feedback is derived from the output of this first stage and is balanced against the negative feedback so that in the mid-frequency range the two feedbacks cancel. The amplification characteristic of the entire amplifier will then tend to be the same as the amplification characteristics of the carefully designed first stage, since any departure of the output from this condition produces an unbalance between the feedbacks that opposes the departure.

Another arrangement is shown schematically in Fig. 40*b*. Here A is an ordinary amplifier to which there is applied a signal voltage e that delivers an output voltage E. A fraction k of the output voltage is balanced against the signal voltage e, and a voltage proportional to the difference between e and kE is reinserted in the amplifier with such polarity that E is altered in a manner that makes $|e - kE|$ less than would

[1] E. L. Ginzton, Balanced Feedback Amplifiers, *Proc. I.R.E.*, Vol. 26, p. 1367, November, 1938.

otherwise be the case. Such an amplifier acts in much the same way as one having negative feedback.[1]

Practical Use of Negative Feedback in Audio-frequency Amplifiers.[2]—Negative feedback is widely used in audio power amplifiers for the purpose of reducing nonlinear

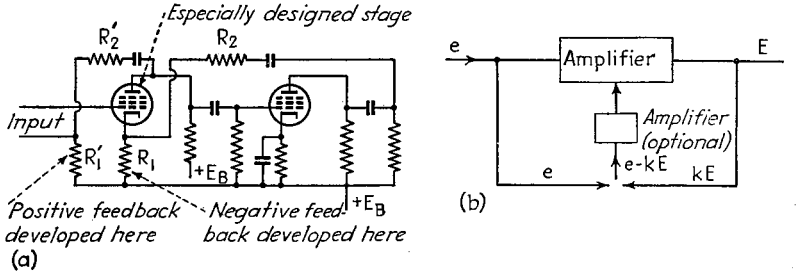
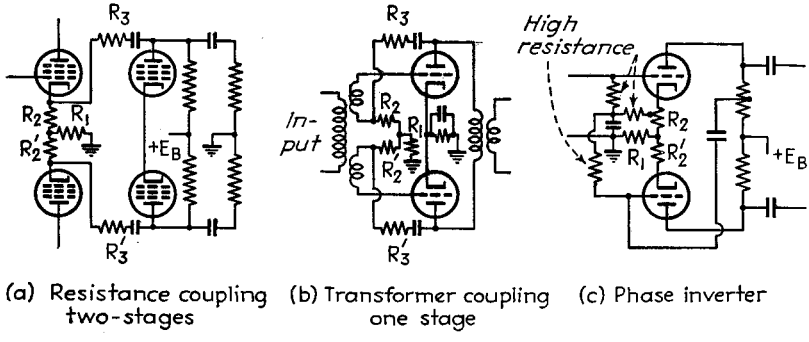


FIG. 40.—Alternative methods of obtaining behavior similar to that obtained with negative feedback.



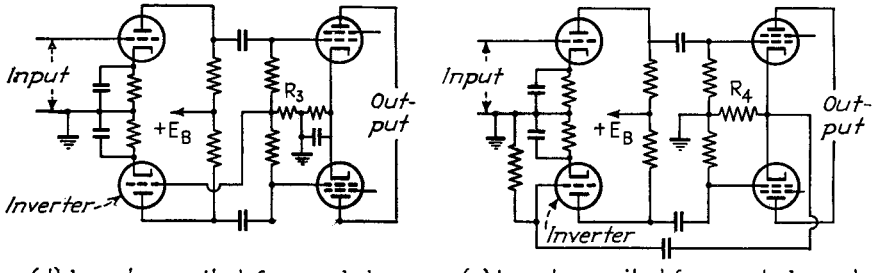


FIG. 41.—Circuits for using feedback to minimize unbalances between the two sides of a push-pull amplifier.

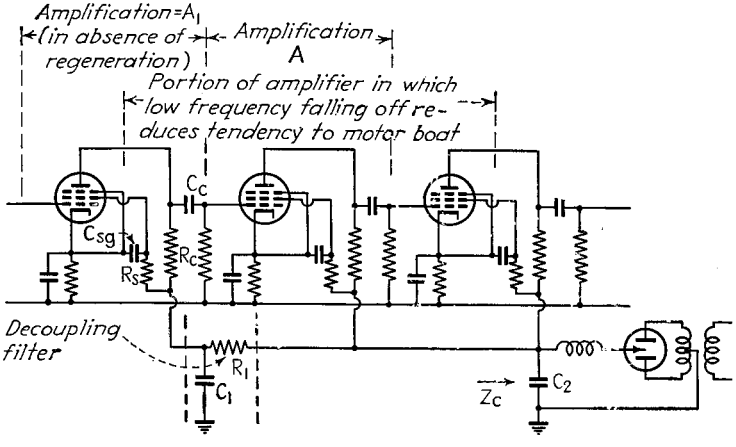
distortion. When pentode and beam tubes are used, negative feedback can also be used to reduce the effective output impedance of the tube to a value sufficiently low to damp out transients in loud-speakers.

[1] Unpublished results of the author show that this circuit has all the important properties of negative feedback circuits. See also P. O. Pedersen, A Distortion-free Amplifier, *Proc. I.R.E.*, Vol. 28, p. 59, February, 1940, who gives a slightly different arrangement.

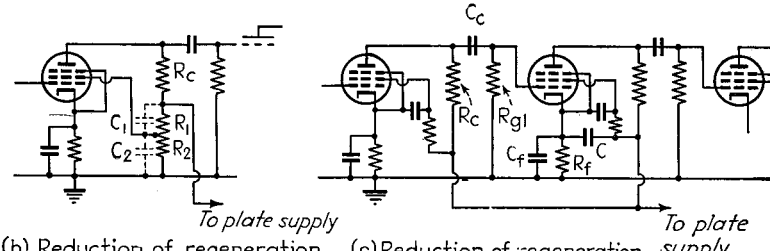
[2] The application of negative feedback to radio transmitters is given in Sec. 9, Par. 1.

Negative feedback is extensively used in telephone repeaters to stabilize gain and reduce distortion and cross-modulation. Gain stability is important in long toll circuits containing many amplifiers, since the allowable variation in gain per individual amplifier with tube replacements and variations of supply voltages is very small. The reduction of cross-modulation obtainable with negative feedback is very important where a number of carrier channels are amplified simultaneously.

Feedback can be employed in push-pull amplifiers to reduce unbalance between the two sides of the push-pull system. Circuits for doing this are illustrated in Fig



(a) Circuit having common plate impedance and decoupling filter



(b) Reduction of regeneration by screen balance

(c) Reduction of regeneration by cathode method

FIG. 42.—A three-stage amplifier with common plate supply having an internal impedance Z_c, and various means for reducing regeneration caused by Z_c.

41. In the first three of these, unbalance in the push-pull stages produces a current through the resistance R_1. This results in the development of a feedback voltage that is applied to the tubes in such a manner as to reduce the difference in the output of the two sides. In these circuits, R_2 is a resistance for developing ordinary negative feedback, and can be short-circuited or by-passed if negative feedback is not desired for other than balancing purposes. In the final two circuits, the excitation of the phase inverter tube is obtained, respectively, from a resistor R_3 used to return the two grid leaks to ground, and from the cathode resistor R_4. These resistors have a voltage developed across them corresponding to the unbalance between the two sides of the amplifier. This unbalanced voltage is then used to excite the grid of the phase-inverter tube in such a way as to make the output of the phase inverter reduce the

unbalance. By employing large gain in the inverter stage, the residual unbalance between stages will normally be less than 10 per cent, even with wide variations in circuit constants.

Feedback is widely used in laboratory amplifiers and in amplifiers incorporated in measuring equipment. By making the feedback factor large and using a resistance feedback network, it is possible to obtain characteristics that approach very closely to perfection. In particular, nonlinear distortion can be made negligible, phase shift reduced virtually to zero, the amplification maintained almost absolutely constant over wide frequency bands, and the gain is substantially unaffected by tube replacements and ordinary variations in supply voltages.[1]

12. Regeneration in Multistage Amplifiers.—When several stages of amplification are connected in cascade, there is the possibility of amplified energy being transferred from stages of high power level back to stages of lower power level. Such energy transfer is termed *regeneration*, and is a form of feedback that is to be avoided, because it gives rise to undesired modifications of the frequency-response characteristic, and in some cases causes oscillations.

Regeneration Arising from a Common Plate Impedance.—Most regeneration troubles occurring in audio-frequency amplifiers arise from the internal impedance of the battery or rectifier-filter system used to supply plate voltage to several stages. The situation arising from the presence of such a common impedance is shown in Fig. 42*a*, where it is apparent that amplified current existing in the plate circuit of the final amplifier stage will develop voltage across the impedance Z_c and that this voltage will be applied to the plate circuits of the stages of lower power level and cause regeneration. This alters the amplification characteristic, and in severe cases gives rise to oscillations.

In a practical analysis of the effect of a common plate impedance, it is necessary to consider only the energy transfer from the final stage of the amplifier back to the plate circuit of the first stage. This is because energy transfer between other stages is negligible by comparison, as a result of the smaller difference in power levels involved. To the extent that this is true, the total effect of regeneration on the entire system is to alter the effective amplification between the grid of the first (input) tube and the grid of the next amplifier tube by the factor[2]

$$\frac{\left.\begin{array}{c}\text{Actual amplification of}\\ \text{first stage}\end{array}\right\}}{\left.\begin{array}{c}\text{Amplification of first}\\ \text{stage with no regeneration}\end{array}\right\}} = \frac{1}{1 - KA\left(\dfrac{A_1}{\beta_1\gamma_1}\right)\dfrac{Z_c}{Z_1}\dfrac{g_m'}{g_m}\beta'\gamma'D} \tag{43}$$

where A = amplification from grid of second tube to the grid of the last tube, as indicated in Fig. 42*a*.

A_1 = actual amplification from grid of first tube to grid of second tube in the absence of regeneration.

β_1 and γ_1 = factors taking into account loss in amplification in first stage resulting from screen and bias impedances, respectively.

$\dfrac{A_1}{\beta_1\gamma_1}$ = amplification of first stage when screen and bias impedances produce no falling off.

[1] F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 27, p. 649, October, 1939.

[2] Equation (43) assumes that the voltage developed across the common impedance produces negligible voltage upon the screen electrode of the first tube. Under practical conditions this assumption is always realized, since the impedance formed by R_sC_{sg} in the screen circuit serves as an additional decoupling filter having proportions such as effectively to isolate the screen even at very low frequencies.

g_m = transconductance of first tube receiving anode voltage from the common plate supply.

g'_m = transconductance of final tube of amplifier.

K = factor such that $Kg'_m\beta'\gamma'$, multiplied by voltage on grid of last tube, gives the a-c current through the common plate impedance (see below).

Z_c = internal impedance of source of anode voltage.

D = transmission factor of decoupling filters in lead from common plate supply to the plate circuit of first tube (see below).

Z_1 = impedance between output of decoupling filter (or output of power supply when decoupling filters are not used) and the plate electrode of the first tube (in Fig. 42*a* $Z_1 = R_c$).

β' and γ' = factors taking into account loss in amplification in last stage resulting from screen and bias impedances, respectively.

The factor K in Eq. (43) is defined by the relation

$$K = \frac{\left\{\begin{matrix}\text{Admittance of circuit element} \\ \text{connecting plate supply to plate} \\ \text{electrode of final tube}\end{matrix}\right\}}{\frac{1}{Z_L} + \frac{1}{R_p}} \tag{44}$$

where Z_L is the total impedance in the plate circuit of the final tube and R_p is the plate resistance of this tube. The value of K depends upon the circuit constants, the frequency, and upon whether the tube is a triode or pentode. At low frequencies, K nearly always exceeds 0.5, and with practical amplifiers is always at least 0.2 under conditions where regeneration from the common plate impedance is important. At the same time K cannot exceed unity in practical cases.

The factor D takes into account the effect of a decoupling filter such as R_1C_1 in Fig. 42*a*. Such a filter is for the purpose of making the feedback voltage reaching the plate circuit of the first tube less than the voltage developed across the common impedance. On the assumption, usually justified, that Z_c is considerably less than R_1 and that the reactance of C_1 is considerably less than the impedance seen in looking from C_1 toward the first tube, one has for a single-stage filter

$$D = \frac{\left(\frac{1}{\omega C_1}\right)}{\sqrt{R_1^2 + \left(\frac{1}{\omega C_1}\right)^2}} \cong \frac{\left(\frac{1}{\omega C_1}\right)}{R_1} \tag{45}$$

Values of D can be obtained graphically with the aid of Fig. 43. When two filter stages are connected in tandem, the effective reduction factor is approximately the product of the reduction factors of the individual filters.

Circuit Design to Minimize Regeneration and Eliminate Motorboating.—Multistage amplifiers should be so designed that there is negligible regeneration in the useful frequency range and so that there is not sufficient regeneration at frequencies outside the normal operating range to produce oscillations.

In order for regeneration to have negligible effect, it is necessary for the factor $KA(A_1/\beta_1\gamma_1)(Z_c/Z_1)(g'_m/g_m)\beta'\gamma'D$ in Eq. (43) to be small compared with unity. To insure absence of oscillation, this quantity must have an absolute value less than unity.[1]

[1] Oscillations will not be present even when $|KA(A_1/\beta_1\gamma_1)(Z_c/Z_1)(g_m'/g_m)\beta'\gamma'D| \gtrsim 1$, provided that under these conditions, the phase of the quantity is such as to give negative rather than positive feedback. For practical purposes, however, the somewhat simpler rule stated is a more convenient guide, and errs on the side of safety. The exact criterion for avoiding oscillation is the same requirement that $A\beta$ must satisfy as discussed in connection with Fig. 32.

Regeneration can be easily made small or negligible for frequencies in and above the normal audio range, by shunting a capacity (C_2 in Fig. 42*a*) across the output of the common plate supply, and, where the gain is large, as is the case with more than two stages, by the use of one or two sections of decoupling filter (R_1C_1 in Fig. 42*a*). However, at low frequencies, the shunt condenser across the power supply is ineffective as a by-pass, and the isolation obtained from the decoupling filters becomes small because of the increasing reactance of condensers at low frequencies. The result in practical amplifiers is that regeneration is frequently sufficient at low frequencies to produce oscillations even when the regeneration has negligible effect upon the ampli-

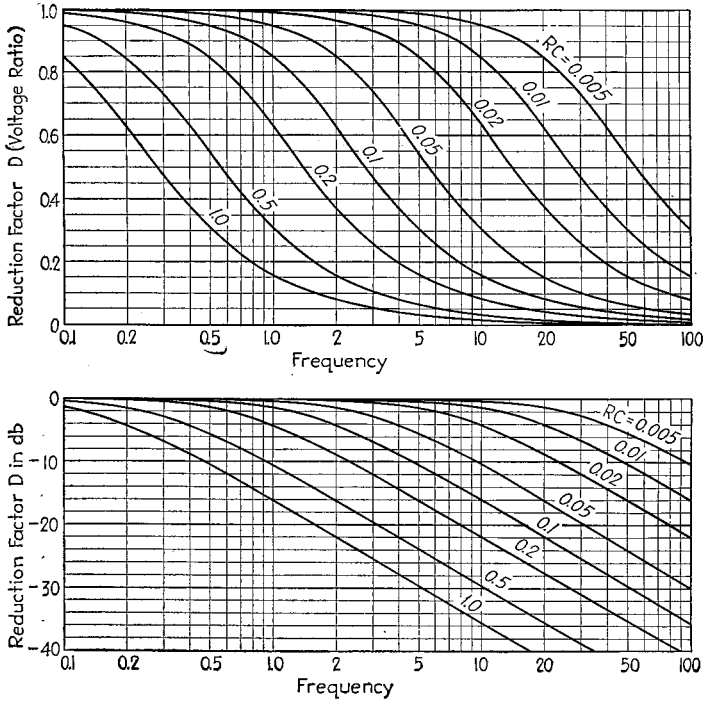


FIG. 43.—Curves giving the reduction D produced by resistance-capacity filter on the assumption that the impedance on the source side is small compared with R_1, and that the impedance on the load side is large compared with $1/\omega C_1$.

fication in the normal frequency range. Such oscillations are referred to as *motorboating*, and commonly have a frequency of only a few cycles per second.

Motorboating cannot exist if $|KA(A_1/\beta_1\gamma_1)(Z_c/Z_1)(g'_m/g_m)\beta'\gamma'D| < 1$. To accomplish this, it is necessary to rely mainly upon the low-frequency falling off in amplification produced by the tubes and circuits between the plate electrode of the first tube and the plate electrode of the final tube.[1] *Accordingly, multistage amplifiers operated from a common plate supply should never be designed to have a better low-frequency response between the plate of the first tube and the plate of the final tube than is actually required. In particular, it is very desirable that whenever possible the design of the plate*

[1] It will be noted that this includes the falling off due to screen grid and bias impedance of all tubes but the first and the effect of coupling networks in all of the plate circuits except in the plate circuit of the final tube.

coupling networks and screen impedances in this section of the amplifier be so coordinated as to give maximum falling off in low-frequency amplification.

A power supply having low internal impedance at low frequencies is also helpful in eliminating motorboating. At low frequencies, the common impedance has a value approximating the equivalent internal d-c resistance corresponding to the d-c voltage regulation of the power supply; *i.e.*,

$$\text{Equivalent } Z_c \text{ at low frequencies} = \frac{\Delta E_B}{\Delta I_B} \tag{46}$$

where ΔE_B is the drop in power-supply voltage produced by increasing the output current by ΔI_B. The common impedance can be made small by using low-resistance chokes in the filter system and employing a rectifier tube with low voltage drop. The common impedance can be made negligible at low frequencies by using a power supply provided with an electronic voltage regulating system that maintains the potential across Z_c constant (see Par. 8, Sec. 8).

The tendency for regeneration and motorboating to exist can be greatly reduced by employing circuits in which the feedback from the common plate impedance to the plate circuit of the first tube is balanced by an equal and opposite feedback to some other point in the amplifier.[1] Arrangements for accomplishing this result are shown in Figs. 42*b* and 42*c*. In the first of these, the screen voltage is derived from a voltage divider R_1R_2 connected across the plate supply, so that any feedback voltage developed across this common impedance will be applied to the screen-grid electrode as well as to the plate circuit.[2] By selection of the proper ratio R_1/R_2 for the voltage divider (which can be readily done by trial), it is possible theoretically to reduce to zero all feedback (and also hum) from the common plate circuit to the amplifier stage thus balanced. This action is independent of frequency if the condenser C_2 that would normally be used to by-pass the screen to ground is balanced by a corresponding condenser C_1 so that $C_2/C_1 = R_1/R_2$. A similar result is achieved in the circuit of Fig. 42*c*, except that here the balancing voltage is applied between the cathode and ground of the following tube.[3] The required circuit proportions are

$$\begin{aligned} \frac{C_f}{C} &= \frac{R_c}{R_{gl}}\left(1 + \frac{R_{gl}}{R_p}\right) \\ \frac{C_c}{C} &= \frac{R_f}{R_{gl}}\left(1 + \frac{R_c}{R_p}\right) \end{aligned} \tag{46a}$$

The notation is illustrated in Fig. 42*c*. The circuit proportions called for by Eq. (46*a*) are of the approximate magnitude that would be used in ordinary audio-amplifier design. When carefully constructed, the balancing systems shown in Fig. 42*b* and *c* will give a reduction in the low-frequency feedback (and power-supply hum) to a particular stage of 40 db or more.

Additional means available for reducing regeneration at low frequencies include the use of push-pull amplification, particularly in the final stage, and the use of separate power supplies for different stages. With push-pull amplification in the final stage, this stage develops voltage across the source impedance only to the extent that the two sides of the amplifier are unbalanced, so that a considerable reduction in the tendency to regenerate is obtained. Where the first stage of the amplifier is also push-pull, the voltage developed across the common impedance is applied to the two

[1] Such circuit arrangements also discriminate equally against hum in the power-supply system.

[2] K. B. Gonser, A Method of Neutralizing Hum and Feedback Caused by Variations in the Plate Supply, *Proc. I.R.E.*, Vol. 26, p. 442, April, 1938.

[3] A discussion of this method of balancing regeneration, together with experimental results, is given by W. Y. Pan, Circuit for Neutralizing Low-frequency Regeneration and Power-supply Hum. *Proc. I.R.E.*, vol. 30, p. 411, September, 1942.

sides of this push-pull stage in the same phase, and so cancels out to the extent that the two sides of the input stage are balanced.

In extreme cases, where the amplification involved is very great or where the amplifier must maintain its gain down to very low frequencies, it is always possible to resort to separate power supplies for the low-level and high-level stages of the amplifier, and thereby reduce the tendency toward regeneration.

Regeneration from Stray Couplings.—Although the most troublesome regeneration problems arising in audio-frequency amplifiers come from the impedance of a common power supply, it is also possible for stray electrostatic and magnetic couplings to cause regeneration. Such sources of regeneration are most effective at the higher frequencies, and can be readily taken care of by proper shielding and by proper spacing and orientation of parts.

Energy transfer between stages through electrostatic couplings can be eliminated by providing grounded shields for glass tubes or grounding the envelopes of metal tubes, by proper placing of wiring and parts that are at widely different power levels, and in severe cases by the use of supplementary electrostatic shielding.

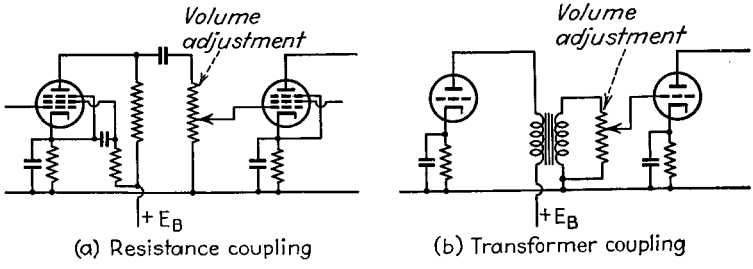


FIG. 44.—Common methods of controlling volume in resistance- and transformer-coupled amplifiers.

The chief source of magnetic coupling is the mutual inductance between transformers associated with stages operating at widely different power levels, as the input and output transformers of an amplifier. Regeneration from this source can be controlled by adequate spacing between transformers, by orientation of cores to reduce mutual inductance, and by use of transformers that are magnetically shielded or otherwise arranged to be relatively immune to external fields.

Another form of energy transfer between stages arises from the grid-plate capacity of the amplifier tube. This has an effect equivalent to modifying the equivalent input impedance of the tube, and is discussed in Par. 25.

13. Amplitude and Volume Control in Audio-frequency Amplifiers.—The volume, amplitude range, peak level, etc., in audio-frequency amplifiers can be controlled by a variety of devices,[1] the most important of which follow.

Manual Control of Volume.—Manually operated volume controls in audio-frequency amplifiers must be so arranged that the frequency response is not adversely affected by the volume setting. The control must also be located at a point of sufficiently low-power level in the amplifier that there is no possibility of stages being overloaded before the control can reduce the volume level.

The standard method of controlling volume in resistance-coupled amplifiers is shown in Fig. 44*a*, and makes use of a grid leak in the form of a high-resistance potentiometer. Such a volume control produces a slight improvement in the high-frequency response at low-volume settings.

[1] A discussion and classification of these devices is given by S. B. Wright, Amplitude Range Control, *Bell System Tech. Jour.*, Vol. 17, p. 520, October, 1938; and A. C. Norwine, Devices for Controlling Amplitude Characteristics of Telephonic Signals, *Bell System Tech. Jour.*, Vol. 17, p. 539, October, 1938.

Volume control in transformer-coupled amplifiers is difficult to obtain without causing serious variations in the frequency response characteristics. When an amplifier contains both resistance- and transformer-coupled stages, it is preferable to control the volume in the resistance-coupled stages. Where this is not possible, the

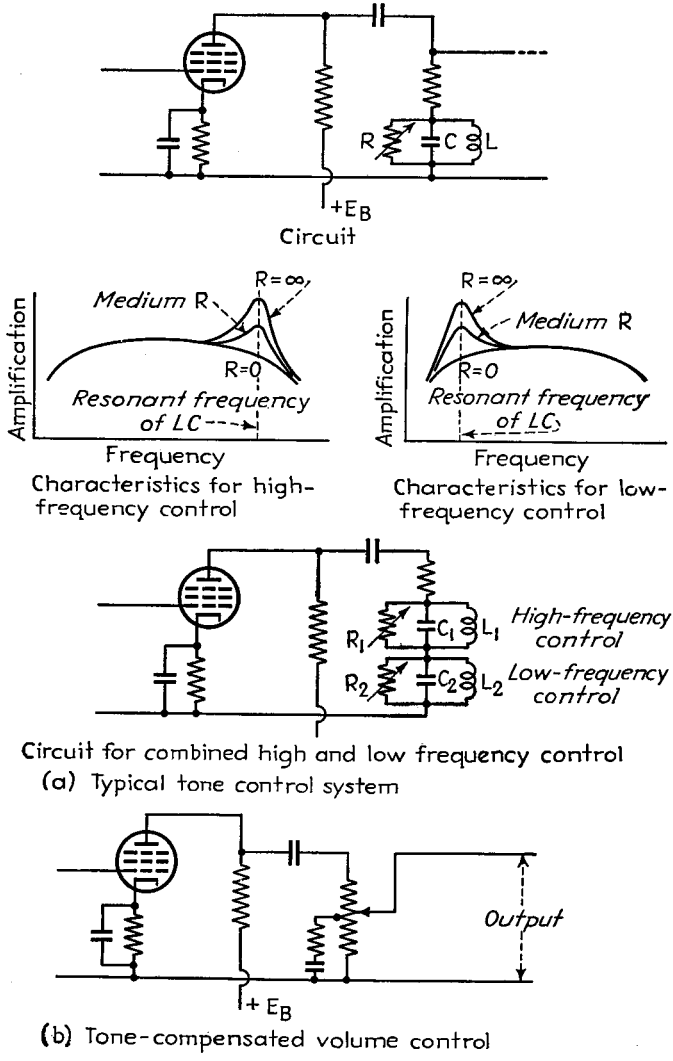


FIG. 45.—Typical tone-control circuits.

usual arrangement is to place a high-resistance potentiometer across the transformer secondary, as in Fig. 44*b*, and preferably employ a transformer designed to operate with resistance loading across its secondary.

Tone Controls and Tone-compensated Volume Controls.—Audio amplifiers are frequently provided with controls for altering the high-frequency and low-frequency characteristics. These are termed *tone controls*, and are for the purpose of modifying

the character of the amplifier output to give new esthetic effects, to compensate for acoustic characteristics, reduce hum, reduce hiss or scratch, etc. Typical tone-control circuits are shown in Fig. 45*a*. Such circuits should preferably exert their control upon the high-frequency and low-frequency amplification without altering the mid-frequency gain, although in many cases this ideal is not achieved.

The characteristics of the ear are such that the apparent loudness of low-frequency tones relative to middle and high-frequency tones is less as the volume level is reduced. In order to correct for this effect, volume controls are sometimes arranged so that as the volume control reduces the gain, the reduction is less for low frequency than for middle and high-frequency components. Such an arrangement is termed a *tone-compensated volume control.* A typical example is shown in Fig. 45*b*.[1]

Volume Expanders and Compressors—Companders.—A volume compressor is a device that reduces the amplification when the signal being amplified is large and increases the amplification when the signal is small. Compressors are used to reduce the volume range in recording sound on film and disk, and are also sometimes used in radio-telephone transmission, etc.

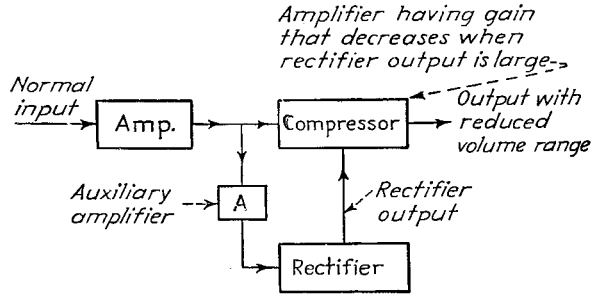


FIG. 46.—Schematic diagram of a volume compressor.

A volume expander is a device that operates in such a way as to increase the amplification when the average power level of the signal is high, and to reduce the amplification when the average power level of the signal is small, thereby causing the loud passages to be still louder and the weak passages to have their weakness accentuated. Expanders are used to counteract the compression of volume range that is necessary in sound recording, etc., and also to introduce added esthetic effects.

Volume expanders (or compressors) are essentially amplifiers in which the amplification is varied in accordance with the envelope amplitude of the signal. This control can be obtained in various ways. One method is illustrated schematically in Fig. 46. Here the signal to be compressed (or expanded) is amplified by an auxiliary amplifier *A* and then rectified in such a way that there is developed across a resistance-condenser combination a d-c voltage that tends to follow the envelope of the signal. This d-c voltage is used to vary the bias, and hence amplification, of a resistance-coupled amplifier tube using a variable-mu tube. By proper attention to circuit details, particularly the time constants, it is possible to control the character of the compression (or expansion) and also the quickness with which the amplification is increased and decreased.[2]

[1] Charts for assisting in the design of tone-compensated volume controls of this type are given by P. A. D'Orio and Rinaldo de Cola, Bass Compensation Design Chart, *Electronics*, Vol. 10, p. 38 October, 1937.

[2] Circuit diagrams (with constants) of volume expanders are given by C. M. Sinnett, Practical Volume Expansion, *Electronics*, Vol. 8, p. 14, November, 1935; Harry Paro, Public Address AVC *Electronics*, Vol. 10, p. 24, July, 1937.

Another method of obtaining volume expansion (or compression) is to introduce feedback in an

A system involving a volume compressor to reduce the volume range for transmission, followed by an expander at the receiving point to restore the original range, is referred to as a compander, and is used in radio-telephone extensions of telephone toll systems to improve the signal-to-noise ratio.[1]

Automatic Volume Control—Vogads.—Automatic volume control (abbreviated A.V.C.) utilizes the rectified audio-frequency currents to control the gain of an amplifier in such a manner that the peak output is maintained at substantially constant level. A.V.C. systems are used for insuring full modulation in modulating a radio transmitter or recording sound on film, for maintaining the output of public-address systems more nearly constant when the speaker turns away from the microphone, etc.[2]

The term vogad is derived from the first letters of the phrase "voice operated gain adjusting device." A vogad is a particular form of A.V.C. system, and is used in radio-telephone extensions of wire telephone systems for the purpose of maintaining full modulation of the transmitter, irrespective of the loudness of the speaker's voice.[3]

Volume Limiter.—A volume limiter is a device that automatically reduces the amplification when the average input volume to the amplifier exceeds a predetermined value, so that the output for all inputs in excess of this value is substantially constant. Volume limiters are used in radio-telephone transmitters to prevent overmodulation.

Volume limiters ordinarily operate by rectifying a portion of the signal and using the rectified output to reduce the gain when the signal exceeds the desired level.

Peak Limiters.—A peak limiter is an amplifier so controlled by rectified voice currents that the gain is quickly reduced and then slowly restored when the instantaneous peak amplitude of the signal being amplified exceeds a predetermined value. A peak limiter differs from a volume limiter primarily in that it is controlled by the instantaneous peaks instead of by the average volume.

Peak Chopper.—A peak chopper is a device that prevents peak amplitudes from exceeding a critical value, and is ordinarily arranged to operate instantaneously.

VIDEO-FREQUENCY AMPLIFIERS

14. General Requirements of Video-frequency (Wide-band) Amplifiers.[4]—Amplifiers for the video-frequency currents of television circuits are required to handle a frequency range from about 60 cycles to about 4 megacycles with substantially uniform gain and negligible phase distortion. The maximum variation in over-all delay

ordinary amplifier by means of a resistance that is supplied by the plate resistance of a triode tube. This resistance, and hence the gain, is then controlled by a bias voltage developed by rectification of the signal. Such an arrangement has the advantage that the negative feedback reduces distortion that frequently occurs in volume expanders and compressors. See B. J. Stevens, Low Distortion Volume Expansion Using Negative Feed-back, *Wireless Eng.*, Vol. 15, p. 143, March, 1938; E. G. Cook, A Low Distortion Limiting Amplifier, *Electronics*, Vol. 12, p. 38, June, 1939; H. H. Stewart and H. S. Pollock, Compression with Feedback, Electronics, Vol. 13, p. 19, February, 1940.

[1] R. C. Mathes and S. B. Wright, The "Compandor"—An Aid Against Radio Static, *Elec. Eng.*, Vol. 53, p. 860, June, 1934; R. A. Heising, Radio Extension Links to the Telephone System, *Bell Sys Tech. Jour.*, Vol. 19, p. 611, October, 1940.

[2] For typical A.V.C. circuits (including constants), see Harry Sohon, Supervisory and Control Equipment for Audio-frequency Amplifiers, *Proc. I.R.E.*, Vol. 21, p. 228, February, 1933; Paro, *loc. cit.*

[3] S. B. Wright, S. Doba, and A. C. Dickieson, A Vogad for Radiotelephone Circuits, *Proc. I.R.E.*, Vol. 27, p. 254, April, 1939; R. A. Heising, *loc. cit.*

[4] Following is a selected bibliography on the theory, design, and construction of video-frequency amplifiers: R. L. Freeman and J. D. Schantz, Video Amplifier Design, *Electronics*, Vol. 10, p. 22, August, 1937, and Vol. 10, p. 52, November, 1937; S. W. Seeley and C. N. Kimball, Analysis and Design of Video Amplifiers, Part II, *R.C.A. Rev.*, Vol. 3, p. 290, January, 1939; A. V. Bedford and G. L. Fredendall, Transient Response of Multistage Video-frequency Amplifiers, *Proc. I.R.E.*, Vol. 27, p. 277, April, 1939; Harold A. Wheeler, Wide-band Amplifiers for Television, *Proc. I.R.E.*, Vol. 27, p. 429, July, 1939; Albert Preisman, Some Notes on Video-amplifier Design, *R.C.A. Rev.*, Vol. 2, p. 421, April, 1938; E. W. Herold, High-frequency Correction in Resistance-coupled Amplifiers, *Communications*, Vol. 18, p. 11, August, 1938; F. A. Everest, Wide-band Television Amplifiers, *Electronics*, Vol. 11, p. 16, January, 1938, and p. 24, May, 1938; G. D. Robinson, Theoretical Notes on Certain Features of Television Receiving Circuits, *Proc. I.R.E.*, Vol. 21, p. 833, June, 1933.

time tolerable in a 441-line television system is commonly of the order of 1,000 microseconds at the low-frequency end of the band and 0.1 microseconds at the high-frequency end. On the basis of 10 amplifier stages being involved, this corresponds to a phase shift per stage not exceeding 2° at 60 cycles, and linearity in the phase-shift characteristic up to the highest frequency of importance, with a deviation not exceeding 14° per stage at 4 mc. The amplification per stage should also be constant to within a few tenths of a decibel.

15. Video-frequency Voltage Amplifiers—Low-frequency Considerations.—Video-frequency voltage amplifiers employ circuit arrangements based upon the ordinary resistance-coupled circuit used for audio-frequency amplification. The severe requirements with regard to uniformity of amplification and delay time that must be met are, however, most satisfactorily realized by modifying the ordinary resistance-coupled amplifier circuit by additional circuit elements that improve the characteristics at low and high frequencies.

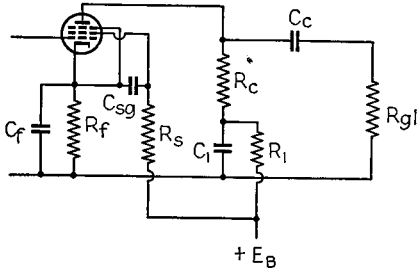


FIG. 47.—Circuit of video-frequency amplifier, showing resistance-condenser combination R_1C_1 for equalizing low-frequency response.

At low-frequencies, the circuit components causing phase shift and falling off in amplification are the coupling condenser, bias impedance, and impedance in the screen-grid circuit. Each of these must have its effect compensated for in the amplifier.

Compensation for Coupling Condenser. The coupling condenser in the plate circuit of an amplifier tube causes the output to fall off and to lead in phase at low frequencies. On the assumption that a pentode tube is used and that the grid-leak resistance is very much larger than the coupling resistance (*i.e.*, $R_{gl} >> R_c$), as is always the case in practical amplifiers, the behavior at low frequencies can be calculated from the equations

$$\text{Phase shift (lead)} = \phi = \tan^{-1}\frac{1}{\omega C_c R_{gl}} \tag{47}$$

$$\text{Falling off in gain at low frequencies} = \cos\phi \tag{48}$$

where C_c is the capacity of the coupling condenser and R_{gl} the grid-leak resistance, as indicated in Fig. 47, and ω is 2π times the frequency. These equations show that the effect of the coupling condenser is more serious on phase shift than on the amplification. Thus a grid leak-condenser combination that causes the output voltage to fall only 1 per cent below the mid-frequency voltage introduces a phase shift of about 8°. Equations (47) and (48) show that if the phase shift is not to exceed 2° at 60 cycles, it is necessary that $R_{gl}C_c > 0.076$. Although this can be achieved by using a large condenser capacity and grid-leak resistance, it is usually more practicable to allow a reasonable phase shift and then provide equalization.

Compensation for the effect of the coupling condenser in the plate circuit can be obtained by the use of a resistance-capacity combination R_1C_1 in series with the plate coupling resistor, as shown in Fig. 47. In the practical case where $R_{gl} >> R_c$ the impedance Z_1 formed by the combination R_1C_1 causes the output to be increased in magnitude by the factor $(R_c + Z_1)/R_c$ and to be shifted in phase in a lagging direction by an amount corresponding to the angle by which the impedance of $Z_1 + R_c$ is leading.

Perfect compensation for the effect of the coupling condenser is obtained when

$$R_c C_1 = R_{gl} C_c \tag{49}$$

and

$$R_1 = \infty \tag{50}$$

Under practical conditions an entirely satisfactory degree of compensation is obtained when $R_1 \geq 10(1/\omega C_1)$.

The effects produced by various adjustments of the compensating circuit R_1C_1 in a typical case are shown in Fig. 48. When Eq. (49) is satisfied, then the results are

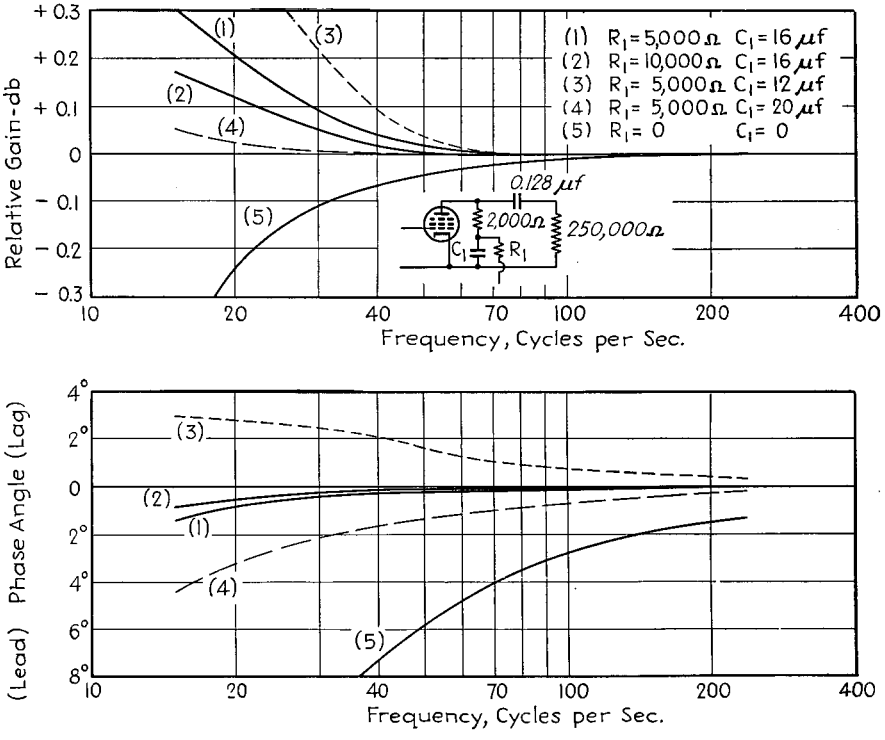


FIG. 48.—Various effects obtainable in equalizing for the effect of plate coupling condenser in video-frequency amplifiers.

more nearly perfect the higher the resistance R_1, as seen by comparison of curves (1), (2), and (5). With a fixed value for R_1, decreasing the capacity C_1 so that the time constant R_cC_1 is slightly less than $R_{gl}C_c$ will make the phase more lagging at low frequencies and tend to cause a rise in amplification, as seen by comparing curves (1) and (3) in Fig. 48. Increasing C_1 to give a larger time constant produces the opposite effect, as seen by comparing curve (4) with curves (1) and (3).

Compensation for Bias Impedance.—The bias impedance R_fC_f in Fig. 47 causes the output voltage to lead in phase and fall off in magnitude by an amount that is excessive even when the largest practicable size of by-pass condenser is used. Bias impedance can be eliminated by obtaining the bias in other ways, but this is commonly not convenient.

The phase shift and change in amplification produced by the combination of the compensating network R_1C_1 and the bias network R_fC_f in Fig. 47 are given by the relation

$$\left.\begin{matrix}\text{Relative output}\\ \text{voltage at low}\\ \text{frequencies}\end{matrix}\right\} = \frac{1 + \frac{Z_1}{R_c}}{1 + g_m Z_f} \tag{51}$$

where Z_1 is the vector impedance of the combination R_1C_1, Z_f the impedance of the combination C_fR_f, R_c the coupling resistance, and g_m the transconductance. A positive phase angle in Eq. (51) represents a leading phase angle for the output voltage.

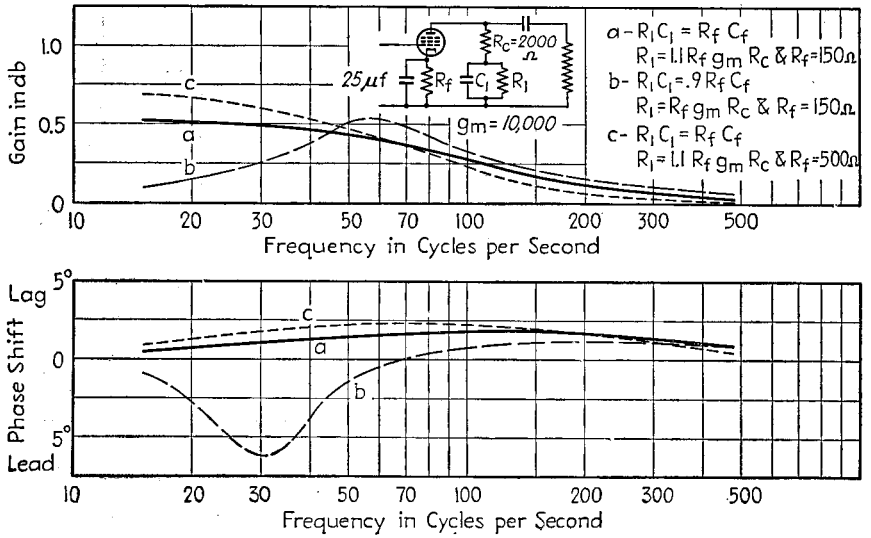


FIG. 49.—Various effects obtainable in equalizing for the cathode impedance in video-frequency amplifiers.

The right-hand side of Eq. (51) will be unity, corresponding to *perfect equalization of the bias impedance* C_fR_f, when

$$R_1 = R_f(g_m R_c) \tag{52}$$

and

$$R_1C_1 = R_fC_f \tag{53}$$

Equations (52) and (53) assume that $R_{gl} >> R_c$, a condition always satisfied in practical amplifiers. When the requirements of Eqs. (52) and (53) are not satisfied, the behavior will vary according to the character and extent of the departure from optimum proportions. The performance is shown in Fig. 49 for several typical cases.

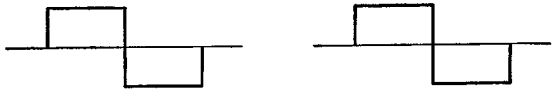
Compensation for Impedance in Screen-grid Circuit.—The proper screen voltage is normally obtained by a screen dropping resistor R_s by-passed to the cathode by a condenser C_{sg}, as shown in Fig. 47. This arrangement places an impedance in the screen circuit that can cause the amplification at low frequencies to shift in phase and fall off in magnitude in the same manner as caused by the coupling condenser. The magnitudes of these effects can be calculated exactly by Eq. (8). Under practical conditions, they are determined mainly by the by-pass condenser C_{sg} and the dynamic screen-cathode resistance R_{sg} of the tube, and can be calculated with sufficient accu-

racy by Eqs. (47) and (48) by substituting C_{sg} and R_{sg} for C_c and R_{gl}, respectively, which appear in these equations. The phase shift will not exceed 2° at 60 cycles if

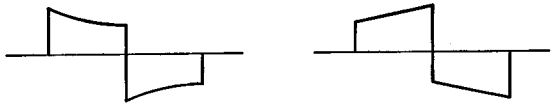
$$R_{sg}C_{sg} > 0.076 \tag{54}$$

Compensation for the screen impedance under these conditions is not necessary. If Eq. (54) is not satisfied, it is necessary to introduce correction in some plate circuit, as discussed below.

Equalization for All Low-frequency Effects.—In practical amplifiers it is desirable to equalize each individual stage as completely as possible rather than attempt to equalize for the deficiencies of several stages at a single point. At the same time it is not feasible to compensate for the effects of the bias impedance and coupling condenser

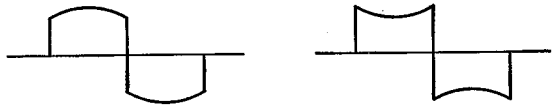


(a) Applied wave



(b) Output wave when phase leads at low-frequency

(c) Output wave when phase lags at low-frequency



(d) Output wave when amplification rises at low-frequency (No phase shift)

(e) Output wave when amplification falls at low-frequency (No phase shift)

FIG. 50.—Effects produced by amplifier imperfections upon a square wave for idealized cases.

in the same stage. The practical method of meeting this situation is to make the effects from the coupling condenser (and screen impedance) small in all stages employing a bias impedance, and then for every two to five stages providing at least one stage in which there is no bias impedance. In this one stage, the resistance combination R_1C_1 in the plate circuit is then proportioned to compensate for the coupling condenser and screen impedance effects of all the stages. This arrangement operates satisfactorily as long as the total phase shift to be corrected is less than 20°.

Experimental Check of Low-frequency Compensation.—After the low-frequency compensating circuits, with proportions as calculated, have been installed in an amplifier, it is always desirable to check the over-all behavior experimentally. Adjustments can then be made to correct for residual defects in performance as necessary.

Low-frequency tests are preferably made by applying the output of a square-wave oscillator to the amplifier input (see Par. 32, Sec. 13). The shape of the resulting output wave can be interpreted in terms of amplifier characteristics at low frequencies, as indicated in Fig. 50. A tilt in the horizontal portion of the output wave is caused

primarily by phase shift at the fundamental frequency; curvature of the horizontal portion is largely caused by a gain characteristic that does not remain constant down to the fundamental frequency of the square wave.

16. Video-frequency Voltage Amplifiers. *High-frequency Considerations.*—High gain over a wide frequency band with substantially constant amplification and constant delay time, as required in video-frequency voltage amplifiers, can be most readily obtained by adding further reactance elements to the ordinary resistance-coupled amplifier circuit. The resulting networks can be classified as two-terminal or four-terminal networks according to whether the plate lead of the amplifier tube and the grid lead to the succeeding tube are connected at the same or different points in the coupling network.

Two-terminal Coupling Networks—General Properties.—Types of two-terminal networks used in television amplifiers are shown in Fig. 51. These all have as their

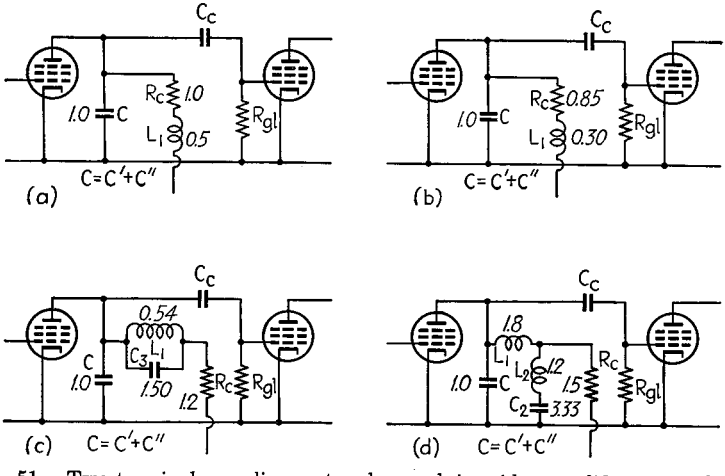


FIG. 51.—Two-terminal coupling networks used in video-amplifiers, together with recommended circuit proportions for optimum high-frequency characteristics. The numbers indicate element impedances at the top frequency f_1 that is to be amplified (reactance of C at f_1 taken as unity).

first element a shunting capacity C across the input that represents the sum of the output capacity of the amplifier tube plus the input capacity of the succeeding tube plus wiring capacity plus any padding capacity that may be added. The coupling network also includes a resistance R_c that provides the coupling impedance at the lower frequencies.

The object of the remaining reactive elements in the two-terminal network is to build up across the terminals of the condenser C the highest possible impedance that is constant in magnitude and has a phase angle proportional to frequency up to the highest frequency to be amplified. Analysis of the properties of networks shows that the highest possible constant impedance that can be built up across a capacity C is $2X_1$, where X_1 is the reactance of the capacity at the highest frequency for which the impedance is to be constant, *i.e.*, the highest frequency to be amplified.

The amplification obtained with a two-terminal coupling network having an impedance Z_c can be taken as[1]

$$\text{Voltage amplification} = g_m Z_c \qquad (55)$$

[1] This equation assumes that the plate and grid-leak resistances are both large compared with Z_c, a requirement readily satisfied by all wide-band amplifiers.

where g_m is the transconductance of the tube. In the region where the coupling impedance is constant, one has $Z_c = R_c$, and Eq. (55) becomes

$$\left.\begin{matrix}\text{Amplification of flat}\\ \text{part of characteristic}\end{matrix}\right\} = g_m R_c \tag{56}$$

Since the maximum possible value of R_c is $2X_1$, one has

$$\left.\begin{matrix}\text{Maximum possible voltage}\\ \text{gain with two-terminal}\\ \text{coupling network}\end{matrix}\right\} = 2g_m X_1 \tag{57}$$

Design of Two-terminal Coupling Networks.—The two-terminal coupling networks having greatest use in video amplifiers are shown in Fig. 51. Suitable circuit propor-

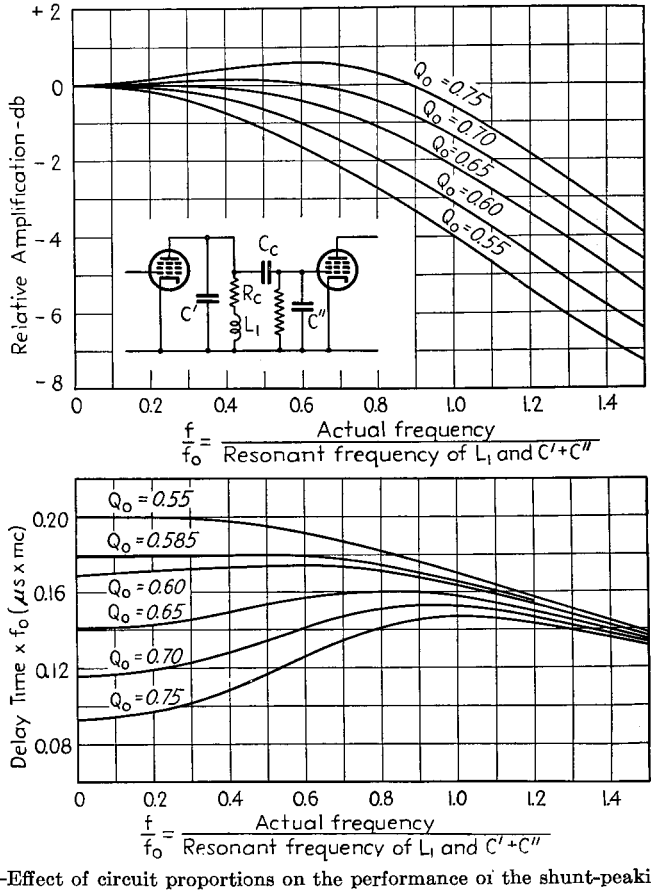


FIG. 52.—Effect of circuit proportions on the performance of the shunt-peaking circuit.

tions for good performance are indicated by the numbers associated with each circuit element. These give the impedance of the element at the top frequency f_1 to be amplified, in terms of the impedance X_1 of the shunting capacity C at this frequency. Thus for Fig. 51*b*, R_c has a resistance equal to 0.85 times the reactance of C at the frequency f_1, and the inductance L is of such size that its reactance at the frequency f_1 is $0.30X_1$.

The suggested circuit proportions are a compromise between proportions giving the most uniform amplification characteristic up to the top frequency f_1 and the proportions giving the most constant delay-time characteristic. This is illustrated by Fig. 52, which shows the amplification and time-delay characteristics of the circuit of Figs. 51*a* and 51*b*. This particular arrangement, which is termed the *shunt-peaking* circuit, is a low Q resonant circuit such as is discussed in Par. 2, Sec. 3, and can have its properties expressed in terms of the resonant frequency f_0 of the inductance and capacity of the circuit, and the circuit Q at resonance (denoted by Q_0). The gain is most nearly constant up to a frequency f_1 when $f_0 = 1.41\ f_1$(*i.e.*, $f_1/f_0 = 0.707$) and Q_0 equals 0.707, corresponding to the proportions of Fig. 51*a*. On the other hand, the

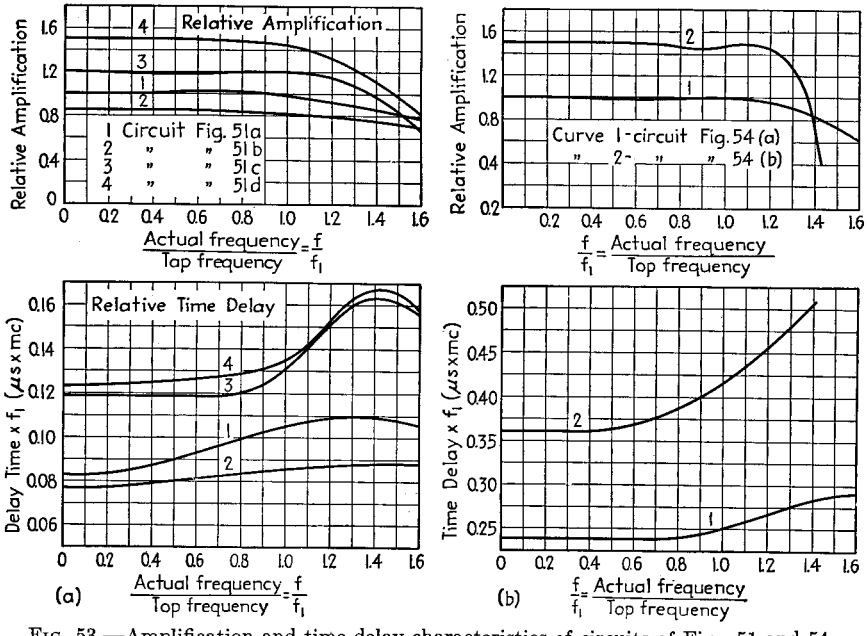


FIG. 53.—Amplification and time-delay characteristics of circuits of Figs. 51 and 54.

time-delay characteristic will be most nearly constant up to a frequency f_1 by making $f_0 = 1.67 f_1$ (*i.e.*, $f_1/f_0 = 0.6$) and $Q_0 = 0.585$. A good compromise is

$$f_0 = 1.825\ f_1 \text{ (i.e., } f_1/f_0 = 0.55)$$

and $Q_0 = 0.65$, corresponding to the proportions of Fig. 51*b*. In general, it will be found that the circuit proportions recommended by different writers will vary somewhat in accordance with the relative emphasis placed upon the importance of uniform amplification, as against uniform time delay, and may differ somewhat from those given in Fig. 51.

Amplifier performance obtained in using two-terminal coupling networks with the proportions indicated in Fig. 51 are given by curves 1 to 4 of Fig. 53, and are also summarized in Table 2. It will be noted that the gain obtainable with satisfactory constancy of amplification and delay time increases with the complexity of the coupling network, and approximates 75 per cent of the theoretical maximum possible gain in the case of the circuit of Fig. 51*d*.

Four-terminal Coupling Networks—General Properties.—Figure 54 shows types of four-terminal coupling networks used in video-frequency amplifiers to obtain desir-

able performance at high frequencies. In these circuits the capacity C' associated with the plate circuit of the amplifier tube is separated from the capacity C'' associated with the grid electrode of the following tube, with the result that the gain obtainable over a given frequency band is greater than when the two capacities are paralleled, as in the case of the two-terminal coupling system.

An analysis of the properties of four-terminal coupling networks shows that the maximum possible amplification that can be obtained depends upon whichever of the capacities C' or C'' is maximum. The properties of tubes are such that normally $C'' > C'$, with the ratio being approximately 2. The effective ratio can be modified somewhat by proper disposal of circuit elements, such as blocking condensers, which have capacity to ground. Assuming that C'' is the larger capacity and denoting the reactance of this condenser as X_1'' at the highest frequency f_1 that is to be amplified, then the maximum theoretically possible amplification obtainable with the four-terminal network under practical conditions is

$$\text{Maximum possible amplification} = 2g_m X_1'' \tag{58}$$

This equation is of the same form as Eq. (57) for the two-terminal case, but since X_1'' in Eq. (58) is the reactance of C'' instead of $C' + C''$, the four-terminal network gives greater gain in accordance with the relation

$$\frac{\left.\begin{matrix}\text{Maximum practical gain}\\ \text{with four-terminal}\\ \text{coupling system}\end{matrix}\right\}}{\left.\begin{matrix}\text{Maximum possible gain}\\ \text{with two-terminal}\\ \text{coupling system}\end{matrix}\right\}} = \frac{C' + C''}{C''} \tag{59}$$

For a given total shunting capacity $C' + C''$, the most favorable division of capacities is $C' = C''$, corresponding to twice as much gain from the four-terminal system as obtainable with a two-terminal network. When $C'' = 2C'$, the four-terminal system has a possible amplification that is $\frac{3}{2}$ times the theoretical possible gain for a two-terminal network.

Design of Networks Used in Four-terminal Coupling Systems.—Four-terminal coupling networks having desirable characteristics for video-frequency voltage amplifiers are shown in Fig. 54. Network proportions that give good performance are indicated by the numbers associated with the circuit elements. These numbers give the impedances of the elements at the top frequency f_1 in terms of the impedance X_1'' of the capacity C'' at the top frequency f_1 to be amplified (assuming that $C'' > C'$). In the event that $C'' < C'$, one uses the same proportions, except that X_1'' is replaced by X_1', the reactance of C' at the frequency f_1, and the circuit is altered by exchanging the points at which the plate and grid electrodes connect to the network (as illustrated in Fig. 57). By the reciprocity theorem, the amplification and phase-shift characteristics will be unchanged by such an exchange of input and output terminals.

The circuits of Fig. 54 call for specified values of capacity ratio C''/C'. In practice, this can be obtained by padding out one or the other of the capacities by means of an adjustable trimming condenser to give the required ratio. Since the impedance level of the coupling network is determined by the largest shunting capacity, the maximum gain is obtained when the circuit is so arranged that the required capacity ratio can be achieved by padding out the smaller of the two condensers rather than by adding capacity in shunt with the larger.

The performance of the four-terminal networks of Fig. 54 with the proportions indicated is given in Fig. 53*b*. The essential properties are summarized in Table 2.

Comparison of Different Coupling Systems.—Examination of Fig. 53 and Table 2 shows that the amplification obtainable with satisfactory constancy of amplification and for small delay error is greater for the four-terminal networks than the two-terminal systems. Also, the amplification obtainable with each type of coupling system

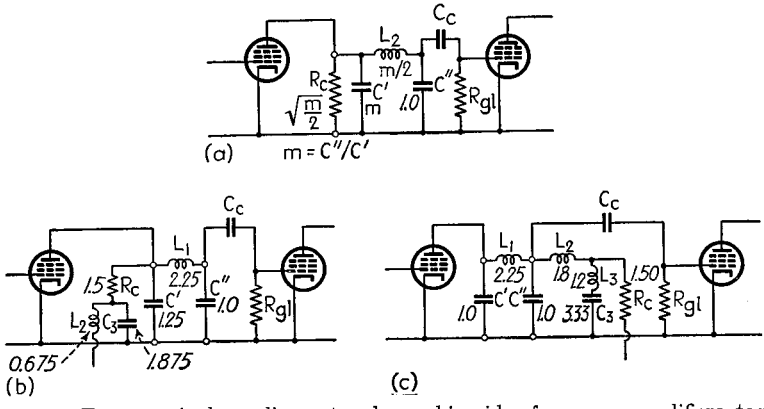


FIG. 54.—Four-terminal coupling networks used in video-frequency amplifiers, together with recommended circuit proportions giving optimum performance. The numbers shown are the element impedances at the top frequency f_1 that is to be amplified (reactance of C'' at f_1 taken as unity). The capacities C' and C'' include the tube capacities.

more closely approaches the theoretical possibilities as the number of reactive elements in the coupling network increases. Four-terminal systems give decidedly more gain.

TABLE 2.—PERFORMANCE OF SOME TYPICAL VIDEO-AMPLIFIER CIRCUITS HAVING HIGH-FREQUENCY COMPENSATION

Circuit	Relative voltage gain	Variation in amplification up to top frequency f_1	Maximum delay time error up to top frequency f_1 mc, in μs
Two-terminal cases			
Fig. 51*a*	1.0	0.02	$0.023/f_1$
Fig. 51*b*	0.85	0.035	$0.0077/f_1$
Fig. 51*c*	1.20	0.02	$0.012f_1$
Fig. 51*d*	1.50	0.035	$0.012/f_1$
Ideal	2.00	0	0
Four-terminal cases			
Fig. 54*a**	1.50	0.02	$0.012/f_1$
Fig. 54*b*†‡	2.25	0.04	$0.055/f_1$
Ideal	3.00	0	0

* For case $m = 2$.

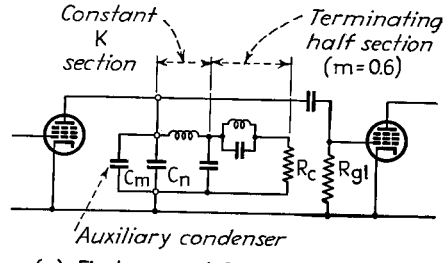
† Figure 54*c* acts substantially the same as Fig. 54*b*, but permits the use of a larger capacity C'.

‡ Amplification and time-delay characteristics calculated on assumption of perfect filter match to load. This assumption is almost perfectly realized up to the top frequency f_1.

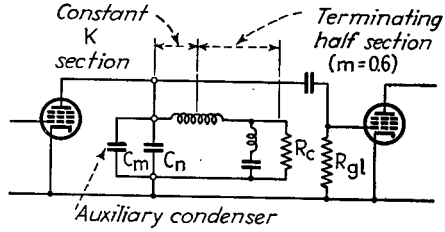
The four-terminal networks are, however, more complicated than the two-terminal ones, are more difficult to adjust, and require more care in construction. Also, the

more complicated networks, particularly the more complicated four-terminal arrangements, have a far sharper cutoff and a much more rapid deterioration of the phase characteristics around the top frequency. This is a disadvantage in most cases, and makes it necessary in practice to use a higher top frequency than would be required with the simple shunt-peaking circuit.

Coupling Networks Considered as Filters.[1]—The two-terminal coupling networks of Fig. 51 can be considered as special cases of general types of two-terminal coupling impedances based upon the properties of low-pass filters. These general two-terminal coupling networks are shown in Fig. 55. They consist of a filter that is matched on an image-impedance basis by an m-derived half section ($m = 0.6$) to a load consisting of coupling resistance R_c. This terminating section is then built out by adding sections on an image-impedance basis so that the input of the filter is at the mid-shunt point of a Type II (see Table 2, Sec. 3) (sometimes called constant k) π section. The input is then shunted by an auxiliary condenser C_m of such size as to make the impedance across the input terminals more nearly constant with frequency than is the mid-shunt impedance of a constant-k section. If X_c is the reactance of the total capacity across the filter input (auxiliary capacity C_m plus first shunting capacity C_n of the filter) at the cutoff frequency of the filter, then, on the assumption that there is a perfect impedance match of R_c to the filter, one has



(a) First general form



(b) Second general form

FIG. 55.—General two-terminal coupling networks based upon low-pass filters. The constant k sections are Type II sections of Table 2, Sec 3.

$$\left.\begin{matrix}\text{Magnitude of}\\ \text{impedance}\end{matrix}\right\} = \frac{R_c}{\sqrt{1 - \left[1 - \left(\frac{R_c}{2X_c}\right)^2\right]\left(\frac{f}{f_c}\right)^2}} \tag{60}$$

$$\text{Phase angle of impedance} = \phi = \tan^{-1}\frac{\left(\frac{R_c}{2X_c}\right)}{\sqrt{\left(\frac{f_c}{f}\right)^2 - 1}} \tag{61}$$

where R_c = load impedance with which filter is designed to operate.

X_c = reactance of total capacity ($C_n + C_m$) across filter input terminals at the cutoff frequency f_c.

f_c = cutoff frequency of filter.

The behavior for various values of $R_c/2X_c$ is shown in Fig. 56. For $X_c = R_c/2$, the impedance that is developed (and hence the amplification obtained) *is precisely constant up to the cutoff frequency f_c of the low-pass filter, and has a value $2X_c$.*[2] *This is the highest constant impedance that can possibly be developed across a capacity C_m over this*

[1] See Wheeler, *loc. cit.*

[2] The condition $X_c = R_c/2$ is obtained when the auxiliary shunting capacity C_m equals the capacity C_n of the filter, since for a constant k section, $(1/2\pi f_c C_n) = R_c$. Hence the total shunting capacity required across the input for uniform impedance is exactly the same as the full shunt capacity of the constant k or Type II section.

frequency range. The most linear phase characteristic is obtained when $X_c = 0.37R_c$, which corresponds, however, to a moderate falling off in impedance as cutoff is approached. The best compromise is when X_c is approximately $0.48R_c$.

The four-terminal networks of Fig. 54 are special cases of a general four-terminal coupling system based on the properties of wave filters, and illustrated in Fig. 57*a*. Here the impedance offered to the plate circuit of the amplifier tube is developed in the same way as in the case of the two-terminal coupling network, but now the impedance that can be developed is greater because part of the shunting capacity provided by the tubes is displaced along the filter to supply the shunting capacity

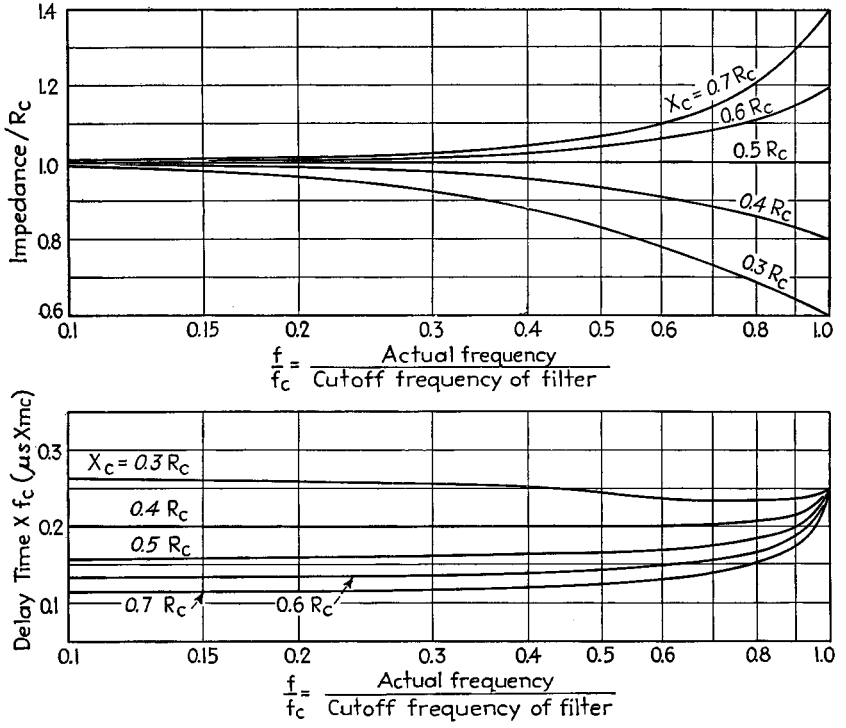
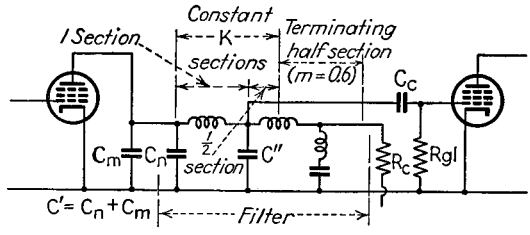


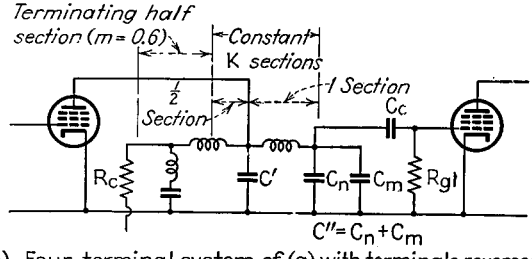
FIG. 56.—Effect of reactance X_c of the total capacity shunted across filter input at the cutoff frequency f_c upon the performance of the general coupling network of Fig. 55.

required at another shunting point. The increase in impedance level, and hence the increase in amplification, obtainable with the four-terminal arrangement depends upon the ratio of capacities C' and C'' associated with the input and output circuits of the coupling network. The impedance level is greatest when $C' = C''$, and is then exactly twice the impedance that can be developed in a two-terminal coupling system having the same total shunting capacity. The maximum possible gain obtained using a four-terminal network under such conditions is therefore exactly twice that obtainable with a two-terminal network. With other capacity ratios the relative advantage of the four-terminal arrangement is given by Eq. (59).

The separation of the grid and plate terminals existing in a four-terminal coupling system produces no effect except to introduce an extra phase shift. This is because the frequencies to be amplified are in the pass band of the filter and so are transmitted

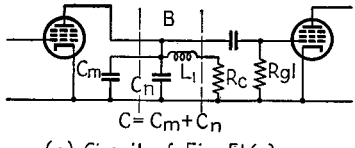


(a) Four-terminal system

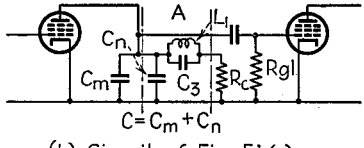


(b) Four-terminal system of (a) with terminals reversed

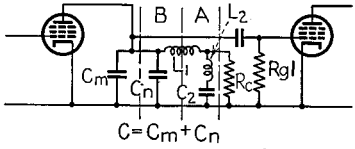
FIG. 57.—General forms of four-terminal coupling systems based upon filters. The constant k sections are Type II sections of Table 2, Sec. 3.



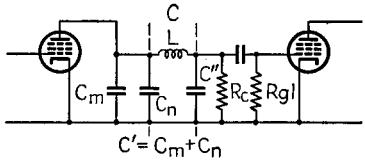
(a) Circuit of Fig. 51(a)



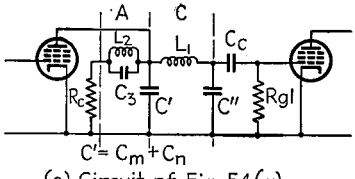
(b) Circuit of Fig. 51(c)



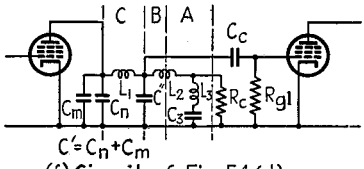
(c) Circuit of Fig. 51(d)



(d) Circuit of Fig. 54(a)



(e) Circuit of Fig. 54(c)



(f) Circuit of Fig. 54(d)

Key for sections
A = Terminating half section
B = Constant K half section
C = Constant K (inverse network) section

FIG. 58.—Circuits of Figs. 51 and 54 redrawn in the form of filter coupling networks. The dotted vertical lines show the division into sections (or half sections). The constant k sections are Type II sections of Table 2, Sec. 3.

through the filter sections between plate and grid without attenuation, but only with phase shift.

The circuits of Figs. 51 and 54 are special cases of the general two- and four-terminal coupling systems, as is apparent when the circuits are redrawn as in Fig. 58. It will be noted in both the two- and four-terminal cases that the performance approaches more closely to the ideal as a terminating half section ($m = 0.6$) is used to match the filter to the coupling resistance, and as enough sections are added to provide full mid-shunt constant-k capacities at the points on the filter with which the input and output terminals are associated.

In four-terminal coupling systems it is theoretically possible to introduce phase-correcting sections between the input and output points, as illustrated schematically

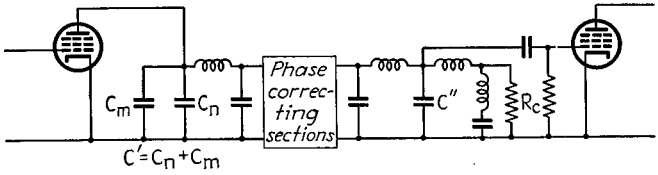


FIG. 59.—Diagram illustrating schematically the use of phase-correcting sections in a four-terminal coupling system.

in Fig. 59. In this way it is possible to obtain absolutely constant time delay, as well as absolutely constant amplification, up to the cutoff frequency of the filter. It is also possible in four-terminal systems to provide an impedance-transforming section when $C' \neq C''$ as indicated in Fig. 60. In this way full use is made of the existing shunting capacities, and maximum possible gain is obtained. Phase-correcting and impedance-transforming sections have not as yet had practical use, however.

Practical Adjustment of High-frequency Coupling Networks.—To obtain the desired high-frequency performance, it is necessary that the coupling network be adjusted in a systematic way. Trial-and-error procedure does not yield satisfactory results because of the many variables involved.

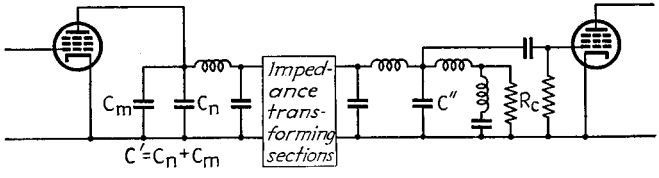


FIG. 60.—Diagram illustrating schematically the use of an impedance transforming-section in a four-terminal coupling system.

In the design of practical television amplifiers the high-frequency coupling network is calculated on the basis of shunting capacities C' and C'' slightly larger than those supplied by the tubes and wiring. Small trimming condensers are then used to pad out the capacities to the required value. In this way, it is possible to take care of variations in the capacities of individual tubes.

In four-terminal coupling systems, the network is built up by measuring the separate inductances, capacities, and resistance involved, and adjusting them separately to the calculated values, with the use of a bridge, Q meter, or other measuring device. In evaluating the capacities, the most accurate results will be obtained when the measurements are made after the coupling circuit has been mounted in final form so that the wiring capacities will be included.

In the case of two-terminal wiring systems one can obtain the proper circuit proportions either by separately measuring the individual circuit elements, as in the four-terminal case, or by using an experimental method to determine when the elements have the correct value. The former method is preferable when the network is complicated; the latter, when it is simple.

The experimental procedure for the shunt-peaking circuit consists in first connecting the input and output tubes in place and biasing the output tube so that it acts as a vacuum-tube voltmeter. The coupling resistance R_c is then short-circuited, and the inductance L_1 is adjusted by trial and error until the curve of amplification versus frequency has its peak at the frequency f_0 corresponding to the desired ratio f_1/f_0. With L_1 shorted, R_c is then adjusted until the amplification at the frequency $Q_0 f_0$ is 0.707 times the amplification at a moderate frequency, where Q_0 is the desired circuit Q (see Fig. 52). The experimental adjustment of the circuit shown in Fig. 51*c* follows an analogous procedure. First C_3 is disconnected, after which L_1 and R_c are adjusted as in the shunt-peaking case to give $Q_0 = 0.62$, and $f_0 = 1.36 f_1$. The proper value for C_3 is then obtained by short-circuiting R_c and adjusting C_3 until the peak of amplification occurs at the frequency $0.95 f_1$.

After the proper circuit proportions have been installed according to calculations or experiment, it is desirable to check the final results by taking a curve of amplification as a function of frequency. If the resulting gain has the expected uniformity up to the top frequency f_1 to be amplified, it is then almost certain that the phase-shift characteristics will likewise be satisfactory. An alternative method of checking the high-frequency characteristics of an amplifier is to apply to the amplifier a steep wave front, such as is obtainable from a square-wave oscillator, and to observe the output wave that results.

Reproduction of High-frequency Transients by Video Amplifiers.[1]—The severe requirements that video amplifiers must meet at high frequencies arise from the need of accurately reproducing abrupt changes in amplitude of the applied voltage, such as is represented by the unit pulse in Fig. 61*a*. The analysis of amplifier behavior under the action of a unit pulse is quite involved, but the general character of the relations that exist between amplifier characteristics and the effects produced in amplifying a pulse are indicated in Fig. 61. It will be noted that the amplifier causes the pulse to rise gradually instead of abruptly, and also that there is a tendency for the amplifier output to overshoot its final value. The character of the behavior depends upon the constants of the amplifier, as indicated in Fig. 61*b*, which applies to the shunt-peaking circuit. It will be noted that the best reproduction of the pulse is when Q_0 is in the range 0.7 to 0.6, which corresponds to gain and time delay characteristics that are most nearly constant.

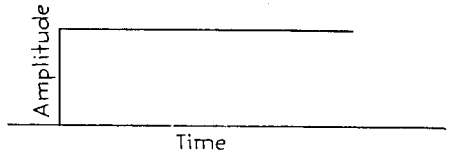
The rate at which the amplifier output rises as the pulse is applied is inversely proportional to the top frequency that can be amplified without serious loss in gain. This means that the greater the band width the more abruptly will the output rise, and hence the more closely will the applied pulse be reproduced.

The effect of applying a pulse to a multistage amplifier is shown in Fig. 61*c*. As the number of stagings increases, the rate of rise of output becomes more gradual, and the tendency to overshoot is more pronounced.

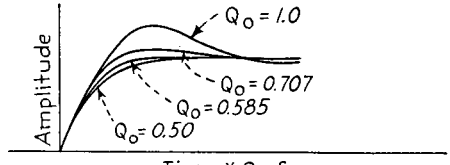
The result of applying a pulse to an amplifier having an ideal filter type of coupling is shown in Fig. 61*c* for the case where there are constant amplification and delay time up to the filter cutoff and no amplification at higher frequencies. For comparison, a dotted curve is shown for a shunt-peaking circuit having the top frequency equal to the cutoff frequency of the filter. The shunt-peaking circuit is superior to the filter circuit in spite of the fact that below its top frequency the shunt-peaking

[1] Bedford and Fredendall, *loc. cit.*

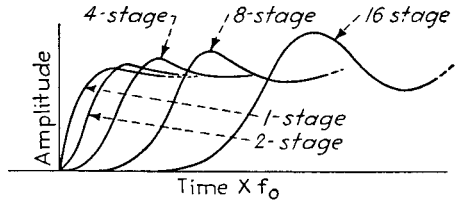
arrangement does not have exactly constant amplification and constant time delay. This superiority is because the shunt-peaking circuit gives amplification at frequencies appreciably greater than the top frequency, whereas the ideal filter does not. This emphasizes the fact that sharp cutoff, such as is obtained with filter coupling systems, is not always desirable.



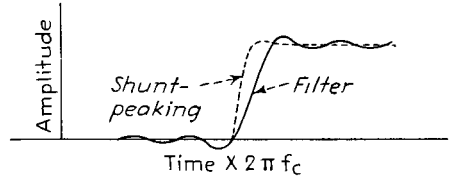
(a) Applied pulse



(b) Output wave for one shunt-peaking stage



(c) Output wave from multistage Amplifier



(d) Output wave from ideal filter coupling system compared with shunt-peaking circuit with same nominal top frequency

FIG. 61.—Response obtained with various amplifying systems when an impulse is applied to the amplifier input.

Combination of High-frequency Coupling Network with Low-frequency Equalizing Circuits.—In practical video amplifiers the low-frequency equalizing circuits are placed on the low potential side of the high-frequency network, as shown in Fig. 62. With this arrangement the presence of the high-frequency network has no effect upon the low-frequency equalization, because at very low frequencies the high-frequency network simply reduces to the coupling resistance R_c. At the same time, the low-frequency equalizing circuits are by-passed to ground at high frequencies by the shunting capacity that is a part of this circuit, and so become inoperative at high frequencies.

Inasmuch as large condensers frequently possess a certain amount of inductance, it is desirable to shunt the equalizing condenser in the plate circuit, and also the bias and screen by-pass condensers, with small mica condensers having a capacity of the order of 0.001 to 0.01 μf, to ensure low impedance to the higher video frequencies.

17. Video-frequency Power Amplifiers.—Video-frequency power amplifiers differ from voltage amplifiers in that the objective is power developed in a load impedance rather than voltage developed across a capacity. The chief use of video-frequency power amplifiers is the production of energy to be transmitted over a cable or transmission line. Amplifiers for this purpose present special problems, since the load impedance is small, and the wide-frequency band involved makes it impossible to use transformers for impedance matching.

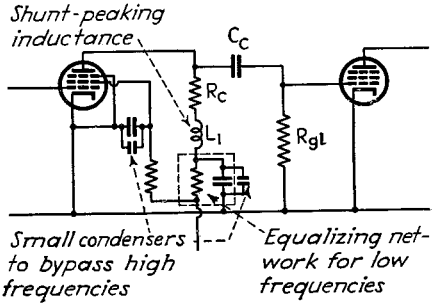


FIG. 62.—Video-amplifier circuit showing combination of high-frequency and low-frequency equalization.

Circuit arrangements for video-frequency power amplification are shown in Fig. 63. Of these, cathode coupling is generally used for reasons discussed below. The d-c plate current of the tube can, if necessary, be blocked from the load by a coupling condenser C_c and shunt-feed resistor R, as shown in Figs. 63*b* and 63*d*. This is to be avoided wherever possible, however, since the shunt-feed resistance reduces the power output obtainable, and the blocking condenser must be of tremendous size if excessive phase shift is to be avoided at low frequencies.[1]

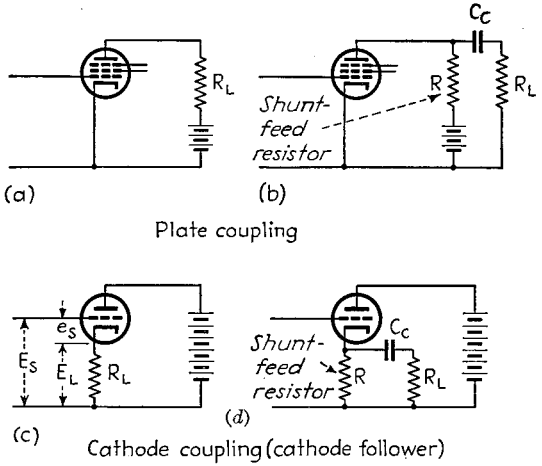


FIG. 63.—Circuits for video-frequency power amplifiers.

The maximum power that a Class A amplifier tube can develop in a load resistance R_L that is small compared with the plate resistance of the tube is

$$\left.\begin{matrix}\text{Power that can be delivered}\\ \text{to load resistance}\end{matrix}\right\} = \frac{I_p^2 R_L}{2} \qquad (62)$$

[1] Values of the order of 1,000 μf are required when the load is a cable.

where I_p is the d-c plate current. Equation (62) assumes that the plate current varies from zero to twice the d-c value during operation. If amplitude distortion is to be kept small, the power actually realizable with practical tubes is of the order of one-half to three-fourths that given by Eq. (62).

Examination of Eq. (62) shows that if a large amount of energy is to be delivered to a fixed load impedance such as that represented by a cable or transmission line, it is necessary that the d-c plate current be large. When a single tube is not capable of delivering the desired power, tubes can be paralleled. The power output that is obtained will then be proportional to the *square* of the number of tubes, with the maximum number of tubes that can be employed limited by the fact that the top frequency to be amplified sets a limit to the capacity that can be allowed in shunt with the load.

Cathode-coupled Power Amplifier.[1]—The *cathode-coupled* (also called *cathode-follower*) circuit of Fig. 63*c* is generally preferred for power amplification at video frequencies. Compared with arrangements that place the load impedance in the plate circuit, the cathode-coupled amplifier has the advantages of (1) permitting one of the load terminals to be grounded; (2) better amplitude and phase characteristics at high frequencies as a result of negative feedback action; (3) higher input impedance as a result of negative feedback; (4) less amplitude distortion as a result of negative feedback; (5) a source impedance presented by the tube to the load that is small enough so that it can be readily matched to a transmission line.

The voltage developed across the load in the cathode-coupled stage of Fig. 63*c* is related to the applied voltage by the equation

$$E_L = g_m Z_{eq} E_s \tag{63}$$

where E_L = voltage developed across the load.
E_s = voltage applied to the amplifier input.
g_m = transconductance of tube.
Z_{eq} = equivalent load impedance formed by the actual load impedance in parallel with a resistance $R_p/(1 + \mu)$, where R_p is the plate resistance and μ, the amplification factor, respectively, of the tube.

Examination of Fig. 63*c* shows that the voltage E_L developed across the load opposes the applied signal E_s as far as potential between the grid and cathode is concerned, so that the circuit exhibits negative feedback characteristic. If the equivalent impedance Z_{eq} and also the transconductance g_m are not too low, the voltage E_L developed across the load will be almost but not quite equal to the applied signal voltage E_s. Under no conditions can the voltage developed across the load exceed the applied signal.

Equation (63) can also be rewritten as

$$E_L = \frac{\mu}{\mu + 1} E_s \frac{R_L}{R_L + \left(\frac{R_p}{1 + \mu}\right)} \tag{63a}$$

The notation is the same as before. Examination of Eq. (63*a*) shows that as far as the load impedance is concerned, the presence of negative feedback cause the tube to act as though it had an amplification factor $\mu/(\mu + 1)$ and a plate resistance[2] of $R_p/(1 + \mu)$.

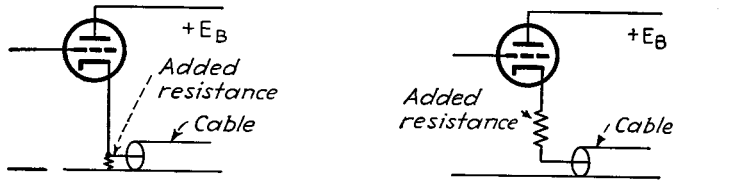
The low equivalent tube impedance that is presented to the load in a cathode-coupled circuit is of considerable practical importance. *First*, it has the effect of improving the response characteristics at high frequencies. *Second*, it makes possible

[1] This discussion follows Albert Preisman, Some Notes on Video Amplifier Design, *R.C.A. Rev.*, Vol. 2, p. 421, April, 1938.

[2] This reduction of equivalent plate resistance by the action of negative feedback is the same here as discussed in Par. 11.

an impedance match between the tube and the load presented by a cable or other transmission line. Such matching eliminates reflections at the sending end of the cable, and so reduces echoes. When a match between a tube and transmission line is required, a triode tube is selected such that with normal operation the equivalent tube impedance $R_p/(1 + \mu)$ will approximate the characteristic impedance of the transmission line. A final adjustment is then made by adding resistance in shunt with the input of the line if the tube resistance is too high or placing a resistance in series if the tube resistance is too low, as shown in Fig. 64.

The negative feedback action in the cathode-coupled stage causes the voltage e_s in Fig. 63c that exists between control grid and cathode to be considerably less than the signal E_s. As a result, any impedance Z_{gk} from cathode to control grid appears



(a) Impedance matching when equivalent tube resistance is too high

(b) Impedance matching when equivalent tube resistance is too low

FIG. 64.—Methods of obtaining exact impedance match between a transmission line and a cathode-follower stage.

to the applied voltage E_s to have a value greater than the actual impedance in accordance with the relation

$$\left.\begin{matrix}\text{Equivalent grid-cathode} \\ \text{impedance}\end{matrix}\right\} = \frac{\text{actual grid-cathode impedance}}{1 - (E_L/E_s)} \tag{64}$$

Equation (64) has three important consequences. *First*, if the grid-leak resistance of the driver stage is connected from grid to cathode, a low resistance that may be needed to carry gas currents will still appear to be a high resistance to the applied signal. *Second*, the effective capacitive loading on the driver stage is reduced according to the relation

$$\left.\begin{matrix}\text{Equivalent input} \\ \text{capacity}\end{matrix}\right\} = \left\{\begin{matrix}\text{actual input} \\ \text{capacity}\end{matrix}\right\} \times \left(1 - \frac{E_L}{E_s}\right) \tag{64a}$$

Third, placing tubes in parallel to increase the output power of a cathode-coupled stage does not make the impedance that is offered to an applied voltage E_s vary inversely as the number of tubes in parallel because with usual circuit proportions $(1 - E_L/E_s)$ is at least roughly inversely proportional to the number of tubes.[1]

High-frequency Characteristics of Video-frequency Power Amplifiers.—The amplification and phase shift of video-frequency power amplifiers at high frequencies can be calculated from the equivalent circuit of Fig. 65. Here the tube is considered as supplying a constant current $g_m E_s$ to an equivalent impedance formed by the parallel combination of the actual load impedance Z_L, the effective internal resistance of the tube, and the shunting capacity C. The internal resistance of the tube is taken as the plate resistance for all cases except cathode coupling,[2] where it is $R_p/(1 + \mu)$.

[1] Impedances between control grid and other electrodes, such as screen grid or plate, are not reduced by the negative feedback action, unless the return circuit for these electrodes is to the cathode instead of to ground.

[2] With cathode coupling, the effective transconductance is, however, taken as the actual transconductance.

The shunting capacity will depend upon the method of coupling. With coupling in the plate circuit (Fig. 63*a*) it consists of the output capacity of the tube. With cathode coupling it consists of the capacity between cathode and ground, plus the cathode-heater capacity when the heater is grounded, or the capacity between the primary and secondary of the filament transformer if the heater is connected to the cathode or if a filament tube is used.

When the actual load impedance is a resistance independent of frequency, as is normally the case, the amplification and the time-delay characteristics will depend upon the ratio X/R, where X is the reactance of the shunting condenser and R is the equivalent load resistance formed by the actual load resistance in parallel with the effective internal resistance of the tube. The amplification and time-delay characteristics that result are shown in Fig. 66. For the same perfection in behavior as normally required with video-frequency voltage amplification, the top frequency should not exceed one-half of the frequency f_0 at which the reactance of the shunting condenser equals the equivalent resistance R. It is to be noted that cathode coupling employing a triode tube will give a better high-frequency characteristic than cathode coupling with a pentode tube or any form of plate coupling, because of the lower effective plate resistance in the equivalent circuit of Fig. 63*c*.

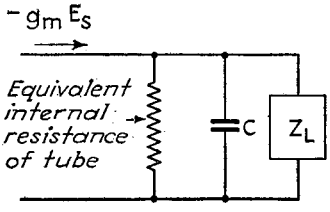


FIG. 65.—Equivalent plate circuit of a video-frequency power amplifier.

Video-frequency Amplifiers for Developing Large Voltages.[1]—Amplifiers in which the objective is to develop the maximum possible output voltage without excessive distortion, irrespective of the applied signal required, are special forms of power amplifiers of practical importance. Such amplifiers are employed to operate cathode-ray tubes in television receivers, and also for grid modulation of television transmitters.

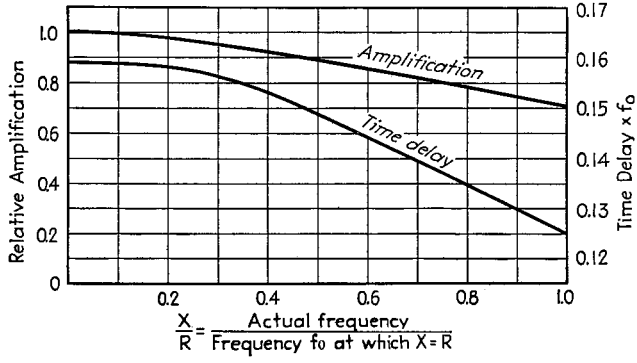


FIG. 66.—Amplitude and time-delay characteristics of a video-frequency power amplifier.

The coupling arrangements used to develop large output voltages depend upon the character of the impedance to which this voltage is applied. When this impedance is a capacity, as in the case of a cathode-ray tube, one can employ one of the usual two- or four-terminal networks, such as are used in voltage amplification, with the capacity of the load taking the place of the input capacity of the next stage in the corresponding voltage amplifier case. On the other hand, when the load across which

[1] Further discussion is given by D. E. Foster and J. A. Rankin, Video Output Systems, *R.C.A. Rev.*, Vol. 5, p. 409, April, 1941.

the large voltage is to be developed consists of a resistance, or a resistance shunted by a capacity, then one is restricted to the circuits of Figs. 54*a* and *b*.

The maximum output voltage that can be developed by an amplifier tube is

$$\text{Peak a-c output voltage} = I_p R_L \tag{65}$$

where I_p = d-c plate current of amplifier.

R_L = load resistance used in coupling network.

When the output capacity of the tube is the biggest capacity involved in the coupling network, paralleling tubes to increase the current I_p in Eq. (65) will make it necessary to use a proportionately lower load resistance R_L, and so will not increase the output voltage obtainable. The only way in which more voltage can be realized under such restrictions is to change the type of tube to one having more favorable characteristics. On the other hand, as long as the capacity associated with the load determines the design of the coupling network, then paralleling tubes will enable increased output voltages to be obtained.

18. Tubes for Video Amplification.—The merit of a tube as a video amplifier depends upon the particular application involved. In the case of video-frequency voltage amplifiers using two-terminal networks, the amount of amplification obtainable over a given band width is proportional to the figure of merit:

$$\left.\begin{array}{l}\text{Figure of merit for}\\ \text{voltage amplification}\\ \text{with two-terminal}\\ \text{coupling networks}\end{array}\right\} = \frac{g_m}{C_g + C_p} \tag{66a}$$

where C_g is the input capacity of the tube (sum of grid-cathode and grid-screen capacities) and C_p is the output capacity (sum of plate-screen and plate-ground capacities). With four-terminal networks, Eq. (66*a*) becomes

$$\left.\begin{array}{l}\text{Figure of merit for}\\ \text{voltage amplification}\\ \text{with four-terminal}\\ \text{coupling networks}\end{array}\right\} = \frac{g_m}{C} \tag{66b}$$

where C is either the input or the output capacity of the tube, whichever is the larger. Finally, in the ideal case of a four-terminal coupling network maintaining the same impedance level at the input and output terminals, the merit is

$$\left.\begin{array}{l}\text{Figure of merit for}\\ \text{voltage amplification}\\ \text{in ideal case}\end{array}\right\} = \frac{g_m}{\sqrt{C_g C_p}} \tag{66c}$$

Examination of these relations shows that, irrespective of the exact conditions, what is desired is a large transconductance in proportion to electrode capacities. In order to achieve this, it is necessary to make the interelectrode spacings of the tube very small.[1] It is of no avail to increase the area of the electrode structures, since doing so increases the capacities in proportion to the gain in transconductance. Commercial tubes with closely spaced electrodes (such as the 1852) are capable of giving several times as much voltage amplification as the corresponding tubes ordinarily used in resistance-coupled amplification.

The figure of merit of an amplifier tube intended to develop as large an output voltage as possible is

[1] See A. P. Kauzmann, New Television Amplifier Receiving Tubes, *R.C.A. Rev.*, Vol. 3 p. 271, January, 1939.

$$\left.\begin{matrix}\text{Figure of merit}\\ \text{for large output}\\ \text{voltage}\end{matrix}\right\} = \frac{I_p}{C_p} \tag{67}$$

When tubes having the same figure of merit according to Eq. (67) differ in input capacity, then the one with the lower input capacity is to be preferred. In particular, triode tubes, because of their high effective input capacity, are always to be avoided, even though their figure of merit may be quite satisfactory. Tube characteristics favoring high merit in Eq. (67) are a large cathode, large electrode areas, and relatively large spacing. These are the characteristics of power tubes, leading to the conclusion that pentode and beam power tubes are best for developing large voltages.

Figures of merit of various representative tubes for voltage-amplifier service and for developing large output voltages are given in Table 3.

TABLE 3.—CHARACTERISTICS OF TYPICAL TELEVISION TUBES

Tube type	Rated d-c plate current, ma	Transconductance at rated plate current, μ mhos	Input capacity, $\mu\mu$f	Output capacity, $\mu\mu$f	Figures of merit for		
					Voltage gain		Large output voltage
					Two terminal	Four terminal	
Receiver pentodes:							
6D6	8.2	1,600	6	6	125	250	1.4
954	2.0	1,400	3	3	230	465	0.7
1851	10.0	9,000	15	5	450	600	2.0
1853	12.5	5,000	10	5	330	500	2.5
Transmitter pentodes:							
802	20.0	2,250	12	8.5	110	190	2.3
803	62.5	4,000	17.5	29	85	140	2.2
837	24.0	3,400	16	10	130	210	2.4
Beam power tubes:							
6L6	72.0	6,000	13	11	250	460	6.6
807	72.0	6,000	11	7	330	550	10.3

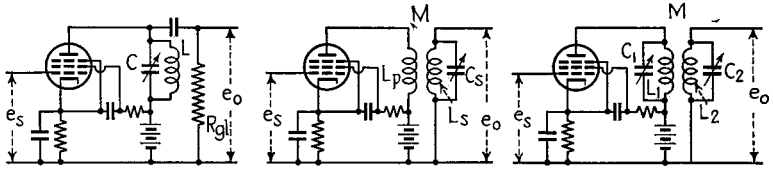
TUNED AMPLIFIERS

19. Tuned Voltage Amplifiers.—A tuned voltage amplifier is a Class A voltage amplifier in which the load impedance is supplied by a resonant circuit. Circuits of typically tuned amplifiers are shown in Fig. 67. Such amplifiers are used to amplify signal frequency voltages in radio receivers, in which case they are referred to as tuned radio-frequency amplifiers, and also to amplify intermediate-frequency voltages in superheterodyne receivers.

The important characteristics of a tuned voltage amplifier are the amplification at resonance, the variation of amplification with frequency in the immediate vicinity of resonance, and the way in which the amplification obtainable changes as the resonant frequency of the tuned circuits is varied.

Characteristics of Simple Tuned Amplifiers.—Tuned amplifiers using only a single resonant circuit to provide the load impedance, as in Figs. 67*a* and 67*b*, have a curve of amplification as a function of frequency that approximates the shape of an ordinary

resonance curve having a Q somewhat lower than the actual Q of the resonant circuit (see Fig. 69). It is accordingly possible to describe the amplification characteristic of such an amplifier in terms of the amplification at resonance and the effective Q of the amplification curve. This leads to the following formulas:



(a) Direct coupling (b) Transformer coupling (c) Band-pass coupling

FIG. 67.—Typical tuned-amplifier circuits.

For direct-coupled circuit of Fig. 67*a*:

$$\left.\begin{matrix}\text{Amplification at}\\ \text{resonance}\end{matrix}\right\} = g_m \frac{\omega LQ}{1 + \frac{\omega LQ}{R_p} + \frac{\omega LQ}{R_{gl}}} \tag{68}$$

$$\frac{\left.\begin{matrix}\text{Effective } Q \text{ of}\\ \text{amplification curve}\end{matrix}\right\}}{\left.\begin{matrix}\text{Actual } Q \text{ of tuned}\\ \text{circuit}\end{matrix}\right\}} = \frac{1}{1 + \frac{\omega LQ}{R_p} + \frac{\omega LQ}{R_{gl}}} \tag{69}$$

For transformer-coupled circuit of Fig. 67*b*:

$$\left.\begin{matrix}\text{Amplification}\\ \text{at resonance}\end{matrix}\right\} = g_m \frac{\omega MQ}{1 + \frac{(\omega M)^2/R_s}{R_p}} \tag{70}$$

$$\frac{\left.\begin{matrix}\text{Effective } Q \text{ of}\\ \text{amplification curve}\end{matrix}\right\}}{\left.\begin{matrix}\text{Actual } Q \text{ of tuned}\\ \text{circuit}\end{matrix}\right\}} = \frac{1}{1 + \frac{(\omega M)^2/R_s}{R_p}} \tag{71}$$

In these equations, g_m is the transconductance and R_p the plate resistance of the amplifier tube; Q is the actual Q of the resonant circuit; and the remaining notation is as illustrated in the figures.

In the case of pentode tubes, the plate resistance R_p is very much greater than the parallel impedance or coupled impedance of the resonant circuit. Also, in practical amplifiers, the grid-leak resistance R_{gl} is likewise much larger than the resonant circuit impedance. Accordingly, with pentode tubes the effective Q of the amplification curve can be taken as approximating the actual Q of the resonant circuit, while the amplification at resonance is given by the following relations:

For direct-coupled circuit of Fig. 67*a*:

$$\left.\begin{matrix}\text{Approximate amplification}\\ \text{at resonance}\end{matrix}\right\} = g_m \omega LQ \tag{72}$$

For transformer-coupled circuit of Fig. 67*b*:

$$\left.\begin{matrix}\text{Approximate amplification}\\ \text{at resonance}\end{matrix}\right\} = g_m \omega MQ \tag{73}$$

With triode tubes, the effective Q of the amplification curve depends greatly upon the coupling between the resonant circuit and the plate circuit of the tube. Also there is an optimum coupling that gives maximum gain. Because of this necessity of impedance matching, the transformer-coupled circuit of Fig. 67*b* is normally used with triodes. The amplification, as given by Eq. (70), is maximum when the resistance $(\omega M)^2/R_s$ coupled into the plate circuit of the tube by the tuned circuit equals the plate resistance of the tube. The effective Q under these conditions is one-half of the actual Q of the resonant circuit. With greater coupling the effective Q is reduced and the amplification becomes less, whereas with smaller coupling, the amplification drops, and the effective Q increases, as shown in Fig. 68.

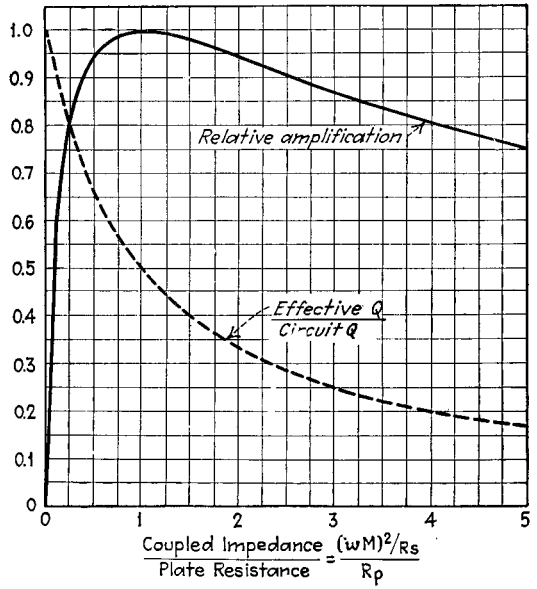


FIG. 68.—Effect of matching between load impedance and plate resistance upon the relative amplification and effective Q of a tuned amplifier.

Band-pass Amplifiers.—Band-pass coupling gives a band-pass type of amplification characteristic as shown in Fig. 69. The shape of the curve obtained, however, differs from the band-pass characteristic of the actual coupled circuits, corresponding instead to the band-pass characteristic obtained with a primary circuit having an effective Q lower than the actual Q of this circuit according to the relation[1]

$$\left.\begin{matrix}\text{Effective } Q \text{ of} \\ \text{primary circuit}\end{matrix}\right\} = \frac{\text{Actual } Q \text{ of primary}}{1 + \dfrac{(1/\omega C_1)^2}{R_1 R_p}} \tag{74}$$

where the notation is as in Fig. 67*c*, with the addition that R_1 is the series resistance of the primary tuned circuit. With pentode tubes, the effective Q closely approximates the actual Q. If Eq. (74) is taken into account, the shape of the amplification characteristic obtained with band-pass coupling can be readily determined with the aid of the principles given in Par. 9, Sec. 3. It was there shown that when the primary and secondary circuit Q's were approximately the same, then the width of

[1] For derivation, see F. E. Terman, "Radio Engineering," 2d ed., p. 216, McGraw-Hill, New York, 1937.

the response band was determined by the coefficient of coupling k according to the approximate relation

$$\frac{\text{Width of response band}}{\text{Mean frequency}} = 1.2k \tag{75a}$$

Substantially constant response is then obtained over the band when the primary and secondary Q's, Q_1 and Q_2, respectively, are so related that

$$\sqrt{Q_1 Q_2} = \frac{1.75}{k} \tag{75b}$$

When the primary and secondary Q's are considerably different, as, for example,

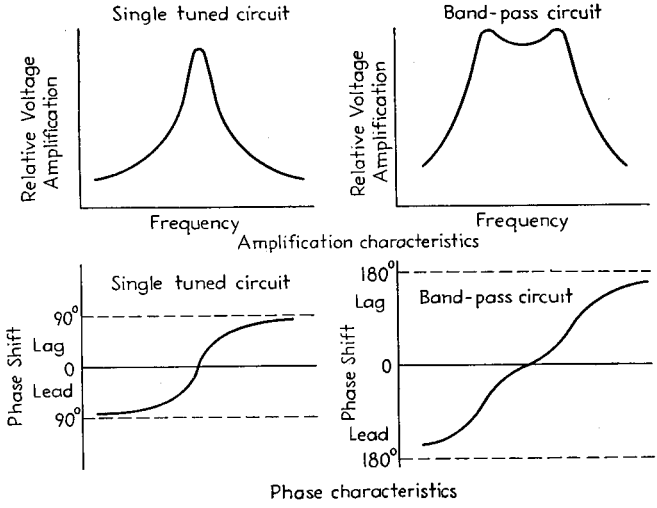


FIG. 69.—Curves showing amplification and phase-shift characteristic as a function of frequency with tuned amplifiers having a single tuned circuit and having a band-pass characteristic.

when practically all the losses of the system are in one of the resonant circuits, then the proper coefficient of coupling is given by the equation

$$\text{Desired coefficient of coupling} = k\sqrt{1 + 0.33a^2} \tag{75c}$$

where k = coefficient of coupling that would be required if the primary and secondary had equal Q's, as given by Eq. (75a).

$$a = \frac{Q_1 - Q_2}{Q_1 + Q_2}.$$

When the two circuits are resonant at the same frequency, the amplification at resonance is

$$\left.\begin{matrix}\text{Amplification at} \\ \text{resonant frequency}\end{matrix}\right\} = g_m k \frac{\omega_0 \sqrt{L_1 L_2}}{k^2 + \dfrac{1}{Q_1 Q_2}} \tag{76}$$

where g_m = transconductance of tube.
k = coefficient of coupling between primary and secondary inductances.
ω_0 = 2π times resonant frequency.
Q_1 = effective Q of primary circuit as obtained from Eq. (74).

Q_2 = Q of secondary circuit.

L_1, L_2 = primary and secondary inductances, respectively.

The amplification is maximum when the coefficient of coupling has the critical value, *i.e.*, when $k = 1/\sqrt{Q_1 Q_2}$, and for this condition becomes

$$\left.\begin{matrix}\text{Maximum possible}\\ \text{amplification}\\ \text{at resonance}\end{matrix}\right\} = g_m \frac{\omega_0 \sqrt{L_1 L_2}\sqrt{Q_1 Q_2}}{2} \tag{77}$$

Band-pass behavior can also be obtained by combining a double-humped curve of one stage having an overcoupled pair of resonant circuits, with a peaked curve of another stage having a single tuned circuit. Such systems are discussed in connection with Fig. 37, Sec. 3. Band-pass action can also be obtained by tuning primary and secondary circuits to slightly different frequencies, or by similarly detuning alternate stages of a multistage amplifier. These last arrangements are very seldom used, however, because of the difficulty of adjusting the circuits properly under field

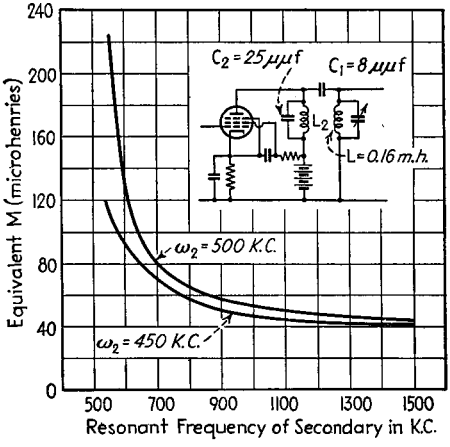


FIG. 70.—Variation of equivalent mutual inductance with frequency in a typical complex coupling system.

conditions, together with the fact that the gain with detuning is less than when the same pass band is obtained with primary and secondary resonant at the same frequency.

Variation of Amplification with Resonant Frequency—Use of Complex Coupling.—With direct and transformer coupling, the amplification tends to be proportional to resonant frequency when the circuit Q is approximately constant. In the case of tuned radio-frequency amplifiers that must be adjustable to amplify signals over a considerable frequency range, this results in greater amplification (with a consequent increased tendency toward regeneration) at the higher frequencies. This undesirable tendency can be controlled by the use of complex coupling systems such as discussed in Par. 8, Sec. 3, in which the equivalent mutual inductance varies with frequency. An example of such a circuit is shown in Fig. 70, and other possible coupling networks are given in Figs. 28 to 34 in Sec. 3. By proper proportioning of these circuits it is possible to control the variation of equivalent mutual inductance with frequency so that the amplification will be the same at two or more widely different frequencies.

Tuned amplifiers involving complex coupled systems can be reduced to an equivalent transformer coupled case by the expedient of reducing these circuits to an equiva-

lent inductively coupled circuit, as discussed in connection with Fig. 35, Sec. 3, and then calculating the behavior of the resulting equivalent circuit by the formulas for transformer coupling.

Regeneration in Multistage Tuned Amplifiers.—Multistage tuned amplifiers must be carefully arranged to avoid regeneration. Coils in the tuned circuits of different stages must be either shielded, or spaced a considerable distance apart with suitable orientation. Tuning condensers must likewise be either shielded or otherwise arranged to minimize coupling between stages. With triode tubes the grid-plate capacity must be neutralized (see Par. 25). High potential leads associated with stages of widely different power levels must be carefully located, and in some cases shielded. By-pass condensers are required to prevent currents from the different stages from flowing through impedances that are common to several stages. Grounds must be arranged so that the chassis does not serve as a common return for circuits of different power levels and so become a source of coupling between stages.

The essential factors involved in locating grounds are illustrated in Fig. 71. If the coils of resonant circuits are returned to ground, as shown by b_3 and b_4, instead

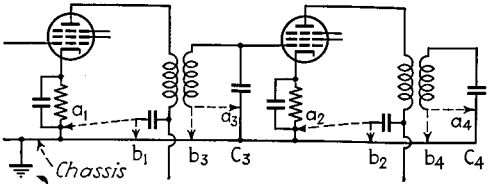


FIG. 71.—Correct and incorrect methods of locating ground connections in a multistage tuned amplifier.

of directly to the condensers, as by a_3 and a_4, then the chassis will carry tuned circuit currents from b_3 to c_3 and b_4 to c_4. The paths of these currents through the chassis will inevitably overlap to some extent, resulting in energy transfer between stages that will cause regeneration. A similar situation exists for the point at which the by-pass condenser returns to ground, the proper arrangements in Fig. 71 being a_1 and a_2 and the improper ones, b_1 and b_2.

The principal effect of regeneration in multistage tuned amplifiers is ordinarily an increase in the effective Q of one or more of the individual tuned circuits together with a slight change in the capacity required to tune these circuits to a given frequency. The increase in effective Q produces an increase in the amplification and in the selectivity.

When tuned radio-frequency amplifiers were first used, regeneration was frequently introduced intentionally in order to increase the amplification and selectivity. The development of improved amplifiers has caused this to be considered poor practice, because the regeneration varies with the resonant frequency, the volume setting, etc., is critical with respect to tube voltages, and causes oscillations when excessive.[1]

When regeneration is carried as far as possible without the production of oscillations, the amplification obtained is greater for weak than for strong signals. This is

[1] The use of negative feedback makes it possible to stabilize the regeneration, and thereby obtain selectivity corresponding to a much higher Q than that of a resonant circuit without the necessity of critical adjustments. See F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 27, p. 649, October, 1939; Shepard Roberts, Stabilized Regenerative Amplifier, *Rev. Sci. Instruments*, Vol. 9, p. 429, December, 1938.

because the third-order curvature of the tube characteristic causes the effective plate resistance of the tube to be slightly greater for large signals than for small signals, thus reducing the regeneration for large signals. Analysis shows that when the regeneration is as great as possible without oscillation in the absence of a signal, the amplification is inversely proportional to the two-thirds power of the signal voltage and directly proportional to the response obtained when no regeneration is present.[1]

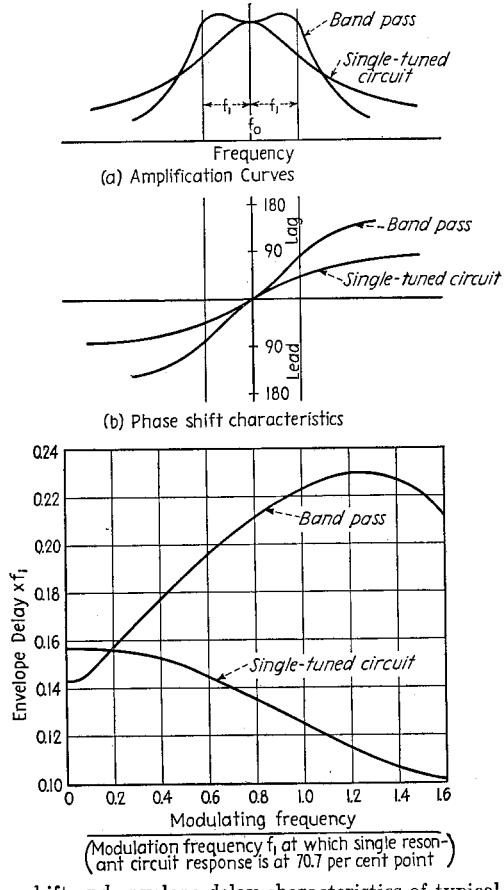


FIG. 71a.—Phase-shift and envelope-delay characteristics of typical tuned amplifiers.

Tubes.—Tuned voltage amplifiers practically always employ pentodes. Such tubes have negligible direct grid-plate capacity and so do not need be neutralized as do triodes. Also, they give more gain per stage and result in an effective Q of the amplification curve that more closely approaches the actual Q of the resonant circuit.

Volume Control.—Volume control of tuned amplifiers is normally obtained by using variable-mu tubes, and varying the bias or the screen voltage. This varies the transconductance and hence the gain, without appreciably affecting the resonant frequency of the circuits.[2] Variable-mu tubes are preferred to those having sharp

[1] Balth. van der Pol, The Effect of Regeneration on the Received Signal Strength, *Proc. I.R.E.*, Vol. 17, p. 339, February, 1929.

[2] The effect on tuning is discussed on p. 473.

cutoff, because they give a less critical control, and introduce less cross-modulation with low volume settings.

Phase and Delay Characteristics of Tuned Amplifiers.—Tuned amplifiers introduce a phase shift that varies with frequency, and so give rise to time-delay effects in the same manner as audio- and video-frequency amplifiers. The characteristic of importance in the case of tuned amplifiers is the variation in phase of the envelope of the wave with modulation frequency, since this introduces a time delay in the envelope that appears as time delay in the output after demodulation. The *envelope delay* for a modulated wave is

$$\left.\begin{matrix}\text{Envelope delay in}\\ \text{seconds}\end{matrix}\right\} = \frac{\Delta\Phi}{\omega_m} \tag{78}$$

where $\Delta\Phi$ = amount side bands are shifted in phase with respect to the carrier frequency.

$\omega_m/2\pi$ = modulation frequency.

In this equation it is assumed that the upper and lower side bands suffer equal phase shifts of opposite polarity, *i.e.*, that the amplifier is adjusted so that its mid-frequency coincides with the carrier frequency.

The envelope delay will be constant only to the extent that the phase shift of the tuned amplifier is a linear function of frequency over the range of side-band frequencies involved. The phase-shift characteristics of typical tuned amplifiers are shown in Fig. 71*a*, together with the corresponding delay characteristics on the assumption of a modulated wave with the carrier centered at the middle of the pass band.

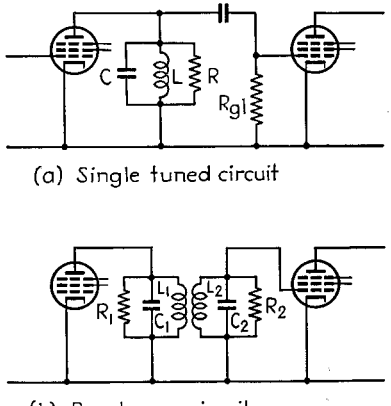


FIG. 72.—Typical circuits for tuned radio-frequency amplification of television signals.

20. Wide-band Tuned Amplifiers for Television Signals.—Tuned amplifiers for television signals differ from ordinary radio-frequency and intermediate-frequency amplifiers in that the carrier frequency is much higher and that the band width is a larger percentage of the carrier frequency. The requirements as to uniformity of amplification and linearity of phase shift within the pass band are also more severe. Intermediate-frequency amplifiers for video signals must likewise provide effective discrimination against the adjacent sound channels.

Tuned Radio-frequency Amplification of Television Signals.—The usual tuned radio-frequency wide-band amplifier consists of either a single resonant circuit or a band-pass arrangement as shown in Fig. 72, with most of the tuning capacity supplied by the tubes, and with resistance loading provided to give the wide band required. Although band-pass circuits are harder to tune to an incoming signal because two circuits must be adjusted, they are better than the single circuit from the point of view of gain, selectivity, uniformity of amplification in the pass band, phase characteristics, etc. Still more complicated coupling arrangements will give further improvement in performance, but are never used because of the added problems involved in tuning.

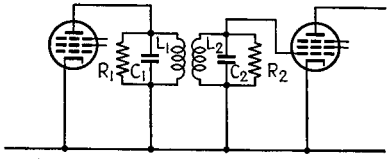
The behavior of tuned amplifiers employing the coupling methods shown in Fig. 72 is discussed in Par. 19. In applying these results, the effect of a resistance R shunted

across a tuned circuit is to make the effective Q of the circuit have the value

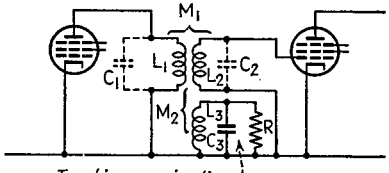
$$\left.\begin{matrix}\text{Effective } Q \text{ of} \\ \text{resonant circuit}\end{matrix}\right\} = Q_1 \frac{1}{1 + \frac{Q_1}{Q_2}} \tag{79}$$

where Q_2 is the actual Q of the resonant circuit without the resistance loading and $Q_1 = R/\omega L$.

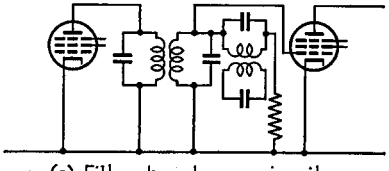
Intermediate-frequency Amplifiers.[1]—The converter of a television receiver ordinarily changes the incoming carrier to an intermediate frequency of the order of 12 mc. The band width that must be handled at this carrier frequency approximates 4.5 mc and so is an unusually large percentage of the carrier frequency. The intermediate-frequency amplifiers for television signals must also provide special means for eliminating the sound associated with the picture signal being handled and the sound of the adjacent television channel. It is also to be noted that inasmuch as ordinary television signals are of the single side-band type wherein one side band is largely suppressed, the carrier frequency is at the edge of the transmission band rather than in the middle.



(a) Simple band-pass



(b) Band-pass with tuned tertiary



(c) Filter band-pass circuit

FIG. 73.—Examples of coupling networks used in intermediate-frequency amplifiers for television signals.

A great variety of circuit arrangements can be employed to give uniform amplification over the required intermediate-frequency band, and at the same time provide reasonable selectivity. Examples in practical use are shown in Fig. 73, and involve at least two resonant circuits in each of which the tuning capacity is supplied largely either by the input or output capacity of an associated tube.

The circuit of Fig. 73*a* is the same as that of Fig. 72*b* and needs no further discussion. The addition of tuned tertiary L_3C_3 containing the circuit damping and coupled to the secondary, as shown in Fig. 73*c*, gives more gain, flatter response within the pass band, and better selectivity than Fig. 73*a*. In this three-circuit system mutual inductance M_1 largely dictates the band width, and M_2 and R determine the flatness of response. The coupling system of Fig. 73*c* is a band-pass analogue of the circuits of Fig. 58, and is one of a class of band-pass circuits that can be built up on filter theory. Such band-pass filters, although having excellent characteristics, are too expensive and complicated to be generally used in television receivers.

In addition to the circuits of Fig. 73, band-pass coupling can also be obtained by suitably detuning successive stages or by combining a double-humped resonance curve of one stage with a peaked curve of the following stage, as in Fig. 37, Sec. 3.

In practical intermediate-frequency amplifiers, it is necessary to provide special trap circuits for suppressing the accompanying sound that is present at the low-

[1] Further discussion, together with designs (including all circuit constants) for intermediate-frequency amplifiers with 4 and 2.3-mc band widths, is given by Garrard Mountjoy, Simplified Television I-F Systems, *R.C.A. Rev.*, Vol. 4, p. 299. January, 1940.

frequency edge of the pass band, and the sound of the adjacent television channel that is present at the other edge of the pass-band. In the case of a 12.75 mc intermediate-frequency carrier, these sound carriers have frequencies of 8.25 mc and 14.25 mc, respectively. Typical methods of providing such attenuation are illustrated in Fig. 74. At *a* suppression is accomplished with the aid of a low resistance resonant circuit *LC* tuned to the undesired frequency and coupled as shown, so that a large resistance will be introduced into the interstage network at the frequency to be suppressed. Circuit *b* employs a resonant trap in the grid lead of the amplifier tube, and circuit *c* makes use of a coupling network that is in series resonance at the undesired frequency and so gives no coupling for this frequency. The arrangement of Fig. 74*d* employs a resonant trap *LC* in the cathode circuit that reduces the amplification at the undesired frequency as a result of the large negative feedback at the resonant frequency of *LC*.

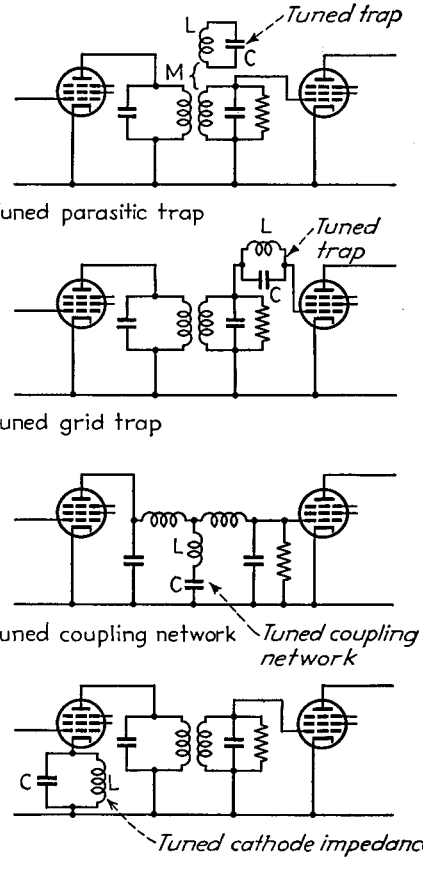


FIG. 74.—Trap circuits for discriminating against the sound signals on the edge of a television picture channel.

In practical intermediate-frequency systems, it is customary to provide about two traps for the accompanying sound, and one or two traps for the interfering sound of the adjacent channel. These can be distributed among the various intermediate-frequency coupling networks and tubes as is convenient.

Envelope Delay.—Envelope delay is of particular importance in tuned amplifiers handling television signals, because any variation in the envelope delay with modulating frequency adds to the total delay error of the system. The relation between the envelope delay and phase shift in the case of a modulated wave is given by Eq. (78). This equation also gives the envelope delay for a signal with suppressed side-band (single side-band television signal), provided that the phase shift is defined as the phase shift of the side-band frequency in question relative to the carrier.

In the case of either ordinary modulated waves or waves with one side band suppressed, the envelope-delay error produced by a single tuned circuit or a band-pass coupling network is much greater than for a stage of properly compensated video-frequency amplification, unless the band width of the tuned amplifier is appreciably greater than is necessary to accommodate the highest side-band frequency (see Fig. 71*a*). In transmitters the envelope-delay error can be kept small by using an extra

wide pass band, although this reduces the power-handling capacity of the tuned amplifier. There is also the possibility in transmitters of compensating for errors introduced by the tuned circuits by suitably proportioning the video-frequency stages. With receivers, the problem of minimizing the delay error in the tuned circuits is more difficult, particularly in the intermediate-frequency amplifier, because of the necessity of obtaining high discrimination against the sound channels on either side of the picture band. Fortunately, only several stages of tuned amplification are required in the intermediate-frequency system of a receiver. By designing the remainder of the television system with the smallest possible delay error, a moderate phase distortion can be tolerated in the intermediate-frequency amplifier without seriously detracting from the reproduced picture.

21. Class C Amplifiers.—A Class C amplifier is an amplifier so operated that the plate current flows in the form of pulses lasting for less than half a cycle. Such

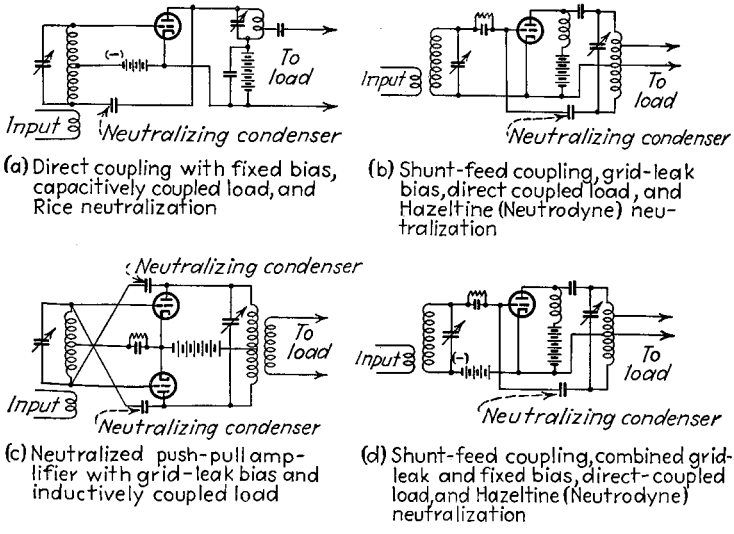


FIG. 75.—Typical circuit arrangements for Class C amplifiers.

amplifiers employ a load impedance consisting of a tuned circuit, and are characterized by relatively high plate efficiency. They are used primarily for the production of large quantities of power at a single frequency, as is necessary, for example, in radio transmitters.

Typical Class C amplifier circuits are shown in Fig. 75. These differ from each other in such respects as method of coupling the load to the tuned circuit, method of neutralization, and the way in which the grid bias is obtained.

Voltage, Current, and Power Relations in Class C Amplifiers Employing Triode Tubes.—The fundamental voltage and current relations existing in a properly adjusted Class C amplifier are shown in Fig. 76. The voltage acting on the grid electrode of the tube consists of a bias voltage E_c greater than the cutoff value, upon which is superimposed an alternating signal voltage E_s that is normally of sufficient amplitude to drive the grid positive at the peak of each cycle by an amount E_{max}, as indicated. The voltage acting on the plate electrode of the tube is the d-c plate supply voltage E_b minus the voltage drop E_L in the tuned load circuit. This drop is sinusoidal because of the resonant character of the load circuit, and with proper adjustment has a peak amplitude almost but not quite equal to the plate-supply voltage. The

minimum instantaneous plate potential E_{min} reached during the cycle is hence small but not zero. The phase relations are such that the minimum instantaneous plate potential E_{min} occurs at the same part of the cycle as the maximum grid potential E_{max}.

The plate and grid currents that flow in the tube at any instant are the result of the combined action of the plate and grid potentials at that instant, and can be obtained from these potentials with the aid of characteristic curves of the tube. These currents are in the form of pulses having angles of flow θ_p and θ_g, respectively, that are less than 180° (Figs. 76*e* and 76*f*). These angles of current flow can be expressed as follows:

$$\cos\left(\frac{\theta_g}{2}\right) = \frac{E_s - E_{max}}{E_s} = \frac{E_c}{E_s} \qquad (80)$$

$$\cos\left(\frac{\theta_p}{2}\right) = \frac{1}{1 + \dfrac{\mu E_{max} + E_{min}}{\mu E_c - E_b}} \qquad (81a)$$

An alternative form for Eq. (81*a*) is

$$\text{Grid bias} = E_c = \frac{E_b}{\mu} + \left(E_{max} + \frac{E_{min}}{\mu}\right) \frac{\cos\left(\dfrac{\theta_p}{2}\right)}{1 - \cos\left(\dfrac{\theta_p}{2}\right)} \qquad (81b)$$

In these equations, E_{max}, E_{min}, and E_b have the definitions given above, and μ is the amplification factor of the tube. The average of the pulses of current flowing to an electrode represents the direct current drawn by that electrode.

The power input to the plate electrode of the tube at any instant is the product of plate-supply voltage and instantaneous plate current, while the corresponding power lost at the plate is the product of instantaneous plate-cathode voltage and instantaneous plate current (see Fig. 76*g*). The difference between these two quantities represents the useful output being developed at the moment. The average input, output, and loss are obtained by averaging the instantaneous powers.

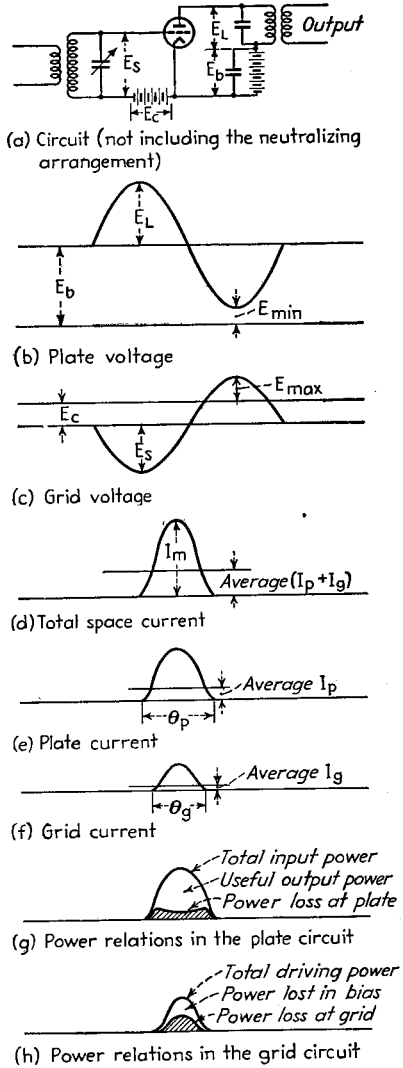


FIG. 76.—Voltage, current, and power relations in a typical Class C amplifier.

The efficiency is the ratio of average output to average input and is commonly of the order of 60 to 80 per cent. The efficiency is high in a Class C amplifier because current is permitted to flow only when most of the plate-supply voltage is used as voltage drop across the tuned load circuit, and only a small fraction is wasted as voltage drop at the plate electrode of the tube.

The instantaneous power absorbed from the exciting voltage applied to the grid electrode of the tube is equal to the product of instantaneous exciting voltage and instantaneous grid current (Fig. 76*h*). A portion of this instantaneous power is dissipated at the grid electrode of the tube, and the remainder is dissipated in the grid-leak resistance (or is used to charge the bias battery).

Calculation of Class C Amplifier Behavior with Triode Tubes—Exact Method.—The first step in the procedure is to calculate the instantaneous plate and grid potentials at various points of the cycle according to the relations

$$\text{Instantaneous plate voltage} = e_p = E_p - (E_b - E_{\min}) \cos \beta \qquad (82a)$$

$$\text{Instantaneous grid potential} = e_g = (E_c + E_{\max}) \cos \beta - E_c \qquad (82b)$$

where β is the number of electrical degrees from the crest of the cycle at which e_p and e_g are to be evaluated and where the remaining notation is as before. For each combination of instantaneous grid and plate potentials, the corresponding instantaneous grid and plate currents can be obtained from characteristic curves of the tube. In this way the plate- and grid-current pulses can be calculated and plotted point by point. Curves of instantaneous power input to the plate circuit of the tube and of power dissipated at the plate electrode can be obtained by multiplying the instantaneous plate current by the plate-supply voltage E_b and the instantaneous plate voltage e_p, respectively. Similarly, the instantaneous power supplied by the exciting voltage E_s is equal to the instantaneous value of exciting voltage $E_s \cos \beta$ times the grid current at that instant. The part of this power that is dissipated at the grid is the product of instantaneous grid current and the grid-cathode voltage e_g at that instant.

The d-c plate and grid currents represent the average of the current pulses over the full cycle, and the mean powers are obtained in the same way by an averaging process. These averages can be obtained by planimetering the areas under the pulses, by plotting the curves on cross-section paper and counting the number of squares under the pulses, or by averaging values calculated for regular steps in β from $\beta = 0$ to $\beta = 180°$.

The exact method of calculating Class C amplifier performance results in an accuracy limited only by the extent to which the tube characteristics used actually apply to the particular tube involved under the operating conditions. The method has the disadvantage of being very laborious, even when the point-by-point calculations are carried out in a systematic manner.[1] Accordingly, the exact method of Class C amplifier analysis is used only where the approximate method given below is not satisfactory, as, for example, in some types of modulated-amplifier calculations.

Calculation of Class C Amplifier Behavior with Triode Tubes—Approximate Method.[2] An approximate calculation of Class C amplifier performance with accuracy sufficient for ordinary design purposes can be obtained without point-by-point calculations by taking advantage of the fact that the total space current $(I_p + I_g)$ can be expressed rather accurately by the relation

$$\text{Total space current} = I_p + I_g = k \left(\frac{e_p}{\mu} + e_g \right)^{\alpha} \qquad (83)$$

[1] A method of organizing the procedure so that results can be obtained with the minimum of time is given by D. C. Prince, Vacuum Tubes as Power Oscillators, *Proc. I.R.E.*, Vol. 11, p. 275, June, p. 405, August, and p. 527, October, 1923.

[2] This follows F. E. Terman and W. C. Roake, Calculation and Design of Class C Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 620, April, 1936.

Other approximate methods of less accuracy but about equal complexity are given by W. G. Wagener, Simplified Methods for Computing Performance of Transmitting Tubes, *Proc. I.R.E.*, Vol. 25, p. 47, January, 1937; W. L. Everitt, Optimum Operating Conditions for Class C Amplifiers, *Proc. I.R.E.*, Vol. 22, p. 152, February, 1934.

Here e_p and e_g are the instantaneous plate and grid potentials, respectively, μ is the amplification factor of the tube (assumed constant), k is a constant, and α is another constant normally having a value very close to $\frac{3}{2}$. To the extent that Eq. (83) holds, the pulses of plate current in a Class C amplifier have a definite form in which the direct-current and fundamental-frequency components are functions only of the angle of plate current flow θ_p, the peak value I_m of the space current, and the exponent α. This relationship has been worked out and is presented in Fig. 77 for values of α between 1 and 2.

The direct-current and fundamental frequency components of the total space currents divide between the grid and plate electrodes. The portions going to the grid electrode can be calculated from the fact that with normal amplifier adjustment, the pulses of grid current usually have a shape that approximates the square of a

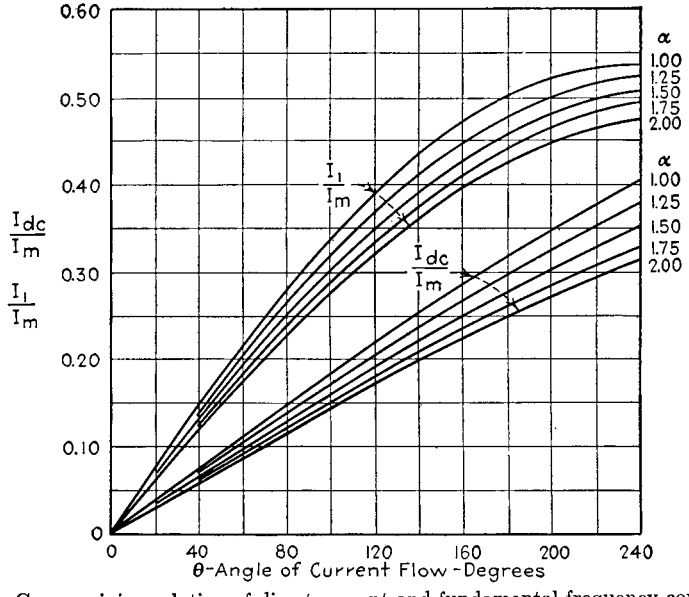


FIG. 77.—Curves giving relation of direct-current and fundamental-frequency components of the space-current pulse as a function of angle of flow θ and the peak amplitude I_m.

section of a sine wave.[1] Accordingly, by assuming that $\alpha = 2$, Fig. 77 can be used to determine the direct-current and fundamental-frequency components of the grid current when the angle of grid-current flow θ_g and the peak d-c grid current corresponding to instantaneous grid and plate electrode potentials E_{max} and E_{min} are known. The components flowing to the plate electrode are obtained by subtracting the parts going to the grid from the corresponding components of the total space current.

The power input to the Class C amplifier is now the product of battery voltage and d-c plate current, or

$$\text{Power input} = E_b I_{dc} \tag{84}$$

[1] This is the case under the usual conditions where secondary emission at the control grid is not excessive. See W. L. Everitt and Karl Spangenberg, Grid-current Flow as a Factor in the Design of Vacuum-tube Power Amplifiers, *Proc. I.R.E.*, Vol. 26, p. 612, May, 1938. With water-cooled tubes, it is sometimes found that the secondary emission may modify the shape of the grid current pulse greatly, and even make the d-c grid current slightly negative. Under these conditions, the grid current wave can be sketched and its d-c value estimated. The fundamental frequency component then has a crest value approximately twice the d-c value.

where I_{dc} is the d-c plate current obtained with the aid of Fig. 77. Likewise, the power delivered to the load is equal to half the product of a-c plate current I_1, obtained with the aid of Fig. 77, and the alternating-current voltage developed across the load, or

$$\text{Power output} = \frac{(E_b - E_{\min})I_1}{2} \tag{85}$$

The plate dissipation is the difference between these two powers, and the efficiency is their ratio.

The required load impedance is

$$\left.\begin{matrix}\text{Load impedance between} \\ \text{plate and cathode}\end{matrix}\right\} = \frac{E_b - E_{\min}}{I_1} \tag{86}$$

The driving power is half the product of the crest values of exciting voltage E_s and the fundamental-frequency component I_{g1} of grid current, or

$$\text{Grid driving power} = \frac{E_s I_{g1}}{2} \tag{87}$$

The portion of this power delivered to the bias is equal to the product $I_g E_c$ of bias voltage and d-c grid current. The difference between the driving power and the bias power is dissipated at the grid electrode of the tube.

The only approximations involved in this analysis of Class C amplifiers is the assumption that μ and the exponent α in Eq. (83) are constant, and that the grid-current pulse follows a square law. Under ordinary conditions, the results are accurate to within about 5 per cent.

An example showing the details involved in calculating Class C amplifier performance by the approximate method is given below.

Design Considerations.—In designing a Class C amplifier, the customary objectives are high efficiency, large power output, and low exciting power. Efficiency can be made high by using a small angle of plate-current flow θ_p and a minimum instantaneous plate potential $E_{\min}$ that is small compared with the plate-supply voltage. Low exciting power is obtained by making $E_{\max}$ small and also considerably less than the minimum plate potential $E_{\min}$, and by adjusting so that the angle of plate-current flow θ_p is large. A large output power is obtained by making θ_p large, combined with $E_{\max}$ and $E_{\min}$ great enough so that the maximum space current I_m is large.

The best compromise between these various conflicting factors involved in the design is obtained when the angle of plate-current flow θ_p is of the order of 120 to 150 electrical degrees. The exciting voltage should drive the grid sufficiently positive so that the plate-current pulse will give an average plate current equal to the rated value. The load impedance should then be adjusted so that the voltage drop in the load with this plate-current pulse gives a minimum plate potential moderately greater than the maximum grid potential.

The maximum allowable peak amplitude I_m of the pulse of space current is the maximum electron emission that the filament can be depended upon to produce throughout its useful life. With tungsten filaments, it is common practice to make I_m very nearly the full emission from the filament, while in the case of thoriated-tungsten and oxide-coated filaments, the deterioration during life is such that I_m must be made much less than the emission initially obtainable from the filament. In the case of thoriated-tungsten filaments, a safety factor of 3 to 10 is typical, while with oxide-coated filaments a still larger margin of safety, such as 5 to 10, must be provided. When no information is available that gives directly the permissible peak space current, this can be taken as being four times the sum of the d-c plate and

grid currents specified by the manufacturer as being permissible in normal Class C amplifier operation. Alternatively, one may assume that the allowable peak milliamperes per watt of heating power is 2 to 4 for tungsten (or 0.65 this value for linear and modulated amplifiers) and 10 to 40 for thoriated and oxide-coated filaments.

Design Procedure for Class C Amplifiers Employing Triode Tubes.—The design of a Class C amplifier can be systematically worked out on paper according to the following steps:

First, select the peak space current I_m on the basis of the electron emission of which the filament is capable, as discussed above.

Second, select a suitable combination of maximum grid potential E_{max} and minimum plate potential E_{min} that will draw this total space current. The minimum plate voltage must not be less than the maximum grid potential, and, if low driving power is important, the minimum plate potential should be appreciably larger than the maximum grid potential, although still relatively small compared with the plate-supply voltage.

Third, decide upon a suitable angle of plate-current flow θ_p, making a reasonable compromise between the high efficiency, small output, and large driving power obtained with small angles of flow, and the large output, small driving power, and low efficiency with large angles. Under most circumstances, the angle of flow will lie between 120 and 150°.

Fourth, calculate the grid bias by the use of Eq. (81*b*). This also determines the signal voltage required since the crest signal voltage is $E_c + E_{max}$.

Fifth, determine the d-c plate current, d-c grid current, plate dissipation, power output, efficiency, grid driving power, etc., either by the exact or by the approximate methods.

Sixth, examine the results obtained to see if they are satisfactory. If not, revise the design, and recalculate the performance.

Seventh, design the tank circuit in accordance with the principles discussed below.

The design procedure and the details involved in calculating the performance by the approximate method can be understood with the aid of the following example:

Example.—A Class C amplifier is to be designed employing a Type 800 tube operating at 1,000 volts plate potential.

The peak emission I_m will be taken as 407 ma, which is arrived at by assuming an electron emission of 100 ma per watt of filament power and a factor of safety of 6. A suitable combination of E_{max} and E_{min} for drawing this current without excessive grid current is $E_{min} = 150$, $E_{max} = 120$. The corresponding peak grid current is 85 ma. The next step is the selection of a suitable angle of flow, which will be taken as 140° as a reasonable compromise between efficiency and output. On the assumption that α equals $\frac{3}{2}$, Fig. 77 gives the direct-current and fundamental-frequency components of the total space current as 0.22 and 0.39, respectively, of the peak space current I_m. The direct-current component of the total space current is hence $407 \times 0.22 = 89.5$ ma, and the crest fundamental-frequency component of the total space current is $407 \times 0.39 = 159$ ma, crest value. It is now necessary to make allowance for the part of the total space current diverted to the grid. With the use of Eq. (80) θ_g is found to be 116°; so from Fig. 77, assuming $\alpha = 2$, the d-c and fundamental frequency components of the grid current are 0.17 and 0.31 times the peak value, or 14.4 ma and 26.3 ma, respectively. The d-c plate current is then $89.5 - 14.4 = 75.1$ ma, and the fundamental-frequency component is $159 - 26.3 = 132.7$ ma crest value. The power input to the plate is found from Eq. (84) to be

$$1{,}000 \times 0.0751 = 75.1 \text{ watts.}$$

The power delivered to the load is obtained from Eq. (85), and is

$$(1{,}000 - 150) \times 0.1327/2 = 56.4 \text{ watts.}$$

The plate loss is $75.1 - 56.4 = 18.7$ watts, and the plate efficiency $56.4 \times 100/75.1 = 75.1$ per cent. The grid bias calculated by Eq. (81*b*) is 134 volts, and can be developed by a grid-leak resistance of $134/0.0144 = 9{,}300$ ohms. The peak exciting voltage is $134 + 120 = 254$ volts, and Eq. (87) shows the exciting power to be $254 \times 0.0263/2 = 3.3$ watts, of which $134 \times 0.0144 = 1.9$ watts is dissipated in the bias, and $3.3 - 1.9 = 1.4$ watts at the grid of the tube. The load impedance is given by Eq. (86) as $(1{,}000 - 150)/0.1327 = 6{,}400$ ohms.

Tank-circuit Design.—The tuned circuit connected between the cathode and plate of the Class C amplifier, commonly called the *tank circuit,* must supply the proper

impedance and must not consume an undue proportion of the power output of the amplifier.

The efficiency of the tank circuit is the fraction of the total power delivered to this circuit by the tube that is transferred to the load, and is

$$\left.\begin{matrix}\text{Tank-circuit} \\ \text{efficiency}\end{matrix}\right\} = \frac{Q_1 - Q_2}{Q_1} 100 \text{ per cent} \tag{88}$$

where Q_1 is the Q of the tank resonant circuit with no load coupled, and Q_2 is the effective Q of the tank circuit when the load is coupled to it. From the point of view of efficiency, it is desirable that the effective Q of the tank circuit, when the load is taken into account, be as low as possible, with the actual Q of the circuit in the absence of load as high as is practicable. A low effective Q tank circuit also reduces the reactive energy that must be stored, and so permits the use of inductances and capacities with smaller volt-ampere ratings, an important item in large amplifiers. At the same time, a low tank-circuit Q results in an appreciable impedance being offered to harmonic components of the plate-current pulse.[1]

The effective Q of the tank circuit is usually made at least 8 or 10, although in the case of some very large push-pull Class C amplifiers, values of effective Q as low as 2 or 3 are employed. With such very low values of Q it is necessary to use a tuning procedure such that the tank circuit is adjusted to offer a resistance impedance to the tube, even though the magnitude of the resulting impedance may not necessarily be maximum for this condition (see Par. 2, Sec. 3).

When the desired Q of the tank circuit has been decided upon, the inductive reactance ωL can then be determined in terms of the required load impedance. In the case of direct coupling with a tank circuit Q not too low, one has

$$\omega L = \frac{\text{required load impedance}}{\text{effective } Q} \tag{89}$$

When ωL is known, the tank-circuit inductance and capacity for a given frequency are then calculated by obvious methods.

A useful relation in the design of tank circuits is the connection that exists between the voltage across the tank circuit, the effective Q, and the power delivered to the circuit. This relation is

$$\left.\begin{matrix}\text{Effective voltage} \\ \text{across tank circuit}\end{matrix}\right\} = \sqrt{P\omega L Q_{\text{eff}}} \tag{90}$$

where P = power, watts, delivered to tank circuit.
ωL = reactance of inductive branch of tank circuit.
Q_{eff} = effective Q of tank circuit, when the effect of the coupled load resistance is taken into account.

Equation (90) applies to all types of tank circuits, including arrangements such as in Figs. 75*b*, 75*c*, and 75*d*, where the plate-cathode connection is made across only a portion of the tank circuit.

[1] The relationship between the harmonic and fundamental frequency voltages developed across the tank circuit and the effective Q for the case where the harmonic voltage is small is approximately

$$\left.\begin{matrix}\text{Per cent second} \\ \text{harmonic}\end{matrix}\right\} = \frac{67}{Q} \frac{\left(\frac{I_2}{I_m}\right)}{\frac{I_1}{I_m}} \tag{88a}$$

where I_1/I_m and I_2/I_m are coefficients obtained from Fig. 86 for the angle θ_p of plate-current flow involved. See E. H. Schulz, R-f Power Amplifier Chart, *Electronics*, Vol. 12, p. 33, December, 1939.

Practical Adjustment of Class C Amplifiers.—The adjustment of Class C amplifiers to realize the design conditions is usually carried out by a cut-and-try process, using the d-c grid current, the d-c plate current, and the output power as guides.

In carrying out the adjustment, one ordinarily starts either with a complete paper design giving the conditions desired, or with data for Class C amplifier operation given by the tube manufacturer. In both cases one knows the proper grid bias, the desired d-c plate current, and the expected power output, and in most instances will also have available the necessary exciting voltage and the approximate d-c grid current. The procedure will then depend upon whether fixed or grid-leak bias is employed.

In the former case, after the bias has been given the proper value, the tank circuit is tuned to resonance with the exciting voltage. The exciting voltage and effective load impedance (*i.e.*, coupling of the load to the tank circuit) are then varied by trial until the expected d-c plate current and output power are obtained without excessive grid current. This procedure can be simplified by keeping in mind that the total space current depends primarily on the maximum grid potential E_{max} and hence upon the excitation, while with a given excitation the minimum plate potential depends upon the load impedance. The ratio of grid current to plate current is accordingly less the higher the impedance in the plate circuit of the tube. The adjustment

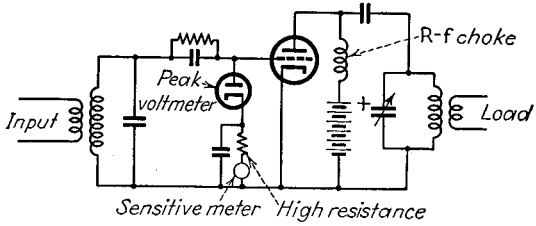


FIG. 78.—Peak voltmeter arrangement for measuring maximum positive potential reached by the grid.

procedure hence involves coupling the load to the tank circuit a reasonable amount, after which the excitation is varied until the total space current approximates the expected value. The effective load impedance is then varied by changing the coupling of the load to the tank circuit until a point is found where decreasing the coupling to the load slightly (*i.e.*, increasing the load impedance) causes the grid current to become excessive. The power output, power input, and grid current are then noted, and the entire procedure is repeated over and over again until the desired operating conditions are obtained. In carrying out the adjustments, it is necessary to pay attention to the plate dissipation (*i.e.*, difference between power input and power output), since there is always danger of damaging the tube by accidentally overheating the electrodes.

When the bias is obtained from a grid-leak resistance, the best procedure is to start with the grid-leak resistance that is required to produce the desired grid bias with the estimated grid current, and then to follow out the procedure outlined above to obtain the best operating point. The actual grid bias that results is then calculated from a knowledge of the grid-leak resistance and the grid current; if it is not the desired value, the grid-leak resistance is altered as necessary, and the process is repeated until the grid bias, space current, power output, and power input approximate the calculated values and the grid current is not excessive

Much of the cut-and-try involved in the foregoing procedure can be eliminated by the use of a peak voltmeter to measure directly the amount the grid is driven positive. Such a voltmeter is described in Par. 15, Sec. 13, and its use is illustrated in Fig. 78.

In making adjustments on a Class C amplifier, it is desirable to check the tuning repeatedly, particularly after each change in the load coupling. With triodes this is done by adjusting the tuning condenser for minimum plate current. This condition normally corresponds to maximum load impedance (*i.e.*, to the lowest possible $E_{\min}$), maximum grid current, and a unity-power-factor load. It also gives the maximum possible power output unless the effective Q of the tank circuit is so low that the points of unity power factor and maximum impedance do not coincide. In this case, the tank circuit is tuned for maximum power into the load.

Neutralization can be accomplished by turning off the plate voltage of the tube (but leaving the tube in place) and applying the exciting voltage. The neutralizing condenser is then adjusted until turning the tank circuit through resonance has no effect on the d-c plate and grid currents of the exciting stage, or until tuning the tank circuit has no reaction on the rectified grid current.

Class C Amplifiers Employing Screen-grid, Beam, and Pentode Tubes.—Screen-grid, beam, and pentode tubes can be operated as Class C amplifiers by making the grid bias greater than the cutoff value corresponding to the screen-grid potential. The performance obtained is then similar to that of triode Class C amplifiers, but with the advantage that no neutralization is required. Also, pentode and beam tubes have the added advantage of requiring smaller driving power than triode tubes. The disadvantage of pentode, beam, and screen-grid tubes is that they are, in general, more expensive than triodes, and are not available in water-cooled types.

The analysis, calculation of performance, and design are the same with screen-grid, beam, and pentode tubes as with triodes except for minor modifications introduced by the screen grid.[1] In particular, the relation between the grid bias and angle of flow is

$$\text{Grid bias} = E_c = \frac{\frac{E_{sg}}{\mu_{sg}} + E_{\max}\cos\left(\frac{\theta}{2}\right)}{1 - \cos\left(\frac{\theta}{2}\right)} \tag{91}$$

where E_{sg} is the screen-grid potential and μ_{sg} is as defined by Eq. (48), Sec. 3. In selecting operating conditions, it is essential that the maximum grid potential $E_{\max}$ should not exceed the screen-grid potential.

The minimum plate potential $E_{\min}$ in the case of beam and pentode tubes must be great enough that a virtual cathode does not form in the tube when the space current has its peak value. This minimum allowable value in practical tubes is of the order of 15 to 20 per cent of the plate-supply voltage. In the case of screen-grid tubes, the minimum plate potential also must not be less than the screen-grid potential, since otherwise the plate will lose secondary electrons to the screen. Otherwise the situation is the same as with triodes, and the analysis can be carried out by either point-by-point calculations or the approximate method based upon the curves of Fig. 76.

In adjusting screen-grid and pentode Class C amplifiers, it is sometimes found that the plate potential has so little effect on the plate current that it is impossible to tune the tank circuit by adjusting for minimum plate current. Under such circumstances, the tuning adjustment can be made to give maximum current in the load.

22. Linear Amplifiers.—A linear amplifier is a tuned amplifier operated with a bias approximating the cutoff value, so that the amplified output voltage that is developed is proportional to the exciting voltage applied to the amplifier. The linear amplifier is a form of Class B amplifier, and its most important use is as a power amplifier of modulated waves.

[1] F. E. Terman and J. H. Ferns, The Calculation of Class C Amplifier and Harmonic Generator Performance of Screen-grid and Similar Tubes, *Proc. I.R.E.*, Vol. 22, p. 359, March, 1934.

Distortion in Linear Amplifiers.—The most important characteristic of a linear amplifier is the relationship between output voltage and exciting voltage, since this shows the extent to which the amplifier is actually linear. A typical characteristic of this type is shown in Fig. 79 and is characterized by an approximately linear characteristic up to a critical exciting voltage, after which the output levels off or *saturates*. The saturation takes place when the alternating voltage developed in the plate circuit across the load impedance has a crest value approaching the plate-supply voltage, and results from the fact that this alternating voltage cannot be greater than the plate-supply voltage, irrespective of how large the applied signal is.

The linearity of the amplifier characteristic below saturation is greatest when the amplifier tube is biased to projected cutoff exactly, as in the case of the Class B audio amplifier (see Par. 10).

Distortion may arise in a system involving a linear amplifier as a result of the variable load that the linear amplifier presents to its exciter. The remedy for this is to employ an exciter of relatively large power capacity, or to use negative feedback or some other system of distortion correction.

The distortion in the modulation envelope that results when a modulated wave is amplified by a linear amplifier can be readily calculated when the linearity curve is known. The procedure is to determine the output voltage at certain points on the modulation cycle and then to use Eq. (24) or (25) to determine the harmonic components. These critical points are the outputs corresponding to carrier excitation, peak of the modulation cycle, trough of the modulation cycle, and when the modulation is 70 per cent of its peak and trough values, as indicated in Fig. 80.

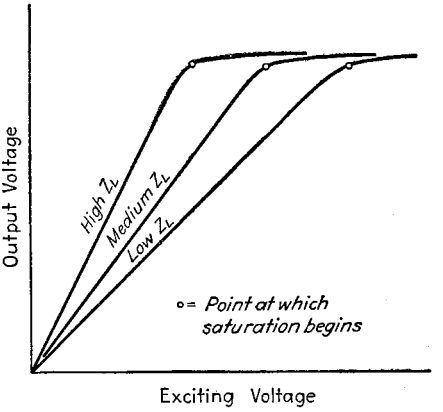


FIG. 79.—Typical linearity curves of linear amplifier with different load impedances in the plate circuit.

Design and Calculation of Linear Amplifiers.—In the design of linear amplifiers, the first step is to determine the bias voltage corresponding to projected cutoff. The bias must be constant, irrespective of the degree of modulation and of the grid current, and so must be obtained from a fixed source or by self-bias. Grid-leak arrangements are not permissible under any circumstances.

A load impedance is then selected such that saturation is just beginning to occur when there is sufficient excitation to develop full load output corresponding to the modulation peaks; *i.e.*, the load impedance is chosen so that with the rated peak power the crest voltage across the load is slightly less than the plate-supply voltage. The type of relationship existing between load impedance, exciting voltage, and output power is shown in Fig. 79. The power output corresponding to the beginning of saturation is inversely proportional to the load impedance, whereas the exciting voltage required to produce saturation is greater as the load impedance is reduced. The proper load impedance in any given case can be obtained either by experiment or by approximate calculations. In the latter case, the method of calculation is exactly that used for Class C amplifiers.

After the load impedance and exciting voltage corresponding to the peak output have been determined, the linearity is obtained by calculating the amplifier output with lesser exciting voltages. This must be done by a cut-and-try process, since the

methods that have been developed for calculating amplifier performance, both the exact and approximate methods, give the load impedance as the answer for an assumed value of alternating voltage between plate and cathode, but cannot be rearranged to do the converse.

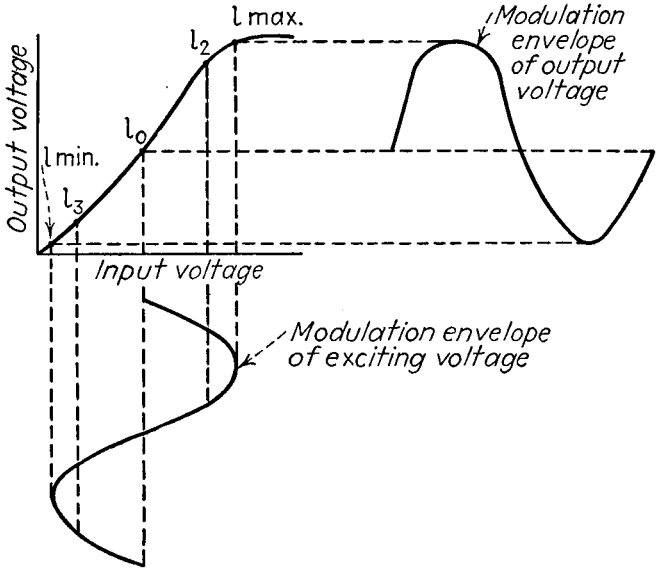


FIG. 80.—Linearity curve of linear amplifier, showing critical points on modulation cycle used in calculating distortion.

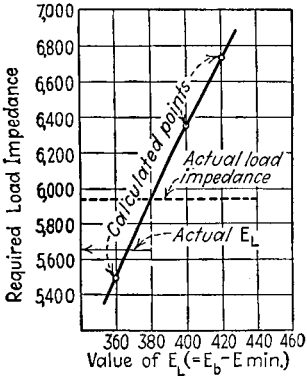


FIG. 81.—Determination of the actual output voltage for a known load impedance by plotting a curve of load impedance required for several assumed values of output voltage.

The procedure for calculating a linearity curve after already having determined the conditions for the peak exciting voltage is to estimate the output voltage expected for a given exciting voltage on the basis that the output voltage is approximately proportional to the exciting voltage. One then calculates the load impedance corresponding to this assumed output voltage, using either the exact or approximate methods of calculation, according to the accuracy desired. If the resulting load impedance determined in this way is greater than the actual load impedance, then one revises the estimate of output voltage to a smaller value (or, conversely, if the calculated load impedance is too small). By repeating this process two or three times, it is possible to draw a curve such as that shown in Fig. 81 and by interpolation to obtain the output voltage that would be obtained for the assumed exciting voltage with the load impedance actually present.

In determining linearity curves, the approximate method of calculation outlined for Class C amplifiers will give the general shape of the characteristic and detect excessive nonlinearity. However, if distortion is to be exactly evaluated, it is necessary to employ the exact method of calculation.

Practical Adjustment of Linear Amplifiers.—The procedure to be followed in adjusting linear amplifiers depends upon the circumstances and the measuring equipment at hand. When ample exciting voltage is available, the combination of load impedance and exciting voltage that gives full peak output with satisfactory linearity can be readily determined experimentally, as can the bias voltage giving best linearity. When the exciting voltage available is barely sufficient to supply the desired carrier, it is then necessary to modulate this carrier, and make use of a cathode-ray oscillograph or a modulation meter to determine when satisfactory linearity is obtained.

When a linear amplifier is excited by a modulated wave, the d-c component of the plate current should not vary with the degree of modulation. Any change in the d-c current with modulation indicates either distortion or a carrier shift (*i.e.*, change in carrier amplitude), although the absence of variation does not guarantee that there is no distortion.

The fact that the d-c current of a linear amplifier should be constant means that the grid bias may be obtained by a cathode resistor.

Efficiency and Power.—The efficiency of an ideal linear amplifier having a linear characteristic and operating so that the plate current flows exactly 180° during each cycle is

$$\text{Efficiency} = \frac{\pi}{4}\left(1 - \frac{E_{\min}}{E_b}\right) 100 \text{ per cent} \tag{92}$$

where E_b is the plate-supply voltage, and $E_{\min}$ is the minimum instantaneous plate voltage during the cycle. The plate efficiency at full output is usually of the order of 50 to 65 per cent under practical conditions. With less than the full output, the efficiency is proportional to the exciting voltage.

When the signal to be amplified is a carrier wave modulated 100 per cent, the carrier amplitude is half the peak amplitude to be handled. The efficiency for the unmodulated carrier is then half of the maximum efficiency, or about 30 per cent under ordinary conditions.

The peak power that can be developed by a tube operating as a linear amplifier is approximately the same as the power that can be developed by the same tube in Class C operation. In the case of completely modulated waves, the peak power is four times the carrier power, so that the carrier power that a linear amplifier is capable of developing is approximately one-fourth of the output obtainable from the same tube operated as a Class C amplifier.

High-efficiency Linear Amplifier (Doherty Amplifier).[1]—The average plate efficiency obtained when an ordinary linear amplifier is handling a modulated wave can be increased by a system devised by Doherty and illustrated schematically in Fig. 82*a*. Here the amplifier is divided into two parts: A_1, which delivers power to the load through a quarter-wave length line, and A_2, which delivers power to the load directly. In order that the outputs of the two amplifiers will add directly at the load, a 90° phase shift, normally obtained with the aid of an artificial quarter-wave line, is introduced in the exciting voltage of one of the amplifiers to compensate for the 90° phase shift of the quarter-wave line connecting the two amplifiers.

Tube A_1 is biased to projected cutoff, as in linear amplifier operation. This tube is operated so that when the exciting voltage is the unmodulated carrier, the tube is just beginning to saturate with a load impedance that is twice the impedance that would be used to develop full output.

Tube A_2 is operated with an exciting voltage and bias adjustment such that at the peak of modulation, *i.e.*, when the exciting voltage is twice the carrier, this tube

[1] W. H. Doherty, A New High Efficiency Power Amplifier for Modulated Waves, *Proc. I.R.E.*, Vol. 24, p. 1163, September, 1936.

develops an output equal to twice the carrier power when operating as an ordinary Class C amplifier, but has its output just reduced to zero when the exciting voltage is the unmodulated carrier. The tank circuit is proportioned so that its impedance is half the impedance that would be used with amplifier A_1 with this tube functioning as a conventional linear amplifier developing its full output. The characteristic impedance of the line is made twice the tank-circuit impedance.

The operation of the high-efficiency system can be described briefly as follows: When the amplitude of the exciting voltage does not exceed the carrier level, A_2 is inoperative, and A_1 functions as a conventional linear amplifier with a load impedance twice the impedance that would be used to develop full output.[1] Tube A_1 therefore gives linear amplification of the exciting voltage during the portions of the modulation cycle where the envelope of the exciting voltage does not exceed carrier level. When the exciting level is greater than carrier level, A_2 also passes current. The voltage developed across the tank circuit is then greater than previously in proportion to the

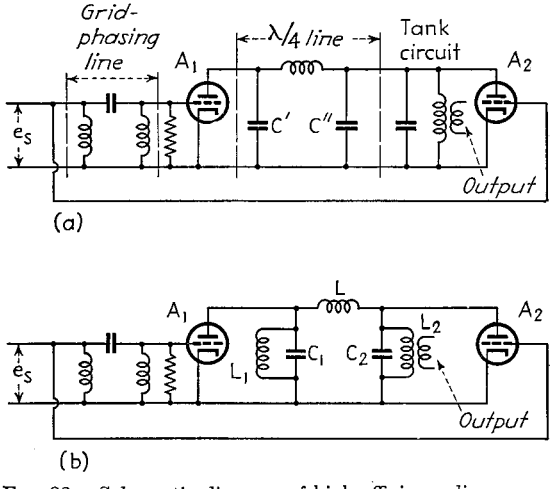


FIG. 82.—Schematic diagram of high-efficiency linear amplifier.

current delivered from A_1 over the quarter-wave line. This results in an increase in the apparent load impedance that the tank circuit offers to the transmission line. Because of the impedance inverting properties of the quarter-wave line, the impedance across the sending end of the line is reduced, thereby causing A_1 to deliver more power to the line in spite of the fact that this tube was already operating under saturation conditions. At the peak of a completely modulated wave, tube A_2 is carrying half of the power, and the apparent load impedance that the tank circuit offers to the line is increased to equality with the characteristic impedance. Tube A_1 then operates with half the impedance in its plate circuit present at carrier level, and so carries the remaining half of the peak power. Curves showing behavior as the exciting voltage varies are given in Fig. 83.

[1] The impedance offered to A_1 by the sending end of a quarter wave line is related to the load impedance by the equation

$$Z_s = \frac{Z_0^2}{Z_r} \tag{93}$$

where Z_s = sending-end impedance offered by the line.
Z_r = impedance of the load connected to the line.
Z_0 = characteristic impedance of the line.
In the case where H_2 is inactive, $Z_r = Z_0/2$, so that $Z_s = 4Z_r$.

The average efficiency of the system is high for all degrees of modulation. When the exciting voltage is an unmodulated carrier, tube A_1 functions as a linear amplifier under saturation conditions corresponding to an efficiency of about 60 per cent. At the positive peaks of modulation, A_1 is still functioning as a linear amplifier under saturation conditions, and A_2 operates as a Class C amplifier with a plate efficiency of the order of 75 per cent. The result is an average efficiency for the modulation cycle that varies only slightly with the degree of modulation, and approximates the efficiency of a linear amplifier operating under peak conditions, *i.e.*, about 60 per cent.

A practical circuit for a high-efficiency amplifier is shown in simplified form in Fig. 82*b*. Here the capacity C'' of the quarter-wave line is combined with the tank circuit, and the capacity reactance of C' is obtained by resonant circuit L_1C_1 detuned a proper amount.

In setting up and tuning the circuit of Fig. 82, the first step is to determine the plate load impedance Z_L required to develop full output from A_1 with this tube

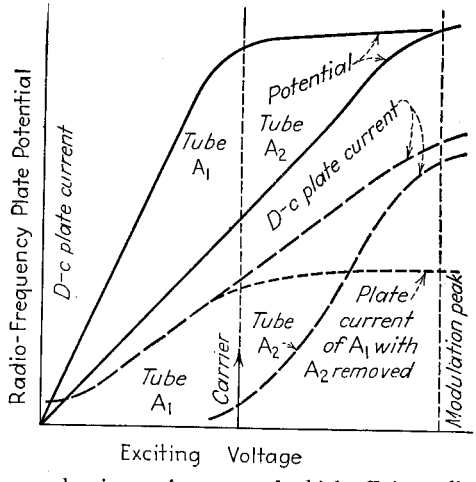


FIG. 83.—Curves showing performance of a high-efficiency linear amplifier.

functioning as a conventional linear amplifier. The characteristic impedance Z_0 of the line must then be equal to Z_L, and L should be such that $\omega L = Z_0 = Z_L$. Resonant circuit L_1C_1 is then tuned by short-circuiting L_2, applying exciting voltage to A_1, and tuning L_1C_1 for minimum plate current. Similarly, the resonant circuit L_2C_2 is tuned by short-circuiting L_1, applying voltage to A_2, and tuning L_2C_2 for minimum plate current. This ensures correct tuning and also ensures that the line is exactly one-quarter wave length long. The characteristic impedance of the line can be checked by the fact that when exciting voltage is applied to A_1 and A_2 is inoperative, the voltages across C_1 and C_2 should be in the ratio of 2:1 when the load impedance across L_2 is $Z_L/2$. An alternative way of checking the characteristic impedance of the line is to make A_2 inoperative, and observe the plate current in A_1 with the appropriate resistance $Z_L/2$ shunted across C_2. The load resistance is then transferred to shunt with C_1 and made $2Z_L$. If the characteristic impedance of the line is the proper value, the plate current of A_1 will not change.

After the circuit adjustments have been made in this way, it remains only to obtain the proper exciting and bias voltages. If these are to be obtained experimentally, a systematic procedure must be followed because of the numerous variables

present. The first step is to make A_2 inoperative and determine the carrier voltage required to cause A_1 to be on the edge of saturation. The bias and excitation on A_2 are then adjusted so that when the carrier is unmodulated, A_2 just shows signs of passing plate current, while with an exciting voltage equal to twice carrier level, A_2 is operating as a normal Class C amplifier delivering half of the total power output and having a d-c plate current approximately 80 per cent of the d-c plate current of tube A_1.[1] It is to be noted that the carrier voltages applied to A_1 and A_2 will normally differ. It is also to be noted that at the peak of modulation it is not necessary for the grid of A_1 to be driven much farther positive than at carrier level. The grid of A_1 can be prevented from going excessively positive at the modulation peaks and thereby drawing an undesirably large grid current by designing the quarter-wave line in the grid circuit to have a high characteristic impedance, and also by using a grid leak with A_1 so that the bias will increase when grid current flows.

The linearity of a high-efficiency amplifier is generally not all that could be desired. Accordingly, when the distortion must be low, as is the case in broadcast transmitters, it is necessary to provide some means of distortion correction, as discussed below.

Distortion Correction by Feedback and Remodulation Systems.—Distortion produced in radio transmitters by linear amplifiers, as well as distortion in the modulator

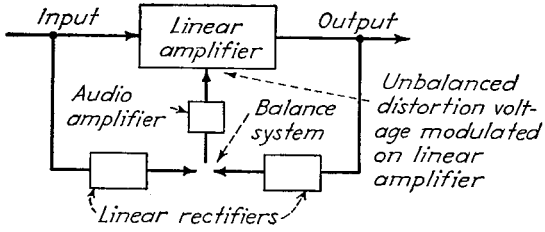


FIG. 84.—Remodulation system for reducing distortion in a linear amplifier.

and audio-frequency system, can be greatly reduced by rectifying a portion of the transmitter output, and properly feeding the resulting audio-frequency developed by the rectifier back into the audio-frequency system. This is a form of negative feedback, and is discussed in greater detail in Par. 1, Sec. 9.

The distortion produced in a linear amplifier can also be corrected by the remodulation system shown schematically in Fig. 84.[2] Here samples of the modulated waves existing in the input and output of the linear amplifier are rectified and the resulting audio-frequency voltages balanced against each other. The adjustment of the balancing device is such that when no distortion is produced in the linear amplifier the unbalanced output will be zero. Distortion, however, produces an output that is amplified and then modulated on the linear amplifier, preferably by grid modulation, in such a manner as to correct for the distortion produced in the amplifier.

Both the preceding arrangements are capable of giving substantially linear over-all performance even when there is considerable nonlinearity in the individual stages involved. These arrangements also greatly reduce hum modulation of the output wave resulting from alternating filament current and from alternating-current hum in the plate supply.

23. Harmonic Generators.—The ordinary vacuum-tube harmonic generator is essentially a Class C amplifier in which the tank circuit is tuned to an harmonic

[1] The difference in plate currents of A_1 and A_2 arises from the fact that A_1 operates as a Class B amplifier, while A_2 operates as a Class C amplifier and so has higher efficiency.

[2] This arrangement is due to F. E. Terman and R. R. Buss, Some Notes on Linear and Grid-modulated Radio-frequency Amplifiers, *Proc I.R.E.*, Vol. 29, p. 104, March, 1941.

component of the pulses of plate current instead of to the fundamental-frequency component.[1] Harmonic generators of this character are used in radio transmitters for developing appreciable amounts of power on harmonic frequencies.

Oscillograms showing voltage, current, and power relations existing in a typical harmonic generator are given in Fig. 85. These are very similar to the corresponding oscillograms for the Class C amplifiers shown in Fig. 76. The significant factors determining the performance of harmonic generators are the same as for the Class C amplifier, namely, the maximum grid potential E_{max}, the minimum plate voltage E_{min}, peak space current I_m, and the angle of plate current flow θ_p. The considerations involved in the design of harmonic generators are also the same as in the case of the Class C amplifier, except that since the harmonic content of the plate current pulse is fairly sensitive to the angle θ_p of plate-current flow, this angle must be carefully chosen. The experimental procedure followed in realizing desired operating conditions for harmonic generators is much the same as in the case of Class C amplifiers.

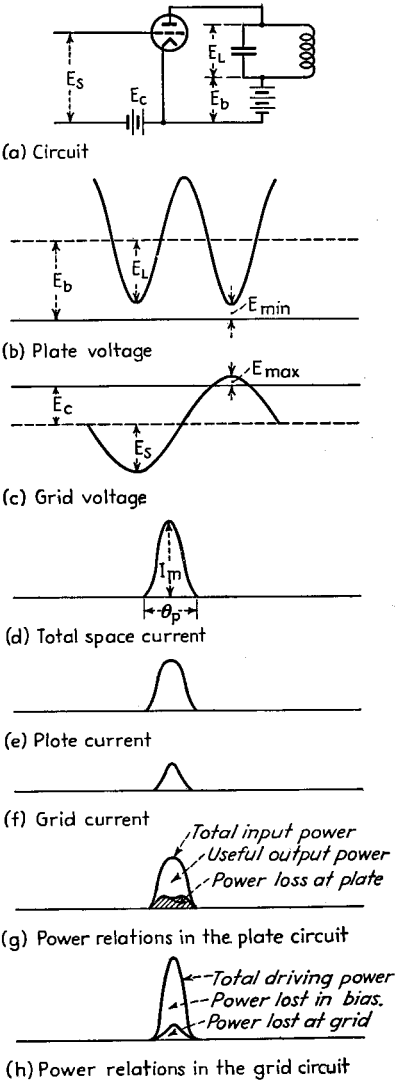


FIG. 85.—Voltage, current, and power relations in a typical Class C harmonic generator.

Factors Governing Harmonic-generator Performance.—The most desirable angle θ_p of plate-current flow is a compromise between conflicting factors. The shorter the length of current pulses in the case of a particular harmonic the higher will be the plate efficiency, but the bias, exciting voltage, and driving power are increased. Also, if the pulse is too long or too short, the output power drops off appreciably. Desirable values for θ_p under practical conditions are given in Table 4.

The harmonic power output that is obtainable decreases with the order of the harmonic. The relative harmonic output obtainable from a given tube compared with Class C output with the same peak space current is given approximately by Table 4. The load impedance is inversely proportional to the output power [Eq. (90)] and so increases with the order of the harmonic, as shown by Table 4. The exciting power required by a harmonic generator is greater than with corresponding Class C

[1] Other types of harmonic generators using vacuum tubes are described by H. J. Scott and L. J. Black, Harmonic Generation, *Proc. I.R.E.*, Vol. 26, p. 449, April, 1938; F. E. Terman, D. E. Chambers, and E. H. Fisher, Harmonic Generation by Means of Grid Circuit Distortion, *Trans. A.I.E.E.*, Vol. 50, p. 811, June, 1931.

operation because of the smaller angle of current flow, and increases rapidly with the order of the harmonic.

Design and Analysis of Harmonic-generator Performance.—The maximum instantaneous grid voltage E_{max} and the peak space current I_m have the same significance and are determined by the same considerations in harmonic generators as in Class C amplifiers. The angle of current flow θ_p is selected in accordance with Table 4, and the bias required to produce a given angle of flow is

$$\text{Grid bias} = E_c = \frac{E_b\left[1 - \cos\left(\frac{n\theta}{2}\right)\right] + E_{min}\cos\left(\frac{n\theta}{2}\right)}{\mu\left[1 - \cos\left(\frac{\theta}{2}\right)\right]} + \frac{E_{max}\cos\left(\frac{\theta}{2}\right)}{1 - \cos\left(\frac{\theta}{2}\right)} \quad (94)$$

TABLE 4.—PLATE-CURRENT PULSE LENGTH AND POWER OUTPUT OF HARMONIC GENERATORS

Harmonic	Optimum length of pulse, electrical degrees at the fundamental frequency	Approximate power output, assuming that normal Class C output is 1.0	Relative load impedance, assuming that Class C case is 1.0
2	90–120	0.65	1.5
3	80–120	0.40	2.5
4	70– 90	0.30	3.3
5	60– 72	0.25	4.0

The notation is the same as in Eq. (81*b*), with the addition that n is the order of harmonic involved.

With E_{min}, E_{max}, E_b, θ_p, grid bias, and the signal voltage determined, oscillograms of plate and grid currents, instantaneous grid and plate losses, power input, etc., can be drawn from characteristic curves of the tube exactly as in the case of a Class C amplifier. Averaging these curves over a full cycle of fundamental frequency then gives power output, plate losses, d-c plate current, etc.

An approximate analysis of triode harmonic-generator behavior can be made by modifying the approximate analysis used with triode Class C amplifiers.[1] It is to be noted first that *the pulse of total space current in an harmonic generator for a given E_{max}, E_{min}, E_b, and θ_p has practically the same shape as in the case of a Class C amplifier.* To the extent that the shape is exactly the same in these two cases, the harmonic content of the space-current pulse is given by Fig. 86. The modifications in the shape of the space-current pulse resulting from the fact that the voltage drop in the plate circuit is a harmonic of the applied frequency instead of being of the same frequency can be allowed for by *subtracting* from the coefficients given by Fig. 86 a correction factor equal to k times the factor obtained from Fig. 87, where

$$k = \frac{E_b - E_{min}}{\mu E_{max} + E_{min}} \quad (95)$$

Here μ is the amplification factor of the tube. The corrected coefficients are then used to calculate the amplifier performance, exactly as the coefficients obtained from Fig. 77 were used in the Class C amplifier case. The details involved are illustrated by the following example:

[1] F. E. Terman, Analysis and Design of Harmonic Generators, *Trans. A.I.E.E.*, Vol. 57, p. 640, November, 1938.

Example.—It is desired to design and calculate the performance of a third-harmonic generator employing a Type 204A tube having $\mu = 23$, a rated plate dissipation of 200 watts, and operating at a plate potential of 1,500 volts.

The first step in the design is the selection of a suitable peak space current and corresponding values of E_{min} and E_{max}. Reference to the data supplied by tube manufacturers indicates that $I_m = 1{,}260$ ma will be satisfactory, and that this could be obtained by $E_{min} = 375$, $E_{max} = 250$. With this combination, the peak grid current will be 225 ma, which is small enough to keep the driving power low, and yet at the same time E_{min} will be small compared with the plate-supply voltage.

The next step is the selection of a value of θ_p. From the discussion given above, combined with an examination of Fig. 86, it appears that $\theta_p = 100°$ would be reasonable. With this value of θ_p and assuming a $\frac{3}{2}$ power law, Fig. 86 then gives $I_0/I_m = 0.161$, and $I_n/I_m = 0.177$. The fractional corrections that Fig. 87 indicates would be applied to these for a $\frac{3}{2}$ law and for $k = 1$ are -0.33 for the d-c term and -0.060 for the third harmonic component. However, since k is

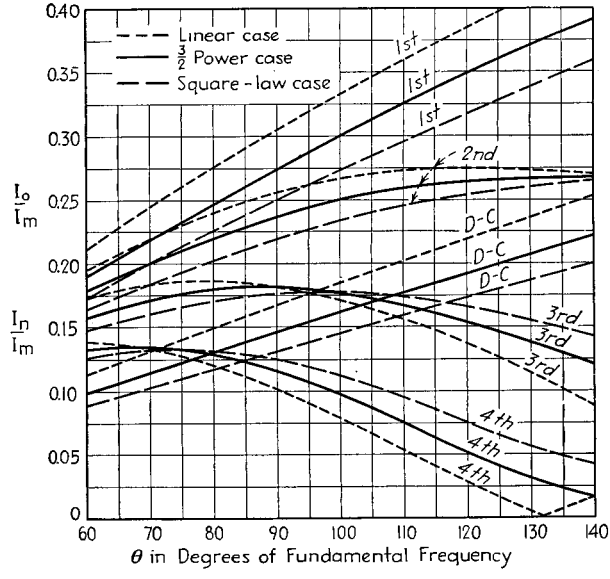


FIG. 86—Curves giving relation of direct-current and harmonic components of the space-current pulse as a function of angle of flow θ and the peak amplitude I_m.

$$\frac{[(1{,}500 - 375)]}{23 \times 250 + 375} = 0.19,$$

the actual corrections to be subtracted are -0.063 and -0.012, respectively. The d-c and third-harmonic components of the total space current are then $1{,}260 \times 0.161(1 + 0.063) = 216$ ma, and $1{,}260 \times 0.177(1 + 0.012) = 226$ ma, respectively. From Eq. (14), the grid bias is found to be 751 volts, which makes the peak signal voltage 1,001 volts. The angle θ_0 of grid-current flow is then given by Eq. (80) as 82.8°. Assuming a square-law grid-current pulse, reference to Fig. 86 shows that for a peak grid current of 225 ma, the d-c, fundamental-frequency, and third-harmonic components of the grid current are 27, 52, and 39 ma, respectively. Since the peak signal voltage is 1,001 volts, the grid driving power is

$$\frac{1{,}001 \times 0.052}{2} = 26 \text{ watts.}$$

If grid-leak bias is used, the leak resistance is $751/0.027 = 28{,}000$ ohms. Next, subtracting the d-c and third-harmonic components of the grid current from the total space current gives 189 and 187 ma, respectively, as the d-c and third-harmonic components of the plate current. This calls for a tank-circuit impedance of $1{,}125/0.187 = 6{,}000$ ohms. The power input to the plate is 284 watts, while the third-harmonic power delivered to the tank circuit is $0.187 \times 1{,}125/2 = 105$ watts. The plate efficiency is hence 37 per cent and the plate dissipation is 179 watts. Inasmuch as this loss is slightly less than the rated value for normal operation, the assumed E_{min}, E_{max}, and θ_p represent a suitable operating condition.

The approximate analysis of harmonic-generator performance can be extended to include pentode, beam, and screen-grid tubes by obvious modifications.[1]

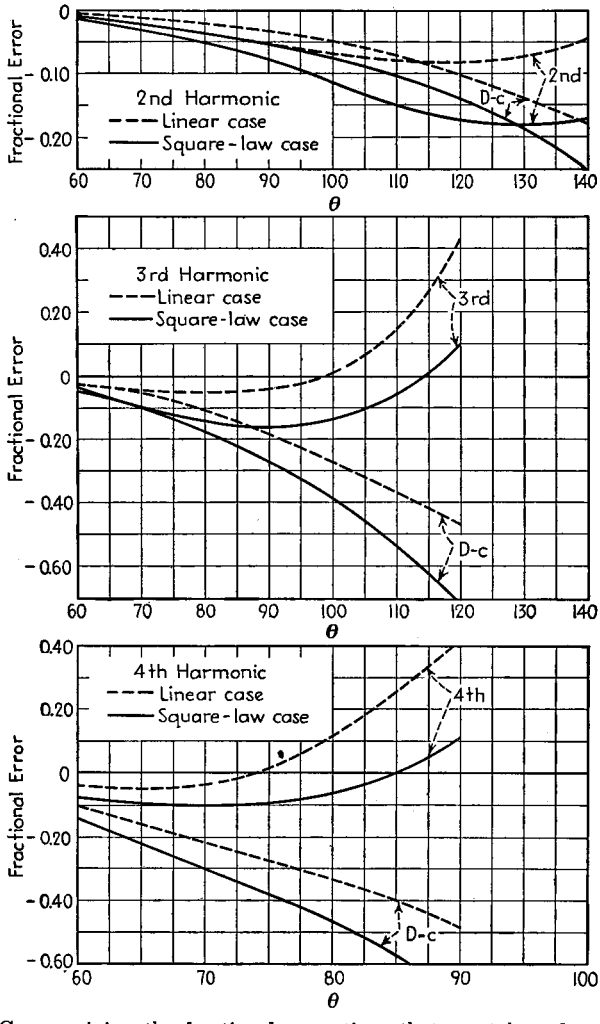


FIG. 87—Curves giving the fractional corrections that must be *subtracted* from the results obtained from Fig. 86 to obtain correct results for harmonic generator action. These curves assume $k = 1$.

MISCELLANEOUS AMPLIFIER PROPERTIES

24. Exact Equivalent Circuit of Class A Amplifier and Application to Analysis of Distortion and Cross-modulation.[2]—The exact behavior of a Class A amplifier can be determined with the aid of the equivalent circuits of Fig. 88.

[1] F. E. Terman and J. H. Ferns, Calculation of Class C Amplifier and Harmonic Generator Performance of Screen-grid and Similar Tubes, *Proc. I.R.E.*, Vol. 22, p. 359, March, 1934.

[2] For further details, see John R. Carson, A Theoretical Study of the Three-element Vacuum Tube,

In the voltage-generator form, the current that flows in the plate circuit as a result of a signal voltage e_g applied to the control grid can be considered as being produced by a series of hypothetical voltages acting in a circuit consisting of the plate resistance of the tube in series with the load impedance, as in Fig. 88*a*. These equivalent generator

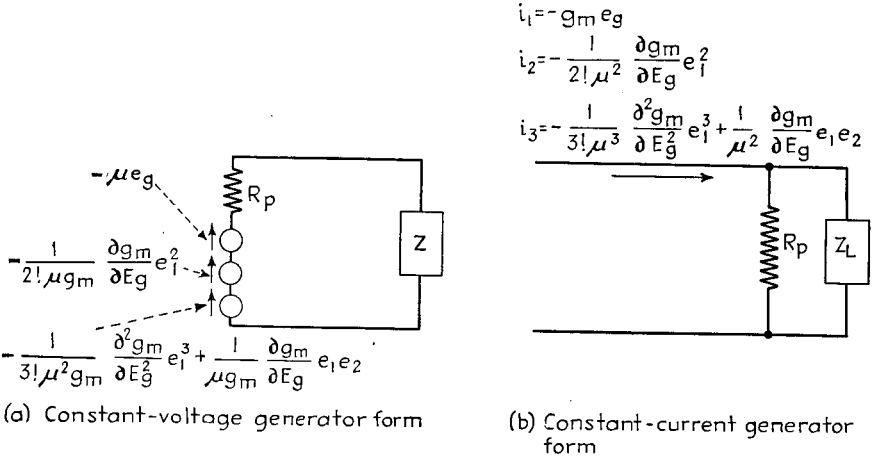


FIG. 88.—Equivalent circuits of a vacuum-tube amplifier, taking into account first-, second-, and third-order components of plate current.

voltages are proportional, respectively, to the first, second, third, etc., powers of the signal voltage and are given by the expressions

$$\left.\begin{array}{l}\text{Equivalent voltage producing}\\ \text{first-order component of}\\ \text{plate current}\end{array}\right\} = -\mu e_g \tag{96}$$

$$\left.\begin{array}{l}\text{Equivalent voltage producing}\\ \text{second-order component of}\\ \text{plate current}\end{array}\right\} = -\frac{1}{2!\mu g_m}\frac{\partial g_m}{\partial E_g}\left(\mu e_g\frac{R_p}{R_p+Z_1}\right)^2 = -\frac{1}{2!\mu g_m}\frac{\partial g_m}{\partial E_g}e_1^2 \tag{97}$$

$$\left.\begin{array}{l}\text{Equivalent voltage producing}\\ \text{third-order component}\\ \text{of plate current}\end{array}\right\} = -\frac{1}{3!\mu^2 g_m}\frac{\partial^2 g_m}{\partial E_g^2}e_1^3+\frac{1}{\mu g_m}\frac{\partial g_m}{E\partial_g}e_1e_2 \tag{98}$$

where e_g = a-c voltage applied to the control grid.
μ = amplification factor of tube.
g_m = transconductance of tube.
E_g = grid bias at the operating point.
R_p = plate resistance at the operating point.

Proc. I.R.E., Vol. 7, p. 187, 1919; F. E. Terman, "Radio Engineering," 2d ed., pp. 153–157, 264–273, McGraw-Hill, New York, 1937.

The discussion in this section assumes either that the amplification factor of the tube is constant [Eqs. (96) to (101)], or that the load resistance is small compared with the plate resistance of the tube [Eqs. (100*a*) and (101*a*)]. Analyses that are more general, but correspondingly more complicated and so less useful, are given by F. B. Llewellyn, Operation of Thermionic Vacuum Tube Circuits, *Bell System Tech. Jour.*, Vol. 5, p. 433, July, 1926; E. Peterson and H. P. Evans, Modulation in Vacuum Tubes Used as Amplifiers, *Bell System Tech. Jour.*, Vol. 6, p. 442, July, 1927; J. G. Brainerd, Mathematical Theory of Four-electrode Tube, *Proc. I.R.E.*, Vol. 17, p. 1006, June, 1929.

Z = load impedance. The subscripts 1, 2, and 3 denote the impedance offered to the first-, second-, and third-order components of the plate current. When any one order of plate current contains components of several frequencies, the appropriate value of Z must be used for each component.

$e_1 = \mu e_g \left(\frac{R_p}{R_p + Z_1}\right)$ = voltage drop produced in plate resistance by first-order component of current.

$e_2 = \left(\frac{Z_2}{R_p + Z_2}\right) \frac{1}{2!\mu g_m} \frac{\partial g_m}{\partial E_g} e_1^2$ voltage drop produced across the load impedance Z_2 by the second-order component of current.

The derivatives in these equations are all to be evaluated at the point on the tube characteristics at which the tube is operated.

In the constant-current generator form, the action taking place in the plate circuit is represented by a number of constant-current generators delivering current to an impedance consisting of the plate resistance in parallel with the load impedance, as in Fig. 88*b*. The currents produced by these constant current-generators are proportional, respectively, to the first, second, third, etc., powers of the applied signal, and are given by the equation.

$$\left.\begin{matrix}\text{Equivalent first-order} \\ \text{current}\end{matrix}\right\} = -g_m e_g \tag{99}$$

$$\left.\begin{matrix}\text{Equivalent second-} \\ \text{order current}\end{matrix}\right\} = -\frac{1}{2!\mu^2} \frac{\partial g_m}{\partial E_g} \left(\mu e_g \frac{R_p}{R_p + Z_1}\right)^2 = -\frac{1}{2!\mu^2} \frac{\partial g_m}{\partial E_g} e_1^2 \tag{100}$$

$$\left.\begin{matrix}\text{Equivalent third-} \\ \text{order current}\end{matrix}\right\} = -\frac{1}{3!\mu^3} \frac{\partial^2 g_m}{\partial E_g^2} e_1^3 + \frac{1}{\mu^2} \frac{\partial g_m}{\partial E_g} e_1 e_2 \tag{101}$$

where $R_p >> Z$, as in pentodes, Eqs. (100) and (101) can be written

$$\left.\begin{matrix}\text{Equivalent second-order} \\ \text{current in pentodes}\end{matrix}\right\} = -\frac{1}{2!} \frac{\partial g_m}{\partial E_g} e_g^2 \tag{100a}$$

$$\left.\begin{matrix}\text{Equivalent third-order} \\ \text{current in pentodes}\end{matrix}\right\} = -\frac{1}{3!} \frac{\partial^2 g_m}{\partial E_g^2} e_g^3 \tag{101a}$$

The first-order effects in the exact equivalent circuits of the Class A amplifier lead to the same results as the equivalent circuits discussed in Par. 3. The second-order effect is a correction of this simple equivalent circuit and depends upon $\partial g_m/\partial E_g$, *i.e.*, upon the curvature of the tube characteristic. The third-order effect is a second-order correction that depends upon the rate of change of curvature of the tube characteristic.

The exact equivalent circuit of the Class A amplifier is useful in calculating distortion, cross-modulation, etc. Its use is not limited to the simple case of resistance load and a sine-wave voltage, as is the method of distortion analysis involving the dynamic characteristic and load line.

Experimental Evaluation of Second- and Third-order Coefficients.—In order to make quantitative calculations based upon Eqs. (96) to (101), it is necessary to evaluate $\partial g_m/\partial E_g$ and $\partial^2 g_m/\partial E_g^2$, as well as the plate impedance, transconductance, etc., at the operating point.

These partial derivatives can be determined experimentally by applying to the grid of the tube a sine-wave voltage of known amplitude when the load impedance

TABLE 5.—FIRST-, SECOND-, AND THIRD-ORDER VOLTAGES IN TWO TYPICAL CASES

	Signal $= E_a \sin \omega t$	Signal $= E_a \sin \omega_a t + E_b \sin \omega_b t$
Equivalent first-order voltage	$-\mu E_a \sin \omega t$	$-\mu(E_a \sin \omega_a t + E_b \sin \omega_b t)$
Equivalent second-order voltage	$-\frac{\mu}{2!g_m}\frac{\partial g_m}{\partial E_g}E_a^2\left(\frac{1-\cos 2\omega t}{2}\right)$	$-\frac{\mu}{2!g_m}\frac{\partial g_m}{\partial E_g}\left(\frac{E_a^2+E_b^2}{2}-\frac{E_a^2\cos 2\omega_a t+E_b^2\cos 2\omega_b t}{2}+E_aE_b\cos(\omega_a+\omega_b)t+E_aE_b\cos(\omega_a-\omega_b)t\right)$
First part of equivalent third-order voltage[1]	$-\frac{\mu}{3!g_m}\frac{\partial^2 g_m}{\partial E_g^2}E_a^3\left(\frac{3}{4}\sin\omega t-\frac{1}{4}\sin 3\omega t\right)$	$-\frac{\mu}{3!g_m}\frac{\partial^2 g_m}{\partial E_g^2}\left\{\left(\frac{3}{4}E_a^3+\frac{3E_aE_b^2}{2}\right)\sin\omega_a t+\left(\frac{3}{4}E_b^3+\frac{3E_a^2E_b}{2}\right)\sin\omega_b t-\frac{E_a^3\sin 3\omega_a t+E_o^3\sin 3\omega_b t}{4}-\frac{3}{4}E_a^2E_b[\sin(2\omega_a+\omega_b)t+\sin(-2\omega_a+\omega_b)t]-\frac{3}{4}E_aE_b^2[\sin(\omega_a+2\omega_b)t+\sin(\omega_b-2\omega_b)t]\right\}$

[1] The second part of the third-order voltage is similar in character to the first-order part, although differing in detail.

in the plate circuit is negligibly small.[1] One then has

$$\frac{\partial g_m}{\partial E_g} = \frac{4\Delta I}{E^2} \tag{102}$$

$$\frac{\partial^2 g_m}{\partial E_g^2} = \frac{24 I_3}{E^3} \tag{103}$$

where ΔI is the change in d-c plate current, I_3 the crest value of the third-harmonic component of the plate current, and E is the crest value of the signal voltage applied to the grid.

The Application of Exact Equivalent Circuit to Simple Cases.—The first-, second-, and third-order components of the equivalent voltage in the exact equivalent circuit for the case of signals consisting of a sine wave and of two superimposed sine waves are summarized in Table 5. An examination of this table shows that the first-order effects account for the undistorted part of the amplified output. The second-order effects give rise to second harmonics, to sum and difference frequencies, and to a direct-current or "rectified" component, all of which are proportional to the square of the signal. The third-order effect gives rise to third harmonics and odd-order combination frequencies. There is also a third-order component having the same wave shape as the applied signal (*i.e.*, as the first-order output) but an amplitude proportional to the cube of the signal, so that when this combines with the first-order part of the output, the result is a lack of proportionality between the amplitude of the input signal and the amplitude of the output. Finally, when the applied signal consists of more than a single frequency, the third-order action causes the part of the output that is at one frequency to depend upon the amplitude of the other components present, giving rise to cross-modulation.

Fourth, sixth, and other higher even-order components give effects similar in character to those produced by the second-order action, namely, even-order harmonics, even-order combination frequencies, and a rectified d-c current. Fifth, seventh, and other higher odd-order components give effects similar to those produced by the third-order action, *i.e.*, odd harmonics, odd-order combination frequencies, lack of proportionality between input and output, and cross-modulation.

Cross-modulation (Cross Talk).—Cross-modulation is said to be present in a circuit when the conditions are such that when two signals of different frequencies are applied, the amplitude of the output of one frequency depends upon the amplitude of the other frequency that is present. In the case where one of the signals is a modulated wave, cross-modulation results in the amplitude of the other wave depending on the envelope amplitude of the modulated wave. This causes the modulation of the first wave to be transferred to the second wave. Cross-modulation occurs in amplifiers as a result of third-order action.

The amount of cross-modulation can be expressed in terms of a *cross-talk* or *cross-modulation* factor that is defined as the percentage modulation that a modulated wave produces on an unmodulated superimposed wave, divided by the percentage modulation of the modulated wave. In the case of pentode and other tubes where the load resistance is small compared with the plate resistance, the cross-talk factor is[2]

$$\left.\begin{matrix}\text{Cross-talk factor} \\ \text{in pentodes}\end{matrix}\right\} = \frac{E_2^2}{2g_m}\frac{\partial^2 g_m}{\partial E_g^2} \tag{104}$$

where E_2 is the crest amplitude of the carrier applied to the grid, whose modulation is

[1] An alternative procedure for measuring the third-order coefficient is given by E. W. Herold, Simple Methods for Checking R. F. Distortion or Cross-modulation of Pentode Amplifiers, *Electronics*, Vol. 13, p. 82, April, 1940.

[2] Stuart Ballantine and H. A. Snow, Reduction of Distortion and Cross Modulation in Radio Receivers by Means of Variable-mu Tetrodes, *Proc. I.R.E.*, Vol. 18, p. 2102, December, 1930.

being transferred to the other carrier frequency. In triode tubes the cross-talk factor depends upon the load impedance to the first- and second-order currents, as well as upon the square of the carrier amplitude and the second derivative of the transconductance.

Distortion of Modulation Envelopes.—It was noted above that third-order action introduces a nonlinear relation between output and input voltage of an amplifier. When the input voltage is a modulated wave, the third-order action therefore causes the output voltage to fail to follow exactly the amplitude variations of the input voltage. This results in a distortion of the modulation envelope, and also makes the degree of modulation of the output voltage differ from that of the applied signal. When the modulation of the applied signal is sinusoidal, and the degree of modulation m is not too great, the second-harmonic distortion introduced in the modulation envelope by third-order action in the case of pentode tubes is

$$\left.\begin{array}{l}\text{Second-harmonic}\\ \text{component of modulation}\end{array}\right\} = \frac{m}{2 + \dfrac{4g_m}{E_0^2 \dfrac{\partial^2 g_m}{\partial E_g^2}}} \qquad (105)$$

where E_0 is the carrier amplitude of the applied signal, and the remaining notation is as above. It will be noted that when the third-order action is large, the distortion approaches $m/2$. Practically, it is found that when modulation distortion is enough to be of significance, cross-modulation is excessive. Cross talk rather than modulation distortion accordingly limits the amount of third-order action permissible over the operating range of an amplifier under ordinary conditions.

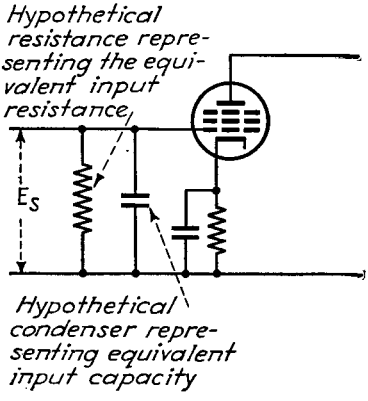


FIG. 89.—Input impedance of a vacuum-tube represented as a resistance shunted by a capacity.

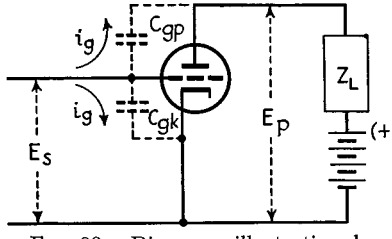


FIG. 90.—Diagrams illustrating how the interelectrode capacities permit grid current to flow when alternating voltage is applied to a negative grid.

25. Input Admittance and Output Impedance of Vacuum-tube Amplifiers. *Input Admittance.*—The input admittance of a vacuum-tube amplifier is defined as the current flowing into the control-grid electrode divided by the voltage that is applied between this electrode and the cathode. The input admittance of a vacuum-tube amplifier can be represented by a capacity shunted by a resistance, as shown in Fig. 89. The capacity is called the input capacity, and the resistance is termed either the input resistance or the input conductance.

Input Admittance of Triodes with Particular Reference to Interelectrode-capacity Effects.—The input capacity and resistance of triodes is determined primarily by the capacities existing between the electrodes of the tube. When an alternating signal voltage is applied between grid and cathode, current flows from grid to cathode and from grid to the plate as a result of the electrostatic capacity between these electrodes,

as indicated in Fig. 90. The input admittance is equal to this total current divided by the voltage applied to the control grid. Since the part of the current that flows to the plate depends upon the difference between the signal applied to the control grid

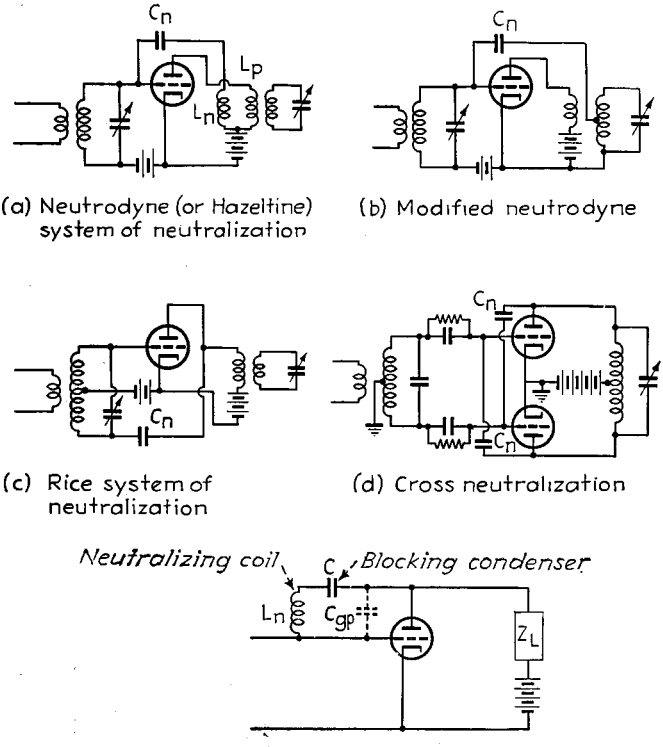


FIG. 91.—Typical neutralizing circuits.

and the amplified voltage developed in the plate circuit, the input admittance is affected by the magnitude and phase of the amplification. The exact relation is[1]

$$\text{Input capacity} = C_{gk} + C_{gp}(1 + A \cos \theta) \tag{106}$$

$$\text{Input resistance} = -\frac{\left(\frac{1}{\omega C_{gp}}\right)}{A \sin \theta} \tag{107}$$

where C_{gp} = grid-plate tube capacity.

C_{gk} = grid-cathode tube capacity.

A = voltage amplification of the tube alone, not taking into account any step-up in voltage existing in the load impedance, such as is present in the case of transformer coupling.

θ = phase angle of load impedance in the plate circuit, taken positive for inductive loads.

[1] For derivation, see F. E. Terman, "Radio Engineering," 2d ed., p. 232, McGraw-Hill, New York. See also J. M. Miller, Dependence of the Input Impedance of a Three-electrode Vacuum Tube upon the Load in the Plate Circuit, *Bur. Standards Sci. Paper* 351; M. von Ardenne and W. Stoff, On the Values and the Effects of Stray Capacities in Resistance-coupled Amplifiers, *Proc. I.R.E.*, Vol. 15, p. 895, November, 1927; E. L. Chaffee, Equivalent Circuits of an Electron Tube and the Equivalent Input and Output Admittances, *Proc. I.R.E.*, Vol. 17, p. 1633, September, 1929.

The input admittance of a triode amplifier depends largely upon the magnitude A and phase θ of the amplification, because with practical tubes the grid-plate capacity has approximately the same magnitude as the grid-filament capacity, and the amplification is normally considerably greater than unity. Under these conditions the input capacity is approximately proportional to the amplification, and is much greater than the interelectrode capacities themselves.

The input resistance takes into account the energy transferred between the plate and grid electrodes through the grid-plate capacity. With capacitive load impedances in the plate circuit (θ negative), energy is transferred directly from the grid to the plate circuit, with the result that the input resistance is positive. On the other hand, with inductive load impedances in the plate circuit (θ negative), amplified energy is transferred from the plate circuit of the tube back to the grid circuit, with the result that the equivalent input resistance is negative. A resistance load impedance in the plate circuit ($\theta = 0$) gives rise to infinite input resistance, corresponding to zero energy transfer between input and output electrodes. The input resistance for a given phase shift θ is inversely proportional to the product of frequency and amplification.

Neutralization of Grid-plate Capacity in Triode Tubes.—Most of the input capacity and practically all the input resistance of triode tubes results from the current that flows through the grid-plate capacity of the tube. The effect of this capacity can be eliminated or neutralized by the arrangements illustrated in Fig. 91. In the neutrodyne circuit the coil L_n is closely coupled to the primary, and so polarized that it applies a voltage to the neutralizing condenser C_n that is of opposite phase from the a-c voltage between plate and cathode. The current through the neutralizing condenser is hence of opposite phase from the current through the grid-plate capacity C_{gp} of the tube, and by proper adjustment of C_n the energies represented by the two currents can be given the same magnitude. The input capacity of such an amplifier with perfect neutralization is $C_{gk} + C_{gp} + C_n$, and the input resistance is infinite. If the neutralizing coil L_n is closely coupled to its primary and the leads have negligible inductive reactance, the neutralization is substantially independent of frequency over a wide frequency band.[1]

An alternative neutralizing arrangement is the Rice circuit (Fig. 91*c*). Here the neutralizing capacity C_n is adjusted so that the current through it neutralizes the effect of the current through the grid-plate capacity as far as the tuned circuit associated with the grid of the tube is concerned. This form of neutralization is theoretically independent of frequency, just as is the neutrodyne type, but has the disadvantage that only half of the signal voltage developed across the input circuit is applied to the grid of the tube, and that neither side of the tuning condenser in the input circuit can be grounded. Also, if several stages are in cascade, Rice neutralization is more likely to cause trouble from parasitic oscillations.

Cross neutralization is used with push-pull amplifiers and requires the addition of no special circuits other than the neutralizing condensers. It can be thought of as a form of neutrodyne that takes advantage of the fact that the voltages on the two sides of a push-pull amplifier are of opposite polarity, thus giving the phase relations required for neutralizing.

In coil neutralization, the neutralizing inductance L_n is resonated with the grid-plate capacity C_{gp} at the frequency for which neutralization is to be effective. In this way the current flowing from control grid to plate is reduced practically to zero, thus eliminating the grid-plate capacity as far as input admittance effects are concerned. In order that the neutralizing coil will not short-circuit the plate-supply

[1] At very high frequencies the variation in the reactance of the leads cannot be neglected. If neutralization is to be maintained over a wide frequency band under these conditions, it is necessary to use modified circuits (see Fig. 20, Sec. 6).

battery, it is necessary to place the blocking condenser C in series with the inductance. The circuit proportions are then arranged so that the combination L_nC offers the inductive reactance required to resonate with C_{gp} at the frequency of neutralization. Coil neutralization is simpler than the other types, but is effective for only one frequency. It is used extensively with transmitters, such as broadcast transmitters. that operate only at a single frequency.

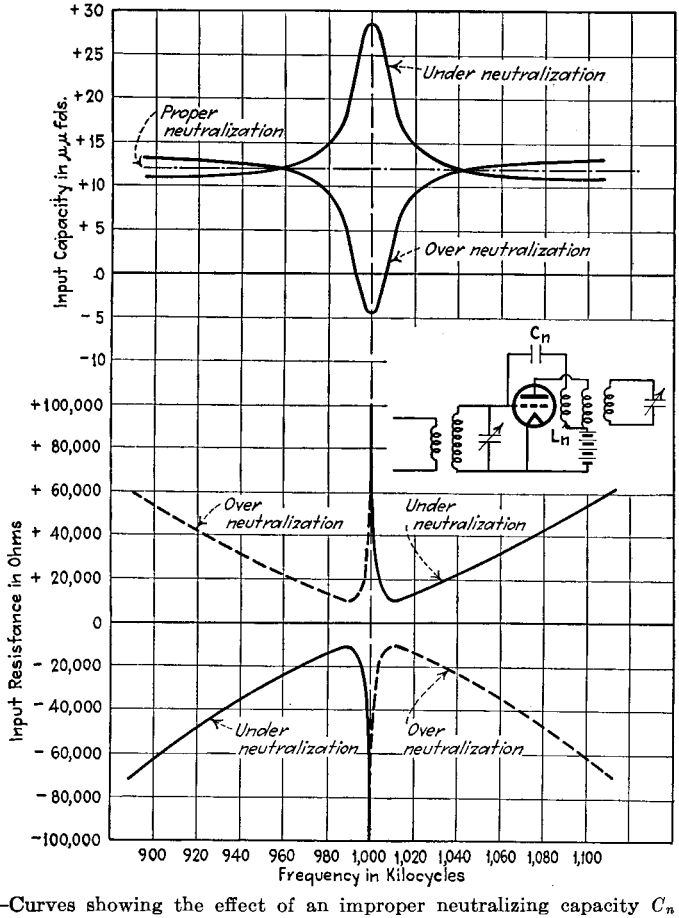


FIG. 92.—Curves showing the effect of an improper neutralizing capacity C_n on input resistance and capacity of a tuned radio-frequency amplifier.

In all neutralizing systems it is necessary to adjust the circuit proportions so that neutralization is exactly correct. Overneutralization produces results similar in character to and just as bad as underneutralization, as illustrated in Fig. 92.

Grounded-grid Self-neutralized Circuit.—The necessity of neutralization can be avoided by means of the grounded-grid circuit of Fig. 93. Here the grid is grounded, the input voltage is applied between cathode and ground, and the output impedance is placed between plate and ground. With this arrangement, capacity currents flowing between plate and grid as a result of the output voltage E_0 developed in the plate circuit do not flow through the input circuit. There is hence no interaction with the source of exciting voltage E_s as a result of grid-plate tube capacity.

Analysis of the voltage and current relations in the grounded-grid circuit shows that since the driving voltage E_s is in series with the external circuit connecting plate and cathode, the tube acts, as far as the output E_0 is concerned, as though it were excited in the usual manner and had an amplification factor $(\mu + 1)$ instead of μ. Also, the exciting voltage delivers some energy directly to the plate circuit, because the amplified plate current flows through E_s in the same polarity as E_s. This places a load on the exciting voltage E_s. This load reduces the gain obtained as compared with a neutrodyne circuit, for example, although appreciable amplification is still possible.

The grounded-grid circuit finds its chief use under conditions where triode tubes must be used, and either (1) it is impractical to obtain the degree of capacity neutralization required for satisfactory operation or (2) the tube or circuit design is such that the natural method of operation is with the grid at ground potential. These conditions are most likely to exist where the tube is to be operated over an enormous frequency range or at a very high frequency, or with special tubes designed for operation at extremely high frequencies.[1]

Input Conductance of Pentodes.—The input conductance of a pentode tube, if negative grid operation is assumed, arises primarily from the transit time of the

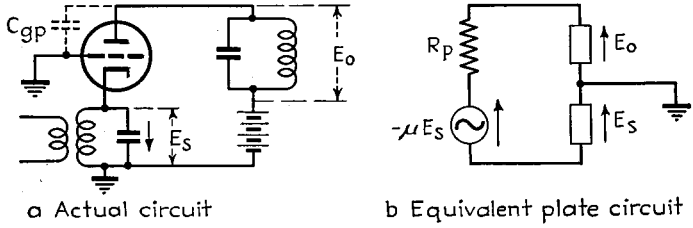


FIG. 93.—Grounded-grid self-neutralized circuit.

electrons and from the fact that the lead from the cathode electrode to the cathode pin on the base of the tube is common to the grid and plate circuits of the tube. Dielectric losses in the grid-cathode and grid-screen capacities are occasionally of importance.[2] The grid-plate capacity contributes very little to the input conductance of a pentode tube, because although this capacity acts the same in pentodes as in triodes, its magnitude is so small (0.005 $\mu\mu$f for a typical voltage amplifier pentode) that the energy transfer taking place between plate and grid is too small to be important.

The input resistance arising from the fact that the electrons require a finite time to travel from cathode to plate is given by[3]

$$\left.\begin{matrix}\text{Input resistance resulting}\\ \text{from finite transit time}\end{matrix}\right\} = \frac{1}{Kg_m f^2 \tau^2} \tag{108}$$

where g_m = transconductance of the tube.

f = frequency.

τ = time required for the electron to travel from the cathode to the grid plane.

The constant K is determined by the grid and plate voltages and by the ratio of transit times from cathode to grid plane and grid plane to anode. The component of input resistance arising from transit time is directly proportional to the square root of the electrode voltages and inversely proportional to the square of the linear dimension of the tube.

[1] Discussion of the grounded grid circuits, with particular reference to high-frequency high-power amplifiers, is given by C. E. Strong, The Inverted Amplifier, *Electronics*, Vol. 13, p. 14, July, 1940.

[2] A discussion of dielectric losses is given by C. J. Franks, Measured Input Losses of Vacuum Tubes, *Electronics*, Vol. 8, p. 222, July, 1935.

[3] W. R. Ferris, Input Resistance of Vacuum Tubes as Ultra-high-frequency Amplifiers, *Proc. I.R.E.*, Vol. 24, p. 82, January, 1936. See Par. 9, Sec. 4 for further discussion of transit time effects.

The inductance of the cathode lead causes the voltage existing across the grid-cathode capacity to differ from the signal voltage applied to the tube by an amount equal to the voltage developed across the lead inductance by the amplified plate current (see Fig. 94). This fact causes the current flowing through the grid-cathode capacity to have a component in phase with the applied voltage.[1] The corresponding input resistance, on the assumption that the reactance of the lead inductance L_K is small compared with the grid-cathode capacity C_{gk}, is

$$\left.\begin{array}{l}\text{Input resistance resulting}\\ \text{from inductance of cathode}\\ \text{lead}\end{array}\right\} = \frac{1}{\omega^2 g_m L_K C_{gk}} \tag{109}$$

It will be noted that the input resistance caused by the inductance of the cathode lead varies with frequency and transconductance in exactly the same manner as does the input resistance resulting from transit time action. In ordinary pentode tubes, these two effects are of the same order of magnitude. The combined action of cathode lead inductance and transit time is of great importance at the higher radio frequencies. This is evident from Table 6, which gives results of measurements on input resistance of a pentode tube.

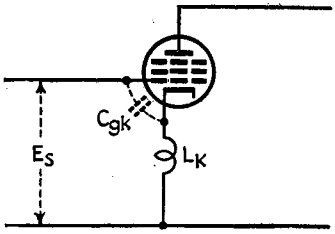


FIG. 94.—The equivalent input circuit of a tube taking into account the inductance of the cathode lead.

The input conductance resulting from the combined effects of cathode-lead inductance and transit time can be neutralized by the arrangement in Fig. 95*a*.[2] Here a small inductance L_n is placed between the cathode terminal on the tube socket and the ground. The grid-return lead is brought directly to the tube socket, and a capacity C_{gg} is connected between grid and ground. Neutralization is obtained when

$$\frac{L_n}{L_k} = \frac{C_{gk}}{C_{gg}} \tag{110}$$

With these proportions, the in-phase component of the current through C_{gg} is equal in magnitude but opposite in phase to the in-phase component of current through C_{gk}, resulting in complete neutralization of the input resistance independently of frequency, provided that $\omega L_k << 1/\omega C_{gk}$. In practical amplifiers, the neutralizing inductance L_n can be a short length of wire of the order of one inch long connecting the cathode terminal of the tube socket to the ground lug, as shown in Fig. 95*b*.

An alternative method of neutralizing the input conductance is to place a small inductance (about 0.1μh) in the screen lead. This method is especially adaptable to converter tubes.

The effect of the cathode lead inductance can be eliminated by using especially constructed tubes having two cathode leads, as shown in Fig. 95*c*.[3] In this way the voltage developed in the cathode lead by the plate current does not appear in series with the voltage applied to the control grid.

Input Capacity of Pentodes, with Particular Reference to the Effect of Control-grid Bias.—The input capacity of a pentode is given by the equation

$$\text{Input capacity} = C_{gk} + C_{gs} + C_{gp}(1 + A\cos\theta) \tag{111}$$

[1] M. J. O. Strutt and A. van der Ziel, The Causes for the Increase of the Admittances of Modern High-frequency Amplifier Tubes on Short Waves, *Proc. I.R.E.*, Vol. 26, p. 1011, August, 1938.

[2] R. L. Freeman, Input Conductance Neutralization, *Electronics*, Vol. 12, p. 22, October, 1939.

[3] Such tubes are discussed by F. Preisach and I. Zakarias, Input Conductance, *Wireless Eng.*, Vol. 17, p. 147, April, 1940.

where C_{gk} = grid-cathode capacity.
C_{gs} = grid-screen capacity.
C_{gp} = grid-plate capacity.
A = voltage amplification (magnitude) of tube alone (not including any voltage step-up in plate coupling network).
θ = phase angle of amplification.

The component $C_{gp}(1 + A \cos \theta)$ is the smallest of the three because of the small grid-plate capacity in pentode tubes, but may be of the order of $1 \mu\mu f$ in radio-frequency

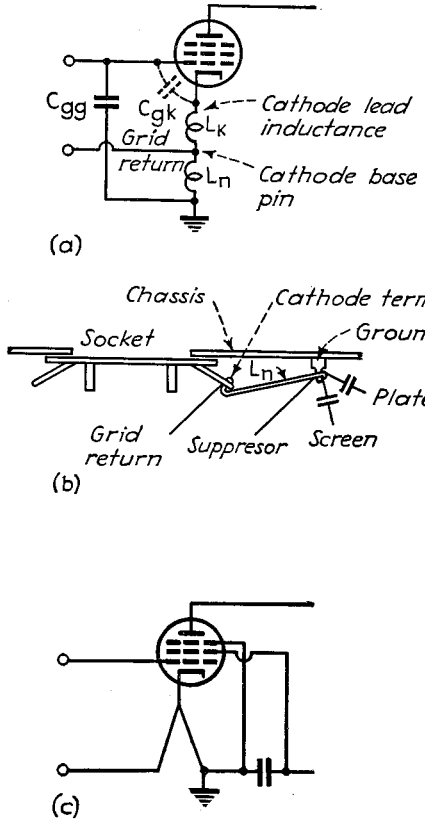


FIG. 95.—Methods of neutralizing the effects of cathode lead inductance and transit time on input resistance.

pentodes when the amplification is large, and somewhat more in audio-frequency pentodes.

The input capacity of a pentode varies somewhat with the grid bias.[1] This is in part due to a change in the component $C_{gp}(1 + A \cos \theta)$ as the grid bias varies the amplification, and in part due to the fact that as the bias varies the plate current the effective position of the space charge surrounding the cathode changes. The total variations produced by varying the grid bias are of the order of 1 to 3 $\mu\mu f$ in typical

[1] R. L. Freeman, Use of Feedback to Compensate for Vacuum-tube Input-capacitance Variations with Grid Bias, *Proc. I.R.E.*, Vol. 26, p. 1360, November, 1938; T. Iowerth Jones, The Dependence of the Inter-electrode Capacitances of Valves upon the Operating Conditions, Wireless Section, *I.E.E.*, Vol. 13, p. 11, March, 1938; also, *Jour. I.E.E.*, Vol. 81, p. 658, 1937.

radio-frequency pentodes. This will detune an ordinary 450-kc intermediate-frequency coupling circuit by about 5 kc.

The change in input capacity is approximately proportional to the change in the transconductance of the tube, and so can be neutralized by the arrangements shown in Fig. 96. Here a resistance R_n is placed between cathode and ground and is so related to the capacity C_{gk} between grid and cathode as to satisfy the relation

$$R_n = \frac{\Delta C}{C_{gk} g_m} \tag{112}$$

where g_m is the transconductance for some particular operating point and ΔC is the increase in input capacity of the tube when the transconductance is g_m over the input capacity when the tube biased to cutoff. When the secondary trimmer condenser C_i is returned to ground, as in Fig. 96*b*, the value of resistance R_n required approximates that normally used for self-bias. On the other hand, if C_{gk} is made large by returning the trimmer condenser of the input circuit to the cathode, instead of to ground, as in Fig. 96*a*, the neutralizing resistance R_n is then of the order of 20 ohms with ordinary r-f pentodes used as intermediate-frequency amplifiers.[1]

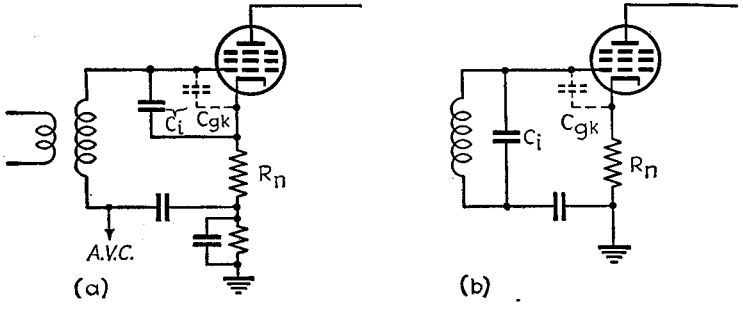


FIG. 96.—Circuits for neutralizing the variation of input capacity of a Class A pentode amplifier with variation in grid bias.

Input Admittance of Amplifier Tube Subjected to a Negative Feedback Voltage.—When a negative feedback voltage is superimposed upon the signal applied to an amplifier, the equivalent input admittance offered by the system to the applied signal voltage is less than in the absence of negative feedback according to the factor

$$\frac{\left\{\begin{matrix}\text{Input admittance with}\\ \text{negative feedback}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Input admittance without}\\ \text{negative feedback}\end{matrix}\right\}} = \frac{1}{1 - A\beta} \tag{113}$$

where $A\beta$ is the feedback factor as defined in connection with Eq. (36). The reduction in input admittance is considerable when $A\beta$ is large, and arises from the fact that most of the signal voltage is used to overcome the negative feedback voltage and only a small fraction is applied across the grid-cathode terminals of the tube.

Practical Importance of Input Admittance in Audio-frequency and Radio-frequency Amplifiers.—In audio-frequency amplifiers, the input capacity is one of the largest shunting capacities in the coupling network in the plate circuit of the preceding ampli-

[1] Another means of compensating for variations in input capacity caused by variation in the control-grid bias by the A.V.C. system is to apply A.V.C. voltage in appropriate amount to the No. 3 grid. See J. F. Farrington, Compensation of Vacuum Tube Input Capacitance Variation by Bias Potential Control, *R.M.A. Eng.*, Vol. 4, p. 13. November, 1937.

fier stage. A pentode is accordingly preferable to a triode tube from the point of view of response obtainable at high frequencies unless the triode is neutralized. In the case of push-pull triodes, neutralization can be easily applied and will reduce the input capacity to the same order of magnitude as that obtained with pentodes, resulting in considerable improvement in high-frequency response.[1] The resistance component of the input admittance is of negligible importance at audio frequencies, since the input resistance is so large as to produce only slight modification of the amplification characteristic, even at the higher audio frequencies with unneutralized triodes.

In tuned radio-frequency amplifiers, the input capacity of the tube is an important factor in fixing the minimum circuit capacity. With unneutralized or partially neutralized triodes, the input capacity varies with frequency, as illustrated in Fig. 92. This dependence of the capacity on frequency can distort the resonance curve of the tuned circuit developing the signal voltage applied to the grid of the tube.

The input resistance at broadcast and lower frequencies is normally caused by the action of the grid-plate tube capacity. With a load impedance supplied by a resonant circuit, the input resistance depends primarily upon frequency, being infinite at resonance and minimum at the 70.7 per cent points of the amplification curve. This minimum value for the case of a simple tuned load circuit is

$$\text{Minimum input resistance} = \frac{(1/\omega C_{gp})}{(A_0/2)} \tag{114}$$

where $1/\omega C_{gp}$ is a reactance of the grid-plate capacity C_{gp} and A_0 is the amplification from grid to plate electrodes at resonance.[2] The input resistance is negative at frequencies below resonance and positive for frequencies greater than resonance, and so tends to distort the resonance curve of the input tuned circuit. If the input resistance is low, oscillations will be produced when the input tuned circuit is resonant at a slightly lower frequency than the output circuit. The effects of the input resistance are so great in tuned amplifiers using triode tubes that such amplifiers must always be neutralized. With pentode tubes, however, the minimum input resistance at broadcast frequencies is so high because of the small grid-plate capacity that it can ordinarily be ignored unless it becomes cumulative through a series of stages.

In the high-frequency and ultra-high-frequency bands, the input resistance is determined primarily by the transit time and cathode lead effects, since these produce effects proportional to the square of frequency, whereas the action of the grid-plate capacity is proportional to frequency. The input resistance at very high frequencies is low enough to produce serious loading, with consequent loss of gain and selectivity.

Output Impedance of Vacuum Tubes.—The output impedance of a vacuum tube is defined as the impedance that the plate circuit of the tube offers to an external voltage applied between the plate and cathode electrodes of the tube. This impedance can be represented by a resistance shunted by a capacity and is the impedance seen by a load looking toward the tube.

In triode tubes the resistance component of the output impedance approximates the plate resistance of the tube up to very high radio frequencies. The output capacity also approximates the sum of the grid-plate and plate-cathode interelectrode capacities. These approximations neglect the effect of dielectric losses, inductance in leads, and energy transfer between the input and output circuits of the tube through the grid-plate capacity, but give satisfactory results for the usual triode operating conditions. In particular, the energy transfer between input and output circuits

[1] P. W. Klipsch, Applying Neutralization to A. F Amplifiers, *Electronics*, Vol. 7, p. 252, August, 1934.

[2] When the tuned load circuit is coupled to the plate circuit of the tube, as with transformer coupling, the value of A_0 is less than the actual voltage amplification of the stage by the factor M/L_s, where M is the mutual inductance between primary and secondary and L_s, the inductance of the secondary.

has a relatively small effect upon the output circuit as compared with the effect on the input circuit because of the difference in power levels.

In pentode and beam tubes the output capacity is equal to the sum of the plate-cathode capacity and the plate-screen capacity. The output resistance, however, depends very greatly upon frequency, being equal to the plate resistance of the tube only at audio and the lower radio frequencies. As the frequency is increased, the output resistance becomes less, first because of dielectric losses associated with the inter-electrode capacities of the tube, and at still higher frequencies as a result of the inductance of the lead wires from the electrodes to the base pins of the tube and capacities between electrodes. The equivalent shunting resistance as measured for a particular pentode tube is shown in Table 6,[1] and it is seen that the output resistance at the higher radio frequencies is far less than the plate resistance of the tube.

26. Tube and Circuit Noise in Amplifiers.—The term *noise* as applied to vacuum-tube amplifiers is commonly used to designate spurious voltages of a random character that represent energy more or less uniformly distributed over an appreciable frequency band. In the case of audio-frequency amplifiers, such noise produces a characteristic hiss.

Thermal Agitation Noise.[2]—The random motion of free electrons in a conductor causes small potential differences to be developed across the terminals of the conductor.

TABLE 6.—INPUT AND OUTPUT RESISTANCE OF A PENTODE AMPLIFIER*

Wave length, meters	Input resistance		Wave length, meters	Output resistance	
	Tube cold, megohms	Normal, megohms		Tube cold, megohms	Normal, megohms
230.0	3.9	3.3	62.5	0.75	0.43
39.5	1.6	0.38	20.4	0.35	0.19
21.2	0.74	0.11	5.05	0.045	0.022
12.4	0.40	0.036			
5.6	0.19	0.0086			

* Data from M. J. O. Strutt and A. Van der Ziel, The Causes for the Increase of the Admittances of Modern High-frequency Amplifier Tubes on Short Waves, *Proc. I.R.E.*, Vol. 26, p. 1011, August, 1938.

This action is termed *thermal agitation*, and the resulting voltage is

$$\left.\begin{array}{l}\text{Square of effective value of}\\ \text{voltage components lying be-}\\ \text{tween frequencies } f_1 \text{ and } f_2\end{array}\right\} = E^2 = 4kT\int_{f_1}^{f_2} R\,df \qquad (115)$$

where k = Boltzmann's constant = 1.374×10^{-23} joule per °K.
T = absolute temperature, °K.
R = resistance component of impedance across which the thermal agitation is developed (a function of frequency).
f = frequency.

[1] From Strutt and van der Ziel, *loc. cit.*

[2] J. B. Johnson, Thermal Agitation of Electricity in Conductors, *Phys. Rev.*, Vol. 32, p. 97, July, 1928; H. Nyquist, Thermal Agitation of Electronic Change in Conductors, *Phys. Rev.*, Vol. 32, p. 110, July, 1928; J. B. Johnson and F. B. Llewellyn, Limits to Amplification, *Elec. Eng.*, Vol. 53, p. 1449, November, 1934; F. C. Williams, Thermal Fluctuations in Complex Networks, Wireless Section, *I.E.E.*, Vol. 13, p. 53, March, 1938, *Jour. I.E.E.*, Vol. 81, p. 751, 1937; F. C. Williams, Coexistent Thermal and Thermionic Fluctuations in Complex Networks, Wireless Section, *I.E.E.*, Vol. 13, p. 327, September, 1938; *Jour. I.E.E.*, Vol. 83, p. 76, 1938.

In the special case where the resistance component of the impedance is constant over the range of frequencies from f_1 to f_2, Eq. (114) reduces to the much simpler form

$$E^2 = 4kTR(f_2 - f_1) \tag{116}$$

In a network involving two or more conductors at different temperatures, the combined thermal agitation effect can be obtained by considering that each impedance involved acts as a generator having the mean square voltage specified by Eq. (115) in series with the impedance.[1] The mean square thermal agitation voltage is by Eq. (116) proportional to the *resistance component* of the impedance across which the voltage is developed, and to be proportional to the band width. The peak amplitude is of the order of 3 to 4 times the equivalent rms value.[2] The energy represented by the thermal agitation voltages developed across a given resistance is uniformly distributed over the entire frequency spectrum from zero frequency to frequencies well above the highest used in communication.

The voltages developed by thermal agitation set a limit to the smallest voltage that can be amplified without being lost in a background of noise. This limit is determined by the band width being amplified and not by the position in the frequency spectrum at which the band is located. The magnitude of the effect can be estimated from the fact that the rms thermal agitation voltage developed across a ½-megohm resistance at 300°K is 6.4 μv for a frequency band of 5,000 cycles.

Noise from Granular Resistances.[3]—Resistances composed of carbon granules (such as an ordinary carbon resistor) generate noise far in excess of the thermal agitation noise when a direct current is passed through the resistance. This high noise arises from fluctuations in the contact resistance between adjacent granules. The noise voltage that results is proportional to the current, and also tends to increase somewhat faster than the resistance.

This effect makes the ordinary carbon resistor unsatisfactory as a plate-coupling resistance in low-level stages of a resistance-coupled amplifier. Carbon resistors are, however, always suitable for bias and screen-grid voltage dropping resistors where by-pass condensers are used.

Tube Noise.—Noise voltages are generated within tubes as a result of a number of actions, the most important of which are (1) random variations in electron emission from the cathode; (2) random variations in the current division between the plate and other positive electrodes, such as the screen grid (this effect is absent in triodes operated with the grid negative); (3) variations in the grid current resulting from positive-ion current. These effects are discussed in Pars. 5 and 13, Sec. 4, where formulas for calculating their magnitudes are also given. Under practical conditions the smallest voltage that can be amplified is sometimes limited by tube noise rather than thermal agitation.

27. Hum.—The term hum is applied to alternating currents appearing in the output of an amplifier as a result of the effect of power-frequency voltages, currents, and fields.

Hum in Audio-frequency Amplifiers.—In audio-frequency amplifiers, hum results from the introduction into the amplifier circuits of currents of the power frequency and its harmonics that are amplified directly by the amplifier. Hum is particularly

[1] F. C. Williams, The Representation and Computation of Fluctuation Voltages, Wireless Section, *I.E.E.*, Vol. 14, p. 325, September, 1939, *Jour. I.E.E.*, Vol. 85, p. 280, 1939.

[2] V. D. Landon, A Study of the Characteristic of Noise, *Proc. I.R.E.*, Vol. 24, p. 1514, November, 1936; The Distribution with Time in Fluctuation Noise, *Proc. I.R.E.*, Vol. 29, p. 50, February, 1941.

[3] C. J. Christensen and G. L. Pearson, Spontaneous Fluctuations in Carbon Microphones and Other Granular Resistances, *Bell System Tech. Jour.*, Vol. 15, p. 181, April, 1936.

troublesome in high-gain audio-frequency amplifiers, since any effects introduced in low-level stages are subjected to large amounts of amplification.

The possible causes of hum in audio-frequency amplifiers are stray electrostatic and magnetic fields, alternating current in the filaments or heaters of the tubes, and poorly filtered power-supply systems. These last two sources of hum are considered in Pars. 9 and 5, Sec. 8. The types of hum that are usually most troublesome in audio-frequency amplifiers are those arising from the presence of electrostatic and magnetic fields.

The chief sources of magnetic fields in audio-frequency amplifiers are the filament leads and the power transformer. These fields induce voltages in coupling transformers, particularly transformers in low-level parts of the amplifier, may induce voltages in closed loops in improperly arranged wiring, and in severe cases may even affect the flow of electrons within the tubes. The most effective way to minimize hum from magnetic fields is to arrange the filament leads in the form of a twisted pair, employ adequate spacing between the low-level parts of the amplifier and the power transformer, and avoid as far as possible the use of coupling transformers in low-level stages. Where coupling transformers cannot be avoided, induced voltages from magnetic fields can be minimized by use of adequate spacing between coupling and power transformers, orientation of the coupling transformers to give minimum mutual inductance, and the use of shielded coupling transformers. It is also helpful to employ a nonmagnetic chassis, or at least to isolate the coupling and power transformers from a sheet-iron chassis by mounting them upon a nonmagnetic base that is then fastened to the chassis.

Electrostatic fields may arise in an amplifier from unshielded wires carrying alternating currents of power frequency, but are more commonly the result of power wires external to the apparatus itself. Hum arising from electrostatic fields can be eliminated by electrostatically shielding all leads that have a high impedance to ground, such as grid leads, by shielding the tubes and the filament wires of low-level stages, and by grounding the chassis to a metal stake driven in wet earth or a water pipe. The use of electrostatic shielding between primaries and secondaries of all transformers (power, input, filament, output, etc.) on the amplifier chassis is also helpful.

Hum in Radio-frequency Amplifiers.—Hum is less troublesome in radio-frequency amplifiers than in audio-frequency amplifiers, since the hum voltages are not amplified directly. However, the same effects that produce hum in audio-frequency amplifiers also will cause any radio-frequency voltage that is being amplified to be modulated at least slightly by the hum.

28. Microphonic Effects.[1]—The term *microphonic*, as applied to vacuum-tube amplifiers, refers to effects produced by mechanical vibration, and may arise from the vibration either of tubes or of the associated circuits.

Microphonic Action in Tubes.—The most important source of microphonic action in amplifiers is the amplifier tube. Vibrations of the tube elements produce variations in plate current that are amplified directly in audio- and video-frequency amplifiers, and that modulate any signal being amplified by radio-frequency amplifiers.

Microphonic effects in tubes can be minimized by preventing vibrations from reaching the tube, by selection of individual tubes for low microphonic action, and by the use of special tubes of unusually rigid construction. The chief way in which vibrations reach a tube is through the base, and this can be greatly reduced by supporting the tube on a shockproof mounting. Individual tubes of the same type may vary as much

[1] D. B. Penick, Measurement and Reduction of Microphonic Noise in Vacuum Tubes, *Bell System Tech. Jour.*, Vol. 13, p. 614, October, 1934; Alan C. Rockwood and Warren R. Ferris, Microphonic Improvements in Vacuum Tubes, *Proc. I.R.E.*, Vol. 17, p. 1621, September 1929.

as 70 db in their tendency to microphonic action, so that there is considerable gain in selecting the least noisy tubes from a large number. A very low susceptibility to vibrations can be obtained by tube constructions that provide unusual rigidity and bracing of the tube electrodes, particularly the filaments and heaters. Improvements of the order of 20 db over standard tubes are obtainable with such special low microphonic tubes.

Microphonic action in tubes sometimes produces a harsh crackling or sputtering type of noise. This is caused by poor contacts between conducting members, or by intermittent electrical leaks across insulation. Such sputter noise contains frequency components in the radio-frequency band, whereas ordinary microphonic noises do not.

Microphonic Action in Circuits.—Mechanical vibration of circuits associated with an amplifier tube will modulate the voltages being amplified. This effect is particularly important in radio-frequency amplifiers and oscillators. The circuit elements that are the most frequent offenders are the coils and variable condensers of resonant circuits. Microphonic effects of this sort can be controlled by rigid construction of the affected circuit elements, by mounting the entire unit on rubber, and by protection against the direct action of sound waves

SECTION 6

OSCILLATORS

POWER OSCILLATORS

1. Circuits for Power Oscillators.[1]—A vacuum-tube oscillator is essentially a vacuum-tube amplifier arranged so that an exciting voltage of the proper magnitude and phase to produce the amplified output is obtained from this output. The tube in such an arrangement acts as an inverter that changes d-c plate power into a-c energy. In power oscillators, the tube is usually operated under conditions of bias, exciting

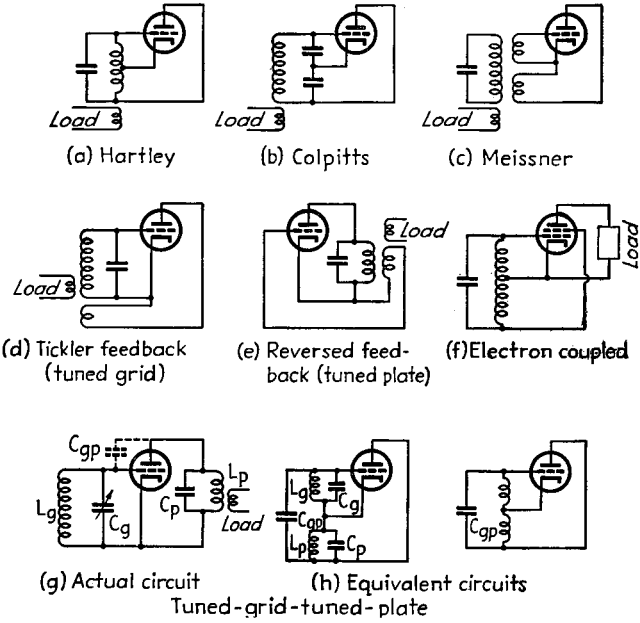


FIG. 1.—Schematic circuits of common types of power oscillators.

voltage, etc., corresponding to Class C amplifier conditions in order to give a large output with good efficiency.

Circuits.—Typical oscillator circuits are shown schematically in Fig. 1. Of these, the ones most frequently used in power oscillators are the Hartley, Colpitts, and the tuned-grid tuned-plate arrangements. In the Hartley and Colpitts circuits an exciting voltage of the appropriate phase is obtained by connecting the grid and plate electrodes to opposite ends of the tank circuit with respect to the cathode connection, with the ratio of exciting voltage to alternating plate-cathode voltage determined by the relative reactances on the two sides of the cathode connection.

[1] For a general discussion of power oscillators, see R. A. Heising, The Audion Oscillator, *A.I.E.E. Jour.*, Vol. 39, p. 365, April; p. 471, May, 1920; D. C. Prince, Vacuum Tubes as Power Oscillators, *Proc. I.R.E.*, Vol. 11, p. 275, June, p. 405, August, and p. 527, October, 1923.

In the tuned-grid tuned-plate circuit, the grid tuned circuit L_gC_g and the plate tuned circuit L_pC_p are both adjusted to offer an inductive reactance at the frequency to be generated. The circuit can accordingly be redrawn as shown in Fig. 1*h*, and is seen to be the equivalent of a Hartley circuit in which the ratio of exciting voltage to alternating plate-cathode voltage is determined by the relative amounts of detuning of the plate and grid circuits. The capacity C_{gp} is usually supplied by the interelectrode capacity of the tube, although at lower frequencies added shunting capacity is helpful.

In the electron-coupled circuit, the cathode, grid, and screen grid of a pentode or screen-grid tube are operated as a triode oscillator, with the screen serving as the anode.[1] Only a small fraction of the space current is intercepted by the screen, but the oscillator circuits are so designed that this will maintain the oscillations properly. The remaining electrons, which represent most of the space current, go on to the plate and produce power output by flowing through the load impedance that is connected in series with the plate electrode. This arrangement has the advantage of making the oscillations independent of the load, and is used where frequency stability is important (see Par. 3).

The various oscillator circuits shown in Fig. 1 are all capable of giving the same power output, efficiency, etc., when used with a particular tube. The choice between them is accordingly one of convenience, when such factors as circuit details, ease of adjustment, tendency for parasitic oscillations to be produced, etc., are taken into account.

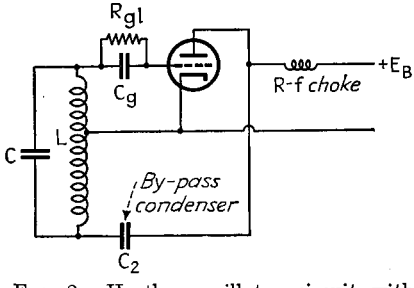


FIG. 2.—Hartley oscillator circuit with shunt feed.

The d-c plate voltage can be introduced in the oscillator either by shunt feed or series feed exactly as in the case of the Class C amplifier. Shunt feed is required in the Colpitts circuit, and is usually preferable in the Hartley circuit.

Bias in power oscillators is practically always obtained by means of a grid-leak grid-condenser arrangement, as shown in Fig. 2, since this makes the oscillator self-starting, and is more likely to give a stable operating point under conditions corresponding to high efficiency operation. The use of a grid leak also tends to make the oscillators self-adjusting at conditions corresponding to good efficiency, and improves the frequency stability.

The grid condenser should have a low enough reactance at the frequency of oscillation so that the alternating voltage drop is small. The condenser should not be too large, however, since then intermittent oscillators may be produced (see below). Also, some capacitive reactance between the grid electrode and the tank circuit is useful in improving the frequency stability (see Par. 3).

Oscillators developing considerable power often have resistances or chokes or both placed in series with leads going to the grid and plate electrodes of the tube in order to eliminate parasitic oscillations (see Par. 5).

Design Considerations.—The voltage and current relations existing in the tube of a power oscillator are exactly the same as in the case of the corresponding Class C amplifier. The performance of the tube can accordingly be determined by calculation as outlined in Par. 21, Sec. 5, except that since the driving power is now obtained from the amplified output, the net output of the oscillator is less than that of the corresponding Class C amplifier by the amount of the driving power.

[1] J. B. Dow, A Recent Development in Vacuum Tube Oscillator Circuits, *Proc. I.R.E.*, Vol. 19, p. 2095, December, 1931.

The circuits of a power oscillator are designed in the same manner as for the Class C amplifier, with the addition that provision must be made so that the grid of the tube obtains from the amplified output an exciting voltage of the proper magnitude and phase. The procedure for laying out the circuits of a power oscillator is illustrated by the following example:

Example.—An oscillator is to be designed to operate under the same conditions as the Class C amplifier in the example of Par. 21, Sec. 5, using a Hartley circuit having an effective Q of 50. Assuming the grid and plate taps are to be at the ends of the tank-circuit inductance, the total alternating voltage across the tank circuit will be the sum of the exciting and alternating plate voltages, or

$$254 + 850 = 1{,}104 \text{ volts.}$$

The filament tap is located so that the ratio between grid and plate voltages is $^{25}/_{850}$. The tank-circuit inductive reactance ωL required is found by Eq. (90), Sec. 5, to be

$$(1{,}104/\sqrt{2})^2/(56 \times 50) = 218 \text{ ohms}$$

for a tank-circuit power of 56 watts. From this the required tank-circuit inductance and capacity for any frequency can be calculated. The tank-circuit circulating current is 1,104/218 = 5.06 amp crest. Since the grid bias required is 134 volts, with a grid current of 14 ma, the grid-leak resistance can be estimated as 134/0.014 = 9,600 ohms. The grid-condenser capacity should be larger than the grid-cathode tube capacity and also large enough to have a low reactance at the operating frequency, but not so large as to cause intermittent oscillations (see below).

If shunt feed is used, as in Fig. 2, the tank-circuit inductance L required to obtain 218 ohms of reactance will be influenced by the fact that the reactance of the radio-frequency choke is in parallel with a portion of the tank circuit. The blocking condenser C_2 for the shunt-feed arrangement must be capable of withstanding the d-c plate voltage E_B and should, at the same time, be large enough to have relatively low reactance compared with the impedance seen by the plate of the tube in looking toward the tank circuit.

If the grid and plate connections were not made to the ends of the coil, the voltage across the tank circuit would have to be increased accordingly, and a higher tank-circuit inductance would be required to obtain the required conditions with the same effective circuit Q. If other oscillator circuits than the Hartley had been employed, minor modifications of this general procedure would be necessary.

Practical Adjustment of Oscillator Circuits.—The procedure to be followed in placing an oscillator in operation depends upon the completeness with which the preliminary design has been worked out. When every detail has been predetermined, including even such things as the proper points on the tank circuit to connect the plate, grid, and cathode leads and the appropriate coupling to the load, it is necessary merely to apply voltage to the oscillator, check the resulting grid bias, power output, frequency, etc., and then make minor readjustments as required to realize exactly the desired behavior.

In the more usual case, the paper design has been carried only to the point where the tank-circuit inductance and capacity, the expected power output, operating grid bias, d-c plate current, etc., are known, while such things as the coupling to the load, the grid-leak resistance, the cathode connection to the tank circuit in Hartley circuits, and the exact tuning in a tuned-grid tuned-plate circuit, are to be determined by trial and error. Under such conditions the adjustment process should start with relatively loose coupling between the load and tank circuit and with the other circuit adjustments at estimated values. After oscillations are obtained, the coupling to the load is increased until the output approximates the rated value and the input power, efficiency, grid bias, etc., noted. Readjustments are then made of grid-leak resistance, exciting voltage, load coupling, etc., as required.

The trial-and-error adjustment of oscillators can be systematized by keeping in mind the fact that oscillators adjusted for reasonable efficiency have the following characteristics: (1) The bias is determined primarily by the exciting voltage. Increasing the excitation increases the bias and reduces the angle of flow while at the same time increasing the maximum grid potential somewhat. (2) The maximum positive grid potential E_{max} is determined both by the exciting voltage and by the grid-leak resistance, being greater as the grid-leak resistance is reduced and as the excitation is increased. (3) The tank-circuit current is determined primarily by the coupling

between the tank circuit and plate circuit of the tube, since the tank current is normally such that the crest alternating voltage developed between plate and cathode is just less than the plate-supply voltage. Hence more coupling between the plate and tank circuits reduces the tank-circuit current. (4) The minimum plate potential E_{min} is determined largely by the coupling between the load and the tank circuit, as is also the d-c plate current. Close coupling increases the dissipation in the load for a given tank-circuit current. This tends to reduce the amplitude of oscillations, thereby increasing the minimum plate potential and increasing the d-c plate current that the plate draws. (5) The grid current is determined by the exciting voltage, grid-leak resistance, and load coupling, and for practical purposes is fixed by the maximum grid voltage E_{max} and the minimum plate potential E_{min}.

2. Miscellaneous Considerations in Power Oscillators. *Starting of Oscillations.*—Oscillators employing grid-leak bias are always self-starting. This is because when the plate voltage is first applied, the grid is at zero potential, and the resulting transconductance is so large that if oscillations will exist at all they will build up, with thermal agitation voltages in the circuit acting as the initiating impulse.[1]

Oscillators employing a fixed bias that places the initial operating point close to or beyond cutoff may not be self-starting, because the plate current, and hence the transconductance, will either be low or zero when the plate voltage is first applied. In such cases oscillations may persist if once started by a momentary reduction in the bias or by a sufficiently large transient voltage induced in the circuits.

Amplitude of Oscillations.—After being started, the oscillations increase in amplitude until the power generated by the tube is equal to the energy dissipated in the load and tank circuits. In practical power oscillators employing grid-leak bias and operating under normal full-load conditions, this equilibrium occurs when the peak alternating voltage developed by the tank circuit between plate and cathode is just slightly less than the plate-supply voltage, corresponding to a small minimum plate potential E_{min}. As the load on the tank circuit is reduced, the output voltage will increase, but even with zero load will never exceed the d-c plate voltage, since this would reduce the minimum plate potential to zero and give zero plate current. Varying the load on an oscillator within reasonable limits therefore has only a small effect upon the amplitude of oscillations. The grid and plate currents are, however, quite sensitive to load, the former increasing and the latter decreasing as the load becomes less.

Intermittent Operation.—When a grid leak is used to develop the bias in an oscillator, it is sometimes found that the oscillations obtained are periodically interrupted at an audio- or radio-frequency rate. Such behavior results when the time constant $R_{gl}C_g$ of the grid leak-condenser combination is too large to permit the bias to readjust itself with the rapidity necessary to follow changes in amplitude produced by random effects. When interrupted oscillations are produced, the rate of interruption is of the order of $1/R_{gl}C_g$ times per second, where R_{gl} and C_g are in ohms and farads, respectively.

Interrupted oscillations can be eliminated by reducing the time constant $R_{gl}C_g$. In practice this means reducing the grid-condenser capacity, since the grid-leak resistance is fixed by the bias requirements.

The tendency for interrupted oscillations to be produced is greatest when the exciting voltage and grid bias are both large, corresponding to conditions with a small angle of plate-current flow and high efficiency.

Blocking.—The phenomenon of blocking appears as a sudden stoppage of oscillations, accompanied by a reversal of grid current, and an increase of plate current to a

[1] Actually, the tendency for oscillations to start will be increased slightly if a few volts of negative bias is applied to the control grid, since in this way the grid current will be initially zero and less energy will be required to sustain a small amplitude of oscillations than when the bias is zero.

value much higher than can be obtained with the full direct-current supply voltage and zero grid potential. A high-power tube is usually destroyed by blocking, since the energy dissipated at the plate is enormous. Blocking is caused by operating conditions that permit secondary electron emission to take place at the grid to such an extent that the grid loses more electrons by secondary emission than it gains from the cathode by direct flow. This causes a reversal of the grid current, resulting in the development of a positive grid-bias voltage by the grid leak, and consequently an excessive plate current flow. In order that blocking may exist, it is necessary that the minimum instantaneous plate voltage and the maximum instantaneous grid potential obtained during the cycle both be high and that the grid leak have a high resistance.

Blocking, when it occurs, is the result of attempting to force the output of the oscillator by increasing the load resistance coupled into the plate circuit. This results in a reduction in the amplitude of oscillations, which increases the minimum plate voltage. If the grid excitation is then increased, the secondary electron emission at the grid will be increased because of the increased positive potential reached by the grid. Under unfavorable conditions this will cause the net grid current to become less, which reduces the grid bias and makes the maximum positive grid potential still greater, causing a further reduction in the grid current, etc.

In order for blocking to occur, the grid must emit secondary electrons. In modern air-cooled tubes the grid structures have usually been so treated that the grid current will never reverse under conditions encountered in ordinary operation, so that blocking cannot occur in these tubes. However, with water-cooled tubes, or where very high electrode voltages are employed, it is not always possible to construct the tube in such a way that negative grid current can be avoided. Such tubes are therefore susceptible to blocking, and either must be operated with considerable care or must be provided with a grid rectifier similar to that shown in Fig. 21, which has the resistance R adjusted so that the rectified current always exceeds the most negative grid current obtainable. In this way the current through the grid leak can never reverse, and all possibility of blocking is avoided.

Wave Shape.—The output of a power oscillator contains harmonics as a result of the fact that the plate and grid currents of the tube flow in the form of pulses. The factors controlling the amplitude of the harmonic voltages developed in the load are the same as in the Class C amplifier. In general it may be said that the harmonics will be reduced if the angle of plate current flow is large (corresponding to low efficiency) and if the effective Q of the tank circuit is large, corresponding to low L/C.

Tubes.—The same tubes are used for power oscillators as for Class C amplifiers.

FREQUENCY STABILITY OF OSCILLATORS—CRYSTAL OSCILLATORS

3. Frequency and Frequency Stability of Ordinary Oscillators.—The frequency generated by a vacuum-tube oscillator adjusts itself to a value such that the voltage that the oscillations apply to the grid of the tube is of exactly the proper phase to produce the oscillations that supply this grid-exciting voltage. This frequency approximates the resonant frequency of the tank circuit,[1] but is also influenced by such factors as the voltages acting on the tube, the effective Q of the tank circuit, the harmonics generated, the resistance and reactance coupled into the tank circuit by the load, etc. All these modifying factors tend to produce small phase shifts between the exciting voltage and the output voltage of the tube, and in order to compensate for these, the oscillator must operate slightly off the resonant frequency of the tank circuit in order to introduce a compensating phase shift.

[1] When there is more than one resonant circuit, as in the tuned-grid tuned-plate arrangement, the frequency tends to be controlled by the circuit in which the circulating reactive volt-amperes are greatest.

Tank-circuit Design for Stable Frequency Operation—Resonant Line Oscillators.—The resonant frequency and the effective Q of the tank circuit are the two most important factors affecting the frequency generated by an oscillator. The generated frequency will vary with the resonant frequency of the tank circuit, while the effect of the tube voltages, load impedance, etc., on the frequency are inversely proportional to the effective Q of the tank circuit.

The resonant frequency of the tank circuit will vary as a result of the temperature coefficient of the inductance and capacity. Included in the temperature effect is the capacity of the tubes that are associated with the resonant circuit and so affect the frequency. In considering the effect of temperature, one must include not only changes in ambient temperature but also the heating of the inductance, capacity, and tubes that occurs during normal operation.

The effect on frequency of unavoidable variations in tube capacity can be minimized by arranging the oscillator circuits so that the electrostatic energy stored in the tube capacities is as small as possible compared with the total electrostatic energy stored in the resonant circuit.[1] This means a high circuit Q, the highest possible ratio of transconductance to changes in tube capacities, and the smallest possible coupling from the tank circuit to the grid and plate electrodes of the tube that will permit oscillations to be maintained. In practical oscillators the ratio of electrostatic energy in the resonant circuit to the electrostatic energy stored in the tube capacities can be made greater the lower the resonant frequency. As a result, an oscillator directly generating the desired frequency can be expected to have poorer frequency stability with respect to temperature changes in the tube, and to tube replacements, than a combination consisting of a harmonic generator and an oscillator operating at a fraction of the desired frequency.

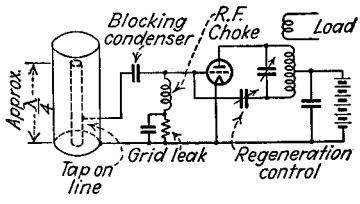


FIG. 3.—Tuned-grid tuned-plate oscillator using resonant line for the tuned-grid circuit.

The effective Q of the tank circuit is important because the percentage change in frequency required to produce a given compensating phase shift between input and output voltages of the tube is inversely proportional to the effective Q. A high effective Q corresponds to a load not too closely coupled, and either to a low L/C ratio in the tank circuit, or an arrangement in which the plate is coupled to only a small part of the tank circuit so that the voltage across the circuit is much greater than the plate-cathode voltage.[2]

High tank-circuit efficiency with high effective Q requires a low loss (high Q) resonant circuit. At frequencies greater than 15 mc to 20 mc, this can be most conveniently realized by the use of resonant lines for the tank circuit.[3] Such lines have high Q (see Par. 16, Sec. 3), are simple to construct, and are mechanically stable. They are also able to dissipate their losses with only a small temperature rise, and where desired can be compensated for temperature changes. An example of a resonant-line oscillator is shown in Fig. 3. This is a tuned-grid tuned-plate arrangement with the line supplying the grid tuned circuit. High frequency stability is achieved by designing the line to have high Q and then connecting the grid so that the section of the line

[1] Walter Van B. Roberts, The Limits of Inherent Frequency Stability, *R.C.A. Rev.*, Vol. 4, p. 478, April, 1940.

[2] G. F. Lampkin, An Improvement in Constant Frequency Oscillators, *Proc. I.R.E.*, Vol. 27, p. 199, March, 1939.

[3] J. W. Conklin, J. L. Finch, and C. W. Hansell, New Methods of Frequency Control Employing Long Lines, *Proc. I.R.E.*, Vol. 19, p. 1918, November, 1931; C. W. Hansell, Resonant Lines for Frequency Control, *Elec. Eng.*, Vol. 54, p. 852, August, 1935; Clarence W Hansell and Philip S. Carter. Frequency Control by Low Power Factor Line Circuits, *Proc. I.R.E.*, Vol. 24, p. 597, April, 1936.

between grid and cathode is only a small part of the total, so that the reactive energy stored in the line will be large.

Effect of Load on Frequency—Master-oscillator and Electron-coupled Arrangements.—The load affects the frequency of an oscillator as the result of the resistance and reactance coupled into the tank circuit. For maximum stability the load should be a resistance, and the coupling to the tank circuit small, so that the oscillator operates only lightly loaded.

The influence of the load on frequency can be minimized by using the oscillator to excite a power amplifier that develops the power required by the load. Such an arrangement is termed a *master-oscillator power-amplifier* system (abbreviated MOPA) and is most effective when the power amplifier consists of several stages employing pentodes, screen-grid tubes, or neutralized triodes. The stage following the master oscillator is then operated as a "buffer amplifier," *i.e.*, with relatively low grid current, so that the loading on the master-oscillator stage is a minimum. Changes in the grid current of the buffer amplifier such as might be produced by variations in the plate-circuit conditions in the amplifier then have minimum effect on the frequency.

The electron-coupled oscillator is equivalent to an oscillator and power amplifier combined in one tube, and can be made to give a frequency practically independent of

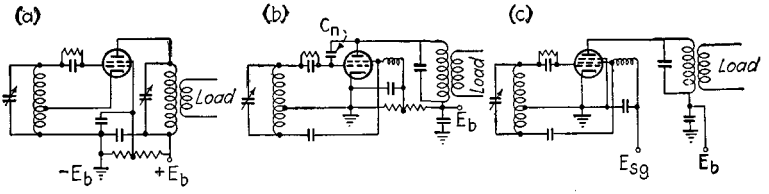


FIG. 4.—Electron-coupled oscillator circuits.

the load impedance that receives the output. To achieve this result, the minimum instantaneous plate potential must be great enough so that there is no tendency for a virtual cathode to form in the tube and return electrons to the screen and cathode, or for secondary electrons to be lost by the plate. This means that the load impedance must not be too great. Further, in the case of screen-grid tubes, the screen grid must be operated at ground potential as at Fig. 4*a*, or the plate-screen capacity must be neutralized as in Fig. 4*b*, in order to prevent capacitive coupling from the plate or output portion to the oscillator section of the tube.

Electron-coupled oscillators using screen-grid tubes generate a frequency practically independent of plate-supply voltage, provided that the ratio of screen to plate voltage is maintained at a suitable value. This desirable result can be obtained by deriving the screen voltage from a voltage divider connected across the plate voltage as in Fig. 4 and adjusting the screen tap by trial until the frequency does not change with a small variation in supply voltage.

Effect of Tube Voltages on Frequency.—The plate and filament voltages acting on the tube affect the frequency by influencing the alternating currents that flow between the tank circuit and the grid and plate electrodes of the tube. This can be minimized by using tank circuits having a high effective Q (low L/C), by the use of grid-leak bias, particularly a high-resistance grid leak, and by employing the closest possible coupling between the different parts of tapped coils, such as are employed in the Hartley circuit, or between the separate coils used in arrangements such as the reversed feedback.

The frequency can be made substantially independent of plate and filament voltages by inserting suitable reactances in series with the leads going to the plate and grid

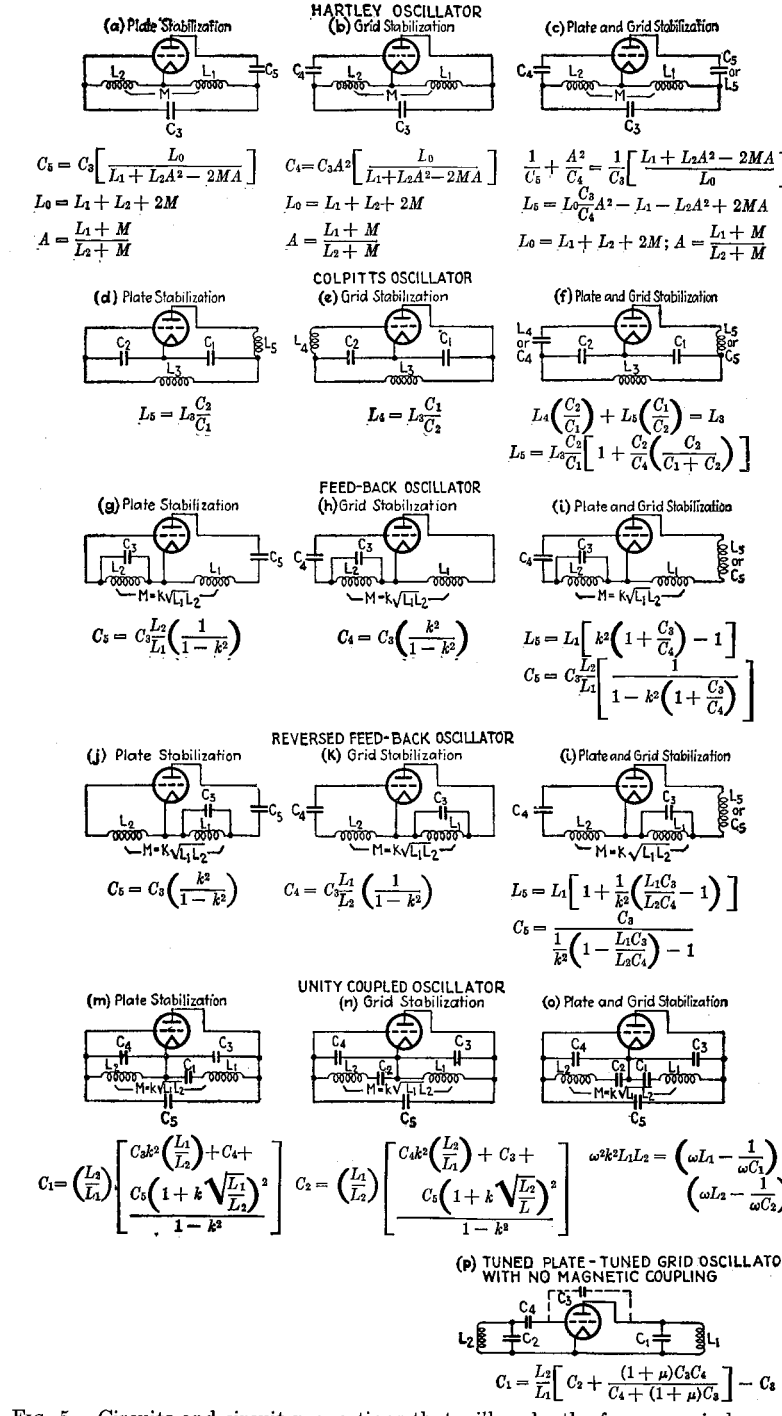


FIG. 5.—Circuits and circuit proportions that will make the frequency independent of tube constants, provided that the Q of the resonant circuit is very high (theoretically infinite).

electrodes.[1] This is shown in Fig. 5, which also gives the formulas for calculating the required reactance in the ideal case of infinite effective Q. Under actual conditions these idealized formulas give only approximate results and the required reactance must be determined by trial. Examination of the circuits of Fig. 5 shows that the grid condenser associated with the grid-leak resistance, or the plate blocking condenser used in shunt-feed arrangements, can be used to supply the required reactance. Thus the frequency of any ordinary oscillating circuit can be made substantially independent of plate and filament voltage for one frequency by a proper choice of these condensers.

When circuits are compensated in the manner illustrated in Fig. 5, the frequency will also be independent of the load, provided that the load is a resistance and is connected either from grid to cathode or from plate to cathode.

The frequency can also be made independent of electrode voltages by means shown in Fig. 6.[2] Here the tap a on the inductance L divides the tank-circuit voltage in the same ratio as do the two condensers C_1 and C_2. A resistance (or coil) is connected between the junction of the condensers and the point a as shown, and the cathode return brought to a point on this resistance that trial shows gives the best frequency stability.

Effect of Harmonics on Frequency Stability.—Harmonic voltages in the circuits of an oscillator adversely affect the frequency stability. This arises from the fact that the harmonics cross-modulate with each other and with the fundamental to produce fundamental frequency currents that are not necessarily in phase with the fundamental frequency currents arising from the normal mode of operation. The phase of the resultant fundamental frequency current affects the frequency of operation, which accordingly becomes sensitive to the amount and distribution of the harmonics in the circuits. The effect of harmonic voltages on frequency stability can be minimized by employing a tank circuit having high effective Q, so that the impedance to the harmonics will be a minimum, and by selecting operating conditions that keep the amplitude of the harmonic voltages and currents at a minimum.

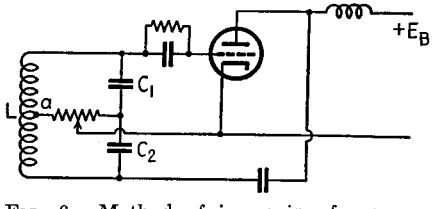


FIG. 6.—Method of improving frequency stability of an oscillator.

4. Crystal Oscillators.[3]—The frequency stability of an oscillator can be made very high by replacing the usual resonant circuit that controls the frequency by a mechanically vibrating piezoelectric quartz crystal and utilizing the piezoelectric effect to give energy transfer between the electric circuits and the mechanical vibrations.

Properties of Piezoelectric Quartz Crystals.—A piezoelectric quartz crystal in its natural state has a hexagonal cross section and pointed ends, as illustrated in Fig. 7.

[1] F. B. Llewellyn, Constant Frequency Oscillators, *Proc. I.R.E.*, Vol. 19, p. 2063, December, 1931; G. H. Stevenson, Stabilized Feedback Oscillators, *Bell System Tech. Jour.*, Vol. 17, p. 458, July, 1938.

[2] Y. Kusunose and S. Ishikawa, Frequency Stabilization of Radio Transmitters, *Proc. I.R.E.*, Vol. 20, p. 310, February, 1932.

[3] Following is a small bibliography on crystal oscillators selected from the very extensive literature on the subject W. G. Cady, The Piezo-electric Resonator, *Proc. I.R.E.*, Vol. 10, p. 83, April, 1922; G. W. Pierce, Piezo-electric Crystal Resonators and Crystal Oscillators Applied to the Precision Calibration of Wavemeters, *Proc. Am. Acad.*, Vol. 59, p. 81, 1923; W. G. Cady, Bibliography on Piezo-electricity, *Proc. I.R.E.*, Vol. 16, p. 521, April, 1928, F. R. Lack, Observations on Modes of Vibration and Temperature Coefficient of Quartz Crystal Plates, *Proc. I.R.E.*, Vol. 17, p. 1123, July, 1929, F. R. Lack, G. W. Willard, and I. E. Fair, Some Improvements in Quartz Crystals Circuit Elements, *Bell System Tech. Jour.*, Vol. 13, p. 453, July, 1934; W. P. Mason, Electrical Wave Filters Employing Quartz Crystals as Elements, *Bell System Tech. Jour.*, Vol. 13, p. 405, July, 1934; W. P. Mason, Low Temperature Coefficient Quartz Crystals, *Bell System Tech. Jour.*, Vol. 19, p. 74, January, 1940; Isaac Koga, Notes on Piezoelectric Quartz Crystals, *Proc. I.R.E.*, Vol. 24, p. 510, March, 1936; C. F. Booth, The Application and Use of Quartz Crystals in Telecommunications, *Jour. I.E.E.*, Vol. 88, Part III, p. 97, June, 1941.

The properties of such a crystal can be conveniently referred to three sets of axes as shown in the figure. The first of these, designated as the optical or Z axis, passes through the apexes of the crystal. The second set, termed the electrical or X axes, consists of the three axes that lie in a plane perpendicular to the optical axis and pass through the corners of the crystal. The third set, termed the mechanical or Y axes, consists of three axes that lie in the same plane as the X axes, but are perpendicular to the faces of the crystal. The exact location of the axes can be determined even when the crystal is imperfectly formed, or when a complete crystal is not available, by using a polariscope or an X-ray spectrometer.

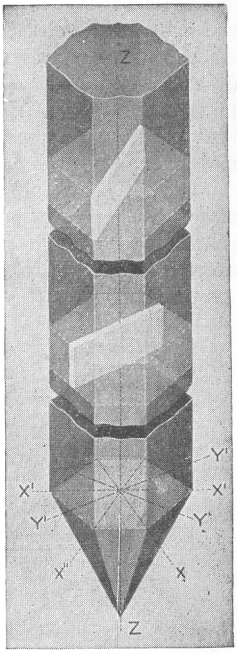


FIG. 7.—Illustration showing the natural quartz crystal and the relation of the electric or X, the mechanical or Y, and the optical or Z axes to the crystal structure. The upper section shows a Y (or 30°) cut plate, while the plate in the center section is X- (or Curie-) cut. The third Y axis, $Y''Y''$, is not shown because the perspective of the drawing makes it coincide with the ZZ axis.

When an electric stress is applied to a quartz crystal in the direction of the X axis, a mechanical stress is produced in the direction of the Y axis at right angles. Conversely, a mechanical stress along the Y axis will cause electrical charges to appear on the faces of the crystal perpendicular to the X axis that goes at right angles to the Y axis involved. The polarity of the electric stress and the direction of the corresponding mechanical force are directly related, a reversal in one causing a reversal in the other. This relationship between electrical stress and mechanical force is termed the *piezoelectric effect*, and provides a means of relating mechanical vibrations to electrical circuits. The piezoelectric effect is present to some extent with electrical stress applied in any direction, since there will nearly always be a component of the stress parallel to an X axis. Similarly, any mechanical force will usually produce some mechanical stress in the direction of a Y axis.

An alternating voltage applied across a quartz crystal will cause the crystal to vibrate, and if the frequency of the applied alternating voltage approximates a frequency at which mechanical resonance can exist in the crystal, the amplitude of vibrations will be very large. Any crystal has a number of such resonant frequencies that depend upon the crystal dimensions, the type of oscillation involved, and the orientation of the plate cut from the natural crystal. Various modes of oscillation are possible in crystal resonators. Those most commonly used are simple longitudinal and shear oscillations, but it is also possible to generate flexural and torsional vibrations and various harmonic modes.[1]

The desirable properties for a piezoelectric resonator to possess are a low temperature coefficient of resonant frequency, a high piezoelectric activity, and a frequency spectrum containing only one resonant frequency in the vicinity of the desired oscilla-

[1] Some examples of crystal oscillators employing such special modes of vibration are given by J. R. Harrison, Piezo-electric Resonance and Oscillatory Phenomena with Flexural Vibrations in Quartz Plates, *Proc. I.R.E.*, Vol. 15, p. 1040, December, 1927; August Hund and R. B. Wright, New Piezo Oscillations with Quartz Cylinders Cut Along the Optical Axis, *Jour. Research*, Vol. 4, p. 383, March, 1930; *Proc. I.R.E.*, Vol. 18, p. 741, May, 1930; N. H. Williams, Modes of Vibration of Piezo-electric Crystals, *Proc. I.R.E.*, Vol. 21, p. 990, July, 1933; Harold Osterberg and John W. Cookson, Piezoelectric Stabilization of High Frequencies, *Rev. Sci. Instruments*, Vol. 5, p. 281, August, 1934; Tsi-Ze Ny, Long-Chao Tsien, and Sun-Hung Fang, Oscillations of Hollow Quartz Cylinders Cut Along the Optic Axis, *Proc. I.R.E.*, Vol. 24, p. 1484, November, 1936.

tion. Temperature can alter the frequency of mechanical resonance through its effects on the density, the linear dimensions, and the moduli of elasticity of the crystal. Inasmuch as some of the elastic constants of a crystal are positive, while others are negative, the temperature coefficient of frequency may be either positive or negative or zero, according to the mode of oscillation, the orientation of the crystal plate, and the shape of the plate.

The term activity applied to a quartz crystal indicates the ease with which the desired oscillations can be excited, and is affected by the orientation of the crystal plate with respect to the axes. The activity will be high when the orientation is such that an electric stress applied across the faces of the crystal will produce a large mechanical stress of the type required to maintain the oscillations.

The resonance properties of a vibrating crystal are determined by the crystal dimensions, orientation, and six elastic constants, and can be calculated for those shapes that are sufficiently simple to permit of mathematical solution. The temperature coefficient of frequency can also be calculated from the crystal orientation and the temperature coefficient of the elastic constants, and conditions giving low or zero temperature coefficients thereby predicted.

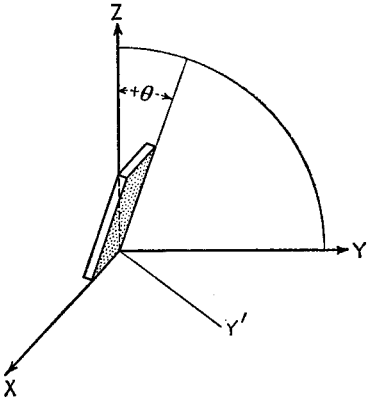


FIG. 8.—Diagram showing how crystal plane is rotated about the X axis to obtain the AT cut.

Common Crystal Cuts.—An infinite variety of shapes and orientations of crystals can be used in crystal oscillators. The crystals used in practice are, however, practically always in the form of plates that are either circular or approximately square if the frequency is high, or sometimes in the form of bars if the frequency is low.

The commonly used orientations of plates with respect to the crystal axes are designated as AT, CT, GT, V, X, Y, etc., cuts.

The AT-cut crystal[1] is extensively used for frequencies ranging from about 500 kc upward to about 10 mc. This type of plate is cut from a plane that is rotated about an X axis so that the angle θ made with the Z axis, as indicated in Fig. 8, is approximately 35.5°. When voltage is applied across the flat sides of such a plate, shear vibrations are set up, causing displacements indicated by Fig. 9*a*. The resonant frequency is related to the thickness of the crystal by the equation

$$\text{Frequency, kc, for } AT \text{ cut} = \frac{1{,}675}{\text{thickness, mm}} \qquad (1)$$

The AT-cut crystal is characterized by substantially zero temperature coefficient at a temperature determined by the exact value of θ, and has a high piezoelectric activity. Furthermore, the elastic coupling between the desired mode of oscillation and harmonics of the width vibration is almost zero, so that the frequency spectrum is very simple. This is in contrast with the complicated frequency spectrum of Y-cut plates.

The CT-cut plate is suitable for operation in the frequency range of 50 to 100 kc.[2] It is essentially an AT-cut plate operating in a mode such that vibrations cause displacements as shown in Fig. 9*b*. The frequency of vibration is accordingly determined

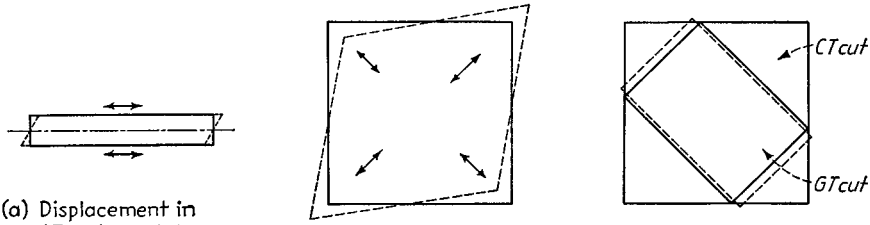
[1] See Lack, Willard, and Fair, *loc. cit.*

[2] S. C. Hight and G. W. Willard, A Simplified Circuit for Frequency Substandards Employing a New Type of Low-frequency Zero-temperature-coefficient Quartz Crystal, *Proc. I.R.E.*, Vol. 25, p. 549, May, 1937.

by the large dimension of the plate, and, in the case of plates, x mm square is given by the equation

$$\text{Frequency, kc, for } CT\text{-cut plates} = \frac{3{,}070}{x} \tag{2}$$

When the plates are square and are cut so that the angle θ in Fig. 8 is approximately 37.5°, the temperature coefficient of vibration will be zero at a temperature approximating room temperature. The CT plate has the very desirable feature that its



(a) Displacement in AT-cut crystal

(b) Displacement in CT-cut crystal

(c) Displacement in GT-cut crystal and relation of GT and CT cuts

FIG. 9.—Displacements produced by the vibrations in various crystal cuts.

frequency can be decreased slightly by grinding in the center of the plate and increased slightly by grinding near the edges, as indicated by D and I, respectively, in Fig. 10*a*. Also, the temperature coefficient can be adjusted either upward or downward by grinding on edges P or N, respectively, as indicated in Fig. 10*b*.

The AT- and CT-cut plates are particular cases of a general class of shear vibrations having zero temperature coefficient, and obtained by rotating the plane of the plate appropriately about two axes.[1] Most of these zero temperature cuts are not particularly useful, however, either because of low activity or because of a complicated

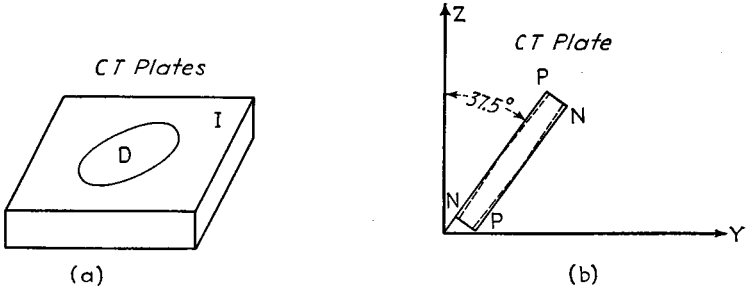


FIG. 10.—Diagram illustrating how the temperature coefficient and resonant frequency of CT plates may be controlled by edge grinding.

frequency spectrum. An exception to this is the V-cut crystal, which has high activity combined with low temperature coefficient.

The frequency of most low-temperature-coefficient cuts depends upon the temperature in the manner illustrated by the AT and CT plates in Fig. 11. The temperature coefficient of frequency is zero only in the vicinity of a particular temperature that is determined by the exact orientation of the plate. An exception

[1] See W. P. Mason, Low Temperature Coefficient Quartz Crystals, *Bell System Tech. Jour.*, Vol. 19, p. 74, January, 1940.

is the *GT* cut, for which the temperature coefficient of frequency is zero over a wide temperature range, as illustrated in Fig. 11.

The *GT* plate[1] is a modified *CT* type in which the plane of the crystal in Fig. 9 corresponds to $\theta = 51.5°$ instead of 37.5°. The crystal is rectangular instead of square with a ratio of width to length of 0.855. The length is in the direction of the diagonal of the *CT* plate, as indicated in Fig. 9*c*. Such a crystal has two principal modes of vibration, both of which are longitudinal, as indicated in Fig. 9*c*, and correspond to resonant frequencies determined by the length and width, respectively. The *GT* crystal uses the width vibration and has a temperature coefficient that depends upon the angle θ of the crystal plane and the ratio of width to length. For each value of θ over a wide range, there is a particular ratio of width to length at which the temperature coefficient at a given temperature is zero. For the particular value of θ corresponding to the *GT* cut, this coefficient is substantially zero over a wide temperature range, as indicated in Fig. 11. The frequency of a *GT* plate is determined primarily

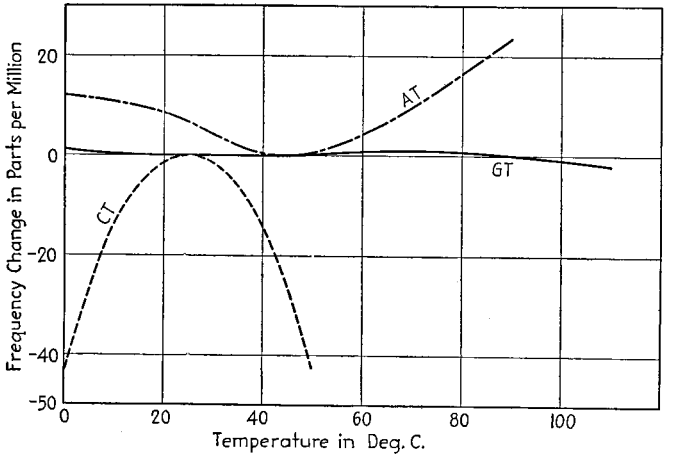


FIG. 11.—Variation of resonant frequency with temperature with various types of crystals.

by the width, while the temperature coefficient is controlled by the ratio of width to length. The frequency and temperature coefficient can hence be adjusted independently.

Y-cut crystal plates[2] were once widely used, but have now largely been displaced by other cuts having superior characteristics. The orientation of a *Y*-cut plate is shown in Fig. 7. Such a plate has one resonant frequency determined by the thickness, and involves the same shear mode of vibration used in *AT*-cut plates. This resonant frequency is given approximately by the equation

$$\left.\begin{matrix}\text{Thickness frequency, kc} \\ \text{of } Y\text{-cut crystal}\end{matrix}\right\} = \frac{1{,}960}{\text{thickness, mm}} \tag{3}$$

Y-cut plates also possess a low-frequency resonance corresponding to a single longitudinal (or width) vibration, having the frequency

[1] W. P. Mason, A New Quartz Crystal Plate, Designated the GT, Which Produces a Very Constant Frequency over a Wide Temperature Range, *Proc. I.R.E.*, Vol. 28, p. 220, May, 1940.

[2] A good discussion of *X*- and *Y*-cut plates is given by Lack, *loc. cit.* Also see Geoffrey Builder and J. E. Benson, Contour Mode Vibrations in *Y*-cut Quartz-crystal Plates, *Proc. I.R.E.*, Vol. 29, p. 182, April, 1941.

$$\left.\begin{array}{l}\text{Width frequency,}\\ \text{kc, of } Y\text{-cut crystal}\end{array}\right\} = \frac{2{,}860}{\text{width, mm, along electric axis}} \qquad (4)$$

Y-cut plates operating at the thickness frequency possess a very complicated frequency spectrum as a result of elastic couplings to harmonics of the width vibration. In a limited frequency range, there are commonly a series of resonant frequencies, as illustrated in Fig. 12. These are affected by the width of the crystal, and although they can be changed by edge grinding to change the width, such grinding merely causes a new series of resonant frequencies to move in to replace those thereby moved out. Thickness vibrations of Y-cut crystals are characterized by a discontinuous temperature-frequency characteristic, as illustrated in Fig. 13. This is a result of elastic couplings to harmonics of the width vibration, and arises from the fact that the width vibration has a negative temperature coefficient of approximately -20 parts in a million per°C, whereas a pure thickness vibration has a positive coefficient. The resultant temperature coefficient of the thickness vibration may vary between -20 and $+100$ parts in a million per°C, with the exact value depending upon the temperature and the ratio of width to thickness. It will be noted from Fig. 13 that in limited temperature ranges the temperature coefficient may be zero. This was the earliest method discovered for obtaining a temperature coefficient of zero.[1]

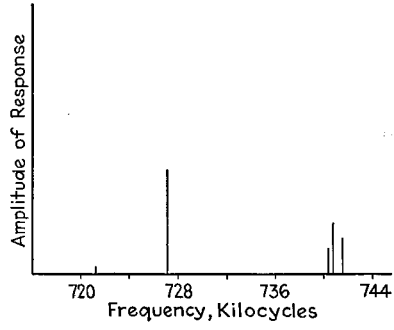


FIG. 12.—Typical frequency spectrum of thin quartz plate showing how several closely spaced resonant frequencies sometimes occur. The length of each vertical line indicates in a rough way the relative tendency to oscillate at that frequency.

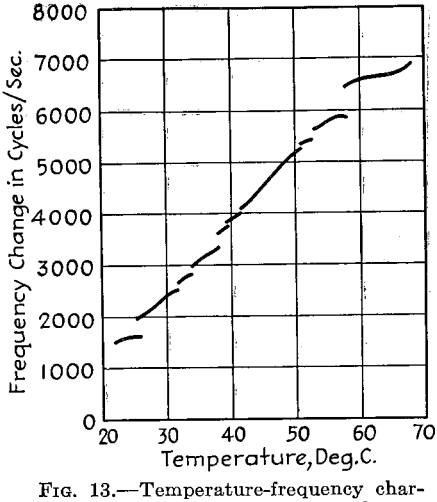


FIG. 13.—Temperature-frequency characteristic of a Y plate showing the discontinuities in frequency and the dependence of temperature coefficient upon the exact temperature.

The X-cut crystal plate is oriented in relation to the crystal axis as shown in Fig. 7. When voltage is applied across the two faces of the crystal, it vibrates with simple longitudinal vibrations, with two possible resonant frequencies corresponding to the thickness of the crystal, or the width in the direction of the Y axis, according to the relation

$$\left.\begin{array}{l}\text{Resonant frequency,}\\ \text{kc, of } X\text{-cut plate}\end{array}\right\} = \frac{2{,}860}{\text{thickness or width, mm}} \qquad (5)$$

The X-cut crystal tends to have a complicated frequency spectrum similar to that obtained with Y cut, and is not quite so active as the Y or AT cuts. The temperature coefficient under ordinary conditions is approximately -20 parts in a million per°C. However, if the width vibration is used, and the length in the direction of the Z axis is very small compared

[1] W. A. Marrison, A High Precision Standard of Frequency, *Proc. I.R.E.*, Vol. 17, p. 1103, July, 1929.

with the width, then the temperature coefficient of frequency is small, and will be zero when the bar is square and has sides that are 0.272 times the width along the mechanical axis.

Equivalent Electric Circuit of Crystal Oscillator.[1]—As far as an electrical circuit associated with a vibrating crystal is concerned, the crystal can be replaced by the electrical network of Fig. 14*b*, in which C_1 represents the electrostatic capacity between the crystal electrodes when the crystal is in place but not vibrating, and the series combination L, C, and R represents the equivalent mass, compliance, and frictional loss of the vibrating crystal, respectively. The impedance offered by the crystal to the electrical circuits accordingly is of the character shown in Fig. 14*c*, being large at the resonant frequency of L and $C + C_1$ and low at a near-by frequency for which L and C are in series resonance. The behavior of electrical circuits containing a crystal can be analyzed by replacing the crystal by its equivalent electrical network and then determining the behavior of the resulting system.

The magnitudes of L, C, and C_1, used to represent the crystal in Fig. 14, depend upon the type of vibration involved, the dimensions of the crystal, and the orientation

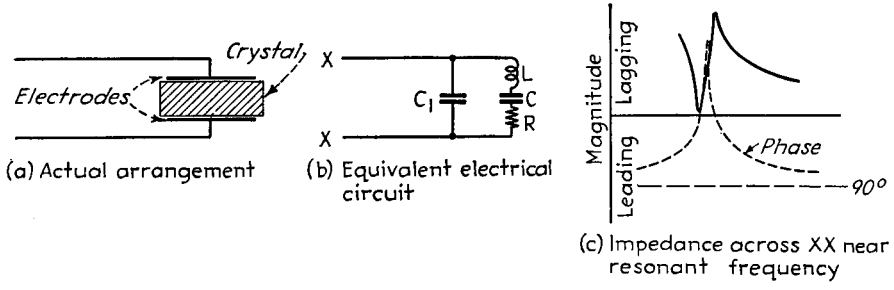


FIG. 14.—Circuit of crystal oscillator together with equivalent circuit in which the crystal is represented by an electric network.

of the crystal plate with respect to the axes. Numerical values for an X-cut plate can be calculated from the equations

$$C = 0.0029 \frac{lw}{t} \mu\mu f \tag{6a}$$

$$C_1 = 0.40 \frac{lw}{t} \mu\mu f \tag{6b}$$

$$L = 118 \frac{t^3}{lw} \text{ henrys (thickness vibration)} \tag{6c}$$

$$= 118 \frac{wt}{l} \text{ henrys (width vibration)} \tag{6d}$$

where w, l, and t are the dimensions in centimeters measured in the direction of the Y axis parallel to the surface of the crystal, in the direction of the X axis perpendicular to this Y axis, and in the direction of the Z axis, respectively. Typical numerical results are shown in Table 1. It will be noted that since the ratio C/C_1 is very much less than unity, the coupling between the crystal and an external electrical circuit associated with the crystal is quite small. The equivalent Q of a crystal resonator is affected by the method of mounting, etc., and is generally very high. Thus etched

[1] K. S. Van Dyke, The Piezo-electric Resonator and Its Equivalent Network, *Proc. I.R.E.*, Vol. 16, p. 742, June, 1928; D. W. Dye, Piezo Electric Quartz Resonator and Its Equivalent Electrical Circuit, *Proc. Phys. Soc. (London)*, Vol. 38, p. 399, 1926; P. Vigoreaux, *Phil. Mag.*, p. 1140, December, 1928; W. P. Mason, An Electromechanical Representation of a Piezo-electric Crystal Used as a Transducer, *Proc. I.R.E.*, Vol. 23, p. 1252, October, 1935.

crystals with plated electrodes, mounted in a vacuum, have been found to give Q's of the order of 500,000.[1]

Crystal Oscillator Circuits.—A quartz crystal can be used to control the frequency of an oscillator by utilizing the crystal to provide the resonant circuit that determines the oscillator frequency. The commonest such circuit arrangement employed with crystals is that shown in Fig. 15*a*. This is equivalent to a tuned-grid tuned-plate circuit, with the crystal supplying the grid tuned circuit. To obtain oscillations, the resonant circuit in the plate is tuned to a frequency higher than that of the crystal, so that the reactance in the plate circuit is inductive at the crystal resonance. The amplitude of oscillation is determined by the amount of inductive reactance in the

TABLE 1.—CHARACTERISTICS OF TYPICAL QUARTZ CRYSTAL

Dimensions:	
Thickness	0.636 cm
Width	3.33 cm
Length	2.75 cm
Resonant frequency (thickness vibration)	430 kc
Equivalent electrical characteristics:	
L = 3.3 henrys	
C = 0.042$\mu\mu$f	
C_1 = 5.8$\mu\mu$f	
R = 4,500 ohms {very approx.}	
Q = 2,300 {very approx.}	

plate circuit, and by the grid-plate tube capacity. In some instances, such as with pentodes, or with crystals operating at low frequencies, it is desirable to supplement the grid-plate tube capacity by an auxiliary condenser C_2, as in Fig. 15*b*. The frequency of oscillation in Fig. 15*b* can be made insensitive to circuit adjustments and to the electrode voltages of the tube by proper choice of the capacities C_2 and C_p. The optimum conditions can be readily realized by first adjusting C_2 so that the generated frequency is maximum and then adjusting C_p for maximum generated frequency, after which the process is repeated in the same order to take care of slight interactions that may be present.[2]

In the circuits of Fig. 15*c* and 15*d*, the crystal is placed in the network that couples the output of the vacuum-tube amplifier to the input, and thereby controls the frequency through its effect on the magnitude and phase of the feedback.[3] In these arrangements, the resonant frequency approximates very closely the series resonant frequency of LC in the equivalent crystal circuit, and so is lower than in the arrangements of Figs. 15*a* and 15*b*.

The circuit of Fig. 15*e* utilizes the crystal as one arm of a bridge, in which the other arms are the resistances R_1, R_2, and R_3.[4] These resistances are arranged so that, together with the equivalent series resistance R of the crystal, the bridge would be in balance except for R_1. This last resistance is provided by a lamp in which the resistance of the filament is sensitive to temperature and hence to the current passing through the bridge. For proper operation, the value of R_1 with the lamp cold is lower than required to balance the bridge, while with the lamp at full brilliancy R_1 is higher than needed for balance. Output from the amplifier tube is applied across

[1] Karl S. Van Dyke, A Determination of Some of the Properties of the Piezo-electric Quartz Resonator, *Proc. I.R.E.*, Vol. 23, p. 386, April, 1935.

[2] See Hight and Willard, *loc. cit.*

[3] An analysis of the circuit of Fig. 15*d* is given by M. Boella, Performance of Piezo Oscillators and the Influence of Decrement of Quartz on the Frequency of Oscillation, *Proc. I.R.E.*, Vol. 19, p. 1252, July, 1931.

[4] See L. A. Meachem, The Bridge-Stabilized Oscillator, *Proc. I.R.E.*, Vol. 26, p. 1278, October, 1938.

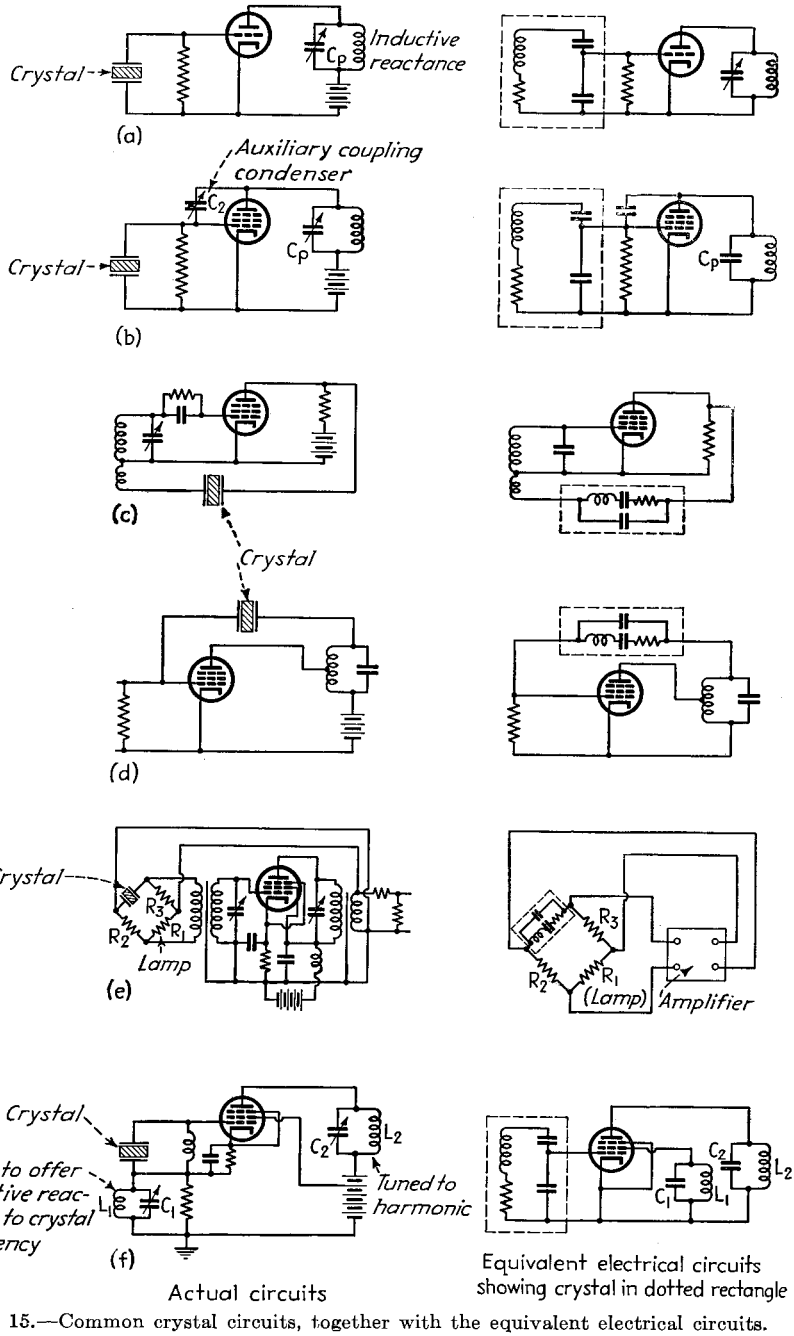


FIG. 15.—Common crystal circuits, together with the equivalent electrical circuits.

one diagonal of the bridge, and input to the same amplifier is derived from the other diagonal. When plate voltage is first applied, the bridge is out of balance because R_1 is too low. Energy is then transferred from the plate of the tube, through the bridge, and back to the grid, thereby starting oscillations at the series resonant frequency of the crystal. As these oscillations build up in amplitude, the resulting current through R_1 increases its resistance, bringing the bridge more nearly in balance and reducing the tendency to oscillate until finally an equilibrium amplitude is reached with a current that causes the bridge to be almost but not quite balanced. If the amplification of the tubes is large, the generated frequency will be substantially independent of the tube constants, the voltages applied to the tube, or the electrical circuits outside the bridge.

One of the factors contributing to the very high frequency stability obtained with the circuit of Fig. 15*e* is that the equilibrium amplitude can be limited so that the tubes are able to function as Class A amplifiers. The frequency stability of any of the other oscillator circuits of Fig. 15 can likewise be increased by using some form of automatic amplitude control that limits the operation to small amplitudes, corresponding to Class A conditions. The disadvantage of doing this, however, is that the power output obtained from the crystal is then quite small.

A circuit combining a crystal oscillator and harmonic generator is shown in Fig. 15*f*. This is commonly used by amateurs, and is known as the tritet circuit. Here the control grid, screen grid, and cathode of a pentode or screen-grid tube function as a triode crystal oscillator with the screen serving as the plate, while the plate tank circuit is tuned to a harmonic. This arrangement is convenient where it is not necessary to attain the utmost in frequency stability.

The frequency of a crystal oscillator can be controlled to the extent of a few parts in a million without regrinding the crystal by adding inductance or capacity in series with the crystal in circuits such as Fig. 15*e*, where the crystal operates at its series resonant frequency, or in shunt in cases where the crystal operates at its parallel resonant frequency as in Fig. 15*a*. The extent of the control obtained in this way over frequency can be analyzed by replacing the crystal by its equivalent resonant circuit and then determining the resonant frequency of the combination of crystal and associated reactances.

Crystal Mountings.[1]—The crystals used in piezo electric oscillators are normally mounted between two electrodes to which the voltage developed by the vacuum tube is applied. In some instances these electrodes are deposited directly on the crystal faces by sputtering or by chemical means. More commonly, however, the electrodes are metal plates. The upper plate in some cases will be in contact with the crystal, either resting lightly or under spring pressure, while in other cases a clearance of a few thousandths of an inch is provided between the upper plate and the crystal. When an air gap is used, it is necessary either that it be small in order to avoid standing supersonic waves or that the entire mounting be evacuated.[2]

The method of holding the crystal in place depends upon the type of crystal, mode of vibration, and the kind of service. For ordinary commercial applications where the frequency stability must be good but ruggedness is also important, it is desirable to have a rigid mounting. This can be obtained with Y- and AT-cut crystals vibrating in their thickness mode by arranging the upper plate so that it touches the crystal only at a few points around the periphery, and then is held firmly against the

[1] O. M. Hovgaard, Application of Quartz Plates to Radio Transmitters, *Proc. I.R.E.*, Vol. 20, p. 767, May, 1932; Vincent E. Heaton and E. G. Lapham, Quartz Plate Mountings and Temperature Control for Piezo Oscillators, *Proc. I.R.E.*, Vol. 20, p. 261, February, 1932; R. C. Hitchcock, Mounting Quartz Oscillator Crystals, *Proc. I.R.E.*, Vol. 15, p. 902, November, 1927.

[2] August Hund, Notes on Quartz Plates, Air Gap Effect, and Audio-frequency Generation, *Proc. I.R.E.*, Vol. 16, p. 1072, August, 1928.

crystal by a spring. Crystals vibrating at low frequencies, such as the *CT* and *GT* cuts, and *X*- and *Y*-cut plates using longitudinal (width) vibrations, possess nodal points at which rigid clamping is possible without introducing appreciable damping.

In primary and secondary standards it is usually preferable to arrange the mounting so that the crystal rests on the lower plate and does not touch the upper plate. This reduces the damping and avoids the possibility of irregularities. The air gap with such mountings is preferably controlled by the use of pyrex or other spacers having the same temperature coefficient of expansion as quartz. When a spaced upper plate is used, it is necessary to employ a retainer to hold the crystal in place, for otherwise it will creep. In some cases the crystal holder is sealed and evacuated.

When the utmost in frequency stability is required, it is customary to use a mounting provided with a heater and thermostat that maintain the crystal at approximately constant temperature. This is desirable even with crystals such as the *AT*- and *AC* cuts having a low temperature coefficient, even though under these conditions the temperature need not be regulated very closely. If high frequency stability is to be obtained with *X*- and *Y*-cut crystals, it is necessary that the design of the thermostat and heater be such as to maintain crystal temperature to 0.01°C or less.[1]

Performance of Crystal Oscillators.—Crystal oscillators are capable of operating in the frequency range 50 to 15,000 kc. The low-frequency limit arises from the difficulty of obtaining large quartz plates, while the high-frequency limit comes from the fact that if the plate is extremely thin it becomes too fragile to be practicable.

The power that can be controlled by a crystal oscillator is limited at low frequencies by the tendency of the crystal to crack from mechanical stress due to large amplitudes of vibration, and at high frequencies by the tendency of the crystal to overheat as a result of dissipation of radio-frequency energy in the crystal. In high-frequency work, the radio-frequency current passing through the crystal is commonly taken as a measure of the load on the crystal, with 100 ma generally considered as a reasonable upper limit. Practical crystal oscillators operating in the range 1.5 to 10 mc commonly develop 5 to 15 watts output. Although larger powers can be obtained, particularly if pentode tubes are used, this causes the frequency stability to be lowered and introduces danger of damaging the crystal. When frequency stability is very important, operation at still smaller outputs is desirable.

The frequency of a crystal oscillator such as is used in commercial broadcast stations will normally remain constant to at least 10 parts in a million over long periods of time, and not infrequently the variations will not exceed one part in a million. When the utmost care is taken to minimize every possible cause of frequency variation, stabilities of the order of one part in 10^7 to one part in 10^8 can be maintained over long time intervals.

MISCELLANEOUS TYPES OF OSCILLATORS AND OSCILLATIONS

5. Parasitic Oscillations.—The term *parasitic* is applied to undesired oscillations occurring in ordinary power amplifiers, oscillators, etc. Parasitic oscillations are very likely to occur when large tubes are employed because of the long leads, large interelectrode capacities, and relatively high values of transconductance involved. Parasitics cause reduction in the power at the desired mode of operation, introduce spurious frequencies, give rise to distortion in linear amplifiers and modulators, may produce spurious side bands, can cause flashovers, etc.

[1] The design of thermostats and heaters is discussed by James K. Clapp, Temperature Control for Frequency Standards, *Proc. I.R.E.*, Vol. 18, p. 2003, December, 1930; W. A. Marrison, Thermostat Design for Frequency Standard, *Proc. I.R.E.*, Vol. 16, p. 976, July, 1928; W. F. Diehl, A New Piezoelectric Quartz Crystal Holder with Thermal Compensation, *R.C.A. Rev.*, Vol. 1, p. 86, October, 1936.

Examples of Parasitic Oscillations.[1]—Parasitic oscillations can be classed as high-frequency, *i.e.*, higher in frequency than the desired mode of operation, low-frequency, and dynatron types.

In the high-frequency parasitic oscillations the leads between the tube and tank circuit serve as the inductance for the parasitic circuits, and the interelectrode capacities of the tube, sometimes augmented by the neutralizing condensers, act as the

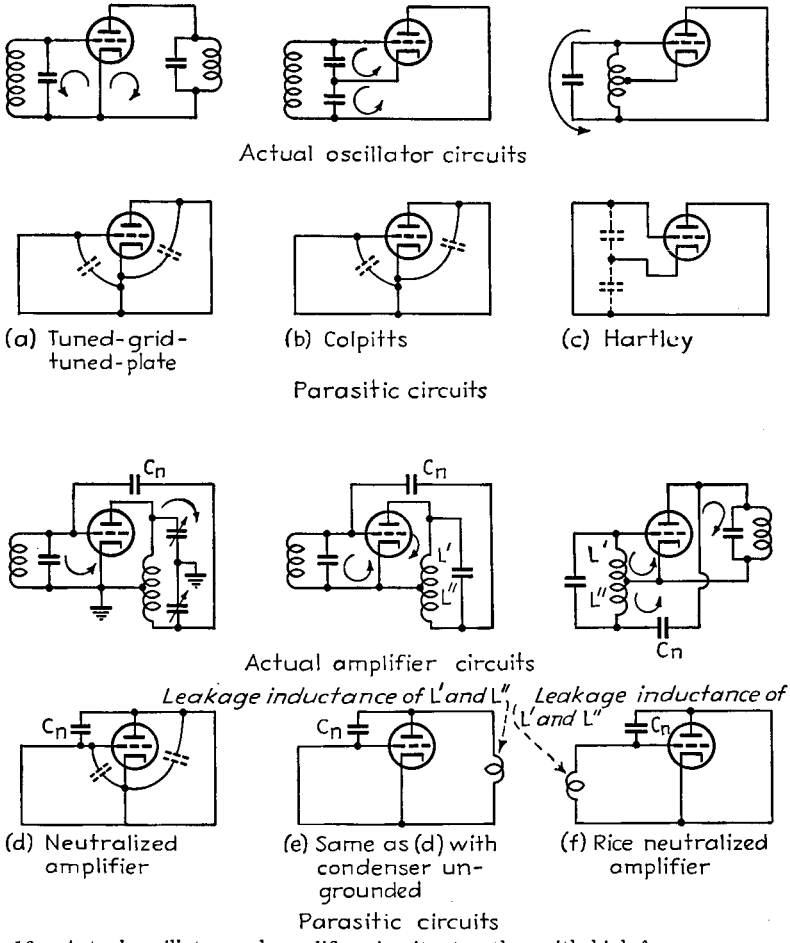


FIG. 16.—Actual oscillator and amplifier circuits, together with high-frequency parasitic circuits that can exist.

capacities. At these high frequencies the normal tank-circuit tuning capacity can be considered as a short circuit, while the tank-circuit inductance acts approximately as an open circuit.

Typical examples of such high-frequency parasitic oscillations are given in Fig. 16. Thus the tuned-grid tuned-plate oscillator circuit (which may be a crystal oscillator circuit) can have a high-frequency parasitic oscillation as shown, in which the induct-

[1] Additional examples of parasitics are given by G. W. Fyler, Parasites and Instability in Radio Transmitters, *Proc. I.R.E.*, Vol. 23, p. 985, September, 1935.

ances of the tuned circuits are supplied by the one-turn loops from the tube electrodes through the tuning capacities back to the cathode. The Colpitts circuit can likewise produce a high-frequency parasitic of the tuned-grid tuned-plate variety, as shown, while the Hartley oscillator circuit can give parasitic operation by acting as a Colpitts circuit with the inductance supplied by the one turn lead from grid through the tuning condenser and back to the plate electrode.

An ordinary Class C amplifier will act as a tuned-grid tuned-plate oscillator circuit, as shown at Fig. 16*d*. It will be noted that the neutralizing condenser not only fails

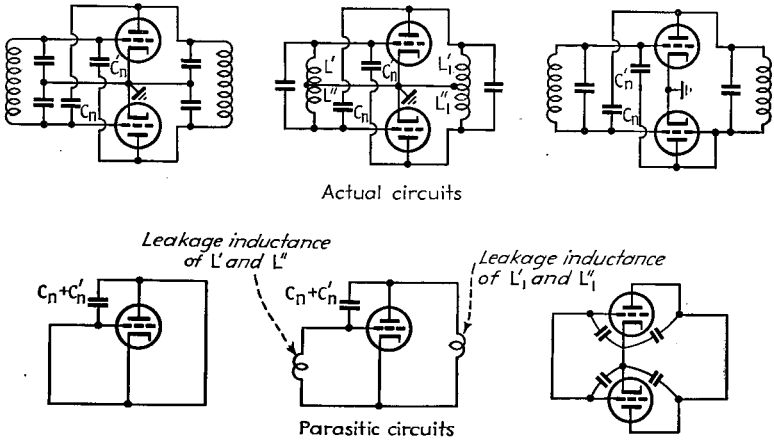


FIG. 17.—Push-pull amplifier circuits and high-frequency parasitic circuits that can exist.

to reduce the energy transfer through the grid-plate capacity of the tube but, as far as the parasitic oscillation is concerned, acts to increase the effective capacity between these electrodes. Even when a different system of neutralization is used, as, for example, in Fig. 16*f*, the circuit can still give a high-frequency parasitic by acting as a tuned-grid tuned-plate oscillator. A push-pull amplifier employing cross-neutralization, with either the coil or condenser center-grounded, has a tendency to develop oscillations of the tuned-grid tuned-plate type, as shown in Fig. 17, with the two tubes acting in parallel. Under these conditions, the neutralizing condensers do not prevent energy transfer from grid to plate, but actually act to increase this transfer. When neither coil nor condenser center is grounded, a high-frequency tuned-grid tuned-plate parasitic is possible unless the neutralization is maintained at the parasitic frequency. Similarly, a neutralized power amplifier employing tubes connected in parallel, as shown in Fig. 18, has a tendency to produce parasitic oscillations of the tuned-grid tuned-plate type, with the two tubes acting in push-pull.

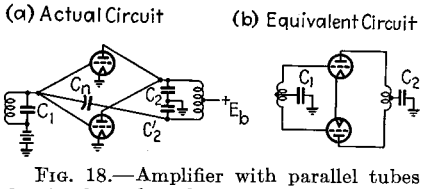


FIG. 18.—Amplifier with parallel tubes showing how the tubes may act in push-pull to produce a high-frequency parasitic.

Parasitic oscillations of relatively low frequency are caused by radio-frequency chokes. Examples are shown in Fig. 19. At *a*, it is seen that a push-pull power amplifier with grid and plate chokes can produce a low-frequency parasitic in which the grid and plate tuning inductances are the corresponding radio-frequency chokes and in which the tubes act in parallel in a tuned-grid tuned-plate circuit. It will be

noted that the neutralization is not effective in reducing this type of parasitic. It is shown at *b* how a Colpitts oscillator with grid and plate chokes can have a low-frequency parasitic.

During the portion of the cycle when the grid of a power oscillator or amplifier is positive, the grid-cathode circuit of some tubes will display negative resistance characteristics as a result of secondary electron emission at the grid. During this portion of the cycle, the negative grid-cathode resistance shunts any parasitic resonant circuit existing between grid and ground, and may introduce high-frequency oscillations of the dynatron type.

Elimination of Parasitic Oscillations.—Parasitic oscillations can be expected as a matter of course in any new design of power amplifier or power oscillator involving large tubes. In the case of Class C amplifiers, the simplest method of investigating the presence of parasitic oscillations of all types except those produced by dynatron action is to remove the exciting voltage, make the grid bias small or even zero, and

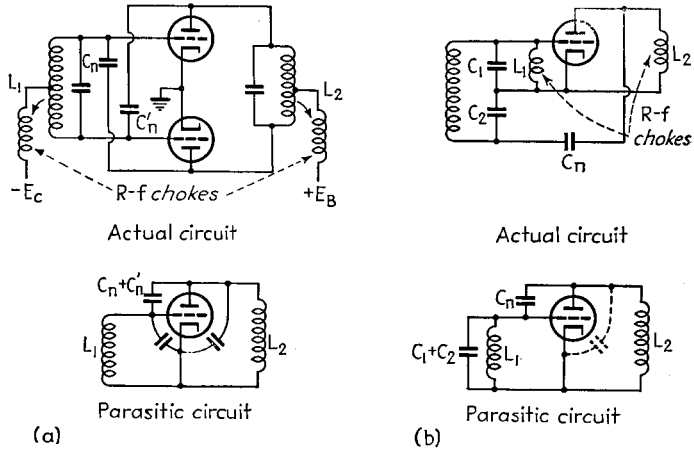


FIG. 19.—Amplifier circuits together with low-frequency parasitic circuits that can exist.

operate the tube at a lowered plate voltage so that the rated plate dissipation will not be exceeded. This ensures a high transconductance and so gives conditions favorable for the excitation of the parasitics, which can then be searched for by means of a neon tube on a stick or simply by drawing a spark to a wire on the end of a stick. If oscillations are present, their character is found by determining the points on the circuit that are at highest voltage and measuring the frequency. The equivalent circuit can then be deduced and remedial means devised. It is frequently found that upon elimination of one parasitic oscillation, another oscillation of a different type will appear, etc.

Various expedients are available for eliminating parasitics. Where the parasitics are of the tuned-grid tuned-plate type, the use of short grid leads and long plate leads, or the insertion of a small choke in the plate lead next to the tube, will cure the trouble. In this way the plate circuit is tuned to a lower resonant frequency than is the grid tank circuit, and oscillations cannot occur. The use of resistors connected in the grid and plate leads next to the tube is also effective, since these resistors are in series with the tank circuits of the parasitics. At the same time they are not in the tank circuit for normal operation, and so have little effect on the desired oscillations.

The use of neutralizing arrangements that maintain neutralization over wide frequency bands will eliminate certain types of parasitics. A schematic circuit of

an ordinary neutralizing system showing how the inductances of the leads will prevent neutralization from being obtained except at a single frequency is given in Fig. 20, together with two symmetrical arrangements in which the lead inductances are so arranged that the neutralization is effective over a wide frequency band.

Low-frequency parasitics normally arise from radio-frequency chokes, and can be eliminated by circuit arrangements that do not require chokes. Where it is not possible to eliminate all chokes, it is desirable to avoid the use of chokes in *both* the

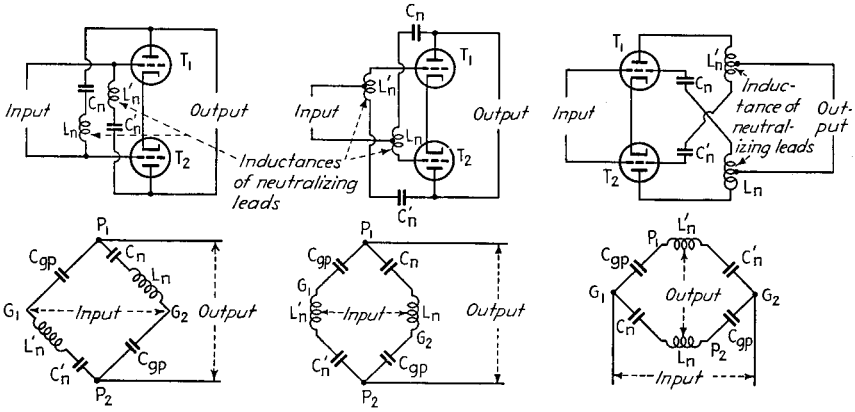


FIG. 20.—Neutralizing systems with push-pull tubes showing how bringing the input and output leads to the proper points will make the neutralization effective over very wide frequency bands.

grid and plate sides of the same tube. Where two chokes must be employed, low-frequency parasitics of the tuned-grid tuned-plate type can be avoided by arranging matters so that the resonant frequency produced in the plate circuit by the plate choke is less than the corresponding resonant frequency resulting from the grid choke.

Parasitics of the dynatron type can be eliminated by connecting a rectifier between grid and cathode, as illustrated in Fig. 21, and adjusting the resistance R in this rectifier so that the equivalent resistance formed by the parallel combination of rectifier and the negative tube resistance will always be positive.

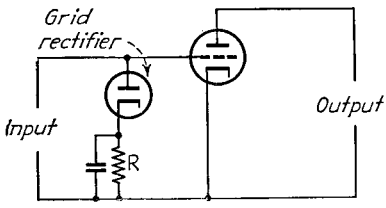


FIG. 21.—Grid rectifier for neutralizing dynatron effect at grid of tube.

Screen-grid and pentode power-amplifier tubes are practically immune from most types of parasitic oscillations, provided that the screen grid is really at ground potential, since then no energy transfer through the tube is possible. However, the circuit from the screen-grid electrode through the by-pass condenser to the cathode has inductance, and it is possible for parasitic oscillations of the tuned-grid tuned-plate type to exist in such tubes, with the screen grid serving the same function as the plate in the ordinary triode circuit.

Parasitic Oscillations in Audio Amplifiers.—Audio-frequency amplifiers, particularly those using large tubes, are frequently troubled with parasitics. Such amplifiers commonly employ triode tubes and are seldom neutralized. This, together with the fact that the input and output transformers are complicated networks involving

inductance and distributed capacities, introduces the possibility of a variety of parasitic oscillations.

6. Laboratory Oscillators. *Special Considerations in Laboratory Oscillators.*—In oscillators used in testing audio amplifiers, etc., the most important considerations are: (1) frequency stability with tube voltage changes, tube replacements and changes in load impedance, (2) good wave shape, and (3) constant output over a wide frequency range. The efficiency is only of secondary importance, since the power required ordinarily does not exceed a few watts, and is usually obtained from a power amplifier that isolates the load from the oscillator proper.

Ordinary oscillator circuits such as the Hartley, electron-coupled, etc., are used extensively in radio-frequency laboratory oscillators. However, many laboratory oscillators, particularly those used in testing audio and video amplifiers, are special types not used for other purposes. The more important of these are described below.

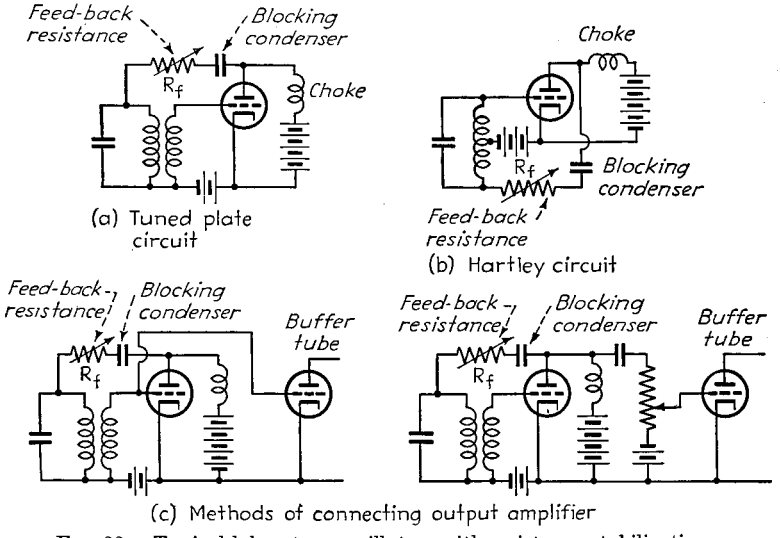


FIG. 22.—Typical laboratory oscillators with resistance stabilization.

Resistance-stabilized Oscillator.[1]—The resistance-stabilized oscillator is a conventional oscillator, usually of the tuned-plate or Hartley types, with the addition of a resistance between the plate and the tuned circuit as shown by R_f in Fig. 22. Such oscillators are widely used in laboratory work at audio and the lower radio frequencies. When properly designed, they are characterized by excellent wave shape, a frequency that is substantially independent of tube voltages and tube replacements, and an output that is substantially constant over wide frequency ranges.

The tube in a resistance-stabilized oscillator is adjusted to operate as a Class A amplifier with fixed bias (usually self-bias) and with the "feedback" resistance R_f in Fig. 22 so high that oscillations are just barely able to start. Upon the application of plate voltage, oscillations then build up until the grid goes positive at the peak of the cycle, when the added losses resulting from the grid current thus drawn prevent further increase in amplitude. For good wave shape, the tube, when considered as an amplifier, must be so adjusted that it will amplify without distortion an alternating voltage

[1] See F. E. Terman, Resistance Stabilized Oscillators, *Electronics*, Vol. 6, p. 190, July, 1933; J. W. Horton, Vacuum Tube Oscillators—A Graphical Method of Analysis, *Bell System Tech. Jour.*, Vol. 3, p. 508, July, 1924.

on the grid having a crest value several volts greater than the grid bias. This corresponds to a bias slightly less than would be used at the same plate voltage for normal Class A operation without grid current. The frequency generated will be substantially independent of the voltages applied to the tube and will not change with tube replacements, provided that the feedback resistance R_f is considerably greater than the plate resistance of the tube.

The design of a resistance-stabilized oscillator is built around the fact that when the impedance of the plate choke is much greater than the plate resistance, oscillations are just barely able to start when the feedback resistance has the value

$$\text{Starting feedback resistance} = R_L(\mu n - 1) - R_p \qquad (7)$$

where R_L = parallel resonant impedance of tuned circuit between feedback resistance and cathode.
μ = amplification factor of tube.
n = ratio of voltage transformation from plate to grid of coil of tuned circuit.
R_p = plate resistance of tube.

In order to obtain certainty of starting, it is desirable to make R_f 5 to 15 per cent less than given by Eq. (7). For good frequency stability the feedback resistance should

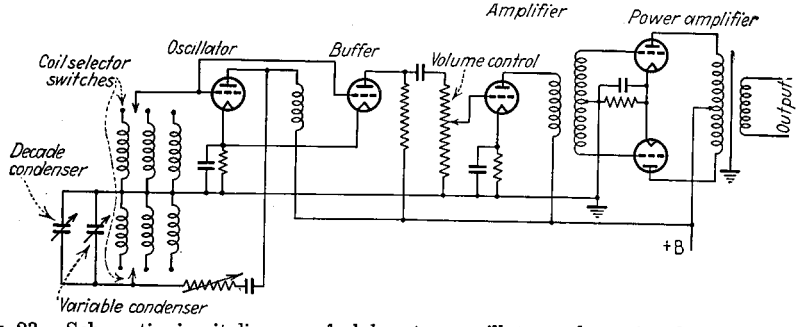


FIG. 23.—Schematic circuit diagram of a laboratory oscillator and associated power amplifier for covering the entire audio-frequency range.

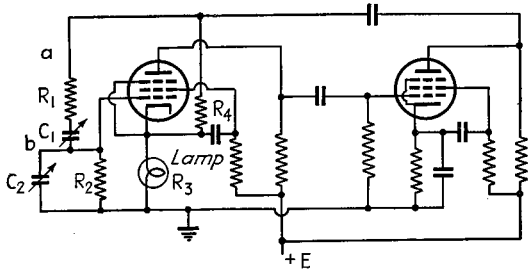
also be at least twice and preferably not less than five times the plate resistance of the tube, and at the same time is preferably not greater than 500,000 ohms. Most satisfactory designs are obtained with triode tubes having an amplification factor of the order of 5 to 8 and coil ratios n of unity. The grid bias should be at least 3 to 5 volts, and the plate potential should be so selected that with a signal having a crest value several volts greater than the bias, the tube will not approach too closely to cutoff on the negative peaks.

The circuit diagram of a complete audio-frequency oscillator of the resistance-stabilized type is shown in Fig. 23. Continuous variation of frequency is obtained by using a variable tuning condenser supplemented by a decade condenser, together with provision for switching coils in and out of the circuit. The feedback resistance is adjustable for each particular frequency to a value slightly less than that for which oscillations start. The proper setting of this resistance will depend upon the frequency and can be given on the frequency calibration chart, or the resistance can be adjusted to a value such that a microammeter in the grid circuit of the oscillator will show a small predetermined grid current.

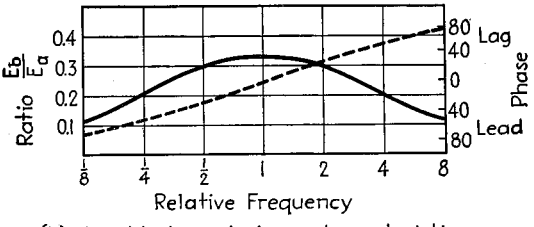
Oscillators with Resistance-capacity Tuning.[1]—In this type of oscillator, the frequency is determined by a resistance-capacity network that provides regenerative

[1] F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 10, p. 649, October, 1939;

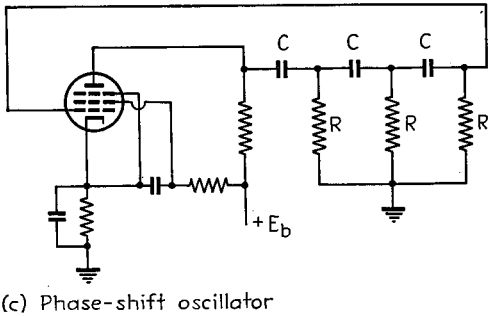
coupling between the output and input of a feedback amplifier, as shown in Fig. 24*a*. Here the network consists of $R_1C_1R_2C_2$, and by proportioning so that $R_1C_1 = R_2C_2$ under all conditions, the ratio of the voltage developed across the input (point *b*) of the amplifier to the voltage existing across the output (point *a*) varies with frequency as shown in Fig. 24*b*. Oscillations tend to take place at the maximum of this curve, which is at a frequency of $1/(2\pi\sqrt{R_1R_2C_1C_2})$ cycles per second.



(a) Circuit of resistance-capacity oscillator



(b) Amplitude and phase characteristics



(c) Phase-shift oscillator

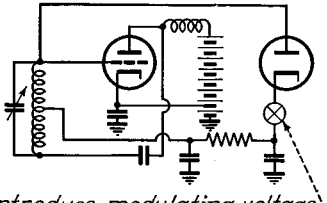
FIG. 24.—Circuit of resistance-capacity tuned oscillator and phase-shift oscillator.

For proper results the amplifier associated with the resistance-capacity network must have negligible phase shift, and provision must be made for controlling the amplitude of the oscillations. The phase shift can be made negligible down to very low frequencies by using as few by-pass condensers as possible, making these of very large capacity, and then employing a large amount of negative feedback. Phase shift at high frequencies can be minimized by designing the amplifier for wide-band response.

H. H. Scott, A New Type of Selective Circuit and Some Applications, *Proc. I.R.E.*, Vol 26, p. 226, February, 1938.

The amplitude can be most satisfactorily controlled by using a lamp to provide the feedback resistance R_3. When the amplitude of oscillations is small, the alternating current through R_3 is likewise small, making the resistance and hence the amount of feedback low. This increases the amplification and allows oscillations to build up. However, increased output increases the current through R_3, which increases the resistance and hence the amount of negative feedback, and reduces the tendency to oscillate. The result is an amplitude that can be made to stabilize at a value substantially constant under widely varying conditions and at a value that does not overload the tubes and distort the wave shape.

Resistance-capacity tuned oscillators have many advantages over beat frequency, resistance-stabilized, and other audio- and video-frequency oscillators used in general laboratory work. The frequency generated is inversely proportional to tuning capacity instead of inversely proportional to the square root of the tuning capacity, as in ordinary oscillators. A 10:1 frequency range can therefore be conveniently obtained from an ordinary variable condenser, with decimal multiplying factors available by changing the resistances R_1 and R_2 in decimal values. The wave shape is extremely good, frequency stability is high, and the output virtually constant over enormous frequency ranges. The oscillations will also readily synchronize in harmonic relation with voltages injected almost anywhere in the circuit, a useful property for certain applications.



(a) Simple amplitude control

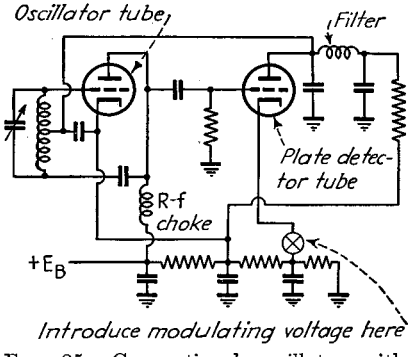


FIG. 25.—Conventional oscillator with amplitude control.

Phase-shift Oscillators.[1]—The phase-shift oscillator is a special type of resistance-capacity tuned oscillator that operates with a single tube. The basic circuit arrangement is illustrated in Fig. 24*c*. Here a three-mesh resistance-capacity phase-shifting network *RC* is connected between the output and input of an amplifier tube, with the circuit so proportioned that the total phase shift between plate and grid terminals is 180° at the frequency of oscillation desired. When the gain of such an amplifier is adjusted either manually or by an automatic-volume-control circuit barely to maintain oscillation, almost pure sine-wave output is obtained. The frequency stability of the circuit with respect to voltage changes is then also very good.

Phase-shift oscillators are particularly suitable for operation at fixed frequencies. and probably represent the simplest satisfactory audio oscillator for this purpose, Variable frequency operation can be obtained by simultaneously adjusting all condensers, or all resistances, from a common control, but the complications thus introduced are such that the arrangement of Fig. 24*a* is usually preferable for variable-frequency operation.

[1] Further information on phase-shift oscillators, including a discussion of modified circuit arrangements, design information, and a sample design, is given by E. L. Ginzton and L. M. Hollingsworth, Phase-shift Oscillators, *Proc. I.R.E.*, Vol. 29, p. 43, February, 1941.

Conventional Oscillators with Special Amplitude Control.[1]—Any conventional oscillator, such as the Hartley or Colpitts, can be made to generate waves with good wave shape and good frequency stability, provided that the amplifier tube is operated as a Class A amplifier, and some form of amplitude control is used to maintain the oscillations on the straight-line part of the tube characteristic.

An example of such an oscillator is shown in Fig. 25. This is an ordinary oscillator in which an automatic-volume-control arrangement operating from the oscillator output is used to increase the plate resistance as the oscillations build up. With proper circuit proportions, equilibrium will be reached with a small amplitude such that Class A operation is realized with no grid current. The result is excellent wave shape, good frequency stability, and substantially constant amplitude of oscillations as the frequency is varied. These advantages are realized to the greatest extent when the automatic volume control is of the type having both delay and amplification.

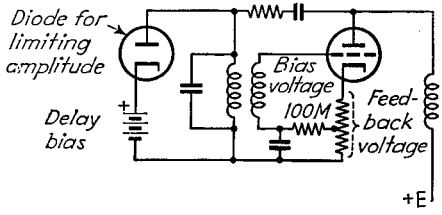


FIG. 26.—Resistance-stabilized oscillator employing feedback, provided with diode for limiting the amplitude.

The amplitude of oscillations can also be limited by means of a diode that is shunted across the tuned circuit, as in Fig. 26,[2] and then biased so that when the oscillations build up to the desired amplitude, current begins to be drawn. This places added load on the circuit and causes an equilibrium to be established. By the use of a large amount of feedback in the amplifier tube, substantially all distortion in the tube is eliminated, and at the same time the tendency to oscillate is stabilized to the point where the oscillations can always be depended upon to start, even when the margin of regeneration is very small. Oscillators of this type can be built to have distortion of the order of 0.01 per cent and less.

Beat-frequency Oscillators.—In the beat-frequency oscillator, voltages obtained from two radio-frequency oscillators operating at slightly different frequencies are combined and applied to a detector or converter tube, as shown schematically in Fig.

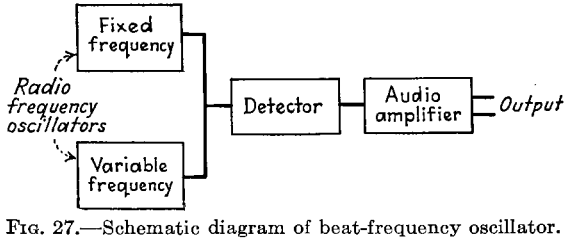


FIG. 27.—Schematic diagram of beat-frequency oscillator.

27. The difference-frequency current that is thus produced represents the output of the oscillator. The practical value of the beat-frequency oscillator arises from the fact that a small percentage variation in the frequency of one of the individual oscillators, such as can be obtained by the single turn of a dial controlling a variable condenser, will vary the "beat" or difference-frequency output continuously from a few cycles per second throughout the entire audio range, or, if desired, through the entire video-

[1] L. B. Arguimbau, An Oscillator Having a Linear Operating Characteristic, *Proc. I.R.E.*, Vol. 21, p. 14, January, 1933; Janusz Groszkowski, Oscillators with Automatic Control of the Threshold of Regeneration, *Proc. I.R.E.*, Vol. 22, p. 145, February, 1934.

[2] See Terman, Buss, Hewlett, and Cahill, *loc. cit.*

frequency range. At the same time, the oscillator output can be made substantially constant as the frequency is varied.

The principal factors affecting the performance of a beat-frequency oscillator are the frequency stability of the individual oscillators, the tendency of the oscillators to synchronize at very low difference frequencies, the wave shape of the difference-frequency output, and the tendency for spurious beat notes to be produced. Good frequency stability requires that the two oscillators be as nearly alike as possible, as to both electrical and thermal characteristics. It is particularly important that differential heating of the oscillator circuits be avoided and that heat-producing elements such as rectifier tubes, etc., be so located as to produce a minimum temperature change in the circuits of the oscillators. In this way, the difference frequency will be unaffected by variations in supply voltage and ambient temperature effects. It is then

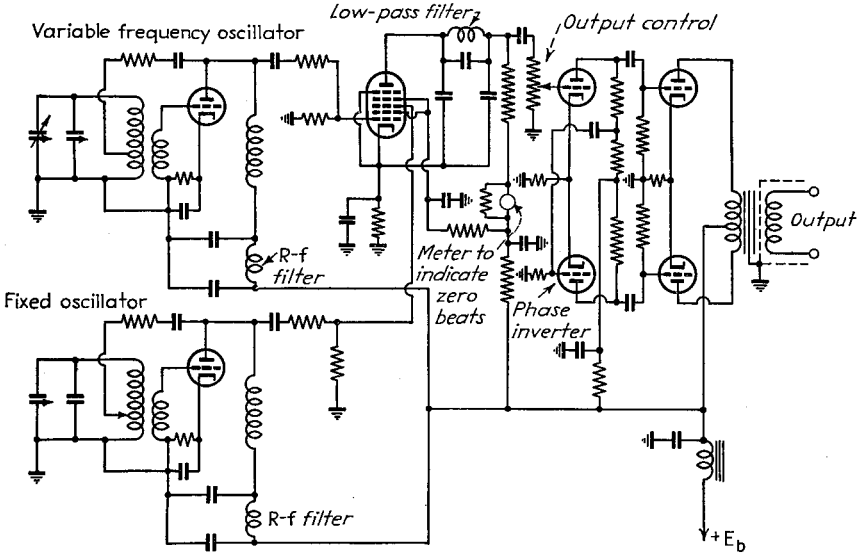


FIG. 28.—Complete circuit diagram of a typical beat-frequency oscillator.

merely necessary to set the frequency of the fixed oscillator so that the calibration is correct at one frequency on the dial as tested by a reed, beats with 60 cycles, etc., in order for the entire dial calibration to be correct. The usual practice is to make the frequency of the fixed oscillator of the order of five times the maximum beat-frequency desired.

With very low difference frequencies, the two oscillators tend to synchronize unless the circuits are properly shielded and isolated. This tendency is to be avoided, since it affects the frequency calibration, and tends to distort the output wave.

The wave shape of the difference-frequency voltages developed by the detector output is determined largely by distortion in the detector, if it is assumed that no troubles from synchronization are present. To ensure low distortion, one of the waves applied to the detector, preferably the one derived from the fixed oscillator, should be free of harmonics and be considerably smaller than the other wave.

Beat-frequency oscillators are commonly troubled with spurious beat notes, sometimes termed "whistles" and "birdies," when the difference frequency is large. These are usually the result of cross-modulation in the audio-frequency amplifier of high-order harmonics of the oscillator frequencies produced by the detector. These

spurious beats can be eliminated by operating the detector tube so as to minimize the production of harmonics, and by placing filters in the output of the detector to prevent the harmonics that do exist there from reaching the following tubes. It is also helpful to operate the audio-frequency stages following the detector with less exciting voltage than that giving full output, and to use Class A operation in these tubes, to minimize the tendency toward cross-modulation.

A typical beat-frequency oscillator circuit is shown in Fig. 28. This particular arrangement uses resistance-stabilized oscillators and a 6L7 mixer tube, although these and other details may be modified. In particular, other types of mixers may be employed, including balanced modulators, and other types of oscillators, notably electron-coupled oscillators, are commonly used.

Dynatron and Other Negative-resistance Two-terminal Oscillators.—Oscillations can be produced by shunting a tuned circuit with a negative resistance as illustrated schematically in Fig. 29*a*, and making the absolute magnitude of this resistance less than the

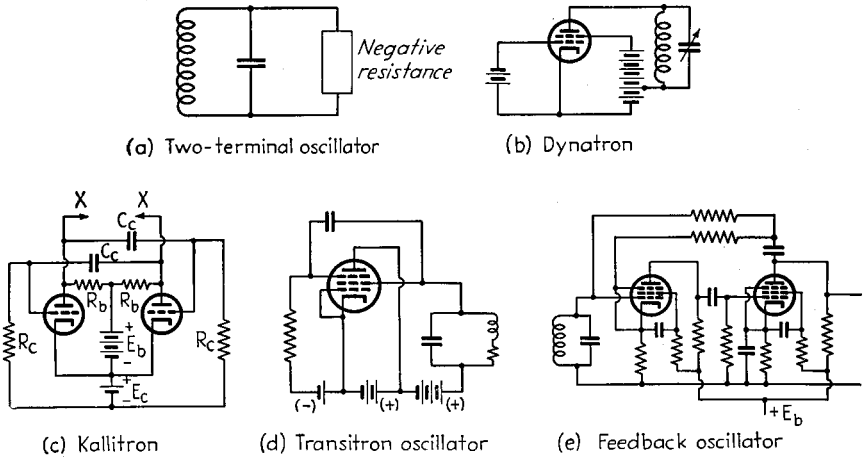


FIG. 29.—Miscellaneous types of two-terminal oscillators.

parallel resonant impedance of the circuit. The negative resistance may be any one of several types.[1] A common arrangement is the dynatron illustrated in Fig. 29*b*, in which the negative resistance exists between plate and cathode of a screen-grid tube as a result of secondary emission at the plate with screen voltages greater than the plate voltage. Another negative-resistance device is the kallitron circuit[2] of Fig. 29*c*, in which two tubes are connected as shown to produce a negative resistance between the terminals *xx*. A third negative-resistance oscillator circuit is shown in Fig. 29*d*, where the negative resistance is obtained by the retarding field or transitron connection.[3] Still another means of obtaining a two-terminal negative-resistance consists in coupling the output of a two-stage resistance-coupled amplifier back to the input, as shown in Fig. 29*e*. In this arrangement, there is considerable advantage in using negative feedback in the amplifier to stabilize the negative resistance.[4]

[1] In particular, see E. W. Herold, Negative Resistance and Devices for Obtaining It, *Proc. I.R.E.*, Vol. 23, p. 1201, October, 1935.

[2] Herbert J. Reich, A Low Distortion Audio-frequency Oscillator, *Proc. I.R.E.*, Vol. 25, p. 1387, November, 1937.

[3] See Herold, *loc. cit.*; also, Cledo Brunetti, The Transitron Oscillator, *Proc. I.R.E.*, Vol. 27, p. 88, February, 1939. See also Par. 14, Sec. 4.

[4] See Terman, Buss, Hewlett, and Cahill, *loc. cit.*

In all the two-terminal arrangements of Fig. 29, the frequency stability is maximum and the wave shape best when the adjustment is such that oscillations can barely exist, and have small amplitude. This desirable condition can be obtained by the use of automatic amplitude control, by some form of limiting device, or by manual control.

A different type of two-terminal oscillator is shown in Fig. 30.[1] Here if

$$R_1 = -R_n = \sqrt{\frac{L}{C}} \tag{8}$$

where R_n is a negative resistance, then the impedance of the circuit $AGGB$ is a resistance at all frequencies with a magnitude that depends upon frequency and which is negative for $\omega < 1/\sqrt{LC}$. By placing a variable resistance R_3 in series with the circuit as shown, the frequency of zero series impedance, and hence the frequency of oscillation, will then be controlled by the variable resistance R_3, which may be a potentiometer or the plate resistance of a tube.

7. Synchronization of Vacuum-tube Oscillators.[2]—Vacuum-tube oscillators tend to synchronize with an injected voltage of approximately the same frequency. This is a result of third-order curvature in the tube characteristics. The amount by which the oscillator frequency will deviate in order to synchronize with the injected oscillations

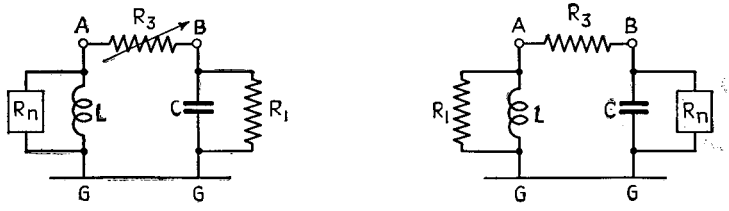


FIG. 30.—Schematic diagram of resistance-tuned oscillator.

increases with the amplitude of the injected voltage and as the frequency stability of the oscillator is reduced. When the frequencies of the oscillator and the injected voltages are approximately in simple harmonic relation, synchronization will tend to occur at the corresponding ratio of integers, although this tendency is less pronounced than when both frequencies are the same. In cases where the injected voltage is almost but not quite large enough to produce synchronization, the frequency of the oscillator is still shifted or "attracted" toward the frequency of the injected voltage.

The tendency toward synchronization can be greatly accentuated by the use of suitable auxiliary equipment to increase the factors causing synchronization.[3] The fundamental principle involved is illustrated in Fig. 31*a*. Here the combination of generated oscillation and injected voltage is applied to a detector, and the rectified output used to actuate a frequency-control system that operates on the oscillator frequency. If synchronization is assumed, the vector sum of the two voltages depends on their relative phase. Thus if the oscillator frequency should for some reason tend to become greater than the value corresponding to synchronization, the resulting advance in phase of the oscillator currents with respect to the injected voltage changes the vector sum and hence the rectified output of the detector, thereby operating the automatic-frequency-control system to decrease the frequency and so restore syn-

[1] W. G. Gordon and R. E. B. Makinson, Resistance-tuned Oscillators, *Wireless Eng.*, Vol. 14, p. 467 September, 1937; Sewell Cabot, Resistance Tuning, *Proc. I.R.E.*, Vol. 22, p. 709, June, 1934.

[2] E. V. Appleton, Automatic Synchronization of Triode Oscillators, *Proc. Cambridge Phil. Soc.* Vol. 21, p. 231; Isaac Koga, A New Frequency Transformer or Frequency Changer, *Proc. I.R.E.*, Vol. 15, p. 669, August, 1927; Balth. van der Pol, The Nonlinear Theory of Electric Oscillations, *Proc. I.R.E.*, Vol. 22, p. 1051, September, 1934;

[3] S. Sabaroff, Frequency Controlled Oscillators, *Communications*, Vol. 19, p. 7, February, 1939.

chronism. The method of Fig. 31*a* can be extended to maintain synchronism between two frequencies that are in harmonic or subharmonic relationship, as shown in Fig. 31*b*. In this way, it is possible to control the relative frequency of oscillator and injected voltage at ratios up to 10:1.

8. Oscillators Capable of Operating at More Than One Frequency.—Oscillators not infrequently have more than one possible frequency of oscillation. Parasitic oscillations (Par. 5) are an important example of this. Another example is the case where a tuned load circuit is coupled to the tank circuit to give a double-humped resonance curve corresponding to two possible frequencies of oscillation.[1]

When two possible oscillation frequencies do not differ greatly, oscillations will normally occur on one or the other frequency, according to whichever has the lowest losses, *i.e.*, whichever frequency corresponds to the lightest loading. If the tendency to oscillate at the two frequencies is equal, then it is a matter of chance which one

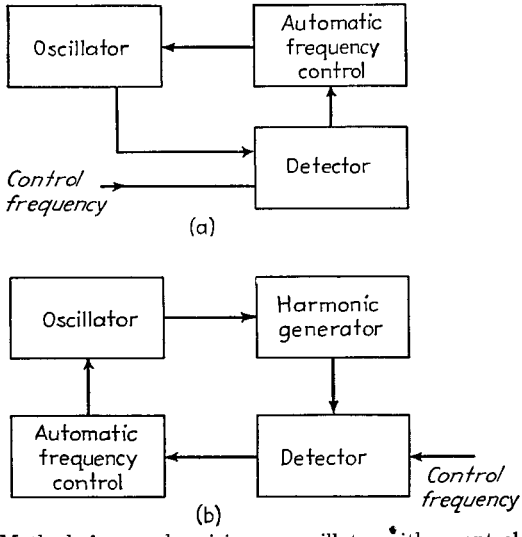


FIG. 31.—Methods for synchronizing an oscillator with a control frequency.

happens to get started, but when a particular frequency is once established, it will suppress the other and persist until interrupted, as, for example, by keying, when the frequency may flip to the other possible value the next time the oscillator is started.

When the possible frequencies of oscillation differ greatly, both tend to occur simultaneously, with the low-frequency oscillation modulating the high-frequency one.

9. Relaxation Oscillators.[2]—The term *relaxation oscillator* is applied to those oscillators in which the frequency is controlled by the charge or discharge of a condenser or inductance through a resistance. Such oscillators are characterized by highly distorted wave shapes and can be used to generate short pulses, saw-tooth, square, or triangular waves according to the detailed circuit arrangements and adjust-

[1] In order that there may be only one possible frequency of oscillation when the load is supplied by such a resonant circuit, it is necessary that the reactive volt-amperes circulating in the secondary circuit be less than the reactive volt-amperes in the primary (reactive volt-amperes = $\omega L \times$ (current squared)). This corresponds to a coefficient of coupling between the two resonant circuits that is less than the reciprocal of the effective Q of the secondary.

[2] For some general considerations relating to relaxation oscillators, see Balth. van der Pol, On Relaxation Oscillations, *Phil. Mag.*, vol. 2, p. 978, November, 1926; Ph. Le Corbeiller, The Non-linear Theory of the Maintenance of Oscillations, *Jour. I.E.E.*, Vol. 79, p. 361, 1936; also, Wireless Section, *I.E.E.*, Vol. 11, p. 292, September, 1936; Donald L. Herr, Oscillations in Certain Non-linear Driven Systems, *Proc. I.R.E.*, Vol. 27, p. 396, June, 1939.

ments. The frequency of relaxation oscillators is also readily controlled by voltages injected into the circuits.

Multivibrator.[1]—The multivibrator is a two-stage resistance-coupled amplifier in which the output of the second stage supplies the input to the first stage, as shown in Fig. 32*a*. Such an arrangement will oscillate because each tube introduces 180° phase shift. The oscillations produced by a multivibrator have a shape such as is shown in Fig. 32*b*. The frequency of oscillations is determined primarily by the time constants of the grid-leak grid-condenser arrangements, and is of the order of magnitude of $1/(R'_gC'_g + R''_gC''_g)$ cycles per second.[2]

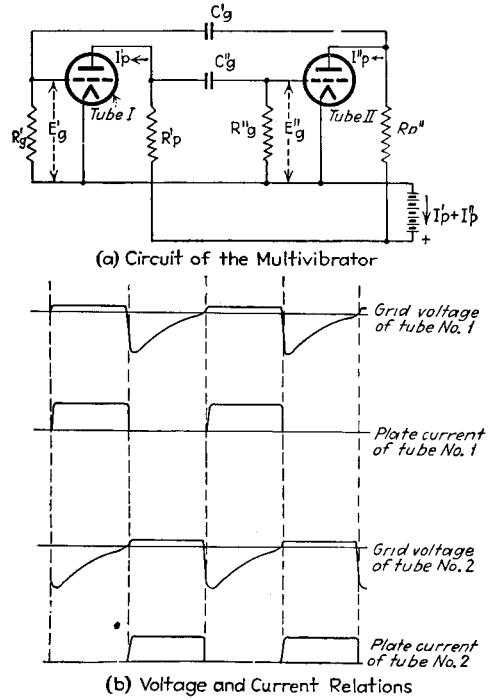


FIG. 32.—Circuit of multivibrator, together with oscillograms showing the way in which the instantaneous grid potential and plate currents vary during the cycle of operation.

The usefulness of the multivibrator arises from the fact that the wave is so irregular that harmonics up to at least the several hundredth are present, and that the frequency is readily controlled by an injected voltage.

Synchronization of Multivibrator with Injected Voltage.—When a voltage is introduced into the circuits of a multivibrator, the oscillations of the latter tend to increase in frequency until the ratio of the injected to multivibrator frequency can be exactly expressed by a ratio of integers.[3] This ratio may be unity, or greater or less than unity.

[1] H. Abraham and E. Block, Mésure en valeur absolué des périods oscillations électriques de haute fréquence, *Annal d. Phys.*, Vol. 12, p. 237, September–October, 1919.

[2] The frequency and pulse width can be calculated accurately by methods described by E. H. B. Bartelink, A Wide-band Square-wave Generator, *Trans. A.I.E.E.*, Vol., 60, p. 371, 1941.

[3] This stability with which the multivibrator oscillations hold in synchronization with an injected voltage in harmonic relation is greatly increased by returning the grid-leak resistances to a high positive bias. The frequency can also be conveniently controlled over a wide range by varying such a bias.

The most important application of the synchronizing possibilities of the multivibrator is in frequency division, *i.e.*, in controlling the multivibrator frequency so that it is $1/n$ of the injected frequency, where n is an integer.[1] When the uncontrolled frequency of the multivibrator is less than the frequency of the injected voltage, then as the amplitude of the injected voltage is progressively increased, the multivibrator

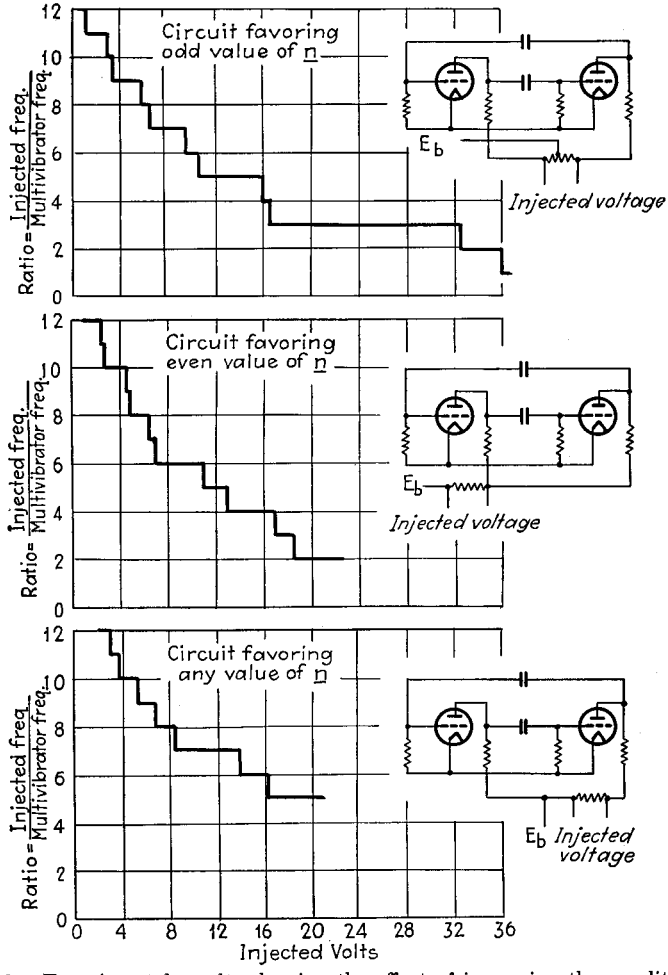


FIG. 33.—Experimental results showing the effect of increasing the amplitude of the synchronizing voltage in a multivibrator and how certain methods of injecting the voltage tend to favor either even or odd harmonics.

frequency will increase in discontinuous steps, *i.e.*, will be "drawn" toward the frequency of the injected voltage. The exact details of the behavior depend upon how the controlling voltage is introduced, and the extent to which the circuits of the two multivibrator tubes are dissymmetrical.[2] With perfect symmetry one can favor

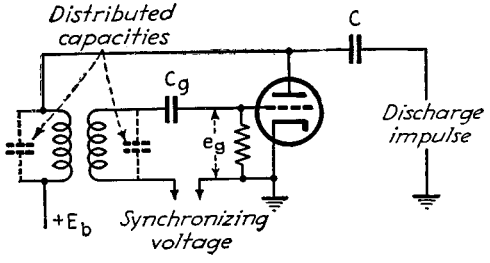
[1] Other methods of frequency division are referred to in Par. 8, Sec. 9.

[2] L. M. Hull and J. K. Clapp, A Convenient Method for Referring Secondary Frequency Standards to a Standard Time Interval, *Proc. I.R.E.*, Vol. 17, p. 252, February, 1929; F. E. Terman, "Measure-

either even- or odd-frequency ratios, or can show no discrimination, according to whether the injected voltage is introduced in the same or opposite phases in the two tubes, or is introduced only in one tube, respectively. This behavior is illustrated in Fig. 33.

By making the time constant of the grid-leak and grid-condenser combinations of the two multivibrator tubes differ, it is possible to make the tendency for synchronization greatest at some particular ratio m/n of integers.

Blocking Oscillators.—The blocking oscillator is widely used in television work to produce pulses that can be synchronized by a control voltage. The circuit of a blocking oscillator is shown in Fig. 34, and the arrangement can be thought of as an ordinary tuned-grid oscillating circuit having a very high ratio of inductance to capacity and very close coupling between the grid and plate circuits, together with



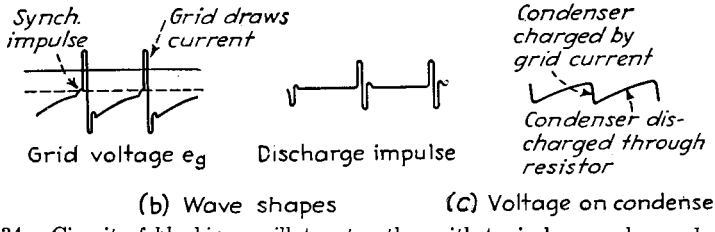


FIG. 34.—Circuit of blocking oscillator, together with typical wave shapes developed at various points.

an extremely high grid-leak resistance. Such an arrangement presents an extreme case of intermittent oscillations, since, as the oscillations start to build up, the grid leak-condenser combination develops a bias far beyond cutoff within a single cycle. Because of the circuit proportions, instability then occurs and the oscillations stop. This is because any tendency for the amplitude to decrease produces progressively greater tendency for the amplitude to become less, since the bias on the grid-leak-condenser combination cannot quickly readjust itself to the reduced amplitude, but rather causes the tube to act momentarily as though it had a fixed bias much greater than cutoff. After the oscillations cease, the grid-leak resistance slowly discharges the charge on the grid condenser until the bias becomes less than cutoff, at which point oscillations start to build up very rapidly, and the cycle repeats. The resulting voltages appearing at various places in the circuit have the wave shape shown in Fig. 34, and it is seen that pulses and saw-tooth waves are generated. The frequency

ments in Radio Engineering," pp. 129–136, 1st ed., McGraw-Hill, New York; Victor J. Andrew, The Adjustment of the Multivibrator for Frequency Division, *Proc. I.R.E.*, Vol. 19, p. 1911, November, 1931.

of the pulses can be controlled by making the uncontrolled frequency slightly less than the desired frequency, which is then injected as shown. For the generation of saw-tooth waves and pulses, the frequency at which the coils oscillate should be considerably higher than the frequency of the relaxation oscillations.

Miscellaneous Types of Relaxation Oscillators.—A large number of relaxation oscillator circuits have been devised. Some typical examples are shown in Fig. 35. In the arrangement *a*, the frequency is determined by the rate at which the condenser C discharges through a resistance corresponding to R_g and R_s in series.[1] The wave shape obtained depends upon the circuit constants and the place of observation. When the resistances external to the tube are large, the voltage at the plate (and also the control grid) is in the form of brief pulses as shown.

The single-tube circuit in Fig. 35*b* operates as follows:[2] When anode voltage is applied, the screen-grid potential rises until space current begins to flow, at which

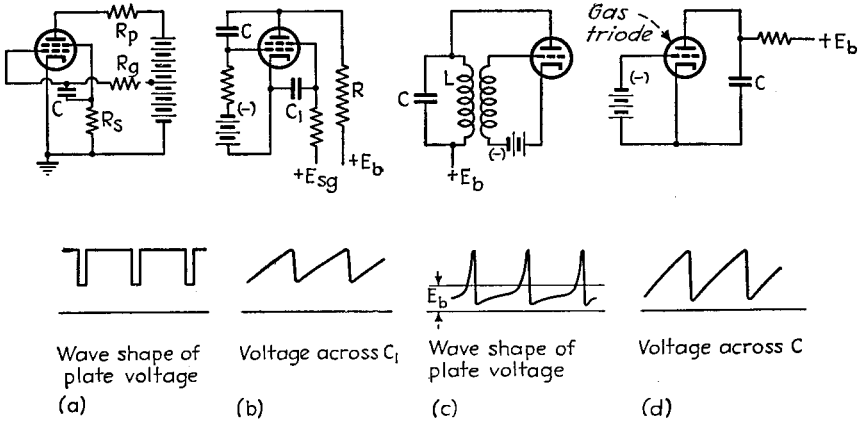


FIG. 35.—Miscellaneous types of relaxation oscillators, together with voltage wave shapes developed in circuits.

point the plate loses electrons due to dynatron action and so rises in potential as a result of the negative plate current flowing through the series resistance R. The plate is coupled to the control grid through condenser C, so that this rise in plate voltage increases the control-grid potential, increasing the space current and so causing a rapid discharge of condenser C_1 through the tube, followed by cutting off of the plate current and repetition of the cycle. The result is that the voltage across C_1 is a saw-tooth wave having very rapid flyback. Various modifications can be introduced, such as the use of pentode tubes to increase the output voltage obtainable.

Any ordinary oscillator tube with very high-impedance tank circuits and close coupling between grid and plate electrodes can be expected to produce relaxation oscillations under appropriate conditions. Thus the tuned-plate oscillator circuit of Fig. 35*c*, in which the tank circuit LC has much higher parallel impedance than the plate resistance of the tube and a Q that is not too low, will produce relaxation oscillations controlled by the rate of charge and discharge of inductance and capacities through the plate resistance of the tube.[3] A typical wave shape obtained with such

[1] R. M. Page and W. F. Curtis, The van der Pol Four-electrode Tube Relaxation Oscillation Circuit, *Proc. I.R.E.*, Vol. 18, p. 1921, November, 1930.

[2] D. H. Black, A New Hard Valve Relaxation Oscillator, *Elec. Comm.*, Vol. 18, p. 50, July, 1939

[3] F. Vecchiacchi, Oscillations in the Circuit of a Strongly Damped Triode, *Proc. I.R.E.*, Vol. 19, p. 856, May, 1931.

an arrangement is shown in the figure, and the oscillations readily synchronize with an injected voltage.

A gas triode (thyratron) tube, connected as shown in Fig. 35*d*, is commonly used to generate saw-tooth waves for sweep circuits of oscilloscopes. Here, when voltage is applied, condenser C charges up through R until the plate voltage is sufficient to cause plate current to start flowing. This gives ionization that causes the grid to lose control, with the result that the plate current becomes very large. The condenser C therefore discharges almost instantaneously to the point where there is not sufficient plate voltage to maintain ionization, whereupon the cycle repeats. The result is a voltage across C that is approximately of saw-tooth shape and that readily synchronizes with an alternating voltage applied to the grid of the gas triode. The highest frequency obtainable in this way is limited by the deionization time of the gas in the tube, and the tube design.

10. Polyphase Oscillators.[1]—It is possible to arrange n tubes in an electrical symmetrical order to obtain n phase oscillations. Examples of two- and three-phase

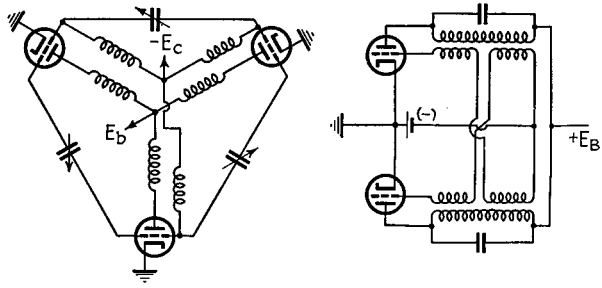


FIG. 36.—Polyphase oscillator circuits.

oscillator circuits are shown in Fig. 36. Such systems oscillate spontaneously in a polyphase manner, provided that there is reasonable symmetry between the tubes and electrical circuits associated with the individual phases and provided that the mutual inductances have been suitably chosen.

ULTRA-HIGH-FREQUENCY OSCILLATORS

11. Triode Oscillators for Very High Frequencies.—The characteristics desired in tubes to be used at ultra-high frequencies are: (1) smallest possible transit time of the electrons; (2) high transconductance in proportion to electrode capacities; (3) short leads of large diameter to minimize lead inductance and power losses in the leads; (4) electrode and lead arrangements that facilitate operation with resonant transmission lines, with the preferred arrangements those in which the leads and electrodes of the tube are as closely as possible extensions of the transmission-line system external to the tube.

Tubes especially designed for amplifier and oscillator operation at ultra-high frequencies are physically small, and have small clearances between electrodes. This is because it can be shown that if all the dimensions of a tube are reduced in the same proportion, the amplification factor, plate resistance, and transconductance are unchanged, while the electrode capacities, lead inductances, and transit time effects

[1] René Mesny, Generation of Polyphase Oscillations by Means of Electron Tubes, *Proc. I.R.E.*, Vol. 15, p. 471, August, 1925; I. Takao, Fundamental Considerations on Four-phase Oscillation Circuits, *Electrotech. Jour.*, Vol. 3. p. 75, April. 1939.

are proportional to the linear dimension.[1] However, these same factors that favor high-frequency operation also limit the power-dissipating ability of the tube, so that the amount of power that can be generated with triode tubes becomes progressively less as the frequency limit is pushed to higher values. Some increase in dissipating ability is obtained by employing cooling fins attached to the plate, and in some cases also to the grid. Water cooling is also employed with larger tubes. However, these expedients at best only help a bad situation.

Tubes especially designed for ultra-high-frequency use are frequently provided with more than one lead from the same electrode. These leads can be connected in parallel to give lowered lead inductance and loss, or they may be used as described in Par. 25, Sec. 5, so that the return currents from different electrodes will be separated. Double-lead arrangements that pass through opposite sides of the glass envelope also enable the tube to become the center portion of a continuous half-wave transmission line (see Fig. 37*e*).

Tank circuits in ultra-high-frequency oscillators are always either resonant transmission lines or cavity resonators. This is because of the high Q and high impedance that can be developed by such circuits, and because of their large physical dimensions in proportion to wave length at resonance. Furthermore, by designing the tube so that the inductance and capacity of the leads and electrodes are extensions of the distributed constants of the line, the inductance and capacity of the tube are less harmful than otherwise. This is because the same amount of total capacity and total inductance will resonate at a higher frequency when in distributed form than when constituting a lumped circuit. Furthermore, in arrangements such as illustrated in Fig. 37*e* only half of the tube inductance and capacity are chargeable to each quarter wave length of the transmission line. Finally, if tubes are so designed that their electrodes represent extensions of resonant-line tank circuits, it is possible to operate these resonant lines at higher order modes, such as three-quarters wave length. The first quarter-wave length section may then be entirely within the tube, and yet one can control the tuning by adjustments of the length of line external to the tube.

Circuits for Ultra-high-frequency Oscillators.—Typical circuits for ultra-high-frequency oscillators are shown in Fig. 37. In all these arrangements lines are used for obtaining resonant circuits.[2] In certain of the arrangements, the cathode is isolated from ground as far as radio-frequency potentials are concerned by cathode chokes (see Figs. 37*a*, *d*, *e*, *f*) or by a quarter-wave resonant line, as in Fig. 37*c*. In other cases, such as Fig. 37*b*, the cathode is brought to ground potential by associating it with a half-wave-length grounded transmission line. It is not possible to ground the cathode directly by means of a by-pass condenser, as at lower frequencies, because of the cathode lead reactance at very high frequencies. In order to obtain proper circuit operation, it is accordingly desirable either to employ a half-wave-length line or to isolate the cathode from ground by means of a choke or a quarter-wave-length line.

Two-tube arrangements such as shown in Figs. 37*c* and 37*f* have the advantage over single tube circuits of increased power output. They also preserve symmetry to ground, and thereby avoid the troubles that otherwise tend to be present from excessive unbalanced currents to ground. Double-tube arrangements also tend to have a higher upper frequency limit than single-tube circuits. This is particularly the

[1] This is the basis of the "Acorn" tubes. See B. J. Thompson and G. M. Rose, Jr., Vacuum Tubes of Extremely Small Dimensions for Use at Extremely High Frequencies, *Proc. I.R.E.*, Vol. 21, p. 1707, December, 1933.

[2] All the wave lengths indicated in the figure are the effective values, and include the loading produced by the capacity and inductance of the tube electrodes and leads. The actual length of line external to the tube will be somewhat less than indicated.

case in Fig. 37*e*, where only half of the interelectrode capacity is chargeable to each quarter-wave-length section of line.

12. Klystron (Velocity Modulation) Oscillators.[1]—A klystron, or velocity-modulation tube, as it is frequently called, has a structure such as shown in Fig. 38. Here heater cathode *A* with associated focusing electrode *C* produces an electron beam

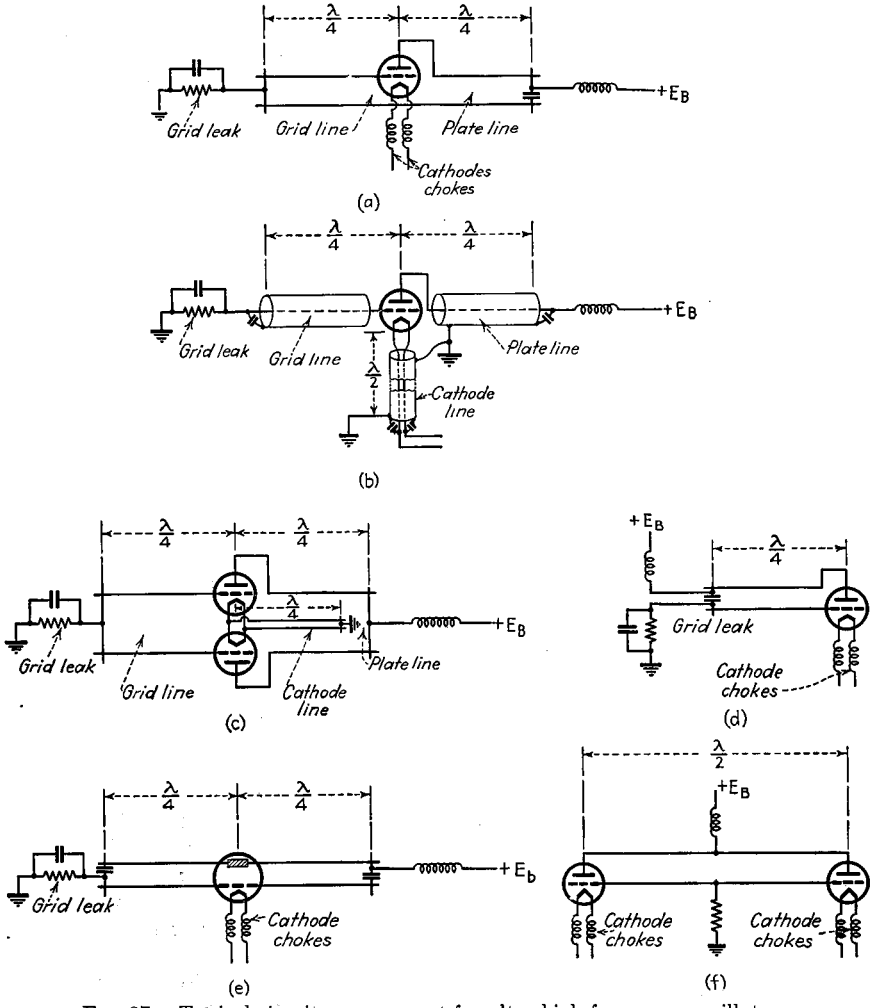


FIG. 37.—Typical circuit arrangement for ultra-high-frequency oscillators.

indicated by arrow *G*. This beam passes through grids *D* in the sides of a reentrant-type cavity resonator, termed the *buncher*, then through a tube termed the *drift tube*, which is at the same electrical potential as the buncher, and finally enters a second

[1] The klystron was invented at Stanford University by the Varian brothers. A good description of the fundamental considerations involved in such tubes is given in the paper by Russell H. Varian and Sigurd F. Varian, A High Frequency Oscillator and Amplifier, *Jour. Applied Phys.*, Vol. 10, p. 321, May, 1939. A similar and parallel development is described by W. C. Hahn and G. F. Metcalf, Velocity Modulated Tubes. *Proc. I.R.E.*, Vol. 27, p. 106, February, 1939.

cavity resonator termed the *catcher*, which is provided with a grid at *E*. The d-c potentials of the various electrodes are arranged as shown. The cathode and its associated focusing electrodes are maintained at a high negative potential with respect to the remaining part of the structure (all of which is at the same d-c potential). The entire arrangement illustrated in Fig. 38 is inclosed in a vacuum. In practical designs this is normally accomplished by providing a glass segment between the cathode and the remaining structure, as indicated schematically in the figure.

The operation of the klystron oscillator is explained as follows:[1] Assume first that oscillations have by some means been already established, so that the buncher cavity resonator is excited, and alternating fields exist between the surfaces *D* of the buncher grids. Then consider an electron that enters the buncher at the instant when this alternating field is zero but just becoming positive. This electron passes through between the two grids *D* of the buncher resonator and travels on through the drift tube and into the catcher resonator with unchanged velocity. Next consider an electron that arrives at the buncher slightly later. This electron finds that there is now a positive field between the grids *D* as a result of the oscillation within the buncher resonator. This later electron accordingly receives acceleration in passing through the buncher, and enters the drift with increased velocity and therefore tends to overtake the earlier electron. Similarly, an electron that arrives at the catcher slightly before the first or reference electron encounters a field between grids *D* that opposes its motion. This early electron hence enters the drift tube with reduced velocity, and tends to drop back and be overtaken by the reference electron. As a result of these actions, the electrons tend to bunch together as they travel down the drift tube, and this effect is maximum at certain distances down the tube, as shown in Fig. 39. The catcher of the klystron is located at such a distance, and accordingly the electrons enter the catcher in pulses, one pulse per cycle. These pulses of electrons correspond to the pulses of current that arrive at the plate of a Class C amplifier, and excite oscillations in the catcher resonator.

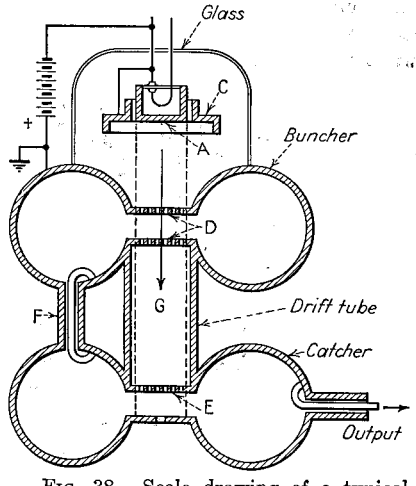


FIG. 38.—Scale drawing of a typical klystron oscillator designed to operate at about 13 centimeters wave length. (*After the Varians.*)

With proper adjustment the amount of power required to produce the bunching effect is relatively small compared with the amount of energy delivered by the electron beam to the catcher.[2] As a result the klystron tube is a power amplifier, and can be

[1] The mathematical theory of klystron tubes and of the bunching principle is given by David L. Webster, Cathode-ray Bunching, *Jour. Applied Phys.*, Vol. 10, p. 501, July, 1939; The Theory of Klystron Oscillations, *Jour. Applied Phys.*, Vol. 10, p. 864, December, 1939; Simon Ramo, The Electronic-wave Theory of Velocity-modulation Tubes, *Proc. I.R.E.*, Vol. 27, p. 757, December, 1939; R. Kompfner, Velocity Modulation, *Wireless Eng.*, Vol. 17, p. 478, November, 1940.

Graphical methods of analyzing bunching are given by D. Martineau Tombs, Velocity-modulated Beams, *Wireless Eng.*, Vol. 17, p. 54, February, p. 202, May, and p. 262, June, 1940; Rudolf Kompfner, Velocity-modulated Beams, *Wireless Eng.*, Vol. 17, p. 262, June, 1940. These two papers also contain many mathematical relations useful to one studying the details of these tubes.

[2] The energy delivered to the electron beam by the bunching effect is discussed by Webster, *loc. cit.*, and Rudolf Kompfner, Velocity Modulating Grids, *Wireless Eng.*, Vol. 19, p. 158, April, 1942. This energy is very small, and most of the required buncher power represents energy dissipated by the resistance losses of the buncher resonator.

used either as the equivalent of a Class C power amplifier or for the amplification of very small radio-frequency voltages. By providing coupling between the catcher and buncher resonators, as shown by F in Fig. 38, a small fraction of the energy developed in the catcher can be diverted to excite the buncher, thus giving an oscillator.

The efficiency of klystron oscillators or power amplifiers compares favorably with other oscillators for generating very high-frequency oscillations. Theory indicates that the maximum possible efficiency is 58 per cent.[1] Practical efficiencies are much less than this, although still substantially greater than obtained with electron oscillators of the Barkhausen type.

Klystron tubes are particularly suitable for use at very high frequencies. The tube dimensions are large in proportion to a wave length. Also, the physical structures involved are inherently capable of dissipating large amounts of heat, so that large powers can be handled. The first klystron oscillator ever built operated at a wave length of 13 centimeters and was physically as large as a power triode tube.

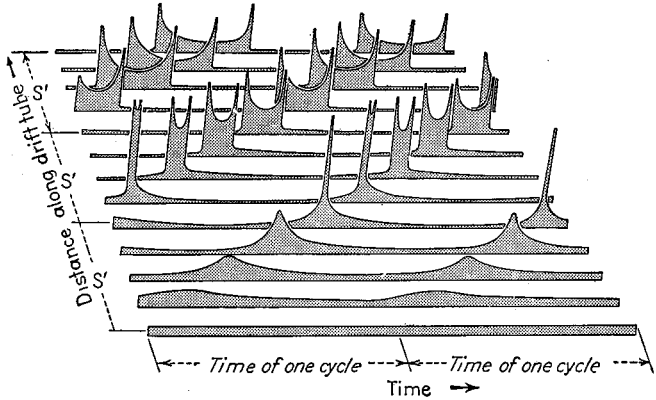


FIG. 39.—Diagram illustrating bunching of electron beam in klystron tube. In this diagram the reference electron is assumed to enter the drift tube at a time corresponding to the lower left-hand corner of the diagram. The width of the various lines represent the variation of electron density with time at various distances along the drift tube, with the distance S' representing the optimum distance for maximum bunching effect.

The klystron is suitable only for very high frequencies because of the impractically large physical sizes required by cavity resonators and the drift tube for lower frequencies.

The frequency generated by a klystron oscillator is controlled primarily by the resonant frequencies of the cavity resonators. These can be adjusted over limited ranges by flexing the sides and by the use of bellows. The resonant frequency can also be controlled to a slight extent by the use of electrical reaction on the resonators, as described in Par. 34, Sec. 3, and is affected by the anode voltage.

Proper operation of a klystron oscillator requires that the anode voltage, beam current, resonant frequencies of cavities, coupling between buncher and catcher (*i.e.*, ratio of buncher to catcher voltages), and also the coupling of the load to the catcher be carefully adjusted for optimum results. The klystron tube is far more critical in such respects than is the ordinary triode tube, because of the fact that the bunching in the electron beam must be maximum at the location of the catcher, and this bunching is affected by the anode voltage, the length of the drift tube, and the exciting voltage across the grids D of the buncher.

[1] Webster, *loc. cit.*

Klystron tubes can be used as detectors, and as power harmonic generators. Theoretical analysis indicates that in performing the latter function, it is theoretically quite feasible to obtain relatively high-order harmonic operation without great loss in anode efficiency or power output.[1]

13. Electron (Barkhausen) Oscillations in Triode Tubes.[2]—Oscillations of extremely high frequencies can be generated in a triode tube by operating the grid at a moderately high positive potential, and placing the plate at zero or a slightly negative potential. Under these conditions most of the electrons attracted by the grid pass between the meshes of the grid toward the plate, after which they slow down and return back toward the grid, to oscillate about the grid at a frequency determined by the electrode voltages and the distances involved. The resulting oscillations are termed *electron* or Barkhausen oscillations.

Mechanism of Operation.—In order for such electron oscillations to develop power, it is necessary that the electrons vibrating about the grid structure tend to do so in synchronism, rather than in purely random phase. The conditions of voltage, cathode emission, etc., which give rise to oscillations in an electron oscillator inherently provide for such synchronization. The mechanism involved can be described briefly as follows:[3] Assume that there is superimposed upon the positive grid voltage, an alternating voltage having a frequency corresponding approximately to the time taken by an electron to travel from cathode to plate. Consider now an electron that leaves the cathode at the instant that this superimposed alternating grid voltage is going through zero and is just beginning to be negative. This electron is attracted to the grid, but the superimposed alternating grid voltage is of such polarity during the transit from cathode to grid as to reduce the acceleration that the electron would otherwise experience. Furthermore, as the electron passes the grid plane, the polarity of the superimposed alternating grid voltage reverses, so that as the electron travels toward the plate, it is subjected to an increased braking force, and comes to rest before reaching the plate. At the instant that the electron starts back toward the grid, the superimposed voltage again reverses; so the return electron is subjected to less acceleration than corresponds to the d-c grid voltage. Upon reaching the grid plane, the superimposed alternating grid voltage reverses again, causing still more slowing down of the electron as it travels away from the grid toward the cathode. This process is repeated over and over again as the electron under consideration oscillates back and forth between the plate and cathode sides of the grid structure. The result is that this electron in its oscillations always works against the alternating voltage assumed superimposed on the d-c grid voltage, and therefore delivers energy to this oscillation. This process continues until the electron ultimately strikes the grid and is lost.

If now one considers a second electron that leaves the cathode a half cycle earlier, *i.e.*, when the alternating voltage on the grid is going through zero and just starting positive, the situation is quite different. The alternating voltage on the grid is of such polarity with respect to this electron as to cause it to accelerate more in traveling

[1] See Webster, *loc. cit.*

[2] The literature on electron oscillations in triode tubes is extensive, but is not in a too satisfactory state for the reader. An excellent survey of work published prior to 1933 is given by E. C. S. Megaw, Electronic Oscillations, *Jour. I.E.E.*, Vol. 72, p. 313, 1933; also Wireless Section, *I.E.E.*, Vol. 8, p. 59, June, 1933, which includes an extensive bibliography. Other references that are recommended are F. B. Llewellyn, "Electron Inertia Effects," Cambridge University Press, 1941, pp. 96–98; F. B. Llewellyn, The Barkhausen Oscillator, *Bell Lab. Rec.*, Vol. 13, p. 354, August, 1935; C. E. Fay and A. L. Samuel, Vacuum Tubes for Generating Frequencies above One Hundred Megacycles, *Proc. I.R.E.*, Vol. 23, p. 199, March, 1935; H. N. Kozanowski, A New Circuit for the Production of Ultra-short-wave Oscillations, *Proc. I.R.E.*, Vol. 20, p. 957, June, 1932; H. E. Hollmann, On the Mechanism of Electron Oscillations in a Triode, *Proc. I.R.E.*, Vol. 17, p. 229, February, 1929; William Alexander, The Retarding Field Oscillator, *Wireless Engr.*, Vol. 19, p. 143, April, 1942.

[3] This follows Llewellyn, *loc. cit.*

from cathode to grid than would be the case without this superimposed oscillation. Furthermore, as the electron passes through the grid toward the plate, the oscillation on the grid reverses, with the result that the braking force tending to slow down the electron as it travels from grid to plate is less by the amount of the superimposed alternating voltage. Accordingly, by the time this electron approaches the plate it is traveling faster than if only the d-c grid voltage were present. This extra velocity represents energy that has been abstracted from the superimposed oscillation. Because of the energy thus obtained from the superimposed oscillation, this electron may acquire sufficient velocity to reach the plate even if the plate is made slightly negative. Hence this particular electron, which is detrimental to the production of oscillations because it abstracts energy, tends to remove itself.

The end result is a sorting-out process, whereby those electrons which deliver energy to the oscillation on the grid oscillate about the grid more cycles than do those electrons which abstract energy. Thus there is a net resultant energy available to sustain oscillations, and there is a tendency for the majority of the electrons in the tube at any one time to be oscillating together.

While the above explanation gives a qualitative picture of the mechanism involved in electron oscillators, the details of the actual behavior are complicated by a number of factors. Among these are the tendency for space charge to form around the cathode and in the vicinity of the plate, the tendency of the grid to intercept the vibrating electrons, and a bunching effect similar to that in the klystron tube whereby electrons out of phase with the main oscillation tend to gain in phase and catch up with the useful electrons, and so finally assist in sustaining the oscillations. Also of fundamental importance is the fact that as a vibrating electron gives up more and more of its energy after successive oscillations, the excursions of its vibrations become progressively less. This causes the electron to drop back in phase (*i.e.*, to oscillate at a lower frequency) until the electron will, if not removed in time, finally get a half cycle out of synchronism and actually begin to absorb rather than to deliver energy.[1]

Properties of Electron Oscillators Using Triode Tubes.—The frequency of the electron oscillation is determined primarily by the dimensions of the tube and the grid potential, but the external tuned circuit, space charges, negative potential of the plate, etc., may alter the frequency by perhaps 30 to 50 per cent. With cylindrical electrodes and equal grid-cathode and grid-plate spacings, the wave length is given approximately by the relation

$$\text{Wave length in centimeters} = \frac{670d}{\sqrt{E_g}} \tag{9}$$

where d is the diameter of the plate in centimeters and E_g is the grid voltage. It is thus apparent that the higher the frequency to be generated the greater must be the grid potential and the smaller the dimensions of the tube.

Typical circuit arrangements suitable for use with electron oscillators are shown in Fig. 40. All these employ a resonant-line tank circuit. The arrangement *a* is the one most commonly employed, but that at *b* has been claimed[2] to give increased output and efficiency and together with the arrangement at *c* appears to be more logical.

The power output and the efficiency of an electron oscillator are quite critical with respect to circuit adjustments and electrode voltages on the tubes. Best results are usually obtained with the plate potential slightly negative and with the cathode emitting only a limited number of electrons so that conditions are intermediate between space-charge limited and temperature-limited space current. The efficiencies

[1] This tendency can be controlled by designing the grid so that it will on the average capture an oscillating electron by the time the phase has shifted sufficiently to cause absorption of energy.

[2] See Kozanowski, *loc. cit.*

obtainable are low because of the inherent nature of the generating process, with values of 1 to 3 per cent being typical for favorable conditions. The power output obtainable is also low because of the low efficiency, and because the grid that must absorb most of the power lost is necessarily small and is a type of structure not capable of dissipating large amounts of energy.

Most electron oscillators employ tubes having cylindrical grid and plate electrodes, but oscillations have been obtained with tubes having plane electrodes, and with tubes having more than one grid.

A given electron oscillator is capable of exciting a variety of types of oscillation. For example, oscillations of approximately twice the frequency given by Eq. (9) are frequently observed, provided that the tuning of the external circuit is suitably adjusted. Such oscillations can be considered as a second harmonic of the fundamental electronic oscillation, in the sense that the oscillation frequency actually generated is about twice the electron vibration frequency.

It is also possible to obtain electronic oscillations in which a resonance determined by the grid spiral plays an important part.[1] In such oscillations the behavior is

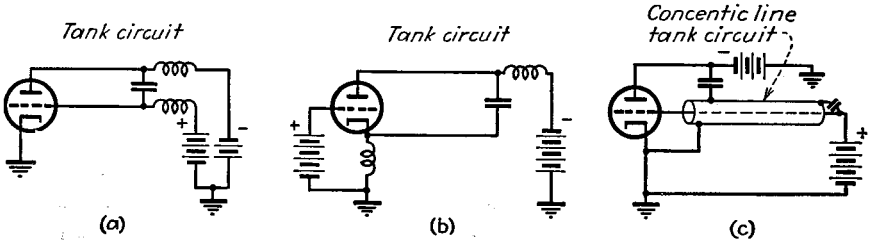


FIG. 40.—Possible circuits for electron oscillations of the Barkhausen type.

affected by the character of the grid spiral, as well as by the external resonant-line circuit and the electrode voltages, and in order to obtain optimum results for a given frequency, all these factors must be properly arranged.[2]

14. The Magnetron Oscillator.—There are two basic types of magnetron oscillators. The first of these involves electronic oscillations analogous to electron oscillations in a triode tube, in which the frequency is determined primarily by the transit time of electrons. The second type of magnetron oscillation is obtained by employing a split-anode arrangement, and taking advantage of the fact that under certain conditions a negative resistance can be obtained between the anodes.

Magnetron Oscillators of the Electronic Type.[3]—A typical oscillator of this type employs a cylindrical magnetron tube with a split anode, associated with a resonant

[1] See E. Pierret, *Compt. rend.*, Vol. 186, p. 1284, 1928.

[2] A good discussion is given by M. J. Kelly and A. L. Samuel, Vacuum Tubes as High-frequency Oscillators, *Trans. A.I.E.E.*, Vol. 53, p, 1504; also, *Bell System Tech. Jour.*, Vol. 14, p. 97, January, 1935.

[3] The literature related to magnetron oscillators of the electronic type is voluminous. It is, however, in a very unsatisfactory state, in that most of it consists of experimental observations for specific cases, together with some theory that is far from complete. This situation is an indication of the fact that the magnetron oscillator is not yet completely understood. The treatment given here represents merely a brief summary of the more important literature on the subject. It is suggested that the reader who wishes to pursue the subject further should consult first the paper by E. C. S. Megaw, An Investigation of the Magnetron Short-wave Oscillator, *Jour. I.E.E.*, Vol. 72, p. 326, 1933; also, Wireless Section, *I.E.E.*, Vol. 8, p. 72, June, 1933, which gives an excellent summary of the state of knowledge at the time of its publication. A later, very valuable summary paper is by G. Heller, The Magnetron as a Generator of Ultra-short Waves, *Philips Tech. Jour.*, Vol. 4, p. 189, July, 1939, which gives a very good account of the mechanisms of both the electronic and negative resistance oscillators. An important article on the operation of magnetrons having two and four anodes, in which the treatment is in terms of a rotating electric field resulting from the generated oscillations, is given by K. Posthumus, Oscillations in a Split Anode Magnetron, *Wireless Eng. and Exp. Wireless*, Vol. 12, p. 126, March, 1935. Other references that are suggested in addition to those mentioned elsewhere in this section are O. Groos, The Magnetic Field Transmitting Valve with Special Reference to the Decimetre Waves, *Elektrische Nachrichten*

transmission line, as shown in Fig. 41*a*, and with the magnetic field so adjusted as to exceed somewhat the critical value at which the anode current is cut off. In the absence of oscillations the electrons emitted from the cathode follow the paths shown dotted in Figs. 41*b* and 41*c*, and have an electron oscillation frequency corresponding to the transit time consumed in this path. As in the triode electron oscillator, there is a mechanism involved in the magnetron tube whereby these oscillating electrons tend to synchronize together.[1] Thus, assume that the magnetic field is just greater than cutoff, and postulate that an alternating voltage having a frequency equal to the frequency of the electron oscillations already exists between the anodes. Consider first an electron emitted as shown in Fig. 41*b*, with the instant of emission corresponding to the moment at which the alternating voltage on the upper plate is zero and just ready to become positive. This electron, instead of following the dotted path, as in the absence of a superimposed alternating voltage, follows the heavy path and reaches the upper anode because of the extra positive potential that this electrode

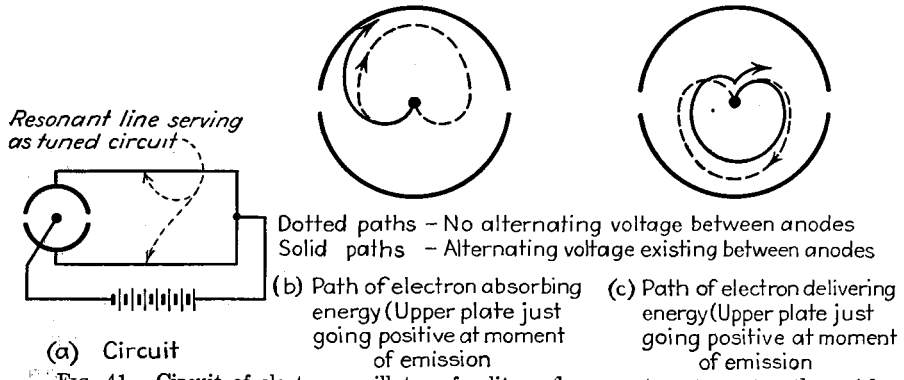


FIG. 41.—Circuit of electron oscillator of split-anode magnetron type, together with electron paths for typical conditions. The magnetic field is perpendicular to the plane of the paper.

possesses during the period that the electron is under the influence of the upper electrode. The added energy that this electron must acquire in order to reach the anode has been absorbed from the superimposed oscillation and so tends to damp it out. Next consider an electron emitted at the same instant as the first electron, but from the side of the cathode indicated in Fig. 41*c*. If there were no superimposed alternating voltage, this electron would follow the dotted path, but because of the alternating voltage between the anodes, the lower anode is less positive than it would otherwise be. The second electron is accordingly attracted less strongly, and hence follows a path such as is indicated by the solid line. This electron does not reach the anode immediately but circles about and comes to rest near the cathode, after which it starts another cycle of oscillations. This electron delivers energy to the superimposed oscillation, and in a complete cycle of oscillation delivers twice as much energy as is absorbed by the first electron. Furthermore, this useful electron is still available for more cycles of oscillation, whereas the first electron was removed at the end of the first half cycle. The end result is seen to be that those electrons that absorb energy from the superimposed oscillation tend to reach the anodes relatively soon and so are

Technik, Vol. 14, p. 325, 1937; G. R. Kilgore, Magnetostatic Oscillators for Generation of Ultra-short Waves, *Proc. I.R.E.*, Vol. 20, p. 1741, November, 1932; I. Wolff, E. G. Linder, and R. A. Braden, Transmission and Reception of Centimeter Waves, *Proc. I.R.E.*, Vol. 23, p. 11, January, 1935.

[1] This explanation was apparently first published by F. E. Terman, "Radio Engineering," 2d ed., p. 389, McGraw-Hill, New York, 1937.

removed before they can do much harm, whereas those which give up energy to the anode circuit, and hence tend to sustain the oscillations, remain and continue to deliver energy for a considerable length of time. Thus a net amount of energy is available to maintain the oscillation assumed to act between the anodes.

This picture of the operation of electron oscillators of the magnetron type should be considered as only qualitative. Thus space-charge effects and their modifying influences are ignored. Likewise, no account is taken of the bunching action that necessarily occurs because electrons emitted at different times from a particular part of the cathode suffer different accelerations during their travel, and therefore tend to bunch under certain conditions. Thus an electron that absorbs energy (the case of Fig. 41*b*), and does not quite reach the anode on the first half cycle of oscillation will advance in phase because of the increased velocity acquired. If this continues for several oscillations, this electron that has been absorbing power gains a half cycle, and is in a position where it becomes a useful electron, and starts delivering power to the external circuit.

The behavior of a magnetron oscillator is complicated by the fact that the electrons do not rotate through exactly 360 degrees around the cathode in making one cycle of oscillation. Accordingly, after a number of cycles have been traversed, an electron that started in such a way as to deliver energy to the superimposed voltage ultimately begins to absorb energy. It is hence necessary to remove the electrons from the tube after they have completed a number of oscillations. This can be accomplished by tilting the axis of the tube slightly with respect to the superimposed magnetic field, as illustrated in Fig. 42*a*.[1] Such a tilt gives the electrons a component of velocity parallel to the axis of the tube, causing them to spiral and ultimately either to pass beyond the ends of the anodes or to strike the anodes. Experiment indicates that the optimum tilt angle is of the order of 3 to 6 degrees in typical tubes. An alternative method of accomplishing the same result is to provide the tube with end plates, as illustrated in Fig. 42*b*.[2] These end plates are made positive with respect to the cathode, and so produce an axial field that causes the electrons to follow a spiral path leading to the end plates, where they are ultimately collected.[3]

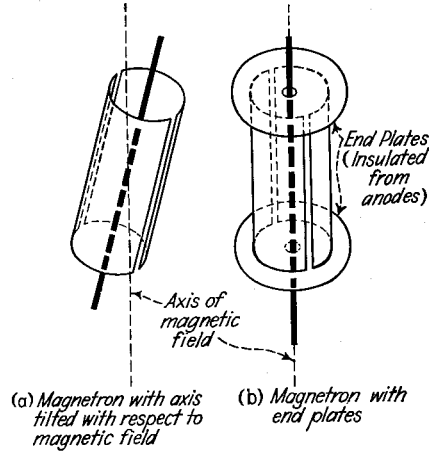


FIG. 42.—Means of causing electrons in a magnetron oscillator to spiral out of the interelectrode space.

The frequency of the oscillations generated by a magnetron oscillator depends upon the electron transit time. This is, however, influenced by the mode of operation of the oscillator, by the presence of space charge, the strength of the magnetic field, and by the design of the tube (as, for example, whether one, two, or four anode segments are employed). Another factor is the tuning of the external circuit, since the

[1] See Kilgore, *loc. cit.*

[2] Ernest G. Linder, Description and Characteristics of the End-plate Magnetron, *Proc. I.R.E.*, Vol. 24, p. 633, April, 1936.

[3] This explanation as to why a magnetron must either be tilted with respect to the magnetic field or be provided with end plates in order to function with maximum effectiveness was presented by the author in "Radio Engineering," 2d ed., p. 390, McGraw-Hill, New York, 1937. Calculations by F. B. Pidduck, Theory of Short Wave Oscillations with the Magnetron, *Wireless Eng.*, Vol. 18, p. 404, October, 1941, indicate that this is apparently the true explanation, and also indicate that on the average an electron should make five or ten oscillations before falling into the anode or an end plate.

magnitude and phase of the voltage developed in this circuit influence the oscillations of the electrons. The wave-length corresponding to the electron oscillations for a cylindrical structure with a magnetic field barely able to cut off the plate current under nonoscillating conditions is given by the empirical equation[1]

$$\lambda \cong \frac{13{,}000}{B} \tag{10}$$

where λ is in centimeters and B is in lines per square centimeter. The magnetic field B appearing in Eq. (10) is determined by the dimensions of the tube and the anode voltage according to the relation,

$$\text{in } B \text{ lines per sq cm} = \frac{6.72}{r}\sqrt{E} \tag{11}$$

where r is the anode radius in centimeters and E is the anode voltage. It is in general found that best results are obtained when the strength of the magnetic field is somewhat greater than the "cutoff" value given by Eq. (11), with a frequency correspondingly higher.

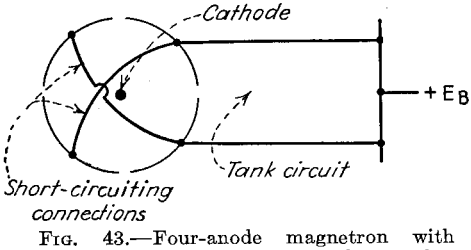


FIG. 43.—Four-anode magnetron with opposite electrodes short-circuited, together with associated transmission line.

Magnetron oscillators can be constructed in a variety of types. It is possible to obtain electron oscillations with the use of a single cylindrical anode, but best results are obtained with a split-anode arrangement, such as illustrated in Fig. 41, or with four anodes, as in Fig. 43, with the opposite anodes connected together. Magnetrons have also been built with more than four anode segments. Very high efficiencies and large power outputs have been claimed for a magnetron consisting of a single cylindrical anode with four quadrantal end plates at each open end of the anode, with the output being taken from the end plates.[2] The tank circuit of a magnetron oscillator is normally a resonant-line arrangement connected externally to the tube, but the line can be placed inside the tube for very high-frequency operation, and can, if desired, be made an integral part of the tube to aid in heat dissipation.[3] It has also been proposed that a cavity resonator be used with a magnetron, and this has been reported to result in a very large power output.[4]

For best operation the filament of a magnetron oscillator should be adjusted so that the current in the tube is emission-limited, or at least so that the space charge about the cathode is far from complete. It is necessary to pay particular attention to the cathode-heating current, because of the fact that some high-velocity electrons are returned from the anode region, and bombard the cathode. This raises the temperature and supplies at least part of the necessary heating power. Under certain operating conditions the energy dissipated in the cathode by such electronic bombardment is greater than the normal cathode heating power, with the result that a cumula-

[1] Theory shows that in the absence of space charge the constant in Eq. (10) is 12,300, while with space-charge saturation it is 16,700. See E. C. S. Megaw, Note on the Theory of the Magnetron Oscillator, *Proc. I.R.E.*, Vol. 21, p. 1749, December, 1933; J. Barton Hoag, A Note on the Theory of the Magnetron Oscillator, *Proc. I.R.E.*, Vol. 21, p. 1132, August, 1933.

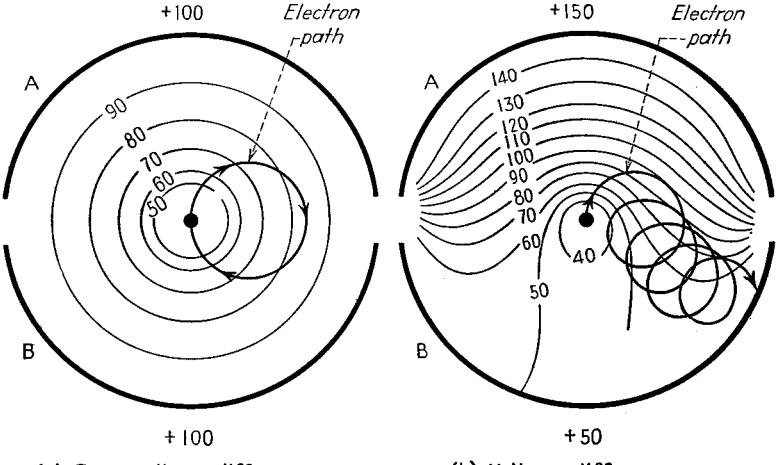
[2] Kinjiro Okabe, Electron-beam Magnetrons and Type-B Magnetron Oscillations, *Proc. I.R.E.*, Vol. 27, p. 24, January, 1939.

[3] E. G. Linder, The Anode-tank-circuit Magnetron, *Proc. I.R.E.*, Vol. 27, p. 732, November, 1939.

[4] K. Owaki, On the Magnetron with a Bowl Type Resonator, *Electrotech. Jour.*, Vol. 4, p. 188, August, 1940.

tive action takes place that results in the destruction of the cathode.[1] It is necessary either that care be taken to operate the magnetron under conditions of anode voltage and current such that this action does not take place or that there be an automatic regulating device that lowers anode voltage whenever the cathode temperature reaches an excessive value, as indicated by excessive emission.

Magnetron oscillators are capable of developing relatively large power output at reasonably good efficiencies at extremely high frequencies. Examination of Eq. (10) shows that very short wave lengths are entirely practical with a magnetic field that can be readily realized. Powers of the order of 100 to 200 watts at wave lengths of 50 to 80 centimeters have been reported with air-cooled tubes at efficiencies of the order of 40 to 70 per cent. Many investigators have obtained smaller amounts of power at much shorter wave lengths with efficiencies of the order of 5 to 25 per cent.



(a) Zero voltage difference between anodes

(b) Voltage difference between anodes 100%

FIG. 44.—Electron paths in split-anode magnetron. The contour lines represent equipotential lines, on the basis that the anode supply voltage is 100 (or 100 per cent).

Magnetron oscillators can operate in a number of modes, and the exact behavior is somewhat critical with respect to various details of circuit adjustments. Thus, if the magnetic field is considerably greater than that corresponding to cutoff as given in Eq. (11), there is a possibility of two or more oscillation frequencies being obtained.[2] These can usually be differentiated by suitable adjustment of the filament emission, but under some conditions several oscillations of different frequencies will be present simultaneously.

Magnetron Oscillators of the Negative Resistance or Dynatron Type.[3]—A split-anode magnetron operated with a combination of anode voltage and magnetic field such that the conditions correspond approximately to cutoff possesses a negative resistance to voltages applied symmetrically between the anodes. This arises from the fact that

[1] E. G. Linder, Excess Energy Electrons and Electron Motion in High-vacuum Tubes, *Proc. I.R.E.*, Vol. 26, p. 346, March, 1938.

[2] J. S. McPetrie and L. H. Ford, An Experimental Investigation of Resonance and Electronic Oscillations in Magnetrons, *Jour. I.E.E.*, Vol. 86, p. 283, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 27, March, 1940.

[3] G. R. Kilgore, Magnetron Oscillators for the Generation of Frequencies between 300 and 600 Megacycles, *Proc. I.R.E.*, Vol. 24, p. 1140, August, 1936; Hsu Chang, The Characteristics of the Negative-resistance Magnetron Oscillator, *Proc. I.R.E.*, Vol. 28, p. 519, November, 1940; A. F. Harvey, Output and Efficiency of the Split-anode Magnetron Oscillating in the Dynatron Regime, *Jour. I.E.E.*, Vol. 84, p. 683, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 161, June, 1939.

the electron paths under these conditions are as illustrated in Fig. 44. Many of the electrons that are attracted by the anode having the highest potential actually are deflected in the course of their travel in such a way as ultimately to arrive at the anode with less potential. This leads to a negative resistance effect, which can be utilized to produce oscillations by the circuit of Fig. 45.

By employing resonant circuits of high impedance it is possible to obtain very high efficiency in split-anode magnetrons of the negative resistance type. In order to achieve this result it is necessary that the magnetic field be considerably greater than cutoff, and that the amplitude of the oscillations across the tuned circuits be large. Then when the superimposed oscillation makes one electrode enough more positive than the battery voltage to exceed cutoff, the other anode (*i.e.*, the anode receiving the current) is at a much lower potential than the battery voltage. The losses in the tube are then small.

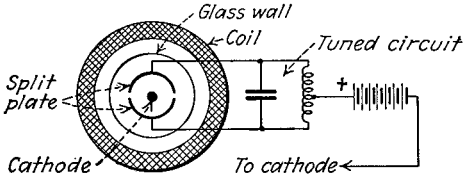


FIG. 45.—Circuit for generating oscillations of the negative-resistance type, with the use of a split-anode magnetron.

Magnetron oscillators of the negative-resistance type, unlike those of the electronic type, generate a frequency determined solely by the tank circuit. Oscillations from audio frequencies upward can be obtained with the upper limit occurring when the transit time of the electrons from emission to arrival at the anode becomes an appreciable part of the cycle. In practice, oscillations up to frequencies of 300 to 600 mc can be developed.

The efficiency of a magnetron oscillator of the negative-resistance type depends largely upon the design of the tank circuit. Under favorable conditions efficiencies of the order of 30 to 60 per cent are possible at low and medium frequencies. Because of these good efficiencies, combined with the fact that the anode structure is of a type capable of dissipating considerable heat, relatively large power outputs can be obtained.

15. Miscellaneous Ultra-high-frequency Tubes. *Diode Oscillator.*[1]—Transit-time effects in a diode tube are such that in certain frequency ranges that occur at very high frequencies, the diode offers a negative resistance between cathode and anode, as explained in Par. 9, Sec. 4. It is possible to utilize this negative resistance to give a negative-resistance type of oscillator similar to those of Par. 6, except that the frequency is very much higher, and the mechanism by which the negative resistance is obtained is different. This type of oscillator does not appear to be very practical, however, since the efficiencies are very low.

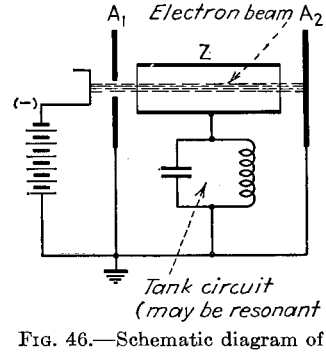


FIG. 46.—Schematic diagram of Heil tube.

The Heil Tube.[2]—The fundamental features involved in the construction of the Heil tube are shown in Fig. 46. A beam of electrons is produced from the cathode that passes through the hole in the anode cylinder A_1, transverses through a hollow

[1] F. B. Llewellyn and A. E. Bowen, The Production of Ultra-high-frequency Oscillations by Means of Diodes, *Bell System Tech. Jour.*, Vol. 18, p. 280, April, 1939; J. S. McPetrie, A Diode for Ultra-H-F Oscillations, *Exp. Wireless and Wireless Eng.*, Vol. 11, p. 118, March, 1934.

[2] A conveniently accessible description of this tube is given by H. E. Hollmann, Theoretical and Experimental Investigation of Electron Motions in Alternating Fields with the Aid of Ballistic Models, *Proc. I.R.E.*, Vol. 29, p. 70, February, 1941. The original paper describing the tube is by A. Heil and O. Heil, *Z. Physik*, Vol. 93, p. 752, 1935.

cylinder Z, after which it is intercepted by a catching anode A_2. The cylinder Z is connected to the anodes through a high-impedance parallel resonant circuit that represents the tank circuit of the oscillator. It can be shown that under certain conditions of electron beam velocity and tube proportions, the interchange of energy between the cylinder Z and the electrons entering and leaving it is such as to result in the delivery of energy to the cylinder at certain very high frequencies. This comes about as a result of a bunching effect that takes place within the cylinder Z, which can be utilized in such a way that the electrons emerging from the cylinder and traveling to the anode A_2 have a distribution with time different from the substantially uniform distribution with time of the electrons entering the cylinder.

Inductive-output Tube.[1]—The essential features of an inductive-output tube are shown in Fig. 47. Here there is a cathode provided with a control electrode and

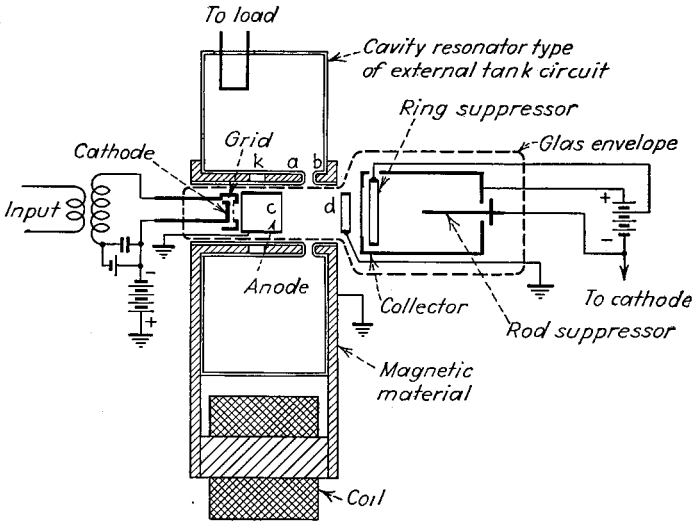


FIG. 47.—Cross-sectional view of inductive-output tube. The shaded area represents magnetic material that, when acted upon by current passing through the coil, produces a flux distribution having a focusing effect on the electron beam that is helpful in preventing the beam from striking the glass walls of the tube.

whatever focusing electrodes are required to produce an electron beam that travels through the anode cylinder C and onward toward the collector electrode. The cathode with its control grid and focusing structure is operated at a high negative potential with respect to the anode electrode C, so that the beam that emerges from C has a high velocity. The number of electrons in this beam is controlled by the grid in the usual manner, the grid electrode performing exactly the same function as the grid of an ordinary triode tube. Thus pulses of electrons travel down the tube at high velocity. These pass by a cavity resonator with cylindrical symmetry, provided with an air gap ab, as shown, that enables the pulses of electrons passing this air gap to induce oscillations in the cavity resonator. The electrons then pass on to the collector electrode, which is operated at a potential moderately positive with respect to the cathode (but much less positive than the anode C). In order to prevent secondary electrons produced at the collector from traveling back down the tube toward electrode C, a suppressing field must be established near the entrance of the collector. One

[1] A. V. Haeff and L. S. Nergaard, A Wide-band Inductive-output Amplifier, *Proc. I.R.E.*, Vol. 28 p. 126, March, 1940.

method of doing this is illustrated in Fig. 47. Here a ring suppressor electrode is placed just inside the entrance to the collector, and a rod at cathode potential is placed along the axis of the collector as shown.

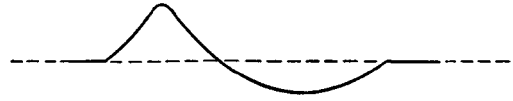
The inductive-output tube, as described in the literature, is suitable for operation at frequencies up to the order of 500 mc, and is capable of giving good efficiency under these conditions. It can be used as a power amplifier, or the output can be coupled back to the input to give oscillator operation. The effectiveness of the inductive-output tube is due in large measure to three factors. First, the inductive coupling arrangement makes it possible to employ a cavity-resonator plate tank circuit. Second, the tube arrangement permits high-velocity electrons to be used, since anode voltages of the order of three or four thousand are feasible. This reduces transit-time troubles that would exist in the corresponding triode tube. Third, the use of a separate collector electrode makes it possible to obtain high efficiency even though the tank circuit produces retarding fields that are only a small fraction of the potential of the beam. This is because the collector electrode can be adjusted to a potential such that the electrons, after passing the gap *ab*, are slowed down and are collected at a relatively low velocity, with correspondingly small energy dissipation.

SECTION 7

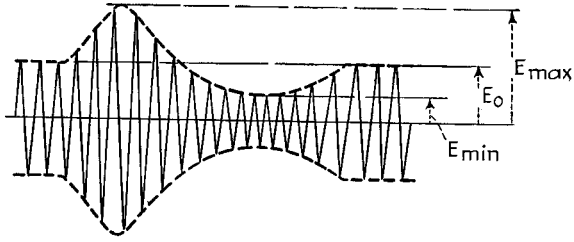
MODULATION AND DEMODULATION

AMPLITUDE MODULATION

1. Composition of Amplitude-modulated Wave.—In amplitude modulation the envelope of the radio-frequency oscillation is varied in accordance with the intelligence being transmitted, as shown in Fig. 1.



(a) Intelligence to be transmitted



(b) Wave modulated by (a)

FIG. 1.—Transmission of intelligence by an amplitude-modulated wave.

The extent of the amplitude variation in an amplitude-modulated wave is expressed in terms of the degree of modulation

For positive peaks (upward modulation):

$$\text{Degree of modulation } m = \frac{E_{\max} - E_0}{E_0} \tag{1}$$

For negative modulation peaks or troughs (downward modulation):

$$\text{Degree of modulation } m = \frac{E_0 - E_{\min}}{E_0} \tag{2}$$

where E_0, $E_{\max}$, and $E_{\min}$ are the average, maximum, and minimum envelope amplitudes, respectively, as indicated in Fig. 1.

With sinusoidal modulation, the positive and negative peak modulations will be the same, but, in general, this is not the case. When the degree of modulation is unity, the wave is said to be completely modulated.

It will be noted that the degree of modulation can never exceed unity on the negative peaks, since the envelope amplitude can never be less than zero. The modulation can, however, exceed unity on the positive peaks, a condition referred to as *overmodulation.*

Analysis of Modulated Wave.[1]—In the case of simple sinusoidal modulation, the equation of the modulated wave can be written as

$$e = E_0(1 + m \sin 2\pi f_s t) \sin 2\pi f t \tag{3}$$

where e = instantaneous amplitude of the wave.
E_0 = average amplitude of the wave.
m = degree of modulation.
f_s = modulating (or signal) frequency.
f = radio (or carrier) frequency.

Equation (3) can be rewritten as the sum of three waves as follows:

$$e = E_0 \sin 2\pi f t + \frac{mE_0}{2} \cos 2\pi(f - f_s)t - \frac{mE_0}{2} \cos 2\pi(f + f_s)t \tag{4}$$

The first component of the modulated wave in Eq. (4) is termed the carrier, and is the same irrespective of the degree of modulation or the modulation frequency. The second and third components are termed *side-band frequencies*, and are the part of the modulated wave that represents the intelligence.

In the general case of nonsinusoidal modulation, the modulation envelope can be expressed as either a Fourier series or a Fourier integral, according to whether the modulation envelope is a periodic or nonperiodic function of time. In either case, there is a carrier having an amplitude equal to the average amplitude of the modulated wave. In addition, there are a pair of side-band components symmetrically arranged about the carrier for each frequency component in the modulation envelope. Each such side-band component has an amplitude that is one-half the amplitude of the corresponding frequency contained in the equation of the modulation envelope, and the frequencies of the pair are greater and less than the carrier frequency by the corresponding modulation frequency component.

The power of a modulated wave is divided between the carrier and the side-band components. The carrier power remains constant, whereas the side-band power varies in accordance with the character and degree of modulation. In the extreme case of complete sinusoidal modulation, the side-band power is 50 per cent of the carrier power, and in other cases is usually less.

Side Bands Required in Telegraph, Telephone, and Picture Transmission.—In the transmission of telegraph signals by the Continental Morse Code, it is theoretically possible to operate telegraph relays, provided that each side band has a width of approximately 0.13 cycles per second for each letter transmitted per minute. In the case of a five-element code such as is employed in printing telegraph systems, the side band need only be three-quarters as wide as with the Continental Morse Code. By employing a synchronous vibrating relay to restore the shape of the received signals transmitted by the printing telegraph code, it is possible to reduce the band width to only 0.05 cycles per letter transmitted per minute.[2] Under the practical conditions usually existing in radio systems, it is always desirable to employ bands much wider than these theoretical minimums.

[1] R. V. L. Hartley, Relations of Carrier and Side-bands in Radio Transmission, *Proc. I.R.E.*, Vol. 11, p. 34, 1923.

[2] A discussion of band-width requirements of telegraph signals is given by F. E. Terman, Some Possibilities of Intelligence Transmission When Using a Limited Band of Frequencies, *Proc. I.R.E.* Vol. 8, p. 167, January, 1930.

Speech of quality corresponding to that commonly existing on long-distance telephone circuits can be obtained by reproducing modulation frequencies from 250 to 2,750 cycles. Speech and music having a quality corresponding to that realized under typical broadcast receiving conditions can be obtained by reproducing modulation frequencies from 100 to 4,500 cycles, while practically perfect reproduction of audible sounds requires a frequency band of approximately 40 to 15,000 cycles. In picture transmission (including television), the maximum modulation frequency required to realize full definition is approximately[1]

$$\left.\begin{matrix}\text{Maximum frequency that} \\ \text{must be transmitted}\end{matrix}\right\} = \frac{1}{2}\frac{w}{h}n^2N \tag{4}$$

where w and h represent the ratio of width to height, n represents the number of lines, and N equals the number of pictures transmitted per second. This equation assumes

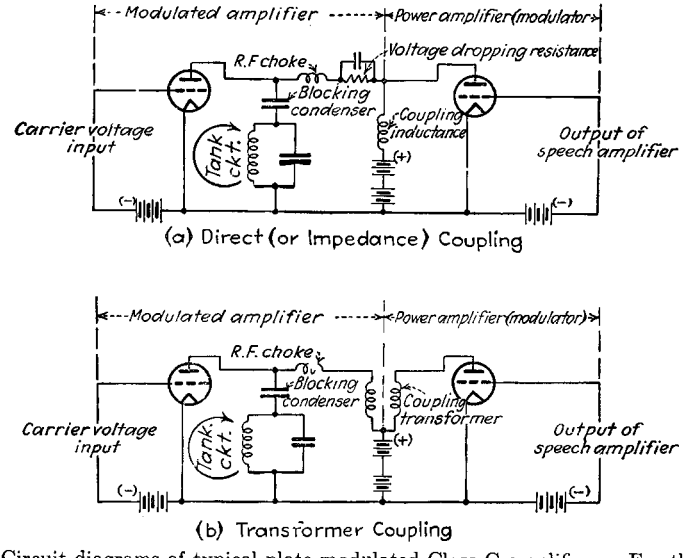


FIG. 2.—Circuit diagrams of typical plate-modulated Class C amplifiers. For the sake of simplicity the neutralizing arrangement for the modulated amplifier is omitted.

that horizontal scanning lines are used, and that the vertical and horizontal resolutions are approximately the same. The lowest frequency that must be reproduced in a picture transmission system is Nk, where k is the number of times the lines are interlaced. In the case of a standard system with 30 frames per second and alternate lines interlaced, this lowest frequency is 60 cycles.

2. Plate-modulated Class C Amplifier.[2]—The plate-modulated Class C amplifier is an ordinary Class C amplifier in which the modulating voltage is superimposed upon the d-c plate-supply voltage. This makes the total effective plate voltage consist of the sum of the d-c plate-supply voltage and the modulating voltage, and so correspond to the desired modulation envelope. Typical circuit arrangements for accomplishing this are shown in Fig. 2.

[1] For further discussion of the band width required to transmit a television signal, see Chap. 6 of V. K. Zworykin and G. A. Morton, "Television," Wiley, New York, 1940.

[2] Plate modulation is frequently called Heising modulation, after its inventor, and is also sometimes referred to as constant-current modulation.

The Class C amplifier in a plate-modulated arrangement has the exciting voltage, grid bias, and tank circuit proportioned so that with a plate voltage equal to twice the actual d-c plate-supply voltage, the tube is operating as a conventional Class C amplifier with good plate efficiency, and with a power output corresponding to four times the carrier power to be developed. When the effective plate voltage is then varied between zero and twice the d-c plate-supply voltage by the modulating voltage, the radio-frequency output voltage will closely follow the plate-voltage variations. It will be noted that these operating conditions correspond to sufficient grid excitation to cause the Class C amplifier to operate under relatively saturated conditions except possibly at the peak of the modulation cycle.

The exciting voltage for a plate-modulated Class C amplifier is preferably obtained from a source having relatively poor regulation, and it is also desirable that the grid bias for the modulated tube be obtained from a grid leak. These features prevent the grid current, and hence power dissipation at the grid, from being excessive at the trough of modulation, and also tend to increase the linearity of modulation. The grid condenser shunting the grid-leak resistance should be small enough so that its reactance at the highest modulating frequency is at least twice the resistance of the grid leak; otherwise distortion will be introduced through inability of the grid bias to follow variations introduced by the modulation.

The plate tank circuit of a plate-modulated amplifier should have an effective Q that is relatively low. Otherwise there will be appreciable discrimination against the higher side-band frequencies. The tank circuit should also have a high impedance in order that the plate efficiency be high. These requirements correspond to a high L/C ratio in the tank circuit, and a relatively close coupling to the load. When the output of the modulated amplifier is used to excite a linear amplifier, the load that the grid of the output tube places upon the modulated stage is variable. It then becomes desirable to load the tank circuit of the modulated stage with sufficient resistance to make the effective Q of the tank circuit reasonably constant in spite of the variable action of the linear amplifier.

The rated d-c plate voltage and carrier power of a tube for plate modulation is less than the corresponding ratings of the same tube used in Class C operation. This is because at the peak of a completely modulated wave the plate voltage effectively applied to the tube is twice the d-c value, and the power output at the modulation peaks is four times the carrier power. Inasmuch as the tube operates at these peak values for only a small fraction of the time, the power ratings for plate modulation are usually about two-thirds of those for Class C operation.

A properly adjusted plate-modulated Class C amplifier produces a modulation envelope that has very little distortion. When distortion does exist in a plate-modulated amplifier, however, it is usually in the form of a flattening off of the positive peaks, as shown in Fig. 3. Such distortion is caused by insufficient electron emission from the cathode of the tube, or distortion in the modulator.

A modulation meter or cathode-ray tube can be used to determine whether distortion is present. A rough check upon the presence of distortion can be obtained by taking advantage of the fact that with linear modulation, the d-c plate current of the tube does not vary when modulation is applied, whereas the d-c plate current will fall with the application of modulation if flattening of the positive peaks is present, and will rise with overmodulation.

Modulator Requirements.—The power amplifier that develops the voltage that is superimposed upon the plate-supply potential of the Class C tube in order to modulate the output is termed the *modulator*. It is usually a Class B or Class AB power amplifier, although in some cases Class A amplifiers are used. The modulator delivers its output to a load having an impedance equal to the ratio of d-c plate voltage to

d-c plate current of the Class C tube. For complete sinusoidal modulation, the modulating power required is 50 per cent of the d-c plate power that must be supplied to the tube during unmodulated conditions. The modulator is accordingly called upon to deliver to the modulated amplifier sufficient energy to generate the side-band power with Class C amplifier efficiency.

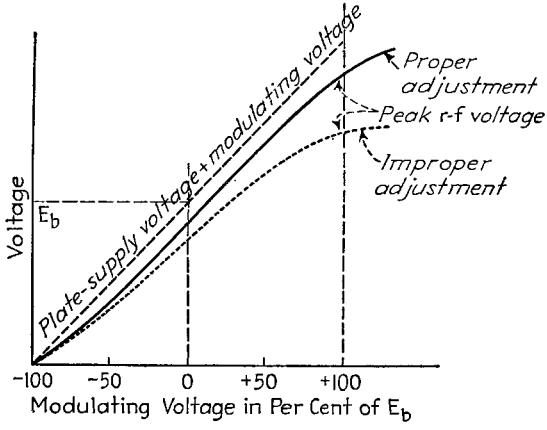


FIG. 3.—Characteristics of plate-modulated amplifiers with proper and improper adjustments.

Complete modulation requires that the peak value of the modulating voltage equal the d-c plate voltage. This requires not only that the modulator have sufficient power capacity but also that there be a proper impedance match. In particular, in the usual case where both modulator and modulated tubes receive their plate voltage from a common supply, it is necessary either to provide an impedance-matching transformer or auto transformer, as in Fig. 2*b*, or to reduce the plate voltage of the modulated tube by means of a resistance by-passed to modulation frequencies, as shown in Fig. 2*a*.

3. Grid-modulated Class C Amplifier.—In the grid-modulated Class C amplifier, the output of a Class C amplifier is controlled by varying the grid-bias voltage in accordance with the modulation, as shown by the simplified circuit arrangement of Fig. 4.

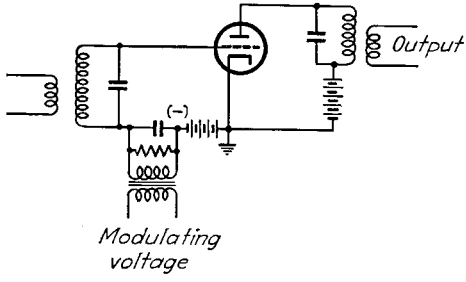


FIG. 4.—Schematic circuit of grid-modulated amplifier. For the sake of simplicity the neutralizing arrangement has been omitted.

The proper adjustment of the Class C amplifier is such that at the crest of the modulation cycle (*i.e.*, when the grid is the least negative), the tube operates under typical Class C amplifier conditions with the full rated Class C output, good plate efficiency, and a grid excitation that is not quite sufficient to give saturation conditions. These conditions can be achieved by usual procedures for designing and adjusting Class C amplifiers.

Complete modulation is then obtained by arranging the relative value of grid-bias voltage and modulating voltage to realize this condition at the positive peak of modulation, and at the same time just barely to cut off the output at the negative peak

of the modulating voltage. The required relation between the various voltages in the grid circuit is as follows

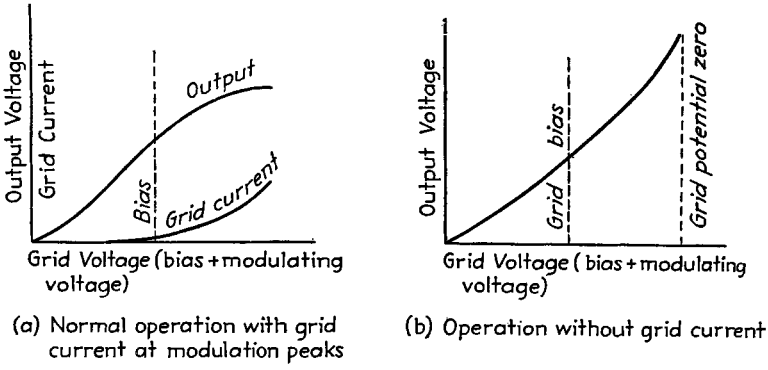
$$\text{Grid voltage for cutoff} = \frac{E_b}{\mu} + E_s \tag{5}$$

$$\text{Bias voltage} = \frac{E_1 + E_s + \frac{E_b}{\mu}}{2} \tag{6}$$

$$\text{Peak modulating voltage} = \frac{E_s + \frac{E_b}{\mu} - E_1}{2} \tag{7}$$

where E_1 = bias voltage required for Class C amplifier operation under conditions corresponding to the peak of modulation.
μ = amplification factor of tube.
E_b = d-c plate-supply voltage.
E_s = peak value of r-f exciting voltage.

The plate efficiency of a grid-modulated amplifier during unmodulated conditions is half of the Class C amplifier efficiency realized at the peak of the modulation cycle,



(a) Normal operation with grid current at modulation peaks

(b) Operation without grid current

FIG. 5.—Characteristics of grid-modulated amplifiers under various conditions.

or between 30 and 40 per cent in typical cases. When the wave is completely modulated, the plate efficiency averages about three-quarters of the Class C amplifier efficiency, or from 45 to 60 per cent for complete modulation. The carrier power that can be obtained from a tube operated as a grid-modulated amplifier is approximately one-fourth of the power obtainable from the same tube operated as an ordinary Class C amplifier.

Distortion in Grid-modulated Amplifier.—The relationship between modulating voltage and envelope of the modulated wave obtained with grid modulation is commonly S-shaped, as in Fig. 5*a*. The tendency for the positive peaks to be flattened off is usually the result of imperfect regulation of the r-f exciting voltage and the modulating voltage at the positive peak of the modulation cycle when the modulated tube draws grid current. It is also possible for flattening to occur as a result of incipient saturation, *i.e.*, because of too much exciting voltage or too high a load impedance. The opposite type of curvature existing in the modulation characteristic at small envelope amplitudes is the result of the finite plate resistance of the Class C tube. This part of the characteristic is more nearly linear the higher the tank-circuit impedance (corresponding to low power outputs) and the lower the plate resistance (and hence the amplification factor) of the tube being modulated.

If a reasonably linear modulation characteristic is to be obtained, it is necessary that the r-f exciting voltage remain substantially constant throughout the modulation cycle in spite of the fact that the Class C tube draws much power from the exciting voltage only at or near the peaks of the modulation. The desired regulation can be obtained either by employing an exciter of relatively large power capacity or by use of an amplitude stabilizing device, as shown in Fig. 6*a*.[1] In this latter arrangement, a diode *D* is shunted across the r-f exciting voltage and the cathode given a positive bias voltage that approximates the peak exciting voltage desired at the crest of the modulation cycle. Any tendency for the exciting voltage to rise to greater values during the remainder of the modulation cycle when the Class C tube draws little or no grid current causes a current to pass through the diode that places an additional load on the exciter, and hence minimizes the tendency for the voltage to change.

The required power capacity, or, more properly, the required internal impedance, of the source of modulating voltage, is determined by the fact that the modulating voltage should not be distorted appreciably by the grid current of the Class C tube

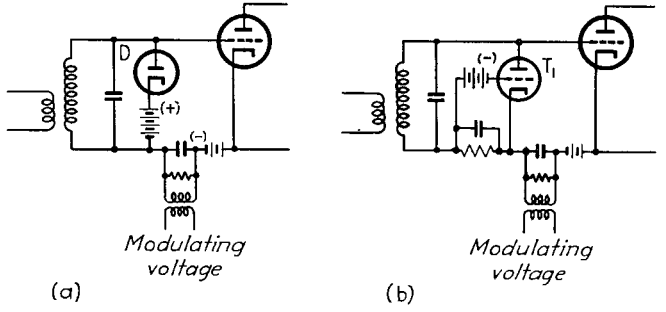


FIG. 6.—Systems for compensating for the effect of grid current in grid-modulated amplifiers.

flowing at the peak of the modulation cycle. The situation here is exactly the same as when one is exciting the grid of an audio power amplifier and drives the grid positive at the peaks of the cycle (see Sec. 5, Par. 7). The modulator power capacity can be reduced to less than would otherwise be required by means of the compensating scheme illustrated in Fig. 6*b*. Here the grid current that tends to flatten the positive peaks of the modulating voltage is used to increase the bias on voltage regulator tube T_1 so that the load that this tube places on the r-f exciting voltage is reduced. This causes the exciting voltage to be increased, thereby tending to compensate for the loss in output and eliminate the distortion that would otherwise be present.

Grid Modulation with No Grid Current.—An extremely linear modulation characteristic can be obtained with grid modulation by operating the modulated amplifier so that the grid is not driven positive even at the peak of the modulation cycle. This avoids the flattening off of output at the peak of modulation, and results in a parabolic modulation characteristic such as is shown in Fig. 5*b*. The curvature of such a characteristic will be less the higher the load impedance and the lower the amplification factor of the tube, and is similar to the curvature of the dynamic characteristic of a single Class A power amplifier tube (see Fig. 17, Sec. 5). Hence, by using a single Class A modulator tube coupled to the grid-modulated amplifier with proper polarity, the two curvatures will be of opposite sign, and by proper circuit proportions can be made to balance. The result is then substantially zero over-all distortion.

[1] The arrangements shown in Fig. 6 were first described by F. E. Terman and R. R. Buss, Some Notes on Linear and Grid-modulated Radio-frequency Amplifiers, *Proc. I.R.E.*, Vol. 29, p. 104, March, 1941.

Grid Modulation of Screen-grid, Beam, and Pentode Class C Amplifiers.—The same principles discussed above in connection with triode tubes also apply to grid modulation in pentode, beam, and screen-grid tubes. There are no differences involved other than obvious ones, such as, for example, the substitution of E_{sg} and μ_{sg} for E_b and μ in Eqs. (5)–(7). The linearity obtained is, however, usually less than that realized with triode tubes.

High-efficiency Grid-modulated Amplifier of Terman and Woodyard.[1]—The low efficiency of an ordinary grid-modulated amplifier can be greatly improved by the arrangement shown schematically in Fig. 7. Here the grid-modulated amplifier is divided into two parts, A_1 and A_2, with the plates of these sections interconnected by a quarter-wave length line and with the grid of A_1 provided with a quarter-wave phasing line so that the radio-frequency outputs will add in the load. This arrangement is analogous to that used in the high-efficiency linear amplifier discussed in Par. 22, Sec. 5, but differs in that instead of applying a radio-frequency exciting voltage

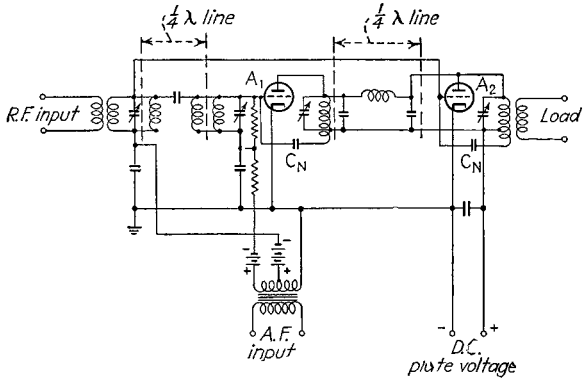


FIG. 7.—Schematic circuit diagram of high-efficiency grid-modulated amplifier.

of variable amplitude, the voltage applied to the grids consists of a radio-frequency exciting voltage of constant amplitude plus an audio modulating voltage.

The operation of the system can be explained briefly as follows: With no modulating voltage applied, tube A_2 is biased so that it just shows signs of plate current, and A_1 is operated with a load impedance and exciting voltage such that the desired carrier power is being generated under Class C amplifier conditions with the grid excitation just beginning to saturate. The required impedance in the plate circuit of tube A_1 is obtained by making the plate tank of A_2 have an impedance that is one-quarter of the impedance to be offered A_1 under carrier conditions, and making the characteristic impedance of the quarter-wave line one-half of this impedance to be offered to A_1. The impedance inverting action of the quarter-wave line will then cause the desired impedance to be presented to tube A_1. Modulating voltage is applied to the grids of the two tubes in the same phase. The relative values of modulating voltage depend upon the circuit adjustments and the tube characteristics,

[1] For further details, see F. E. Terman and J. R. Woodyard, A High-efficiency Grid-moduated Amplifier, *Proc. I.R.E.*, Vol. 26, p. 929, August, 1938.

Another system of high-efficiency grid modulation that is somewhat less practical than the one illustrated in Fig. 7 has been described by F. E. Terman and F. A. Everest, Dynamic Shift Grid Bias Modulation, *Radio*, p. 22, July, 1936. In this modulation system, the tube is operated at reduced plate and bias voltages during periods of no modulation. When modulation is applied, the bias and plate voltages are simultaneously increased in accordance with the envelope amplitude of the modulating voltage in such a manner that the ratio of plate voltage increment to bias increment is equal to the amplification factor of the tube.

but in most cases tube A_1 will require a somewhat larger voltage than tube A_2. The radio-frequency exciting voltage on the two tubes will also not necessarily be the same.

With proper adjustments, A_1 operates so that during the negative half cycle of the modulating voltage, this tube functions as a conventional grid-modulated amplifier giving full Class C amplifier efficiency at carrier level, and correspondingly less efficiency during the negative portion of the modulation cycle. On the positive half of the modulation cycle, tube A_2 also delivers output to the load impedance. This extra power received to the load causes the apparent load impedance at the receiving end of the line to increase, and because of the impedance inverting action of the quarter-wave line interconnecting the two tubes, causes the impedance that is offered to the plate circuit of A_1 to become less. As a result, A_1 increases its power output while still maintaining full Class C amplifier efficiency. At the positive peak of the modulation cycle tubes A_1 and A_2 deliver equal amounts of power to the load, both operate under ordinary Class C amplifier conditions, and the alternating voltages across the two ends of the quarter-wave plate-coupling line are the same.

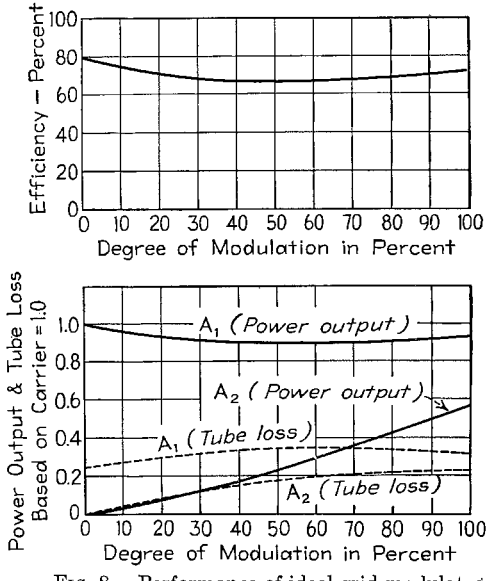


FIG. 8.—Performance of ideal grid-modulated amplifier, on the basis of a Class C amplifier efficiency of 80 per cent.

The average efficiency of this arrangement is high. Tube A_1 operates as a Class C amplifier of good efficiency at carrier level and during the positive half of the modulation cycle. Its efficiency falls off only at the troughs of the modulation cycle when the power involved is low. Tube A_2 also operates with full Class C amplifier efficiency at the modulation peaks when it is carrying a considerable amount of power. Although its efficiency drops to half that of a Class C amplifier before the tube becomes inoperative, the power handled becomes less as the efficiency drops.

The performance of a high-efficiency grid-modulated amplifier in an idealized case is shown in Fig. 8, on the assumption that the normal Class C amplifier efficiency is 80 per cent. It will be noted that the average efficiency is high for all degrees of modulation and that although A_1 and A_2 are called upon to develop the same peak power, the average output and average dissipation of A_2 is much less than for A_1.

The design and adjustment of the quarter-wave line interconnecting the plate circuits of the two tubes, and of the associated tank circuits and coupling to load, are carried out exactly as in the case of the high-efficiency linear amplifier discussed in Par. 22, Sec. 5. After this has been done, it is still necessary to determine the proper exciting and modulating voltages for the two tubes. These can be determined by starting with unmodulated conditions and A_2 inoperative. A_1 is excited and biased so that it gives the desired carrier output with grid saturation just beginning to show. The amount the bias must be increased to cut the output of A_1 to zero is then determined, and this increase in bias taken as the peak modulating voltage that will be

required for A_1. An exciting voltage such as would be used for Class C operation for A_2 is applied to this tube, and the bias is then adjusted so that there is just a trace of plate current flowing. This is the grid bias that A_2 employs. The difference between this grid bias, and the bias that would be used with the same r-f exciting voltage to give full Class C amplifier operation on tube A_2 with grid saturation just beginning to show, is the peak modulating voltage required by A_2. With these appropriate modulating voltages on the two tubes, the total power output at the positive crest of modulation should be four times the carrier power The linearity obtained with these modulating voltages can now be checked, and if it is not satisfactory, new values of exciting voltages should be employed and the design worked over again until satisfactory results are obtained.

The linearity of modulation of the high-efficiency grid-modulated amplifier is not particularly high under practical conditions. For service requiring low distortion, such as broadcast work, it is accordingly necessary to employ negative feedback.

4. Miscellaneous Types of Modulated Amplifiers.—Most modulated amplifiers are either of the grid- or plate-modulated type, but many other arrangements are used

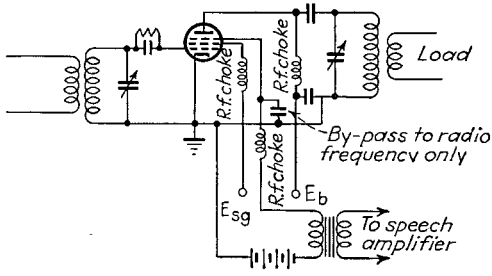


FIG. 9.—Circuit of suppressor-grid-modulated amplifier.

to a limited extent, and still other arrangements of merit have been proposed and tested in the laboratory. The more important of these miscellaneous modulation methods follow.

Suppressor-grid Modulation.—The output of a pentode Class C amplifier can be controlled by applying to the suppressor grid a modulating voltage superimposed upon suitable bias as illustrated in Fig. 9. The operation of this arrangement makes use of the fact that as the suppressor-grid potential is made increasingly negative, a virtual cathode forms between suppressor and screen. The plate current and hence the output then become less the more negative the suppressor voltage.

In the practical adjustment of the suppressor-grid modulated amplifier, the tube is first adjusted to give rated Class C amplifier operation under conditions where the suppressor grid is slightly positive. This condition corresponds to the crest of the modulation cycle, and represents an output approximately four times the carrier power that will be obtained. The suppressor potential required to reduce the output substantially to zero is then determined experimentally, giving the conditions at the negative modulation peak. Likewise, the suppressor voltage at which the output first begins to drop off is determined. The bias voltage is then the average of these last two potentials, and the peak modulating voltage is half their difference.

The plate efficiency of a suppressor-grid modulated amplifier is approximately half the efficiency obtained in normal Class C amplifier operation. The over-all efficiency is somewhat less because of the screen-grid losses. It is possible, however, to make the efficiency approach that obtained with Class C amplifier operation by using the same circuit as employed in the high-efficiency grid-modulated amplifier of

Fig. 7, except that the modulating voltages are now applied to the suppressors instead of to the control grids.

Typical characteristics of an ordinary suppressor-grid modulated amplifier are shown in Fig. 10. The linearity of modulation is not particularly high. Also, the screen current rises during the negative portion of the modulation cycle, so that care must be taken to avoid exceeding the allowable screen dissipation when suppressor-grid modulation is used.

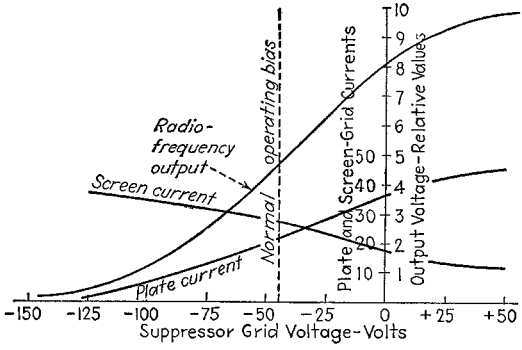


FIG. 10.—Linearity curve of typical suppressor-grid modulated amplifier. Note that it is impossible to obtain complete modulation without introducing appreciable distortion.

Screen-grid Modulation.—Modulation is occasionally obtained by applying the modulating voltage to the screen grid of the tube superimposed upon the screen voltage. This is analogous to plate modulation of a triode, and requires considerably less modulating power in proportion to output than with ordinary plate modulation. At the same time, the modulation characteristic is not particularly linear, and the plate efficiency has the same low value obtained with control-grid modulation.

Cathode Modulation.[1]—In cathode modulation the modulating voltage is applied between cathode and ground of a Class C amplifier as shown schematically in Fig. 11.

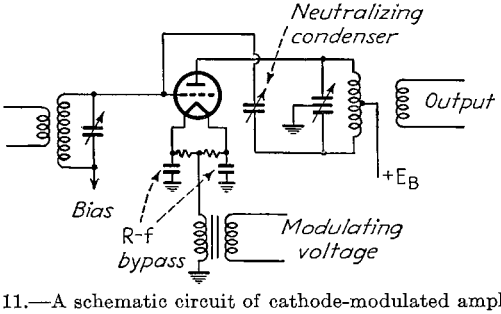


FIG. 11.—A schematic circuit of cathode-modulated amplifier.

This arrangement is essentially a combination of control-grid and plate modulation, with the former predominating. The plate efficiency, modulating power, and carrier power obtainable from a given tube are accordingly intermediate between the corresponding cases of pure plate and pure grid modulation, with a tendency to be more like the latter than the former.

[1] An extensive discussion of cathode modulation with practical circuit details for specific tubes is given by Frank Jones, "Cathode Modulation," Pacific Radio Publishing Company, San Francisco.

Loss (Absorption Methods) of Modulation.—Loss modulation is accomplished by using the modulating voltage to vary the energy absorbed from the generated radio-frequency oscillations. This method of modulation was one of the first ever used in radio communication. In spite of the fact that many methods of carrying out loss modulation have been proposed, the only arrangements of this type giving satisfactory performance are two recently developed types shown in Figs. 12 and 13.

In the circuit of Fig. 12*a*,[1] T_1 is a Class C amplifier obtaining its plate current through a choke L having high inductance to modulation frequencies. A losser tube T_2 is shunted across the tank circuit of the Class C amplifier, and has its bias adjusted so that with no modulation half of the output of the Class C tube is dissipated in the losser, and half goes to the load. When modulating voltage is applied to the losser tube, the load placed on the tank circuit L_1C_1 changes. The Class C tube would then

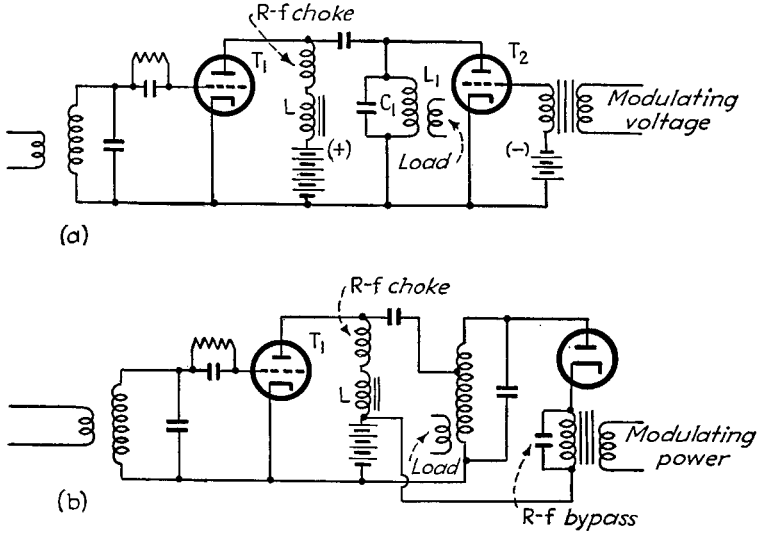


FIG. 12.—Circuits for loss modulation using constant-current choke.

draw a varying current if it were not for the inductance L, and there would be little effect upon the voltage across the tank circuit L_1C_1. However, since the inductance L maintains the current drawn from the plate-supply voltage substantially constant, varying the bias on T_2 will produce variations in the plate voltage of T_1. In particular, when T_2 is driven to cutoff, the plate voltage of T_1 will reach approximately twice the plate-supply voltage, and furthermore, the losses in T_2 will be zero, thereby enabling the positive peaks of modulation to be carried. At the same time, when the modulating voltage on T_2 is at its positive peak, the plate voltage on T_1 will be practically zero (assuming T_1 has low impedance), giving negligible output and negligible dissipation in the losser tube. The result is a modulating system in which the over-all efficiency under carrier conditions is half the efficiency of a Class C amplifier, while the efficiency increases during modulation. The carrier power obtainable is half the rating of the Class C tube operating as a plate-modulated amplifier, and the losser tube must be capable of dissipating a power equal to the carrier power.

The arrangements of Fig. 12*a* can be modified as shown in Fig. 12*b*, where the losser tube has been replaced by a diode in which a positive bias voltage and a modulating voltage are simultaneously applied to the cathode. By making the radio-

[1] The systems of modulation in Fig. 12 were devised by the author.

frequency voltage applied to the plate of the diode somewhat greater than the plate-supply voltage of the Class C tube, it is possible to obtain the cathode bias from the plate-supply voltage. With this arrangement, practically all the power absorbed by the losser circuit during the unmodulated intervals is returned to the plate supply in the form of rectified d-c current. The result is an over-all efficiency very nearly that of a Class C amplifier. The modulating power required is approximately one-half of the power that would be required to plate-modulate the same amount of carrier energy, while the carrier power that can be developed approximates one-half the rating of the same tube when operated as a plate-modulated Class C amplifier.

The arrangement illustrated in Fig. 13 has been termed load-impedance modulation.[1] Here the tank circuit of an oscillator or Class C amplifier is coupled to the load through a quarter-wave transmission line L_1, while losser tubes T_2 are shunted across the receiving end of a second quarter-wave line L_2 shunting the load. Modulation is applied to the control grids of the losser tubes, thereby varying the impedance at the receiving end x of the line L_2 in accordance with the modulation. This varies the impedance across y in an inverse manner as a result of the impedance inverting property of a quarter-wave line, which in turn varies the load upon the Class C

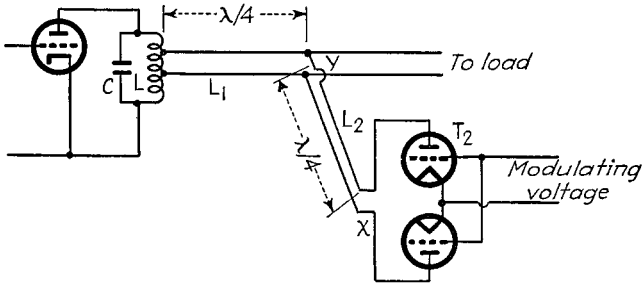


FIG. 13.—Schematic circuit for load-impedance modulation.

amplifier. The power delivered by the tank circuit LC to the line L_1 accordingly varies with the modulating voltage applied to the losser tubes, as does likewise the fraction of this power that is dissipated in the losser tubes. The combined result is a modulating system giving approximately the same efficiency and requiring the same modulating power as a grid-modulated amplifier. The system has been proposed for use in modulation of television transmitters as an alternative to grid modulation. In this application, the plate efficiencies of the two arrangements are about the same, as is also the carrier output in proportion to tube capacity. The band width is either comparable or slightly inferior to that with bias modulation, while the modulation characteristic is somewhat more linear.[2]

Modulation Systems in Which Side Bands and Carrier Are Generated Separately.—An arrangement of this type is illustrated schematically in Fig. 14*a*, and has been given the name "phase opposition modulation."[3] Here tube T_1 is excited by the unmodulated carrier, while T_2 receives its excitation from the output of a balanced modulator and operates essentially as a linear amplifier. The load is coupled to the tank circuit of tube T_2, and this circuit is in turn coupled to the separate tank circuit of T_1. The adjustment is such that during unmodulated conditions when T_2 is inoperative, T_1 operates as a conventional Class C amplifier with sufficient excitation

[1] W. N. Parker, A Unique Method of Modulation for High-fidelity Television Transmitters, *Proc. I.R.E.*, Vol. 26, p. 946, August, 1938.

[2] Hans Roder, Analysis of Load-impedance Modulation, *Proc. I.R.E.*, Vol. 27, p. 386, June, 1939.

[3] L. F. Gaudernack, A Phase-opposition System of Amplitude Modulation, *Proc. I.R.E.*, Vol. 26, p. 983, August, 1938.

on its grids to give complete saturation. During the positive half of the modulation cycle, tube T_2 delivers output to the tank circuit of this tube, and the phasings are so arranged that the output adds to the output that T_1 produces in the load. As far as T_1 is concerned, this is equivalent to an increase of the effective load impedance of the resonant circuit coupled to T_1, which increases the equivalent resistance that is coupled into the tank circuit of T_1, and so lowers the impedance of this tank circuit. The result is that T_1 produces additional output and helps carry the peaks of modulation. Conversely, on the negative half of the modulation cycle, the output of T_2 is of such a phase as to oppose the currents that T_1 delivers to the load. This reduces the resulting current flowing in the tank circuit of T_2 with consequent decrease of the impedance coupled into the resonant circuit of T_1. This increases the impedance offered to the plate circuit of T_1 and reduces the output of T_1. The conditions are such that at the positive peaks of a completely modulated wave, tubes T_1 and T_2 carry equal power, while at the negative peaks, T_1 is called upon to supply only the losses of its own tank circuit, and T_2 operates into a short circuit. The efficiency of the

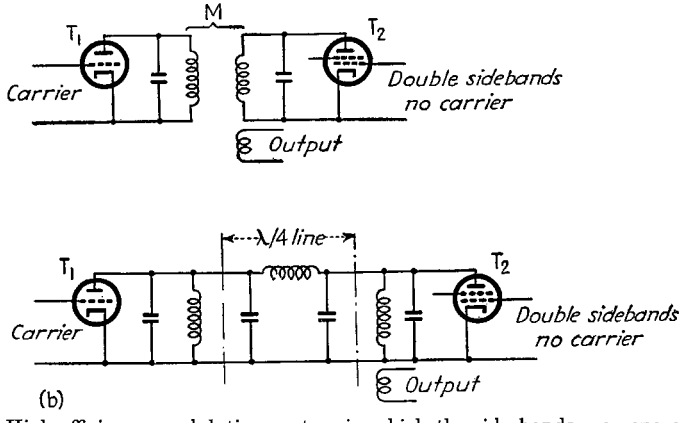


FIG. 14.—High-efficiency modulation system in which the side bands are generated separately from the carrier.

arrangement during periods of no modulation corresponds to that of an ordinary Class C amplifier, and drops off slightly with modulation, but will normally exceed 50 per cent even during periods of complete modulation.

A modification of this system is shown in Fig. 14*b*, where the two tubes are interconnected by a quarter-wave line adjusted as in the high-efficiency grid-modulated amplifier (Par. 3), instead of using inductive or capacitive coupling between separate tank circuits.[1] The operation of the system is essentially the same as in Fig. 14*a*. Tube T_1 generates carrier power during unmodulated conditions. When positive modulation is applied, T_2 delivers side-band power to the load in phase with the output derived from T_1, thereby increasing the effective impedance at the load end of the line. This lowers the impedance offered to the plate of T_1, and causes T_1 to increase its output to help carry the modulation peaks. Similarly, during the negative half of the modulation cycle, T_2 delivers output in phase opposition to the output derived from T_1. This reduces voltage across the load end of the line and hence the effective impedance at this point, resulting in an increase in impedance offered to T_1 and a reduction in output of T_1.

With both the preceding systems, it is desirable that T_2 function as a constant-current generator, so that pentode or screen-grid tubes are preferable. However,

[1] A. W. Vance, A High-efficiency Modulating System, *Proc. I.R.E.*, Vol. 27, p. 506, August, 1939.

where such tubes are not practicable, it is possible to substitute triodes and introduce compensation for the finite plate resistance by applying to T_2 a small amount of carrier along with the two side bands. This carrier can be obtained by slightly unbalancing the balanced modulators used to develop the side bands applied to T_2.

A modulation system has been proposed for broadcasting in which the carrier is radiated from a central antenna, while a polyphase figure of eight radiation pattern rotating at audio frequencies is radiated at the side-band frequencies from a combination of four vertical antennas at the corners of a square centered on the carrier antenna and appropriately phased.[1] In such an arrangement, the carrier tube can operate at full efficiency at all times, and since it supplies most of the power, the over-all efficiency of the system has a high average value.

High-efficiency Modulation System of Dome.[2]—This system combines absorption modulation with a linear amplifier and low-level modulation, as illustrated schematically in Fig. 15. Here tube T_3 is an ordinary plate-modulated amplifier. Tube T_2 is biased to act as a linear amplifier, but is provided with sufficient exciting voltage so that grid saturation occurs at approximately carrier level of exciting voltage. Tube T_1 is a losser tube connected to the receiving end of a quarter wave length

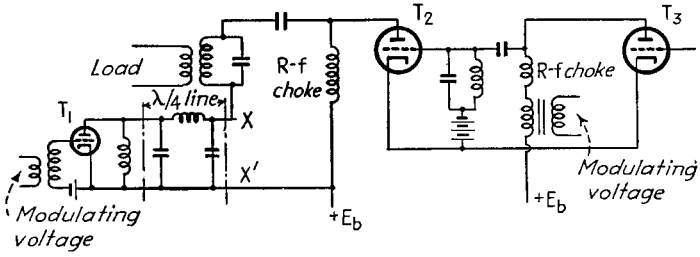


FIG. 15.—High-efficiency modulation system developed by Dome.

transmission line, the input of which is in series with the tank circuit of T_2. The grid of the losser tube T_1 is supplied with modulating voltage of such polarity that when the modulation on T_3 is at its positive peak the grid of T_1 is negative. The bias on T_1 is so adjusted that with zero modulating voltage, approximately half the output power of T_2 is delivered to the transmission line, and the remaining half goes to the tank circuit and thence to the load. The operation can be explained as follows: When the modulating voltage on T_3 is negative, tube T_2 acts as an ordinary linear amplifier. The resulting negative modulation is helped out by the fact that the modulating voltage on the grid of T_1 is positive, which increases the impedance across terminals xx of the line, thereby reducing the proportion of the output power of T_1 that is developed in the load. On the other half of the modulation cycle, the grid of T_1 is driven to cutoff at the peak, causing the impedance of the transmission line at xx to be negligible. This halves the total impedance between the plate and cathode of T_2, and also causes all the output power of T_2 to be delivered to the load, thereby taking care of the modulation peak. By proportioning the circuits so that under carrier conditions the radio-frequency voltage applied to the plate of T_1 is greater than the plate-supply voltage, a considerable part of the power delivered to the transmission line at xx is returned to the plate-supply voltage in the form of rectified d-c current, rather than being absorbed in the losser tube. The over-all efficiency of the combination is then of the order of 50 to 60 per cent.

[1] J. F. Byrne, Polyphase Broadcasting, *Trans. A.I.E.E.*, Vol. 58, p. 347, July, 1939; Paul Loyet, Experimental Polyphase Broadcasting, *Proc. I.R.E.*, Vol. 30, p. 213, May, 1942.

[2] R. B. Dome, High-efficiency Modulation System, *Proc. I.R.E.*, Vol. 26, p. 963, August, 1938.

Outphasing System of Modulation.[1]—This arrangement takes advantage of the fact that when the side-band frequencies are shifted 90° from the phase position existing in an amplitude-modulated wave, the envelope of the resulting wave is substantially constant. Such a wave can be amplified without distortion and with high plate efficiency by a Class C amplifier, after which it can be converted back into an amplitude-modulated wave by a suitable phase shift of the carrier with respect to the side-band components.

A practical arrangement for carrying out the necessary operations is illustrated schematically in Fig. 16. Here the Class C power-amplifier system is divided into two parallel parts. Side bands are separately generated by means of balanced modulator

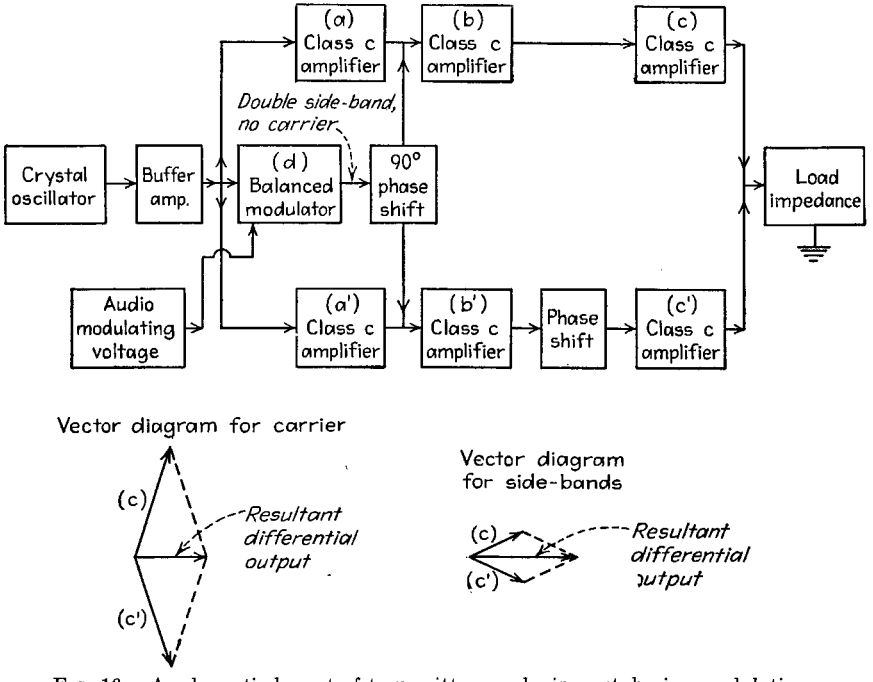


FIG. 16.—A schematic layout of transmitter employing outphasing modulation.

d, and after a 90° phase shift are inserted in equal amounts but in opposite phase in the two power-amplifier systems, as shown. The outputs of the two power-amplifier systems are then differentially combined in the load impedance so that the net input to the load is the vector difference of the outputs of the two amplifiers. The side-band components of the two amplifiers add directly in the load, since these side-band frequencies were introduced in opposite phase in the two parts of the amplifier. On the other hand, the carrier frequency in the load is almost canceled out, but is prevented from being zero by the insertion of a small phase shift in the lower amplifier. The resultant differential carrier is approximately in quadrature with the carrier voltages of the individual amplifiers, as shown by the vector diagrams in Fig. 16. The required 90° change in phase of carrier in the output is thereby achieved. By properly coordinating the amount of side-band energy inserted into the system and the phase shift in the one Class C amplifier system, complete modulation can be obtained.

[1] H. Chireix, High Power Outphasing Modulation, *Proc. I.R.E.*, Vol. 23, p. 1370, November, 1935.

The outphasing system of modulation is capable of giving substantially Class C amplifier efficiency throughout the modulation cycle. With proper adjustments, it also gives relatively low distortion.

Controlled-carrier Systems of Modulation.[1]—In controlled-carrier arrangements, the amplitude of the carrier wave is automatically varied in accordance with the degree of modulation as averaged over a short interval of time. When the modulation is small or zero, the carrier is made small, but as the modulating voltage is increased to a large value, then the carrier is increased accordingly. A typical wave that results is illustrated in Fig. 17.

Methods of producing controlled-carrier waves are shown in Fig. 18. At *a*, the plate-supply voltage of a plate-modulated amplifier is varied by a saturable reactor that is controlled by the d-c plate current of the Class B amplifier. When the modulating voltage is small, the reactor saturation is likewise small, thereby placing a high reactance in series with the power-supply system. On the other hand, with a large

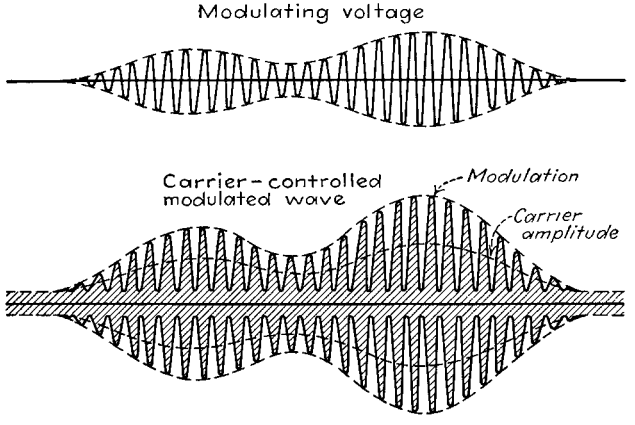


FIG. 17.—Wave transmitted from a station employing controlled carrier.

modulating voltage there is a large d-c current to saturate the reactor, thereby allowing the d-c voltage applied to the modulated tube to be large, and giving a correspondingly large carrier. In the arrangement at *b*, the plate circuit of the Class B modulator is connected between cathode and ground of the plate-modulated Class C tube as far as d-c and very low-frequency currents are concerned. With little or no modulation, the Class B tube consumes a large portion of the available supply voltage, while with large modulation the Class B tube draws more current in proportion to voltage. This permits a larger fraction of the total plate-supply voltage to be applied between the plate and cathode of the plate-modulated stage and so increases the carrier output.

The use of a controlled-carrier system reduces the average power of the modulated wave, and hence the average power loss in the modulated tubes. Also, when such a wave is applied to a conventional linear amplifier, it is possible to obtain approximately twice as much peak power in a typical case with the same average plate dissipation as would be obtained with an ordinary modulated wave. Controlled carrier has the disadvantage, however, that there is distortion when the modulating voltage suddenly increases in amplitude, since the carrier amplitude cannot be changed instantly. Furthermore, it is necessary that the receiver used in the reception of a controlled-carrier wave be provided with automatic-volume-control circuits having time con-

[1] G. W. Fyler, Phone Transmission with Voice-controlled Carrier Power, *QST*, Vol. 19, p. 9, January, 1935; I. A. Mitchell, Controlled Carrier Modulation, *Radio*, Vol. 17, p. 6, March, 1935.

stants suitably related to those of the transmitter. Otherwise violent distortion will occur when the carrier fluctuates in amplitude.

Van der Bijl Modulated Amplifier.—This type of modulator consists of an ordinary Class A amplifier, to the grid of which are applied a small radio-frequency carrier voltage and a large modulating voltage. Modulation takes place as a result of the fact that curvature of the $I_p - E_g$ characteristic of the tube causes the amplification of the radio-frequency voltage to vary during the modulation cycle, as shown in Fig. 19. This arrangement can be analyzed by the methods outlined in Par. 24, Sec. 5.[1]

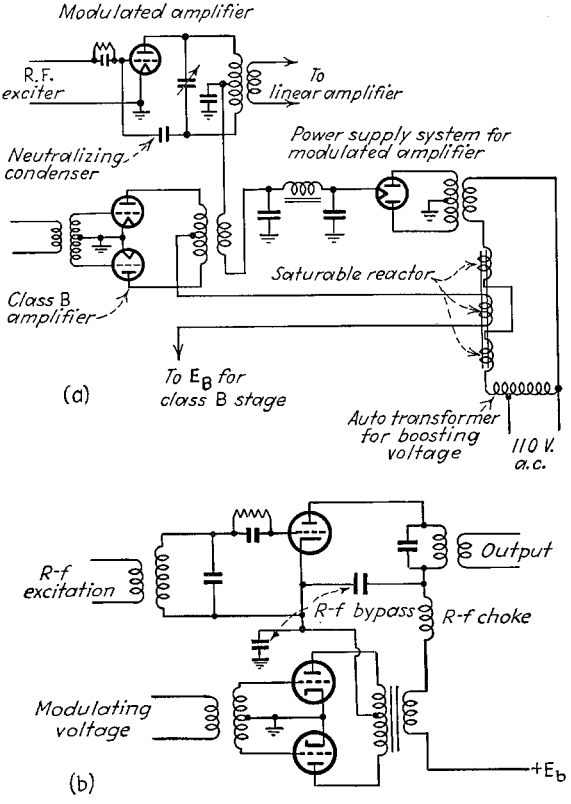


FIG. 18.—Methods of developing controlled-carrier wave.

It is found in this way that the side bands are the result of second-order curvature, and that for best results the load impedance in the plate circuit should be negligible to the modulating frequency, and be one-third of the average plate-resistance to the modulated wave.

The van der Bijl modulated amplifier is relatively easy to adjust, has stable characteristics, and requires negligible exciting and modulating power. At the same time its plate efficiency is low, and a high degree of modulation cannot be obtained without distortion. The system is used extensively in carrier telephone communication in balanced modulators.

[1] See also John R. Carson, The Equivalent Circuit of the Vacuum Tube Modulator, *Proc. I.R.E.*, Vol. 9, p. 243, June, 1921.

5. Modulated Oscillators.—In a modulated oscillator, the modulating voltage is used to vary the amplitude of the generated oscillations. This is in contrast with modulated amplifiers, where the modulation is accomplished by varying the amplification of a constant exciting voltage.

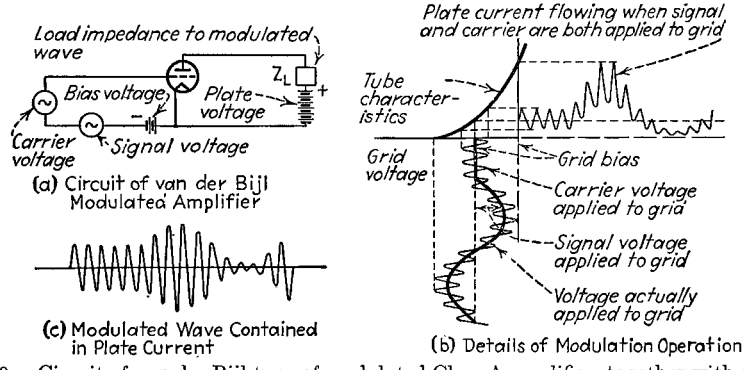


FIG. 19.—Circuit of van der Bijl type of modulated Class A amplifier, together with oscillograms showing details of operation.

Plate-modulated Oscillators.—The most common type of modulated oscillator is that in which plate modulation is used in the same manner in which a Class C amplifier is plate modulated. If the oscillator obtains its bias from a grid-leak grid-condenser arrangement and is adjusted to operate at high efficiency, the amplitude of oscillations is always such that the crest value of the alternating plate-cathode voltage is just slightly less than the d-c plate voltage. Accordingly, if the plate voltage is varied by

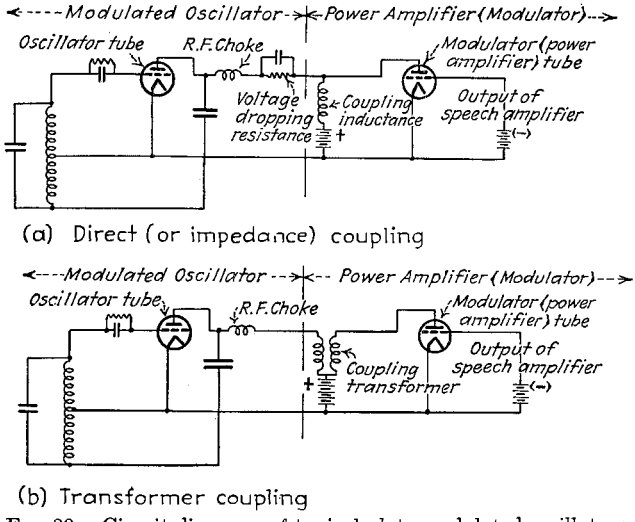


FIG. 20.—Circuit diagrams of typical plate-modulated oscillators.

means of a modulating voltage, the amplitude of the generated oscillations will faithfully follow the modulation.

Typical circuit arrangements for a plate-modulated oscillator are shown in Fig. 20, and are analogous to those of Fig. 2 for the plate-modulated amplifier. The power

relations and efficiency obtained with this type of modulation are the same as for the corresponding plate-modulated Class C amplifier.

The linearity of a plate-modulated oscillator is as good as or better than that of any other modulating system. To obtain good linearity, it is necessary merely that the oscillator be adjusted to give efficient operation, that the reactance of the grid condenser be at least twice the grid-leak resistance at the highest modulating frequency, and that the effective Q of the plate tank circuit be low enough to avoid discrimination against the higher modulation frequencies.

Miscellaneous Types of Modulated Oscillators.—Although plate modulation is by far the commonest method used to modulate oscillators, other means are occasionally employed. Thus the output of an electron-coupled oscillator can be modulated by applying modulating voltage simultaneously to the screen grid and plate in the ratio of the d-c voltages acting on these electrodes, as shown in Fig. 21. Oscillators having automatic amplitude control can be modulated by inserting modulating voltage in series with the voltage that acts to control the amplitude at the point indicated by x in Fig. 25, Sec. 6. The systems of modulation of Figs. 12 and 13 can also be applied to oscillators by obvious modifications.

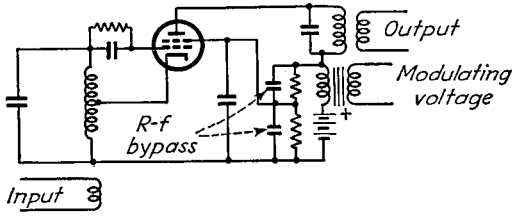


FIG. 21.—Combined plate and screen modulation of electron-coupled oscillator.

6. Comparison of Modulation Systems.—Modulated amplifiers are practically always used in preference to modulated oscillators, except occasionally at the lower radio frequencies. This is because in modulating an oscillator, there is inevitably a certain amount of incidental frequency modulation introduced. This results in spurious side bands that have sufficient intensity at all except the lowest radio frequencies to introduce undesired consequences in most applications.

The plate-modulated Class C amplifier is the most widely used of all systems of modulation. It gives excellent linearity, but has the disadvantage of requiring relatively large modulating power. The ordinary grid-modulated amplifier is used to a limited extent. It requires only a small modulating power, but has low plate efficiency and relatively high distortion. The high-efficiency form of the grid-modulated amplifier combines high over-all efficiency and economy of installed tube capacity to a greater degree than any other modulating system except perhaps the outphasing system. Although comparatively new, this high-efficiency system is already finding its way into commercial use. The outphasing system of modulation has been used in some of the high-power French broadcasting stations, but the circuits are so complicated and the adjustments so critical that it is probably less desirable than other comparable systems. The modulating systems of Fig. 14, in which the side bands are separately generated, have slightly less over-all efficiency than the high-efficiency grid-modulated arrangement, and are fully as complicated, so appear to offer no particular advantage. The loss-modulation systems described in Figs. 12 and 13 and also the Dome system have good to fair characteristics, but their use has so far been confined to equipment being tested experimentally. Cathode modulation is used to a considerable extent in amateur circles, but being partly control grid and partly plate modulation, it is

essentially a compromise inferior either to pure plate or pure grid modulation, whichever is considered most desirable for the circumstances at hand.

The modulation systems capable of giving the most linear modulation up to high degrees of modulation are the plate-modulated Class C amplifier, the grid-modulated Class C amplifier operated without grid current, and the outphasing system. A high degree of linearity is no longer of major importance in practical modulating systems, however, since with the development of negative feedback, it is now possible to sacrifice linearity for efficiency and still obtain low over-all distortion.

7. Carrier Suppression Systems and Single Side-band Generation.—The side bands of a modulated wave can be obtained without the carrier by use of a *balanced modulating* system such as illustrated in Fig. 22.[1] Here T_1 and T_2 are two modulated amplifiers to which carrier voltage is applied in the same phase and modulating voltage in opposite phase. If the system is symmetrical, the phase relations are such that no carrier voltage appears in the output, and side bands do appear in the output.

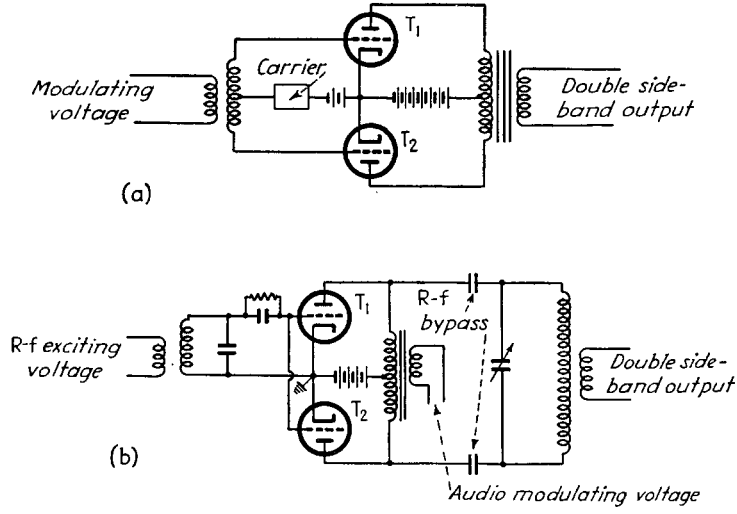


FIG. 22.—Balanced modulator arrangements for generating a double side-band wave with carrier suppressed.

A single side band can be produced by passing the output of a balanced modulator through a filter. Thus, if the carrier is 20 kc and the modulating frequencies are in the range 250 to 3,000 cycles, as in the case of carrier telephone communication, the output of the balanced modulator will contain a lower side band, extending from 19,750 to 17,000 cycles, an upper side band from 20,250 to 23,000 cycles, and whatever residual carrier comes through as a result of dissymmetry in the balanced arrangement. If this output is then passed through a filter that transmits only one of the side bands, as, for example, a band-pass filter in which the pass band extends from 20,250 to 23,000, one obtains a pure single side band. The highest carrier frequency for which a single side band can be obtained in this way is limited by the maximum sharpness of cutoff that can be obtained in the filter. However, once a single side band has been produced at a low carrier frequency, it can be shifted to a higher frequency region by modulating the single side band upon a higher carrier frequency and then separating out one of the resulting side bands by means of a second filter. Thus, if the single side band in the preceding example is used to modulate a 500,000-cycle carrier with the use of a balanced modulator, the output of this operation would contain two

[1] Other balanced modulator circuits are given in Par. 8.

side bands representing frequency ranges 520,250 to 523,000 and 480,250 to 483,000 cycles. It is then a relatively simple matter to separate one of these side bands from the other with a filter. The process can be repeated again by modulating this side band upon a still higher carrier frequency if the side band is to be displaced higher in the frequency spectrum. By carrying out the displacement of the side band in steps in this way, the percentage difference in frequency between the upper and lower side bands can always be made great enough for ordinary filters to function effectively.[1]

In some types of communication, a single side band is employed with a reduced carrier amplitude and a residual second side band. The frequency spectrum of such an asymmetric single-side-band system is shown in Fig. 23. Such arrangements are used in television, and have been proposed for broadcast transmission as a means of reducing the frequency band. Signals of this type can be produced by delivering the output of an ordinary modulator to a filter that is so designed that the carrier frequency is accurately located on the edge of the transmission band of the filter in the manner shown in Fig. 23.

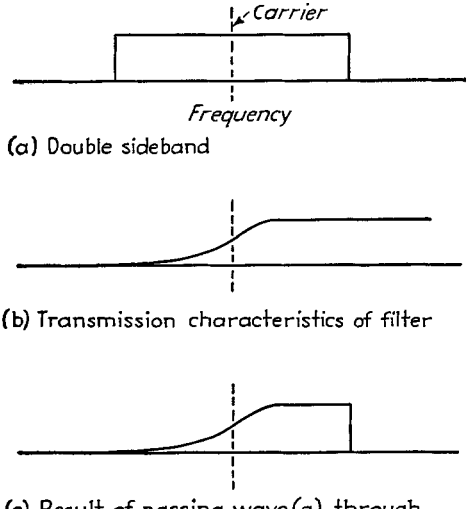


FIG. 23.—Schematic diagrams illustrating generation of single side-band signal with reduced carrier amplitude and residual second side band.

8. Nonlinear Modulators—Copper Oxide Modulators.[2]—Modulation can be accomplished by simultaneously applying audio- and radio-frequency voltages to a nonlinear circuit element. Practical modulators of this type ordinarily employ copper oxide rectifiers as the nonlinear device. Such rectifiers offer a very low resistance to voltages of one polarity and an extremely high resistance to potentials of opposite polarity, thereby having a very pronounced nonlinear action.

Copper oxide modulators find their widest application in telephone work and in laboratory apparatus. In such uses, it is customary to employ four copper oxide rectifier units arranged in a circuit such that the output of the modulator contains side bands but no carrier. Typical circuit arrangements are illustrated in Fig. 24. In practical operation these make use of a carrier-frequency voltage that is considerably greater than the audio voltage that is to be modulated upon the carrier. As a result, the carrier acts essentially as a high-frequency switch, causing the individual rectifiers to conduct or fail to conduct the applied modulating voltage according to the polarity of the applied radio-frequency voltage, with the modulating voltage having very little control over the action because of its small amplitude. In the

[1] Details of systems for producing single side bands located at high frequencies are given by A. A. Oswald, A Single Side-band Radiotelephone System, *Proc. I.R.E.*, Vol. 26, p. 1431, December, 1938; R. A. Heising, Production of Single Sideband for Trans-Atlantic Radio Telephony, *Proc. I.R.E.*, Vol. 13, p. 291, June, 1925.

[2] A more detailed discussion of copper-oxide rectifiers, together with methods of analyzing their behavior, is given in the following articles: R. S. Caruthers, Copper Oxide Modulators in Carrier Telephone Systems, *Trans. A.I.E.E.*, Vol. 58, p. 253, June, 1939; Leo L. Beranek, Applications of Copper Oxide Rectifiers, *Electronics*, Vol. 12, p. 15, July, 1939; E. Peterson and L. W. Hussey, Equivalent Modulator Circuits, *Bell System Tech. Jour.*, Vol. 18, p. 32, January, 1939.

circuit of Fig. 24*a* the action of such a carrier voltage causes the path from the input to output of the modulator to be periodically short-circuited every half cycle, while in Fig. 24*b* the path is open-circuited every half cycle. In the arrangement of Fig. 24*c* the radio-frequency voltage causes the direction in which the modulating voltage is transmitted to the output to be reversed every half cycle of the carrier voltage. In every case the resulting output contains the two side bands with the carrier absent. There are also present additional side bands corresponding to harmonics of the carrier, but these can be filtered out by appropriate circuits.

The side-band energy developed in the output is all derived from the energy applied to the modulator by the modulating voltage, and is equal to this energy less the resistance loss in the copper oxide units. These losses can be as low as 2 db at low frequencies, and will reach 8 to 9 db at frequencies of the order of 4 mc.

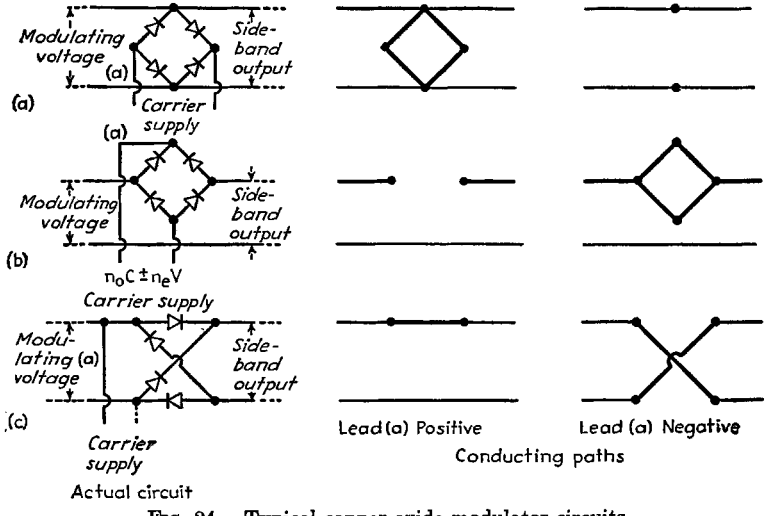


FIG. 24.—Typical copper oxide modulator circuits.

Copper oxide rectifiers are suitable for operation up to frequencies of at least 4 mc. At these high frequencies it is necessary to operate at an impedance level of the order of 50 ohms to minimize the effect of internal capacities. At lower frequencies impedance levels of 600 to 1,000 ohms can be used, and result in increased efficiency.

DETECTION

9. Diode Detection of Amplitude-modulated Waves.[1]—Detection, which is also sometimes referred to as *demodulation*, is the process of recovering from a wave the intelligence that was originally modulated upon it. In the case of amplitude modulation, detection accordingly means deriving from the modulated wave a voltage that varies in accordance with the modulation envelope. In all practical cases, this is accomplished by rectification of the modulated wave.

The most widely used detector of modulated waves is the diode rectifier. The circuit of a simple detector of this type is shown in Fig. 25*a*, where R is the resistance

[1] The discussion given here for diode detectors was worked out by the author with the aid of C. K. Chang, and represents an extension of the methods of Harold A. Wheeler, Design Formulas for Diode Detectors, *Proc. I.R.E.*, Vol. 26, p. 745, June, 1938. Additional information on diode detection is given by E. Roberts, Straight-line Detection with Diodes, *Jour. I.E.E.*, Vol. 75, p. 379, 1934; Wireless Section, *I.E.E.*, Vol. 9, p. 325, September, 1934; C. E. Kilgour and J. M. Glessner, Diode Detection Analysis, *Proc. I.R.E.* Vol. 21, p. 930, July, 1933.

to which the rectifier output is delivered and C is a condenser that by-passes radio-frequency voltages, but has a high impedance to the modulation-frequency potentials developed across R.

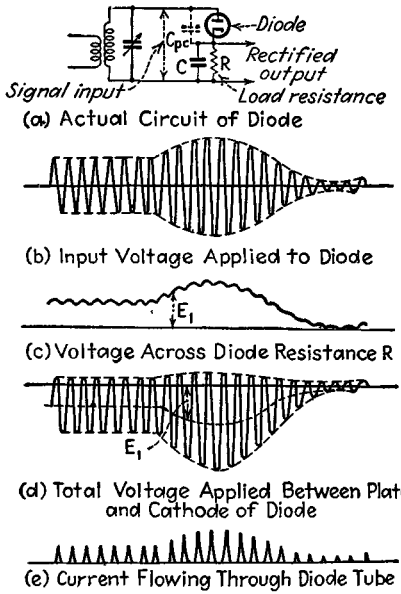


FIG. 25.—Circuit of simple diode detector with oscillograms illustrating mechanism of operation.

The action of such a detector is shown in the oscillograms of Fig. 25. At each positive peak of input voltage, the plate is more positive than the cathode, causing a pulse of a plate current to flow that charges the condenser C to a potential that is only slightly less than the peak voltage of the signal input. Between peaks, the plate of the diode is less positive than the cathode as a result of the negative charge on the condenser C, so that no current flows through the tube and part of the charge on the condenser C leaks off through the resistance R, to be replenished by the next pulse of current.

Efficiency of Rectification.—The ratio of d-c voltage developed across the output of a diode to the peak amplitude of the applied carrier is termed the efficiency of rectification of the carrier.

$$\left.\begin{matrix}\text{Efficiency of rectification of carrier}\end{matrix}\right\} = \eta_0 = \frac{\text{d-c voltage across diode load}}{\text{peak amplitude of applied carrier}} \quad (8)$$

Similarly, the efficiency with which the modulation is rectified can be expressed by the equation

$$\left.\begin{matrix}\text{Efficiency of rectification of modulation}\end{matrix}\right\} = \eta_m = \frac{\left\{\begin{matrix}\text{peak modulation-frequency voltage}\\ \text{developed across diode load impedance}\end{matrix}\right\}}{\left\{\begin{matrix}\text{(degree of modulation) } \times \text{ (peak amplitude of applied carrier)}\end{matrix}\right\}} \quad (9)$$

The efficiency η_0 of rectification of the carrier depends upon the amplitude of the applied signal and the ratio of diode load impedance to internal plate resistance of the diode tube. It tends to be constant for large signal voltages, but falls off when the applied potential is of the order of a few volts or less. The value of η_0 with large signals depends upon the ratio of load resistance to internal plate resistance of the diode, and is normally of the order of 70 to 90 per cent, as shown in Fig. 26.

The efficiency η_m with which the modulation is rectified usually differs only very slightly from the efficiency with which carrier is rectified, particularly when the efficiency of rectification is high. For ordinary design purposes, one can assume that the two are the same unless the load offers an impedance to modulation-frequency currents that is very much less than the d-c resistance, in which case η_m will be slightly less than η_0.

Distortion in Diode Detectors—A-C–D-C Impedance Ratio.—An ideal diode detector will exactly reproduce in its output the modulation envelope of the applied signal. When a practical diode detector is operated under conditions where the efficiency of rectification is high and is substantially constant over the range of amplitudes represented by the variations of the modulation envelope, the output will contain very

little distortion, provided that one satisfies the relation

$$\frac{\left\{\begin{array}{l}\text{Diode load impedance to}\\ \text{modulation-frequency currents}\end{array}\right\}}{\left\{\begin{array}{l}\text{Diode load impedance to}\\ \text{direct currents}\end{array}\right\}} \geqslant \left\{\begin{array}{l}\text{degree of modulation of}\\ \text{applied signal}\end{array}\right. \tag{10a}$$

In symbols, this equation can be written

$$\frac{Z_m}{R_0} \geqslant m \tag{10b}$$

The ratio Z_m/R_0 is commonly referred to as the *a-c–d-c impedance ratio* of the diode load.

The character of the distortion resulting from violating the criterion represented by Eq. (10) is shown in Fig. 27. In the case where the a-c impedance Z_m is a resistance, but is less than the d-c resistance, as is the case in the circuits of Fig. 29*a* when

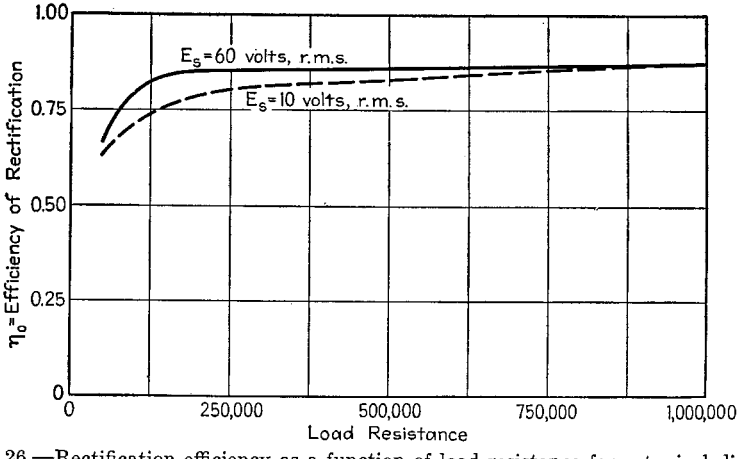


FIG. 26.—Rectification efficiency as a function of load resistance for a typical diode.

the condenser C_2 is an adequate by-pass to modulation frequencies, then the negative peaks of the output voltage are clipped off at a point where the instantaneous amplitude corresponds to an instantaneous degree of modulation allowed by Eq. (10). This is shown in Fig. 27*c*. When such clipping occurs, the amount of distortion that results is approximately

$$\left\{\begin{array}{l}\text{Approximate rms}\\ \text{distortion}\end{array}\right\} = \frac{\text{actual modulation} - \text{modulation allowed by Eq. (10)}}{2 \times \text{actual modulation}} \tag{11}$$

When the diode load impedance Z_m to the modulation frequency is less than the d-c resistance and also has a phase angle, as will always be the case at high frequencies where the shunting condenser cannot be neglected, then violation of Eq. (10) results in a diagonal clipping, as in Fig. 27*d*. The distortion introduced in this way is less than with flat clipping of the negative peaks, but is by no means negligible.

When the criterion of Eq. (10) is satisfied, the distortion is small, although there will be some residual distortion as a result of variation of the efficiency of rectification with instantaneous envelope amplitude. This is commonly under 3 per cent, with the value tending to be smallest when the applied signal is large, the degree of modulation small, and the diode load impedance high compared with the diode plate resist-

ance. The exact magnitude of this type of distortion can be determined for any particular case by experimental methods given below.

Input Impedance of Diode Detectors—Reduction in Modulation of Input Signal by Detector.—A diode offers to the applied carrier voltage a resistive input impedance given by the relation

$$\text{Input resistance to carrier} = \frac{R_0}{2\eta_0} \qquad (12)$$

where R_0 is the d-c resistance of the diode load impedance and η_0 is the efficiency with which the carrier is rectified. Similarly to a side-band frequency the diode acts as though it had an input impedance

$$\left.\begin{matrix}\text{Diode input impedance} \\ \text{to side-band frequency}\end{matrix}\right\} = \frac{|Z_m|}{2\eta_m} \qquad (13)$$

where $|Z_m|$ is the magnitude of the diode load impedance to the modulation frequency involved and η_m is the efficiency of rectification of the modulation. The phase angle of the input impedance offered to the upper side band is the same as the angle of Z_m, and for the lower side band is the negative of this phase angle.

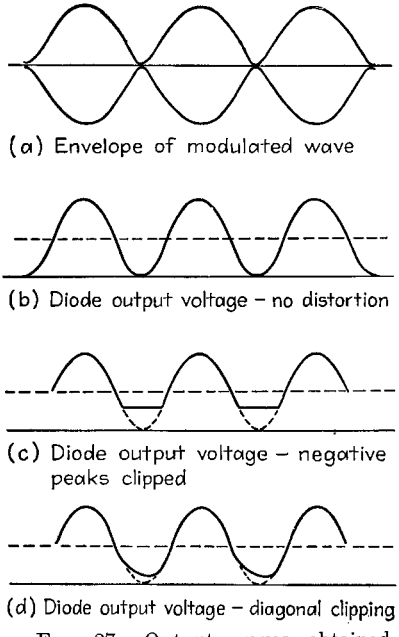


FIG. 27.—Output waves obtained from a diode, indicating different types of distortion.

The input impedance of the diode causes the amplitude of the carrier and side bands at the diode terminals to be less than if the tube were inoperative. The amount of reduction is determined by the ratio of internal impedance of the exciting voltage to the input impedance of the diode, and is not necessarily the same for the carrier and side bands. The reduction in carrier amplitude caused by the diode loading may accordingly be accompanied by a change in the degree of modulation.[1] The quantitative relations involved are as follows:

$$\left.\begin{matrix}\text{Carrier voltage} \\ \text{applied to diode}\end{matrix}\right\} = E = \frac{\dfrac{R_0}{2\eta_0}}{Z_s + \dfrac{R_0}{2\eta_0}} \qquad (14)$$

$$\left.\begin{matrix}\text{Degree of modulation} \\ \text{at diode terminals}\end{matrix}\right\} = m = \frac{\dfrac{Z_m}{2\eta_m}}{\left(Z_s' + \dfrac{Z_m}{2\eta_m}\right)} \frac{Z_s + \dfrac{R_0}{2\eta_0}}{\dfrac{R_0}{2\eta_0}} m_0 \qquad (15)$$

where E_0 = carrier voltage at diode terminals with tube inoperative (*i.e.*, filament unlighted).
m_0 = degree of modulation at diode terminals with tube inoperative.
Z_s = internal impedance of source of exciting voltage at the carrier frequency.
Z_s' = internal impedance of source of exciting voltage at the side-band frequencies.

[1] This change in the degree of modulation of the applied signal is in addition to any side-band trimming caused by the resonant circuit supplying voltage to the diode.

The quantities R_0, η_m, η_0, and Z_m are defined as before. In Eqs. (14) and (15), it is to be noted that Z_s, Z'_s, and Z_m may have phase angles, and so must be treated as vectors. In Eq. (15), the phase angles to be used for Z'_s and Z_m are either the pair of values for the upper side band or the pair of values applying to the lower side band.

Examination of Eq. (15) brings out the fact that when $Z_m/R_0 < 1$ and the source impedance is substantially the same to the carrier and side bands, the diode reduces the degree of modulation of the voltage existing at its terminals. This helps prevent the degree of modulation at the diode terminals from exceeding the allowable value given by Eq. (10).

Radio-frequency Voltage in Detector Output.—A radio-frequency voltage appears across the shunt condenser in Fig. 25*a* as a result of radio-frequency currents that flow through the plate-cathode capacity C_{pc} of the diode and the radio-frequency components of the pulses of plate current. Although this voltage is small, it is commonly not negligible. To prevent it from passing on into the audio-frequency amplifier following the detector output, it is customary to provide a radio-frequency filter R_1C_1 as shown in Fig. 29. The capacity C_1 in this filter is usually the same size as C, and the resistance R_1 is a portion of the d-c resistance of the diode load impedance.

On the assumption that the radio-frequency voltage in the output is small, one has[1]

$$\text{R-f carrier voltage across } C = E_0 \left[\left(\frac{C_{pc}}{C}\right)^2 + \left(\frac{\frac{1}{\omega C}}{\frac{R_0}{2\eta_0}} \right)^2 \right]^{1/2} \tag{15a}$$

The presence of the radio-frequency filter reduces this voltage by the factor $(1/\omega C_1)/R_1$

The first term in the parentheses on the right-hand size of Eq. (15*a*) takes into account the effect of the current through the plate-cathode capacity of the diode, while the second term is the contribution of the pulses of plate current. These components are in quadrature, and with practical circuits designed for operation on voice-modulated waves commonly have the same magnitude somewhere in the frequency range 75 to 200 kc. At higher frequencies the first term will dominate; at lower frequencies, the second.

Experimental Determination of Diode Characteristics.—The efficiency with which a diode detector rectifies the carrier can be readily determined by applying a sine-wave voltage to the diode. The applied signal can be of any frequency, even 60 cycles, it merely being necessary that the capacity shunting the diode load impedance have negligible reactance at the frequency of measurement. The d-c voltage across the load can be conveniently obtained from a knowledge of the d-c resistance R_0 and the reading of a microammeter in series with this resistance.

The rectification efficiency to the modulation for the case where the load impedance Z_m to the modulation frequency is a resistance less than the d-c resistance R_0 can be obtained experimentally as follows: the d-c rectified voltage E_0 is first observed with an applied signal corresponding to carrier amplitude, and with a diode load impedance equal to the d-c resistance. The load resistance is then reduced to Z_m, and a positive-bias voltage is applied to the cathode in series with the load impedance as shown in Fig. 28, having the value

$$\text{D-c bias} = \left(1 - \frac{Z_m}{R_0}\right) E_0 \tag{16}$$

The amplitude of the applied signal is now varied to simulate the variation in envelope amplitude present during modulation and the resulting voltage across the load resist-

[1] This relation was derived and verified experimentally by the author and C. K. Chang.

ance Z_m plotted as in Fig. 28. The rectification efficiency η_m to the modulation is given by the slope of the resulting curve, as shown.

The residual distortion occurring in detection when the conditions specified by Eq. (10) are satisfied can be obtained by treating the curve such as that of Fig 28 that gives the variation of output voltage with variations in signal amplitude about

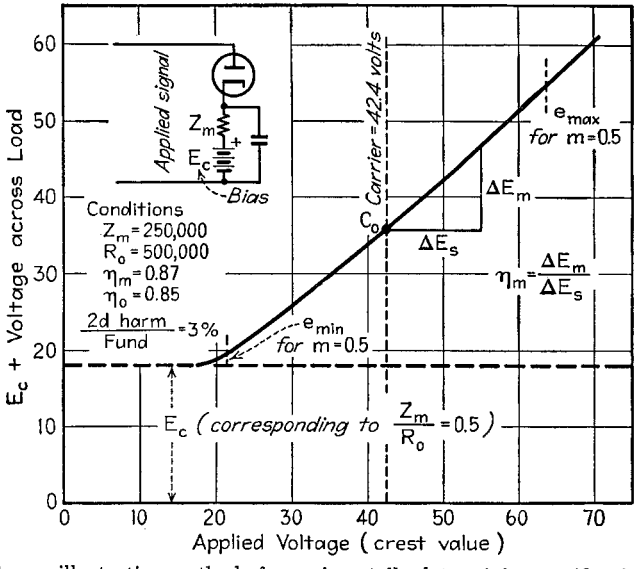


FIG. 28.—Curves illustrating method of experimentally determining rectification characteristic when the a-c-d-c impedance ratio is less than unity.

the carrier value, as though these curves were dynamic characteristics of an amplifier; *i.e.*

$$\frac{2d \text{ harmonic}}{\text{Fundamental}} = \frac{e_{max} + e_{min} - 2e_0}{2(e_{max} - e_{min})} \tag{17}$$

where e_{max} = output voltage between cathode and ground when signal is $(1 + m)$ times carrier.

e_{min} = output voltage between cathode and ground when signal is $(1 - m)$ times carrier.

e_0 = output voltage between cathode and ground when signal is the carrier.

These voltages are indicated in Fig. 28, and all include the bias E_c as part of e_{max}, e_{min}, and e_0.

Practical Diode Circuits and Factors Controlling Their Performance.—Circuits of typical diode detectors as used in practice are shown in Fig. 29. These differ from the simple arrangement of Fig. 25*a* in that a radio-frequency filter R_1C_1 is provided, and d-c voltages are blocked from the output by means of the condenser C_2. In the circuit of Fig. 29*b*, provision is also made for automatic volume control by separating the d-c component from the modulation-frequency component of the output with audio-frequency filter R_3C_3. The radio-frequency filter consumes some of the rectified output of the detector, so that the net voltage available at the output terminals is less than the voltage developed across the full load impedance. The presence of the blocking condensers C_2 and C_3 also causes the load impedance to modulation frequencies to be less than the resistance offered to d-c rectified currents. This intro-

duces the possibility of negative peak clipping at low and moderate modulation frequencies, with an additional diagonal clipping possible at high frequencies because of the shunting effect of the condenser C at high audio frequencies.

In proportioning circuits such as are shown in Fig. 29, the input condenser C is commonly 50 to 100 $\mu\mu$f. This value is fixed by the fact that condenser C should be considerably larger than the plate-cathode capacity of the diode and, at the same time, should have the highest possible impedance to modulation frequencies. The d-c resistance $R_0 = R_1 + R_2$ is selected on the basis that this resistance must be at

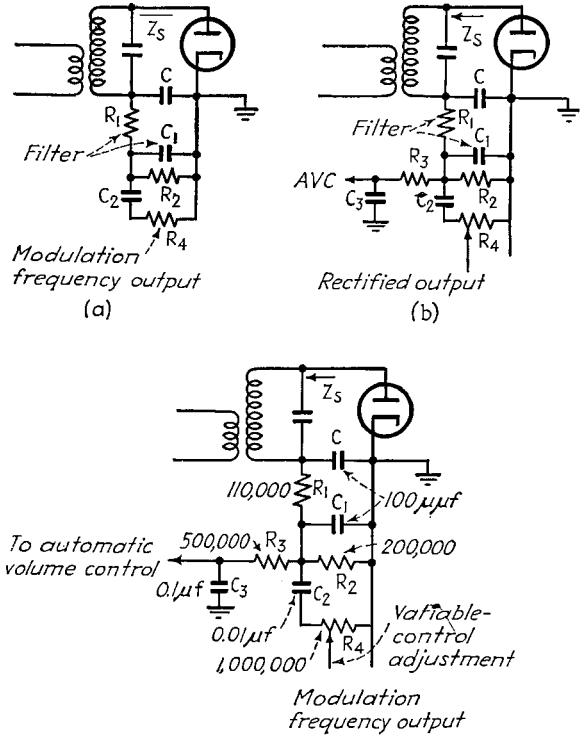


FIG. 29.—Typical design of diode detector in which provision is made for an automatic-volume-control voltage and for obtaining the modulation-frequency output free of direct-current voltages.

least as great as the reactance of the input condenser C at the highest modulation frequency that is to be rectified without appreciable frequency distortion. The resistances R_3 and R_4 in the audio-frequency shunts are made as high as possible, and are commonly arranged so that the resistance R_2 across which they are shunted is appreciably less than the total d-c resistance of the diode load. In this way, the impedance Z_m offered to modulation frequencies at the terminals of condenser C will differ less from the d-c resistance across the same terminals than if the audio-frequency shunts were connected directly across condenser C and were of lower resistance.

The most important characteristics of a diode detector are (1) the maximum degree of modulation as a function of modulation frequency that the incoming signal may

have, without some form of negative peak clipping; and (2) the variation of output voltage as a function of modulation frequency for an incoming signal of constant amplitude and degree of modulation. These two characteristics depend upon both the input impedance of the diode and the internal impedance of the source of exciting voltage, and make it necessary to consider the diode tube with its load circuit and input circuit as a single system.

The allowable degree of modulation that an incoming signal may have without introducing peak clipping can be calculated by the following steps:

1. Determine the ratio of modulation at the input terminals of the diode to the modulation of the incoming signal for the case where the diode tube is inoperative. This takes into account reduction in the degree of modulation caused by side-band trimming arising from the selectivity of the tuned input circuit.
2. Calculate with the aid of Eq. (15) the reduction in degree of modulation produced by the input impedance of the diode.
3. Multiply the factors obtained in steps (1) and (2) to get the total loss in degree of modulation arising from side-band trimming plus input impedance effects of the diode.
4. Calculate the allowable degree of modulation at the input terminals of the diode by Eq. (10).
5. Divide the result from (4) by the factor obtained in (3) to obtain the degree of modulation of the incoming signal that makes the modulation at the diode terminals equal the maximum allowable value.

The results of such calculation for the detector proportions of Fig. 29*c* are given in Fig. 30.

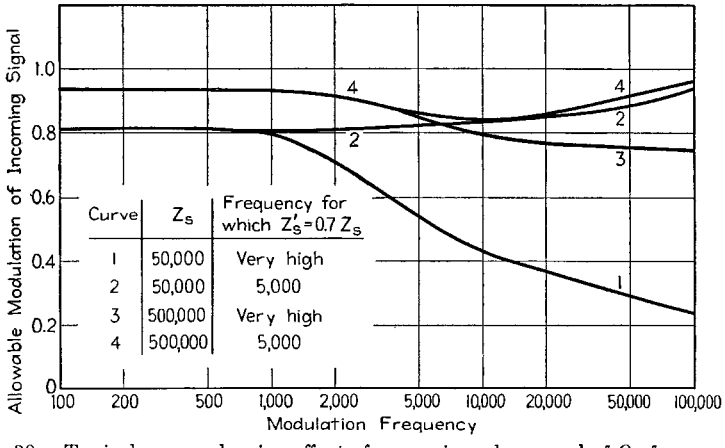


FIG. 30.—Typical curves showing effect of source impedance and of Q of source impedance on the allowable degree of modulation a signal may have without peak clipping. This figure assumes the detector circuit of Fig. 29*c*.

The maximum degree of modulation that the incoming signal may have without peak clipping depends primarily upon the impedance ratio Z_m/R_0 of the diode input, and upon the ratio of the source impedance to the input impedance of the diode. When the internal impedance of the voltage applied to the diode is much higher than the input impedance of the diode, then Eq. (15) shows that the diode reduces the degree of modulation by the factor Z_m/R_0 (assuming $\eta_m = \eta_0$). There is then no possibility of negative peak clipping, since even a completely modulated incoming signal has its degree of modulation reduce to the point where Eq. (10) is satisfied. This corresponds to curve 3 in Fig. 30, which is to be compared with curve 1. On the other hand, when the source impedance Z_s and (Z_s') is much less than the input impedance of the diode, then the input impedance of the diode does not affect the degree of modulation, and peak clipping occurs when the modulation of the incoming signal exceeds Z_m/R_0. Side-band trimming in the tuned input circuit reduces the tendency

toward peak clipping as shown by comparing curves 1 and 2 in Fig. 30, and also curves 3 and 4, respectively.

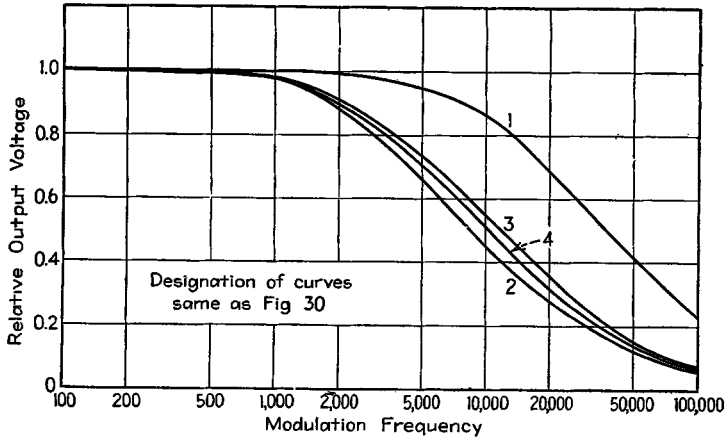


FIG. 31.—Curves corresponding to those of Fig. 30, giving relative output voltage of the detector as a function of modulation frequency, and assuming constant amplitude and modulation of the incoming signal.

The relative output voltage as a function of modulation frequency is given by the way in which the modulation at the diode terminals varies with modulation frequency, as calculated in step 3 above. This characteristic is affected by the way in which the ratio Z_m/Z'_s varies with modulation frequency and by the side-band trimming produced by the selectivity of the exciting circuit at the higher modulation frequencies. When the internal impedance of the r-f source is much higher than the input impedance of the diode, the relative output voltage will then be proportional to Z_m/R_0 (assuming $\eta_m = \eta_0$). This is the case in curve 3, Fig. 31. On the other hand, when the internal impedance of the source is much less than the input impedance of the diode, the only variation in output voltage with modulation frequency is the side-band trimming produced by the selectivity of the r-f exciting circuit at high modulation frequencies. Curves 1 and 2 of Figs. 30 and 31 correspond to this case. The behavior in intermediate cases depends upon the relative r-f source and diode input impedances and upon their relative variation with frequency. In general, when the internal impedance of the source and the input impedance of the diode are of the same order of magnitude, then the

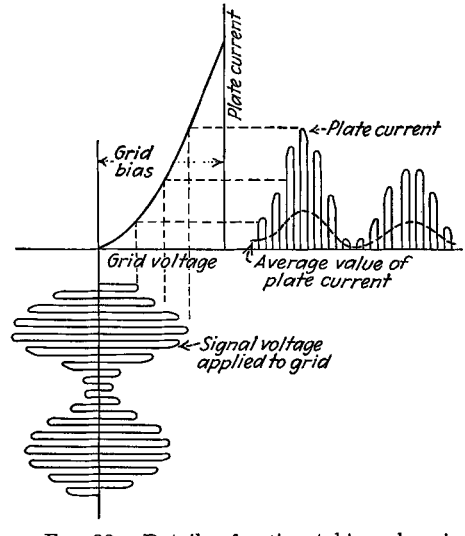


FIG. 32.—Details of action taking place in plate rectifier, showing how a modulated wave applied to a grid adjusted to cutoff will cause plate-current pulses having an average value that varies in accordance with the modulation envelope.

falling off in output at high modulation frequencies will be either greater or less than that caused by side-band trimming in the input circuit, according to whether the load impedance of the diode varies more or less rapidly, respectively, with modulation frequency than does the internal impedance of the exciting voltage.

In general, circuit relations that result in the output falling off at the higher modulation frequencies will also correspond to a high allowable degree of modulation at high modulation frequencies, and vice versa. A comparison of Figs. 30 and 31 illustrates this.

10. Miscellaneous Methods of Detection.—Although diode detectors are used in the great majority of cases in rectifying modulated waves, other methods of detec-

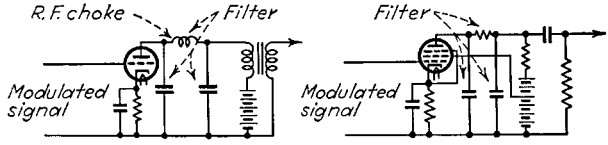


FIG. 33.—Typical plate-detector circuits.

tion are possible and have a limited use. The more important of these are described below.

Plate Detection of Large Signals.[1]—In plate detection (also termed bias and anode detection), an ordinary amplifier tube is biased approximately to cutoff. A radio-frequency signal voltage applied to the grid of such a tube will give pulses of plate current on the positive half cycles and no current on the negative cycles, as shown in Fig. 32. The resultant average plate current is then dependent upon the average amplitude of the applied signal, thus giving demodulation. This current develops an output voltage by being passed through an ordinary audio-coupling system, as shown in Fig. 33.

As far as the audio-frequency load impedance is concerned, the modulation-frequency currents in a plate detector output act as though they were produced by a generator voltage E_r acting between plate and cathode in series with a certain equivalent plate resistance R_d. This leads to the equivalent circuits shown in Fig. 34, which have the same form as the equivalent amplifier circuits of Figs. 1 and 88, Sec. 5, but differ in that the equivalent generator voltage and internal voltage are now determined by different factors.

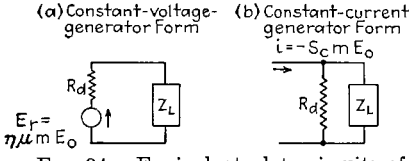


FIG. 34.—Equivalent plate circuits of plate detector. Note the similarity to the corresponding equivalent amplifier circuits of Fig. 1, Sec. 5.

The modulation-frequency voltage E_r that can be considered as giving rise to th modulation-frequency current is given by the equation

$$E_r = \eta\mu m E_0 \tag{18}$$

where m = degree of modulation of the applied signal.
E_0 = carrier amplitude of the applied signal.
μ = amplification factor of tube.

[1] The discussion of plate detection given here is based on a modification of analysis originated by Stuart Ballantine, Detection at High Signal Voltages: Plate Rectification with the High Vacuum Triode, *Proc. I.R.E.*, Vol. 17, p. 1153, July, 1929.

$$\eta = \frac{(\partial I_p/\partial E)_{E=E_0}}{\mu(\partial I_p/\partial E_p)_{E=E_0}} = \text{efficiency of rectification.}^1$$

E = applied signal (crest amplitude).

In practical plate detectors the rectification efficiency is of the order of 0.8 to 0.9, provided that the carrier is not too small.

The equivalent plate resistance R_d of the plate detector is the dynamic or incremental plate resistance of the tube as measured by an audio frequency when the radio-frequency carrier voltage is acting on the tube. Its numerical value is usually about three times the plate resistance of the same tube when operated as an ordinary amplifier.

The ratio $\eta\mu/R_d = S_c$ corresponds to the transconductance of amplifiers and is termed the *conversion transconductance.* It is of particular importance in the case of pentode tubes, in which case the output voltage developed across the load impedance Z_L in the plate circuit is[2]

$$\text{Output voltage} = mE_0S_cZ_L \tag{19}$$

where E_0 is the carrier voltage applied to the grid and m is the degree of modulation. The numerical value of S_c in ordinary tubes is approximately 0.3 times the transconductance of the same tube operated as an ordinary amplifier.

The audio-frequency coupling network of a plate detector is designed on the basis of the equivalent circuit of Fig. 34, exactly as though an audio amplifier is being dealt with, except that it is desirable to insert a radio-frequency filter between the plate and the output, as shown in Fig. 33. The bias voltage may be obtained from a fixed source or may be self-bias. In the latter case, the bias resistance should be such that the tube operates at cutoff when the rated carrier voltage is applied. Resistance coupling is normally used with pentode tubes, while transformer coupling is customarily with triodes. Resistance coupling is to be avoided in the latter case, since it introduces an a-c–d-c impedance ratio problem not present with transformer coupling or with the pentode tube.

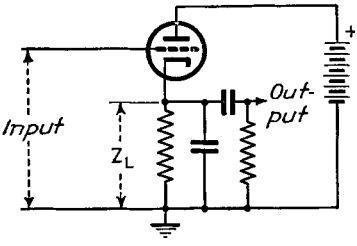


FIG. 35.—Circuit of infinite input-impedance detector.

Compared with a diode, plate detection has infinite input impedance, and gives more output voltage in proportion to applied signal. At the same time plate detection has greater distortion, is more critical as to the range of conditions giving satisfactory operation, and can develop only a limited available output voltage without excessive distortion. In particular, it is possible to obtain only about one-third as much output voltage from a given tube acting as a plate detector as when operating as a Class A audio amplifier. It is also much easier to obtain automatic volume control from diode detectors.

Infinite Input Impedance Detector.—This detector employs the circuit of Fig. 35. When a signal voltage is applied, a rectified plate current flows that builds up a voltage

[1] The efficiency of detection η can be measured by applying to the grid of the detector an alternating voltage (of any convenient frequency) equal to the carrier voltage E_0 for which η is desired and noting the d-c plate current. The crest value of this alternating grid voltage is then increased by a small increment ΔE_s, after which the plate voltage is altered by any amount ΔE_b such that the d-c plate current is the same value as before the addition of the grid-voltage increment. The ratio $\Delta E_b/\Delta E_s$ is then $\mu\eta$.

[2] Equation (19) assumes that the effective plate resistance R_d of the tube is much higher than the actual load resistance. If this is not true, then Z_L must be taken as the impedance formed by the actual load impedance when shunted by the effective plate resistance of the tube.

across the impedance Z_L that is only slightly less than the envelope amplitude. The end result is much the same as with diode detection, except that the input impedance is infinite. The peak modulation-frequency voltage obtainable across the output without overloading is quite large, approaching half the plate-supply voltage. The infinite input impedance detector is subject to negative peak clipping when the degree of modulation exceeds the a-c–d-c impedance ratio of the load. For proper operation, the infinite input impedance detector should be adjusted so that it operates reasonably close to cutoff when no signal is applied. It is also necessary that a reasonably large

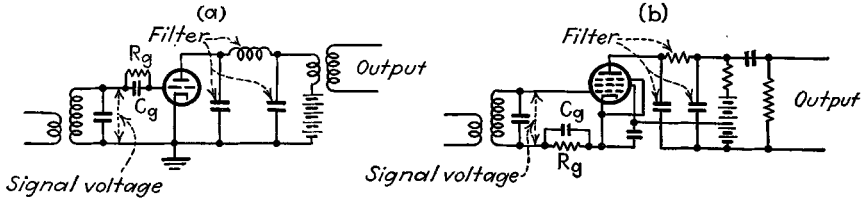


FIG. 36.—Circuits of grid-leak detector.

signal voltage be applied if distortion is to be minimized. A voltage for automatic-volume-control purposes cannot be readily obtained.

Grid-leak Power Detector.[1]—A typical circuit of a grid-leak power detector is shown in Fig. 36. Here rectification takes place in the grid circuit, with the grid and cathode functioning as a diode detector and with the rectified voltage being developed across the grid leak-condenser combination R_gC_g. The mechanism of operation is illustrated in Fig. 37, where the voltage existing between the grid and cathode of the tube

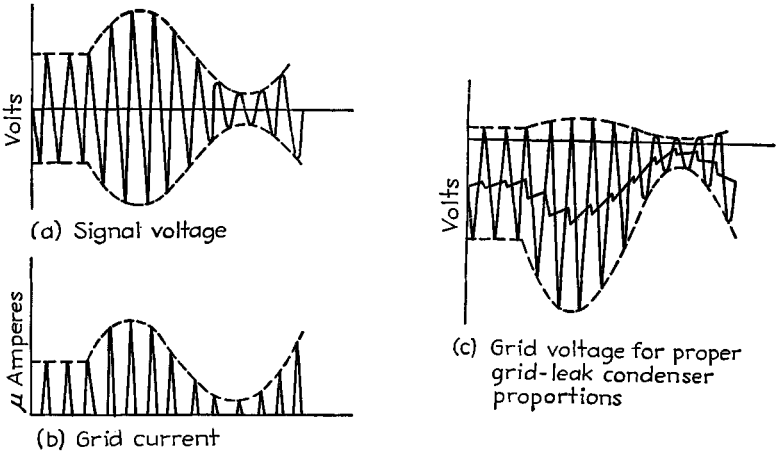


FIG. 37.—Details of action taking place in grid circuit of grid-leak power detector of Fig. 36.

is seen to consist of the radio-frequency signal voltage plus the voltage across R_g arising from the rectified grid current. This latter potential, which represents the desired output, is amplified in the plate circuit, where it can be separated and utilized as desired.

A grid-leak power detector corresponds roughly to a diode detector plus one stage of audio-frequency amplification. It has the disadvantage, however, that the maxi-

[1] F. E. Terman and N. R. Morgan, Some Properties of Grid Leak Power Detection, *Proc. I.R.E.*, Vol. 18, p. 2160, December, 1930.

mum carrier voltage that can be handled is of the order of 40 per cent of the peak alternating voltage that could be applied to the same tube operating as a Class A audio amplifier, and the modulation-frequency output voltage obtainable without excessive distortion is likewise about 40 per cent of the voltage obtainable from the same tube acting as an audio amplifier. These limitations arise from the fact that with applied signals larger than this, plate rectification becomes pronounced on the peaks of modulation, and combines with the effect of grid rectification to flatten the positive peaks.

A modified form of grid-leak power detector using the Wunderlich tube is illustrated in Fig. 38. Here diode type of rectification takes place between the grid and the cathode, with the use of a full-wave center-tapped connection. The advantage of this arrangement is that the radio-frequency signal voltage is applied to the two coplanar grids in phase opposition, and so does not affect the plate current. The result is that a Wunderlich tube in this connection is capable of developing an output voltage approximately the same as when the tube functions as a Class A amplifier with the two grids in parallel.[1]

Square-law (Weak-signal) Detection.[2]—The term *square law* is applied to a detector in which the rectified output is proportional to the square of the effective value of the applied signal. Any rectifier will act as a square-law device when the applied signal is sufficiently small, since over a limited range the voltage-current characteristic can always be approximated by a section of a parabola. For this reason, square-law detectors are sometimes referred to as weak-signal detectors.

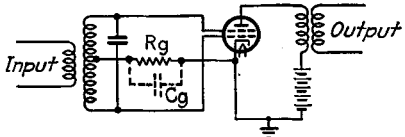


FIG. 38.—Circuit of grid-leak detector using Wunderlich tube.

The most common type of square-law detector is an amplifier so adjusted that the operating range falls on the curved portion of the grid-voltage plate-current characteristic curve. When a signal is applied to such an amplifier, the positive half cycles are amplified more than the negative half cycles, causing an increase in the d-c plate current and also introducing frequency components in the output current not present in the applied signal. Other types of square-law detectors that sometimes are used are crystal detectors (see below) and grid-leak detectors. The latter are similar to the grid-leak power detector except that the grid-leak resistance is chosen on a different basis.

The behavior of square-law detectors can be analyzed by the power-series method and the resulting equivalent circuit discussed in Par. 24, Sec. 5, and depends upon second-order effects. These effects can be briefly summarized as follows: There is a rectified d-c current proportional to the square of the effective value of the applied signal, together with second harmonics of each frequency component of the applied signal, and sum and difference frequencies formed by every possible combination of frequencies contained in this signal. The amplitude of each second-harmonic com-

[1] Additional discussion of the theory of this system of detection is given by F. E. Terman, Further Description of the Wunderlich Tube, *Radio Eng.*, Vol. 12, p. 25, May, 1932.

[2] The following is a list of references selected from the extensive literature on the detection of weak signals: Stuart Ballantine, Detection by Grid Rectification with the High-vacuum Triode, *Proc. I.R.E.*, Vol. 16, p. 593, May, 1928; John R. Carson, The Equivalent Circuit of the Vacuum-tube Modulator, *Proc. I.R.E.*, Vol. 9, p. 243, June, 1921; E. L. Chaffee and G. H. Browning, A Theoretical and Experimental Investigation of Detection for Small Signals, *Proc. I.R.E.*, Vol. 15, p. 113, February, 1927; F. M. Colebrook, The Rectification of Small Radio-frequency Potential Differences by Means of Triode Valves, *Exp. Wireless and Wireless Eng.*, Vol. 2, p. 946, December, 1925; F. E. Terman, Some Principles of Grid-leak Grid-condenser Detection, *Proc. I.R.E.*, Vol. 16, p. 1384, October, 1928; F. E. Terman and T. M. Googin, Detection Characteristics of Three-element Vacuum Tubes, *Proc. I.R.E.*, Vol. 17, p. 149, January, 1929.

ponent is proportional to the square of the amplitude of the corresponding fundamental part of the applied signal, while each combination frequency has an amplitude proportional to the product of the amplitudes of the two individual components involved. Thus, when the applied signal consists of a simple sine wave $E_s \sin 2\pi ft$, the output contains d-c and second-harmonic components proportional to E_s^2. With a heterodyne signal represented by $E_1 \sin 2\pi f_1 t + E_2 \sin 2\pi f_2 t$, the output will contain a rectified d-c current proportional to $\frac{(E_1^2 + E_2^2)}{2}$, sum and difference combinations having frequencies $f_1 + f_2$ and $f_1 - f_2$, respectively and amplitudes proportional to $E_1 E_2$ and second-harmonic terms of frequencies $2f_1$ and $2f_2$ with amplitudes proportional to E_1^2 and E_2^2, respectively. Finally, when the signal is a modulated wave, the fact that the rectified output is proportional to the square of the amplitude of the signal causes the output to be a distorted reproduction of the modulation envelope. With sinusoidal modulation, the distortion consists of a second harmonic of the modulation, and has a percentage $100 \times m/4$. For purposes of analysis, the modulated wave can be thought of as consisting of three frequency components, the carrier and two side bands. The output then contains the rectified direct-current term, second harmonics of each of the three components, two difference-frequency terms formed by the carrier combining with the two side bands to give the desired modulation-frequency output, and a second harmonic of the modulation frequency arising from the difference frequency formed by the two side-band components.

Square-law detectors are of importance for several reasons: First, any detector becomes a square-law device when the applied signal is sufficiently small. Also, square-law detectors give a distortionless difference frequency in the case of heterodyne signals. Finally, square-law detectors are widely used in measurement work because they have a simple and exact law of behavior, and do not generate high-order harmonics and combination frequencies such as exist when rectifiers are used that produce current pulses that are chopped-off sections of sine waves.

Crystal Detectors.[1]—In a crystal detector, use is made of the rectifying action that can be obtained with contacts involving certain crystals. Although a great variety of crystals can be used for this purpose, the most common are galena, silicon, iron pyrites, and carborundum. With galena, a fine pointed wire resting lightly on the crystal is employed. Silicon and iron pyrites can use a somewhat higher pressure, and so are more stable, but are at the same time less efficient as a rectifier of small signals. Carborundum will operate satisfactorily with still more pressure, and is very stable but correspondingly less sensitive.

Crystal rectifiers are effective in rectifying weak signals, provided that a sensitive point on the crystal is located. Crystal rectifiers have the disadvantage, however, of being quite delicate and variable. The only use of crystal detectors at the present time is in ultra-high-frequency work, because they will function at frequencies where transit time effects in diodes are excessive. In such applications silicon has been recommended as a good compromise between sensitivity and stability.[2]

Copper Oxide Detectors.—Copper oxide rectifiers find extensive use as detectors in certain special applications, notably as heterodyne detectors at carrier frequencies. In such applications, copper oxide elements are arranged in the same circuits used for copper oxide modulators (Fig. 24), with the signal applied to one pair of input terminals and the local oscillator to the other pair. The difference frequency then appears at the output as the lower side band.

[1] An excellent survey of crystal detectors is given by F. M. Colebrook, The Rectifying Detector, *Exp. Wireless and Wireless Eng.*, Vol. 2, p. 330, March, p. 394, April, p. 459, May, 1925.

[2] See G. C. Southworth and A. P. King, Metal Horns as Directive Receivers of Ultra-short Waves, *Proc. I.R.E.*, Vol. 27, p. 95, February, 1939, which also describes constructional details of a silicon detector for experimental use

Copper oxide rectifiers have been proposed as substitutes for diode detectors in ordinary radio receivers, but have not been used commercially for this purpose.[1]

11. Frequency Conversion by Heterodyne Action.—The carrier frequency of a signal can be changed by superimposing a local oscillation, and rectifying the resulting wave. The detector output then contains a wave having the same modulation as the original signal, but a carrier frequency that is the difference between the frequency of the signal and the superimposed oscillation.[2] Such an arrangement is known as a frequency converter, and the process of obtaining the new frequency is referred to as heterodyne action.

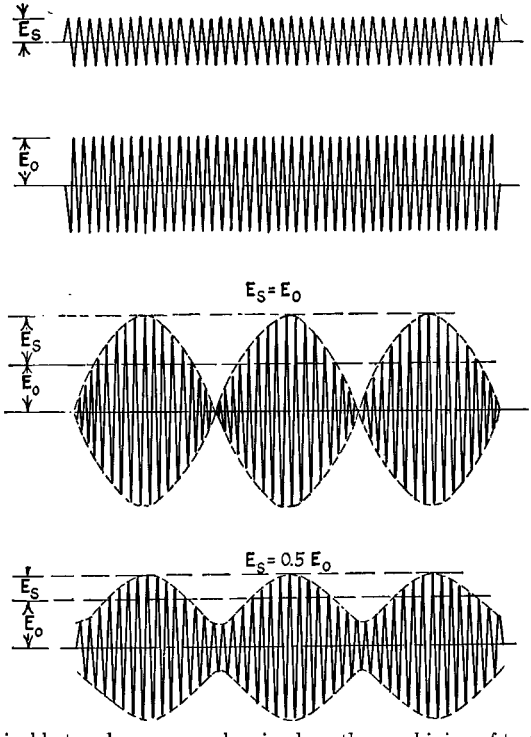


FIG. 39.—Typical heterodyne waves showing how the combining of two waves of slightly different frequencies results in a wave having an envelope that pulsates in amplitude at the difference frequency of the component waves.

Envelope of Heterodyne Signal.—The envelope of a wave formed by superimposing two sine waves pulsates at the difference frequency, is as shown in Fig. 39. The equation of the envelope is

$$\left.\begin{matrix}\text{Instantaneous amplitude} \\ \text{of envelope}\end{matrix}\right\} = \sqrt{E_s^2 + E_0^2 + 2E_sE_0 \cos \omega t} \qquad (20)$$

where E_s and E_0 are the crest amplitudes of the two waves and $\omega/2\pi$ is their difference in frequency. This envelope can be considered as having a constant component upon

[1] L. O. Grondahl and W. P. Place, Copper-oxide Rectifier Used for Radio Detection and Automatic Volume Control, *Proc. I.R.E.*, Vol. 20, p. 1599, October, 1932.

[2] There is also in the detector output another component modulated in the same manner as the signal but having a carrier frequency that is the sum of the signal and superimposed frequency. This sum frequency component is only very occasionally made use of, however.

which is superimposed a difference-frequency component and various harmonics of the difference frequency, according to the equation[1]

$$\left.\begin{matrix}\text{Envelope} \\ \text{amplitude}\end{matrix}\right\} = a_0 E_0(1 + m_1 \cos \omega t - m_2 \cos 2\omega t + m_3 \cos 3\omega t \ldots) \quad (21)$$

where E_0 is the amplitude of the larger oscillation and a_0, m_1, m_2, etc. are coefficients depending on E_s/E_0, as given in Fig. 40.

This equation shows that when one signal is much larger than the other, the fluctuations in amplitude of the envelope are substantially sinusoidal, proportional to the

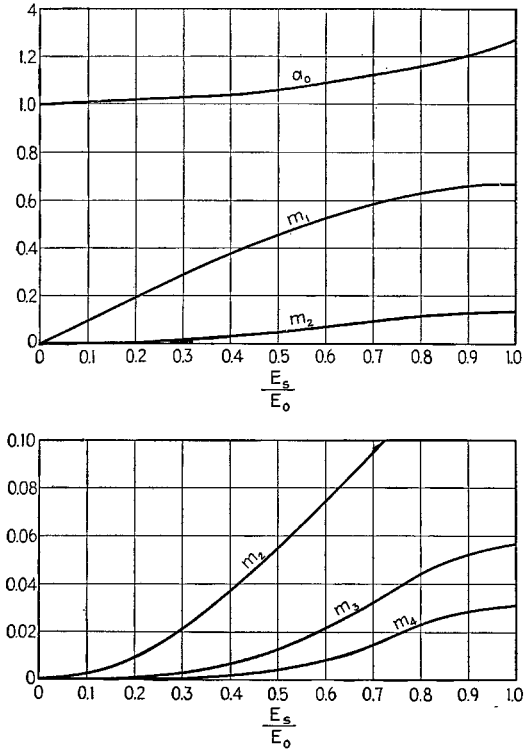


FIG. 40.—Coefficients for use in Eq. (21) giving the amplitudes of the various frequency components contained in the envelope of a heterodyne wave.

weaker signal, and substantially independent of the amplitude of the stronger signal. When the two signals are equal, the magnitude of the difference frequency output is affected somewhat by the amplitude of both signals, and the distortion in the envelope reaches a maximum of 20 per cent second harmonic.

In its most general form, the equation of the envelope of a heterodyne signal consisting of a series of superimposed waves

$$e = E_0 \sin \omega_0 t + E_1 \sin (\omega_1 t + \varphi_1) + E_2 \sin (\omega_2 t + \varphi_2) + \ldots$$

[1] F. M. Colebrook, The Frequency Analysis of the Heterodyne Envelope, *Exp. Wireless and Wireless Eng.*, Vol. 9, p. 195, April, 1932; F. E. Terman, Linear Detection of Heterodyne Signals, *Electronics*, Vol. 1, p. 386, November, 1930.

is given by the equation

$$\text{Envelope} = [E_0^2 + E_1^2 + E_2^2 + \ldots + 2E_0E_1 \cos[(\omega_1 - \omega_0)t + \varphi_1] + 2E_0E_2 \cos[(\omega_2 - \omega_0)t + \varphi_2] + \ldots + 2E_1E_2 \cos[(\omega_1 - \omega_2)t + (\varphi_1 - \varphi_2)] + \ldots]^{1/2} \quad (21a)$$

Linear and Square-law Detection of Heterodyne Waves.—When a heterodyne signal consisting of two sine waves is rectified by a square-law detector, the detector output always contains a sinusoidal difference frequency component free of harmonics, irrespective of the relative amplitudes of the signal and superimposed oscillation. A linear detector, on the other hand, gives a rectified output proportional to the envelope of the wave rather than to the square of the envelope, and so develops output components according to Eq. (21). The difference-frequency output will then be substantially free of harmonics only if the superimposed oscillation is much larger than the signal that is to have its frequency changed. If the two have equal amplitudes, there is 20 per cent second-harmonic distortion.

In the general case of a heterodyne signal containing many frequency components, a square-law detector will develop in its output a component for every sum and difference frequency combination that can possibly be formed from the frequency components involved in the wave applied to the detector. The relative amplitude of each of these sum and difference frequencies is the product of the amplitudes of the corresponding components contained in the applied signal.

With linear detectors, the situation is somewhat more complicated. In general, the output will contain frequency components corresponding to the Fourier series analysis of the envelope of the applied signal. In addition, there will also be sum and difference frequencies and various combination frequencies representing higher order combinations than simple sums and differences of the applied radio frequencies.

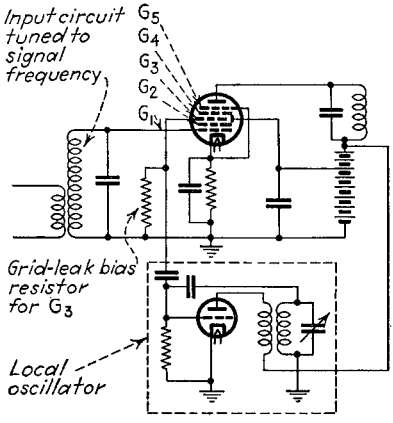


FIG. 41.—Circuit of typical pentagrid-mixer tube.

12. Special Converter (or Mixer) Tubes.[1]—The demodulator used in a frequency converting system can be any type of detector or modulator. Extensive use is made of special tubes, however, of which the most important are the pentagrid mixer, pentagrid converter, and triode-hexode converter.

Pentagrid Mixer (6L7).[2]—The pentagrid mixer is a special five-grid tube employed in the circuit arrangement illustrated in Fig. 41. The inner grid G_1 functions as an ordinary control grid, to which the signal voltage is applied. The second and fourth grids, G_2 and G_4, are connected together and function as a screen grid. The third grid, G_3, is biased negatively and has voltage applied to it from a separate local oscillator. Grid G_5 is an ordinary suppressor grid, giving the plate circuit of the tube a characteristic corresponding to that of a pentode. In operation, G_1 controls the space current drawn from the cathode in accordance with the signal voltage, and the oscillator acting on G_3 serves as a switch that allows these electrons to pass on to

[1] An excellent discussion of converters is given by E. W. Herold, The Operation of Frequency Converters and Mixers for Superheterodyne Reception, *Proc. I.R.E.*, Vol. 30, p. 8, February, 1942.

[2] C. F. Nesslage, E. W. Herold, and W. A. Harris, A New Tube for Use in Superheterodyne Frequency Conversion Systems, *Proc. I.R.E.*, Vol. 24, p. 207, February, 1936.

the anode or causes them to be returned to the screen region, according to whether the oscillator voltage is positive or negative, respectively. The result is equivalent to modulating the oscillator voltage upon the signal frequency, with the result that a difference frequency (or side-band) current is developed in the anode circuit.

The pentagrid mixer tube has the merit of giving very low interaction between the local oscillator and the signal-frequency circuits, even at very high frequencies. It has the disadvantage of requiring a separate oscillator.

Pentagrid Converter.—The pentagrid converter combines the local oscillator and demodulator in a single tube, as shown in Fig. 42. Here G_1 and G_2 function as the control grid and anode of an electron-coupled local oscillator. G_3 and G_5 function as screen grids, while G_4 is biased negatively and has the signal voltage applied to it. The action of the oscillator causes pulses of plate current to arrive at G_4 at the peak of each oscillator cycle. A virtual cathode is then formed in front of G_4, and the signal voltage on this grid controls the number of electrons that the plate is able to draw from this virtual cathode which appears at each positive peak of the oscillator voltage and then disappears between pulses. The current actually arriving at the plate is hence modulated by both oscillator and signal voltages, giving a difference-frequency or lower side-band component corresponding to the desired output frequency.

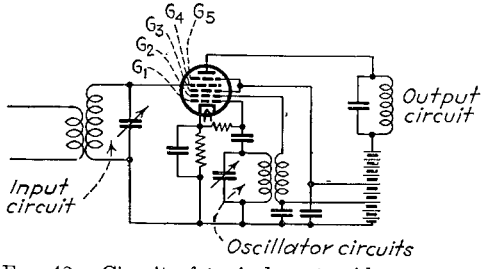


FIG. 42.—Circuit of typical pentagrid-converter tube.

The pentagrid converter tube has the advantage of simplicity in that one tube serves as both converter and oscillator. At the same time, it has three fundamental defects:[1] *First*, the transconductance of the oscillator section is relatively low, which introduces design difficulties in the oscillator circuit at high frequencies. *Second*, the oscillator frequency will change slightly with variation in the bias on the signal grid. This occurs because the bias on the signal grid affects the number of electrons returning from the virtual cathode to the vicinity of the actual cathode, and so alters the effective capacity between the oscillator grid G_1 and the space charge near the cathode. This variation in frequency with bias on the signal grid is sufficient to be troublesome at high frequencies. *Third*, the coupling between G_4 and the pulsating virtual cathode between it and G_3 causes oscillator-frequency currents to flow through the signal-frequency circuits associated with grid G_4. This leads to undesirable operating characteristics, as, for example, lowered efficiency of conversion, overloading of the signal grid, etc. The effect is particularly important at high frequencies where percentage difference in oscillator and signal frequency is small.

The effect of the signal-grid bias on oscillator frequency can be eliminated by modifying the design of the pentagrid converter as shown in Fig. 43. In this arrangement, typified by the 6SA7 tube, a pair of collector plates is mounted at the side rods of the screen-grid structure G_2 as shown, and the geometry is such that most of the electrons returning toward the cathode from the virtual cathode between G_3 and G_2 are deflected sidewise and intercepted by these end plates. The result is that the voltages applied to the signal grid G_3 have little effect upon the space charge around the cathode, and hence on the oscillator frequency.

[1] There is also, in addition to these defects, a certain amount of coupling between oscillator and signal circuits, even at low frequencies, as a result of the fact that the voltage on the signal grid affects the current flowing to the oscillator grid G_2. See Paul W. Klipsch, Suppression of Interlocking in First Detector Circuits, *Proc. I.R.E.*, Vol. 22, p. 699, June, 1934.

Triode-hexode Mixer.[1]—The electrode arrangement of a triode-hexode mixer, together with a typical circuit arrangement, is shown in Fig. 44. Here G_1 and the anode A_1 are on one side of a flat cathode and function in a triode oscillator circuit. On the other side of the same cathode there is an extension of the oscillator grid G_1, followed by a screen grid G_2 that surrounds a signal grid G_3, with a second (or mixer)

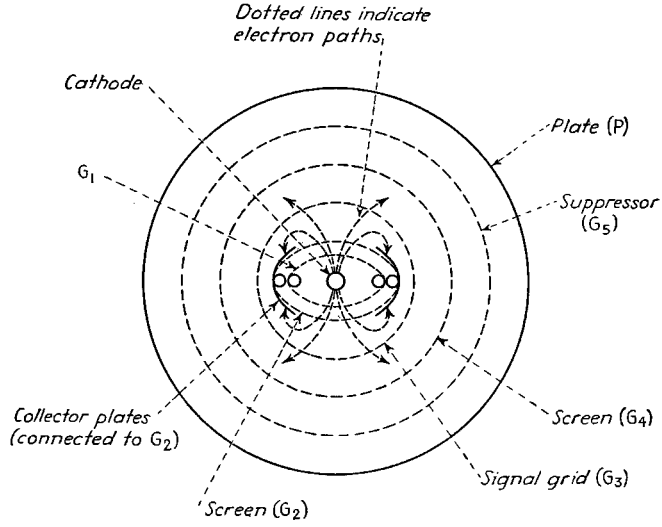


FIG. 43.—Pentagrid converter provided with collector plates that serve to prevent electrons returned by grid G_4 from reaching the cathode region.

anode A_2. This side of the tube operates in much the same manner as a pentagrid converter, with the voltage that the oscillator section develops on the grid G_1 causing pulses of space current that form a virtual cathode on the cathode side of the signal grid G_3. The triode-hexode mixer has the advantage over the pentagrid tube in that the oscillator can be designed for higher transconductance. Furthermore, the

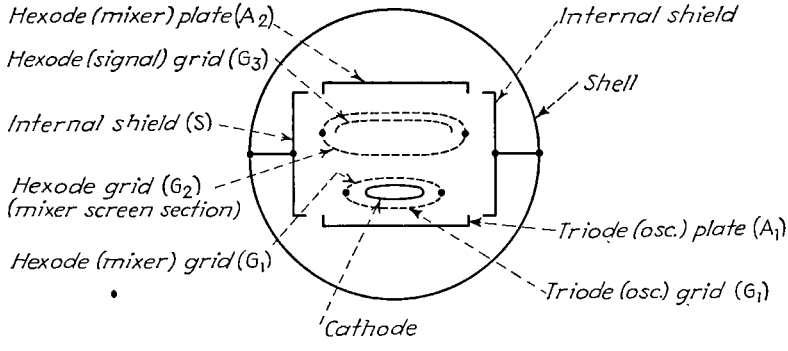


FIG. 44.—Electrode arrangement of triode-hexode mixer.

shield S connected to the cathode eliminates all interaction between the oscillator and triode sections, so that the oscillator frequency is independent of bias on the signal grid or tuning in the signal-grid circuits.

[1] E. W. Herold, W. A. Harris, and T. J. Henry, A New Converter Tube for All-wave Receivers, *R.C.A. Rev.*, Vol. 3, p. 67, July, 1938.

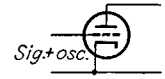
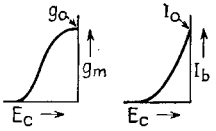
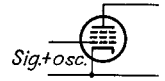
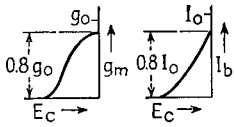
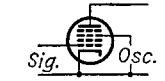
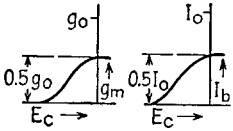
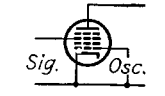
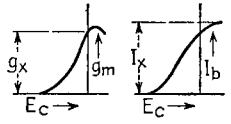
13. Gain of Converter Systems.—The ratio of the difference-frequency voltage developed in the output of a converter to the signal voltage applied to the input is termed the converter gain, and is[1]

$$\text{Converter gain} = S_c Z_L \tag{22}$$

where S_c equals conversion transconductance (defined in the same way as in plate rectification) and Z_L is the load impedance in the plate circuit to the difference frequency.

The value of conversion gain depends upon the variation of transconductance of the tube over the cycle of the local oscillator. With optimum operating conditions,

TABLE 1.—CONVERSION GAIN AND NOISE IN CONVERTER SYSTEMS

Type of detector	Typical characteristics	Approx. conversion transconductance	Approx. equivalent noise resistance
1. Triode plate detection 		$0.28g_0$	$13/g_0$
2. Pentode plate detection 		$0.23g_0$	$\dfrac{15 + 21\dfrac{I_0}{g_0}}{g_0}$
3. Pentagrid mixer 		$0.14g_0$	$\dfrac{120\dfrac{I_0}{g_0}}{g_0}$
4. Pentagrid converter and triode-hexode mixer 		$0.28g_x$	$\dfrac{57\dfrac{I_x}{g_x}}{g_x}$

g_m = transconductance, mhos.
g_0 = transconductance, mhos, of tube with zero bias on G_1 (inner grid), when tube is connected to act as triode (*i.e.*, all grids but G_1 connected to anode).
I_b = plate current, amp.
I_0 = total space current with zero bias on G_1, amp.
g_x = transconductance in case 4 when oscillator voltage is at the positive crest and signal grid (G_3) has normal operating bias, mhos (g_x approximates g_0 in practical cases).
I_x = plate current under conditions for which g_x is defined, amp.
E_c = voltage on oscillator grid.

[1] Equation (22) assumes that the load resistance is small compared with the effective plate resistance of the converter tube. This assumption is very closely realized in all arrangements where the plate current is substantially independent of plate voltage. In case this is not true, then Z_L in Eq. (22) should be taken as the resistance formed by the actual load impedance shunted by the equivalent plate resistance of the tube defined as in plate rectification.

values of conversion transconductance for different types of converters are listed in Table 1.[1] The gain that can be expected in a converter tube is roughly 0.3 times the gain of the same tube operating under normal conditions as a conventional amplifier.

The conversion gain can be reduced by increasing the negative bias on the signal grid. To facilitate such control, the signal grid of most converter tubes is provided with a variable-mu characteristic.

14. Noise in Converter Tubes.[1]—The tube of a converter stage produces noise voltages in its output as a result of random fluctuations in the space current of the tube and random variations in the way the space current divides between the plate and the other positive electrodes. The fundamental factors involved are the same as discussed in Par. 5, Sec. 4, but the situation is complicated by the fact that the space current, and hence the noise, vary over the oscillator cycle. There is also a tendency for virtual cathodes to exist in pentagrid arrangements, which likewise affects the noise.

Theoretical formulas based on reasonable assumptions have been derived for the noise that can be expected with various types of converters. These are given in Table 1, where the noise is expressed in terms of the resistance that, when connected between signal grid and cathode, would give a thermal agitation noise that after frequency conversion would equal the noise actually produced within the tube. The results of these calculations represent the lower limit of noise that can be expected, and, as shown in Table 2, agree satisfactorily with observed noise levels.

Examination of Tables 1 and 2 shows that the triode plate-rectifier type of converter has the least noise and the maximum conversion transconductance, other things equal. The pentode plate detector is next, with the special mixer tubes used in broadcast receivers distinctly inferior from the point of view of noise level. It is also to be noted that a high ratio of transconductance to space current reduces the noise level, so that television tubes such as the 1852 have much lower noise output than ordinary tubes.

TABLE 2.—EQUIVALENT GRID-RESISTANCE REPRESENTING CONVERTER TUBE NOISE*

Tube	Equivalent grid noise resistance	
	Measured	Calculated
6SA7 pentagrid converter	210,000	220,000
6L7 pentagrid mixer	210,000	230,000
6J5 triode	5,800	3,700
1853 television pentode	13,000	18,000
1852 television pentode	3,000	3,400

* From E. W. Herold, l.c.

The equivalent grid resistance representing the noise of a converter tube is always greater than the corresponding equivalent resistance representing the noise generated in the same tube functioning as an ordinary Class A amplifier (see Par. 5, Sec. 4). Thus the 1852 tube has equivalent noise resistance of 710 ohms when operating as a pentode amplifier, as compared with 3,000 ohms when functioning as a pentode plate-rectifier type of converter. Accordingly, when the internal impedance of the signal voltage applied to the converter input does not greatly exceed the equivalent noise resistance of the converter, there is an advantage in employing a stage of radio-frequency amplification before conversion is attempted. In this way, the main source

[1] E. W. Herold, Superheterodyne Converter System Considerations in Television Receivers, *R.C.A. Rev.*, Vol. 4, p. 324, January, 1940; Operation of Frequency Converters and Mixers for Superheterodyne Reception, *Proc. I.R.E.*, Vol. 30, p. 84, February, 1942.

of noise will be thermal agitation in the input circuit, instead of tube noise. Examination of the numerical values in Table 2 shows that the only practical case when r-f amplification will not reduce the noise level appreciably is when triode or pentode plate-rectifier converters are used. In such arrangements, practical values of internal impedance of the voltage applied to the converter tube can be made considerably greater than the equivalent noise resistance, except possibly in television systems.

15. Miscellaneous. *Regenerative Detectors.*—In grid-leak and plate detectors, regeneration is sometimes introduced by coupling from the plate circuit of the detector back to the radio-frequency input. Typical regenerative circuits of this character are shown in Fig. 45, and make use of the fact that there are amplified signal currents in the plate circuit along with the rectified output. The effect of regeneration introduced in this way is equivalent to a reduction of the radio-frequency resistance of the resonant input circuit. This increases the radio-frequency voltage applied to the detector, and also produces a corresponding increase in selectivity.

When regeneration is carried as far as possible without oscillations being produced, the resulting increase in amplitude is very great for extremely weak signals, and is less, although still considerable, for strong signals. This behavior arises from the

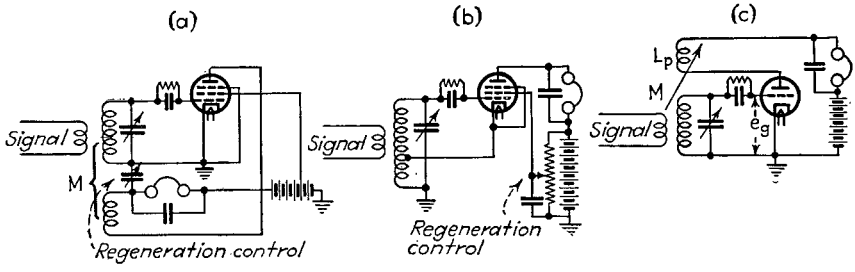


FIG. 45.—Circuits for regenerative and oscillating detectors.

third-order curvature of the tube characteristics, which causes the effective transconductance of the tube to be slightly less for large signals than for small, thus reducing the regeneration for large signals. Analysis shows that when the regeneration is made as large as possible without oscillations in the absence of applied signal, then the gain from regeneration is inversely proportional to the two-thirds power of the signal voltage and directly proportional to the response obtained when no regeneration is present.[1] The maximum possible regeneration without oscillation is termed *critical regeneration.*

Although regeneration represents an inexpensive means of increasing the radio-frequency amplification, it increases the selectivity excessively, requires critical adjustments that depend on the signal frequency, and introduces oscillations with consequent interference and whistles whenever the regeneration is accidentally made excessive. Since the development of satisfactory radio-frequency amplifiers, regenerative detectors have found relatively little application.

Oscillating Detectors.—The oscillating detector is a heterodyne detector, commonly of the grid-leak type, in which the detector tube generates the local oscillation as well as functioning as the detector. The circuit arrangements employed are the same as shown in Fig. 45 for regenerative detectors, but the regeneration is increased to the point where oscillations are generated.

Oscillating detectors are widely used in the reception of continuous-wave telegraph signals to obtain an audio beat note. At all except the lowest radio frequencies,

[1] See Balth. van der Pol, The Effect of Regeneration on the Received Signal Strength, *Proc. I.R.E.*, Vol. 17, p. 339, February, 1929.

the percentage difference between the frequency of the incoming signal and the local oscillation is so small that when the resonant input circuit is adjusted to give the required oscillator frequency, it is also for all practical purposes tuned to the desired incoming signal.

When the regeneration control of an oscillating detector is adjusted so that the oscillations are barely able to sustain themselves, the signal applied to the grid of the tube undergoes a large increase in amplitude through regenerative action. This comes about because with weak oscillations there is a critical balance between the energy loss in the resonant circuit and the energy supplied to this circuit by the regeneration. The superposition of a small signal voltage upsets this balance, causing large increases and decreases in the amplitude of the oscillation according to whether the incoming signal momentarily aids or opposes, respectively, the oscillations. The end result, as far as the incoming signal is concerned, is as though one had critical regeneration, although the adjustments are not critical, since the oscillations automatically assume an amplitude that picks out the critical condition, and maintains it with complete stability. The effective amplification of the incoming signal obtainable in this way is very large, values of the order of 15,000 having been observed under favorable conditions.[1] The amount of amplification of this character obtainable is greatest when the oscillations are just able to maintain themselves and when the circuits initially have low losses.

When the oscillating detector employs a triode tube, trouble is sometimes encountered from a sustained audio-frequency howl that occurs when the adjustment is such that oscillations are just barely maintained. This is known as "threshold" or "fringe" howl, and is to be avoided, since it occurs under conditions for which the oscillating detector is most sensitive. Threshold howl may occur in grid-leak arrangements when the audio-frequency load impedance in the plate circuit of the triode is inductive, and it can be cured either by using resistance coupling or by shunting the inductive load with a sufficiently low resistance. Threshold howl does not ordinarily occur when pentode tubes are employed.[2]

Distortion of Modulation Envelope by Phase Shifts and Alteration of Relative Magnitudes of Different Components.—A modulated wave consists of a carrier having an amplitude at least equal to the peak instantaneous amplitude formed by the sum of the side-band components, and of upper and lower side bands of equal magnitude and possessing symmetrical phase relations with respect to the carrier. If the side bands are of unequal amplitude, or if they have a phase dissymmetry with respect to the carrier, or if the sum of their peak instantaneous amplitude exceeds the carrier, then distortion of the modulation envelope will result. Examples of such distortion are shown in Fig. 46. At (*a*) are shown the envelopes obtained as phase dissymmetry is introduced by shifting the carrier by $\alpha°$ from its normal position with respect to the side bands, while (*b*) gives the effect of varying the carrier amplitude. It will be noted that large distortions in the envelope are possible.

The envelope for any case can be calculated by first writing the modulated wave as the sum of the carrier plus side-band frequencies. Thus, for the case of simple sinusoidal modulation

$$e = A \cos(\omega t + \alpha) + B \cos[(\omega + v)t + \beta] + C \cos[(\omega - v)t - \gamma] \qquad (23)$$

where $\omega = 2\pi$ times carrier frequency.
$v = 2\pi$ times modulating frequency.
α, β, γ = phase angles of various frequency components.
A, B, C = magnitudes of various components.

[1] H. A. Robinson, Regenerative Detectors, *QST*, Vol. 17, p. 26, February, 1933.

[2] The mechanism of threshold howl is described by L. S. B. Alder, Threshold Howl in Reaction Receivers, *Exp. Wireless and Wireless Eng.*, Vol. 7, p. 197, April, 1930.

The envelope is then obtained by substituting in Eq. (21a). An alternative method of obtaining the envelope that is frequently used in the case of simple sinusoidal modulation is by the graphical addition of vectors, as shown in Fig. 47.[1] Here the carrier and various side bands are drawn as vectors with appropriate amplitudes, and

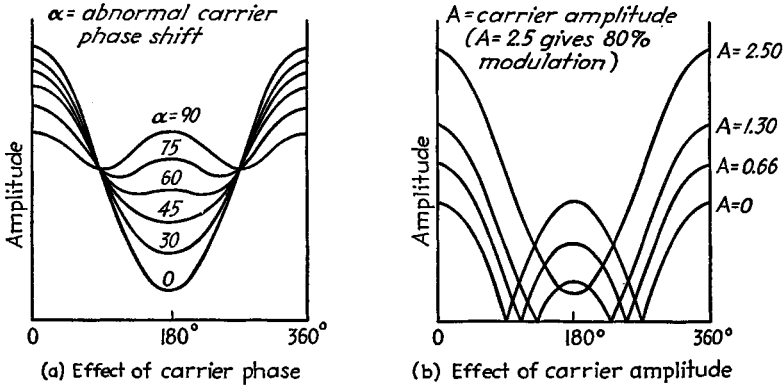


FIG. 46.—Envelope shapes for one modulation cycle, showing effect of varying carrier phase and carrier amplitude (amplitude of individual side-band components constant in all cases).

with relative phases corresponding to the relative positions at $t = 0$. Then vector B, representing the upper side band, is assumed to rotate counterclockwise, relative to A at a frequency $v/2\pi$, and vector C rotates clockwise at the same rate. The amplitude of the modulation envelope at any value of time t_0 is then obtained by adding the

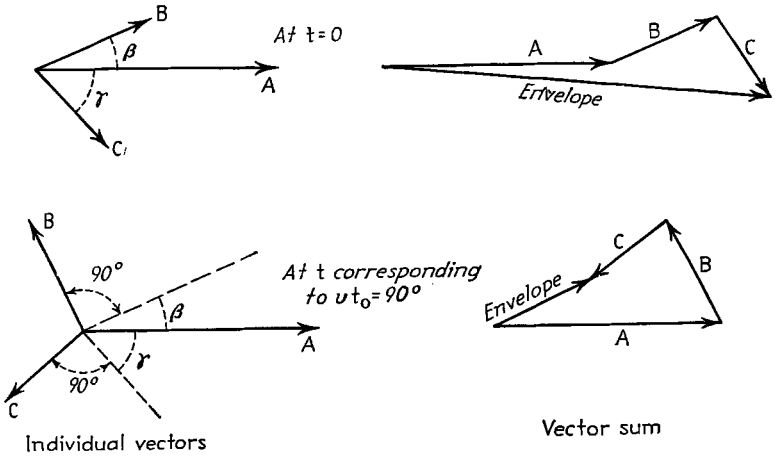


FIG. 47.—Vector construction for obtaining amplitude of envelope of a wave containing three frequency components.

three vectors when the angle of rotation is vt_0 radians. Several such additions are indicated in Fig. 47 for different values of t_0.

The relations between carrier and side band that exist in a normal amplitude-modulated wave can be upset by selective fading or by resonant circuits of the trans-

[1] Further details are given by E. A. Laport, Characteristics of Amplitude Modulated Waves, *R.C.A. Rev.*, Vol. 1, p. 26, April, 1937.

mitter or receiver that fail to be adjusted to give a symmetrical response curve centered at the carrier frequency.

Behavior of Detectors When Applied Signal Consists of Two Modulated Waves.—Four cases of this type are to be distinguished:

1. Difference between carrier frequencies so high as to be an inaudible frequency.
2. Difference between carrier frequencies audible.
3. Difference between carrier frequencies too small to be audible.
4. Carrier frequencies identical.

When the carrier frequencies differ sufficiently that their difference is inaudible, then if a linear detector is used and the radio-frequency envelope of one signal is considerably greater than the envelope of the other signal, the modulation of the weaker of the two signals tends to be suppressed insofar as the detector output is concerned. The factor by which the modulation output of the weaker signal is reduced below the output voltage it develops in the absence of the strong signal is

$$\frac{1}{2}\,\frac{\text{weak-signal amplitude}}{\text{strong-signal amplitude}}$$

This suppression of the weaker modulation is equivalent to an increase in the effective selectivity of the system supplying voltage to the detector, and represents an important property of a linear detector.[1] This effect is present only if the detector is linear, and is entirely absent in the case of square-law detectors.

When the difference between the carrier frequencies of the desired and undesired signals is in the audible range, a number of undesired components of audible frequency appear in the output of both square-law and linear detectors.[2] The component having the largest amplitude is the difference frequency between the two carriers, and this is the most disturbing component when it exceeds 100 to 200 cycles.

When the frequencies of the two carriers differ by less than about 50 cycles, the most disturbing components in the output are the difference frequencies formed by the carrier of the strong or desired signal heterodyning with the side-band frequencies of the weaker or undesired signal to produce what can be termed *side-band noise.* When the carrier frequencies differ by only a few cycles a second, this side-band noise gives rise to the characteristic flutter commonly heard when two or more broadcast stations are simultaneously transmitting on approximately the same frequency. When the two carriers are both weak enough so that there is a background of noise, the noise level in the detector output will also flutter at a frequency corresponding to the difference frequency between the two carriers, and may in some cases produce an effect more annoying than the side-band noise.

In the event that the two signals come from stations that have their carriers synchronized and that are modulated with identical programs, the detector output will not ordinarily represent a distortionless reproduction of the original modulation unless one of the carrier amplitudes is much weaker than the other. This is because the relative phase with which the two carriers and their respective side-band components combine in the detector input depends upon the distance to the transmitter,

[1] For further information and details of the method of analyzing this phenomenon, see R. T. Beatty, Apparent Demodulation of a Weak Station by a Stronger One, *Exp. Wireless and Wireless Eng.*, Vol. 5, p. 300, June, 1928; S. Butterworth, Note on the Apparent Demodulation of a Weak Station by a Stronger One, *Exp. Wireless and Wireless Eng.*, Vol. 6, p. 619, November, 1929; E. V. Appleton and D. Boohariwalla, The Mutual Interference of Wireless Signals in Simultaneous Detection, *Exp. Wireless and Wireless Eng.*, Vol. 9, p. 136, March, 1932.

[2] Further discussion on this and the following cases is contained in the following papers: Charles B. Aiken, Theory of the Detection of Two Modulated Waves by a Linear Rectifier, *Proc. I.R.E.*, Vol. 21, p. 601, April, 1933; A Study of Reception from Synchronized Broadcast Stations, *Proc. I.R.E.*, Vol. 21, p. 1265, September, 1933; The Effect of Background Noise in Shared Channel Broadcast, *Bell System Tech. Jour.*, Vol. 13, p. 333, July, 1934. See also Hans Roder, Superposition of Two Modulated Radio Frequencies, *Proc. I.R.E.*, Vol. 20, p. 1962, December, 1932.

time differences in the transmission of the program, and the side-band frequency involved. The result is that, when the carrier amplitudes are of approximately the same order of magnitude, certain side-band frequencies will tend to cancel while others will be reinforced; furthermore, at certain locations, the two carriers will also tend to cancel and thereby distort the envelope of the wave applied to the detector input. In order for distortion of this sort to be imperceptible under the worst practical conditions, it is necessary that one carrier have an amplitude at least four times that of the other carrier.

Detection of Single Side-band Signals.—Detection of single side-band systems is accomplished by superimposing upon the single side band a local oscillation having the same frequency as that of the suppressed carrier, and then rectifying the combination wave that results. The frequencies contained in the single side-band heterodyne with the superimposed local oscillation to give beat frequencies corresponding to the original modulation frequencies. When the amplitude of the local oscillation is large compared with the amplitude of the single side-band signal, the output obtained upon rectification will be a substantially distortionless reproduction of the original modulation, provided that the local oscillation has the correct frequency. It makes no difference in this case whether a linear or square-law detector is used. However, if the local oscillation is not much larger than the single side-band signal, then the detector output will contain distortion currents of importance in the form of difference frequencies formed by the frequency components within the single side band. Distortion will also result if the local oscillator does not have exactly the correct frequency. In the case of speech, it has been found that the intelligibility is not seriously impaired, however, as long as the local oscillator is within 10 cycles of the correct frequency, even though there may be some loss in naturalness.

FREQUENCY MODULATION

16. Frequency and Phase-modulated Waves.[1]—In frequency modulation the intelligence is transmitted by varying the instantaneous frequency $d(\theta/2\pi)/dt$, where ϑ is the instantaneous angular velocity. When the variation of angular velocity is sinusoidal, corresponding to sinusoidal frequency modulation, the equation of the wave is

$$e = E_0 \sin (\omega t + m_f \sin pt) \tag{24}$$

where E_0 = crest amplitude of the wave.

$\omega = 2\pi$ times frequency in the absence of modulation.

$p = 2\pi$ times the frequency (or rate) at which the frequency of the wave is varied (*i.e.*, the modulating frequency).

$$m_f = \frac{\left\{\begin{array}{l}\text{variation of radio frequency away from}\\ \text{the mean frequency}\end{array}\right\}}{\text{modulating frequency}} = \text{modulation index.}$$

Components of Frequency-modulated Wave.[2]—The wave represented by Eq. (24) can be expressed in the form

$$\begin{aligned} e = E_0\{J_0(m_f) \sin \omega t &+ J_1(m_f)[\sin (\omega + p)t - \sin (\omega - p)t] \\ &+ J_2(m_f)[\sin (\omega + 2p)t + \sin (\omega - 2p)t] \\ &+ J_3(m_f)[\sin (\omega + 3p)t - \sin (\omega - 3pt)] \\ &+ \cdots\cdots\cdots\cdots \end{aligned} \tag{25}$$

[1] A very readable general discussion on the subject is given by W. L. Everitt, Frequency Modulation. *Trans. A.I.E.E.*, Vol. 59, p. 613, November, 1940.

[2] John R. Carson, Notes on the Theory of Modulation, *Proc. I.R.E.*, Vol. 10, p. 57, February, 1922; Balth. van der Pol, Frequency Modulation, *Proc. I.R.E.*, Vol. 18, p. 1194, July, 1930; Hans Roder, Amplitude, Phase, and Frequency Modulation, *Proc. I.R.E.*, Vol. 19, p. 2145, December, 1931.

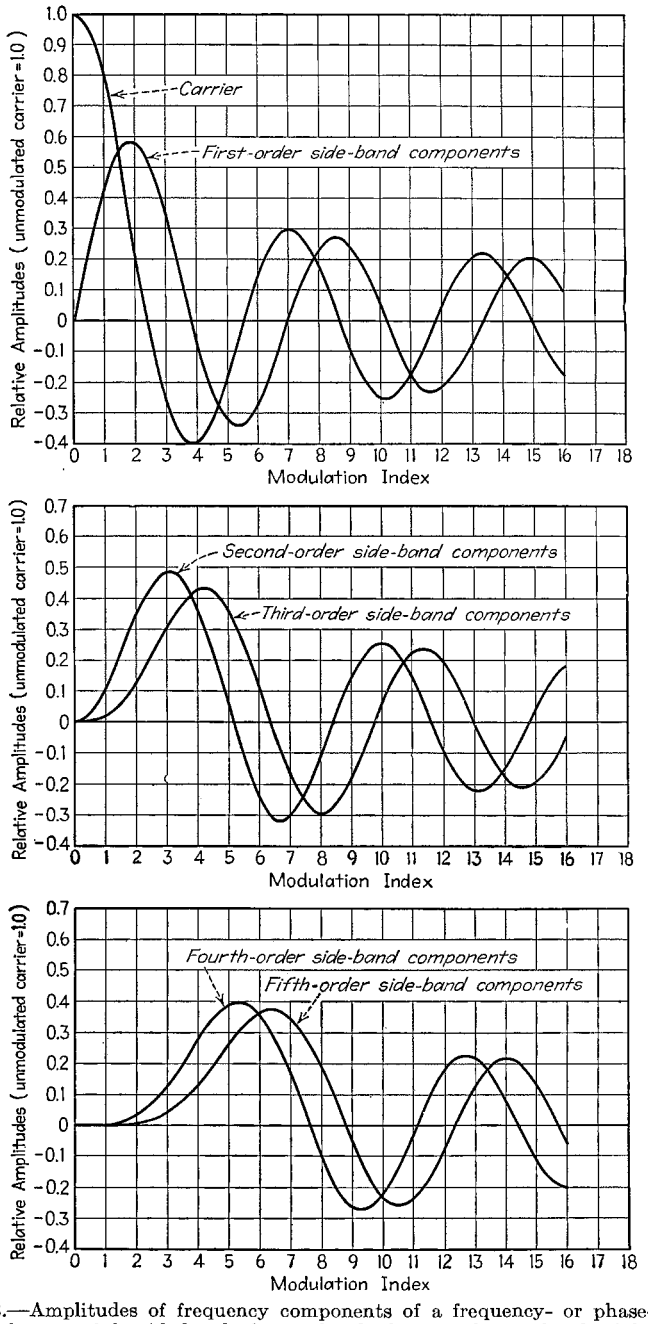


FIG. 48.—Amplitudes of frequency components of a frequency- or phase-modulated wave. In the case of the side bands the amplitude shown is the amplitude of the individual side-band component and not of the pair of companion side bands taken together.

where $J_n(m_f)$ means the Bessel function of the first kind and the nth order, with argument m_f. The frequencies present in Eq. (25) are ω, $\omega \pm p$, $\omega \pm 2p$, etc. The first component is analogous to the carrier frequency of an amplitude-modulated wave, but differs in that the amplitude now depends upon the modulation index. The second pair of frequencies corresponds to the side bands existing with amplitude modulation, but differs in that the phase relation to the carrier has been shifted 90°, and also the amplitude of these components is not directly proportional to the modulating force except when the modulation index is very small. The remaining components represent higher order side bands that are not present in amplitude modulation and that, when of appreciable amplitude, make the frequency range required to accommodate the frequency-modulated wave greater than when the same intelligence is transmitted by amplitude modulation.

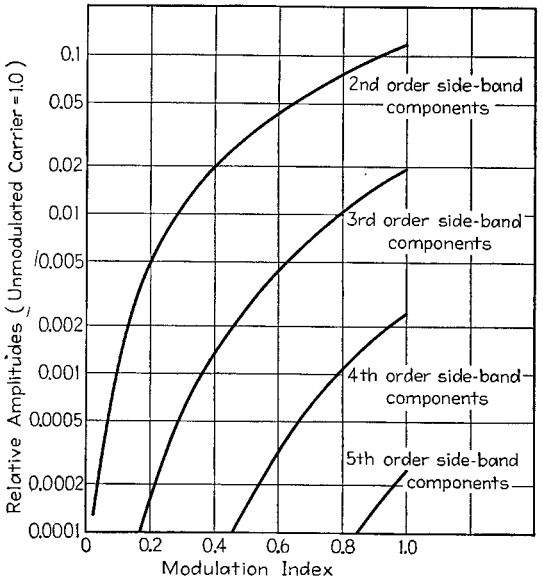
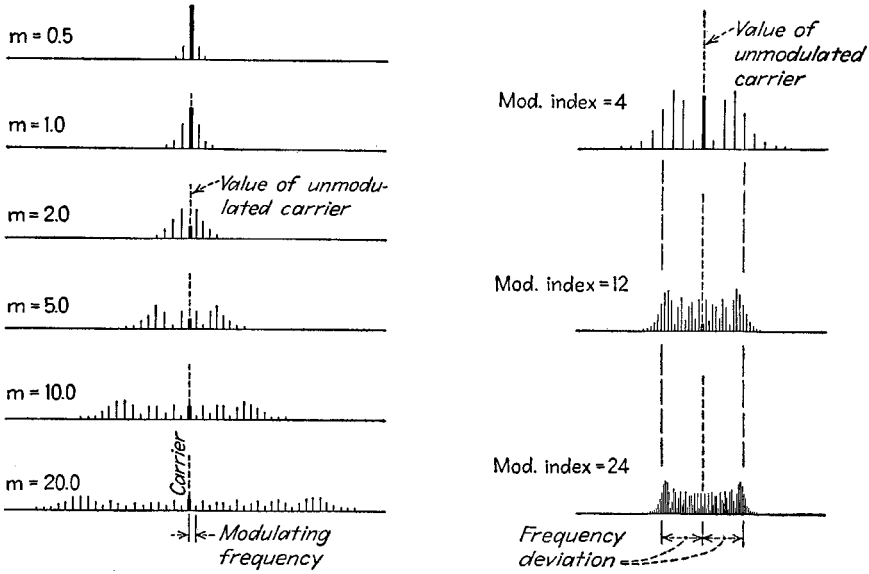


FIG. 49.—Continuation of Fig. 48, showing to an enlarged scale the amplitude of the higher order side-band components when the modulation index is small.

The amplitude of the different frequency components depends upon the modulation index m_f, and can either be calculated with the aid of a table of Bessel's functions or be obtained from Figs. 48 and 49. The character of the frequency spectra obtained under different conditions with frequency modulation is illustrated in Fig. 50. When the modulation index is less than 0.5, *i.e.*, when the frequency deviation is less than half the modulating frequency, the second and higher order side-band components are relatively small, and the frequency band required to accommodate the essential part of the signal is the same as in amplitude modulation. On the other hand, when the modulation index m exceeds unity, *i.e.*, when the variation in radio frequency away from its mean value is greater than the modulating frequency, there are important frequency components extending a distance on each side of the mean or carrier frequency by an amount approximately $(m + 1)$ times the modulation frequency. The spacing of the frequencies within the essential band under these conditions is equal to the modulating frequency, and so is less the lower the modulating frequency.

Frequency-modulated Wave with Complex Modulation.[1]—When the instantaneous frequency is varied in a more complex manner than that corresponding to sinusoidal modulation, the frequency spectrum becomes very complicated. Thus, where there are two modulating frequencies $p_1/2\pi$ and $p_2/2\pi$, each having modulation indices m_1 and m_2, frequency components as shown in Table 3 are present. The side frequencies present include not only those that would be obtained with each modulation acting separately, but also include combination frequencies with amplitudes equal to the products of Bessel's functions having orders equal to the order of the side-band frequencies involved.



(a) Frequency spectra with increasing frequency deviation and constant modulating frequency

(b) Frequency spectra with constant frequency deviation but variable modulating frequency

FIG. 50.—Frequency spectra of frequency-modulated waves under various conditions.

Although complex modulation increases the number of frequency components present in a frequency-modulated wave in an alarming way, it does not widen the frequency band occupied by the energy of the wave. This is because the modulation index of each modulating component will in general be reduced when the total modulation is divided between two or more components. In fact, when the maximum frequency deviation produced by a complex modulation is fixed, the energy of the wave tends, if anything, to be concentrated in a narrower band than with a simple sinusoidal modulation having the same frequency deviation. The total band required will accordingly approximate either twice the maximum frequency deviation away from the mean or twice the modulating frequency, whichever is greatest.

Phase Modulation.—In phase modulation, intelligence is transmitted by varying the phase of the transmitted wave. In the case of sinusoidal phase modulation, the phase of the wave is

$$\text{Phase} = \varphi = \varphi_0 + m_p \sin pt \tag{26}$$

[1] Murray G. Crosby, Carrier and Side Frequency Relations with Multi-tone Frequency or Phase Modulation, *R.C.A. Rev.*, Vol. 3, p. 103, July, 1938.

TABLE 3.—FREQUENCY COMPONENTS OF WAVE THAT IS FREQUENCY-MODULATED SIMULTANEOUSLY BY TWO MODULATING FREQUENCIES

Type of component	Angular velocity of component	Relative amplitude
Carrier	ω	$J_0(m_1)J_0(m_2)$
Simple side bands of p_1	$\omega \pm p_1$	$J_1(m_1)J_0(m_2)$
	$\omega \pm 2p_1$	$J_2(m_1)J_0(m_2)$
	.....	.....
Simple side bands of p_2	$\omega \pm p_2$	$J_0(m_1)J_1(m_2)$
	$\omega \pm 2p_2$	$J_0(m_1)J_2(m_2)$
	.....	.....
Combination frequencies	$\omega + p_1 \pm p_2$	$J_1(m_1)J_1(m_2)$
	$\omega - p_1 \pm p_2$	$J_1(m_1)J_1(m_2)$
	.....	.....
	$\omega + p_1 \pm 2p_2$	$J_1(m_1)J_2(m_2)$
	$\omega - p_1 \pm 2p_2$	$J_1(m_1)J_2(m_2)$
	.....	.....
	$\omega + 2p_1 \pm p_2$	$J_2(m_1)J_1(m_2)$
	$\omega - 2p_1 \pm p_2$	$J_2(m_1)J_1(m_2)$
	.....	.....

where φ_0 is the phase of the wave in the absence of modulation, m_p (termed the modulation index) is the phase shift in radians produced by the modulation, and $p/2\pi$ is the modulating frequency. The equation of a phase-modulated wave is

$$e = E_0 \sin (wt + \varphi_0 + m_p \sin pt) \tag{27}$$

Equations (24) and (27) are the same except that the modulation index is now defined differently. The frequency components contained in a phase-modulated wave are hence the same as for a frequency-modulated wave with the same modulation index. In particular, if the modulation index m_p is less than 0.5, *i.e.*, maximum phase deviation less than one-half radian, there is only a single pair of side bands of appreciable amplitude, but for each additional radian in the modulation index, an additional order of side bands of appreciable amplitude will be produced.

The essential difference between phase and frequency modulation is that in the former the modulation index is independent of modulating frequency, while in the latter it is inversely proportional to the modulating frequency. This means that with a fixed modulating voltage, the band required to accommodate a phase-modulated signal is proportional to the modulating frequency, while with frequency modulation the band occupied is independent of modulating frequency, except when the modulation index is small.

The Effect of Frequency Multiplication on Phase- and Frequency-modulated Waves.— When a phase- or frequency-modulated wave is passed through a harmonic generator, the effect is to increase the modulation index by a factor equal to the frequency multiplication obtained. It is accordingly possible to generate the phase- or frequency-modulated wave at a moderate frequency, keeping the modulation index small in order to obtain linearity, and then increasing the modulation index to a desired large value by use of a suitable number of harmonic generators.

17. Production of Frequency- and Phase-modulated Waves. *Phase Modulation.* A number of methods that have been proposed for producing phase-modulated waves are shown schematically in Fig. 51. At (*a*), an amplitude-modulated wave is combined with a much larger unmodulated carrier that has its phase shifted 90° from the phase of the carrier of the amplitude-modulated wave. The sum of the two waves is then passed through a limiter to remove the residual amplitude fluctu-

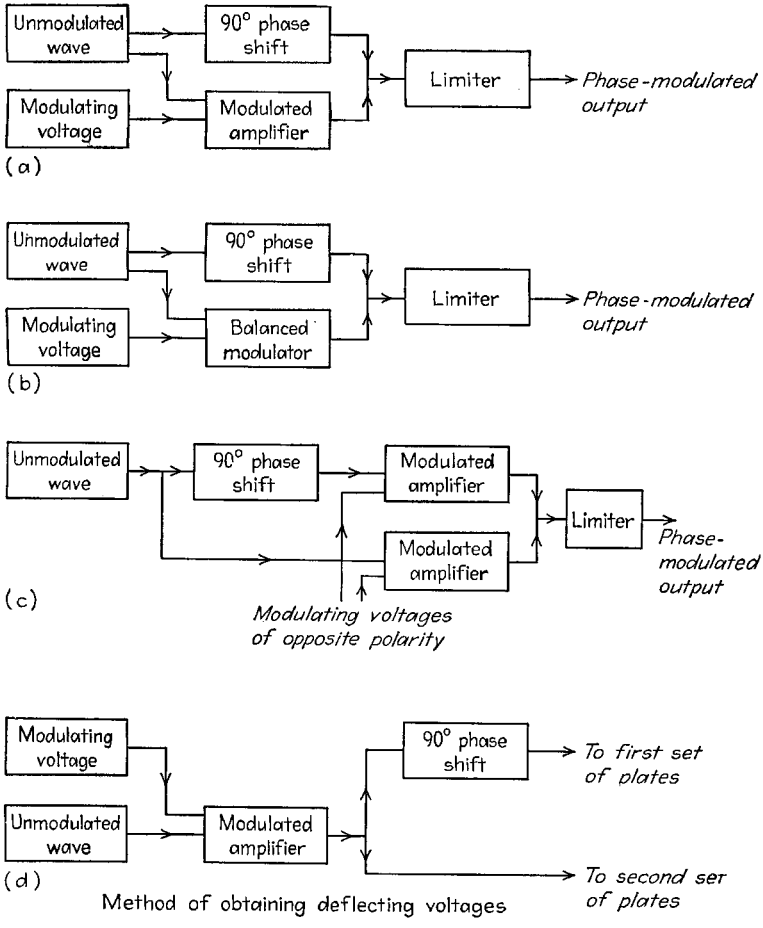


FIG. 51.—Methods of generating a phase-modulated wave.

ations. By making the amplitude of the modulated wave small compared with that of the unmodulated carrier, the phase shift produced by the amplitude modulation will be almost exactly proportional to the modulating voltage.

A modification of this arrangement is shown at Fig. 51*b*, where the output of a balanced modulator, instead of an amplitude-modulated wave, is combined with the unmodulated carrier. This arrangement has the advantage that it gives twice as much phase shift for the same degree of linearity as does the arrangement of Fig. 51*a*.[1]

In the phase-modulation system of Fig. 51*c*, two Class C amplifiers are modulated with voltages of opposite polarity, and are excited with radio-frequency voltages differing 90° in phase. The combined output of the system can be made constant, but will vary in phase during modulation as a result of the fact that the relative contributions from the two amplifiers are varied by the modulation.[2]

In the arrangement of Fig. 51*d*, two equal radio-frequency voltages having a phase difference of 90° are applied to the two sets of deflecting plates of a cathode-ray tube, to give a circular trace. By amplitude-modulating these deflecting voltages, the diameter of the circle will vary in accordance with the modulation. The anode target of the cathode-ray tube is divided into two parts as shown, by two lines that follow the equation[3]

$$\text{Radius} = a\theta \tag{28}$$

where a is a constant. As the diameter of the circular trace varies, the phase of the a-c voltage developed between the two parts of the target by the beam current will vary linearly with the amplitude of the circle. The maximum phase shift possible is $\pm 360°$ for each revolution of the spiral, so that by employing a number of turns a large phase shift can be obtained.[4]

Any system of frequency modulation will produce a phase-modulated wave by the use of a network that makes the modulating potential proportional to modulating frequency.

Frequency Modulation.—There are two principal methods used to produce frequency-modulated waves. The first of these makes use of a system of phase modulation, but before applying the modulating voltage to the system, transmits this voltage through a network that gives an output inversely proportional to frequency.[5] This gives a modulation index that varies with modulating frequency in the way required for frequency modulation, so that the output of the phase-modulating system then gives the desired frequency-modulated wave. The arrangement has the advantage that all oscillators involved can be crystal-controlled, with resulting high stability of the mean frequency. The disadvantage of the arrangement is that the modulation index that can be directly obtained and still maintain linearity of modulation is small, so that a number of stages of harmonic generation must be used if a large modulation

[1] In this arrangement, the distortion is largely third harmonic, and will be less than 5 per cent if the phase shift does not exceed $\pm 25°$. See D. L. Jaffe, Armstrong's Frequency Modulator, *Proc. I.R.E.*, Vol. 26, p. 475, April, 1938; D. I. Lawson, Frequency Modulation, *Wireless Eng.*, Vol. 17, p. 388, September, 1940.

By predistorting the amplitude of the modulating voltage amplitude to accentuate the peaks, a somewhat larger phase shift is permissible. See Samuel Sabaroff, System of Phase and Frequency Modulation, *Comms.* Vol. 20, p. 11, October, 1940.

[2] Murray G. Crosby, Communication by Phase Modulation, *Proc. I.R.E.*, Vol. 27, p. 126, February, 1939.

[3] This curve is known as an Archimedean spiral.

[4] For further details see Robert E. Shelby, A Cathode-ray Frequency Modulation Generator, *Electronics*, Vol. 13, p. 14, February, 1940.

[5] Such a system, based on the phase modulator of Fig. 51*b*, is described by Edwin H. Armstrong, A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, *Proc. I.R.E.*, Vol. 24, p. 689, May, 1936.

index is to be obtained ultimately. The relationship between the phase shift m_p in radians, and the equivalent frequency deviation Δf away from the mean frequency is

$$\Delta f = m_p \times \text{modulating frequency.} \tag{29}$$

The second common system of frequency modulation utilizes the modulating voltage to control the frequency of the generated oscillations. The frequency control is usually accomplished[1] with the aid of a reactance tube acting in shunt with the oscillator circuit, as used automatic frequency-control systems (see Par. 5, Sec. 9). Application of the modulating voltage to the grid of the reactance tube will vary the reactive current drawn by the plate electrode of the tube, and so will affect the oscillator frequency. This method of frequency modulation is capable of giving a large modulation index, but has the disadvantage that the mean frequency cannot be stabilized accurately without the use of a relatively complicated automatic-frequency-control system that is referred to an auxiliary crystal oscillator.[2]

Details involved in the production of frequency-modulated waves in commercial equipment are discussed in Par. 8, Sec. 9.

18. Detection of Frequency- and Phase-modulated Signals.—Frequency- and phase-modulated signals are received in the same way, since for a given modulation index there is no difference between them. The fact that the modulation index varies differently with modulating frequency in the two cases can be taken into account by an auxiliary network having a suitable variation of transmission with frequency. Thus, if a phase-modulated wave is to be received on a frequency-modulated receiver, one must modify the modulation-frequency components obtained in the receiver output by a network that will make the amplitudes inversely proportional to modulation frequency.

Demodulation of a frequency-modulated wave is accomplished by distorting the frequency spectrum of the wave in a manner that causes the envelope to fluctuate in accordance with the intelligence involved. The wave can then be rectified by an ordinary detector. The usual detecting systems are also so arranged that they will not respond to amplitude modulation contained in the original wave.

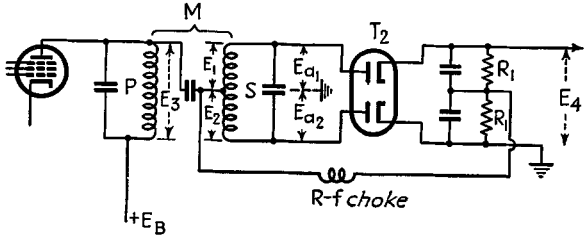
Practical Frequency-modulation Detectors.—Two types of frequency-modulation detectors are in common use. The first, shown in Fig. 52, makes use of a frequency discriminator similar to that employed in ordinary automatic-frequency-control systems (see Par. 5, Sec. 9). Here a double diode T_2 is so connected that the output voltage E_4 is the difference of the rectified outputs of the individual diodes. Excitation is obtained from an arrangement of two coupled circuits P and S resonant at the same frequency and connected as shown. The rectified output voltage E_4 of such an arrangement varies with frequency in the manner shown in Fig. 52*b*. This characteristic results from the fact that at resonance the voltage developing across S differs 90° in phase from the voltage across P. The potential E_{a1} and E_{a2} acting on the anodes are then of equal magnitude, as shown by the first vector diagrams of Fig. 52*c*. However, for frequencies slightly off resonance, the secondary phase is greater or less than 90°, with the result that the r-f voltage applied to one anode is now larger and the voltage to the other anode smaller, as seen from the second

[1] An alternative method of varying the oscillator frequency consists in shunting a portion of the tank circuit of the oscillator by the input of an eighth wave-length line and then varying the resistance load at the end of the line in accordance with the modulating voltage with the aid of a vacuum tube. In such an arrangement, the input reactance of the line varies with the modulation, and the equivalent shunt resistance at the input terminals stays substantially constant. See Austin V. Eastman and Earl D. Scott, Transmission Lines as Frequency Modulators, *Proc. I.R.E.*, Vol. 22, p. 878, July, 1934.

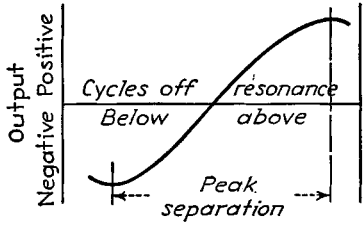
[2] Details relative to reactance-tube modulators which are important in design of such devices are discussed by Murray G. Crosby, Reactance-tube Frequency Modulators, *QST*, June, 1940, also *R.C.... Rev.*, Vol. 5, p. 89, July, 1940; C. F. Shaeffer, Frequency Modulator, *Proc. I.R.E.*, Vol. 28, p. 66, February, 1940.

and third vector diagrams of Fig. 52c. The result is a differential output E_4 that varies almost linearly over a considerable range of frequencies as shown.

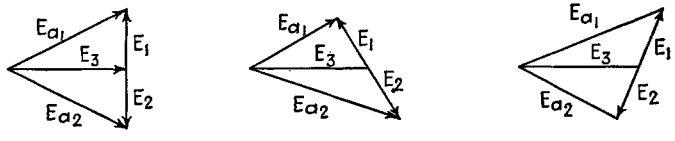
In the discriminator of Fig. 52a, the diode produces a loading effect that is equivalent to shunting a resistance $R_1/4\eta$ across the primary P and R_1/η across the full secondary S, where η is the efficiency of rectification. The voltages E_3, E_1, E_2, E_{a1}, and E_{a2} can be calculated on the basis of ordinary coupled-circuit theory by including the effect of this loading on the circuit Q's. When the coupling between primary and secondary does not exceed 0.75 times the critical coupling, the separation between



(a) Circuit



(b) Variation of output with frequency



(c) Vector diagrams

FIG. 52.—Discriminator and detector arrangement for receiving a frequency-modulated wave without responding to amplitude modulation.

the peaks of the discriminator characteristic in Fig. 52b is approximately equal to (carrier frequency)/Q_s, where Q_s is the effective Q of the secondary, if the diode loading is taken into account.[1] For greater coupling, the peaks are somewhat farther apart than given by this calculation. The useful frequency range of the discriminator is from one-third to two-thirds of the separation between peaks, with the exact value depending on the degree of linearity that must be maintained over the operating range.

The second common system for demodulating frequency-modulated signals is shown in Fig. 53. Here the discriminator consists of a primary L_1C_1 tuned to the

[1] Hans Roder, Theory of the Discriminator Circuit for Automatic Frequency Control, *Proc. I.R.E.*, Vol. 26, p. 590, May, 1938.

carrier frequency, and two secondary circuits L_2C_2 and L_3C_3 that are tuned to frequencies slightly above and slightly below the carrier, respectively. With this arrangement each diode develops across its load impedance R_2 a voltage that varies with frequency as shown in Fig. 53*b*, while the output E_4 is the difference of these, and so varies as shown. The result is substantially the same as in Fig. 52. For proper operation, the primary circuit Q should satisfy the empirical relation

$$Q_p = \frac{\text{carrier frequency}}{3 \times (\text{frequency deviation from mean})} \tag{30}$$

The Q's of the secondary circuits should be twice that of the primary. With these

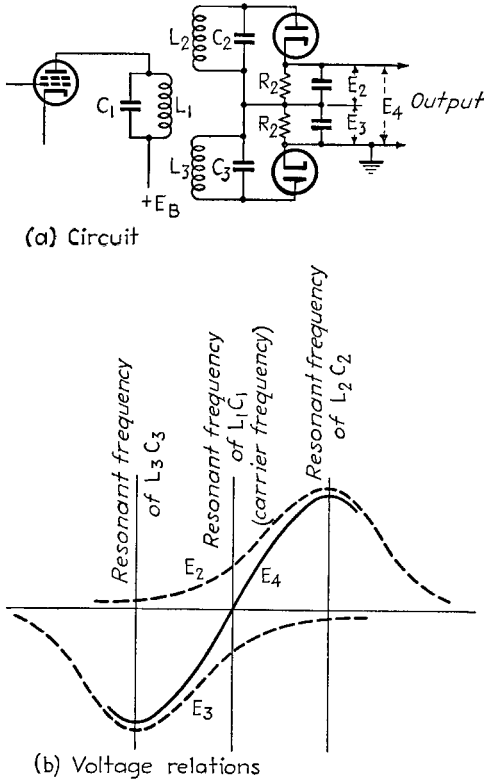


FIG. 53.—Balanced detector with detuned input circuits for receiving a frequency-modulated wave while not responding to amplitude modulation.

proportions, the output will be a linear function of frequency over the range of deviation specified.[1]

The arrangements of Figs. 52 and 53 do not respond to amplitude modulation of the incoming carrier, because the effects of amplitude variation are balanced out in the output. As a consequence, these systems have an inherent discrimination against noise and interfering signals that are weaker than the incoming frequency-modulated carrier. This is discussed in Par. 10, Sec. 9.

Miscellaneous Methods of Demodulating Frequency-modulated Signals.—Demodulation of frequency-modulated signals may be accomplished in a number of ways in

[1] M. G. Crosby, Reactance Tube Frequency Modulators. *R.C.A. Rev.*, Vol. 5, p. 89, July, 1940.

addition to those shown in Figs. 52 and 53. A simple method is to detune an ordinary receiver slightly so that the edge of the selectivity curve is at the frequency of the incoming carrier. This causes the components of one side band to be reduced in amplitude in proportion to their difference from the carrier frequency, while the components of the other side band are correspondingly increased in amplitude. The result is that the envelope amplitude is no longer constant, but fluctuates in accordance with the modulation. When the selectivity curve has a substantially linear amplitude and phase characteristic over the frequency range that includes all the important components of the frequency-modulated wave, the amplitude modulation of the envelope that results is a distortionless reproduction of the modulation of the original signal.[1]

Another method of demodulation consists in passing a portion of the frequency-modulated wave through an artificial line that gives the same delay time for all components, and then combines this delayed wave with a portion of the undelayed wave. With constant delay time, the phase shift will be different for the different side-band frequencies, so that when the delayed and undelayed currents are combined they will add in a manner that depends on frequency. Rectification of the resultant by a square-law detector will give an output that can be made to have relatively little distortion by properly selecting the operating conditions.[2]

Demodulation of a frequency-modulated wave can also be carried out by separating the carrier by a sharply tuned selective circuit, as, for example, a crystal filter. This carrier is then amplified, shifted in phase by 90°, and recombined with the frequency-modulated wave. The side-band components now have the phase required to produce amplitude modulation, and the wave can be rectified by an ordinary detector.[3]

[1] Hans Roder, Effects of Tuned Circuits upon a Frequency Modulated Signal, *Proc. I.R.E.*, Vol. 25, p. 1617, December, 1937. Additional discussion of the subject is given by Victor J. Andrew, The Reception of Frequency-modulated Radio Signals, *Proc. I.R.E.*, Vol. 20, p. 835, May, 1932; J. G. Chaffee, The Detection of Frequency-modulated Waves, *Proc. I.R.E.*, Vol. 23, p. 517, May, 1935.

[2] This system was suggested to the author by Dr. J. R. Woodyard. Still another method credited to Dr. Woodyard is described in his paper Application of the Autosynchronized Oscillator to Frequency Demodulation, *Proc. I.R.E.*, Vol. 25, p. 612, May, 1937.

[3] For further details, see M. G. Crosby, Communication by Phase Modnlation, *Proc. I.R.E.*, Vol. 27, p. 126, February, 1939.

SECTION 8

POWER-SUPPLY SYSTEMS

RECTIFIER-FILTER SYSTEMS FOR ANODE POWER

1. Rectifiers.—Most rectifiers used for anode-power-supply systems are either high-vacuum diodes or mercury-vapor hot-cathode types. In addition, mercury-arc and ignitron rectifiers are sometimes used for high power, and copper oxide, selenium, and cold-cathode gaseous rectifiers are occasionally employed for developing small amounts of anode power.

High-vacuum Rectifiers.—The high-vacuum thermionic rectifier is a diode tube consisting of cathode and anode electrodes. The rectifying action results from the fact that when the anode is positive with respect to the cathode, electrons flow to the anode, while if the anode is negative the tube becomes nonconducting. Further details relative to diode rectifier tubes are given in Par. 6, Sec. 4.

The important characteristics of a high-vacuum diode rectifier are the allowable peak and average plate currents, the allowable inverse voltage, and the voltage drop in the tube. The peak plate current represents the maximum electron emission that the cathode can be counted upon to supply during the entire useful life of the tube and still maintain a complete space charge around the cathode. The allowable average current represents the d-c output current that can be continuously carried without overheating the tube or in other ways reducing the life below the normal expectancy. Since the rectifier never allows the current to flow more than half the time, the average plate current will never exceed one-half of the peak current, and tubes are often designed for a still lower ratio of average to peak current in order to provide for transient peaks.

The maximum allowable inverse plate voltage is the largest negative voltage that may be applied to the plate with safety, and determines the direct-current voltage that can be obtained from the rectifier tube. The exact relationship between d-c output voltage and allowable inverse voltage depends upon the rectifier circuit used. The d-c potential will always be less than the inverse voltage (except for voltage-multiplying circuits), and in certain rectifier connections is only $1/\pi$ times as great.

The voltage drop is an important factor in determining the voltage regulation of the rectifier-filter system, and also fixes the amount of plate dissipation that must be provided in the tube design. The voltage drop is determined by the current, the cathode area, and the cathode-anode spacing. In tubes where the inverse voltage rating is of the order of 2,000 volts and less, very small spacings can be employed, with a resulting voltage drop that is remarkably low in proportion to the current.

High-vacuum rectifiers are used almost to the complete exclusion of other types for develping d-c output voltages up to about 400 volts at currents up to several hundred milliamperes. Compared with the hot-cathode mercury-vapor type, they have the advantage in low-power applications of being able to stand more abuse, of having voltage drops that are little if any greater, and of not producing radio-frequency transients.

Hot-cathode Mercury-vapor Rectifier Tubes.[1]—The hot-cathode mercury-vapor rectifier is a diode tube containing mercury vapor in equilibrium with liquid mercury.

[1] A discussion of hot-cathode mercury-vapor rectifier tubes used in radio work is given by H. C. Steiner and H. T. Maser, Hot-cathode Mercury-vapor Rectifier Tubes, *Proc. I.R.E.*, p. 67, January, 1930.

When the plate of such a tube is positive, the mercury vapor is ionized by collision with the electrons, producing positive ions that neutralize the space charge of the electrons emitted from the cathode. The result is that the full cathode emission current can be drawn to the anode with a potential drop in the tube of the order of only 10 to 15 volts, even when the anode-cathode spacing is large, and some of the electrons are emitted from recesses in a heater-type cathode. The low voltage drop existing in a hot-cathode mercury-vapor tube permits the use of oxide-coated cathodes with their high thermionic efficiency, since positive-ion bombardment of the cathode will not cause serious deterioration until the drop in the tube is of the order of 22 volts. Further details relative to the characteristics of hot-cathode mercury-vapor tubes are given in Par. 17, Sec. 4.

The important characteristics of hot-cathode mercury-vapor rectifier tubes are the allowable peak and average plate currents, and the maximum safe inverse plate voltage.[1] The allowable peak plate current represents the maximum instantaneous current that it is permissible for the tube to carry in normal operation. In order to ensure satisfactory operation during the entire life of the tube, the total electron emission provided by the cathode must be much greater than the rated peak current. The safe average current is determined by the allowable heating of the plate and the current that can be drawn continuously from the cathode, and still realize a reasonable life expectancy. The maximum allowable inverse plate voltage is the greatest negative potential that can exist on the plate while substantially complete freedom from arc-back within the tube is maintained. The value of this voltage is influenced greatly by the construction of the tube, tends to be greater with filament as compared with heater cathodes, and is less as the pressure of the mercury vapor is increased (*i.e.*, as the temperature of the condensed mercury is increased).

Hot-cathode mercury-vapor rectifier tubes require considerably more care in operation than do high-vacuum rectifiers. The cathode must be brought to normal operating temperature before plate voltage is applied. The temperature of the condensed mercury must also lie within the proper range, for if the temperature is too great there is danger of arc-back, and if it is too small, there will not be sufficient ionization to give space-charge neutralization. It is also necessary that hot-cathode mercury-vapor tubes be carefully protected from even momentary short circuits and overloads, since if the peak current ratings are exceeded for even a few seconds, the cathode of the tube may be permanently damaged.

Hot-cathode mercury-vapor rectifier tubes are almost universally used where high voltage and high currents are required. In such applications they have the advantage over high-vacuum rectifier tubes of much greater efficiency, lower first cost, better voltage regulation, and lower filament power. The chief disadvantage is the greater care that must be employed in their use.

Miscellaneous Types of Rectifiers.—Mercury-arc rectifiers[2] have been used to a limited extent in high-power radio transmitters, particularly in Europe. The types used ordinarily are provided with grids that enable the output voltage to be controlled and regulated, and give means for high-speed protection in case of trouble.

Mercury-pool rectifiers of the ignitron type have been proposed for high-power radio uses.[3] As compared with mercury-arc rectifiers, they have the advantage of being less expensive in first cost and maintenance, are more flexible, and are adapted

[1] In some cases, there is also specified a maximum allowable instantaneous current. This is the maximum current that can be allowed to flow momentarily under infrequent special circumstances without permanently damaging the cathode. It is useful in indicating the circuit impedance that is required to protect the rectifier against faults, such as short circuits.

[2] For further information on such tubes, see S. R. Durand and O. Keller, Grid Control of Radio Rectifiers, *Proc. I.R.E.*, Vol. 25, p. 570, May, 1937.

[3] C. B. Foos and W. Lattemann, High Voltage Mercury-pool Tube Rectifiers, *Proc. I.R.E.*, Vol. 24, p. 977, July, 1936.

to circuits that require the cathodes associated with different anodes to be at different potentials.

The cold-cathode gaseous rectifier makes use of the fact that the current that ionization causes to flow between two electrodes immersed in a low-pressure gas is roughly proportional to the cathode area. Hence if one electrode has a very small, and the other a very large area, a rectifying action takes place. Such rectifiers are used to a limited extent in connection with vibrator power supplies. They have the advantage of needing no filament heating power, but do require a relatively high voltage drop to start the flow of current, and introduce high-frequency transients.

Copper oxide and, more recently, selenium rectifiers are used for a variety of miscellaneous applications such as rectifiers for producing bias voltages, for exciting the fields of loud-speakers, for battery charging, and in a few cases for producing plate power. The copper oxide rectifier makes use of the fact that when a thin film of cuprous oxide is formed upon a metallic copper surface, the resistance that this film offers is small for currents flowing in one direction and high for currents going the opposite way.[1] This rectifying action is very stable, and the only effect of long use is a slight increase of the resistance in the conducting direction during the first 10,000 hours of use.

Selenium rectifiers utilize the rectifying action obtained when a properly processed selenium film is formed on a metallic surface such as iron.[2] Selenium rectifiers have characteristics similar to those of copper oxide rectifiers with respect to stability and long life. The resistance in the forward direction is less, however, so that the rectifier efficiency and the current carrying capacity for a given physical size are greater.

Parallel Operation of Rectifiers.—High-vacuum, copper oxide, and selenium rectifiers may be operated in parallel without any particular precautions other than ensuring that the units in parallel are reasonably similar in characteristics.

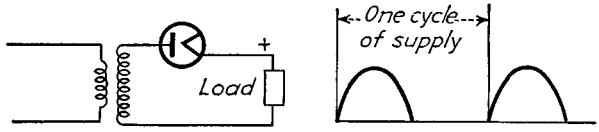
Rectifiers involving ionization, however, give difficulty in parallel operation, because if one tube ionizes first, the voltage across that tube then drops and the other tube will fail to ionize. In the case of hot-cathode mercury-vapor tubes this situation can be remedied by placing in series with the plate of each tube a resistance of such a size as to produce a voltage drop of the order of 10 to 30 volts with the peak current being carried by the tubes. Cold-cathode gaseous rectifiers will not operate satisfactorily in parallel even with series resistance. Where more current is required than can be handled by one tube, it is necessary to operate each tube from a separate transformer winding or to deliver the output of each tube to separate filters having a common output.

2. Rectifier Circuits.—The circuits in which rectifiers are used can be conveniently classified according to whether the power source is single-phase or polyphase. Single-phase rectifier circuits are used to develop anode power for radio receivers, public-address systems, small radio transmitters, etc., whenever the power required is small, *i.e.*, of the order of 1 kw or less. Polyphase rectifiers are used for higher powers. They have the advantage of utilizing the possibilities of the transformers and rectifier tubes more completely than the single-phase arrangements but the disadvantage of being more complicated and requiring a polyphase source.

Single-phase Rectifier Circuits.—The various types of rectifier connections that may be employed with a single-phase power source are shown in Fig. 1, together with the

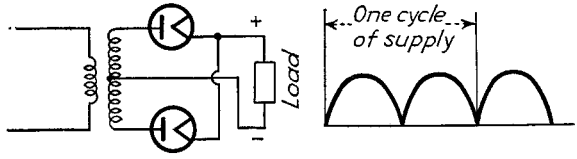
[1] For further information, see L. O. Grondahl and P. H. Geiger, A New Electronic Rectifier, *Trans. A.I.E.E.*, Vol. 46, p. 357, 1927.

[2] Detailed information on selenium rectifiers is given by Carole A. Clarke, Selenium Rectifier Characteristics, Application and Design Factors, *Elec. Comm.*, Vol. 20, No. 1, p. 47, 1941; J. E. Yarmack, Selenium Rectifiers and Their Design, *Trans. A.I.E.E.*, Vol. 61, p. 488, July, 1942; Selenium Rectifiers for Closely Regulated Voltages, *Elec. Comm.*, Vol. 20, No. 2, p. 124, 1941; E. A. Richards, The Characteristics and Applications of the Selenium Rectifier, *Jour. I.E.E.*, Vol. 88, Part III, p. 238, December, 1941, and Vol. 89, Part III, p. 73, March, 1942.



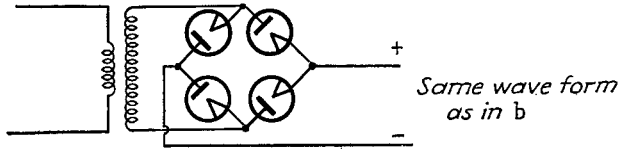
Half-wave (single-phase) output

(a) Half-wave rectifier circuit



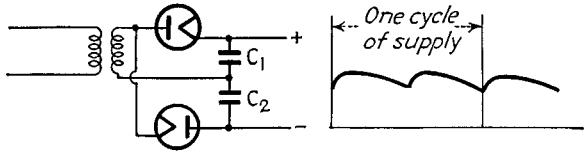
Full-wave (two-phase) output

(b) Full-wave circuit with center-tapped transformer

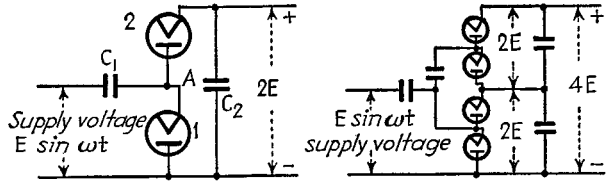


(c) Full-wave bridge circuit

Voltage multiplying circuits



(d) Conventional voltage doubling circuit



(e) Cascade doubler (f) Cascade quadrupler

FIG. 1.—Rectifier circuits operating from a single-phase power source.

wave form of the voltage that is developed across a resistance load. Of these, the full-wave circuit with center-tapped transformer is by far the most widely used with tube rectifiers, and the full-wave bridge circuit is generally preferred with copper oxide and selenium rectifiers. As compared with the center-tapped circuit, the bridge arrangement requires twice as many tubes operated at half the inverse voltage, gives greater d-c power output in proportion to the transformer kva rating, but requires a filament transformer having three separate well-insulated windings instead of a single winding. The half-wave circuit is used only when very small currents are required because the secondary current always flows through the transformer secondary in the same direction, thereby tending to saturate the iron.

The voltage-multiplying circuits provide the possibility of developing a d-c voltage greater than the crest alternating voltage of the source. Such circuits are used when such high voltages are to be developed that it is not convenient to build a transformer for the full voltage, or when it is desired to obtain sufficient anode voltage to operate properly small tubes from a 110-volt a-c source without using transformers. In the conventional voltage-doubling circuit of Fig. 1*d*, the condensers C_1 and C_2 are charged on alternate half cycles, and are then arranged so that the voltages of these two condensers add, thereby giving the possibility of an output voltage that is twice the peak voltage of the a-c supply. The cascade doubler of Fig. 1*e* produces a result equivalent to that of the conventional voltage-doubling circuit, but has the advantage that there is a common terminal between the supply and the output, so that both may be grounded simultaneously. The operation of this circuit can be explained as follows: If tube 2 is disconnected, then condenser C_1 would charge to the peak supply voltage and the potential at the point A would fluctuate between zero and twice the peak supply voltage. If now tube 2 is connected, the voltage between A and ground will tend to charge condenser C_2 to the peak voltage existing at A, thereby giving the possibility of twice the peak supply voltage being obtained in the output. Such a cascading arrangement can be extended indefinitely.[1] A four-stage system developing a d-c output potential of four times the peak supply voltage with small load currents is shown in Fig. 1*f*.[2]

Polyphase Rectifier Connections.—Polyphase rectifiers ordinarily operate from a three-phase power system.[3] Although an unlimited number of circuit combinations are possible, the circuits of Fig. 2 include all the arrangements ordinarily encountered in radio work. They can be conveniently subdivided into systems giving three- and six-phase output waves, *i.e.*, ripple frequencies of three and six times the supply, respectively.

The three-phase output developed by connections *a* and *b* of Fig. 2 is much more nearly constant than is the output of the full-wave single-phase circuits of Fig. 1, and the ripple frequency is three times the supply frequency instead of only twice as great. The star connection has the advantage over the broken star of slightly better utilization of the transformer capacity, but the disadvantage that in order to avoid d-c saturation in the core, it is necessary to use a single three-phase transformer, whereas with the broken star it is possible to employ three single-phase transformers, provided that each transformer has two secondary windings. With the broken star, it is also permissible to use a Y connection in the primary.

[1] Such cascade systems are described by J. D. Cockcroft and E. T. Walton, Experiments with High Velocity Positive Ions: (1) Further Developments in the Method of Obtaining High Velocity Positive Ions, *Proc. Roy. Soc. (London)*, A, Vol. 136, p. 619, 1932. See also E. B. Moullin, External Characteristic of Thermionic Rectifier, *Jour. I.E.E.*, Vol. 80, p. 553, 1937; also, Wireless Section, *I.E.E.*, Vol. 12, p. 156, June, 1937.

[2] Another type of voltage quadrupler is described by William W. Garstang, A New Voltage Doubler, *Electronics*, Vol. 4, p. 50, February, 1932.

[3] Information on rectifiers operating from a two-phase supply is given by R. W. Armstrong, Polyphase Rectification Special Connections, *Proc. I.R.E.*, Vol. 19, p. 78, January, 1931.

The six-phase output wave obtained with connections *c* to *f* in Fig. 2 is even more nearly constant than the three-phase output wave, and the ripple is now six times the supply frequency. The six-phase single Y circuit makes the most effective use of the transformer and tube of any of the polyphase circuits shown, and also has the lowest inverse voltage of any of the circuits. The only disadvantage of this circuit is that it requires four separate well-insulated windings on the filament transformer

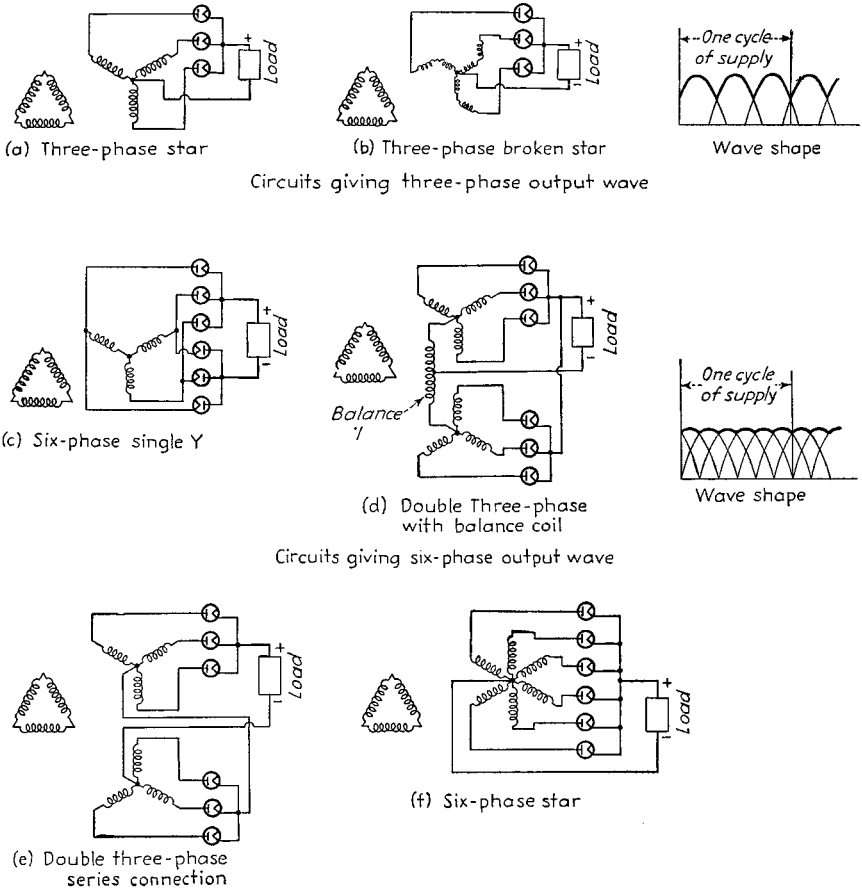


FIG. 2.—Rectifier circuits operating from a three-phase power source and developing three-phase and six-phase output waves.

Although a delta primary connection is shown in Fig. 2*c*, a Y connection is also permissible.

The double three-phase arrangement with balance coil requires a somewhat higher kva transformer rating than does the six-phase Y, but is decidedly superior in this respect to the circuits giving a three-phase output. If the balance coil is short-circuited, this arrangement reduces to a six-phase star, an undesirable arrangement from the point of view of the effectiveness with which the transformer and tubes are utilized. In order that the balance coil may operate satisfactorily, it must have sufficient inductance to maintain current flowing continuously through each half of the

coil.[1] The double three-phase arrangement with series connection gives twice as much voltage and half the current as the same transformers and tubes arranged with the balance coil, while the effectiveness with which transformers and tubes are utilized, the inverse voltage, etc., are the same in the two cases.

The permissible transformer arrangements for the systems giving a six-phase output vary with the rectifier connection. In the case of the six-phase single-Y rectifier circuit, either three separate single-phase transformers or one three-phase transformer can be used. With the double three-phase arrangements, however, it is not permissible to use single-phase transformers unless these are arranged so that each transformer has two secondaries divided between the two stars. Six single-phase transformers cannot be used because d-c saturation of the core would result. Delta primary connections are shown in Figs. 2*c* to 2*f*, but it is permissible to use a star primary for the six-phase single Y, and the same is true of the double three-phase circuits provided that a single three-phase transformer or three single-phase transformers are used. If separate three-phase transformers are used to supply each secondary star of a double three-phase system, then a delta primary is necessary.

It is possible to operate three- and six-phase rectifier systems from two special transformers arranged in a T connection. When this is done, slightly greater transformer kva must be provided, but since two large transformers are often cheaper than three small transformers of slightly less total rating, there is economy in the T connection.[2]

Rectifier circuits providing output waves of twelve and more phases can be readily obtained from a three-phase supply, and are widely used in mercury-arc rectifiers. They have the advantage of much less ripple voltage and a higher ripple frequency than the six-phase circuits, but are seldom employed in radio work because of the greater number of rectifier tubes required.

Transformer Insulation Requirements.—The insulation that must be provided for the secondary of the power transformer in the circuits of Fig. 1 depends upon the rectifier connections involved. On the assumption that the negative side of the rectified output is grounded, then in the center-tapped arrangement, the middle of the transformer secondary requires negligible insulation, and the ends must withstand half of the total secondary voltage. In contrast with this, the bridge circuit requires that the center of the transformer withstand half and the terminals the full secondary voltage, whereas in the conventional voltage-doubling circuit, the lower end of the transformer must be insulated to stand the full secondary voltage to ground while the other end must be capable of handling twice the secondary voltage.

In the polyphase circuits of Fig. 2 the insulation that is required also varies with the connection. In all cases except the three-phase star and the double three-phase arrangement with balance coil, there is a winding or a portion of a winding subjected to a greater voltage to ground than the crest alternating voltage developed between the terminals of that particular coil. Polyphase rectifier circuits employing single-phase transformers designed for power-line service can accordingly be expected to overstress the insulation in many cases.

The secondaries of the filament transformers must be well insulated from ground, and from each other. Thus in Fig. 1*b*, the secondary is subjected to half the transformer voltage, and in Figs. 1*a* and 1*c*, each filament winding must stand a voltage to ground equal to the full secondary voltage of the power transformer. In Fig. 1*c*, this same insulation is also required between individual filament windings. In the polyphase circuits of Figs. 2*a*, *b*, *d*, and *f*, the filament winding is subjected to the full

[1] Further discussion of the requirements that a balance coil must meet in order to operate satisfactorily are given in Par. 3.

[2] For further information see R. W. Armstrong, Polyphase Rectification Special Connections, *Proc. I.R.E.*, Vol. 19, p. 78, January, 1931.

voltage to neutral of the power transformer, and in Fig. 2*e* the voltage stress is twice that developed to neutral by one set of transformers. With the six-phase single Y, the filament windings must all be insulated to stand the secondary delta voltage of the power transformer both to ground and to each other.

Combination Rectifier Circuits.—Two different types of rectifier connections are sometimes combined in a single arrangement to provide different output voltages. Examples are shown in Fig. 3. The arrangement at *a* combines the full-wave bridge rectifier circuit of Fig. 1*c* with the center-tapped circuit of Fig. 1*b* to give two d-c outputs developing voltages in the ratio of 2:1.[1] Analogously, the circuit of Fig. 3*b* combines the six-phase single-Y circuit of Fig. 2*c* with the three-phase star of Fig. 2*b* to give two output voltages in the ratio of 2:1. Such arrangements provide a simple

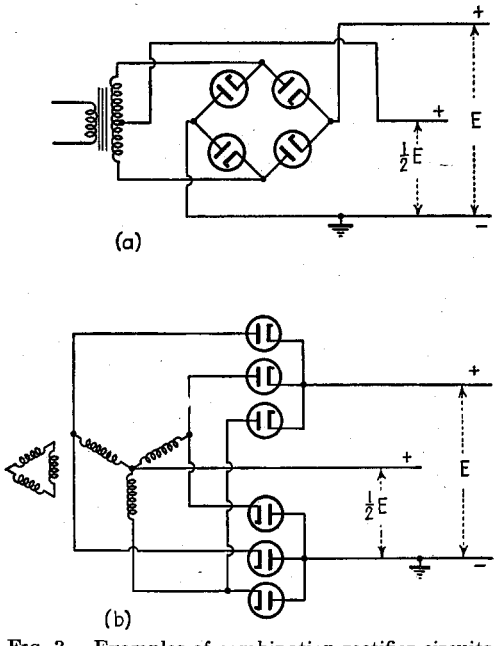


FIG. 3.—Examples of combination rectifier circuits.

means of obtaining a lower voltage output of comparatively small current rating without the addition of appreciable accessory equipment.

3. Performance of Rectifiers When Used with Filter Systems Having Series-inductance Inputs.—The output of a rectifier is normally delivered to a filter that smooths out the pulsations and produces a substantially constant d-c voltage. The filters used for this purpose can be divided into two types, *i.e.*, those which present a series inductance to the rectifier output and those which present a shunt condenser to the output terminals of the rectifier. Either arrangement can be used with any of the rectifier circuits of Figs. 1 and 2 except that the voltage-multiplying circuits and the half-wave connection of Fig. 1 require condenser input. Inductance-input systems utilize the tubes and transformers more effectively than condenser-input arrangements and so are used wherever large powers are involved. They also have better voltage regulation and so are employed where the load on the system varies, as in Class B amplifiers.

[1] For further details of a practical character see G. E. M. Bertram and R. S. Quimby, A Duplex Plate Supply Using Type 83 Tubes, *Q.S.T.*, Vol. 17, p. 31, March, 1933.

Voltage and Current Relations in Rectifier Tubes and Transformers.—The basic voltage and current relations existing with an inductance-input filter can be understood by considering the ideal case of an infinite input inductance combined with a transformer having zero leakage inductance and resistance. Under these conditions the output current delivered by the rectifier system is constant, the current simply shifting from anode to anode during the cycle. The resulting current waves in typical cases are shown in Fig. 4. The fact that the current waves in the transformer are not sinusoidal causes the heating of the transformer windings to be greater for a given d-c power output from the rectifier than would be the case if the same amount of a-c power were

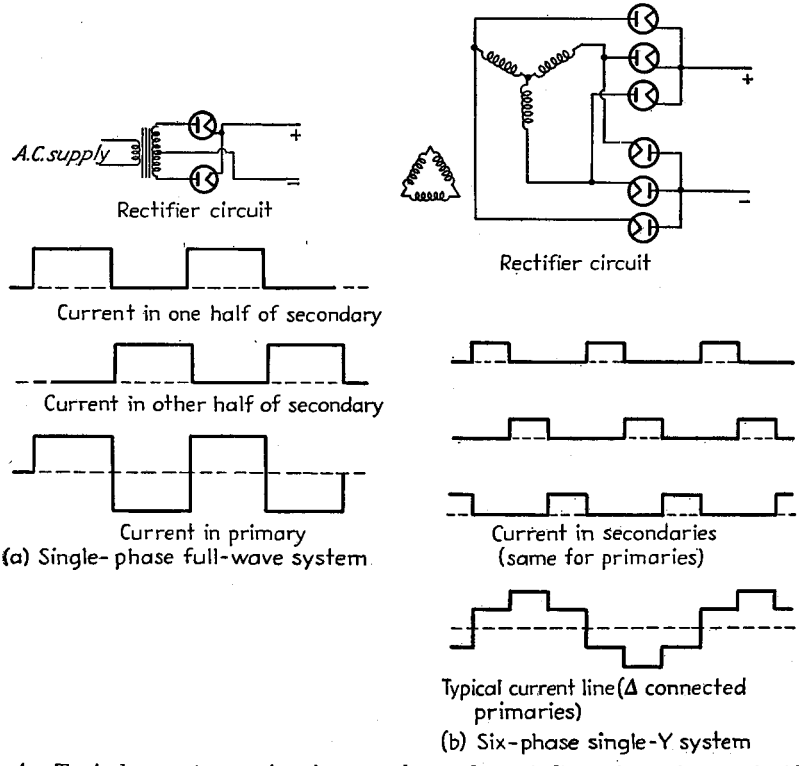


FIG. 4.—Typical current waves in primary and secondary windings of transformers for ideal inductance-input system.

drawn by a resistance load. This makes it necessary to design the transformer windings more generously for rectifier applications than would otherwise be the case. The ratio of d-c output to required kva capacity is termed the transformer utilization factor, or *utility factor*. Its value depends upon the rectifier connection, and is also commonly not the same for the primary and secondary windings.

The voltage developed across the input terminals of the filter system is the same as the voltage developed across a resistance load. The character of the output voltage therefore depends upon the number of phases in the output wave, as shown in Figs. 1 and 2, and is not affected by the particular way in which the rectifier is connected to develop this number of phases.

Quantitative Behavior of Ideal Rectifier System with Inductance-input Filter.—The quantitative conditions existing in ideal rectifier-filter systems, assuming zero trans-

former resistance and leakage inductance, zero tube resistance, and an infinite input inductance are given in Table 1.[1] This table gives all the information essential for determining the voltage and kva capacity required of each transformer winding, the peak and average currents through the rectifier tubes, the maximum inverse voltage that the rectifier tube must stand, and the harmonic components contained in the output voltage. A simple calculation showing the use of the table in a practical case is given in Par. 6.

TABLE 1

	Rectifier circuit							
	Single-phase, full-wave center-tapped connection (Fig. 1b)	Single-phase, full-wave bridge (Fig. 1c)	Three-phase star (Fig. 2a)	Three-phase broken star (Fig. 2b)	Six-phase single Y (Fig. 2c)	Double three-phase with balance coil (Fig. 2d)	Double three-phase, series connection (Fig. 2e)	Six-phase star (Fig. 2f)
Voltage relations (d-c component of output voltage taken as 1.0):								
a. Rms value of transformer secondary voltage (per leg)	1.11*	1.11	0.855	0.985	0.428	0.855	0.428	0.740
b. Maximum inverse voltage	3.14	1.57	2.09	2.09	1.05	2.42	1.05	2.09
c. Lowest frequency in rectifier output (F = frequency of power supply)	$2F$	$2F$	$3F$	$3F$	$6F$	$6F$	$6F$	$6F$
d. Peak value of first three a-c components of rectifier output:								
Ripple frequency	0.667	0.667	0.250	0.250	0.057	0.057†	0.057	0.057
Second harmonic of ripple frequency	0.133	0.133	0.057	0.057	0.014	0.014	0.014	0.014
Third harmonic of ripple frequency	0.057	0.057	0.025	0.025	0.006	0.006	0.006	0.006
e. Ripple peaks with reference to d-c axis								
Positive peak	0.363	0.363	0.209	0.209	0.0472	0.0472	0.0472	0.0472
Negative peak	0.637	0.637	0.395	0.395	0.0930	0.0930	0.0930	0.0930
Current relations:								
f. $\frac{\text{Average current per anode}}{\text{Peak anode current}}$	0.500	0.500	0.333	0.333	0.333	0.333	0.333	0.167
g. $\frac{\text{Average current per anode}}{\text{Direct-current in load}}$	0.500	0.500	0.333	0.333	0.333	0.167	0.333	0.167
h. $\frac{\text{Peak current per anode}}{\text{Direct-current in load}}$	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000
Transformer requirements (d-c output power = 1.0):								
i. Primary kva	1.11	1.11	1.21	1.05	1.05	1.05	1.05	1.28
j. Secondary kva	1.57	1.11	1.48	1.71	1.05	1.48	1.48	1.81
k. Average of primary and secondary kva	1.34	1.11	1.35	1.38	1.05	1.26	1.26	1.55

NOTE: This table assumes that the input inductance is sufficiently large to maintain the output current of the rectifier substantially constant, and neglects the effects of voltage drop in the rectifier and the transformers.

* Secondary voltage one side of center tap.

† The principal component of voltage across the balance coil has a frequency of $3F$ and a peak amplitude of 0.500. The peak balance coil voltage, including the smaller higher harmonics, is 0.605.

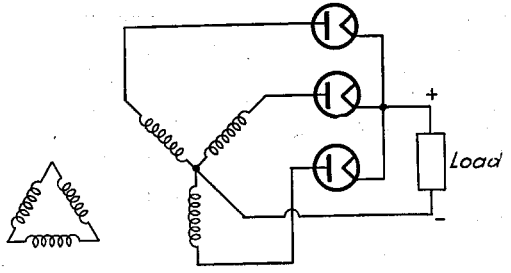
Voltage Regulation in Input-inductance Systems.—The output voltage of a rectifier-filter system employing a properly designed input inductance falls off with increasing load as a result of (1) resistance in the filter; (2) resistance of the rectifier tubes;[2]

[1] Most of the data in this table are from Armstrong, *loc. cit.* For corresponding information on additional connections see Armstrong, *loc. cit.;* D. C. Prince and F. B. Vodges, "Mercury Arc Rectifiers and Circuits," McGraw-Hill, New York, 1927; O. K. Marti and H. Winograd, "Mercury Arc Power Rectifiers," McGraw-Hill, New York, 1930.

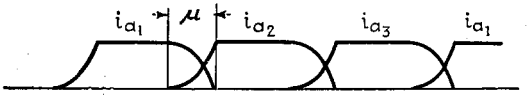
[2] With hot-cathode mercury-vapor tubes, the voltage drop is constant irrespective of the current. It is accordingly permissible to assume here that the tube has zero resistance and that the output potential is always 10 to 15 volts less than would be calculated on the basis of an ideal tube.

In grid-controlled and ignition rectifiers, the starting voltage at which the tube fires may be much

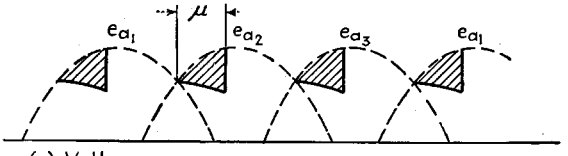
(3) resistance in the transformer; (4) leakage reactance of the supply transformer. The various resistances in the circuit reduce the output voltage without affecting the wave shape in the system. In calculating the effect of the resistances in the circuit, the tube drop is determined by the number of tubes in series and the tube drop required to carry the anode currents. The voltage loss in the transformer resistance is calculated on the basis of the primary and secondary currents in the transformer at any one time and the transformer resistances through which this current must flow. With



(a) Three-phase half-wave rectifier circuit



(b) Current of individual anodes



(c) Voltage waves

FIG. 5.—Effect of leakage inductance in the transformer on the behavior of a polyphase rectifier.

choke-input systems, the tube and transformer resistance drop is constant throughout the cycle, even though the current shifts between tubes and transformer windings.

The leakage reactance of the transformer prevents the current from shifting instantly from one winding to another as in the ideal case in Fig. 4. The situation in a typical case is shown in Fig. 5*b*, where μ represents the interval required to transfer the current. During this transition period, the output voltage, instead of following the open-circuit potential of the more positive anode, as in the case of an ideal rectifier,

greater than the voltage drop required to maintain the current, once the flow has started. This causes the time at which the tube fires to be later in the cycle than in the case of an ideal tube, with a resultant lowering of the output voltage by an amount that depends upon the firing voltage, *i.e.*, upon the delay in firing time.

assumes a value that is the average of the open-circuit voltages of the two windings that are simultaneously carrying current, as shown in Fig. 5c. This causes the average voltage of the output to be less than if there were no leakage inductance by an amount indicated by the shaded areas of Fig. 5c. The calculation of this voltage loss is complicated, and depends both upon the rectifier and transformer connections. The reader wishing to make such computations can find the details in books devoted solely to the subject of rectifiers.[1]

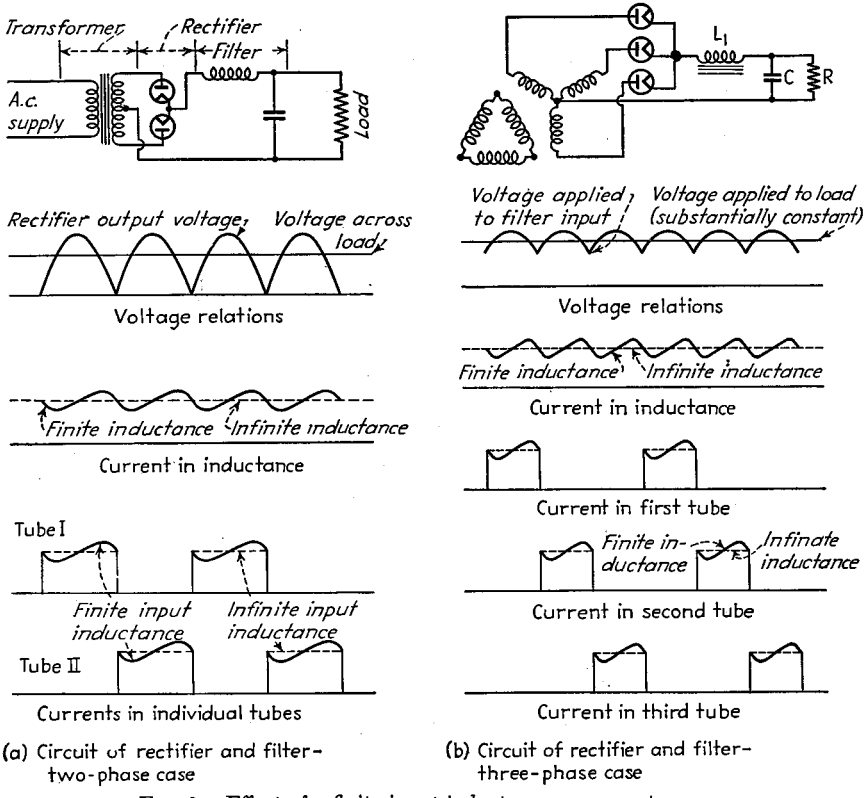


FIG. 6.—Effect of a finite input inductance on current waves.

Effect of Finite Input Inductance.—When the input inductance is large but not infinite, the output current of the rectifier will vary somewhat, increasing when the output voltage developed by the rectifier is greater than the d-c component of the output, and dropping when it is less. The behavior in two typical cases is shown in Fig. 6, and it is apparent that the peak current that the tubes must now carry to develop a given d-c output current is greater than when an infinitely large choke is used.

Most of the fluctuation in output current with finite inductance is composed of current of the lowest ripple frequency. To the extent that this is true, the increase in peak rectifier current resulting from a finite input inductance is given by the equation

$$\frac{\text{Peak current with finite input inductance}}{\text{Peak current with infinite input inductance}} = 1 + \frac{E_1}{E_0}\frac{R_{\text{eff}}}{\omega L_1} \qquad (1)$$

[1] Prince and Vodges, *op. cit.*; Marti and Winogard, *op. cit.*

where E_1/E_0 is the ratio of lowest-frequency ripple component to the direct-current voltage in rectifier output as given by Table 1, R_{eff} is the effective load resistance (actual load plus filter resistance plus equivalent tube and transformer resistances), and ωL_1 is the reactance of the input inductance *at the lowest ripple frequency.* The fact that the input inductance is finite instead of infinite has practically no effect upon the output voltage, provided that the input inductance exceeds the critical value discussed below.

The minimum input inductance that will give normal input-inductance operation of the rectifier system is the smallest inductance that will maintain a continuous flow of current from the rectifier output. The requirement for continuous current flow is given by the approximate relation[1]

$$\frac{\omega L_1}{R_{\text{eff}}} \geq \frac{E_1}{E_0} \qquad (2)$$

where the notation is the same as in Eq. (1). In the important practical case of a 60-cycle single-phase full-wave circuit, Eq. (2) becomes

$$L_1 = \frac{R_{\text{eff}}}{1{,}130} \qquad (3)$$

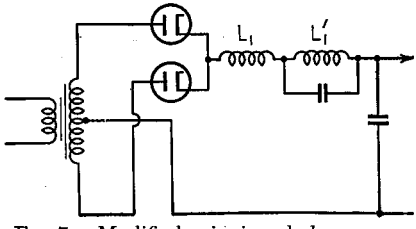


FIG. 7.—Modified swinging-choke arrangement.

The minimum allowable input inductance as given by Eq. (2) is termed the *critical inductance.*[2] When the inductance is less than the critical value, the system acts as a condenser-input system. The d-c voltage developed across the filter input is then greater than given by Table 1, and the regulation is poorer than would be the case with adequate input inductance.

When the direct current drawn from the system varies, it is necessary to satisfy Eq. (2) or Eq. (3) at all times if good voltage regulation is to be maintained. In order to do this with very small load current without excessive inductance, it is necessary to place a bleeder resistance across the filter output to ensure that R_{eff} in Eq. (2) will never exceed a reasonable value.

With single-phase full-wave systems, the input inductance required for light loads, even with a reasonable bleeder resistance, is still greater than desirable. One way of meeting this situation in an economical manner is to use a moderate-sized input inductance and then to make the air gap so small that the required incremental inductance is obtained with a reasonable bleeder current. With large currents, the core saturates, and the inductance will drop, but Eq. (2) is still satisfied. An input inductance adjusted to operate in this way is sometimes referred to as a *swinging choke.* Another method[3] of maintaining a continuous flow of current from the rectifier without employing an excessively large inductance is shown in Fig. 7, and consists of a swinging choke L_1 so small as to be inadequate when operating alone at small currents, in series with another small choke L_1' that is tuned to the fundamental ripple frequency.[4]

[1] This assumes that the entire fluctuation of current in the input inductance is a result of the lowest ripple voltage contained in the rectifier output acting against the impedance of the input inductance. This assumption is very closely realized because the higher frequency components of current in the input inductance are quite small and because, in practical filters, the shunting condenser following the input inductance has very small impedance to the ripple frequency compared with the reactance of the input inductance. A more complete analysis is given by C. R. Dunham, Some Considerations in the Design of Hot-cathode Mercury-vapor Rectifier Circuits, *Jour. I.E.E.*, Vol. 75, p. 278, 1934, Vol. 76, p. 421, 1935 (also Wireless Section, *I.E.E.*, Vol. 9, p. 275, September, 1934, and Vol. 10, p. 108, June, 1935).

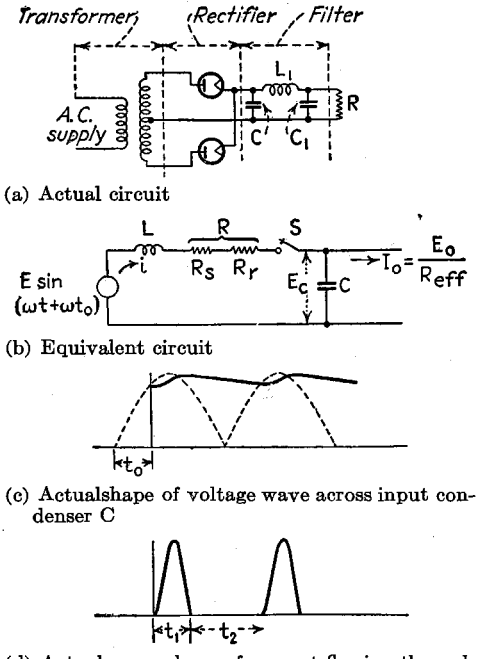
[2] For further discussion, see F. S. Dellenbaugh, Jr., and R. S. Quimby, The First Filter Choke—Its Effect on Regulation and Smoothing, *QST*, Vol. 16, p. 26, March, 1932.

[3] See Reuben Lee, Solving a Rectifier Problem, *Electronics*, Vol. 11, p. 39, April, 1938.

[4] If the inductance L_1' is not too small, the swinging choke L_1 in Fig. 7 can be eliminated.

In rectifiers where the voltage required to initiate a discharge between cathode and anode is greater than the subsequent voltage drop when carrying current, as is the case in all grid-controlled and ignition mercury-pool arrangements when the adjustment is such as to give delayed firing, the input inductance required to maintain a continuous flow of current and give proper operation is greater than called for by Eq. (2).[1] Failure to employ sufficient input inductance in such cases not only will make the voltage regulation poorer but also may cause erratic behavior, such as irregular firing of the tubes.

4. Performance of Rectifiers When Used with Filter Systems Having Shunt-condenser Input. *Mechanism of Operation.*—When the input to the filter presents to the rectifier a shunt condenser instead of a series inductance, the behavior is greatly modified. Each time the crest alternating voltage of the transformer is applied to one of the rectifier anodes, the input condenser charges up to just slightly less than this peak voltage. The rectifier then ceases to deliver current to the filter until another anode approaches its peak positive potential, when the condenser is charged again. During the interval when the voltage across the input condenser is greater than the potential of any of the anodes, the voltage across the input condenser drops off nearly linearly with time, because the first filter inductance draws a substantially constant current from the input condenser. A typical set of voltage and current waves is shown in Fig. 8.



(a) Actual circuit

(b) Equivalent circuit

(c) Actualshape of voltage wave across input condenser C

(d) Actual wave shape of current flowing through tube

FIG. 8.—Actual and equivalent circuits of condenser-input rectifier-filter system, together with oscillograms showing operation under typical conditions.

Quantitative Relations.—The detailed action that takes place in a condenser-input system depends in a relatively complicated way upon the load resistance in the rectifier output, the input capacity, the leakage reactance and resistance of the transformer, and the characteristics of the rectifier tube. For purposes of analysis, the actual rectifier circuit, such as that of Fig. 8*a*, can be replaced by the equivalent circuit of Fig. 8*b*. Here the rectifying action of the tube is replaced by a suitable switch S that is assumed to be closed whenever one of the rectifier anodes is conducting current and that is open the remainder of the time. The inductance L of the equivalent circuit is the leakage inductance of the transformer, measured in Fig. 8 across one-half the secondary winding with the primary short-circuited, and the equivalent resistance R is the corresponding transformer resistance R_s measured in the same way, plus a fixed resistance R_r that takes into account the voltage drop in the rectifier tube. The input condenser of the

[1] A comprehensive discussion of the input inductance required in such cases, including formulas for calculating the critical inductance, is given by W. P. Overbeck, Critical Inductance and Control Rectifiers, *Proc. I.R.E.*, Vol. 10, p. 655, October, 1939.

filter system is C, and the first inductance L_1 is assumed to draw a constant current I_0 equal to the d-c voltage E_0 developed across the input condenser divided by a resistance equal to the sum of the actual load resistance plus the resistance of the filter inductances.

The most important use of condenser-input filter systems is with single-phase full-wave rectifier circuits supplying power to radio receivers, small public-address systems, etc. The performance of such rectifier systems can be determined by a relatively

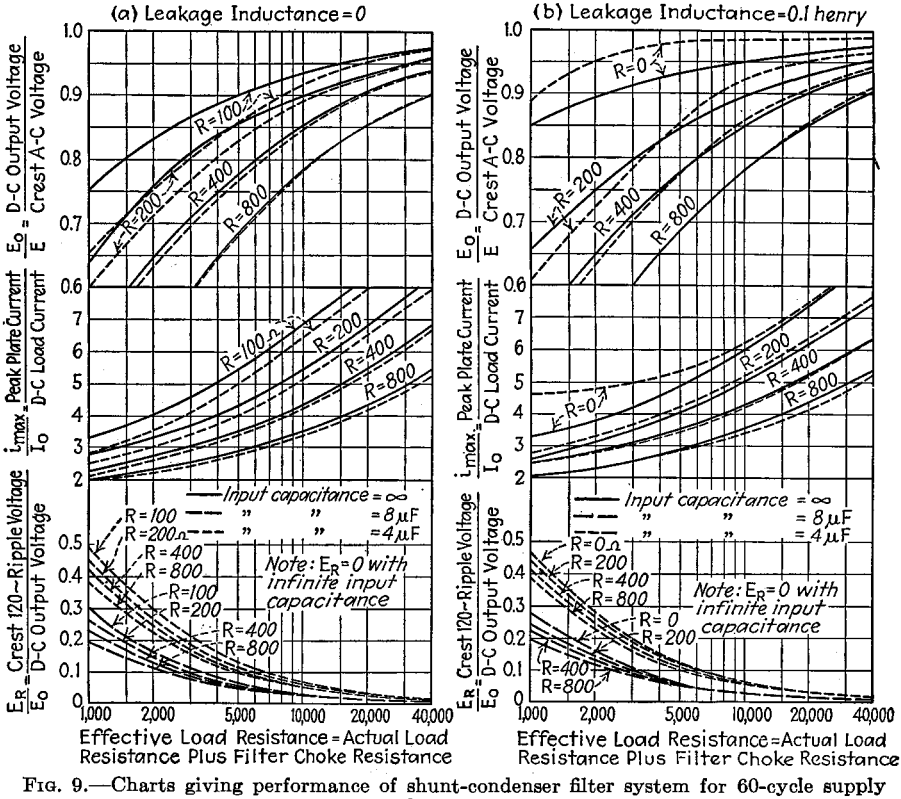


FIG. 9.—Charts giving performance of shunt-condenser filter system for 60-cycle supply source and full-wave rectifier.

complicated analysis based on the equivalent circuit of Fig. 8*b* and leads to the results given in Fig. 9.[1] The approximations involved in Fig. 9 are as follows: (1) The rectifier

[1] For details of the analysis, see F. E. Terman, "Radio Engineering," 2d ed., pp. 492–498, McGraw-Hill, New York, 1937, or the thesis by Dr. R. L. Freeman, Analysis of Rectifier-filter Circuits, Stanford University Library, 1934.

Analyses for a variety of special cases of condenser-input systems have been developed. Although none of these take into account all of the factors involved and at the same time present the results in convenient form for practical use, they are helpful in estimating the behavior that can be expected. The following references, with a brief summary of the ground covered by each, will he helpful to the reader wishing to go into the subject more deeply:

E. B. Moullin, The External Characteristic of a Diode Rectifier, *Jour. I.E.E.*, Vol. 80, p. 553, 1937; also, Wireless Section, *I.E.E.*, Vol. 12, p. 156, June, 1937. Analysis is made of single-phase half-wave and full-wave circuits, assuming a linear tube characteristic, zero generator impedance, and a resistance load connected directly in shunt with the input capacity. Equations are derived for power output, power loss in diode, effective value of diode current, and voltage regulation characteristic. The results are extended to voltage multiplier systems. N. H. Roberts, the Diode as Half-wave, Full-wave and Voltage-doubling Rectifiers, *Wireless Eng.*, Vol. 13, p. 351, July, 1936, which covers the same

tube is assumed to begin passing current as soon as the anode becomes positive with respect to the cathode; (2) the rectifier is replaced by a fixed resistance during its conducting period; (3) the current flowing into the first inductance is assumed to be constant throughout the cycle; (4) the voltage at the primary terminals of the power transformer is assumed to have perfect regulation. The errors that result from the last three of these approximations are small under practical circumstances, and the first approximation is also satisfied in the case of high-vacuum, copper oxide, and selenium rectifiers. The case of hot-cathode mercury-vapor tubes can be handled by assuming that the tubes have zero resistance and then subtracting from the calculated output 15 volts drop for each tube that is in series. An example of the use of the curves in Fig. 9 is given in Par. 6.

Discussion of Shunt-condenser Systems.—Examination of Fig. 9 shows that the d-c voltage developed by the rectifier operating with a shunt-condenser filter system depends primarily upon the load resistance and the internal impedance of the power source, as represented by R and L in the equivalent circuit of Fig. 8*b*. The ratio of peak to average current through the rectifier tubes and transformer secondaries also depends upon these same quantities, with the peaks being smaller in proportion to the d-c current when the internal impedance is high and a large load current is drawn from the rectifier. The regulation of the d-c voltage is controlled largely by the internal impedance of the source, which should be small if the regulation is to be good. The ripple voltage is greater the lower the load resistance (*i.e.*, the greater the current drawn from the rectifier), and the smaller the input condenser, but is roughly independent of the source impedance. Oscillograms showing the effects of circuit parameters on the detailed behavior of a typical shunt-condenser input system are shown in Fig. 10.

Compared with filters having a series input inductance, the use of a shunt condenser results in the development of a higher d-c voltage across the rectifier output, together with a smaller ripple in this voltage. At the same time the shunt-condenser arrangement results in poorer voltage regulation, a much poorer utilization factor for the transformer, and a relatively high peak current through the rectifier tube in proportion to average current.

Condenser input systems find their chief use in single-phase full-wave rectifier systems developing small amounts of power at reasonably high output voltages under conditions where voltage regulation is not important. In particular, shunt-condenser input systems are almost universally used for supplying anode power to radio receivers and small public-address amplifiers where the total power involved is of the order of 50 watts or less, at voltages of 250 to 400 volts.

High-vacuum rectifiers are usually preferred to hot-cathode mercury-vapor tubes in condenser-input systems, because the life of the high-vacuum tube is not so critical

---

ground as Moullin and likewise assumes a linear tube characteristic, but extends the analysis to include the calculation of ripple voltages for the idealized case of zero reactance in the source. W. H. Aldous, The Characteristics of Thermionic Rectifiers, *Wireless Eng.*, Vol. 13, p. 576, November, 1936. Calculation of anode dissipation, output voltage, and peak current, on the assumption of a three-halves power-tube characteristic, an infinite input capacity, and zero reactance in the source of alternating voltage. C. M. Wallis, Half-wave Rectifier Circuits, *Electronics*, Vol. 11, p. 12, October, 1938. Analysis of voltage and current wave forms for half-wave rectifier of gaseous conduction type, assuming that the voltage drop in the tube is independent of current and that the voltage required to initiate the discharge is no greater than this drop. It is also assumed that the power transformer has zero resistance and leakage inductance. Analysis is given for the load resistance in shunt with input condenser (including the case where the shunt condenser is reduced to zero) and for the load resistance in series with an inductance (but no condensers in system). M. B. Stout, Analysis of Rectifier Filter Circuits, *Elec. Eng.*, Vol. 54, p. 977, September, 1935. Methods of calculating current wave forms are given for a variety of conditions where the current is both continuous and discontinuous. In the latter case, the tube drop and impedance of the source are neglected. D. L. Waidelich, The Full-wave Voltage-doubling Rectifier Circuit, *Proc. I.R.E.*, Vol. 29, p. 554, October, 1941. This article gives an analysis of the full-wave voltage-doubling circuit with a resistance load connected directly across the input condensers and with zero tube drop and source impedance assumed.

with respect to peak current. When hot-cathode mercury-vapor tubes are used, it is often desirable to place between the rectifier output and the shunt condenser a small series inductance that is not large enough to satisfy Eq. (2). Such a choke is seen from the equivalent circuit of Fig. 8 to have the same effect as increasing the leakage inductance of the transformer. It will therefore reduce the peak current passing through the rectifier tubes at the cost of poorer voltage regulation and reduced output voltage. Cold-cathode gaseous rectifiers are always used with shunt-condenser filter systems. No special precautions need be taken with regard to such tubes, since the gaseous ionization permits the momentary passage of very large currents.

Inductance-input systems in which the input inductance is less than the critical value act as condenser-input systems in which the effective source inductance L assumed acting in the equivalent circuit of Fig. 8*b* consists of the actual transformer

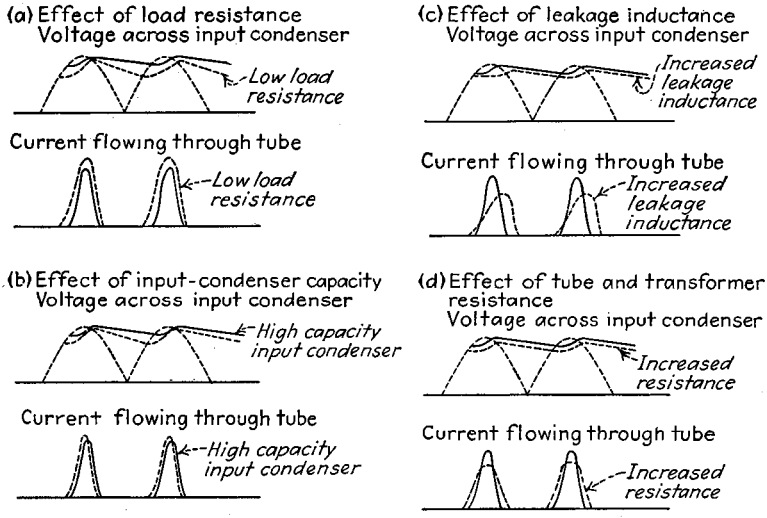


FIG. 10.—Effect of circuit parameters on behavior of rectifier operated with condenser-input filter.

inductance plus the input inductance of the filter. The essential distinction between series-inductance and shunt-condenser systems is that the former refers to operation in which the current flows continuously from the rectifier output into the filter system, and the latter refers to systems in which the current flows intermittently from rectifier into the filter. If intermittent action is present, then the system is classified as a shunt-condenser arrangement even though there may be a series inductance of less than the critical value present.

Shunt-condenser systems are sometimes operated with the load resistance connected directly across the condenser. Such an arrangement is practical when the load current is so small in proportion to the condenser capacity that the variation in voltage across the input condenser is small. The performance of such cases for a single-phase full-wave circuit can be obtained with reasonable accuracy with the aid of Fig. 9.

The performance curves of Fig. 9 can be extended to cover the single-phase half-wave rectifier of Fig. 1*a* by reading the curve on the basis of a leakage inductance that is half the actual inductance and a shunt capacity half the value really present. The ripple voltage in such cases has the same frequency as the supply.

5. Filters.—A filter is ordinarily placed between the output of a rectifier and the load resistance for the purpose of smoothing out the voltage fluctuations of the rectifier output. Such filters consist of alternate series impedances that oppose the flow of alternating currents and shunt condensers that by-pass these currents. Typical filter arrangements are given in Fig. 11. The side of the circuit with the series impedances may be placed in either lead of the filter. However, if one output terminal is grounded, and it is essential that the hum voltage in the output be extremely small, then the filter chokes must be placed in the ungrounded lead.[1]

For purposes of discussion and analysis, filters are ordinarily divided into sections, each of which consists of a series impedance associated with a shunt impedance as indicated in Fig. 11. In making this division, a series inductance is considered part of the first section, and a shunt capacity across the filter input is not included in the first section.

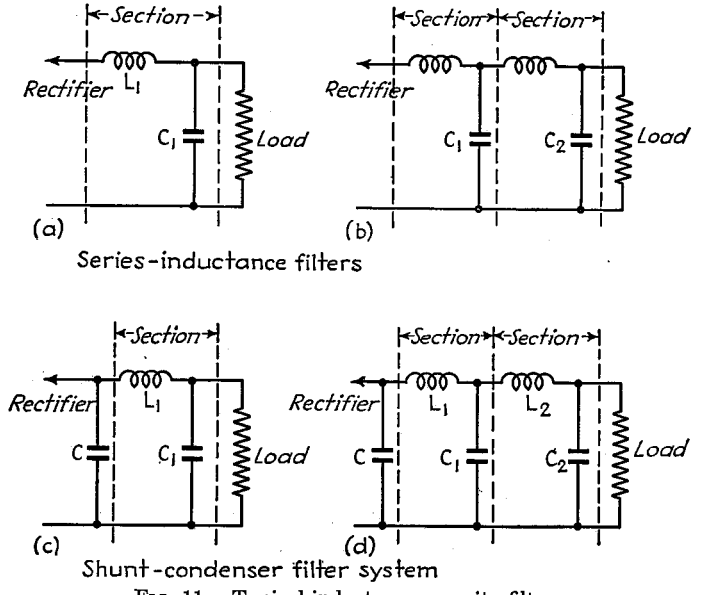


FIG. 11.—Typical inductance-capacity filters.

Inductance-capacity Filters.—Most filter sections are composed of a series inductance and a shunt capacity. In practical filters of this type, the reactance of the shunt capacity is much smaller than the reactance of the series inductance and of the load resistance. To the extent that this is true, such a filter section reduces each frequency component of voltage applied to the input side according to the relation

$$\frac{\left\{\begin{matrix}\text{Alternating voltage}\\ \text{across output of section}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Alternating voltage applied}\\ \text{to input of section}\end{matrix}\right\}} = \frac{1}{(\omega^2 LC - 1)} \tag{4}$$

where L = series inductance of filter section.
C = shunt capacity of section.
$\omega/2\pi$ = frequency of ripple voltage involved.

[1] This is because of electrostatic capacity of the transformer secondary to ground. See F. E. Terman and S. B. Pickles, Note on a Cause of Residual Hum in Rectifier Filter Systems. *Proc. I.R.E.*, Vol. 22, p. 1040. August, 1934.

Results of Eq. (4) are given in Fig. 12 for the usual case where the ripple components are harmonics of 60 cycles.

When a filter consists of more than one section, the total reduction in ripple voltage produced by the combination is very nearly the product of the voltage-reduction factors of the individual sections, or what is the same thing, the sum of the db attenua-

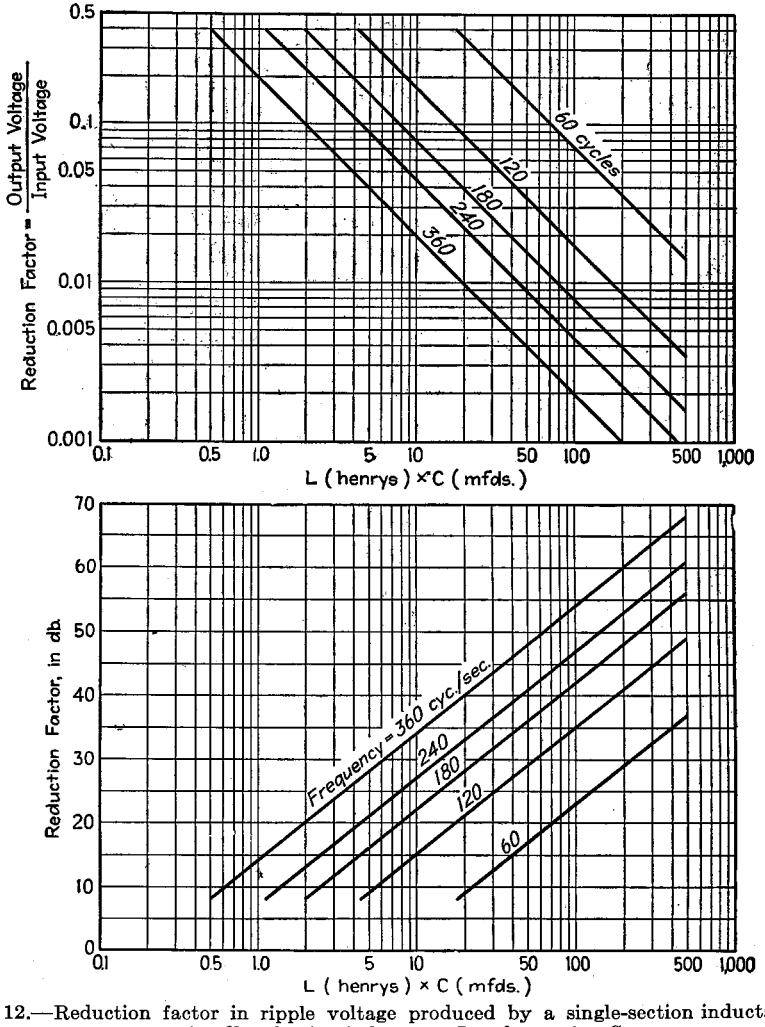


FIG. 12.—Reduction factor in ripple voltage produced by a single-section inductance-capacity filter having inductance L and capacity C.

tions of the individual sections. In the case of multisection filters, the most efficient way to divide a given amount of inductance (or capacity) between the sections is to do so equally.[1]

Resistance-capacity Filters.—A resistance-capacity filter section is shown in Fig. 13*a* and consists of a series resistance followed by a shunt condenser. Such sections

[1] L. B. Hallman, Jr., A Note on the Simple Two-element Low-pass Filter of Two and Three Sections, *Proc. I.R.E.*, Vol. 21, p. 1603, November, 1933.

are frequently used where the current is so small that the loss in voltage in a large resistance will not be excessive. Resistance-capacity sections also find use where it is necessary to drop the voltage, as, for example, in obtaining screen voltage from the same power supply that operates the plate electrodes of the tube. In practical resistance-capacity filters, the reactance of the shunting condenser is always made small compared with the series resistance of the filter and the load resistance to which the

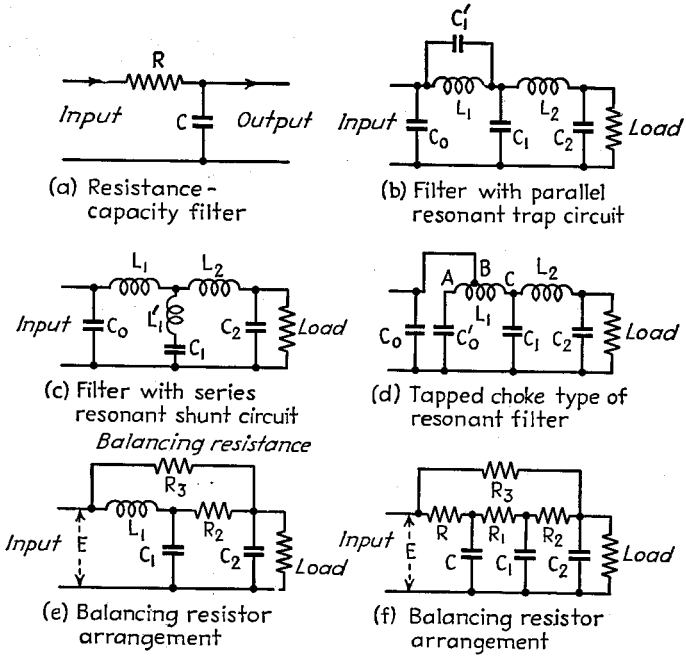


FIG. 13.—Miscellaneous filter circuits.

filter output is delivered. Under these conditions, the resistance-capacity filter produces a reduction in a-c voltage according to the relation

$$\frac{\left\{\begin{matrix}\text{Alternating voltage}\\ \text{across output of section}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Alternating voltage applied}\\ \text{to input of section}\end{matrix}\right\}} = \frac{1}{R\omega C} \qquad (5)$$

where R is the series resistance and C is the shunt condenser (see Fig. 13*a*). The results of Eq. (5) are given in convenient form in Fig. 14 for the usual case where the ripple frequencies involved are harmonics of 60 cycles.

Filters Employing Resonant Elements or Involving Balancing Arrangements.—The effectiveness with which a filter suppresses a particular ripple frequency can be greatly increased by using resonant impedance elements or by balancing arrangements. Thus tuning the series inductance to parallel resonance with the undesired ripple frequency or the shunt condenser to series resonance, as shown in Fig. 13, will greatly increase the attenuation of this frequency. Such arrangements have the disadvantage, however, of requiring a careful adjustment of circuit constants in order to be fully effective, and also offer less attenuation to the higher ripple frequencies.

Balancing systems are shown in Figs. 13*d*, 13*e*, and 13*f*. In the tapped-choke arrangement of Fig. 13*d*, the capacity of the condenser C_0' and the location of the tap on

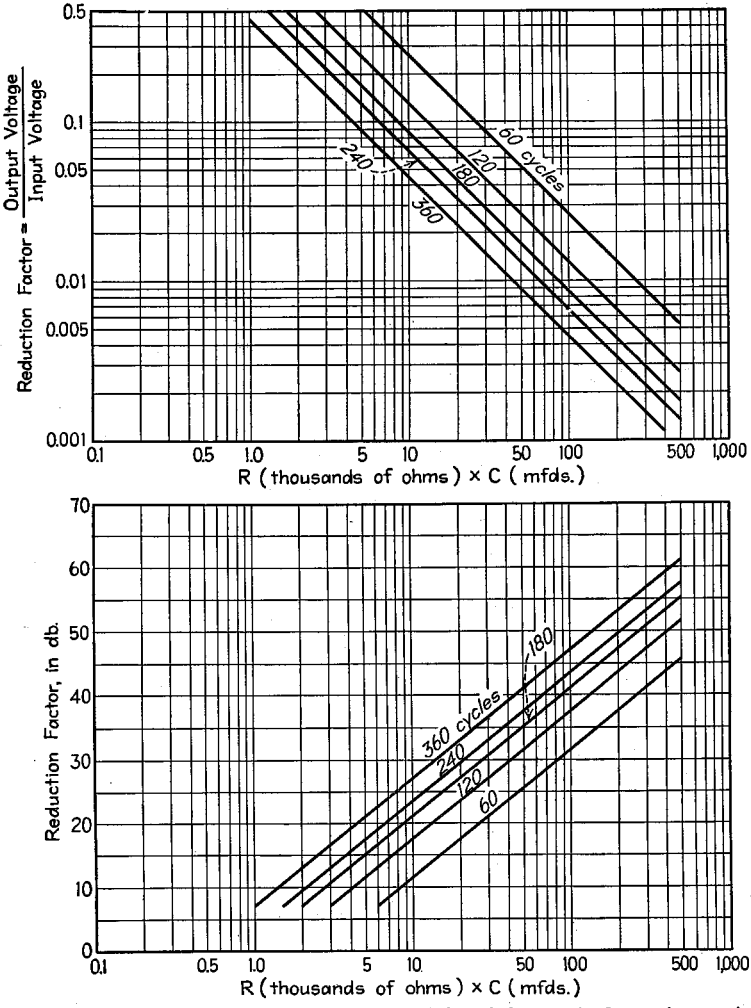


FIG. 14.—Reduction factor in ripple voltage produced by a single-section resistance-capacity filter having a resistance R and a capacity C.

inductance L_1 should be such as to satisfy the relation[1]

$$X_c = (n + 1)X_L \qquad (6)$$

where n = ratio of turns in section BC to number of turns in section AB of coil L_1.
X_c = reactance of condenser C_0' at the lowest ripple frequency.
X_L = reactance of the choke between points A and B at the lowest ripple frequency.

[1] This formula assumes that there is zero leakage inductance between the two parts of the choke. Experimental data showing the hum reduction that can be achieved in practice is given by B. F. Miessner, Hum in All-electric Radio Receivers, *Proc. I.R.E.*, Vol. 18, p. 137, January, 1930.

There is no ripple current in the section BC of the choke, because the current flowing through the section AB induces in the other section sufficient voltage to balance exactly the ripple voltage applied to the choke. Tapped chokes were once widely used in small rectifier-filter systems, but went into the discard with the coming of inexpensive high-capacity electrolytic condensers.

In the balancing arrangement of Fig. 13*e*, advantage is taken of the fact that there is 180° phase difference between the input voltage E and the ripple voltages developed across C_1. The ripple current delivered to C_2 through resistance R_2 by E can hence be balanced by a corresponding current delivered to C_2 by the resistance R_3, provided that

$$R_3 = \frac{R_2}{\left\{\begin{array}{l}\text{Reduction factor of section } L_1C_1 \\ \text{as given by Eq. (4)}\end{array}\right\}} \tag{7}$$

The circuit of Fig. 13*f* is similar to that of Fig. 13*e*, except that the single inductance-capacity section is replaced by two resistance-capacity sections, in each of which the shunting reactance is much less than the series resistance.[1]

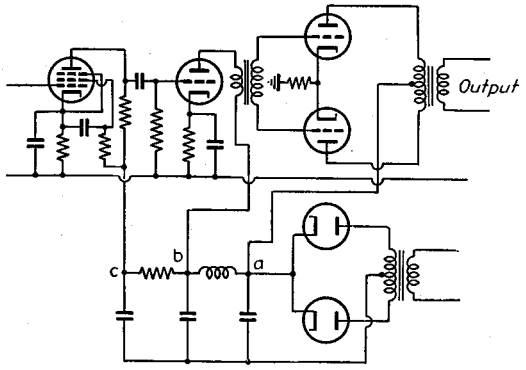


FIG. 15.—Example of graded filter in which the amplifier stages operating at lower power level are provided with progressively increased filtering.

Graded Filters.—A rectifier-filter system is commonly used to supply anode voltage to the several stages of an amplifier. In such an arrangement the amount of ripple or hum voltage that can be allowed to exist in the voltage applied to the various stages is less the lower the power level, because hum introduced at a low power level is amplified by the subsequent stages. This situation makes it desirable to arrange the power-supply system so that the voltages applied to different tubes operated from the system undergo differing amounts of filtering. An example of such a *graded filter* is shown in Fig. 15. Here the output stage is shown receiving its voltage directly from the input condenser, which is permissible because of the high power level at which the output stage operates, combined with the hum balancing action of the push-pull connection. Progressively greater filtering is provided for the lower level stages. In designing such an arrangement, one starts by arranging matters so that the hum voltage applied to the final stage is as large as, but no larger than permissible. Then the next filter section (*ab* in Fig. 15) is proportioned to give greater attenuation to the ripple voltage than the amplification that the ripple voltage undergoes in the final stage between *a* and *b*. Similarly, the filter section *bc* is proportioned to give at least as much reduction in hum as there is gain in the amplifier between points *b* and *c*.

[1] H. H. Scott, Simple Improvements in R-C Power-supply Filters, *Electronics*, Vol. 12, p. 42, August, 1939.

A graded filter reduces to the lowest possible value the amount of current that must be carried by the series-impedance arms of the filter, and so results in substantial economies over an arrangement in which the full rectifier output is subjected to the maximum amount of filtering. A graded filter also provides isolation between stages, thereby reducing low-frequency regeneration and motorboating in the case of audio-frequency amplifiers.

Filter Inductances and Capacities.—The inductance coils used in a filter must have laminated iron cores with an air gap that is sufficient to prevent the direct-current magnetization from saturating the core. The insulation between the winding and the core must be capable of withstanding the full d-c voltage delivered to the rectifier output (unless the filter is always to be used on the grounded side of the circuit). The inductance that is effective in the filter is the incremental inductance, which depends both upon the direct and the alternating magnetization of the core, as discussed in Par. 23, Sec. 2. In estimating the alternating magnetization that can be expected, it is normally permissible to assume that all inductances except that in the first section of an input-inductance filter carry only very small alternating currents. The first choke is subjected to appreciable alternating magnetization, however, since it carries a current that is given approximately by[1]

$$\left.\begin{matrix}\text{A-c current in} \\ \text{input inductance}\end{matrix}\right\} = \frac{E_1}{\omega L_1} \tag{8}$$

where E_1 is the amplitude of the lowest frequency component of the ripple voltage and ωL_1 is the reactance of the input inductance at this frequency.

The condensers used in filters must be capable of *continuously* withstanding a direct-current voltage that is equal to the crest alternating-current voltage applied to the rectifier. Ordinarily a single condenser should be used to withstand the entire voltage, rather than several condensers in series. When condensers are in series, the direct-current voltage stress divides between them in proportion to their leakage resistances rather than their dielectric strength, and the leakage resistances are variable and uncertain. If filter condensers are to be connected in series, it is advisable to shunt them with high resistances proportioned to divide the voltage in the correct ratios. Electrolytic condensers are ordinarily used where the peak voltages do not exceed 400 to 500 volts. Such capacitors have very low cost in proportion to capacity, although there is an accompanying disadvantage of limited life. Paper condensers are used when voltages exceeding 400 are involved, and also even with lower voltages if long life is important.

Filter condensers used in condenser-input systems associated with amplifiers and radio receivers are normally subjected to a brief interval of overvoltage when the power is first turned on. This is particularly the case when filament tubes are used in the rectifier and heater tubes are used in the power stages operated from the rectifier-filter system, because then the rectifier becomes fully operative before the load comes on. To avoid trouble under these conditions, it is necessary to provide a factor of safety in the voltage ratings of the condensers or to place somewhere in the system a voltage-regulating condenser (see Par. 34, Sec. 2).

6. Examples of Rectifier-filter Calculations.—The use of Table 1, Fig. 9, and the various equations relating to filters are illustrated by the following examples:

Example 1.—It is desired to design a three-phase star rectifier-filter system to operate from a 60-cycle power supply and to deliver a direct-current output of 2,500 volts and 0.4 amp with a ripple that must not exceed 2 per cent.

[1] This assumes that the reactance of the condenser in the first section is small compared with the reactance of the inductance, a condition very nearly realized under most practical circumstances. It also neglects currents of higher ripple frequencies, but these are so small and so phased that their effect is secondary.

If the direct-current resistance of the filter inductances is neglected, the rectifier must deliver a direct-current output voltage of 2,500 volts, and Table 1 shows that the rms voltage that each secondary leg must develop is $2,500 \times 0.855 = 2,135$ volts. Since the utilization factors of the primary and secondary, as given by Table 1, are 0.827 and 0.675, respectively, each leg of the primary must have a rating of

$$\frac{2,500 \times 0.4}{(3 \times 0.827)} = 403 \text{ watts}$$

and each leg of the secondary, a rating of $2,500 \times 0.4/(3 \times 0.675) = 493$ watts. Tentative calculation based on Eq. (4) shows that the filter of Fig. 11*a* with $C_1 = 1.0$ μf and $L_1 = 9.8$ henrys will keep the ripple voltage down to 2 per cent and will be generally satisfactory. Reference to Table 1 shows that the peak anode current would be 0.4 amp for infinite input inductance, and the maximum inverse voltage that each rectifier must stand is $2,500 \times 2.09 = 5,225$ volts. Type 866 mercury-vapor tubes will meet these requirements. In actual practice, the secondary voltage of the transformer would be made greater than 2,135 volts by perhaps 10 per cent to compensate for the loss of voltage caused by the resistance of the filter inductance, the voltage drop in the rectifier, the leakage reactance of the transformer, and the transformer resistance.

Example 2.—A load of 4,000 ohms is to be supplied by direct current from a condenser-input rectifier-filter system. The power transformer develops a potential of 300 volts rms on each half of the secondary. The effective leakage inductance and resistance are 40 mh and 125 ohms, respectively, and a Type 80 rectifier tube is used. The input condenser is 4 μf, and is followed by a single-section filter consisting of a 10-henry choke with 400 ohms resistance and an 8-μf condenser across the output. Determine the performance of the system.

The first step in the solution is the determination of the effective load and source resistances. The former is the actual load resistance plus the choke resistance, or $4,000 + 400 = 4,400$ ohms. The average resistance of the rectifier tube is (from tube manual) approximately $62/0.125 = 498$ ohms, so that the source resistance is $125 + 498 = 623$ ohms. Reference to Fig. 9 shows that for these values of source and load resistances and input capacity one has

	For $L = 0$	For $L = 0.1$	For $L = 0.040$
E_0/E	0.73	0.73	0.73
I_{max}/I_0	2.9	3.00	2.95
E_R/E_0	0.12	0.115	0.117

The last column is obtained by interpolation between the first two. From this column one has

D-c voltage across input condenser $= 0.73 \times 300\sqrt{2} = 309$
D-c voltage across actual load $= 309 \times 4,000/4,400 = 281$
D-c current in load $= 309/4,400 = 0.0703$ amp
Peak space current $= 0.0703 \times 2.95 = 0.207$ amp
Ripple voltage across input condenser $= 0.117 \times 309 = 36$ volts crest

$$\text{Ripple voltage across load [from Eq. (4)]} = \frac{36}{(754)^2 \times 10 \times 8 \times 10^{-6} - 1} = 0.79 \text{ volt crest}$$

MISCELLANEOUS

7. Anode Power from Low-voltage Direct-current Sources. *Vibrator Systems.*[1]—Direct current at high voltage can be obtained from a storage battery by using a vibrating contact to change the direct current delivered by the battery into alternating current that can be stepped up in voltage by a transformer and rectified. A typical arrangement of this type is shown in Fig. 16. Here a vibrating reed is provided, with contacts connected in the circuit in such a manner that the storage-battery voltage is first applied across one-half of the primary winding of a transformer, and then in the opposite direction across the other half. This induces an alternating voltage in the secondary having a value determined by the battery voltage and the transformer ratio. The reed is kept in vibration at its frequency of mechanical resonance by the magnet M, which is so arranged that when the reed is drawn to the magnet, the terminals of the latter are short-circuited, and the reed is allowed to spring back. The resistances R and capacity C_1 shown in Fig. 16 are for the purpose of minimizing sparking at the contacts, and the buffer condenser C associated with the secondary, as well as the radio-

[1] An excellent discussion of commercial types of vibrators is given in Vibrators, *Electronics*, Vol. 9, p. 25, February, 1936.

frequency chokes, are essential to control transient voltages and to prevent radio interference.

The secondary voltage developed by the transformer associated with the vibrator can be rectified by a rectifier or by an extra set of contacts on the reed, as in Fig. 17. The latter arrangement is termed a *synchronous vibrator*, and eliminates the cost of the rectifier tube and the voltage drop in the tube, but is more expensive and more likely to get out of order. Both high-vacuum and cold-cathode rectifiers are used in association with vibrator systems that require a tube.

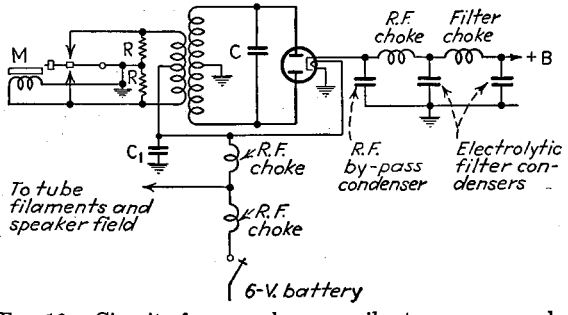


FIG. 16.—Circuit of nonsynchronous vibrator power supply.

The vibrating reed normally has a resonant frequency of the order of 90 to 150 cycles. The associated transformer and filter system are designed accordingly, instead of for a 60-cycle fundamental. In the design, it is to be kept in mind that the wave shapes of the voltages involved are square rather than sinusoidal. It is important that the transformer have the lowest possible leakage inductance in order to minimize sparking at the contacts, and it is desirable that the transformer be designed to operate at a low flux density.

Vibrator power supplies are used to operate radio receivers from low-voltage storage batteries, particularly automobile and aircraft equipment, and are also employed for small transmitters. Compared with the dynamotor, a vibrator power supply is less expensive but has shorter life, and tends to produce radio interference (often referred to as "hash").

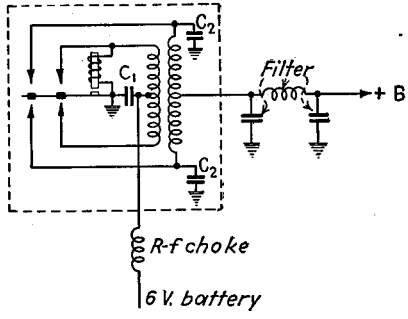


FIG. 17.—Circuit of synchronous vibrator power supply.

Dynamotor.—A dynamotor is a combined motor generator having two or more separate armature windings and a common field. One of the armature windings is operated from a low-voltage direct-current power source, commonly a storage battery, to give motor operation, while the other windings then serve as generators to produce the desired direct-current voltages.

Dynamotors are built in various sizes and ratings, ranging from small permanent magnet types intended for six-volt storage batteries, developing plate power for radio receivers and very small transmitters, to large machines supplying plate power for good-sized radio transmitters.

When plate power is obtained from a dynamotor or other equivalent generator of the commutator type, it is necessary to place a filter in the output of the machine to eliminate ripple voltages arising from the finite number of commutator bars and from miscellaneous irregularities.

8. Self-rectifying Circuits.—In a self-rectifying circuit, alternating voltage is applied directly to the plate of an oscillator or power amplifier. On the positive half of the cycle, output voltage will be developed, while on the negative half cycle, the tube will be inoperative. The usual self-rectifying circuit employs two tubes in each stage, so arranged that the two plates are supplied with alternating voltages that are 180° out of phase, and a large inductance L is inserted in the common filament return lead, as shown in Fig. 18. When the inductance is large, the total current drawn by the two tubes will be approximately constant, and the amplitude of the generated oscillations will be reasonably constant.

The radio-frequency output from a self-rectifying arrangement normally possesses both amplitude and phase modulation. The amount of amplitude modulation depends upon the size of the filter choke and is given approximately by the equation

$$\text{Amplitude modulation} = \frac{2}{3\sqrt{1+\left(\frac{\omega L}{R}\right)^2}} \qquad (9)$$

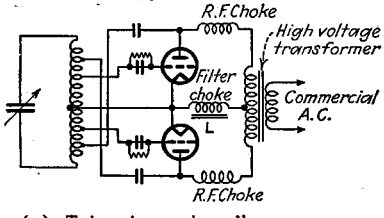
where ωL = reactance of filter inductance at twice the supply frequency.

R = equivalent d-c load resistance

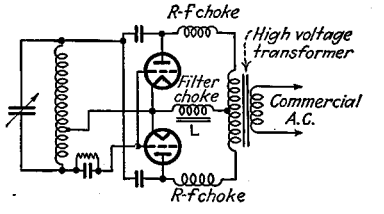
$$= \frac{\left\{\begin{matrix}1.11 \times \text{rms voltage on each} \\ \text{half of transformer secondary}\end{matrix}\right\}}{\text{d-c current through } L}$$

The phase modulation arises from the fact that the tubes operate alternately, and is less the more closely the two sides of the circuit are balanced.

Self-rectifying circuits find use where the complications of a rectifier and filter are to be avoided. They are also used to some extent in radio telegraphy for the purpose of creating a number of side bands in the radiated wave in order to minimize fading.



(a) Tubes in push pull



(b) Tubes in parallel

FIG. 18.—Typical self-rectifying oscillator circuits.

9. Voltage-regulating Systems.—The voltage developed by an anode-supply system may be made approximately independent of the a-c line voltage by (1) an electronic voltage stabilizer employing a series regulating tube, (2) magnetic voltage regulator utilizing magnetic saturation, (3) a neon voltage-regulating tube, (4) a ballast lamp. The last three of these arrangements can also be utilized to regulate the filament voltage obtained from a transformer.

Electronic Voltage Regulators.[1]—A typical regulator of this type is shown in Fig. 19, which also gives constants suitable for a small power-supply system.[2] The system operates in such a way as to cause the output voltage E_0 of the system to be substantially independent of the load impedance connected across E_0, or of the d-c voltage E developed by the filter output.[3] This is because any fluctuation in the output voltage

[1] An extensive discussion of electronic voltage regulators of various types is given by F. V. Hunt and R. W. Hickman, On Electronic Voltage Stabilizers, *Rev. Sci. Instruments*, Vol. 10, p. 6, January, 1939. Special types of electronic voltage regulators having extremely high stability in order to meet the severe requirements of electron microscopes are described by A. W. Vance, Stable Power Supplies for Electron Microscopes, *R.C.A. Rev.*, Vol. 5, p. 293, January, 1941.

[2] This design is given in A Voltage Regulator for D-C Power Supplies, *R.C.A. Mfg. Co. Application Note* 96, Aug. 24, 1938.

[3] The regulating system accordingly acts as a filter section by causing the hum voltage in the output E_0 to be much less than in the input E to the regulator system.

will vary the potential E_1 applied to the grid of T_2, and hence the grid-cathode potential, since the regulating action of neon tube T_3 keeps the cathode potential E_3 of T_2 constant. Any variation in output voltage is accordingly applied to T_2, and after amplification by this tube affects the grid potential of T_1 in such a way as to produce a change in voltage drop in this tube that opposes the change in output voltage. The value of output voltage that the system attempts to maintain is determined by voltage E_1 and hence can be readily controlled by the setting of the potentiometer R_2.

The fact that the output potential E_0 obtained in Fig. 19 tends to be independent of the load placed on the system *causes E_0 to act as though it had very low internal impedance.*[1] This property is very helpful in eliminating regeneration and motorboating in multistage audio amplifiers and in preventing coupling between different types of equipment operating from the same supply system.

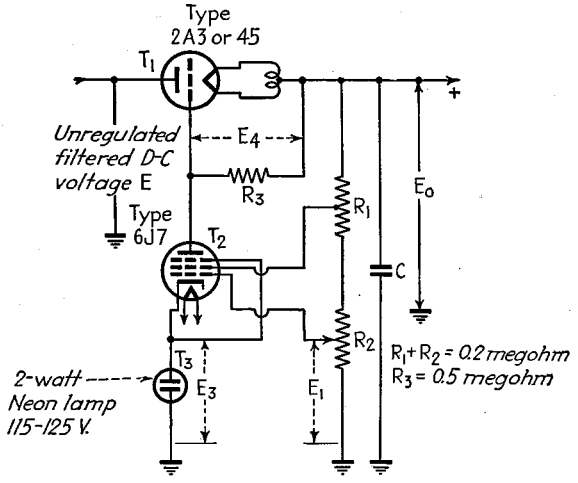


FIG. 19.—Circuit diagram of typical voltage regulator of the electronic type.

Voltage Regulators Using Magnetic Saturation.[2]—The fundamental principles involved in this type of voltage regulator can be explained with the aid of Fig. 20. Here T_1 is an ordinary transformer, and T_2 is a transformer designed so that when the primary winding is placed in series with T_1, the core of T_2 will be at least moderately saturated over the range of applied voltage for which constant output voltage is desired. The secondary windings are so poled as to oppose each other, with the saturated transformer developing most of the secondary voltage. When the alternating voltage applied to such an arrangement increases in amplitude, the division of voltage between the two transformers changes, with the saturated transformer T_2 taking proportionately less of the total, as shown in Fig. 20*b*. This causes the secondary voltages to become more nearly equal, and by properly proportioning the system, the output will remain substantially constant over an appreciable range of applied voltage, as shown.

The actual details of magnetic voltage regulators can be varied in many respects. Thus it is possible to combine the two transformers in Fig. 20*a* into a single three-

[1] This low internal impedance is effective only at low and moderate frequencies, such that the neon lamp is able to follow the cyclic variations in current. A low impedance can be assured at higher frequencies by shunting the output with a large condenser C, as shown in Fig. 19.

[2] An excellent survey of regulators of this type is given by K. J. Way, Voltage Regulators Using Magnetic Saturation, *Electronics*, Vol. 10, p. 14, July, 1937.

legged transformer. The windings can also be arranged in various combinations, including auto-transformer connections. A condenser is sometimes placed across the saturated transformer, as shown dotted in Fig. 20*a*, for the purpose of improving the voltage regulation with variation in the applied voltage and with load changes and for improving the wave shape of the output voltage.

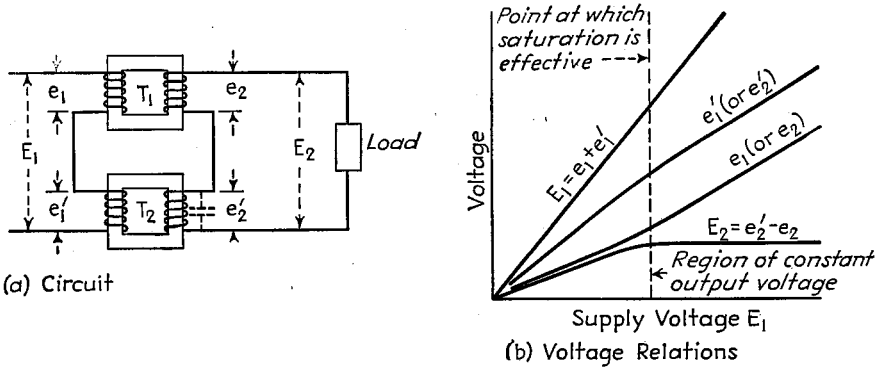


FIG. 20.—Circuit and relations existing in a typical voltage regulator of the magnetic-saturation type.

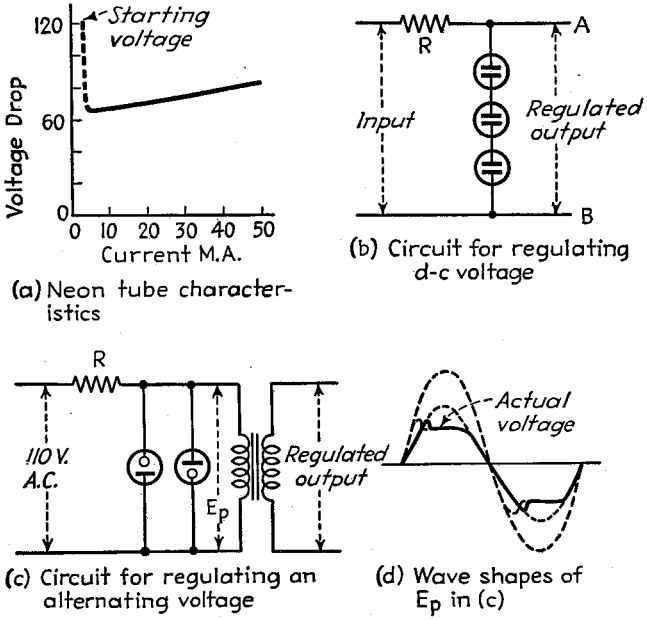


FIG. 21.—Methods of utilizing a neon tube for voltage regulation.

Voltage regulators of the magnetic-saturation type are capable of maintaining the output voltage constant to ± 1 per cent, with variations in applied voltage of ± 15 per cent. Such regulators are widely used in conjunction with filament and plate-supply transformers, and are commercially available in ratings from a few watts to 25 kilowatts.

Neon-tube Voltage Regulators.—The fact that neon and similar cold-cathode gaseous tubes give a substantially constant voltage drop over a wide range of current, as shown

in Fig. 21*a* can be utilized in a variety of ways to regulate voltage. Thus in Fig. 21*b*, a series of neon tubes are placed across terminals *AB* where the d-c voltage is to be maintained constant, and the arrangement is then fed through a series resistance R that absorbs any fluctuation in supply voltage. Such an arrangement not only causes the output voltage to be substantially independent of the supply voltage but also tends to make the output independent of the load impedance placed upon the system. It accordingly gives the output of the system a very low internal impedance, and reduces hum.

Another type of voltage-regulating system is shown in Fig. 21*c*.[1] Here neon tubes are placed back to back across the 110-volt supply to a transformer, and a resistance *R* is placed in series with the line. This limits the peak voltage applied to the transformer to a fixed value, irrespective of the supply voltage, and so tends to keep the wave at the terminals of the transformer constant, as shown at Fig. 21*d*.

Voltage Regulators Employing Ballast Resistances.[2]—A ballast resistance is a resistance element having a temperature coefficient of resistivity such that over an appreciable range of voltage, the current through the resistance is independent of voltage Such a characteristic is illustrated in Fig. 22. A ballast resistance placed in series with the filament of a tube or the primary of a transformer will accordingly absorb moderate voltage fluctuations and maintain the current through the tube constant.

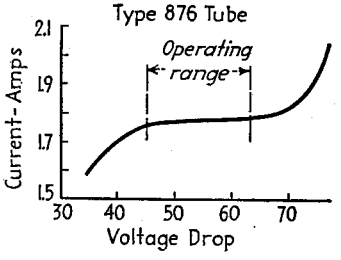


FIG. 22.—Characteristics of a typical ballast tube, showing how the tube operates as a constant-current device over an appreciable voltage range.

10. Problems Arising from Use of Alternating Current to Heat Cathodes.—When alternating current is used to heat the cathode of a vacuum tube, it is to be expected that the amplified output of the tube will contain a certain amount of hum. This hum is either in the form of alternating components in the plate current that are in harmonic relationship to the heating voltage or in the form of a corresponding modulation of the voltages and currents being amplified by the tube.

Hum from Alternating Filament Current.[3]—The chief causes of hum in filament tubes are (1) alternating voltage drop in the filament; (2) magnetic effect of the alternating filament current on the space current; (3) temperature variations of the cathode.

The effect of an alternating voltage drop in a filament cathode upon the plate current can be minimized by connecting the grid- and plate-return leads to a point having the same potential as the center of the filament. This can be done by the use of either a center-tapped resistance across the filament or a center-tapped winding on the filament transformer. Connecting the grid- and plate-return leads in this way eliminates the principal cause of hum having the same frequency as the filament heating current. However, even with the center-tapped filament connection, there still remains a residual component of hum having twice the supply frequency. This results from the fact that with a center-tapped connection, the half of the filament that is negative at the moment supplies more electrons to the anode than does the positive half, and since the electron flow is proportional to the three-halves power of the electrostatic field strength, the current from the negative end of the field is increased more by the voltage

[1] G. F. Lampkin, A Simple A-C Voltage Regulator, *Electronics*, Vol. 10, p. 30, Autust, 1937.

[2] For further information see S. G. Gordon, Ballast Tubes as Automatic Voltage Regulators, *Electronics*, Vol. 15, p. 26, January, 1942.

[3] W. J. Kimmell, The Cause and Prevention of Hum in Receiving Tubes Employing Alternating Current Direct on the Filament, *Proc. I.R.E.*, Vol. 16, p. 1089, August, 1928.

drop than the current drawn from the positive end is decreased. Hence each time there is a voltage drop, which is twice per cycle, the space current increases.

The magnetic field produced by the filament current introduces a double-frequency hum by virtue of the fact that the electrostatic field deflects the electrons from the paths that they would otherwise follow, and so tends to decrease the plate current slightly. This effect is exactly out of phase with that arising from voltage drop in the filament, and by proper design the two can be made to cancel each other for the normal operating condition.

Alternating filament current produces its maximum heating effect at the peaks of the cycle. The filament temperature and hence the emission accordingly fluctuate at twice the supply frequency. However, with ordinary filaments, the heat capacity is such that there is extremely small change in temperature with a 60-cycle supply. Temperature fluctuations are therefore usually unimportant as a source of hum, particularly if a full space charge is present.

Factors favoring a low double-frequency hum with a-c heating of a filamentary cathode are (1) filament designed for relatively low voltage and high current; (2) a *V* or *W* arrangement of the filament wires to reduce the magnetic effect and to permit exchange of electrons between opposite ends of the filament in order to minimize the effect of voltage drop in the filament; (3) tube design such that the magnetic and voltage drop components of hum are approximately equal at the usual operating point.

Filamentary cathodes are always used in large high-vacuum tubes, in many hot-cathode mercury-vapor rectifiers, and to some extent in small power tubes. They are seldom used in radio-frequency or audio-frequency voltage amplifiers or in detector tubes, except low-filament power tubes intended for battery operation.

Polyphase and Quadrature Arrangements for Filament Heating.—Large transmitting tubes having a multistrand filament are frequently arranged so that the different strands can be excited from different phases of a polyphase source. This results in a material reduction in hum.

In large hot-cathode mercury-vapor tubes employing filamentary cathodes, it has been found desirable to make the filament current 90° out of phase with the voltage applied to the anode of the rectifier tube. In this way, the plate current flows when the voltage drop in the filament is minimum. The tube then acts more nearly as though it had an equipotential cathode, and so gives more uniform behavior along the length of the filament than would otherwise be the case. This is important in the case of hot-cathode mercury-vapor tubes, because the average anode drop is only 10 to 15 volts, and a few volts' difference in potential between the ends of the filament will greatly influence the behavior.

Hum with Heater Cathodes.[1]—The use of a heater-type cathode removes most of the factors that produce hum in filament tubes, but there is still present a residual hum as a result of the following causes: (1) magnetic field produced by the heater current; (2) leakage resistances from heater to other electrodes; (3) electrostatic fields within the tube arising from unshielded portions of the heater or heater leads; (4) electron emission from the heater.

The magnetic field of the heater introduces a double-frequency hum, just as in filament tubes. This can be minimized by using a high-voltage, low-current heater and employing proper geometry in the heater so that a minimum of external field is produced.

[1] J. O. McNally, Analysis and Reduction of Output Disturbances Resulting from the Alternating-current Operation of the Heaters of Indirectly Heated Cathodes, *Proc. I.R.E.*, Vol. 20, p. 1263, August, 1932.

Leakage paths from the heater to other electrodes cause voltages of the power-supply frequency to be developed between the electrodes and ground.[1] The most important leakage of this character is that from heater to cathode. Any leakage current of this character must flow through the bias resistance between cathode and ground, and will produce a voltage that is effectively applied between grid and cathode. Hum from this source can be eliminated by more adequate by-passing of the bias resistance, by grounding the cathode, and in some cases by properly biasing the heater with respect to the cathode. Leakage from the heater to other electrodes, notably the grid, is usually the result of getter material deposited between the leads of the tube. The remedy in such cases is either to try another tube, lower the impedance between the affected electrode and ground, or bias the heater with respect to ground.

Unshielded portions of the heater produce electrostatic fields that may influence the space current and result in 60-cycle hum. Electrostatic fields from the heater leads may also have direct capacitive coupling to other electrodes, and result in the production of hum voltages that are not negligible if the electrode impedance to ground is large.

When low hum is desired and the heater is to be operated with alternating current, it is desirable to select the tube type carefully. Special low-hum tubes are available to meet very severe requirements, and commercial tubes used in broadcast receivers vary greatly from type to type as a result of differences in heater geometry, shielding, etc. Some differences in hum between individual tubes of the same type and make can be expected as a result of chance differences in leakage films. Where complete absence of hum is essential, as in very high-gain audio amplifiers, the use of direct current for the heater is recommended.

11. Bias Voltage.—The grid bias for voltage amplifiers and small power amplifiers is nearly always derived from the plate-supply system. The commonest method of doing this is to use a resistance between cathode and ground that makes the cathode positive by the resistance drop produced by the *total space current.* This arrangement is satisfactory wherever the total space current is substantially constant during operation, as in Class C, Class A, linear, and most types of modulated amplifiers. It is not suitable for Class B and Class AB audio amplifiers, however. To avoid negative feedback type of regeneration, the cathode biasing resistor must be shunted by a condenser such that the impedance between cathode and ground to the alternating currents being amplified satisfies the relation[2]

$$g_m Z_f << 1 \tag{10}$$

where g_m is the transconductance of the tube and Z_f is the impedance between cathode and ground to the currents being amplified. In the case where the tube under question is excited by transformer coupling, negative feedback can be avoided by the use of a decoupling network RC, as illustrated in Fig. 23. In this arrangement, the resistance R corresponds to a grid leak and can be quite high. The capacity C must be large enough that its reactance at the lowest frequency for which feedback is to be avoided is appreciably less than the resistance R. Since R is much greater than the bias resistance, a much smaller condenser is required in this circuit than in the usual arrangement consisting of a bias resistor shunted by a condenser.

[1] See Heater Cathode Leakage as a Source of Hum, *Electronics*, Vol. 13, p. 48, February, 1940; also, Notes on Hum in Heater-type Tubes, *R.C.A. Application Note* 88.

[2] This applies directly to the case of a Class A pentode tube, and is based upon the analysis given in Par. 4, Sec. 5. In the more general cases, the requirement is that the alternating voltage developed across the bias impedance by the space current be small compared with the signal voltage applied to the grid of the tube.

When the bias is obtained from a cathode resistor, it is necessary that the power-supply system be capable of developing an output potential greater than the desired anode-cathode voltage by the value of the bias potential.

The circuit of Fig. 24 employs the resistance drop in the filter choke to develop the bias. In this way, use is made of a potential that is otherwise wasted, and the total voltage that must be developed by the output of the filter is only that required by the anode of the tube. With this circuit, it is necessary to employ a resistance-capacity filter *RC* to prevent the hum voltage developed across the filter chokes from reaching the grid of the tube.

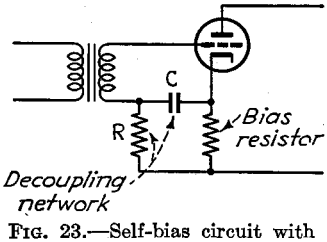


FIG. 23.—Self-bias circuit with decoupling network.

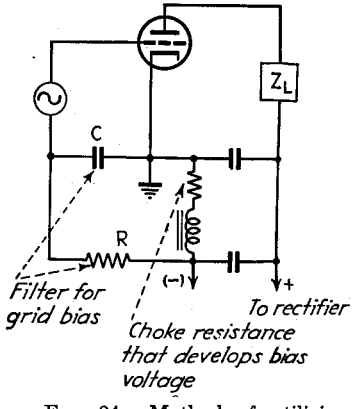


FIG. 24.—Method of utilizing the voltage drop in the filter choke to produce grid bias.

Small batteries are sometimes used for bias purposes, particularly in portable equipment and in laboratory apparatus. These may be either small dry cells or specially constructed bias cells having an extremely low current capacity and high internal resistance but very long shelf life.

Bias voltages for large power tubes can be obtained from a self-bias arrangement, direct-current generators, rectifier-filter systems, and in the case of Class C amplifiers and oscillators, by grid-leak grid-condenser combinations. When rectifier-filter systems are employed, it is essential that a load resistance be connected across the output of the filter to provide a path by which d-c grid current can return to the cathode. With direct-current generators, a filter must be employed between the generator and tube to eliminate commutator ripple.

SECTION 9

RADIO TRANSMITTERS AND RECEIVERS

TRANSMITTERS EMPLOYING AMPLITUDE MODULATION

1. Radio-telephone Transmitters. *Broadcast Transmitters.*—Broadcast transmitters represent the highest development of radio-telephone transmitters with respect to stability of carrier frequency, band width, low distortion and noise, etc. Such transmitters normally consist of a crystal oscillator followed by several buffer amplifier stages, a modulated amplifier, audio-frequency modulating system, etc., plus accessories such as protection equipment, monitoring facilities, etc.[1]

The crystal oscillator of a broadcast transmitter is operated at the carrier frequency, and normally employs a zero-temperature-coefficient crystal operating at low-power level and under conditions such that the very greatest in frequency stability will be realized. The buffer amplifiers are for the purpose of isolating the crystal from the modulated amplifier, and employ screen-grid, pentode, or neutralized triode tubes. Power amplifiers (including buffer stages) operating at lower power levels than the modulated stage are of the Class C type, while all power amplifiers following the modulated stage must be linear amplifiers.

Modulation may be accomplished either in the last stage of radio-frequency amplification or at a lower power level. The former is termed *high-level modulation;* the latter, *low-level modulation.* In low-level systems, the radio-frequency amplifiers following the modulated stage are linear amplifiers, and are usually of the high-efficiency type when appreciable energy is involved.

High-level modulation usually takes the form of plate modulation with a Class B audio amplifier. There has, however, recently been developed a commercial high-level system employing the high-efficiency grid-modulated amplifier discussed in Par. 3, Sec. 7.

A variety of modulation methods are in use for low-level modulation, including plate, conventional control-grid and suppressor-grid modulation, in this order of popularity. Almost any system that gives complete modulation can be employed, since it is possible to take care of distortion with the aid of negative feedback.

Short-wave and Ultra-high-frequency Radio-telephone Transmitters.—The only essential difference between broadcast and short-wave telephone transmitters is that in the latter case the crystal oscillator operates at a subharmonic of the desired frequency, and harmonic generators as required are used between the crystal oscillator and the modulated stage. In many applications of short waves, it is not necessary to meet the same standards of performance as in broadcast work, with the result that buffer amplifiers are sometimes reduced in number, or even eliminated, and other simplifications are permitted that result in reduced band width, more distortion, and greater noise than would be tolerable in broadcast work.

[1] Typical broadcast transmitters of various types are described in the following papers: J. B. Coleman and V. A. Trouant, Recent Developments in Radio Transmitters, *R.C.A. Rev.*, Vol. 3, p. 316, January, 1939; R. E. Coram, Improved Design for Five-kilowatt Broadcast Transmitter, *Bell Lab. Rec.*, Vol. 17, p. 7, September, 1938; J. R. Poppele, Design and Equipment of a Fifty-kilowatt Broadcast Station for WOR, *Proc. I.R.E.*, Vol. 24, p. 1063, August, 1936; J. A. Chambers, L. F. Jones, G. W. Fyler, R. H. Williamson, E. A. Leach, and J. A. Hutcheson, The WLW 500-kilowatt Broadcast Transmitter, *Proc. I.R.E.*, Vol. 22, p. 1151, October, 1934.

Transmitters for short waves, unlike those used in the ordinary broadcast band, are nearly always designed for operation at a number of frequencies, with facilities provided for relatively rapid frequency change.

Transmitters for use at ultra-high frequencies frequently employ resonant-line oscillators instead of crystal control.[1] Such oscillators are sometimes used to generate directly the required power output, but in other cases a power amplifier, and sometimes a buffer amplifier and power amplifier, are employed in association with the resonant-line oscillator to provide isolation.

Negative Feedback in Radio-telephone Transmitters.—The performance of a radio-telephone transmitter can be improved in many respects by rectifying a portion of the output wave and feeding the resulting rectified output back into the audio-frequency system of the transmitter in phase opposition to the audio-frequency voltages being

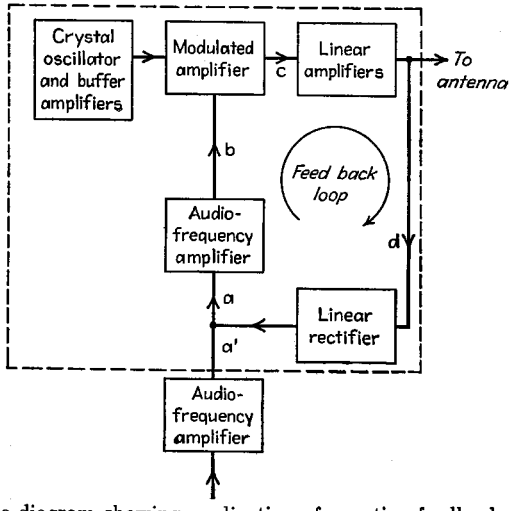


FIG. 1.—Schematic diagram showing application of negative feedback system to a radio-telephone transmitter.

amplified, as illustrated schematically in Fig. 1.[2] This is a form of negative feedback of the type discussed in Par. 11, Sec. 5, and to the extent that the rectifier is distortionless, does for the modulation envelope of the transmitter output what negative feedback in audio amplifiers does for the output wave in audio systems. These consequences can be summarized briefly as follows: The audio gain is reduced so that more amplification is required in the modulator. Amplitude, phase, and frequency distortion generated anywhere within the loop *abcd* in Fig. 1 are reduced as far as their effects upon the transmitter output are concerned, and all forms of noise modulation, including hum, induced anywhere in the part of the transmitter enclosed by the dotted rectangle in Fig. 1 are likewise discriminated against.

The essential quantitative relations involved in negative feedback systems applied to transmitters are as follows:

[1] Examples of such transmitters are described by W. I. Harrington and C. W. Hansell, The Hawaiian Radio-telephone System, *Elec. Eng.*, Vol. 54, p. 822, August, 1935; J. E. Smith, F. H. Kroger, and R. W. George, Practical Application of an Ultra-high-frequency Radio-relay Circuit, *Proc. I.R.E.*, Vol. 26, p. 1311, November, 1938.

[2] The remodulation system of distortion correction described in Par. 22, Sec. 5, can likewise be used to improve the characteristics of an individual linear amplifier stage.

$$\frac{\text{A-f amplitude of modulation envelope of output wave}}{\text{A-f voltage at input of feedback loop (point } a' \text{ in Fig. 1}} = \frac{A}{1 - A\beta} = -\frac{1}{\beta}\frac{1}{1 - \frac{1}{A\beta}} \tag{1}$$

$$\frac{\text{Amplitude distortion in output envelope with feedback}}{\text{Amplitude distortion in output envelope without feedback}} = \frac{1}{1 - A\beta} \tag{2}$$

$$\frac{\text{Noise (hum) modulation of output envelope with feedback}}{\text{Noise (hum) modulation of output envelope without feedback}} = \frac{1}{1 - A\beta} \tag{3}$$

where A is the ratio of the audio-frequency component of the modulation envelope of the transmitter output to the audio-frequency voltage at point a in the modulating system (see Fig. 1) and β is the ratio of the audio-frequency voltage that the rectifier output superimposes on the audio system at a to the audio component of the modulation envelope at d.

It will be noted that these equations are the same as Eqs. (36), (37), and (38) of Sec. 5 that apply to ordinary feedback. In particular, the quantity $A\beta$ is a measure of the amount of feedback, exactly as in the corresponding audio-frequency case. When the feedback factor $A\beta$ is large, Eq. (1) shows that the modulation envelope of the output tends to reproduce exactly the audio-frequency voltage at point a' in Fig. 1, with corresponding reduction in noise modulation, and frequency and phase distortion of the output envelope.

The reduction in distortion, noise, etc., resulting from the use of negative feedback is exactly proportional to the reduction in gain caused by feedback and is accordingly commonly measured in this way. In a typical broadcast transmitter, the amount of feedback that is used is generally of the order of 15 to 30 db.

Negative feedback employed with transmitters introduces the possibility of oscillation, exactly as does feedback in corresponding audio-frequency systems. The criterion for avoiding oscillation is the same in the two cases—namely, the feedback factor $A\beta$, when plotted in the complex plane, must not enclose the point (1,0) (see Fig. 32, Sec. 5). In practice, this normally means that at frequencies where the phase of $A\beta$ has shifted 180° from the normal value obtained in the mid-frequency range and so transformed the negative feedback into positive feedback, the magnitude of $A\beta$ must have dropped to less than unity. A typical situation is shown in Fig. 2.

The variation in magnitude and phase of $A\beta$ with frequency includes not only the effects of the audio-frequency amplifier in the feedback loop *abcd* of Fig. 1 but also the changes in the modulation envelope produced by the radio-frequency circuits in this loop. Thus side-band trimming in the tuned circuits will reduce the magnitude of A, and phase shifts of the side-band frequencies with respect to the carrier will change the phase of the modulation envelope.

The effect of the radio-frequency circuits on the phase of the modulation envelope is particularly important in negative feedback systems. In the case of a symmetrical resonance curve with the resonant frequency adjusted to the carrier frequency of the modulated wave, *the phase shift of the modulation envelope is the same in degrees at the modulation frequency as the phase shift of the corresponding side bands in degrees at radio frequency.* Thus a simple tuned circuit carrying a wave modulated at a frequency such that the side bands suffer a symmetrical phase shift of 45° will cause the modulation envelope to be likewise shifted 45° at the modulation frequency. As a consequence of this behavior, each simple resonant circuit in the radio-frequency system can intro-

duce phase shifts up to 90° at high modulation frequencies, and a simple band-pass circuit will give rise to phase shifts that can reach 180°. At low modulation frequencies, the radio-frequency circuits will introduce negligible phase shift of the modulation envelope. This gives the result illustrated in Fig. 2.

The circuits of a transmitter must be designed with considerable care if a large amount of feedback is to be employed. The fundamental principles involved are discussed in Par. 27, Sec. 3. The important requirement is that the transmission around the feedback loop at high frequencies shall not fall off too rapidly. This normally means that the attenuation must not change more rapidly than 10 db per octave. One way to achieve this is to give the radio and audio systems a much broader frequency char-

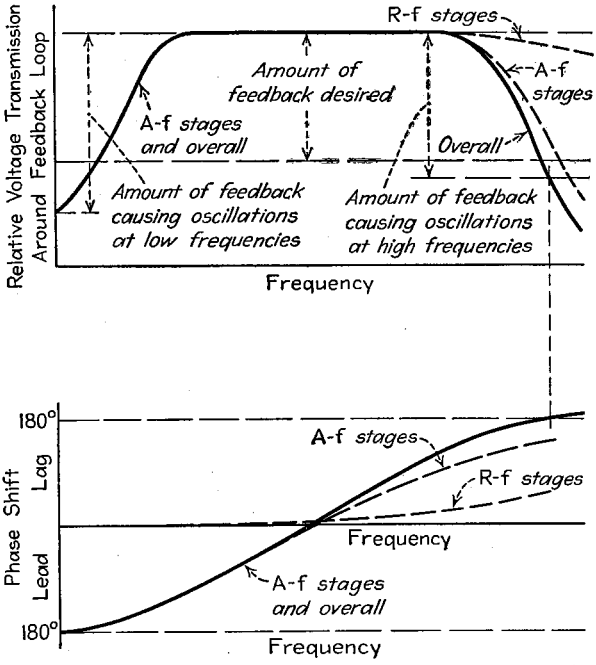


FIG. 2.—Hypothetical phase and amplitude characteristics of a broadcast transmitter and their relation to the amount of negative feedback that can be employed without producing oscillation.

acteristic than is actually required, and then to introduce supplementary attenuation in the feedback loop to control the attenuation characteristics at high frequencies. An alternative to this procedure is to design the audio- and radio-frequency circuits for the normal transmission band and then to provide auxiliary high-frequency amplification, or some form of high-frequency peaking network in the feedback circuit to compensate for the tendency of the transmission around the feedback loop to fall off rapidly at high frequencies.

The usual rectifier in feedback systems is a diode detector. The load circuits of the diode must be proportioned so that negative peak clipping is avoided. This can be achieved by appropriate circuit design, by the use of direct coupling, or by applying a suitable positive bias on the diode plate. The detector should be operated with a relatively large input voltage in order to minimize distortion. A full-wave rectification circuit is desirable, since then the carrier component is largely balanced out of

the output, making it possible to reduce the size of the shunting condensers in the r-f filter in the detector output.

The most important practical benefits gained from the use of negative feedback in transmitters are reduction in distortion and in noise modulation. The reduction in distortion makes it possible to sacrifice linearity of modulation and amplification for efficiency, and in particular makes it practicable to use high-efficiency systems that without feedback would have such high distortion as to be unsatisfactory for most applications. The reduction in noise modulation resulting from negative feedback makes it practicable to use alternating filament current in the radio-frequency tubes and to employ a power-supply system with relatively light filtering and still keep the hum modulation in the transmitter output down to a reasonable value.

These benefits of negative feedback result in improved performance at lower cost. At the same time, negative feedback is not a cure-all for all transmitter troubles, and there are certain things that it will not accomplish. In particular, negative feedback will not prevent distortion arising from overmodulation. Also, when there is side-band trimming at high modulation frequencies, the benefits of negative feedback will be effective only if the degree of modulation is not excessive. If the wave is fully modulated and there is side-band trimming in a stage following the modulation, negative feedback will attempt to overmodulate the transmitter to correct for the falling off in output, and will thereby introduce distortion.

Special Features of Audio-amplifier Systems.[1]—The audio-frequency amplifier systems used for modulation of radio-telephone transmitters are often provided with such features as peak limiters, volume compressors, vogads (or automatic volume control), etc. Peak limiters are universal in modern broadcast transmitters, and are for the purpose of preventing overmodulation with consequent adjacent channel interference on occasional peaks of very high intensity. Volume compressors are used to raise the level of average modulation that can be maintained without overmodulating during periods of high intensity. They improve the signal-to-noise ratio, particularly during weaker passages, at the expense of lowered volume range. Volume compressors are used in some broadcast transmitters and are very common in long-distance short-wave radio-telephone circuits. Vogads are for the purpose of maintaining full modulation at the transmitter irrespective of the amplitude of the incoming audio signal that is to be modulated upon the transmitter. Vogads are standard equipment in radio extensions of land telephone circuits.

Audio-frequency systems used in radio-telephone transmitters are sometimes intentionally designed with an amplitude-frequency characteristic that is not constant. In particular, transoceanic telephone circuits are normally arranged so that the high frequencies are accentuated before modulation on the transmitter, and are then brought back to their proper proportion at the receiver. Similar arrangements have been proposed for frequency modulation, and for the sound accompanying television signals. Preemphasis of the high frequencies in this way increases the signal-to-noise and signal-to-distortion level at the higher audio frequencies.[2] This gain is important because the high frequencies have relatively low amplitude, and so normally have an inherently less favorable ratio of signal to noise and distortion than is desirable. The low modulation frequencies are sometimes reduced in relative amplitude, or entirely eliminated, in radio systems designed to transmit speech where the emphasis

[1] A good description of the audio frequency equipment involved in a modern broadcast station is given by H. A. Chinn, Broadcast Studio Audio-frequency Systems Design, *Proc. I.R.E.*, Vol. 27, p. 83, February, 1939. The use of such special devices as vogads, vodas, compounders, codons, etc., in transoceanic radio-telephone circuits as described by R. A. Heising, Radio Extension Links to the Telephone System, *Bell Sys. Tech. Jour.*, Vol. 19, p. 611, October, 1940.

[2] Typical experimental results showing the improvement in signal-to-noise ratio from high-frequency pre-emphasis are reported by J. L. Hathaway, High Frequency Pre-Emphasis, *Electronics*, Vol. 12, p. 29, November, 1939.

is upon intelligibility rather than naturalness. Such discrimination against the lower speech frequencies reduces considerably the total audio-frequency energy that must be handled, and so makes it possible to increase the degree of modulation for the more important frequencies. The loss in intelligibility is small, and the gain in signal-to-noise ratio great.

Noise (Hum) in Transmitters.—The term *noise* as applied to radio transmitters includes all spurious modulations of the output voltage such as may arise from hum effects, vibration, thermal noise, etc. The amount of hum present is commonly expressed in terms of the level in decibels of the rms hum modulation with respect to 100 per cent modulation. In some cases the different frequency components composing the noise are weighted according to the sensitivity of the ear before the rms value is determined.

In broadcast transmitters the noise generated within the transmitter should be so low that the noise voltages on the incoming program lines control the noise level in the transmitter output. This requires that the noise generated within the transmitter have an unweighted value of at least 60 db below 100 per cent modulation.

The chief source of noise in broadcast transmitters is a-c hum modulation arising from the use of alternating filament current and from imperfectly filtered plate-supply voltages. Hot-cathode mercury-vapor rectifier tubes also may introduce "hash" unless the rectifier circuits are properly shielded and provided with radio-frequency filters. The most effective method of reducing noise modulation in a transmitter is to employ negative feedback, and this represents one of the chief benefits gained from negative feedback in transmitters.[1]

In aircraft transmitters, vibration of tubes and circuit elements will modulate the output wave and introduce noise. This is commonly handled by rubber shock mountings, inclosing the transmitter to protect it from sound waves, and by the use of negative feedback.

Synchronization of Broadcast Transmitters.—Two or more broadcast transmitters are occasionally modulated with identical programs and then operated with their carriers synchronized. Synchronization can be accomplished in several ways.[2] One method is to transmit to the transmitters an audio- or carrier-frequency wave that is a subharmonic of the desired carrier, and then to generate the latter by the use of harmonic generators. Another method is to derive from one transmitter by means of multivibrators a low frequency that is a subharmonic of the desired carrier, and then to transmit this over wires to the other transmitter, using harmonic generators to obtain the carrier frequency for the latter station. Still another arrangement[3] consists in generating the carrier of each station independently by very stable crystal oscillators. The two carrier frequencies are then compared by a monitoring station midway between the two transmitters and any difference in frequency transmitted to one of them. The operating personnel then makes whatever manual adjustments are required to maintain synchronism within at least one cycle in five seconds.

In order to obtain proper reception of a program when synchronized transmitters are employed, it is necessary that one signal dominate. Otherwise the phase differences with which the carrier and side bands combine at the receiver introduce the possibility of serious distortion. Experiments have indicated that when two transmitters are modulated with identical programs, it is necessary that the carrier ratio be

[1] Noise may also be introduced into the output in the form of phase modulation. Such modulation is not affected by negative feedback, and will introduce noise in the signal that is received at a distant point, as discussed in Par. 3.

[2] L. McC. Young, Present Practice in the Synchronous Operation of Broadcast Stations as Exemplified by WBBM and KFAB, *Proc. I.R.E.*, Vol. 24, p. 433, March, 1936.

[3] G. D. Gillett, Some Developments in Common Frequency Broadcasting, *Proc. I.R.E.*, Vol. 19, p. 1347, August, 1931.

of the order of 10 to 12 db for the distortion to be undetectable with perfect synchronization and 20 db if the carriers differ in frequency by one cycle every 5 to 20 seconds.[1] It is also necessary that time-delay effects in the audio-frequency program circuits be equalized so that the programs will be modulated upon the carrier simultaneously. Time differences of the order of 50 microseconds will produce distortion and time differences of 5 milliseconds will give rise to noticeable "echoes."

Single-side-band and Asymmetric-side-band Transmissions.—Single-side-band transmission is used both in long- and in short-wave transoceanic extensions of telephone systems. In transmitters of this type the single side band is generated as described in Par. 7, Sec. 7, and is then raised to the desired power level by linear amplifiers.[2] In some single-side-band systems, the carrier is transmitted at a reduced level (normally 10 to 25 db below its normal value for double-side-band transmission) for the purpose of operating the automatic frequency control on the receiver oscillator and also the automatic-volume-control system of the receiver.

Single-side-band transmission has the advantage over double-side-band operation in that the amount of side-band energy is greater in proportion to the power rating of the transmitter. The frequency band occupied by the signal is also half as great, and the signal-to-noise ratio is improved because of the narrower band and increased side-band power. The disadvantage of single side-band operation is that the transmitters and receivers are quite complicated.

The use of asymmetrical side-band transmission, as discussed in Par. 7, Sec. 7, has been suggested for broadcast work. Such a system occupies only about 60 per cent of the band width of double-side-band systems, but has the disadvantage of a more involved transmitter and of requiring a special tuning procedure at the receiver to place the carrier properly at the edge instead of the center of the receiver transmission band. Asymmetrical-side-band systems have not been used in commercial broadcasting, although they are used in television.[3]

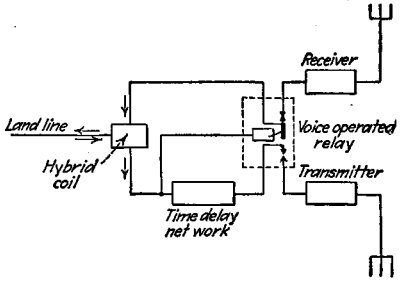


FIG. 3.—Circuit involving voice-operated relays and time-delay network for making possible two-way radiotelephone communication using the same frequency for transmission in both directions.

Secrecy (Privacy) Equipment.—In radio extensions of telephone systems, it is desirable to avoid the possibility of the public listening in on the conversation. This can be accomplished in a variety of ways. One method is to divide the audio-frequency signals to be modulated upon the transmitter into several narrow frequency bands, separating these by filters and then interchanging them (and also sometimes simultaneously inverting) by heterodyne action before modulation upon the transmitter. It is also possible to cause certain of these frequency bands to suffer time delay. By expedients such as these, it is possible to make a system that is practically impossible to decipher, particularly if means are provided for automatically changing the method of rearranging the audio frequencies every few moments, according to a

[1] Charles B. Aiken, A Study of Reception from Synchronized Broadcast Stations, *Proc. I.R.E.*, Vol. 21, p. 1265, September, 1933; K. A. Norton, Note on Synchronized Broadcast Stations WJZ and WBAL, *Proc. I.R.E.*, Vol. 22, p. 1087, September, 1934.

[2] A description of such a transmitter is given by A. A. Oswald, A Short-wave Single-sideband Radiotelephone System, *Proc. I.R.E.*, Vol. 26, p. 1431, December, 1938.

[3] An extensive description of the problems involved in asymmetric broadcasting is given by P. P. Eckersley, Asymmetric-side-band Broadcasting, *Proc. I.R.E.*, Vol. 26, p. 1041, September, 1938. A slightly modified form of asymmetric broadcasting is also described by N. Koomans, Asymmetric-sideband broadcasting, *Proc. I.R.E.*, Vol. 27, p. 687, November, 1939.

prearranged system. All transoceanic extensions of radio-telephone systems are provided with some form of secrecy equipment.

Simultaneous Two-way Telephone Conversation on the Same Frequency.[1]—Radio-telephone channels that are links in wire telephone systems sometimes employ the same frequency for transmission in both directions. This can be accomplished by using a hybrid coil at the land-line terminations to separate the incoming and outgoing channels exactly as in the case of four-wire telephone toll circuits. Voice-controlled relays are provided at each termination to prevent the transmitter and

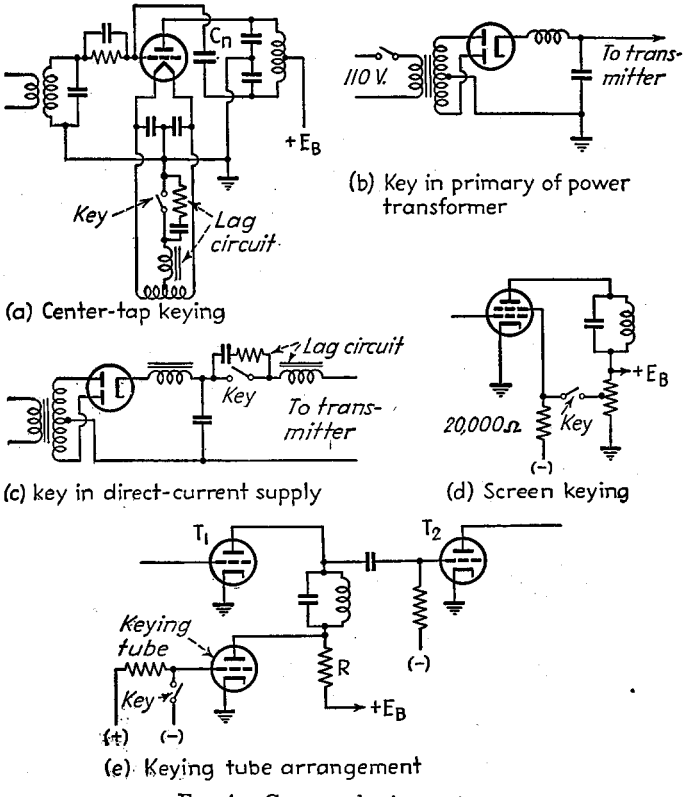


FIG. 4.—Common keying systems.

receiver from being simultaneously connected to the system. Normally, the receiver is connected and the transmitter is nonoperative, but audio signals coming in from the land line will cause a relay operation that disconnects the receiver and places the transmitter in the circuit (see Fig. 3).[2] A time-delay network such as is illustrated in Fig. 3 is desirable so that the relay operation may be completed before the speech actually arrives at the transmitter. A switching system such as is illustrated in Fig. 3 is termed a *vodas*.

[1] S. B. Wright and D. Mitchell, Two-way Radiotelephone Circuits, *Proc. I.R.E.*, Vol. 20, p. 1117, July, 1932; S. B. Wright, The Vodas, *Elec. Eng.*, Vol. 56, p. 1012, August, 1937; Heising, *loc. cit.*

[2] Arrangements of this type are used in most radio links that attempt to provide two-way talking service, even when different transmission frequencies are used in the two directions. In such cases, the voice-controlled relays eliminate the possibility of echoes or singing from unbalances in the hybrid coil terminations.

Voice-controlled relays are sometimes used where the microphone is associated directly with the transmitter, and take the place of the send-receive button that would otherwise be required.

2. Radiotelegraph Transmitters.—Radiotelegraph transmitters differ from telephone transmitters primarily in that in place of the modulator, a keying system is provided to turn the transmitter on and off, and that Class C amplifiers are used throughout, even where linear amplifiers would be used in the corresponding telephone transmitter. In telegraph work it is generally not necessary to have the transmitter particularly free from noise and hum modulation, and in fact self-rectifying circuits that have relatively large hum are frequently used.

Keying Methods.—Keying of a telegraph transmitter is nearly always accomplished at low or moderate power levels so that the power that need be controlled is small. Oscillator keying is occasionally employed in order to achieve break-in operation without the necessity of extremely careful shielding, but in general, keying in one of the subsequent amplifiers is preferable in order to minimize phase and frequency modulation. The keying operation is normally performed with the aid of a relay, although in low-power transmitters the key is sometimes inserted directly in the transmitter circuits.

Typical methods that can be used to control the output of telegraph transmitters are shown in Fig. 4. Probably the commonest method is to place the key or relay in the cathode return lead, as shown in Fig. 4*a*. With the cathode return open, the cathode assumes a potential that is approximately the same as the plate, while the control grid is at ground potential, thereby cutting off the plate current. In such an arrangement, the filament transformer must be well insulated, and since a high voltage is developed across the contacts, a relay is normally used. Another arrangement frequently used in small transmitters is to key in the primary of the power-supply transformer. This arrangement requires that the filtering in the rectifier output be light and that hand keying be employed. Otherwise the filter will round off dots and dashes and cause the code characters to become indistinct. Other keying methods used to some extent include keying in the d-c plate circuit, biasing the suppressor-grid (or screen grid) negative, and using a blocking bias on the control grid sufficient to reduce the output of a Class C amplifier to zero even when normally excited.

The relay in keying systems is sometimes replaced by a tube. An example of one of the many ways in which such a *keying tube* may be employed is shown in Fig. 4*e*. With the contacts closed, the keying tube is biased beyond cutoff, and normal operation takes place. When the key is open, however, the grid of the keying tube is slightly positive, causing this tube to draw a large plate current through the series resistance R and hence to reduce the potential that is applied to the plate of T_1. This reduces the output of T_1 to a value insufficient to drive the grid of the Class C amplifier tube T_2 above cutoff.

Tubes on the high-power side of the keyed stage must be operated so that with the exciting voltage removed the plate current will not be sufficient to make the plate dissipation of the tube excessive. This requires either that the Class C amplifier stages have fixed bias, self-bias, or a combination of these two. Grid-leak bias is seldom permissible.

Unless the oscillator is keyed, the low-power portions of the transmitter operate continuously. It is important that this part of the telegraph transmitter be carefully shielded to prevent radiation during the spacing periods. If such a "back wave" has appreciable intensity, the keyed characters are hard to read.

Key Clicks.—A properly designed radio-telegraph transmitter will produce radiation in the form of clean-cut dots and dashes having rounded edges as shown in Fig. 5*b*.

If the energy rises and falls too rapidly, as in Fig. 5c, the abrupt changes in amplitude introduce high-order side bands. These represent energy at a frequency differing greatly from the carrier frequency, and although weak, they are capable of introducing disturbances in the form of clicks or thumps in neighboring receivers, even when these receivers are tuned to other frequencies. Key clicks can be eliminated by (1) rounding off the code characters by a suitable network or lag circuit associated with the key or (2) arranging so that the keyed signals pass through highly selective radio circuits before radiation takes place. Examples of lag circuits are shown in Figs. 4a and 4c.

Key clicks in near-by receivers are sometimes caused by radio-frequency transients produced in the power circuits by the keying operation. These transients may become modulated upon the transmitter or may be radiated directly by the power lines. They can be readily eliminated by the use of suitable radio-frequency filters associated with the key, and by placing filters in the power lines.

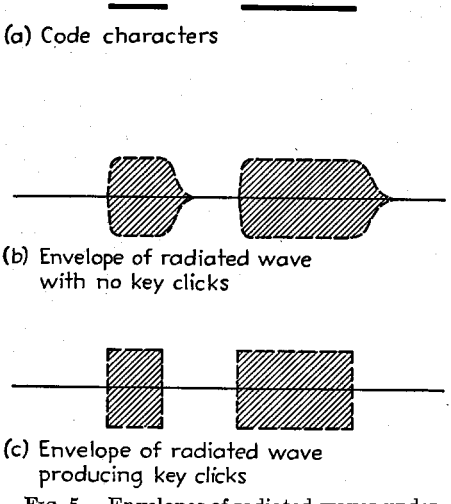


FIG. 5.—Envelopes of radiated waves under conditions favorable and unfavorable for the production of key clicks.

Keying Transients in Power-supply Systems.[1]—Keying causes the load placed on a power-supply system by the keyed stages to change abruptly. This produces low-frequency transients that distort the shape and character of the transmitted pulses, particularly if the keying speed is high.

The fundamental relations that exist in the case of a rectifier-filter system with inductance input are illustrated in Fig. 6. When the key contacts are closed, the keyed stages suddenly draw power, causing the output voltage E to drop momentarily and then oscillate about the ultimate steady-state value at a frequency that corresponds approximately to the resonant frequency of the filter inductance and capacity. When the key contacts are opened, the load on the filter system is suddenly removed, and the voltage across the output rises momentarily above normal until the current through the filter inductance can cease flowing, after which the voltage falls off at a rate corresponding to the discharge of the filter condenser through whatever bleeder resistance or other load is still present.[2]

The keying speed has a great effect upon the shape of the pulses, as is apparent from Fig. 6. With slow keying, the transients occupy only a small part of the total dot length and so affect only the first portion of the code character. However, as the speed of transmission increases, the transients last for a bigger fraction of each pulse, until finally with very high speeds the pulses overlap, and the distortion becomes extremely great.

The important characteristics of the keying transients are (1) the magnitude of the drop in voltage when the key contacts close, (2) the rise in voltage above normal when the key contacts are opened, and (3) the time that elapses from the opening or

[1] Reuben Lee, Radiotelegraph Keying Transients, *Proc. I.R.E.*, Vol. 22, p. 213, February, 1934.

[2] In the case of a generator power-supply system, negative current can flow through the source of power, and an oscillation also occurs when the circuit is broken as well as when the circuit is made.

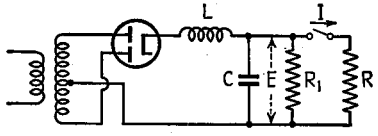
closing of the key to the bottom of the dip, or the peak of the rise, as the case may be. The exact calculations for these properties are quite involved, but can be carried out by methods that have been developed.[1] The following approximate formulas are, however, satisfactory for most practical calculations:

$$\text{Fractional dip in voltage} = \frac{\Delta E_1}{E_0} = \frac{\sqrt{\frac{L}{C}}}{R} \tag{4}$$

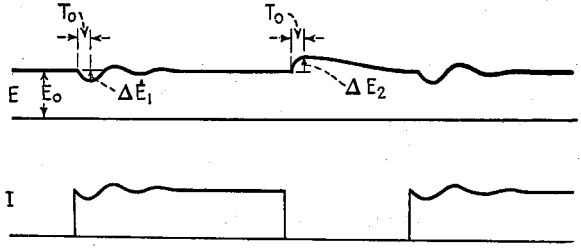
$$\text{Fractional rise in voltage} = \frac{\Delta E_2}{E_0} = \frac{\Delta E_1}{E_0} \tag{5}$$

$$\left.\begin{matrix}\text{Time required to reach} \\ \text{peak or trough}\end{matrix}\right\} = T_0 = \frac{\pi}{2}\sqrt{LC} \text{ seconds} \tag{6}$$

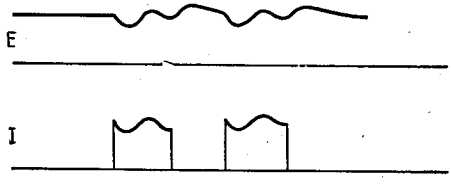
where ΔE_1, ΔE_2, E_0, and T_0 have the meanings in Fig. 6, L and C are the filter inductance and capacity in henrys and farads, respectively, R the resistance load placed



(a) Rectifier circuit



(b) Slow-speed keying



(c) High-speed keying

FIG. 6.—Distortions of code characteristics by keying transients. The cases shown here represent small distortion compared with what is possible in improperly designed systems.

on the power supply by closing the key. These equations assume that the system comprising L, C, and R is oscillatory to transients, a condition realized in all filter systems of practical proportions. These equations also neglect the damping of the transient oscillation during the first quarter cycle. Because of these approximations,

[1] See Lee, *op. cit.*

they lead to calculated values of dip and rise that are larger than the true value, and to calculated values of T_0 that are slightly small. However, these errors are small if the transient oscillation is reasonably underdamped.

Examination of Eq. (4) shows that for a small voltage dip with a given load, the filter should have small inductance and high capacity. At the same time, high-speed keying requires that T_0 be small, corresponding to a small value of LC, *i.e.*, to a poor filter.

Multistage filters in the power supply give rise to much more complicated transients than exist in a single-stage filter. However, if the various sections are reasonably similar, the action is not greatly different from that of a single stage, particularly with respect to the initial dip on closing the key contacts.

Frequency Diversity in Radio-telegraph Transmitters.—Fading of signals at the receiver in short-wave radio-telegraph communication can be greatly reduced by modulating the transmitted wave in some manner to give side-band components. In this way the transmitted energy is divided among several frequencies, thereby reducing the likelihood that the entire signal will fade out. This method of reducing fading is termed *frequency diversity.* The usual means of obtaining frequency diversity is by amplitude modulation of the transmitter output, commonly achieved by the use of a 500-cycle source of plate power for the final power amplifier, with either a full-wave self-rectifying circuit or an ordinary full-wave rectifier with little or no filter.[1] More recently, frequency modulation of the transmitted wave has been used to distribute the transmitted energy between several frequencies.[2] Frequency modulation has the advantage that it gives a higher average output power for a given peak power. A suitable form of frequency modulation consists of a frequency deviation of ± 800 cycles with a modulating frequency of 400 cycles.

3. Miscellaneous. *Suppression of Harmonic Radiation.*—Inasmuch as Class C and linear amplifiers generate harmonics, provision must be made to prevent these harmonic voltages from being radiated. This is especially important in high-power transmitters, since then a harmonic radiation that is only a very small fraction of the desired radiation will still be of sufficient strength to produce interference with other communications up to a reasonable distance.

Harmonic energy may be directly radiated from the circuits of the transmitter or from miscellaneous wires that are coupled to the transmitter circuits. Such direct radiation from the transmitter and associated circuits is usually small, but cannot be neglected in dealing with high powers, particularly with high-frequency transmitters. Direct radiation can be prevented by completely shielding the entire transmitter (normally by the use of a metal cabinet) and by providing radio-frequency filters in power and control leads that pass through the shield.

Harmonic energy may also be delivered directly to the antenna by the output amplifier of the transmitter and then radiated in the normal manner. This action can be kept small by use of the following expedients: (1) a high-Q tank circuit; (2) coupling systems from the final tank circuit to antenna that discriminate against harmonics; (3) radio-frequency filters between final tank circuit and the antenna.

[1] Theoretical and experimental studies indicate that there is an optimum modulation frequency for any given set of conditions. This ranges between about 100 and 750 cycles depending on the circumstances. In some cases, simultaneous modulation by two frequencies that are in the ratio three to one gives greater freedom from fading than can be obtained with a single modulation frequency. With such double modulation, the optimum value of the higher frequency will range from 100 to 750 cycles depending on the distance, the ionosphere conditions and the frequency of the transmitted carrier. For further information see A. L. Green and Geoffrey Builder, Control of Wireless Signal Variations, *Jour. I.E.E.*, Vol. 80, p. 610, 1937 (also Wireless Section, *I.E.E.*, Vol. 12, p. 104, June, 1937); A. L. Green and O. O. Pulley, Control of Phase-fading in Long Distance Radio Communication, *Jour. I.E.E.*, Vol. 80, p. 623 (also Wireless Section, *I.E.E.*, Vol. 12, p. 117, June, 1937).

[2] For example, see C. W. Hansell and G. L. Usselman, A 200-kilowatt Radiotelegraph Transmitter, *R.C.A. Rev.*, Vol. 2, p. 442, April, 1938.

The ratio of harmonic to fundamental frequency voltage developed across the tank circuit is given by Eq. (88*a*), Sec. 5, which shows that this ratio is inversely proportional to the tank circuit Q. However, the maximum feasible Q is limited by considerations of coil efficiency and phase shift, and except in very small transmitters, the use of the highest practicable Q will not reduce the harmonics sufficiently. The next step in harmonic reduction is to select a suitable method of coupling to the tank circuit. Thus, although the coupling method shown in Fig. 7*a* gives approximately the same

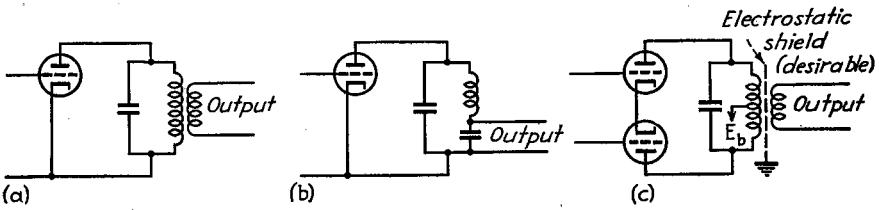


FIG. 7.—Coupling systems that give varying degrees of discrimination against harmonics.

percentage of harmonic voltage in the output as exists in the voltage across the tank circuit, the coupling circuit of Fig. 7*b* reduces the harmonic percentage by the factor $1/n^2$, where n is the order of the harmonic. Also, perfectly symmetrical coupling to a balanced push-pull amplifier, as shown in Fig. 7*c*, will transfer only odd harmonics to the output, since the push-pull symmetry prevents even-harmonic voltages from appearing across the tank circuit.[1]

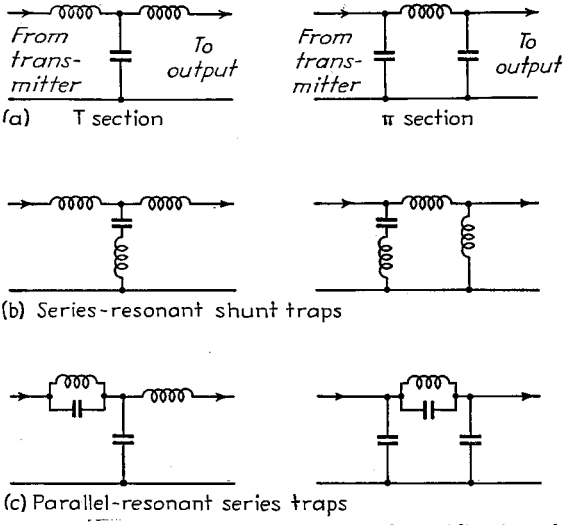


FIG. 8.—Filter networks for suppressing harmonics and providing impedance matching.

Additional reduction of harmonics beyond that obtained by the use of high-Q tank circuits and properly chosen coupling systems can be obtained by placing one or more low-pass filters between the tank-circuit output and the antenna. These filters are normally designed to serve the dual purpose of harmonic suppression and

[1] Even harmonic voltages do exist, however, between ground and the two extreme ends of the tank circuit. In order to prevent these voltages from transferring energy to the transmission line by electrostatic coupling, in which the two wires of the transmission line act in parallel as one side of a circuit having a ground return, it is necessary to place an electrostatic shield between the tank circuit and the coupling coil, or in some other way minimize capacitive coupling.

impedance matching. Examples of suitable filters are shown in Fig. 8. The circuits at *a* are simple *T* and π low-pass sections, and may be either symmetrical or non-symmetrical, according to the impedance matching requirements. These simple circuits can be modified by using series resonant shunt arms, or parallel resonant series arms, as shown at *b* and *c*, respectively. Each such resonant branch is intended to discriminate against a particular harmonic. The design of these coupling networks is discussed in Par. 25, Sec. 3.

An ingenious modified form of filter that is both simple and effective is shown in Fig. 9.[1] Here the transmission line, or output circuits of the transmitter, are shunted by the input of two auxiliary transmission lines. One of these is a quarter of a wave length long and short-circuited at its receiving end, while the second is one-twelfth wave length long, open-circuited at its receiving end, and having its input shunted by a coil that tunes the input of this line to parallel resonance at the fundamental frequency of the transmitter. The quarter wave-length line acts as a short circuit to all even harmonics, and the second line acts as a short circuit to the third harmonic. Neither line has appreciable effect upon the fundamental, because to this frequency the input impedance of both lines is an extremely high resistance. This method of

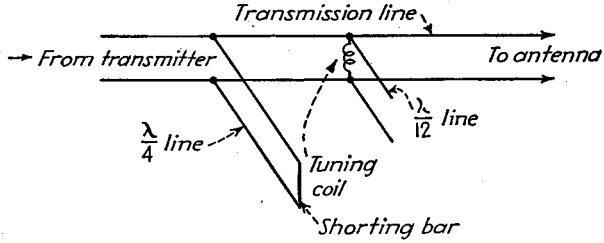


FIG. 9.—Suppression of harmonics by resonant transmission lines.

eliminating harmonics is particularly suitable for transmitters operating at a fixed frequency, such as broadcast transmitters.

Adjacent Channel Interference from High-order Side Bands.[2]—The output of a radio-telephone transmitter will interfere with communications in adjacent channels if the output contains high-order side bands. High-order components can result from overmodulation of the transmitter or from distortion in the audio-modulating system that causes harmonics of high audio frequencies to be modulated on the carrier. The former is the most serious in practice, and is handled by the use of peak limiters, automatic-volume-control systems, and alarms that indicate to the operator every time the transmitter modulation exceeds 100 per cent.

Phase and frequency modulation is a troublesome cause of high-order side bands in short-wave transmitters. This is discussed further below.

Incidental Phase and Frequency Modulation.—Phase or frequency modulation is often present in the output of a telephone or telegraph transmitter as a by-product produced by the modulation or keying actions.[3] Such incidental phase and frequency modulation is to be avoided, since it frequently gives rise to high-order side bands that will cause adjacent channel interference. Phase modulation is usually more troublesome than frequency modulation, for the modulation index in the former

[1] W. H. Doherty and O. W. Towner, A 50-kilowatt Broadcast Station Utilizing the Doherty Amplifier and Designed for Expansion to 500 Kilowatts, *Proc. I.R.E.*, Vol. 27, p. 531, September, 1939.

[2] I. J. Kaar, Some Notes on Adjacent Channel Interference, *Proc. I.R.E.*, Vol. 22, p. 295, March, 1934. Adjacent channel interference produced by frequency modulation signals is discussed by L. J. Black and H. J. Scott, Modulation Limits in FM, *Electronics*, Vol. 13, p. 30, September, 1940.

[3] W. A. Fitch, Phase Shift in Radio Transmitters, *Proc. I.R.E.*, Vol. 20, p. 683, May 1932.

case is independent of frequency but in the latter instance varies inversely with frequency.

It is also possible for noise and hum effects to produce incidental phase modulation, which, although not disturbing the modulation envelope at the transmitter, will cause noise and hum to be present in the modulation of the signal when received at a distant point under conditions where the ionosphere disturbs the phase relations between carrier and side bands.

Modulation or keying of a transmitter can cause phase modulation in one of several ways. In telegraph transmitters, the sudden change in output of the keyed stages may react back upon the unkeyed portion of the transmitter to cause a sudden change in phase or "phase whip" that represents an appreciable amount of energy at very high-order side-band frequencies.[1]

In radio-telephone transmitters, phase modulation will result if some of the modulated wave leaks back into an unmodulated buffer amplifier or oscillator section of the transmitter. Inasmuch as limiting action is always present, such mixing of the modulated and unmodulated waves generates a phase-modulated wave as discussed in Par. 17, Sec. 7. The mixing of modulated and unmodulated energy required to produce such phase modulation can occur as a result of imperfect shielding in the transmitter or because of an incompletely neutralized modulated amplifier.

Another cause of phase modulation in radio-telephone transmitters is failure to tune the tank circuit of the modulated amplifier accurately to the carrier frequency. Incorrectly tuned tank circuits of linear amplifiers can also cause phase modulation, but to a much smaller degree.

Tuning and Wave Changing in Radio Transmitters.—In short-wave transmitters, it is nearly always necessary to be able to operate on two or more frequencies, and in some cases operation over continuous bands of frequencies of considerable width is necessary. In broadcast and long-wave transmitters, it is usually necessary to tune the tank circuits over at least a limited frequency range.

Tuning in low-level stages is ordinarily accomplished with the aid of variable condensers and coil switching, following practices similar to those employed in all-wave receivers. In high-level stages, particularly in stages where the power is so great that water-cooled tubes are employed, tuning and wave changing are more complicated. Adjustment over a limited frequency range can be made by a copper disk or ring (doughnut) rotated in the magnetic field of the coil. Step-by-step changes of somewhat greater magnitude can be obtained by short-circuiting turns of the tuning inductance and by varying the tuning capacity in fixed steps. Large wave-length changes necessitate coil switching. Condensers used in high-power stages usually have air dielectric, and compressed gas types have advantages of compactness combined with continuous adjustment.

In transmitters adjustable over a large continuous frequency band, trouble is often experienced from parasitic circuits that are coupled to the various tank circuits and so absorb power at the resonant frequency of the parasitic circuit.[2] The result of such power absorption is low output from the particular tank circuit in question, and in some cases the production of unbalance in push-pull stages. Whenever the output of a stage is low over a small portion of its tuning range, parasitic circuits should be looked for. Common sources of parasitic circuits are unused tuning coils, circuits involving by-pass condensers and their associated leads, etc.

Monitoring and Protection of Radio Transmitters.—Provision is always made in all except very small radio transmitters for monitoring the transmitter behavior and

[1] An analysis of the frequency spectrum that results when phase or frequency modulation is produced by keying is given by Balth. van der Pol, Frequency Modulation, *Proc. I.R.E.*, Vol. 18, p. 1194, July, 1930.

[2] P. A. Ekstrand, Parasitic Circuits, *Electronics*, Vol. 11, p. 26, October, 1938.

protecting against faults. Monitoring equipment normally includes meters to read the plate and grid currents of the various stages and the transmitter output. Frequency monitors are also employed in all broadcast transmitters and large short-wave transmitters. In radio-telephone installations, a meter indicating the audio-frequency level is essential, and some device for showing when the transmitter is overmodulated is desirable. In telegraph transmitters, a provision is commonly made by which the transmitted signals can be listened to, and with high-speed transmission a monitoring tape recorder is almost essential.

All large transmitters are provided with rather elaborate protective equipment. This is to protect against failure of the normal control circuits to operate properly, to ensure safety to life, and to take care of such failures in operation as short circuits, overloading, flashback in the rectifier tubes, failure of cooling water circulating system, etc.

RECEIVERS FOR AMPLITUDE-MODULATED WAVES

4. Broadcast and Short-wave Superheterodyne Receivers for Amplitude-modulated Waves.—The usual radio-telephone receiver is of the superheterodyne type. Such a receiver can be represented schematically as shown in Fig. 10, and consists

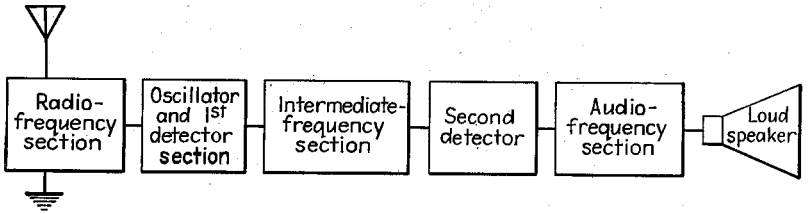


FIG. 10.—Schematic diagram of superheterodyne receiver.

of a radio-frequency section, converter (first detector) and oscillator, intermediate-frequency amplifier, second detector, audio-frequency amplifier, and loud-speaker.[1]

Radio-frequency Section.—The radio-frequency section includes the coupling from the antenna to the grid of the first tube, and also any amplifier stages that amplify the incoming signal before its frequency is changed. One of the chief purposes of the radio-frequency section is to provide selectivity against image and intermediate-frequency signals.

The antenna-to-grid coupling should be arranged so that the transfer constant is relatively uniform over the entire frequency band and is not too small. It must also permit the tuned input circuit to track properly with other adjustable circuits in the receiver. With home receivers operating in the band 550 to 1,500 kc and employing an ordinary antenna, it is customary to use a complex coupling system involving a high-inductance primary resonated at a frequency just below the band to be received. Suitable arrangements of this type are shown in Figs. 29 to 34 of Sec. 3. These arrangements can be made to give relatively uniform antenna to grid voltage step-up over the desired band, and have the merit that the maximum coupling outside the band is at a frequency where there are usually few interfering signals present. The effective coupling between antenna and grid is usually made quite small. Although this reduces the voltage step-up between antenna and first grid, it allows wide latitude in the electrical constants of the antenna without preventing tracking of the first tuned

[1] A summary of characteristics of broadcast receivers now in use, together with discussions of the significance of these characteristics under practical conditions, is given by A. Van Dyck and D. E. Foster, Characteristics of American Broadcast Receivers as Related to the Power and Frequency of Transmitters, *Proc. I.R.E.*, Vol. 25, p. 387, April, 1937; D. E. Foster, Receiver Characteristics of Special Significance to Broadcasters. *Communications*. Vol. 19, p. 9, May, 1939.

circuit with the remaining tuned circuits, and also prevents the voltage that reaches the grid of the first tube in the case of strong local signals to exceed the value that the automatic-volume-control system is able to handle without distortion.

Typical coupling methods used with loop antennas are shown in Figs. 11*a* to 11*c*. Here the situation is complicated by the fact that the receiver must usually also be capable of operating with an outdoor antenna. When a loop is employed, it is permissible to use closer coupling between the antenna and the tuned input circuit than would be possible with outdoor antennas, since the loop constants can be fixed by the manufacturer.

In short-wave bands, the usual antenna coupling system is a simple mutual inductance as shown in Fig. 11*d*, commonly proportioned to operate from a two-wire transmission line of about 100 ohms characteristic impedance.

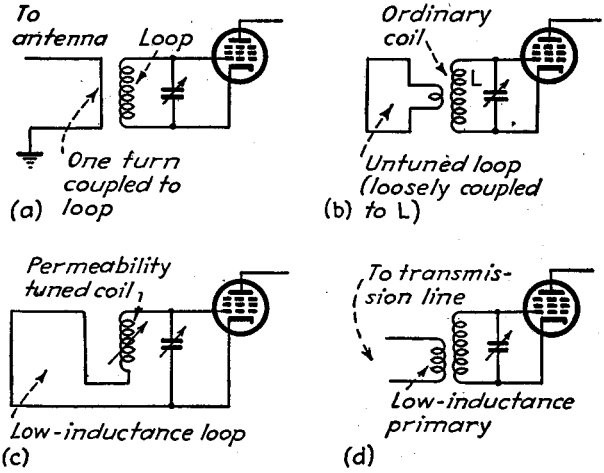


FIG. 11.—Input circuits for broadcast receivers operating with loop antennas and in the short-wave bands.

The use of one or more stages of tuned radio-frequency amplification in the r-f section increases the selectivity between the antenna and the input to the converter. This reduces image and intermediate-frequency responses and cross-modulation. It also raises the signal level at the converter input and thereby tends to minimize the undesirable consequences of converter noise. At the same time, tuned radio-frequency amplification has the disadvantage that the same tube placed in the intermediate-frequency system will give more amplification, more adjacent channel selectivity, and will lower manufacturing cost. Tuned radio-frequency amplification is accordingly employed only in the more expensive receivers, or where special requirements must be satisfied.

The resonant circuits of the tuned radio-frequency section are normally adjusted by means of a variable air-dielectric condenser. Permeability tuning is used to some extent, however, particularly in automobile radios, and in receivers employing push-button control.

Converter—Oscillator Section.—The converter-oscillator section (sometimes termed the first detector) has for its purpose the conversion by heterodyne action of the incoming signal to the intermediate frequency. The various converter systems employed in practice are discussed in Par. 12, Sec. 7, with arrangements in which the oscillator and detector are combined in a single tube being customary. However,

such converters have a high noise level that tends to dominate the signal-to-noise ratio unless a stage of tuned radio-frequency amplification is employed. The triode plate-rectifier type of converter with separate oscillator is sometimes used because of its lower noise level.

Intermediate-frequency Section.—The intermediate-frequency section consists of the coupling networks and amplifier tubes between the plate of the converter and the input terminals of the diode detector. The intermediate-frequency section usually includes one or two stages of amplification, and provides most of the gain existing between the antenna and the detector. This amplification per stage is high, because, since no provision need be made for tuning over a wide frequency range, a high L/C ratio can be employed in the coupling circuits. In addition to providing a large part of the total amplification, the intermediate-frequency section also provides most of the adjacent channel selectivity possessed by the receiver.

Coupling circuits in the intermediate-frequency amplifier stages are usually band-pass circuits consisting of primary and secondary separately tuned to the desired frequency and coupled together with slightly more than critical coupling.[1] The final coupling network delivering voltage to the diode detector must be especially designed to take into account the input impedance of the diode. The resonant frequencies of the intermediate-frequency circuits are adjusted to the desired value either by trimmer condensers or by magnetic slugs composed of powdered iron.

The choice of intermediate frequency is a compromise between the fact that a high frequency provides a better image discrimination and gives fewer intermediate-frequency harmonics lying in the broadcast band, but at the same time gives less discrimination against signals of intermediate frequency. Present practice is to use a value approximately 455 kc for all broadcast receivers. This value is higher than is really necessary for the standard broadcast band of 550 to 1,500 kc, but is none too great for the short-wave bands. In receivers designed exclusively for short waves, intermediate frequencies of the order of 2 mc are sometimes employed to get increased image discrimination, although this sacrifices adjacent channel selectivity.[2]

Detector.—The final or second detector of a superheterodyne is almost universally a diode, which is normally arranged to provide a voltage for automatic-volume-control purposes along with the modulation-frequency output. Such a detector requires a minimum of several volts to operate it properly, and the radio-frequency, converter, and intermediate-frequency sections should be designed to provide this output with the weakest signal the receiver is intended to operate on satisfactorily.

Audio-frequency Section.—The audio-frequency section commonly consists of a stage of voltage amplification, followed by a stage of power amplification, and includes the manual volume control and tone control. The power stage may employ a single pentode tube, or push-pull triodes or pentodes. A limited amount of negative feedback is sometimes employed in the power stage, particularly if pentodes are used, since feedback is useful in lowering the equivalent internal resistance that the pentode tube presents to the loud-speaker.

Power Sources.—The sources of power used for operating receivers are 110 volts alternating current, 110 volts direct current, 6-volt storage battery, and dry cells. Some types of receivers can be operated from several sources, as, for example, the a-c–d-c sets, which will function on all 110-volt lighting circuits, and some of the dry-

[1] When special requirements are to be met, more complicated coupling systems based on wave-filter theory can be used. An example of such an arrangement is given by J. B. Moore and H. A. Moore, I-F Selectivity in Receivers for Commercial Radio Services, *R.C.A. Rev.*, Vol. 4, p. 319, January, 1940.

[2] A discussion of factors determining the choice of a suitable intermediate frequency is given by F. E. Spaulding, Jr., Design of Superheterodyne, Intermediate-frequency Circuits, *R.C.A. Rev.*, Vol. 4, p. 485, April, 1940.

battery operated sets, which are arranged so that they can be plugged into a 110-volt line.

Sets operating from 110-volt a-c power lines employ a-c filament current, with heater-type tubes being used throughout except for the rectifier and possibly the power amplifier. The plate-supply system in such cases is a rectifier-filter arrangement employing a condenser-input filter, with the speaker field winding often used as the first filter choke. The power tubes in such receivers frequently receive their plate voltage directly from the input condenser, particularly when push-pull amplification is employed. This reduces the current that must be carried by the first choke and so lowers its cost.

The a-c–d-c sets omit the power transformer, but normally leave the rectifier tube in the circuit so that the d-c line must be plugged in with the proper polarity if plate current is to flow. Half-wave rectification is usually employed, although in some of the more expensive sets a voltage-doubling circuit is used with a-c operation. The cathodes of tubes in a-c–d-c sets are always of the heater type, and employ heaters arranged in a series parallel combination across the 110-volt line. It will be noted that the chassis of an a-c–d-c set is likely to be at the full line voltage above ground, and so must be properly protected to prevent it from becoming a hazard to life.

Auto receivers use the 6-volt car storage battery to operate the filaments of tubes, and employ a vibrator or dynamotor for supplying plate power.

Completely self-contained sets operating from dry batteries normally use separate plate and filament batteries, and obtain the necessary grid biases by self-bias arrangements. Such sets are often arranged with a rectifier tube, filter, etc., so that they can be operated from 110-volt sources by throwing a switch. The filament power with such 110-volt operation is commonly rectified d-c current.

5. Miscellaneous Features and Considerations Relating to Radio Receivers. *Automatic Volume Control.*—Practically all broadcast and similar receivers are provided with an automatic-volume-control system (abbreviated A.V.C.) to maintain the carrier voltage at the detector approximately constant. This is accomplished by biasing the grids of the radio-frequency, intermediate-frequency, and converter tubes negatively with a d-c voltage derived by rectifying the carrier. An increase in the signal hence increases the negative bias and tends to counteract the increased signal by reducing the amplification, and vice versa. In this way, the receiver may be tuned from strong to weak stations without the necessity of resetting the manual volume control, and also variations in signal strength due to fading are smoothed out.

Automatic-volume-control action is normally obtained by deriving from the usual diode detector a d-c voltage proportional to the amplitude of the carrier at the diode input terminals, and free of modulation.[1] A typical circuit arrangement is shown in Fig. 13*a*, and consists of an ordinary diode detector to which there has been added provision for filtering out the modulation-frequency currents from the d-c rectified voltage. The time constants of the filter circuits should be great enough so that the lowest modulation frequencies do not reach the A.V.C. output, but at the same time be small enough so that the rectified bias will follow moderately rapid changes in carrier amplitude. Time constants of the order of $\frac{1}{5}$ to $\frac{1}{10}$ second are customary.

An ideal automatic-volume-control system would not affect the receiver amplification until the incoming signal reached a level sufficient to produce an adequate voltage at the input of the second detector, and then with larger signals would maintain this output constant, as shown by curve *a* of Fig. 12. In contrast with this, the simple A.V.C. system of Fig. 13*a* has a characteristic such as given by *b*, in which

[1] Such systems were first devised by H. A. Wheeler. See his paper, Automatic Volume Control for Radio Receiving Sets, *Proc. I.R.E.*, Vol. 16, p. 30, January, 1928.

the receiver gain starts to fall off as soon as the diode begins to develop output. The actual control characteristics can be made to approach the ideal more closely by arranging so that the A.V.C. voltage controls the gain of a considerable number of tubes, and then employing delay action so that the A.V.C. system does not function until the signal reaches a predetermined level (curves *c* and *d* of Fig. 12). Amplification of the A.V.C. voltage, as by the use of a separate intermediate-frequency amplifier for operating the A.V.C. system, also permits still further improvement in the control characteristics (curve *e* of Fig. 12).

Delay action can be obtained by the use of a second diode arranged as shown in Fig. 13*b*. With no signal the conductivity of this second diode causes the A.V.C. bus to assume the potential of the cathode of this diode, *i.e.*, a moderate negative bias. This bias is maintained on the A.V.C. system until the rectified signal develops a direct-current voltage greater than the second cathode voltage to ground. The second

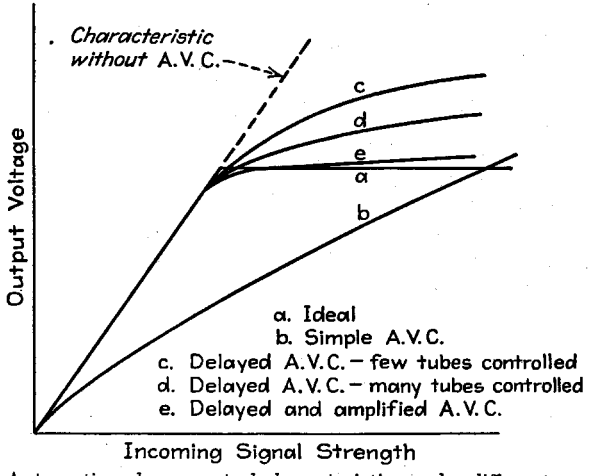


FIG. 12.—Automatic-volume-control characteristics under different conditions.

diode then becomes nonconducting, and the automatic-volume-control system functions in the usual manner. Delay can also be obtained by using a separate A.V.C. diode that is biased, and so does not develop output until the signal exceeds the bias. This arrangement, however, has the disadvantage that the delay and hence the receiver gain vary with changes in the average degree of modulation.

Amplification of the A.V.C. voltage used to control the receiver gain, combined with delay, will make the characteristics approach much more closely the ideal. Amplification can be obtained by the use of an A.V.C. diode operating from a separate intermediate-frequency amplifier stage that supplies more signal to the control diode than reaches the detector diode, as shown in Fig. 13*c*, or by amplifying the d-c control voltage developed by the diode, as shown in Fig. 13*d*. Here the rectified output of the diode D_1 is applied between grid and cathode of amplifier tube T_1. This tube has its cathode returned through a large resistance to a point considerably negative with respect to ground. The initial adjustment is such that the cathode of T_1 is at a slightly positive potential with respect to ground by an amount determined by the delay action desired. Rectified output produced by an incoming signal biases the grid of T_1 negative by an amount proportional to the amplitude of the incoming signal. This reduces the plate current of T_1, and causes the cathode voltage to decrease by an amount corresponding to the amplification that the output of diode

D_1 undergoes in tube T_1 with coupling resistance R. As long as the cathode of T_1 is positive with respect to ground, the A.V.C. bus remains at ground potential, but as soon as the cathode of T_1 becomes negative with respect to ground, the conductivity between cathode and diode plate D_2 causes the A.V.C. bus to assume the potential of the cathode. This combines amplified control with a delay action determined by the positive initial bias voltage of the cathode of T_1.

Operation of automatic-volume-control systems from a separate branch of the intermediate-frequency amplifier has advantages beyond giving a means for obtaining amplified A.V.C. Thus the selectivity of the A.V.C. branch of the intermediate-frequency system can be made less than for the detector branch, which facilitates tuning the receiver properly without a tuning indicator. Also, by applying A.V.C. voltage to an intermediate-frequency stage subsequent to the dividing point in the

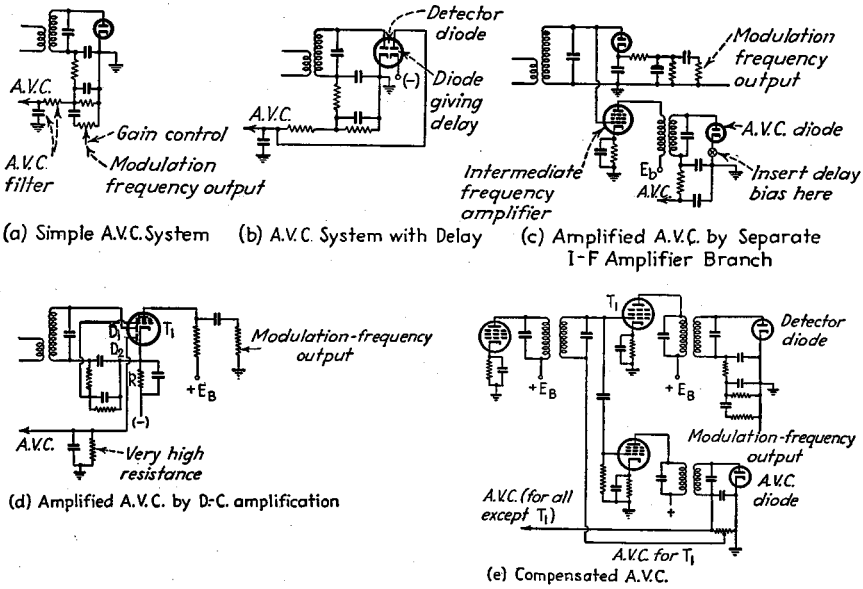


FIG. 13.—Automatic-volume-control systems using diode tubes.

intermediate-frequency system such as T_1, as illustrated in Fig. 13*e*, it is possible to compensate for the usual rising characteristic of an A.V.C. system (see Fig. 12) and maintain the voltage at the detector input terminals substantially independent of signal amplitude.

While diode detectors are nearly always employed in A.V.C. systems, it is possible to use plate detection. A typical example is illustrated in Fig. 14,[1] where the A.V.C. voltage is developed by the output of direct-current amplifier T_2 that is operated from the d-c voltage built up across the bias resistance by the rectified plate current of the detector tube T_1. For best performance, T_2 should be operated so that with no signal it is biased beyond cutoff by the amount of delay that is desired. As compared with diode arrangements, plate-detection systems have all the limitations of plate detectors, combined with the fact that unless rather unusual cathode voltages are employed an extra tube is necessary in order to obtain a control voltage of the proper polarity.

[1] For additional examples, see Dorman D. Israel, Sensitivity Controls—Manual and Automatic, *Proc. I.R.E.*, Vol. 20, p. 461, March, 1932.

The automatic-volume-control voltage is normally applied to all the intermediate-frequency amplifier tubes, to the converter, and to all radio-frequency stages. The more tubes that are controlled the less will the output of the receiver vary with the strength of the incoming signal. This is illustrated in Fig. 12 and is one of the reasons that receivers with many tubes perform better than do less elaborate receivers. The tubes that are controlled by the A.V.C. system are always of the variable-mu type in order to reduce the possibility of cross-modulation. In some cases, there is an advantage, from the point of view of low cross modulation and distortion, of applying

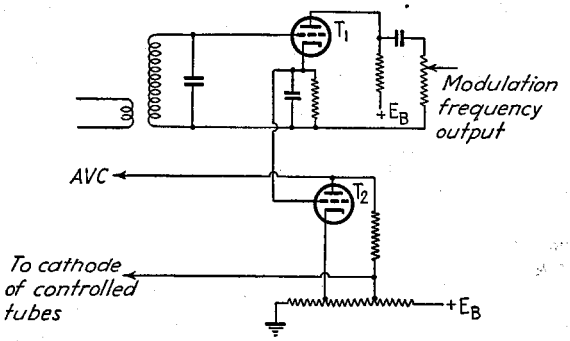


FIG. 14.—Automatic-volume-control system using plate detector.

less control voltage to the tube operating at the highest power level than to the remaining tubes that are controlled by the A.V.C. system.

Manual Volume Control.—All receivers are provided with a manual volume control for the purpose of enabling the user to adjust the level of the reproduced sound. In receivers having automatic volume control, the manual control practically always consists of a potentiometer in the grid circuit of the first or second audio-frequency amplifier stage following the detector. In the rather unusual case where the receiver is not provided with automatic volume control, the manual control operates by varying the grid bias or the screen-grid voltage of variable-mu tubes of the radio-frequency, converter, and intermediate-frequencies sections, combined in some cases with an antenna input control.[1]

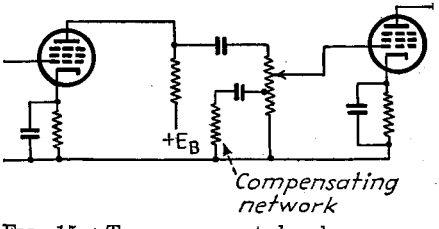


FIG. 15.—Tone-compensated volume control.

The sound output of a radio receiver will appear more natural if the volume of the low frequencies and, to a lesser extent, of the high frequencies is reduced more slowly than is the volume of the middle range of frequencies. This is a result of the fact that as the level of the reproduced sound is reduced, the relative sensitivity of the ear becomes less for the low and high frequencies. The amount of compensation required to counteract this effect is quite large at the low frequencies, being of the order of 20 db under typical conditions. This compensation can be automatically introduced by the use of a *tone-compensated volume control*, such as is illustrated in Fig. 15.

Tone Control.—Most radio receivers are provided with means, termed a *tone control*, by which the listener may reduce the relative amplitude of the higher frequencies in the reproduced sound. This is accomplished by a suitable audio-frequency network,

[1] See Israel, *loc. cit.*, for further details of arrangements of this type.

the most common arrangement consisting of the combination of a capacity and variable resistance shunted across an audio-frequency coupling impedance, as shown in Fig. 16.

Push-button Tuning.—There are three basic ways in which push-button tuning is carried out, namely, (1) systems in which operating the push button disconnects the ordinary tuning system and replaces it with pretuned circuit elements, (2) arrangements in which the pressure on the push button rotates the tuning shaft to the desired position, and (3) systems in which the push button controls an electric motor that drives the tuning shaft.[1]

The first of these methods is the one most commonly employed, and may involve either replacing the main tuning condenser sections with preset trimmers, or replacing each entire tuned circuit with resonant circuits adjusted to the desired frequency by means of permeability tuning.

The second method is the cheapest of the three, but requires an undesirably large pressure on the buttons, and so is used only in the least expensive sets. The third method is particularly suitable where remote control is desired, and is the most

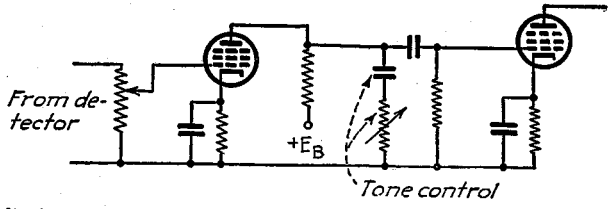


FIG. 16.—Audio-frequency system of a radio receiver, showing tone control and manual volume control.

expensive of the three.[2] In the second and third types of push-button systems, where operation involves rotating the control shaft of the tuning system, contacts are always provided so that the receiver output is short-circuited while the push button is being operated. Otherwise the receiver produces disagreeable sounds as the tuning passes through various stations on the way to its ultimate position.

In push-button tuning, it is very important that the resonant circuits involved, particularly the oscillator circuit, be very stable. Appreciable variation in resonant frequency, as a result of either aging or temperature variations, will result in mistuning of the signals when push-button operation is employed.

Tuning Methods—Range Change—Band Spread.—The radio-frequency and oscillator circuits of a receiver are normally tuned by a gang variable condenser. Permeability tuning, *i.e.*, varying the inductance of a coil by moving an iron dust slug, is used to a considerable extent in the antenna circuits of automobile radios, and there are a few sets on the market in which all the tuning is accomplished by varying the permeability of the coil. Permeability tuning has the advantage of compactness, freedom from vibration, and stability of the resonant frequency, but at the present time is more expensive.

[1] A system in which the drive is obtained from a solenoid instead of a motor is described by J. H. Little and F. X. Rettenmeyer, A Five-band Receiver for Automobile Service, *Proc. I.R.E.*, Vol. 29, p. 151, April, 1941.

[2] Remote control is usually carried out with a long cord. Control methods using induction fields and a low-frequency carrier transmitted over the power line have been proposed. See Robert G. Herzog, Mystery Control, *Communications*, Vol. 18, p. 20, October, 1938; S. W. Seeley, H. B. Deal, and C. N. Kimball, Teledynamic Control by Selective Ionization with Application to Radio Receivers, *Proc. I.R.E.*, Vol. 26, p. 813, July, 1938; Charles N. Kimball, A New System of Remote Control, *R.C.A. Rev.*, Vol. 2, p. 303, January, 1938.

In all-wave receivers, the tuning range is normally changed by switching coils in and out of the circuit, although in some two-range receivers a tapped coil is used. Coils for each tuning range are provided with their own trimmers.

In short-wave reception, there is an advantage in being able to spread out the tuning range, so that each of the comparatively narrow short-wave broadcast bands covers the full dial length. Such *band-spread* tuning arrangements have two forms: (1) those in which the main tuning control is set to a particular position and the band spread then obtained by an auxiliary variable vernier trimming condenser; and (2) systems in which auxiliary reactances are switched into the circuit so that the full variation of the normal tuning control will produce only a small percentage change in frequency. The former method is used in communication receivers, and the latter is the method commonly employed in broadcast receivers. The second method can be carried out as illustrated in Fig. 17, with the use of auxiliary fixed condensers in the case of capacity tuning and auxiliary fixed inductances in the case of permeability tuning.[1]

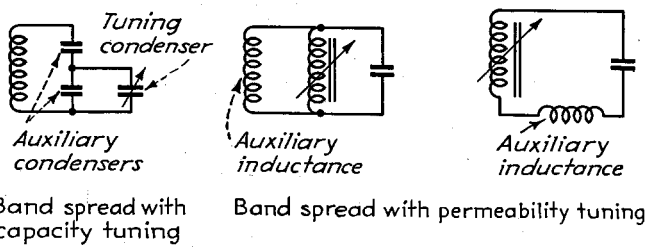


FIG. 17.—Circuit arrangements for obtaining band spread with the regular tuning adjustment.

The chief problem involved in band-spread tuning is maintenance of the frequency calibration. The frequency range is usually so small that even a small percentage change in the calibration from temperature effects or aging can readily shift the desired band entirely outside the tuning range of the receiver.

Stability of Frequency Calibration—Crystal Controlled Receivers.—Stability of the receiver frequency calibration can be obtained by careful attention to the mechanical and electrical design of the resonant circuits. In particular, dielectrics associated with the circuits are potential sources of trouble, and fixed condensers that are involved not only must be of the type having cyclic behavior and a low temperature coefficient but also should be chosen so that their temperature coefficient is the negative of that possessed by the combination of tuning coil and tube.[2]

The oscillator circuit is the most important from the point of view of stability, since as long as the radio-frequency circuits are not too sharp, they can vary appreciably without giving trouble. Where communication is carried on with the use of only a few fixed frequencies, as in point-to-point services, the superheterodyne oscillator is frequently a crystal oscillator, with a crystal being provided in the set for each of the fixed frequencies.

[1] For further discussion of band spreading see F. H. Woodbridge, Band Spread—Its Effect on the "Tuning Rate," *Wireless Eng.*, Vol. 17, p. 394, September, 1940; Dudley E. Foster and Garrard Mountjoy, Short Wave Spread Bands in Automobile and Home Receivers, *Proc. I.R.E.*, Vol. 30, p. 222, May, 1942.

[2] John M. Miller, Thermal Drift in Superheterodyne Receivers, *Electronics*, Vol. 10, p. 24, November, 1937.

Special temperature compensation to take care of the warming-up period is also desirable. This can be obtained by using in the oscillator circuit a small fixed condenser with a negative temperature coefficient and associated with a small heating resistor that is turned on and off with the receiver. See M. L. Levy, Frequency Compensation, *Electronics*, Vol. 12, p. 15, May, 1939.

Spurious Responses in Superheterodyne Receivers.—In superheterodyne receivers, there is the possibility of obtaining a variety of spurious responses. Some of these appear in the form of whistles having a pitch dependent upon the tuning of the receiver; others cause an interfering program to be heard at unexpected places on the dial. The principal sources of spurious responses are (1) image-frequency signals; (2) signals of intermediate frequency; (3) harmonics of the intermediate-frequency generated by the second detector; and (4) harmonics of the incoming signal generated in the converter tube.

Image signals are signals having a frequency that is greater than the frequency to which the receiver is tuned by twice the intermediate frequency (assuming the beating oscillator operates at a higher frequency than the desired signal). If such signals reach the converter tube, they combine with the local oscillator to produce a difference frequency that is exactly equal to the intermediate frequency, and so appear in the receiver output along with the desired program. Image responses can be eliminated by providing sufficient selectivity between the antenna and the converter tube to prevent a signal of image frequency from reaching the converter tube with appreciable amplitude. For satisfactory elimination of image response, it

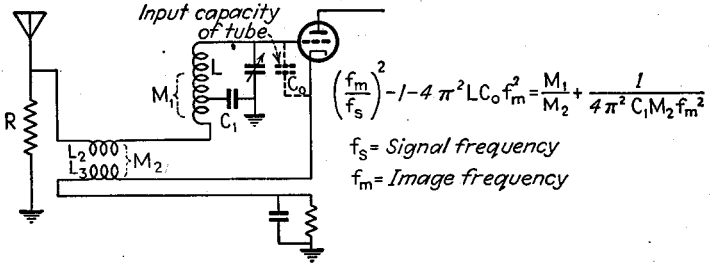


FIG. 18.—Typical circuit for suppressing image signals, together with equations for maximum suppression.

is necessary that the image voltage reaching the converter input be at least 30 db below the desired signal. Since the incoming image signal may be somewhat stronger than the desired signal, the circuits between the antenna and the converter input should give a discrimination exceeding 50 to 60 db against the image signal relative to the signal to which the receiver is tuned. The discrimination obtainable against the image signal is greater the higher the ratio of intermediate to signal frequency. For this reason, the intermediate frequency in all-wave broadcast receivers is made as high as possible (*i.e.*, about 455 kc). Intermediate frequencies of the order of 2 mc are often used in receivers intended for the reception of only short-wave signals. The discrimination against image signals is increased by the use of one or more stages of tuned radio-frequency amplification, since this provides additional selectivity between the antenna and the converter input. Discrimination against image signals can also be improved greatly by the use of special antenna coupling circuits. An example of such an arrangement is shown in Fig. 18.[1] Here the antenna is brought to ground through a coupling coil L_2, a portion of the tuning inductance L, and a capacity C_1 of the order of ten times the maximum tuning capacity. The resistance R is for broadening the antenna resonance. The effect of the image voltage produced at the grid of the tube is largely neutralized by the voltage that the antenna current induces in the cathode coil L_3. With proper circuit proportions, the suppression

[1] Further discussion of this and other circuits, including those which suppress signals of intermediate frequency as well as image signals, is given by Harold A. Wheeler, Image Suppression in Superheterodyne Receivers, *Proc. I.R.E.*, Vol. 23, p. 569, June, 1935.

can be made almost complete at two carrier frequencies and will be high at frequencies in between. At the same time, the desired carrier is not balanced out. The necessary design formulas are given in Fig. 18, and the suppression obtained is independent of the antenna constants.

Radio signals of intermediate frequency will be heard in the output of a radio receiver if these signals are able to reach the input of the converter tube in appreciable amplitude. Response to such signals can be prevented by selecting the intermediate frequency so that it differs from the frequency of any strong local signal that is likely to be present by providing adequate selectivity in the radio-frequency section and, in particularly troublesome cases, by employing a trap circuit in the network coupling the antenna to the grid of the first tube.

Whistles can occur in a superheterodyne receiver when the harmonics of the intermediate frequency that are present in the output of the second detector get coupled back into the radio-frequency section. Thus, when a receiver having an intermediate frequency of 455 kc is tuned to approximately 910 or 1,365 kc, the second and third harmonics, respectively, of the intermediate frequency will produce an audible beat note with incoming signals. The remedy is to provide filters in the output of the second detector that will localize the harmonics of the intermediate frequency, and also to arrange the wiring so that coupling between the radio-frequency and second-detector circuits is a minimum.

When very strong signals are being received at a frequency that is approximately twice the intermediate frequency, it is possible for high-order modulation action to produce an output of the type $2s - o$, where s and o are the signal and local oscillator frequency, respectively. When the signal s is twice the intermediate frequency, this gives a frequency that is amplified by the i-f amplifier, but the exact frequency varies differently with tuning, *i.e.*, with variation in o, than does the intermediate frequency produced by normal converter action and so beats with the i-f produced by normal action. Since this type of spurious response is generated within the converter, it can not be eliminated by selectivity in the radio frequency. This undesired response is troublesome only when very strong signals are being received, and can be prevented by reducing the antenna pickup or providing a trap that discriminates against the troublesome signal.

The spurious responses discussed above represent the types that are most troublesome in superheterodyne receivers. Other types of combination frequencies are possible, and are occasionally encountered.[1]

Cross-talk and Cross-modulation.—Cross-talk or cross-modulation of the type discussed in Par. 24, Sec. 5, can occur when a receiver is tuned to a strong signal and there is also present another strong signal of a not too different frequency. Cross-talk then manifests itself by the fact that during intervals when there is no modulation on the desired carrier, one hears of the modulation of the undesired signal, but if the desired carrier is absent, the undesired program is no longer heard.

Such cross-talk is a result of third-order curvature in the plate-current characteristic of the first or possibly the second tube of the receiver. Most cross-talk problems are eliminated by the use of variable-mu tubes, since these have very low third-order curvature. Cross-talk may still be present, however, when signals of the order of a volt or more are induced in the antenna. This type of trouble can be minimized by making the transfer constant from antenna to first grid low so that the voltage

[1] A comprehensive discussion covering all the possible types of interfering responses is given by Howard K. Morgan, Interfering Responses in Superheterodynes, *Proc. I.R.E.*, Vol. 23, p. 1164, October, 1935. In this paper, it is shown that all the spurious responses, including the more common ones as well as the unusual types, correspond to frequencies of the interfering signal that are $no \pm ns$, $no \pm i/n$, ni, or $no + s/n$, where o is the oscillator frequency, s the frequency to which the receiver is tuned, i the intermediate frequency, and n the harmonic order (*i.e.*, an integer such as 1, 2, etc.).

at the grid of the first tube will not be too great for the A.V.C. system to handle. In extremely difficult cases, such as where a very powerful near-by station is involved, a tuned trap in the antenna circuit can be used.[1]

External Cross-modulation.[2]—The term *external cross-modulation* is applied to spurious frequencies generated by nonlinear or rectifying contacts in conduit, ground connections, light wires, etc., in the vicinity of the receiver that carry currents induced by the signal.[3] When strong radio signals are present, such nonlinear contacts develop new frequencies in appreciable amplitude, and these may then reach the receiver by conduction, or by direct radiation, causing signals to be heard at unexpected places on the dial. The most common new frequencies produced are harmonics of the transmitted carrier and simple combination frequencies of two carriers of the type $a \pm b$, $a \pm 2b$ and $2a \pm b$ and of three carriers of the type $a + b \pm c$, $a \pm b + c$, etc. In the case of third-order combinations it has been found that if the product of the three field strengths combining to give the cross modulation (*i.e.*, $E_a E_b^2$, $E_a^2 E_b$, or $E_a E_b E_c$, as the case may be) exceeds 0.001 volts,[3] then serious external cross modulation is likely to be observed.[4] In the case where the rectifying contact is on a power line, there is the possibility of strong hum modulation being produced on the carrier frequency of the station.

External cross-modulation is appreciable only when the carriers involved are strong, *i.e.*, of the order of one volt. This fact makes it undesirable to locate two transmitting stations so close to each other that the strong signal areas overlap. In particular, two transmitters operating from one antenna, or located on the same premises, are to be avoided.

When the presence of cross-modulation is suspected, one should first make sure that the trouble is external to the receiver by using traps to prevent strong local signals from reaching the receiver and also by checking the combination frequencies observed with those calculated as being most likely to be present in the case of cross-modulation. If the presence of external cross-modulation is confirmed, then the antenna and ground circuits should be checked for poor contacts, a capacity filter should be tried across the power line, and possible poor contacts in conduits and wiring investigated. If these do not remedy the difficulty, a new antenna with shielded lead-in wires should be tried.

Noise and Signal-to-noise Ratio.[5]—The ratio of signal-to-noise energy in a receiver output is one of the most important characteristics of a receiver.[6] Possible sources of noise are static and other disturbances induced in the antenna, thermal agitation in the tuned input circuit, tube noise (including particularly converter noise), and man-made disturbances of various types.

Noise energy is normally uniformly distributed over a considerable frequency band. The first requirement for a good signal-to-noise ratio is, therefore, that the

[1] Discussion of the especially difficult problem of operating receivers in close proximity to transmitters is given by P. C. Sandretto, Some Principles in Aeronautical Ground Radio Station Design, *Proc. I.R.E.*, Vol. 27, p. 5, January, 1939.

[2] D. E. Foster, A New Form of Interference—External Cross Modulation, *R.C.A. Rev.*, Vol. 1, p. 18, April, 1937.

[3] Thus poor connections in the receiving antenna system, such as in the ground connection, are particularly common sources of trouble. See A. James Ebel, A Note on the Sources of Spurious Radiations in the Field of Two Strong Signals, *Proc. I.R.E.*, Vol. 30, p. 81, February, 1942.

[4] See A. V. Eastman and L. C. F. Horle, The Generation of Spurious Signals by Nonlinearity of the Transmission Path, *Proc. I.R.E.*, Vol. 28, p. 438, October, 1940.

[5] The limitations to the sensitivity of radio receivers that result from noise, particularly at very high frequencies, are comprehensively discussed by E. W. Herold, An Analysis of the Signal-to-noise Ratio of Ultra-high-frequency Receivers, *R.C.A. Rev.*, Vol. 6, p. 302, January, 1942; D. O. North, The Absolute Sensitivity of Radio Receivers, *R.C.A. Rev.*, Vol. 6, p. 332, January, 1942.

[6] A good discussion of noise levels that can be expected at high frequencies is given by K. G. Jansky, Minimum Noise Levels Obtained on Short-wave Radio Receiving Systems, *Proc. I.R.E.*, Vol. 25, p. 1517, December, 1937. In this paper it is stated that a carrier to noise ratio of 18 db will give intelligible speech with 100 per cent modulation and an audio band of 5,000 cycles.

response band of the receiver be no wider than necessary to accommodate the signal.

The ultimate limit of signal-to-noise ratio is obtained when the receiver band width has the minimum possible value and all the noise in the receiver output is caused by thermal agitation in the input circuit to the first tube. The closeness with which this limit is approached can be determined by short-circuiting or detuning this input circuit and noting the amount of residual noise that still remains.[1] The residual present may be caused by (1) noise in the converter tube; (2) a noisy radio-frequency tube; and (3) inefficient tuned input circuit. The remedies for the last two sources of noise are obviously a better tube and more efficient circuits, respectively. The remedy for the first source of noise is either to use a stage of radio-frequency amplification between the antenna and the converter tube, so that both the signal and the thermal agitation noise in the input circuit will be raised to a level that will swamp the converter noise, or to employ a type of converter having low inherent noise, such as the triode plate detector. Discussion of converter noise is given in Par. 14, Sec. 7.

When a set is designed so that the thermal agitation in the input circuit determines the set noise, the best signal-to-noise ratio is obtained when the coupling between the antenna and the input circuit greatly exceeds an impedance match. The pickup sensitivity of the antenna should then be enough so that the normal static, cosmic hiss, and man-made disturbances induced in the antenna will deliver more noise to the first tube than the thermal-agitation voltages in the input circuit. This condition can be checked by replacing the antenna with an equivalent dummy impedance and noting whether or not the noise level drops appreciably.[2]

The disturbing effect of background noise in radio-telephone systems can be lowered by volume compression of the transmitted signals and by reduction in the receiver sensitivity during silent periods. Volume compression causes the weak components of the signal to be transmitted at a higher modulation level than would otherwise be the case, and so results in an improvement in the signal-to-noise ratio.[3] Lowering of the receiver sensitivity during silent periods reduces the deafening or masking effect of noise, and has been found to improve the intelligibility of telephone communication by an amount equivalent to approximately 5-db in transmitter power.[4]

When receivers are used near internal-combustion engines, as in the case of automobile and aircraft receivers, the noises from the ignition system are usually the limiting factor in determining signal-to-noise ratio. In automobiles, the ignition noise is reduced by placing suppressor resistances in the central distributor wire and sometimes in each individual spark-plug lead, by bonding various parts of the automobile to prevent irregular contacts, by inclosing the receiver in a metal case, and by employing by-pass condensers to prevent ignition disturbances from reaching the battery and traveling through the automobile wiring to dome lights, etc. In airplane work, it has been found necessary to inclose the entire ignition system in an electrostatic shield.

Vibration Effects—Microphonics.—Vibration can cause noise to appear in the output of a receiver either as a result of microphonic action in the tubes or through vibration of circuit components, particularly the plates of variable condensers. Vibra-

[1] F. B. Llewellyn, A Rapid Method for Estimating the Signal-to-noise Ratio of a High Gain Receiver, *Proc. I.R.E.*, Vol. 19, p. 416, March, 1931.

[2] A rough indication is obtained by simply disconnecting the antenna and noting the effect on noise output.

[3] The quantitative relation between the amount of compression and the improvement in signal-to-noise ratio is discussed by E. L. E. Pawley, The Effect of Volume Compression on the Tolerable Noise Level in Electrical Communication Systems, *Wireless Eng.*, Vol. 14, p. 12, January, 1937.

[4] C. C. Taylor, Radiotelephone Noise Reduction by Voice Control at Receiver, *Elec. Eng.*, Vol. 56, p. 971, August, 1937.

tion effects are especially important in receivers on airplanes, automobiles, tanks, etc., and also are troublesome when the receiver chassis is in close proximity to the loudspeaker. In the latter case a sustained audio-frequency howl may be present as a result of feedback around a path that includes an acoustic link.

The effects of vibration can be minimized by the use of rigid construction, by rubber mounting, and by inclosing the entire receiver in a case to protect it against sound waves traveling in air.[1] In some cases tubes and circuits operating at low power levels are placed on a separate subchassis that is connected to the main chassis through rubber supports.

Tracking and Alignment of Tuned Circuits in Radio Receivers.—Nearly all receivers are adjusted to the desired signal by means of variable condensers mounted on a common shaft to provide a single control. In the radio-frequency section of a receiver, the coils and variable condenser sections are made as nearly alike as possible. Exact alignment of the radio-frequency circuits at the high-frequency end of the tuning range is then obtained by the use of an adjustable trimmer condenser in shunt with

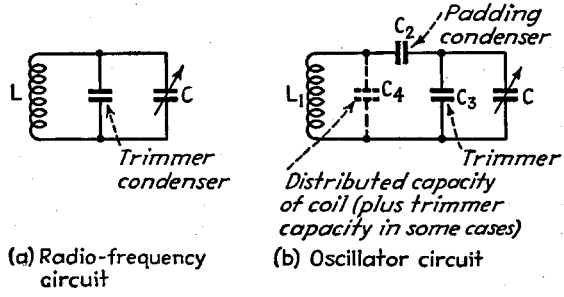


FIG. 19.—Trimmer systems for tuned radio frequency and superheterodyne oscillator circuits.

the regular tuning condenser, shown in Fig. 19*a*. In receivers covering only a single frequency band, it is possible to align at the low-frequency end of the tuning range by bending sectors on the end plates of the variable condensers. In all-wave receivers, it is necessary to depend for alignment at the low-frequency end of the tuning ranges either on careful manufacturing control that ensures uniformity of coils and condensers or on individually adjusting the inductance of each coil by shifting the position of the end turn until low-frequency alignment is obtained.

Oscillator circuits of superheterodyne receivers must be arranged so that for all tuning adjustments the difference between resonant frequencies of the radio-frequency and oscillator tuned circuits closely approximates the intermediate frequency. This result is normally obtained with satisfactory accuracy by using a variable tuning condenser for the oscillator section that is identical with the condensers used in the radio-frequency sections, but that employs an oscillator coil having less inductance than the inductance used in the radio-frequency sections, combined with series and shunt trimming condensers as shown in Fig. 19*b*. By properly proportioning such an arrangement, it is possible to obtain exactly correct tracking at three frequencies in any tuning range, with relatively small tracking error between these frequencies, as shown in Fig. 20. It is apparent that the maximum error in tracking is so small as to be of no consequence, provided that the selectivity of the radio-frequency sections is not excessively high.

[1] A discussion of antivibration mountings, with particular reference to aircraft conditions, is given by L. B. Hallman, Jr., Aircraft Radio Vibration, *Communications*, Vol. 20, p. 5, May, 1940.

The proper oscillator circuit proportions to obtain tracking can be calculated by the following formulas:[1]

Case 1. When $C_4 = 0$, or $C_4 << C_2$ (the Usual Case).

$$\begin{aligned} C_2 &= C_0F_0^2\left(\frac{1}{n^2} - \frac{1}{l^2}\right) \\ C_3 &= \frac{C_0F_0^2}{l^2} \\ L_1 &= L\left(\frac{l^2}{m^2}\right)\frac{(C_2 + C_3)}{C_2} \end{aligned} \tag{7}$$

Case 2. When $C_3 = 0$.

$$\begin{aligned} C_2 &= \frac{C_0F_0^2}{n^2} \\ C_4 &= \frac{C_0F_0^2}{l^2 - n^2} \\ L_1 &= L\left(\frac{l^2}{m^2}\right)\frac{C_2}{C_2 + C_4} \end{aligned} \tag{8}$$

Case 3. When C_4 Is Known.

$$\begin{aligned} C_2 &= A\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{C_4}{A}}\right) \\ C_3 &= \frac{C_0F_0^2}{l^2} - \frac{C_2C_4}{C_2 + C_4} \\ L_1 &= L\left(\frac{l^2}{m^2}\right)\frac{(C_2 + C_3)}{C_2 + C_4} \end{aligned} \tag{9}$$

Case 4. When C_3 Is Known.

$$\begin{aligned} C_2 &= \frac{C_0F_0^2}{n^2} - C_3 \\ C_4 &= \frac{C_2B}{C_2 - B} \\ L_1 &= L\left(\frac{l^2}{m^2}\right)\frac{(C_2 + C_3)}{C_2 + C_4} \end{aligned} \tag{10}$$

Check Formulas.—Equation for oscillator frequency:

$$f_1 = m\sqrt{\frac{f^2 + n^2}{f^2 + l^2}} \tag{11}$$

Equations for l^2, m^2, and n^2, in terms of oscillator constants:

$$\begin{aligned} l^2 &= \frac{C_0F_0^2}{C_3 + \dfrac{C_2C_4}{C_2 + C_4}} \\ m^2 &= \frac{C_0F_0^2}{\left(\dfrac{L_1}{L}\right)\left(C_4 + \dfrac{C_2C_3}{C_2 + C_3}\right)} \\ n^2 &= \frac{C_0F_0^2}{C_2 + C_3} \end{aligned} \tag{12}$$

[1] This follows Laboratory Series Report UL-8, RCA Radiotron Company. Other equivalent formulas are given by V. D. Landon and E. A. Sveen, A Solution of the Superheterodyne Tracking Problem, *Electronics*, Vol. 5, p. 250, August, 1932; Hans Roder, Oscillator Padding, *Radio Eng.*, Vol. 15, p. 7.

where f_0 = intermediate frequency, in mc.
F_1, F_2, F_3 = frequencies at which exact tracking is to be obtained, mc.

$$a = F_1 + F_2 + F_3.$$
$$b^2 = F_1F_2 + F_1F_3 + F_2F_3.$$
$$c^3 = F_1F_2F_3.$$
$$d = a + 2f_0.$$
$$l^2 = \frac{b^2d - c^3}{2f_0}$$
$$m^2 = l^2 + f_0^2 + ad - b^2.$$
$$n^2 = \frac{c^3d + f_0^2l^2}{m^2}$$

C_0 = maximum tuning capacity in radio-frequency stage, $\mu\mu$f.
F_0 = lowest frequency to which the radio-frequency circuit is to be tuned, mc.

$$L = \frac{25{,}330}{C_0F_0^2}, \text{ or, if } L \text{ is known, then } C_0F_0^2 = \frac{25{,}330}{L} \quad (L \text{ in microhenrys}).$$

$$A = C_0F_0^2\left(\frac{1}{n^2} - \frac{1}{l^2}\right). \quad \text{Required only for Case 3.}$$

$$B = \frac{C_0F_0^2}{l^2} - C_3. \quad \text{Required only for Case 4.}$$

Of these cases, Case 1 is of most practical importance. This is because it is preferable to place the trimmer condenser in shunt with the tuning condenser rather than in shunt with the coil in order to take care of variations in minimum capacity of the variable tuning condenser. When this is done, C_4 consists only of the coil distributed capacity, and is so small that it can be considered as merged with C_3 with negligible error.

The best average oscillator tracking is obtained when the frequencies of exact tracking (commonly termed cross-over frequencies) are so chosen that the maximum deviations of the tracking curve within the tuning range are all the same, as shown in Fig. 20.[1]

In all-wave receivers, each coil is provided with its own trimmer and padding condensers, and the combination of coil and condensers is switched across the variable tuning condenser. Each shunt trimming condenser is adjustable, while the series trimmers used in the oscillator circuits are usually fixed in value except in the case of the broadcast band.

After the constants required for the oscillator circuit have been calculated and the proper values inserted in the circuit, it is necessary to check the final adjustment of the trimmers by experiment. In doing this it is to be noted that with a given coil, adjustment of the shunt trimmer determines the high-frequency cross-over point, and the series padding condenser is the principal factor controlling the low-frequency cross-over. The middle-frequency cross-over is determined by the inductance of the oscillator circuit, with too much inductance causing the cross-over to occur at a frequency that is high. Details involved of the experimental alignment of receivers are discussed in Par. 37, Sec. 13.

The alignment and tracking problems involved in permeability tuning are analogous to those encountered where tuning is by means of a variable condenser. The

---

March, 1935; A. L. M. Sowerby, Ganging the Tuning Controls of a Superheterodyne Receiver, *Wireless Eng. and Exp. Wireless*, Vol. 9, p. 70, February, 1932.

Charts giving the solution of the design equations are presented by A. L. Green and R. Scott-Payne, Superheterodyne Tracking Charts, *Wireless Eng.*, Vol. 19, p. 243, June, 1942; p. 290, July, 1942. Also see discussion by G. W. O. Howe, Superheterodyne Receiver Tracking, *Wireless Eng.*, Vol. 19, p. 141, April, 1942.

[1] A discussion of how to achieve this result is given by M. Wald, Ganging Superheterodyne Receivers, *Wireless Eng.*, Vol. 18, p. 146, April, 1941.

chief differences are that trimmer slugs replace trimmer condensers, and variation in the fixed capacity of a permeability tuned circuit corresponds to change in the coil inductance with condenser tuning.

Tuning Indicator.—A tuning indicator is for the purpose of giving a visual indication of when the receiver is tuned so that the carrier is at the center of the response band of the receiver. Tuning indicators are useful in receivers provided with automatic volume control, since the automatic-volume-control system tends to maintain the loudness of the output substantially unchanged, even when the receiver is so badly mistuned as to distort the reproduced signals because of unequal transmission of the various side-band components.

Several forms of tuning indicators are in common use. One of these consists of a visual indicator tube (see Par. 14, Sec. 4) in which the luminous area is a sector of a circle having an angular spread determined by the negative bias applied to a pair of control electrodes. By deriving this bias from the automatic-volume-control system,

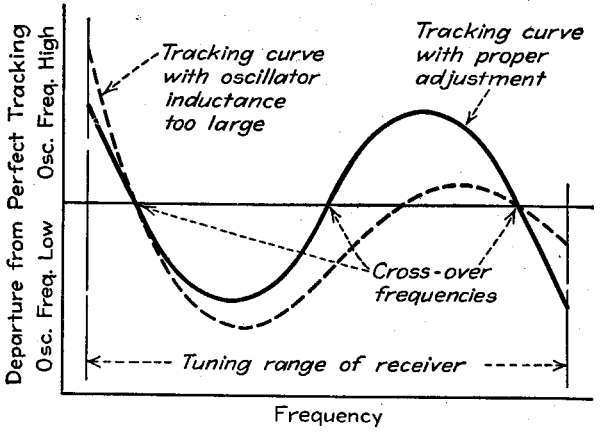


FIG. 20.—Typical tracking curves of superheterodyne oscillator.

the size of the luminous area serves as an aid to tuning. Other tuning indicators consist of a meter indicating the d-c current going to the plates of the tubes that are controlled by the A.V.C. system, while still another type employs a light valve operated by the d-c currents to the controlled tubes.

Special Problems Involved in High-fidelity Receivers.[1]—The usual broadcast receiver will not reproduce frequencies higher than 3,500 to 5,000 cycles. If high-fidelity reception is to be obtained, it is necessary to modify the receiver to respond to modulation frequencies up to 10 to 15 kc. This requires that the design of the audio-frequency, intermediate-frequency, and radio-frequency sections be modified to provide for increased band width. It is also necessary, if the reproduction is to sound natural, that the low-frequency range be extended in proportion to the extension of the high-frequency cutoff. A common working rule is that the product of lowest and highest frequency reproduced should be from 400,000 to 600,000.

Increasing the response band of the receiver introduces problems not present to the same degree in ordinary broadcast receivers. Thus amplitude distortion must be kept lower, because the amount of distortion that can be tolerated without produc-

[1] Discussions of the general problems involved in high-fidelity reproduction are given by Stuart Ballantine, Quality Radio Broadcast Transmission and Reception, *Proc. I.R.E.*, Vol. 22, p. 564, May, 1934; A. N. Goldsmith, Conditions Necessary for an Increase in Usuable Receiver Fidelity, *Proc. I.R.E.*, Vol. 22, p. 6, January, 1934; E. H. Scott, Scott High Fidelity Receivers, *Proc. I.R.E.*, Vol. 29, p. 295, June. 1941.

ing a noticeable degradation of the program is less the wider the frequency band.[1] Similarly, the ratio of signal to noise must be higher, because noise becomes increasingly objectionable as the fidelity is improved.[2] A filter is also necessary to eliminate 10-kc beats with adjacent-channel carriers. This filter is preferably of a type having a very narrow stop band that is centered on 10 kc but that does not appreciably affect frequencies somewhat higher or lower.[3]

High-fidelity reception is satisfactory only when the desired signal is so strong as to override all noise and interfering signals. When this is not the case, more satisfactory reception will be obtained by narrowing the band and sacrificing fidelity. This situation has led to the development of receivers in which the band width of the intermediate-frequency amplifier can be adjusted by the user to give the highest fidelity feasible under the receiving conditions existing for a particular station at the moment.[4] The band-width control can be accomplished in a variety of ways, one method being described in Fig. 37, Sec. 3. Means have also been devised that will vary the band width automatically in accordance with the strength of the received signal, but these have not been used except in laboratory receivers.[5]

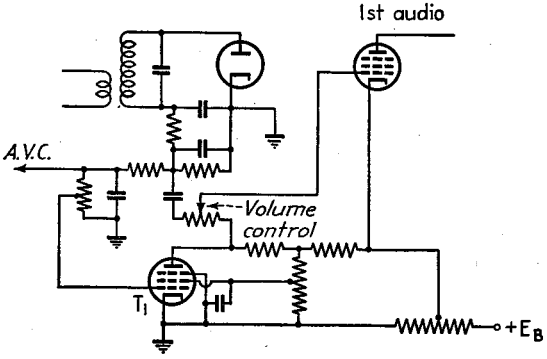


FIG. 21.—Circuit diagram of typical muting system.

Muting Systems—The Codan.—When automatic volume control is employed with a receiver having considerable sensitivity, a disagreeable amount of noise will be heard in the receiver output when no carrier is present, either as one listens for a station not on the air or as one tunes from one station to another. Arrangements for suppressing this noise have been given such names as *muting, quieting, or squelch systems, tuning silencers,* and *interchannel noise suppressors.* Such devices are also sometimes referred to as *codans,* from the first letters of the phrase "carrier-operated device antinoise."

A typical muting system is shown in Fig. 21. Here tube T_1 is so arranged that it biases the grid of the first audio tube beyond cutoff unless the grid bias of T_1 approaches or exceeds cutoff. By using the A.V.C. system to bias T_1, it is then possible to make the receiver inoperative until a carrier of predetermined amplitude is present. A

[1] Frank Massa, Permissible Amplitude Distortion of Speech in an Audio Reproducing System, *Proc. I.R.E.*, Vol. 21, p. 62, May, 1933.

[2] Experimental data showing the effect of band width on allowable noise are given by C. B. Aiken and G. C. Porter, Receiver Band-width and Background Noise, *Radio Eng.*, Vol. 15, p. 7, May, 1935.

[3] Such a filter is described by W. Baggally, An Improved Carrier Interference Eliminator, *Wireless Eng. and Expt. Wireless*, Vol. 12, p. 647, December, 1935.

[4] Harold A. Wheeler and J. Kelly Johnson, High Fidelity Receivers with Expanding Selectors *Proc. I.R.E.*, Vol. 23, p. 594, June, 1935.

[5] G. L. Beers, Automatic Selective Control, *Proc. I.R.E.*, Vol. 23, p. 1425, December, 1935; H. F. Mayer, Automatic Selectivity Control, *Electronics*, Vol. 9. p. 32, December, 1936; J. F. Farrington, Receiver with Automatic Selectivity Control Responsive to Interference, *Proc. I.R.E.*, Vol. 27, p. 239, April, 1939.

modification of this arrangement consists in operating tube T_1 from a separate branch of the intermediate-frequency amplifier that delivers its output to a second diode. By making this auxiliary intermediate-frequency branch very selective, the signal-to-noise ratio will be much higher for the noise-suppressor diode than for the detector diode. It is possible in this way to set the threshold of the system at a level where the signal is so far down in the noise as to be barely usable. This arrangement also has the advantage that the receiver delivers no output until the receiver is tuned exactly to the desired carrier frequency, thus making accidental mistuning practically impossible.

Muting systems find use in receivers for commercial communication circuits and in high-priced broadcast receivers.

Automatic Frequency Control.[1]—The difficulty of accurately tuning a receiver having automatic volume control, as well as difficulties caused by drifts in the frequency of the oscillator in receiving short-wave signals, can be eliminated by an arrangement that will automatically shift the frequency of the beating oscillator so as to produce an intermediate frequency of almost exactly the proper value, provided that the tuning is only roughly correct. This is termed *automatic frequency control* (abbreviated A.F.C.).

Automatic-frequency-control systems are composed of two essential parts. First, there is a section termed the *discriminator*, which develops a d-c voltage proportional to the extent that the intermediate frequency differs from the desired value. This voltage is used to control a second tube, termed a *reactance tube*, which is arranged to draw a reactive current having a value dependent upon the control voltage developed by the discriminator. The reactance tube is shunted across the oscillator tuned circuit, and the system is so arranged that when the intermediate frequency does not have the desired value, there results a control action on the reactance tube that shifts the frequency of the oscillator so that the intermediate frequency is closer to the desired value. It is a simple matter in this way to reduce the deviations of the intermediate frequency by a factor of 100 to 200.

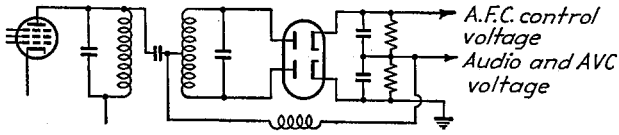
The two most commonly used discriminator circuits are shown in Fig. 22. These are essentially the same as the frequency-modulation detector circuits of Figs. 52 and 53, Sec. 7, with the difference that what was the audio output voltage in the case of frequency modulation becomes the d-c control voltage in the corresponding A.F.C. system. However, inasmuch as an exactly linear amplitude characteristic is not essential and the separation between the two peaks can be much less than in the case of frequency-modulated signals, the circuit proportions are different in the two cases. It is further to be noted that the frequency range over which the control is effective in A.F.C. systems is somewhat greater than the peak separation, whereas in the detection of frequency-modulated signals, the useful frequency range is less than this separation.

The reactance tube that is controlled by the A.F.C. bias consists of a pentode tube in which the plate-cathode circuit is in shunt with the tuned circuit of the oscillator, and the control-grid electrode is supplied with an exciting voltage derived from the alternating voltage existing across the tuned circuit but 90° out of phase with it. In such an arrangement the amplified grid voltage acting in the plate circuit draws an alternating plate current that is 90° out of phase with the voltage across the oscillator tuned circuit. The tube then acts as a shunting reactance having a magnitude depending upon the amplification of the tube, and hence upon the bias developed by the discriminator system. The numerical value of this reactance is

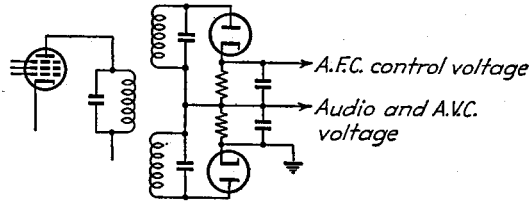
[1] D. E. Foster and S. W. Seeley, Automatic Tuning, Simplified Circuits, and Design Practice, *Proc. I.R.E.*, Vol. 25, p. 289, March, 1937; Hans Roder, Theory of the Discriminator Circuit for Automatic Frequency Control, *Proc. I.R.E.*, Vol. 26, p. 590, May, 1938; Charles Travis, Automatic Frequency Control, *Proc. I.R.E.*, Vol. 23, p. 1125, October, 1935; R. L. Freeman, Improvements in A.F.C. Circuits, *Electronics*, Vol. 9, p. 21, November, 1936.

$$\left.\begin{matrix}\text{Reactance of plate-}\\ \text{cathode circuit}\end{matrix}\right\} = \frac{E_1}{I_1} = \frac{1}{kg_m} \tag{13}$$

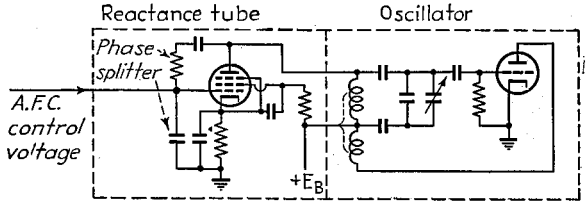
where g_m is the transconductance of the tube, k the absolute value of the ratio of alternating grid voltage to alternating plate-cathode voltage E_1, and I_1 is the alter-



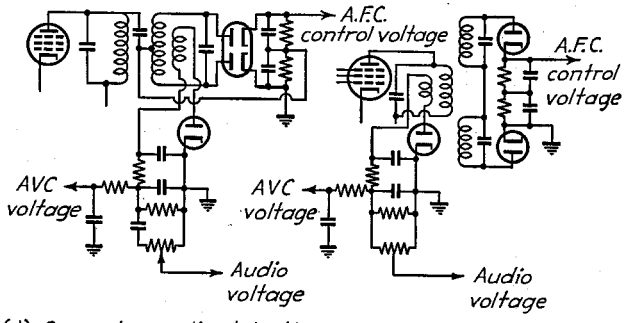
(a) Coupled-circuit type of discriminator



(b) Detuned-circuit type of discriminator



(c) Reactance tube and oscillator



(d) Separator audio detection

FIG. 22.—Circuits of common automatic-frequency-control systems.

nating current drawn by the tube. The required 90° phase shift of the grid exciting voltage of the reactance tube is normally obtained by a resistance-capacity phase splitter as shown in Fig. 22. For proper operation, it is necessary that the phase difference between the plate-cathode and control-grid voltages be exactly 90°. Otherwise the tube will draw a resistance component of current, and so shunt across the

oscillator tuned circuit a positive or negative resistance having a magnitude that will depend upon the A.F.C. bias. Such a resistance causes the amplitude of oscillations to depend upon the frequency deviations that are being corrected, and so is to be avoided.

The discriminator circuits of Fig. 22 can be used to provide detector and alsc A.V.C voltage, as well as the A.F.C. control potential. However, the detector output represents a distorted reproduction of the modulated wave because of the asymmetric action of the discriminator with respect to the two side bands. Accordingly, when high fidelity is important, it is necessary to employ a separate diode detector, as shown in Fig. 22*d*, for developing the audio output.

6. Code and Communications Receivers.—Receivers used for code and short-wave telephone signals in amateur and commercial services are usually modified

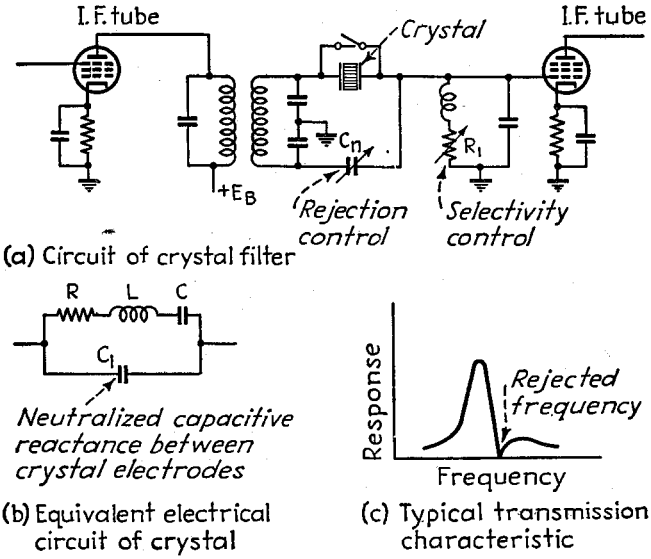


FIG. 23.—Typical crystal coupling arrangement used in single-signal receiver.

superheterodyne receivers in which certain refinements such as crystal filters, noise suppressors, i-f beating oscillators, etc., are added. Such receivers are commonly termed *communications receivers*, and their operation requires a degree of technical knowledge and skill greater than that possessed by the usual broadcast listener.

Code Reception.—Code receivers usually are ordinary superheterodyne receivers to which there has been added an oscillator that produces audio beats with the intermediate frequency, and thus delivers code signals in the form of an audio whistle. When the beating oscillator is turned on for code reception, it is necessary that the automatic-volume-control system that would normally be used in the reception of telephone signals be disconnected and a manual volume control substituted. Otherwise the local oscillator amplitude controls the receiver sensitivity.

Code reception is occasionally obtained by using a receiver consisting of an oscillating detector, preferably preceded by one stage of radio-frequency amplification and followed by one or two stages of audio amplification. Such a receiver has nearly as great sensitivity as a superheterodyne, but possesses less selectivity, greater tendency for cross-modulation, and is also not suitable for telephone signals. Although once widely used, it is accordingly now largely obsolete.

Single-signal Receiver.—A single-signal receiver is a superheterodyne provided with a crystal filter in the intermediate-frequency amplifier to give a high and controllable selectivity. A typical circuit arrangement is shown in Fig. 23*a*, in which the crystal is used as a coupling link between two tuned circuits. The transmission as a function of frequency with such a crystal coupling arrangement is illustrated in Fig. 23*c*. The transmission is maximum at the frequency for which the crystal is in series resonance (resonance of L and C in the equivalent electrical circuit), and is very poor at a slightly different frequency, such that the crystal is in parallel resonance (resonant frequency of L, C, and C_1 in equivalent electrical circuit). The selectivity obtained with the circuit of Fig. 23*a* can be varied by changing the resistance R_1, a high resistance causing the selectivity to approach that of the crystal alone, while a low resistance

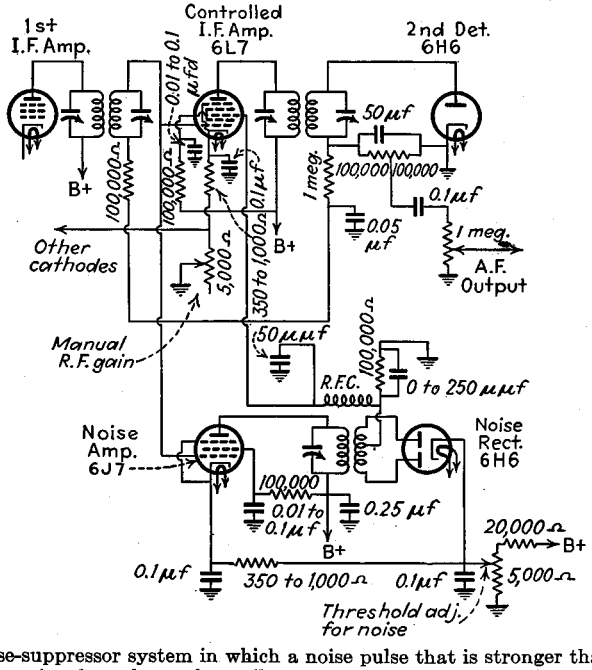


FIG. 24.—Noise-suppressor system in which a noise pulse that is stronger than the desired signal can be made to silence the receiver momentarily.

broadens the response curve. In this way band widths from 100 cycles or less up to 5,000 cycles can be obtained.

The frequency at which the transmission through the crystal is very low can be controlled by varying the neutralizing capacity C_n, which acts to neutralize the crystal capacity C_1. This rejected frequency can be made to lie anywhere from a few hundred to a few thousand cycles on either side of the desired frequency, according to the degree of neutralization and according to whether the system is over- or underneutralized.

Noise Suppression.—Communications receivers are frequently provided with special arrangements for suppressing impulse type of noise, such as that produced by ignition systems. These systems may take two forms. In the first, illustrated in Fig. 24, an auxiliary intermediate-frequency amplifier (labeled "noise amp.") is connected with its grid in parallel with the grid of the final intermediate-frequency amplifier (a 6L7 tube in Fig. 24) and delivers its output to an auxiliary rectifier. The direct-current output of this auxiliary rectifier biases the third grid of the 6L7 final

intermediate-frequency amplifier tube. With proper adjustment, a noise voltage having a peak amplitude greater than the signal being received will develop enough bias to make the final intermediate-frequency tube inoperative, thus silencing the receiver for the duration of the noise. Although this causes some distortion, the reception is much improved over that obtained without the noise suppressor.

A somewhat simpler although less effective form of noise suppressor consists of a simple limiter in the audio-frequency output of the receiver.[1] An example is shown in Fig. 25, and consists of a diode shunted across an audio-frequency coupling network and biased with a voltage slightly greater than the normal audio-frequency signal at

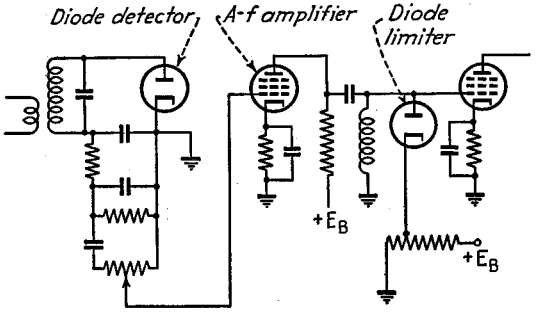


FIG. 25.—Limiter system for noise suppression.

that point. Any voltage in excess of the bias on the diode limiter will then be short-circuited by this tube, thus limiting the noise to a peak amplitude only slightly greater than that of the signal. To be effective, polarities must be arranged so that a large pulse voltage applied to the detector tends to make the plate of the limiting diode positive.

7. Miscellaneous Types of Receivers and Receiving Systems. *Tuned Radio-frequency Receivers.*—At one time, many broadcast receivers were of the tuned radio-frequency type. In such sets, the incoming signal is amplified by two or three stages of tuned-radio frequency amplification and then applied directly to the detector without undergoing a frequency change. Such receivers have the disadvantage of poor

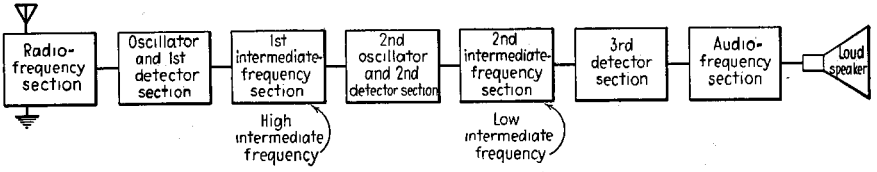


FIG. 26.—Schematic diagram of triple-detection receiver.

selectivity and low sensitivity in proportion to the number of tubes employed. They are accordingly practically obsolete.

Triple-detection Receivers.—The triple-detection receiver is a modified superheterodyne having two converters and two intermediate frequencies, as indicated schematically in Fig. 26. The first intermediate frequency is relatively high, to give effective image suppression, and the second intermediate frequency is comparatively low in order to increase the adjacent-channel selectivity.

[1] The effectiveness of amplitude limiters is highest when the frequency selectivity in the circuits after the limiter is greater than the selectivity preceding the limiting. See M. Wald, Noise Suppression by Means of Amplitude Limiters, *Wireless Eng.*, Vol. 17, p. 432, October, 1940.

Triple-detection receivers provide a combination of greater image discrimination and adjacent selectivity than can be realized in an ordinary superheterodyne. Such receivers are used extensively in short-wave commercial communication services, but are too complicated and too expensive for ordinary broadcast service.

Tone-corrected Receiver.—Proposals have been made from time to time that the radio receiver be designed to take advantage of the apparent demodulation of a weak signal that occurs in the presence of a strong signal when a linear detector is used (see Par. 15, Sec. 7). The idea is to employ such high i-f selectivity that even strong interfering carriers will be reduced to less amplitude than the carrier of the signal to which the receiver is tuned. Linear detection will then suppress the interfering signal almost entirely. Trimming of the desired side bands caused by the selective circuits is then corrected by audio-frequency equalization, so that all modulation frequencies of the desired carrier are reproduced equally. The term "stenode" has been applied to such an arrangement.

Analysis of this system of reception shows that when side bands of the desired and undesired carriers do not overlap, there is a material reduction in the interference but that if the side bands associated with the two signals do overlap, little or no benefit results.[1]

Single-side-band Receivers.—In reception of single-side-band signals in which the carrier has been suppressed, it is necessary to reinsert the carrier at the receiver before detection, as discussed in Par. 15, Sec. 7.

In order that speech may sound natural, it is necessary that the reinserted carrier be within 20 cycles of the correct value. At carrier frequencies up to several megacycles, this accuracy can be obtained by the use of stable crystal oscillators at transmitter and receiver. At higher transmission frequencies, however, crystal oscillators will not maintain their frequency with sufficient accuracy over long periods of time under commercial conditions. It is then necessary to transmit a pilot frequency to the receiver for the purpose of indicating at the receiving point the correct frequency to be reinserted. This pilot frequency can be transmitted as a carrier of reduced amplitude, or as some other frequency related to the original transmitted carrier. The pilot frequency is used either to control automatically the frequency of the local oscillations by some arrangement as described in connection with Fig. 22, or by mechanical means; or it can be separated by a crystal filter, amplified, passed through a limiter, and then in this reconditioned form used in place of the receiver oscillator.[2]

Receiving Systems for Minimizing Fading—Diversity and Musa Receivers.—The fading of broadcast and short-wave signals can be minimized in a number of ways as follows: (1) the use of automatic volume control; (2) frequency diversity; (3) space diversity; (4) polarization diversity; (5) Musa receiving system.

Automatic volume control is a great help in smoothing out variations in signal strength caused by fading, but is not a complete solution. Automatic volume control cannot be used in code receivers employing a local oscillator, and even with telephone signals does not prevent the amplitude from falling below the noise level. Also, automatic volume control does not prevent quality distortion.

[1] E. B. Moullin, An Outline of the Action of a Tone Corrected Highly Selective Receiver, *Proc. I.R.E.*, Vol. 21, p. 1252, September, 1933; F. M. Colebrook, High Selectivity Tone-corrected Receiving Circuits, H. M. Stationery Office, London, 1932.

[2] Further information on single-side-band receivers for short-wave operation is given by A. A. Roetken, A Single Sideband Receiver for Short Wave Telephone Service, *Proc. I.R.E.*, Vol. 26, p. 1455, December, 1938; F. A. Polkinghorn and N. F. Schlaack, A Single Sideband Short-wave System for Transatlantic Telephony, *Proc. I.R.E.*, Vol. 23, p. 701, July, 1935; N. Koomans, Single Sideband Telephony Applied to the Radio Link between the Netherlands and the Netherlands East Indies, *Proc. I.R.E.*, Vol. 26, p. 182, February, 1938.

Mechanical devices involving motors, or watt-hour devices, for actuating the A.F.C. system are described in the first two of these references.

Frequency diversity is based on the observed fact that frequencies differing by as little as 100 cycles tend to fade independently. Frequency diversity is extensively used in code transmitters, with both amplitude and frequency modulation employed, as discussed in Par. 2. By modulating the code transmitter in this way, energy is radiated simultaneously on several frequencies, and it is seldom that all of these will fade out simultaneously.

Space diversity[1] makes use of the experimentally observed fact that signals induced in antennas 5 to 10 wave lengths apart fade independently. When three or more such antennas are employed with separate receivers, it is extremely unlikely that a resulting signal obtained by combining the three outputs will fade out completely. In code receivers making use of this principle, the outputs of the three receivers are separately rectified and then added. The resulting d-c voltage is amplified, and then used to key a local tone oscillator, as shown in Fig. 27, to give code characters suitable for transmission over a telephone line. In the reception of telephone signals, the audio-frequency outputs of the three receivers are combined directly, preferably with

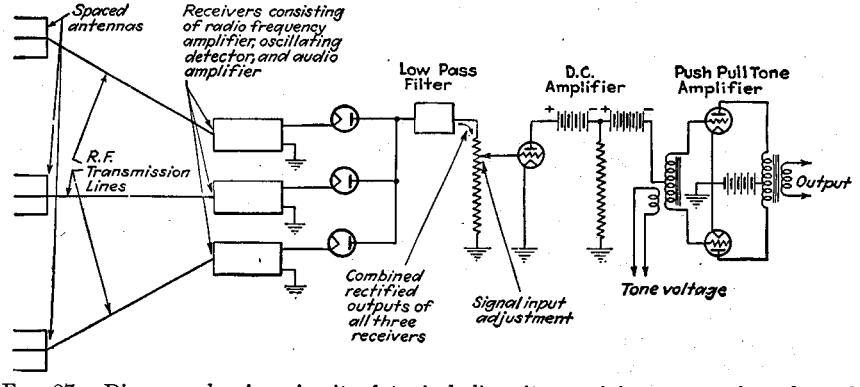


FIG. 27.—Diagram showing circuit of typical diversity receiving system for telegraph signals.

the use of square-law detectors in each receiver. The A.V.C. voltages developed by the three receivers are also added directly, and the combined voltage used to control simultaneously the gain of the three receivers, as in Fig. 28. In this way, the channel that receives the loudest signal dominates the situation at the moment, and the other channels contribute little or nothing to the output in the way of either noise or signal.

Polarization diversity is analogous to space diversity but makes use of the fact that the horizontally and vertically polarized components of a received signal do not fade together.

The Musa receiving system makes use of the fact that short-wave signals arrive at a receiver at certain preferred vertical angles that are relatively stable over an appreciable time interval, that the signal at a particular angle fades in magnitude but possesses little if any quality distortion, and that the envelope delay of the incoming signal is greater the higher the angle of arrival. The Musa system employs a directional antenna of the type described in Par. 23, Sec. 11, with the vertical directivity steerable by adjusting the phases with which the outputs of the individual antennas are

[1] Detailed discussions of diversity receiving systems of this type are given by H. H. Beverage and H. O. Peterson, Diversity Receiving System of R.C.A. Communications, Inc., for Radiotelegraphy, *Proc. I.R.E.*, Vol. 19, p. 531, April, 1931; J. B. Moore, Recent Developments in Diversity Receiving Equipment, *R.C.A. Rev.*, Vol. 2, p. 94, July, 1937; H. O. Peterson, H. H. Beverage, and J. B. Moore, Diversity Telephone Receiving System of R.C.A. Communications, Inc., *Proc. I.R.E.*, Vol. 19, p. 562, April, 1931.

combined. Several channels are normally derived from the same antenna system, with the vertical directivity of each channel individually steerable. One of these channels is set for the vertical angle corresponding to the best reception at the moment;

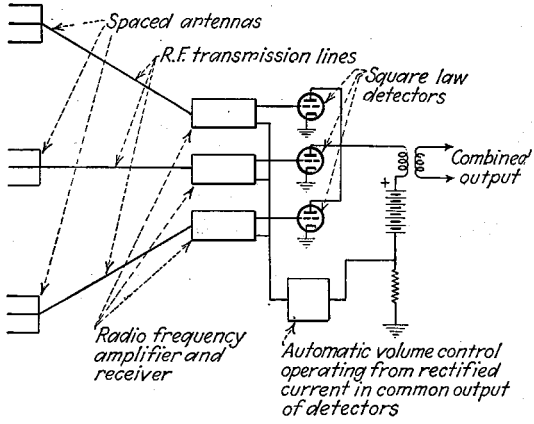


FIG. 28.—Schematic diagram of diversity telephone receiving system.

another, at the angle for second best reception, etc., while one is reserved for exploring the conditions at different vertical angles and for monitoring. The outputs of the separate channels are combined after rectification, with sufficient phase delay inserted in the lower angle channels so that the envelopes of all the channels will be in the

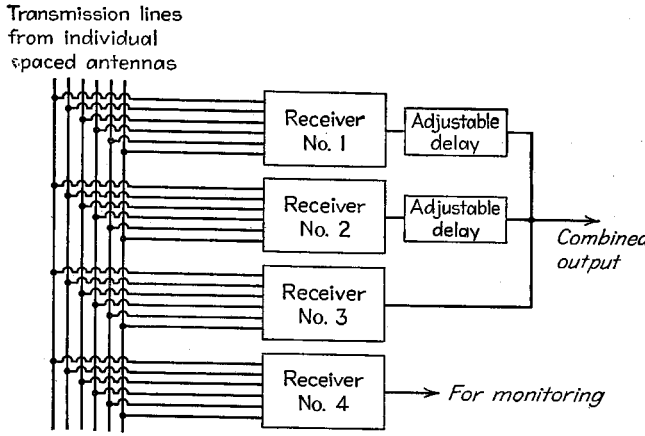


FIG. 29.—Schematic diagram of Musa receiving system.

same phase when added. If three channels are added, there is practically never a time when at least one will not be above the noise level, thus giving full diversity. A schematic diagram of the system is shown in Fig. 29.[1]

[1] Details are given by F. A. Polkinghorn, A Single Sideband Musa Receiving System for Commercial Operation on Transatlantic Radiotelephone Circuits, *Bell System Tech. Jour.*, Vol. 19, p. 306, April, 1940; H. T. Friis and C. B. Feldman, A Multiple Unit Steerable Antenna for Short-wave Reception, *Proc. I.R.E.*, Vol. 25, p, 841, July, 1937.

The Musa receiving system is the most effective arrangement that has been devised for the reception of short-wave signals. It not only eliminates fading but also avoids most of the quality distortion associated with ordinary fading. Furthermore, the use of directivity steering makes it possible to employ much greater vertical directivity than would otherwise be permissible. In the most elaborate installation of this type that has been made, the improvement in signal-to-noise ratio resulting from the added vertical directivity is 12 db over the ratio for a single directional antenna.

Superregenerative Receivers.[1]—Superregeneration is a form of regenerative amplification in which the circuit is alternately made oscillatory and nonoscillatory at a low radio-frequency rate. When this operation is properly carried out, tremendous amplification results, a one-tube superregenerative detector being capable of reaching the thermal-agitation noise level of the tuned input circuit.

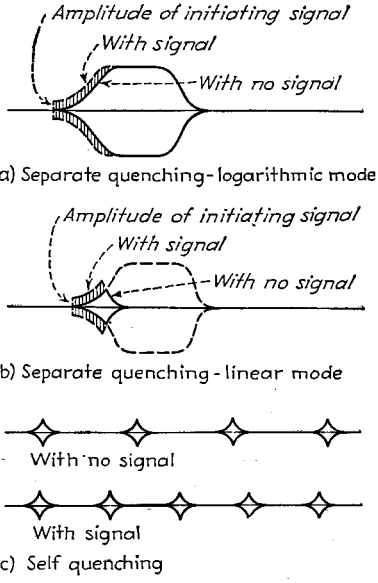


FIG. 30.—Envelopes of oscillations produced with different types of superregeneration.

Three modes of operation can be distinguished in superregeneration, namely (1) separate quenching, logarithmic mode, (2) separate quenching, linear mode, (3) self-quenching.

In separately quenched systems, an oscillator operating at a low radio frequency alternately allows oscillations to build up in the regenerative circuit and then causes them to die out, or be "quenched." In the logarithmic mode, the operating conditions are so chosen that the oscillations are able to build up to an equilibrium value before being quenched. In the absence of a signal, thermal-agitation noises in the input circuit produce the initiating voltage that starts the build-up process. However, when there is present an incoming signal larger than the thermal-agitation voltages, this signal provides the initiating pulse for the build-up period, and causes the equilibrium to be reached sooner than when the initiating pulse is smaller. This is illustrated in Fig. 30*a*, and has the effect of advancing the starting time of the oscillations. This causes the area under the envelope of the oscillations to increase by an amount indicated by the shaded area in Fig. 30*a*. If the oscillations are rectified by a detector, the detector output will accordingly give an indication of the incoming signal. In the case of a linear detector, the separately quenched logarithmic mode of operation gives

$$\frac{\left.\begin{matrix}\text{Detector output with signal}\\ \text{envelope amplitude } E_1\end{matrix}\right\}}{\left.\begin{matrix}\text{Detector output with signal}\\ \text{envelope amplitude } E_2\end{matrix}\right\}} = K_1 \frac{L}{R} \log_e \left(\frac{E_1}{E_2}\right) \tag{14}$$

[1] The discussion of theory given here follows the presentation of Frederick W. Frink, The Basic Principles of Superregenerative Reception, *Proc. I.R.E.*, Vol. 26, p. 76, January, 1938. For additional information on the subject, see Edwin H. Armstrong, Some Recent Developments in Regenerative Circuits, *Proc. I.R.E.*, Vol. 10, p. 244, August, 1922; Hikosaburo Ataka, On Superregeneration of an

where K_1 is a constant, L is the inductance of the circuit in which the oscillations build up, and R is the effective value of the negative resistance in this circuit during the build-up period. The relation between signal amplitude and detector output, as given by Eq. (14), is shown graphically in Fig. 31. When the signal is a modulated wave, the variations in the rectified output are a highly distorted reproduction of the modulation envelope, since the positive half cycles of modulation are practically suppressed, whereas troughs are accentuated, particularly with deep modulation. At the same time, the characteristic of Fig. 31 gives a very pronounced limiting action that is helpful in suppressing noise pulses of large amplitude but small duration.

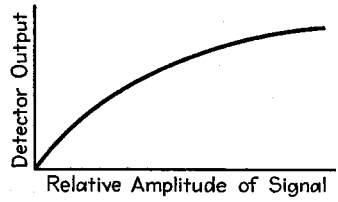


FIG. 31.—Relation of output of superregenerative system to amplitude of applied signal.

The second or linear mode of operation differs from the logarithmic mode in that the quenching frequency is so chosen in relationship to the circuit constants that the oscillations are not able to build up to the equilibrium value before being quenched. The resulting action is shown in Fig. 30*b*. As before, an incoming signal causes the starting time of the build-up period to be advanced by an amount determined by the amplitude of the signal, but now the earlier starting time also causes the maximum amplitude that is reached to be greater. Analysis of this method of operation shows that

$$\text{Detector output} = K_2\omega_r \frac{L}{R^2}\left(\epsilon^{\frac{R}{2L}t_1} - 1\right)E \qquad (15)$$

where E is the amplitude of the incoming signal, ω_r is 2π times the resonant frequency of the regenerative circuit, t_1 is the length of time the oscillations are allowed to build up, K_2 is a constant, and L and R are as in Eq. 14). It is seen from Eq. (15) that in this method of operation the rectified output is directly proportional to the incoming signal. This avoids distortion of the modulation envelope, but has the disadvantage that the limiting action has been lost and that the adjustments are quite critical if good sensitivity is to be obtained.

In self-quenching operation, a grid leak-condenser combination is used in the regenerative circuit, with proportions such that interrupted oscillations are generated (see Par. 2, Sec. 6). In the absence of an incoming signal, oscillations are initiated by the thermal-agitation noises, build up to a critical amplitude, and then die out. An incoming signal larger than the thermal noises then causes the build-up time to be advanced, so that the peak is reached sooner than would otherwise be the case, and the oscillations die out sooner. The result is that the frequency of interruption (the quench frequency) increases with the signal strength. At the same time the amplitude reached by the oscillations and the shape of their envelope are not affected by the signal strength, as shown in Fig. 30*c*. Analysis of the rectified output resulting under these circumstances shows that it follows the same logarithmic law given by Eq. (14) and shown in Fig. 31.

For best results, the quenching frequency of a superregenerative receiver must be carefully chosen. In the logarithmic separately quenched mode of operation, the quenching frequency should be as high as possible and still allow the oscillations to build up to their equilibrium amplitude before being quenched. With separately quenched linear operation, the quenching frequency should be as high as possible while still not permitting the oscillations to build up to equilibrium value with small

Ultra-short-wave Receiver, *Proc. I.R.E.*, Vol. 23, p. 481, August, 1935; M. G. Scroggie, The Super-regenerative Receiver, *Wireless Eng.*, Vol. 13, p. 581, November, 1936.

or moderate incoming signals. The proper quenching frequency in this case is quite critical. In self-quenching arrangements, the circuit adjustments should be such that the critical amplitude at which the oscillations begin to die out is reasonably large, and the quenching frequency should then be as high as possible and still allow the oscillations to die out below the noise level before a new build-up cycle begins.

Typical superregenerative circuits are shown in Fig. 32. In the separately quenched arrangement, the low-frequency oscillator acts as a modulator for the regenerative tube, allowing it to generate oscillations on one-half of he modulation cycle and causing a nonoscillatory condition to exist during the other half of the modulation cycle.

Receivers employing superregeneration find their chief usefulness in the wavelength range 0.5 to 10 meters. For such wave lengths, superregeneration provides a simple means of obtaining a very large amount of radio-frequency amplification at frequencies that are difficult to amplify by conventional methods. Also, when two

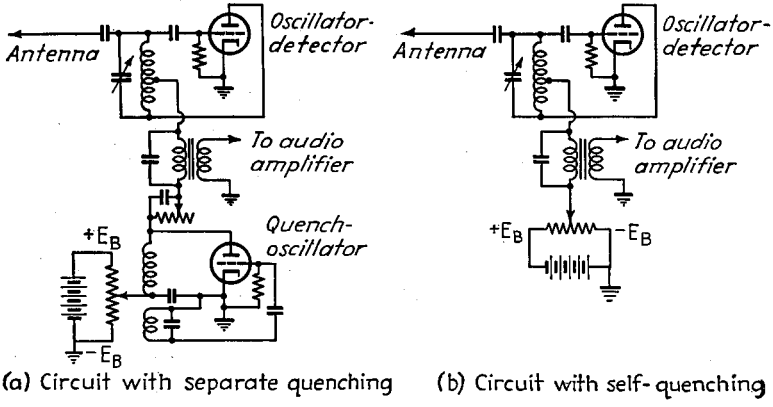


FIG. 32.—Typical circuits of superregenerative receivers.

signals are present, the stronger one controls the initiation of the build-up period, and so suppresses any weaker signals that may be present. With logarithmic operation, the limiting feature is very valuable in suppressing noise. Superregeneration at the same time has the disadvantage of providing poor selectivity[1] and of developing a characteristic hiss in the absence of a signal. This hiss represents amplified thermal agitation noise, and disappears only when a carrier of greater amplitude than thermal agitation is present.

TRANSMITTING AND RECEIVING SYSTEMS EMPLOYING FREQUENCY MODULATION

8. Frequency-modulated Transmitters. *Frequency-modulation Transmitters Based upon Reactance Tube.*—The commonest method of obtaining frequency modulation consists in using a reactance tube to vary the frequency of an oscillator in accordance with the modulated voltage, as described in Par. 17, Sec. 7. The oscillator of such a system is normally operated at a frequency of a few megacycles and at a low power level. The frequency is then increased to a desired value by means of a chain of

[1] This poor selectivity can be avoided by the use of tuned radio-frequency amplification before superregeneration takes place. In some cases a superheterodyne arrangement has been used to provide selectivity, with superregeneration being employed in the second detector to give limiting action against noise. Thus see Stewart Becker and L. M. Leeds, A Modern Two-way Radio System, *Proc. I.R.E.*, Vol. 24, p. 1183, September, 1936.

harmonic generators, and the required output power is obtained through the amplification of the harmonic generators supplemented by additional Class C power amplification as required. The modulated oscillator is operated at a low power to reduce the power-handling capacity required in the reactance tube. Operation of the modulated oscillator at a moderate radio frequency makes it possible to use only a small frequency swing (because of the subsequent frequency multiplication), and increases the inherent frequency stability obtainable in the oscillator.

High stability of the mean frequency can be achieved only by continuous and automatic monitoring, with the use of a crystal as a reference. A typical arrangement is illustrated schematically in Fig. 33. Here the crystal monitor output, after passing through a frequency tripler, is applied to a converter along with output derived from the transmitter. To give a suitable difference frequency, the crystal monitor operates at such a frequency that the difference between three times the crystal frequency and the desired mean frequency of the frequency-modulated transmitter is some predetermined intermediate frequency, for example, 450 kc. The output of the converter is applied to a discriminator of the type used in frequency-

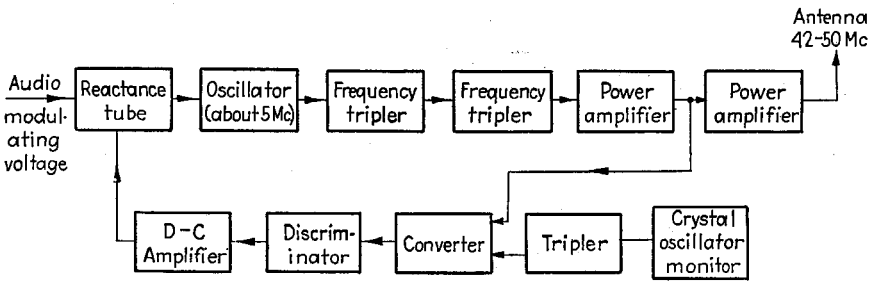


FIG. 33.—Typical arrangement whereby the mean frequency of a frequency-modulated broadcast transmitter is automatically regulated by means of a crystal monitor.

modulation receivers, which is so proportioned that when the transmitter is exactly at the correct frequency, the discriminator develops zero output. Frequency deviations of the transmitter will, however, vary the difference frequency and so will cause the discriminator to develop a d-c output voltage having a polarity determined by the sense of the transmitter drift. This d-c output voltage, after suitable d-c amplification, is applied to the grid of the reactance in such a way as to modify the frequency of the oscillator to correct for the drift in frequency.

The frequency obtained with a transmitter arranged as in Fig. 33 is closely but not absolutely controlled by the monitoring crystal. This is in part because the control that the crystal monitor exerts upon the transmitted frequency is dependent upon the center frequency of the discriminator circuit, and in part because the frequency-control system only acts to make the transmitted frequency approach closer to the desired value. In order for any control effect at all to be present it is necessary that there be some frequency deviation. Factors that make for high frequency stability are a low frequency for the oscillator, large direct-current amplification between the output of the discriminator and the reactance tube, oscillator design for maximum inherent frequency stability, discriminator circuits designed for minimum temperature coefficient, and heat-oven regulation of the oscillator and discriminator tuned circuits.[1] When reasonable attention is paid in the design to minimizing

[1] For further discussion see E. S. Winlund, Drift Analysis of the Crosby Frequency-modulated Transmitter Circuit, *Proc. I.R.E.*, Vol. 29, p. 390, July, 1941.

frequency drift, frequency stability adequate to meet all ordinary requirements can be obtained without undue difficulty.

Closer control of the transmitted frequency from a crystal monitor can be obtained by the arrangement illustrated schematically in Fig. 34.[1] Here a reactance tube controls an oscillator that operates at one-eighth of the output frequency. The output of the oscillator is applied to a chain of frequency dividers[2] consisting of ten stages, each halving the frequency, thus giving an output of the order of 5,000 cycles. The monitoring crystal likewise has its output applied to a chain of frequency dividers, and the crystal frequency and number of associated dividers are so chosen that when the transmitter output is exactly on the correct frequency, the outputs of the two systems of dividers will be exactly the same frequency. The outputs of these two

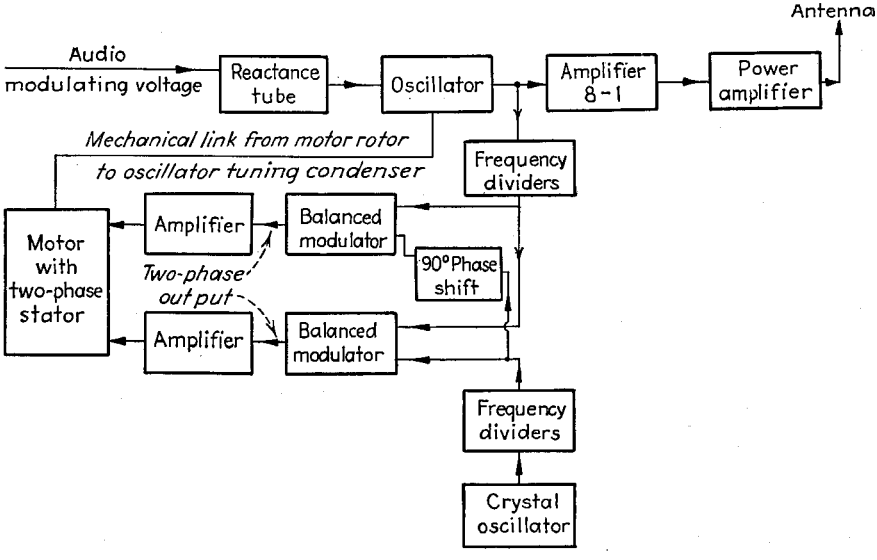


FIG. 34.—Schematic diagram illustrating method by which the mean frequency of a frequency-modulated transmitter can be made to have an exact relation to the frequency of a crystal monitor, irrespective of all other factors.

systems of dividers are applied to two balanced modulators, in the input of one of which a 90° phase shift is inserted as shown. The two systems of balanced modulators accordingly develop an output difference frequency proportional to the deviation in frequency of the transmitter from the desired value. The outputs of the two systems of balanced modulators are in quadrature, and therefore comprise a two-phase system

[1] See J. F. Morrison, A New Broadcast-transmitter Circuit Design for Frequency Modulation, *Proc. I.R.E.*, Vol. 28, p. 444, October, 1940.

[2] The frequency dividers used in this system are those described by R. L. Miller, Fractional-frequency Generators Utilizing Regenerative Modulation, *Proc. I.R.E.*, Vol. 27, p. 446, July, 1939. These frequency dividers are essentially balanced modulator systems in which the frequency to be divided is applied to one set of input terminals, while the half-frequency obtained from the output is applied to the other set of input terminals. Modulation of the half-frequency output upon the input produces a frequency component that is half the input frequency, so that the output is capable of supplying the input required by the second set of input terminals. In the frequency-modulation transmitters that have been built utilizing frequency dividers in the control of the transmitted frequency, copper oxide modulators have been employed.

Additional information on frequency dividers of this type is given by R. L. Fortescue, Quasi-stable Frequency-dividing Circuits, *Jour. I.E.E.*, Vol. 84, p. 693, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 171, June, 1939: H. Sterky, Frequency Multiplication and Division, *Proc. I.R.E.*, Vol. 25, p. 1153. September, 1937.

with the sense of rotation depending upon whether the transmitter frequency is above or below the desired value. This two-phase output, after amplification, is applied to the stator windings of a two-phase induction motor, the rotor of which is geared to the tuning control on the oscillator, as indicated. Hence when the output frequency deviates from the value corresponding to that set by the crystal monitor, the rotor turns in a direction depending on the sense of the deviation, and this rotation continues until the tuning of the master oscillator has been readjusted to eliminate the frequency deviation. When this has been achieved there is no difference in frequency between the outputs of the systems of frequency dividers; so no voltage is applied to the two-phase motor, and the control system comes to rest. As compared with the arrangement of Fig. 33, this system is somewhat more complicated, but it has the advantage that the output frequency is controlled absolutely by the crystal monitor, with resulting increased frequency stability.

Frequency-modulation Transmitters Employing Phase-shift Modulation.—The details of frequency-modulated transmitters employing the phase-shift method of modulation described in Fig. 51, Sec. 7, depend upon the required frequency swing, and the lowest

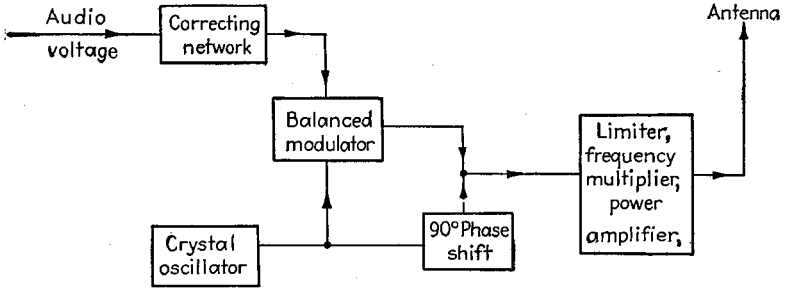


FIG. 35.—Simplified schematic block diagram showing transmitter employing phase-shift method to produce a frequency-modulated wave, in which the transmitted frequency is derived from the modulator output by direct frequency multiplication.

modulating frequency that must be reproduced without excessive distortion. When a large frequency swing is not required and the lowest frequency that need be modulated on the system is relatively high, as is the case in police radio and similar communication systems, an arrangement of the type illustrated schematically in Fig. 35 is suitable. The frequency swing that can be obtained in this way, assuming that the lowest modulating frequency is 200 cycles and that the maximum phase deviation away from the mean that can be obtained is about one-half radian, is 100 cycles [see Eq. (29), Sec. 7]. On the basis of a transmitted frequency that is approximately 250 times the modulator frequency, the maximum frequency deviation from the mean that can be obtained is of the order of 25 kc.

When greater frequency deviations are required, or when the lowest modulating frequency is less, as is the case in high-fidelity broadcast transmitters, the arrangement illustrated in Fig. 36 is employed.[1] Here the output of the modulator is raised to a high value by a series of frequency multipliers that serve to increase the frequency deviation. This output is then reduced to a comparatively low value by heterodyning with crystal controlled oscillations of slightly different frequency. This permits the use of additional frequency multiplication before the desired frequency is obtained, so that a greater frequency deviation can be obtained than if the output frequency is derived directly from the modulator by simple frequency multiplication. At the

[1] Edwin H. Armstrong, A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, *Proc. I.R.E.*, Vol. 24, p. 689, May, 1936.

same time an arrangement of the type shown in Fig. 36 has the disadvantage that a small percentage change in the frequency of either crystal oscillator will make a much larger percentage change in the mean frequency of the transmitter output. The stability is hence less than for arrangements of the type illustrated in Fig. 34, although by employing well-designed crystal oscillators it is possible to meet commercial requirements.

9. Receivers for Frequency-modulated Signals.—The more important ways in which a frequency-modulated receiver differs from a receiver for amplitude-modulated signals are as follows: (1) The second detector of the ordinary amplitude-modulated receiver is replaced by a frequency-modulated detector, normally a discriminator of the type illustrated in Fig. 52, Sec. 7. (2) One or two voltage-limiting stages are provided in the intermediate-frequency amplifier just preceding the second detector (discriminator). (3) The total gain from antenna to the input terminals of the discriminator (final detector) is preferably greater than required in an ordinary

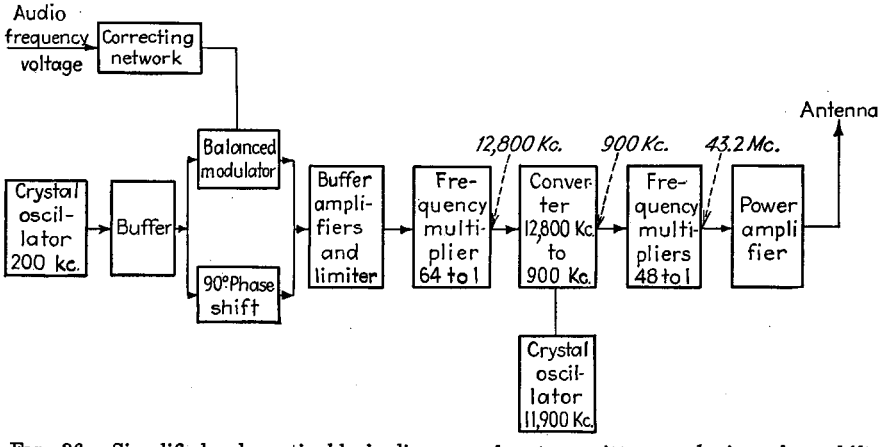


FIG. 36.—Simplified schematic block diagram of a transmitter employing phase-shift modulation and arranged to produce a high-frequency deviation.

amplitude-modulation receiver. (4) The intermediate-frequency amplifier must have a response band much wider than that required with amplitude-modulation receivers in order to accommodate the total range through which the frequency of the signal may be varied. (5) The intermediate frequency is usually quite high, of the order of 4 megacycles or more. (6) Such expedients as two intermediate frequencies, corresponding to the ordinary triple-detection receiver, are frequently used to keep within limits the amount of intermediate-frequency amplification that must be obtained at any one frequency.

The limiters in a frequency-modulation receiver are for the purpose of removing any amplitude modulation that may be present in the wave applied to the discriminator because of noise, interference, etc. The removal of amplitude modulation reduces the noise and interference developed in the receiver output as discussed below. A single stage can be used to accomplish the limiting, but for really good results it is desirable to employ two stages. The circuit diagram of a two-stage limiter is shown in Fig. 37. The first tube T_1 is a sharp cutoff pentode, and is operated at a screen voltage such that a bias of approximately 4 volts will drive the grid to cutoff. Sufficient amplification is provided between the antenna and the input to T_1 so that the signal applied to T_1 will be appreciably in excess of 4 volts under normal conditions.

The grid-leak grid-condenser combination RC will then develop sufficient bias to allow the grid of T_1 to be driven slightly, but only slightly, positive at the peak of each cycle. This makes the output voltage of tube T_1 relatively independent of the signal strength, because this tube is operated under "saturation" conditions. In Fig. 37 the first limiter stage T_1 is resistance-coupled to a second limiting stage T_2, which is operated in a similar manner and removes the smaller residual variations in amplitude. For best operation, the time constant of RC should be relatively short, while the time constant of the combination $R_1R_2C_1$ should be somewhat larger.

The full possibilities of frequency modulation in reducing noise and interference can be realized only if sufficient signal is always present at the limiter input to provide effective limiting action. This requires considerably more amplification between antenna and final detector (discriminator) than is needed in the ordinary amplitude-modulated receiver. Furthermore, the intermediate-frequency stages must be wide band in order to accommodate the wide frequency swing of the frequency-modulation signal, and this tends to reduce the gain per stage. At the same time, frequency-modulation stations operate at such a high-frequency part of the spectrum that the amount of gain that can be obtained from a radio-frequency stage is relatively small.

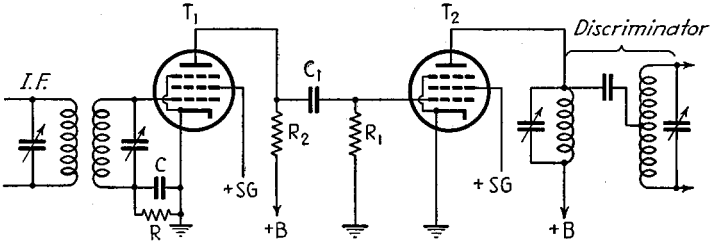


FIG. 37.—Circuit of two-stage limiter suitable for use in a frequency-modulation receiver.

The result of all this is that the gain the intermediate-frequency amplifier must provide, even when two stages of radio-frequency amplification are used, is greater than it is practicable to handle without trouble from regeneration. This has led to the use in many cases of triple-detection receivers, thus giving two intermediate frequencies between which the required i-f gain can be divided. Such triple-detection receivers require considerable care in design, because with two oscillators and two intermediate frequencies, the possibilities of spurious responses through oscillator harmonics, and through nonlinearities in the converters, are greatly increased. An ingenious method of overcoming the need of two oscillators in the case of a frequency-modulation receiver for the 42 to 50 mc broadcast band is illustrated in Fig. 38.[1] Here the same oscillator T_3 is used as the beating oscillator for the first and second converters. The frequency relations required between the incoming signal, the oscillator, the first intermediate frequency, and the second intermediate frequency are given in the figure. In such a system the radio-frequency input circuit L_1C_1, the oscillator tuning L_2C_2, and the tuning of the first intermediate-frequency circuit L_3C_3 are all varied from a common ganged control.

When a single intermediate frequency is used, the intermediate frequency is commonly of the order of 4 to 5 megacycles. In the case of triple-detection systems, one of the intermediate frequencies is usually of this order, and the other may be higher, as illustrated in Fig. 38, or alternatively one may design[2] the receiver to have a second

[1] For further details see J. A. Worcester, Jr., Recent Improvements in Frequency-modulation Receiver Design, *R.M.A. Tech. Bull.* 2, Nov. 12, 1940.

[2] An example of such a receiver is given by H. E. Thomas, The Development of a Frequency-modulated Police Receiver for Ultra-high-frequency Use, *R.C.A. Rev.*, Vol. 6, p. 222, October, 1941.

intermediate frequency of about 450 kc, similar to that of an ordinary receiver. The optimum value for the intermediate frequencies depends upon the width of the band over which the receiver must tune, the likelihood of interfering signals being present at other frequencies, etc. In the case of the frequency-modulation broadcast band of 42 to 50 mc, it has been found that if a single intermediate frequency is used, the best value is at least 4 mc.[1] Values up to 8 or 10 mc are still better in reducing spurious response, but entail some sacrifice in gain per stage and in frequency stability.

10. Miscellaneous Features of Frequency-modulation Systems. *Noise Reduction Realized with Frequency-modulation Systems.*[2]—One of the most valuable properties of frequency-modulation systems is the improvement of signal-to-noise ratio that is obtained as compared with amplitude-modulation systems. This improvement can be explained as follows. The noise voltages encountered in radio-frequency circuits

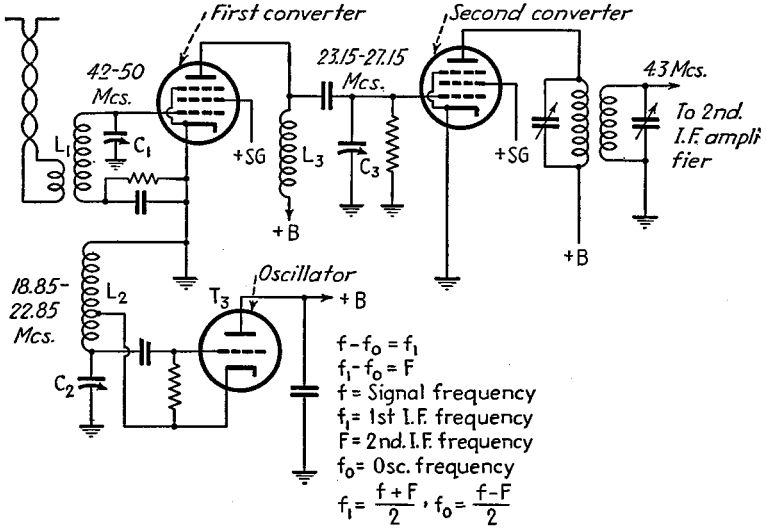


FIG. 38.—Circuit of input section of triple detection receiver in which a single-beat oscillator is used for both frequency conversions.

can be considered as being modulated both in amplitude and in frequency. When such voltages are superimposed upon an ordinary amplitude-modulated signal, disturbances in the amplitude of the modulation envelope are produced corresponding to the noise, and these disturbances appear in the output of the detector. In the case of frequency-modulated signals the presence of noise produces two effects that must be considered. In the first place, the noise causes an amplitude disturbance, just as in the case of amplitude-modulation systems. In the second place, the frequency-modulated component of the noise produces variations in the frequency swing of the desired signal, and hence superimposes a noise frequency modulation upon the signal. The first effect, *i.e.*, the amplitude modulation produced by noise, can be eliminated by the use of a receiving system that is insensitive to amplitude

[1] Further discussion on the subject is given by Dudley E. Foster and John A. Rankin, Intermediate-frequency Values for Frequency-modulated-wave Receivers, *Proc. I.R.E.*, Vol. 29, p. 546, October, 1941.

[2] For further discussion and for details of the analysis leading to the results stated here, see Murray G. Crosby, Frequency Modulation Noise Characteristics, *Proc. I.R.E.*, Vol. 25, p. 472, April, 1937; Hans Roder, Noise in Frequency Modulation, *Electronics*, Vol. 10, p. 22, May, 1937; John R. Carson and Thornton C. Fry, Variable Frequency Electric Circuit Theory With Application to the Theory of Frequency-modulation, *Bell System Tech. Jour.*, Vol. 16, p. 513, October, 1937.

variations. The ordinary discriminators used in the detection of frequency-modulated signals tend to be nonresponsive to amplitude variations, and when supplemented by the use of limiters in the intermediate-frequency system, amplitude disturbances can be almost completely removed. The disturbances in the frequency swing produced by the noise are, however, still present. Their effect on the receiver output can be made small, however, by employing a frequency swing of the transmitted signal that is large. In this way the frequency swing caused by the signal can be made as large as desired in proportion to that arising from noise, with corresponding possibilities for a large signal-to-noise ratio in the output.

The details of the noise-suppressing action, and the amount of improvement obtained, depend upon the carrier-to-noise ratio, the frequency deviation ratio (*i.e.*, ratio of maximum frequency swing from the mean to the maximum audio frequency of the modulation), and upon whether the noise is of a random type or in the form of pulses (ignition noise). When the carrier amplitude is much greater than the noise peaks, one has[1]

For random noise:

$$\frac{\left\{\begin{matrix}\text{Signal-to-noise ratio with} \\ \text{frequency modulation}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Signal-to-noise ratio with} \\ \text{amplitude modulation}\end{matrix}\right\}} = \sqrt{3}\,\frac{\{\text{peak frequency swing}\}}{\left\{\begin{matrix}\text{maximum audio modu-} \\ \text{lating frequency}\end{matrix}\right\}} \tag{16}$$

For pulse noise:

$$\frac{\left\{\begin{matrix}\text{Signal-to-noise ratio with} \\ \text{frequency modulation}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Signal-to-noise ratio with} \\ \text{amplitude modulation}\end{matrix}\right\}} = 2\,\frac{\{\text{peak frequency swing}\}}{\left\{\begin{matrix}\text{maximum audio modu-} \\ \text{lating frequency}\end{matrix}\right\}} \tag{17}$$

Equation (16) applies for both rms and peak signal-to-noise ratio in receiver output, while Eq. (17) applies only for peak values.

Equations (16) and (17) apply only when the carrier-noise ratio exceeds a certain critical or threshold value. For ratios below this threshold value the improvement is less, and may even be no better than in an amplitude-modulation system. The effect of the carrier-to-noise ratio upon the improvement obtained in the signal-to-noise ratio existing in the output of a frequency-modulation system, as compared with a corresponding amplitude-modulation system, is shown in Fig. 39[2] for the case of random noise. It is seen that there is a critical ratio of carrier to noise, below which there is little if any improvement in the ratio of signal to peak noise existing in the output of the frequency-modulation system. This value of carrier-to-noise ratio that must be exceeded in order to realize the improvement indicated by Eq. (16) is greater the higher the frequency-deviation ratio (because this means a wider response band of the receiver, and hence the acceptance of more noise by the receiver), and is also greater with impulse noise than with random noise. The critical ratio in the case of random noise is of the order of 2 to 4 in wide-band systems, and several times as great with pulse noise. The rms noise appearing in the receiver output behaves as indicated by the dotted lines in Fig. 39. In contrast with peak noise, the rms noise drops off only slowly as the carrier-to-noise ratio becomes less than the critical value. Even with carrier-to-noise ratios of unity there is some improvement in signal-to-rms noise effected by the use of frequency modulation. This difference between rms noise output and peak noise output is particularly pronounced when the deviation ratio of the frequency-modulation signal is small.

[1] See Crosby, *loc. cit.*

[2] From Crosby, *loc. cit.*

Frequency-modulation receiving systems having a limiter in the intermediate-frequency amplifier system produce a limiting action on the frequency swing. In conjunction with the selectivity of the audio- and intermediate-frequency circuits, this limits the maximum value of noise that can possibly appear in the output to a value that is the maximum value of signal output obtainable (as fixed by the limiter) divided by the square root of the deviation ratio in the case of random noise, and to a value not greater than this maximum possible signal divided by the deviation ratio in the case of pulse noise. This action is equivalent to amplitude limiting in amplitude-modulation systems, and is particularly effective in reducing the disturbances resulting from impulse noise. Associated with this limiting action is a signal-depressing effect that causes the signal likewise to be reduced in the presence of a large noise. The principal effect of the peak limiting action is hence not to improve the signal-to-noise ratio, but rather to prevent large noise peaks from producing a deafening effect.

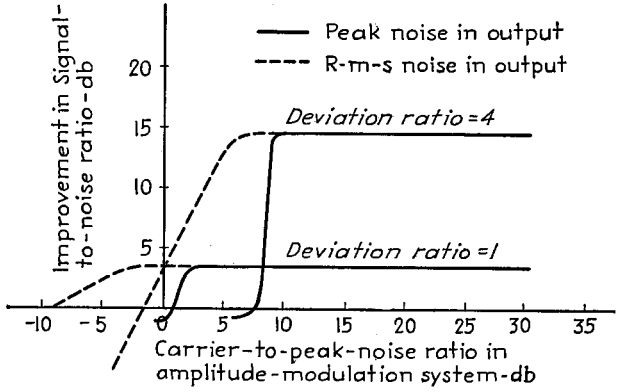


FIG. 39.—Experimental results showing effect of carrier-to-noise ratio upon the improvement in signal-to-noise ratio appearing in the receiver output for the case of frequency modulation as compared with the corresponding amplitude modulation system.

For good performance and best noise reduction, the selectivity curve of the radio-frequency system of the receiver should be symmetrical, and the discriminator should be accurately centered to the middle of the response curve.[1] The best noise reduction cannot be obtained if the selectivity curve is unsymmetrical, even if a limiter is used.

Common-channel Interference and Cross-modulation.[2]—When two frequency-modulated transmitters located within range of a receiver are operated on the same frequency, the particular signal that is the strongest tends to suppress the weaker signal almost entirely and prevent its modulation from appearing in the receiver output. This suppression is almost perfect if the amplitude of the stronger signal is at least twice the weaker signal. The suppression is quite good, however, even with ratios of signal strength much closer to unity. The result is that it is possible to locate frequency-modulation stations much closer to each other geographically than would be possible in the case of amplitude-modulation stations operating on the same frequency.

Analysis of the residual interference that can be produced by two frequency-modulation stations operating on a common channel shows that this interference is of two principal types. First, there is cross-talk, which causes the modulation of the weaker signal to appear in the receiver output even when the stronger is unmodulated.

[1] Vernon D. Landon, Impulse Noise in F-M Reception, *Electronics*, Vol. 14, p. 26, February, 1941.

[2] Harold A. Wheeler, Common-channel Interference between Two Frequency-modulated Signals, *Proc. I.R.E.*, Vol. 30, p. 34, January, 1942; Two-signal Cross Modulation in a Frequency-modulation Receiver, *Proc. I.R.E.*, Vol. 28, p. 537, December, 1940.

The cross-talk interference is very small provided that the desired signal amplitude is at least a small amount larger than the undesired signal amplitude. The second type of common-channel interference corresponds to a beat note between the two signals. The beat-note interference varies in frequency in accordance with the difference in instantaneous frequency deviation of the two signals, and with large frequency swings will be inaudible during a portion of the time. The amplitude of the beat note is proportional at any moment to the instantaneous frequency difference. The beat-note interference gives a sizzling or spitting sound, and is maximum when the two signals have equal amplitudes. The amount by which the desired signal must be stronger than the weaker signal to render beat-note interference unnoticeable or unobjectionable is of the order of 10 to 20 db, and is greater the smaller the frequency swing of the system.[1]

Cross-modulation of a desired frequency-modulated signal by an undesired frequency-modulated signal in an adjacent band will occur only to the extent that the undesired signal possesses amplitude modulation and the receiver is incidentally sensitive to amplitude modulation. It is possible for a form of beat-note interference to exist between signals in adjacent channels. Also, the desired signal may be depressed in amplitude if the undesired signal is strong enough to overload the receiver at the input stage before the radio-frequency selectivity has had an opportunity to reduce the undesired signal to an amplitude less than that of the desired signal. However, adjacent channel interference is much less a problem in frequency-modulation systems than in amplitude-modulation arrangements.

Relative Merit of Wide-band and Narrow-band Frequency-modulation Systems.—Frequency-modulation systems having a high frequency swing (or deviation) should be used when a high signal-to-noise ratio is desired. On the other hand, systems with small frequency deviation, such as 5 to 20 kc (narrow-band systems), are best when a relatively low signal-to-noise ratio is tolerable, and the longest possible range is important.[2]

As a result of this situation wide-band frequency-modulation systems are indicated in the case of high-fidelity broadcast work. On the other hand, in frequency-modulation systems devoted to communication, as, for example, police radio, a relatively narrow-band system gives best results, because the readability of the signal is not seriously impaired by a moderate amount of noise, and range is very important.

Feedback Applied to Frequency-modulation Systems.[3]—If the output voltage of a frequency-modulation receiver is used to frequency-modulate the local oscillator of the receiver with such a phase as to reduce the frequency difference between the local oscillator and the incoming signal, a form of negative feedback is achieved, in which the effective frequency swing of the intermediate-frequency wave is diminished by the feedback action. When the feedback effect is large (*i.e.*, large reduction of frequency swing), the output voltage of the receiver tends to become independent of the amplitude of the incoming signal, of the receiver gain, of the discriminator effectiveness, of the local oscillator voltage, etc. Also, distortion between the converter input and the receiver output is greatly reduced. At the same time the noise-suppression effect obtained is almost as great as in the corresponding wide-band reception of the same signal, even though the negative feedback system makes a wide-band receiver unnecessary to receive a signal with a large frequency swing.

[1] Raymond F. Guy and Robert M. Morris, NBC Frequency-modulation Field Test, *R.C.A. Rev.*, Vol. 5, p. 190, October, 1940.

[2] Murray G. Crosby, Band Width and Readability in Frequency Modulation, *R.C.A. Rev.*, Vol. 5, p. 363, January, 1941; The Service Range of Frequency Modulation, *R.C.A. Rev.*, Vol. 4, p. 349, January, 1940; p. 504, April, 1940; M. L. Levy, Narrow Band vs Wide Band in F-M Reception, *Electronics*, Vol. 13, p. 26, June, 1940.

[3] J. G. Chaffee, The Application of Negative Feedback to Frequency-modulation Systems, *Proc. I.R.E.*, Vol. 27, p. 317, May, 1939.

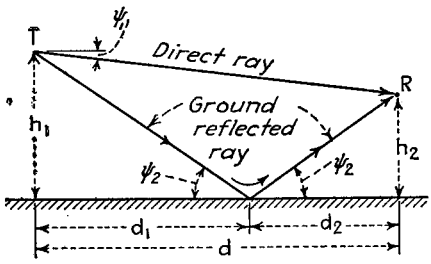
SECTION 10

PROPAGATION OF RADIO WAVES

GROUND-WAVE PROPAGATION AND GROUND EFFECTS

1. Factors Affecting Ground-wave Field Intensity.—The ground wave is here considered to be that portion of the radio wave that is affected by the presence of the ground and that would account for all the energy reaching the receiver if it were not for ionospheric waves (sky waves) and tropospheric waves.

The ground wave can conveniently be divided into two components, a *surface wave* and a *space wave*. The surface wave travels along the surface of the earth. The space wave is the result of two component waves as illustrated in Fig. 1, namely, a direct wave and a ground-reflected wave. When the transmitting and receiving antennas are both at the earth's surface, these two components of the space wave are equal in magnitude and opposite in phase, thereby canceling and leaving the surface wave as the only component of the ground wave. This is the case of "ground-wave" transmission of broadcast frequencies. However, as the antennas are raised the amplitude of the space wave rapidly increases, and it soon becomes the principal part of the ground wave. This is the condition existing at ultra-high frequencies when the antennas are raised a few wave lengths or more above the earth.



$$\tan \psi_2 = \frac{h_1 + h_2}{d} = \frac{h_1}{d_1} = \frac{h_2}{d_2}$$

$$\tan \psi_1 = \frac{h_1 - h_2}{d}$$

FIG. 1.—Diagram showing direct and ground-reflected components of the space wave under circumstances where the curvature of the earth can be neglected. The curvature of the wave paths produced by the variation of refractive index of the earth's atmosphere with height is also neglected.

The magnitude of the ground wave, and of its individual components, the space and surface waves, is influenced by the resistivity and dielectric constant of the earth, the frequency, the height of the transmitting and receiving antennas, the earth's curvature, the distance to the transmitter, and the variation of refractive index of earth's atmosphere with height. The electrical constants of the earth affect the rate of attenuation of the surface wave, and also the reflection coefficient to which the ground-reflected component of the space wave is subjected. The height of the transmitting and receiving antennas affects the relative amplitudes and phases of space and surface waves, and so influences the resulting field. The curvature of the earth makes it necessary for both the space and surface waves to diffract around the earth in order to reach a distant receiving point that is obscured by the curvature of the earth. The variation of density of the earth's atmosphere with height, particularly the variation of moisture content, causes the refractive index of the atmosphere to decrease slowly with height above earth. This causes the paths followed by the waves to be slightly curved, instead of straight as shown in Fig. 1. This curvature

is in the same direction as the curvature of the earth's surface, but is usually somewhat smaller in amount.

2. The Surface Wave.[1]—The surface wave is of importance because it represents the whole of the ground wave when both transmitting and receiving antennas are located at the surface of the earth. It accounts, for example, for the daytime coverage of broadcast stations.

Plane Earth Conditions (Short Distance).—When the distance from the transmitting antenna is not too great the curvature of the earth's surface can be neglected. If the heights of transmitting and receiving antennas are low enough that the respective numerical height q_1 and q_2 as calculated by Eq. (14) satisfy the relation $(q_1 + q_2) < 0.01$, then the surface wave is given by

$$E_{su} = \frac{2E_0}{d} A \tag{1}$$

where E_{su} = field intensity of surface wave in same units as E_0.
d = distance, in same unit of distance as used in E_0.
A = factor taking into account the effect of the earth, given by Fig. 2.
E_0 = a constant determined by the field radiated along the horizontal as discussed below.

[1] The material in Pars. 2 and 3 closely follows K. A. Norton, The Calculation of Ground-wave Field Intensities over a Finitely Conducting Spherical Earth, *Proc. I.R.E.*, Vol. 29, p. 623, December, 1941 (also *F.C.C. Rept.* 39920, Mar. 18, 1940). Alternative presentations are given by Charles R. Burrows and Marion C. Gray, The Effect of the Earth's Curvature on Ground-wave Propagation, *Proc. I.R.E.*, Vol. 29, p. 16, January, 1941; T. L. Eckersley, Ultra-short-wave Refraction and Diffraction, *Jour. I.E.E.*, Vol. 80, p. 286, March, 1937.

A solution for the field produced by a short vertical antenna at the surface of a plane earth of finite conductivity was first obtained by A. Sommerfeld, The Propagation of Waves in Wireless Telegraphy, *Ann. Physik*, Vol. 28, p. 665, March, 1909. Little attention was paid to it, however, until it was used to explain ground-wave attenuation of broadcast waves by R. H. Barfield, The Attenuation of Wireless Waves over Land, *Jour. I.E.E.*, Vol. 66, p. 204, January, 1928. Later analyses of the same problem made by different methods by H. Weyl, A. Sommerfeld, in 1926, Balth. van der Pol and K. F. Niessen, and W. H. Wise differed slightly from Sommerfeld's original solution because of an error in sign in the original derivation discovered by K. A. Norton, The Propagation of Radio Waves over a Plane Earth, *Nature*, Vol. 125, p. 954, June, 1935. Work by Charles R. Burrows, The Surface Wave in Radio Propagation over Plane Earth, *Proc. I.R.E.*, Vol. 25, p. 219, February, 1937, gave experimental verification that the revised analysis was correct, and that a wave of the Zenneck type is not produced at the earth's surface by a vertical radiator, as was indicated by Sommerfeld's 1909 paper.

Those who may wish to study the literature on ground-wave propagation will find the following supplementary bibliography useful, and also complete when references given in these papers are included: Balth. van der Pol and H. Bremmer, The Diffraction of Electro-magnetic Waves from an Electrical Point Source Round a Finitely Conducting Sphere, with Applications to Radio Telegraphy and the Theory of the Rainbow, *Phil. Mag.*, Vol. 24, p. 141, July, 1937, p. 825; Supplement, November, 1937, Vol. 25, p. 817, June, 1938, Vol. 27, p. 261, March, 1939; T. L. Eckersley and G. Millington, Application of the Phase Integral Method to the Analysis of the Diffraction and Refraction of Wireless Waves Round the Earth, *Phil. Trans. Roy. Soc. (London)*, Vol. 237, No. 778, p. 273, June, 1938; The Diffraction of Wireless Waves Round the Earth, *Phil. Mag.*, Vol. 27, p. 517, May, 1939; The Experimental Verification of the Diffraction Analysis of the Relation between Height and Gain for Radio Waves of Medium Length, *Proc. Phys. Soc. London*, Vol. 51, p. 805, September, 1939; Marion C. Gray, Horizontally Polarized Electro-magnetic Waves over a Spherical Earth, *Phil. Mag.*, Vol. 27, p. 421, April, 1939; H. Weyl, The Propagation of Electro-magnetic Waves over a Plane Conductor, *Ann. Physik*, Vol. 60, p. 481, November, 1919; Balth. van der Pol and K. F. Niessen, The Propagation of Electro-magnetic Waves over a Plane Earth, *Ann. Physik*, Vol. 6, p. 273, August, 1930; B. Wwedensky, The Diffractive Propagation of Radio Waves, *Tech. Phys. U.S.S.R.*, Vol. 3, p. 195, November, 1936; W. Howard Wise, The Physical Reality of Zenneck's Surface Wave, *Bell System Tech. Jour.*, Vol. 16, p. 35, January, 1937; Asymptotic Dipole Radiation Formulas, *Bell System Tech. Jour.*, Vol. 8, p. 662, October, 1929; Charles R. Burrows, Radio Propagation over Plane Earth—Field Strength Curves, *Bell System Tech. Jour.*, Vol. 16, p. 45, January, 1937; p. 574, October, 1937; K. A. Norton, The Physical Reality of Space and Surface Waves in the Radiation Field of Radio Antennas, *Proc. I.R.E.*, Vol. 25, p. 1192, September, 1937; The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, *Proc. I.R.E.*, Vol. 24, p. 1367, October, 1936; The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, *Proc. I.R.E.*, Vol. 25, p. 1203, September, 1937; D. O. Rice, Series for the Wave Function of a Radiating Dipole at the Earth's Surface, *Bell System Tech. Jour.*, Vol. 16, p. 101, January, 1937

The constant E_0 is the *free-space* field produced at unit distance from the transmitter in free space with the same antenna currents as are actually present when the antenna is instead near the earth. When the transmitting antenna is vertical, and short compared with $\lambda/4$, so that the field radiated is proportional to the cosine of the angle of elevation, then $2E_0 = 300\sqrt{P}$ millivolts per meter at one km, or $2E_0 = 186.4\sqrt{P}$ millivolts per meter at one mile, where P is the radiated power in kilowatts. With other vertical antenna arrangements the value of $2E_0$ should be modified by the ratio of the field strength produced by the actual antenna for one kilowatt of power, to the field strength produced by the short vertical antenna.

The quantity A takes into account the effect of the losses in the earth upon the surface wave, and depends in a relatively complicated way upon the frequency, dielectric constant, and conductivity of the earth, and the actual distance. By making simplifying assumptions that do not introduce appreciable error under the conditions existing in practical radio communication, this reduction factor can be

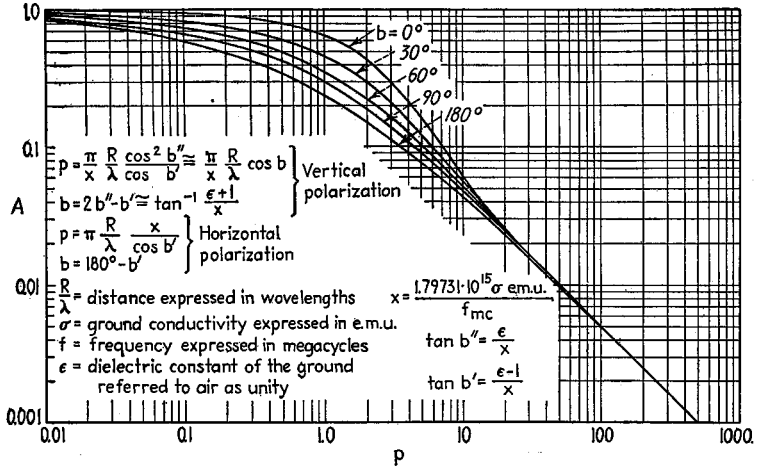


FIG. 2.—Factor A appearing in Eq. (1), which takes into account the effect of ground losses of the surface wave.

expressed in terms of two parameters, the *numerical distance* p and the *phase constant* b, defined as follows:

Vertically polarized waves:

$$p = \frac{\pi}{x}\frac{d}{\lambda}\frac{\cos^2 b''}{\cos b'} \cong \frac{\pi}{x}\frac{d}{\lambda}\cos b \tag{2a}$$

$$b = (2b'' - b') \cong \tan^{-1}\left(\frac{\epsilon + 1}{x}\right) \tag{2b}$$

Horizontally polarized waves:

$$p = \pi\frac{d}{\lambda}\frac{x}{\cos b'} \tag{3a}$$

$$b = 180° - b' \tag{3b}$$

where $x = \dfrac{1.80 \times 10^{15}\sigma}{f_{mc}}$. (4)

$$b' = \tan^{-1}\frac{\epsilon - \cos^2\psi_2}{x} \cong \tan^{-1}\frac{\epsilon - 1}{x}. \tag{5}$$

$$b'' = \tan^{-1}\epsilon/x. \tag{6}$$

f_{mc} = frequency, mc.
σ = earth conductivity in electromagnetic units.
ϵ = dielectric constant of the earth (taking dielectric constant of the air as unity).
λ = wave length in same units as the distance d.
ψ_2 = angle of incidence of ground-reflected wave with earth.

The relationship between p and b and the reduction factor A is given graphically in Fig. 2, which applies to both vertically and horizontally polarized waves if the appropriate definition of p and b is used in each case. The relationship for $b \precsim 90°$ can also be expressed with fair approximation by the empirical relation

$$A \cong \frac{2 + 0.3p}{2 + p + 0.6p^2} - \sqrt{\frac{p}{2}}\,\epsilon^{-\frac{5p}{8}} \sin b \qquad (7)$$

The conditions $b = 0$ for vertical polarization, and $b = 180°$ for horizontal polarization, represent conditions for which the earth can be considered as offering a resist-

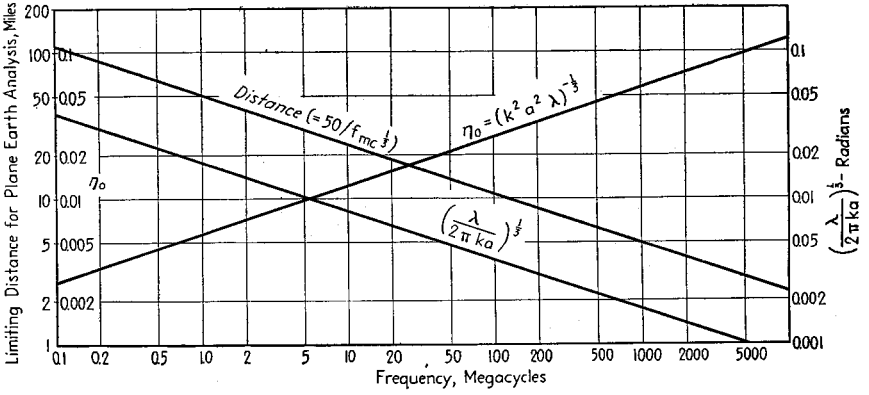


FIG. 3.—Various factors important in the theory of surface-wave propagation.

ance impedance to the flow of the radio-frequency current. The condition $b = 90°$ for either polarization similarly represents the condition where the earth offers a capacitive impedance to the flow of current. It is to be noted that b'' is the power-factor angle of the impedance offered by the earth to the flow of current.

The numerical distance p is proportional to the actual distance, and is a measure of the distance from the transmitter in units that depend upon f^2/σ and the power factor of the earth's impedance. Examination of Fig. 2 and Eq. (7) shows that at large numerical distances ($p > 10$), the factor A is inversely proportional to distance, and has a value approximating $1/2p$. Under these conditions the strength of the surface wave is inversely proportional to the square of the distance.

The assumption of a plane earth involved in Eq. (1) is permissible up to distances of about $50/f_{mc}^{\frac{1}{3}}$ mile. This limiting distance is plotted in Fig. 3.

Strength of Surface Wave at Distances Great Enough so that the Earth's Curvature Cannot Be Neglected.—At large distances from the transmitter the curvature of the earth produces a bulge that requires a bending of the surface-wave path. Some of the required bending is obtained as a result of refraction in the earth's atmosphere (see Par. 4). The remainder of the required bending must be obtained by diffraction. The variation of the strength of the surface wave with distance under these conditions is determined by the wave length, the ground constants, and the effective radius of

the earth (corrected to take into account the refraction in the earth's atmosphere), and can be expressed in terms of x, b', and b'' as defined by Eqs. (4), (5), and (6), and a parameter K given by

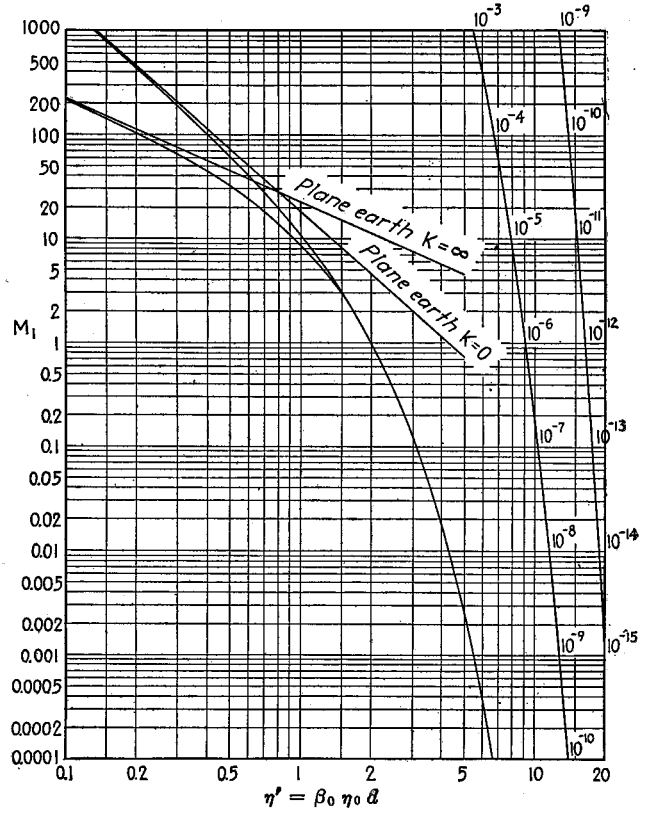


FIG. 4.—Constant M_1 for use in Eq. (10).

Vertical polarization:

$$K = \left(\frac{\lambda}{2\pi ka}\right)^{1/3} \left(\frac{x \cos b'}{\cos^2 b''}\right)^{1/2} \tag{8a}$$

Horizontal polarization:

$$K = \left(\frac{\lambda}{2\pi ka}\right)^{1/3} \left(\frac{\cos b'}{x}\right)^{1/2} \tag{8b}$$

where λ = wave length.

a = earth radius in same units as λ.

k = factor taking into account refractive index of earth's atmosphere.

x, b', and b'' are as above.

The factor k is discussed in Par. 4. Under ordinary conditions it has a value of approximately 1.33, corresponding to ka = 5,280 miles, and

$$\left(\frac{\lambda}{2\pi ka}\right)^{1/3} = \frac{0.0178}{f_{mc}^{1/3}} \tag{9}$$

The relation expressed by Eq. (9) is given in Fig. 3.

The variation of field strength as a function of distance for great distances is

$$\left.\begin{matrix}\text{Field at large}\\ \text{distances}\end{matrix}\right\} = E_{(\eta'=2)}M_1 \tag{10}$$

where M_1 is given by Fig. 4, and represents the relative field as a function of the distance parameter η', and $E_{(\eta'=2)}$ is the strength of the ground wave at a distance d'

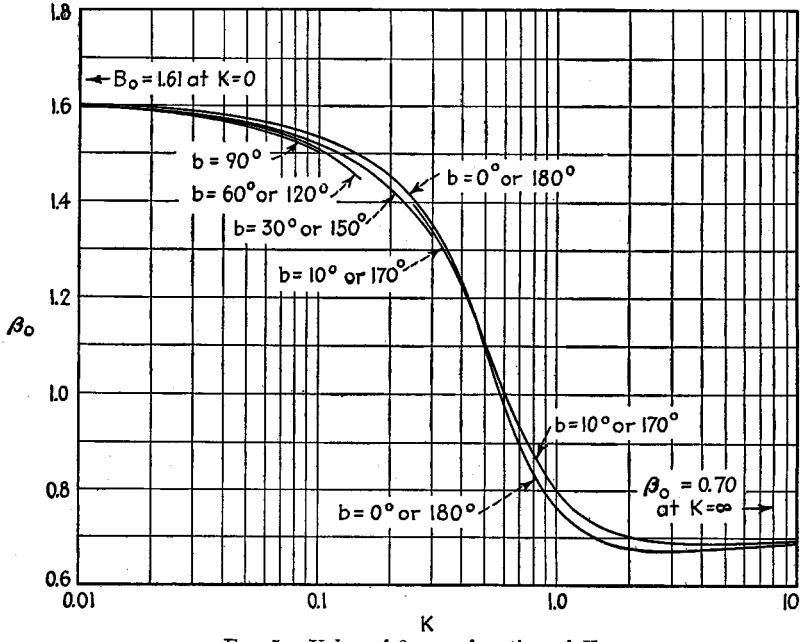


FIG. 5.—Value of β_0 as a function of K.

corresponding to $\eta' = 2$. The factor η' is given by

$$\eta' = \beta_0\eta_0 d \tag{11}$$

and $\eta' = 2$ at a distance d' given by

$$d' = \frac{2}{\beta_0\eta_0} \tag{12}$$

The constant β_0 is given in Fig. 5, and η_0 is given by Fig. 3 when the distance d in Eq. (11) is expressed in miles. The field $E_{(\eta'=2)}$ at the distance $\eta' = 2$ is

$$E_{(\eta'=2)} = 2E_0\eta_0\gamma \tag{13}$$

The parameter γ is given graphically as a function of K and b in Fig. 6.

In making use of Eq. (10) and Fig. 4, it is to be noted that for distances appreciably less than d' (*i.e.*, for $\eta' < 2$) one must interpolate between two values of K, the lower one of which is valid for very large values of K (very low frequencies or very conducting earth) and the upper one valid for small values of K (very high frequencies or poorly conducting earth). This interpolation is made in such a manner as to obtain a smooth transition between the field intensity curve obtained on the assumption of

a flat earth with the aid of Fig. 2, and the single curve for great distances obtained from Fig. 4 for $\eta' > 1.5$.

Summary of Procedure for Calculating Intensity of Surface Wave.—At distances less than given by Fig. 3 the curvature of the earth can be neglected, and the surface wave calculated by Eq. (1) and Fig. 2. At very large distances the field strength and

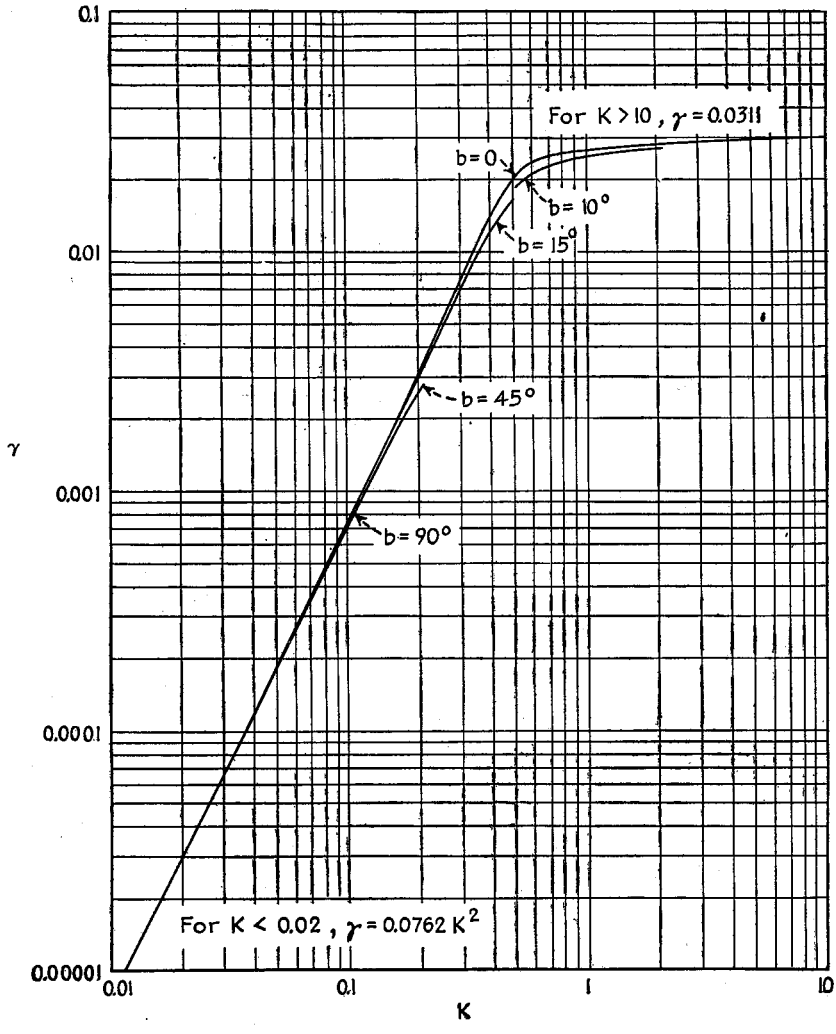


FIG. 6.—Value of the parameter γ.

distance corresponding to $\eta' = 2$ are calculated with the aid of Eqs. (12) and (13), and Eq. (10) and Fig. 4 are used to give the variation in field intensity with distance for distances such that $\eta' > 1.5$. A smooth transition curve is then drawn joining the curves for short and long distances, which transition curve is interpolated between the curves that would be derived with the aid of Fig. 4 for very large and very small values of K.

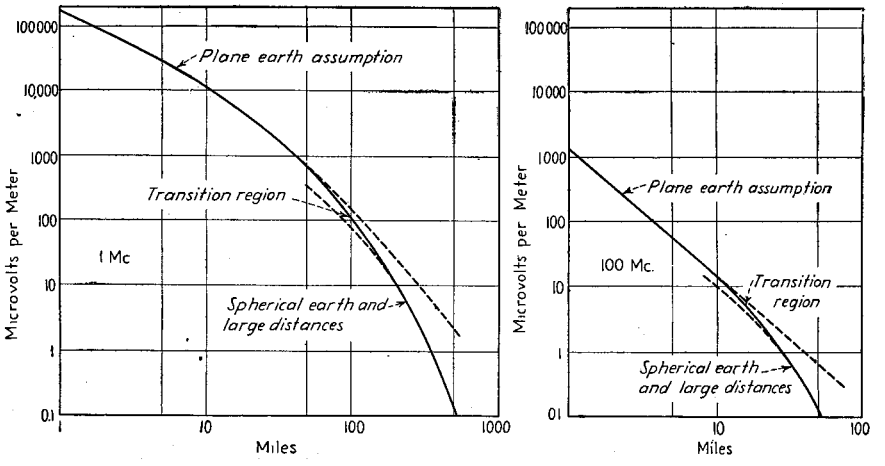


FIG. 7.—Strength of surface waves as a function of distance as determined for two typical cases (1 mc and 100 mc) by methods outlined in Par. 2 (transmitted power 1 kw from short vertical antenna).

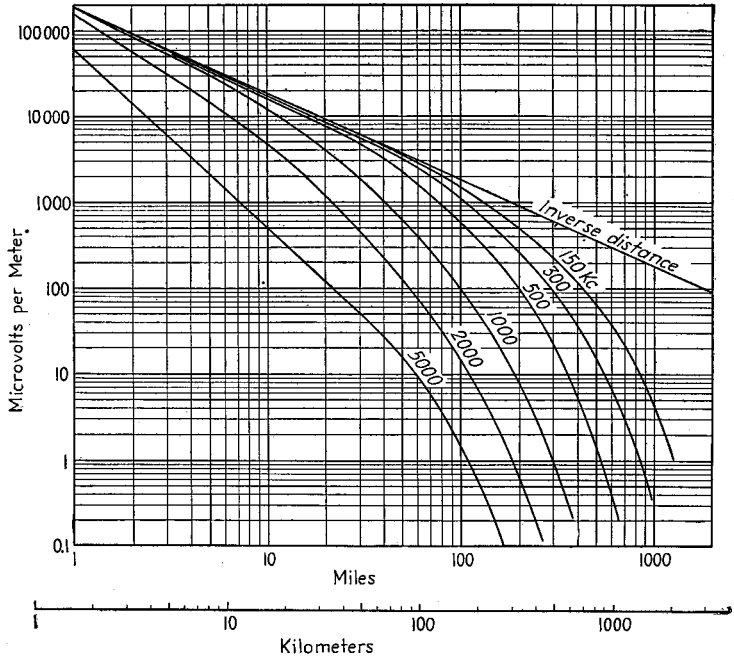


FIG. 8.—Strength of surface waves as a function of distance with a vertical antenna for good earth ($\sigma = 10^{-13}$ emu and $\epsilon = 15$).

Two examples of field-strength curves obtained in this way are given in Fig. 7. These show that at distances such that the curvature of the earth must be taken into account, the field strength dies off much more rapidly with distance than would be the case if the earth were perfectly flat. Also, at higher frequencies the effect of earth curvature first appears at shorter distances, and tends to be more pronounced.

Curves giving surface wave intensity with vertical polarization as a function of distance for frequencies in the range 150 to 5,000 kc are given in Figs. 8, 9, and 10, for good earth, poor earth, and sea water, respectively.[1] Additional curves are given in Figs. 50 and 68.

3. Ground-wave Intensity with Elevated Antennas.[2]—When the transmitting and receiving points are at the surface of the ground the two components that add up to

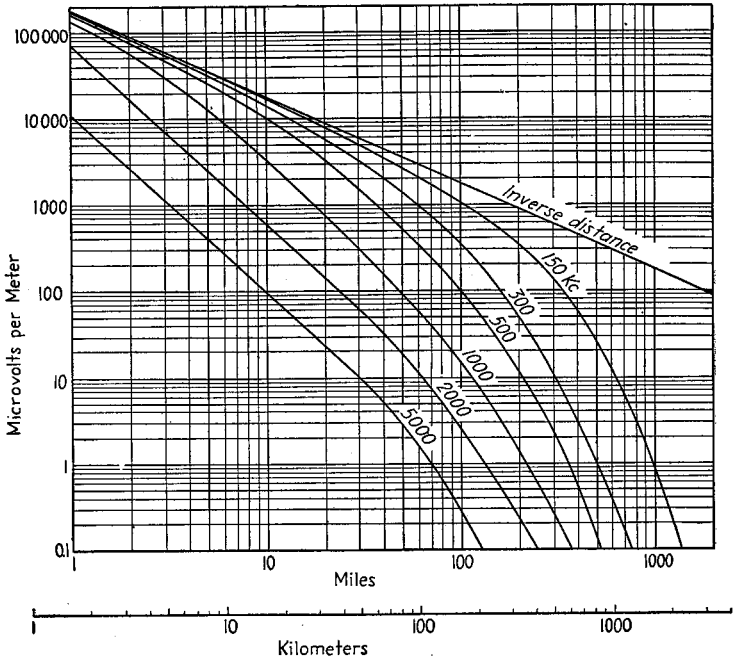


FIG. 9.—Strength of surface waves as a function of distance with a vertical antenna for poor earth ($\sigma = 2 \times 10^{-14}$ emu and $\epsilon = 5$).

give the space wave are of equal magnitude and opposite phase at the surface of the earth, and so give zero resultant space wave. As the height of one or both of the antennas is increased, however, the space wave becomes increasingly strong. For moderate heights the surface and space waves tend to be of comparable magnitude, and the resultant field is the vector sum of the surface and space waves. As the antennas are raised still higher, the space-wave intensity continues to increase until finally one may neglect the surface wave and consider only the space wave.

In problems involving ground-wave propagation with elevated antennas it is convenient to express the antenna heights in terms of a "numerical height" q, which is defined as

[1] Figures 8 and 10 are from Report of Committee on Radio Wave Propagation, *Proc. I.R.E.*, Vol. 26, p. 1193, October, 1938. Figure 9 has been calculated by the methods outlined here.

[2] This presentation follows K. A. Norton, as does Par. 2, and the references given there also apply here.

Vertical polarization:

$$q = \frac{2\pi h}{\lambda}\left(\frac{\cos^2 b''}{x \cos b'}\right)^{1/2} \tag{14a}$$

Horizontal polarization:

$$q = \frac{2\pi h}{\lambda}\left(\frac{x}{\cos b'}\right)^{1/2} \tag{14b}$$

The notation is the same as in Eqs. (4) and (5) with the addition that h/λ is the height measured in wave lengths. The antenna height in wave lengths corresponding to $q = 1$ is given in Fig. 11 as a function of frequency for several typical earth conditions.

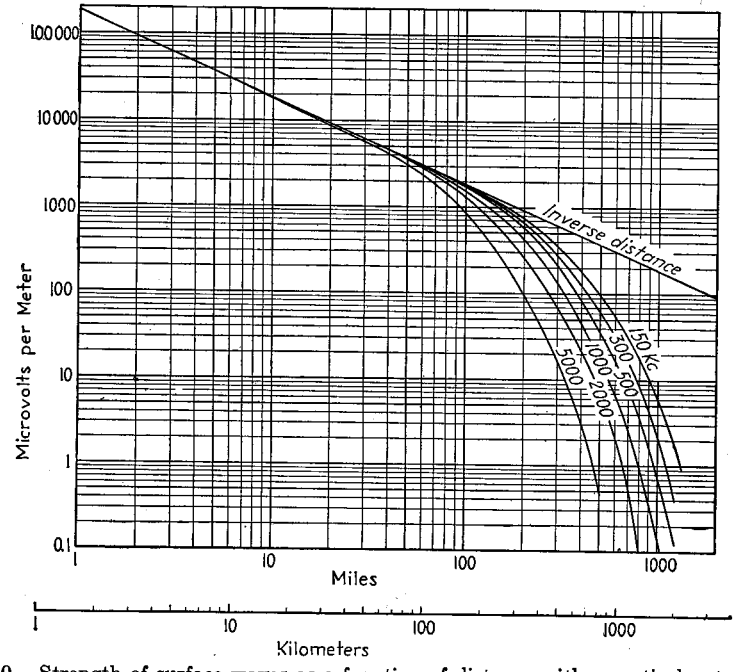


FIG. 10.—Strength of surface waves as a function of distance with a vertical antenna for sea water ($\sigma = 4 \times 10^{-11}$ emu and $\epsilon = 80$).

It will be noted that this height is much less with horizontally than vertically polarized waves, particularly at the lower frequencies.

Considerations Relating to Earth Curvature.—A receiving point is said to be above the line of sight of the transmitter if it is possible for a direct ray to pass from transmitter to receiving point without being intercepted by the bulge in the earth's surface, if the variation of the refractive index of the earth's atmosphere with height is taken into account. As explained in Par. 4, the reduction in dielectric constant with height causes the path of the direct ray to be curved slightly in the same manner as the earth's surface but to a less extent, so that "line-of-sight" condition exists slightly beyond the horizon as determined on the basis of a straight-line path. This effect can be taken into account by assuming that the space waves propagate along straight-line paths, and that the earth has an effective radius slightly larger than its actual radius. Under average conditions this effective radius is approximately 1.33 times the actual radius.

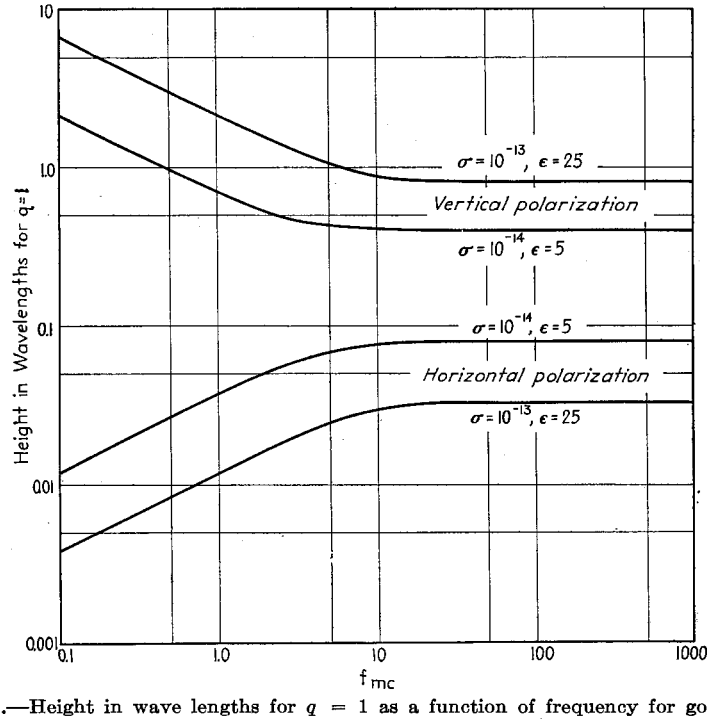


FIG. 11.—Height in wave lengths for $q = 1$ as a function of frequency for good earth ($\sigma = 10^{-13}$ emu) and poor earth ($\sigma = 10^{-14}$ emu).

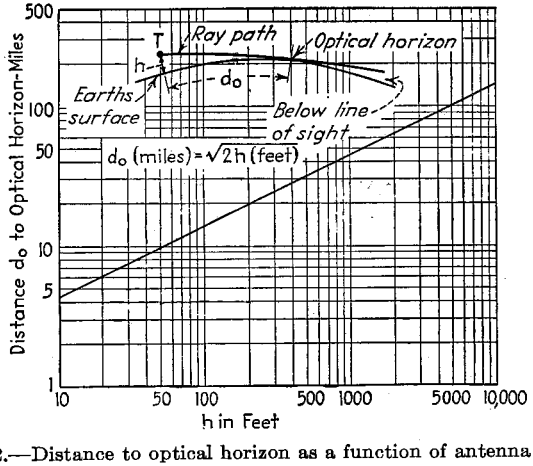


FIG. 12.—Distance to optical horizon as a function of antenna height.

The distance d_0 to the optical horizon with an antenna height h, as illustrated in Fig. 12, is

$$\left.\begin{matrix}\text{Distance to optical}\\ \text{horizon}\end{matrix}\right\} = d_0 = \sqrt{2kah} \tag{15a}$$

where h is the antenna height, a the earth's radius, and k the factor by which the

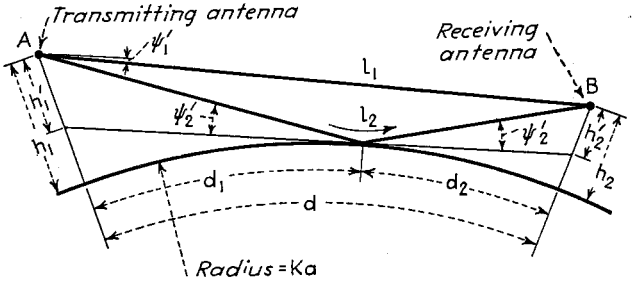


FIG. 13.—Curved-earth geometry.

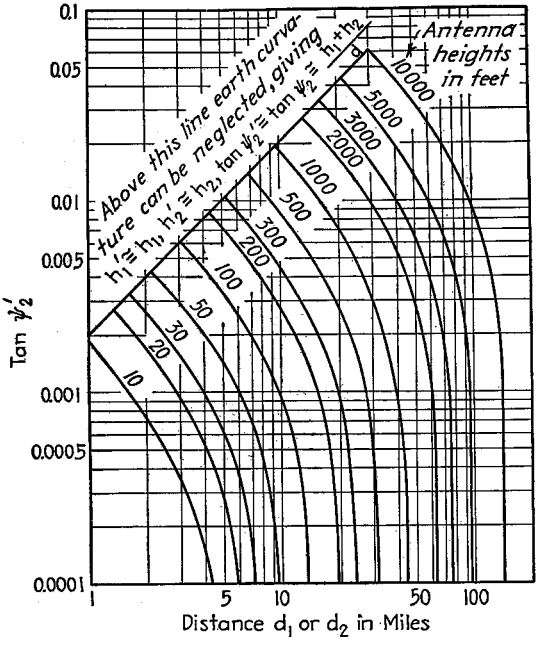


FIG. 14.—Chart for determining d_1 and d_2 when total transmission distance d (see Fig. 13) and the antenna heights are known. The charts are used on a cut-and-try basis by assuming a value of ψ_2' and then determining the corresponding distances d_1, d_2, corresponding to antenna heights h_1 and h_2, and then obtaining $d = d_1 + d_2$. The trial value of ψ_2' is then revised as required until the total distance d comes out to be the actual transmitting distance. These curves assume $k = 1.33$.

radius must be increased to take into account refraction by the earth's atmosphere. When d_0 is in miles, h in feet, and k is taken as 1.33, then

$$d_0 = \sqrt{2h} \tag{15b}$$

This is plotted in Fig. 12.

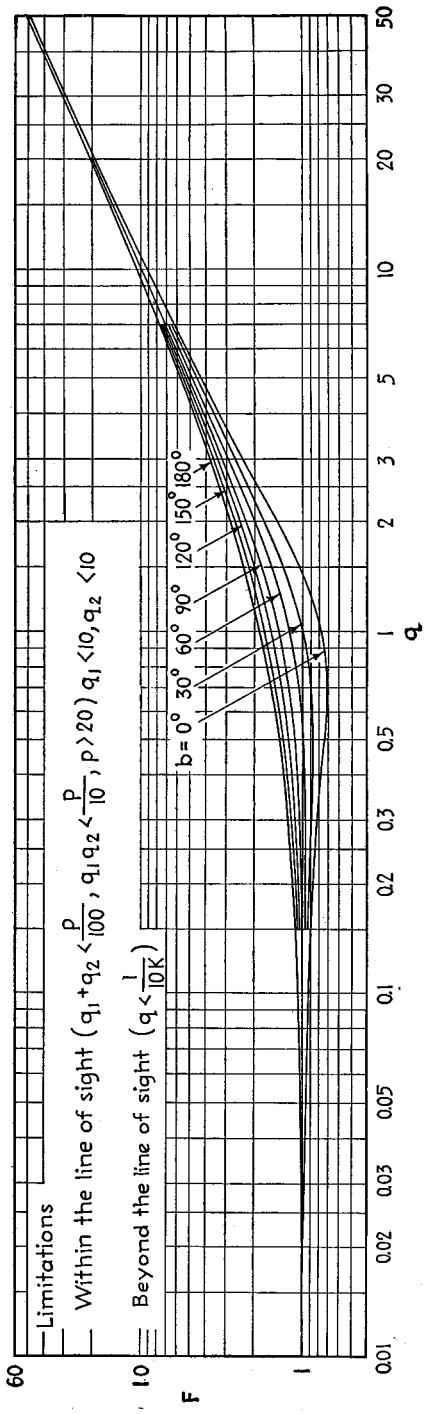


FIG. 15.—Curves for determining values of F factors for use in Eq. (18) for small or moderate values of numerical height within limits indicated.

When transmitting and receiving heights are h_1 and h_2, the maximum distance d_L, over which a wave can propagate between them without being intercepted by the earth, is

$$d_L = \sqrt{2h_1} + \sqrt{2h_2} \qquad (16)$$

When a wave is reflected from the surface of the earth the geometrical relations that exist are shown in Fig. 13, with the point of reflection at the earth so located that the angle of incidence ψ'_2 equals the angle of reflection, also designated by ψ'_2 in the figure.

Relations existing in Fig. 13 are (assuming $h < 20{,}000$ feet)

$$h'_1 = h_1 - \frac{d_1^2}{2ka} \qquad (17a)$$

$$h'_2 = h_2 - \frac{d_2^2}{2ka} \qquad (17b)$$

$$\tan \psi'_2 = \frac{h'_1 + h'_2}{d} = \frac{h'_1}{d_1} = \frac{h'_2}{d_2} \qquad (17c)$$

$$\tan \psi'_1 = \frac{h'_1 - h'_2}{d} \qquad (17d)$$

The notation is the same as previously used and as illustrated in Fig. 13. The value of ψ'_2 and d_1 (or d_2) corresponding to a given total distance $d = d_1 + d_2$ and antenna heights h_1 and h_2 can be easily obtained with the aid of Fig. 14, the use of which is explained in the legend of the figure. In practical calculations such as these, it is customary to take $k = 1.33$, so that $ka = 5{,}280$ miles.

Effect of Antenna Height.—With transmitting and receiving antenna heights h_1 and h_2 such that the corresponding numerical heights q_1 and q_2 as calculated by Eq. (14) cause $(q_1 + q_2) > 0.01$, then the field strength will differ from the intensity of the surface wave.

The general formula for ground-wave field strength with elevated antennas is so complicated that it is necessary to consider only a limited range of conditions at a time. The majority of situations involving elevated antennas can be calculated from the

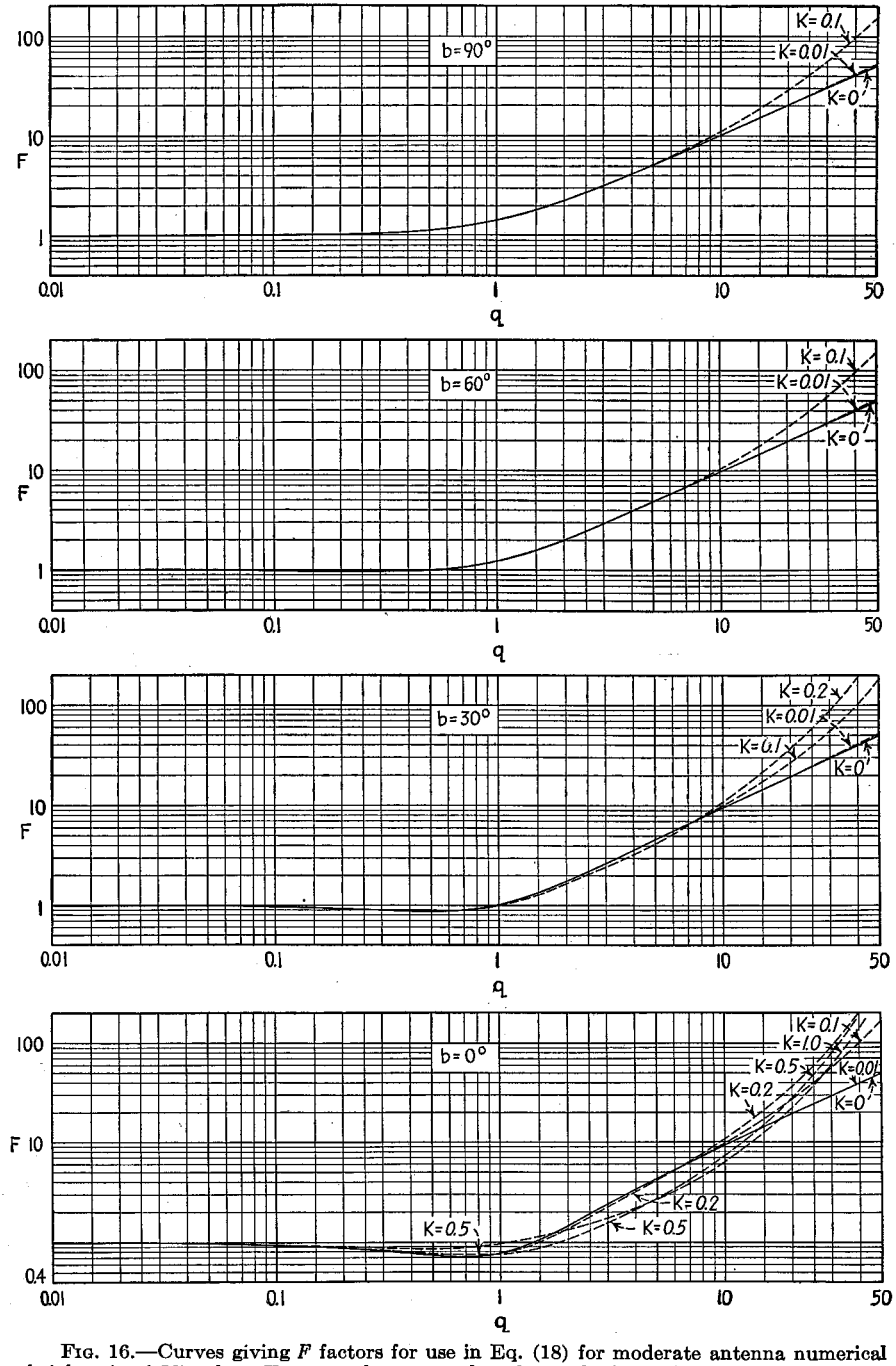


FIG. 16.—Curves giving F factors for use in Eq. (18) for moderate antenna numerical heights ($q < 50$) when K cannot be assumed to be negligibly small. (The curves for $K = 0$ correspond to Fig. 15.)

following relation:

$$E = F_1 F_2 E_{su} \tag{18}$$

where E is the field strength at the receiving antenna, E_{su} the surface-wave intensity at the point on the earth's surface below the receiving antenna as calculated by Par. 2, and F_1 and F_2 are height factors for transmitting and receiving antennas, respectively.

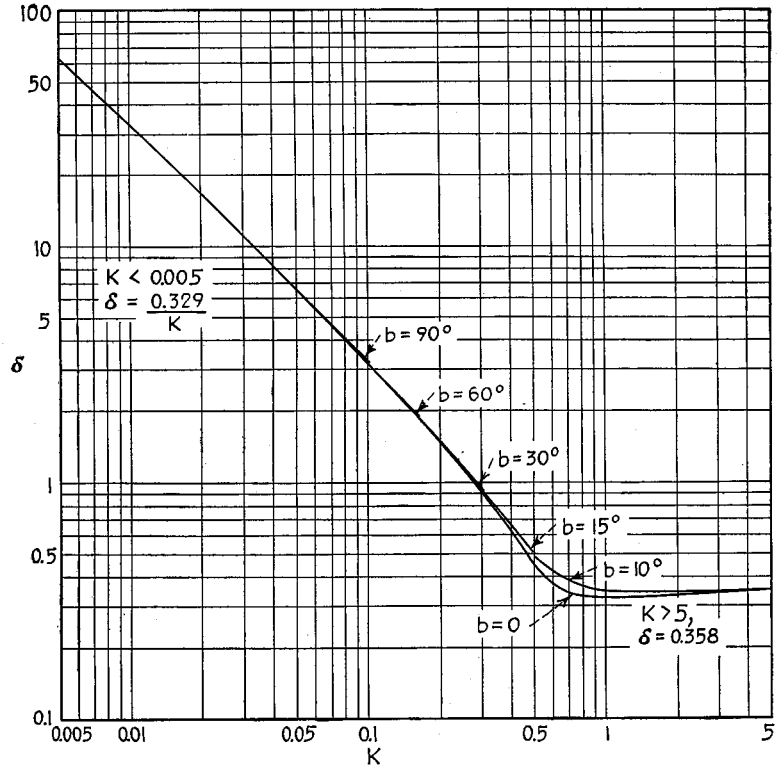


FIG. 17.—Value of δ for use in Eq. (20).

Equation (18) can be used only at distances great enough in relation to the antenna height, and also great enough in absolute magnitude, to satisfy simultaneously all the conditions

$$(q_1 + q_2) < \frac{p}{100}, \qquad q_1 q_2 < \frac{p}{10}, \qquad p > 20 \tag{19}$$

where q_1 and q_2 are the numerical heights of transmitting and receiving antennas, respectively, as defined in Eq. (14) and p is the numerical distance as defined in Eq. (2a) or (3a).

Subject to the limitations imposed by requirements (19), the factors F_1 and F_2 are given by Fig. 15 when the curvature of the earth can be neglected. This limiting distance is about $(50/f_{mc}^{1/3})$ miles, and is given in Fig. 3. The factors obtained from Fig. 15 also apply for distances up to the optical horizon, even when the earth curvature cannot be neglected, provided that $q_1 < 10$ and $q_2 < 10$, and also can be used for receiving points below the line of sight when $q_1 < 1/10K$ and $q_2 < 1/10K$, where K is defined by Eq. (8).

When the receiving point is *below* the line of sight, and the antenna height is a moderate value, $q < 50$, then the F factors are given by Fig. 16. For still greater antenna heights (but below the line of sight) one has

$$F = \delta f_1(H) \tag{20}$$

where δ is given graphically in Fig. 17 and $f_1(H)$ by Fig. 18. The function H is defined by

$$H = \frac{h\beta_0^2}{(ka\lambda^2)^{1/3}} \tag{21a}$$

The height $h_{H=1}$ in feet corresponding to $H = 1$ is

$$h_{H=1} = \frac{30{,}000}{\beta_0^2 f_{mc}^{2/3}} \tag{21b}$$

The notation in Eqs. (21) is the same as previously used. These below-line-of-sight calculations are accurate when the distance exceeds $d_L + 1.5/\beta_0\eta_0$ [notation same as in Eqs. (12) and (16)].

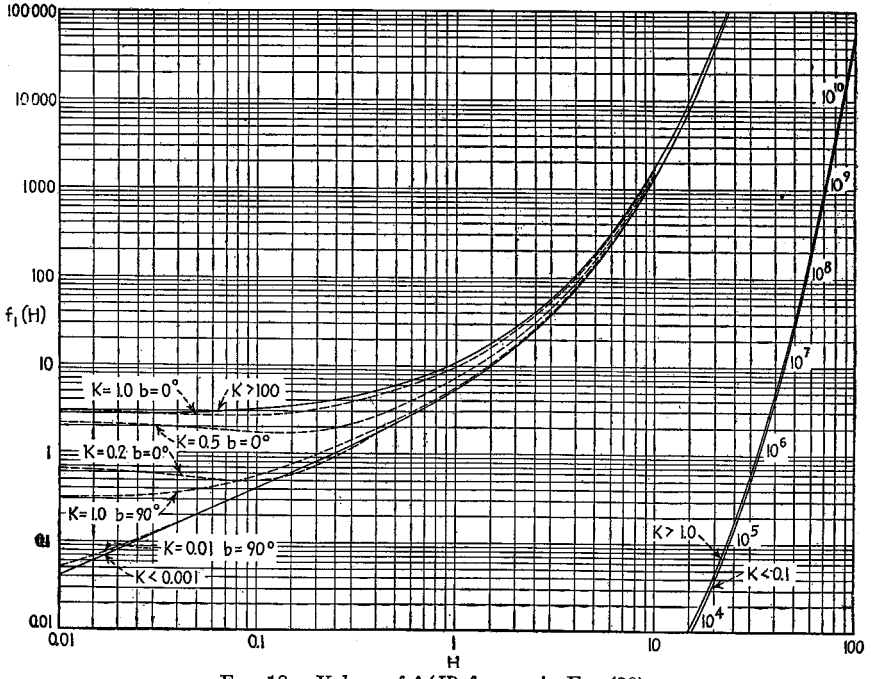


FIG. 18.—Values of $f_1(H)$ for use in Eq. (20).

The only cases where Eq. (18) cannot be used to calculate field strength are (1) when the distance is so small or the height so great that conditions (19) are not satisfied; (2) where conditions (19) are satisfied and the receiving point is above the line of sight, but the height of one (but not both) of the antennas is great enough so that $q > 10$ and at the same time the earth curvature cannot be neglected; (3) the receiving point is at a considerable height above earth and at the same time below the line of

sight, but not far enough away for the transmission distance to exceed $d_L + 1.5/\beta_{07\%}$ as defined above. The general solution for Case 1 can be obtained by calculating the magnitude and phase of surface and space waves independently, as described by K. A. Norton in his 1941 *I.R.E.* paper (or *F.C.C. Report* 39920), and then combining these two components of the ground wave, taking into account their phase relations. In the special case where the numerical heights q_1 and q_2 of both transmitting and receiving antennas exceed about 3, the surface wave may be neglected. The field strength can then be calculated by methods described below, even though the distance is small.

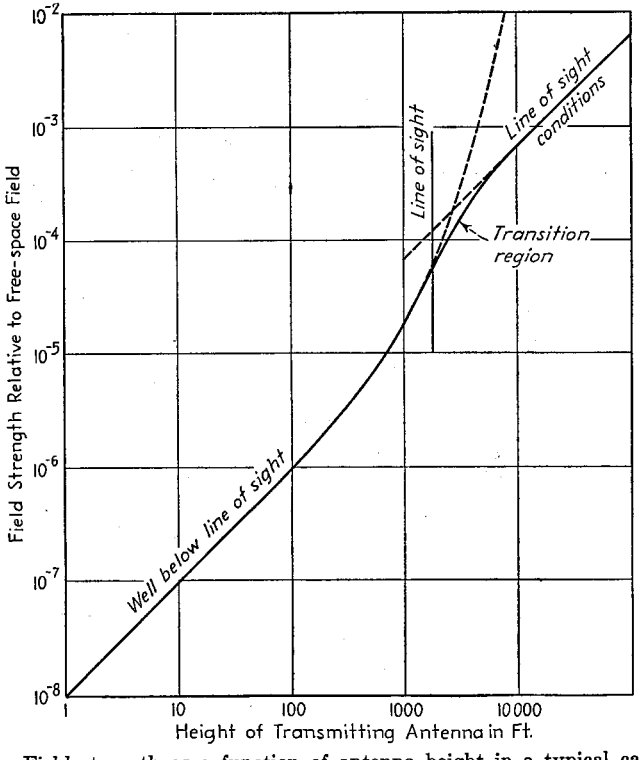


FIG. 19.—Field strength as a function of antenna height in a typical case, showing method by which the field strength at points near the line of sight is obtained by drawing a smooth transition curve between a calculated curve for points well above the line of sight and a calculated curve for points well below the line of sight.

In Case 2, when one antenna is at the surface of the earth, one can obtain at least an approximate result by neglecting the earth curvature and using Eq. (18) and Fig. 16, and assuming $K = 0$; this yields a field smaller than the true value. If one height is low, but not zero, an approximate result can also be obtained by neglecting the surface wave as discussed below. The limitations introduced by (3) can be eliminated by drawing a smooth transition curve between the field strength calculated for line-of-sight conditions and the field strength calculated when the receiving antenna is well below the line of sight, as shown in Fig. 19.

Space-wave Calculations with Receiving Antenna above the Line of Sight.—A large percentage of the conditions encountered in ultra-high-frequency propagation represent propagation to a receiving point above the line of sight, and under conditions

where the received field is almost entirely a space wave. The space wave and surface waves will have equal intensity when the height of the lower antenna corresponds to a numerical height q somewhere in the range $1 < q < 2.5$. Accordingly, calculations of field strength based on space-wave calculations and ignoring the surface wave entirely will not be greatly in error until the numerical height of one or both of the antennas is less than 1.5. At heights below about $q = 1.5$ the field strength, if both space and surface waves are taken into account, will not be greatly different from that calculated for $q = 1.5$ on the basis of the space wave alone.

Space-wave calculations are particularly important in the case of horizontally polarized waves at ultra-high frequencies. This is because, with horizontal polarization, the numerical height corresponding to a given physical height is much greater than with vertical polarization, and also the numerical distance corresponding to a given physical distance is likewise much greater. The maximum antenna height at which the surface wave must be taken into account is hence absurdly small with horizontally polarized waves; *i.e.*, the surface wave can be neglected under nearly all practical conditions with such waves.

The general formula for calculating the intensity of the space wave at an elevated receiving point B produced by an elevated transmitter located at A under conditions where B is above the line of sight with respect to A (see Fig. 13) is

$$\text{Space wave} = \frac{E_0}{d}\cos^3\psi_1'\left[1 \pm \frac{\cos^3\psi_2'}{\cos^3\psi_1'}Dr/\underline{\rho - \theta}\right]$$

$$= \frac{E_0}{d}\cos^3\psi_1'\left[1 + \left(\frac{\cos^3\psi_2'}{\cos^3\psi_1'}Dr\right)^2 \pm 2\frac{\cos^3\psi_2'}{\cos^3\psi_1'}Dr\cos(\rho - \theta)\right]^{1/2} \qquad (22)$$

where E_0 = field that would be produced at unit distance if the antenna were in free space and the currents were the same as are actually present.

$$D = \left(\frac{1}{1 + \dfrac{2h_1'h_2'}{kad\tan^3\psi_2'}}\right)^{1/2} = \text{a factor taking into account divergence of reflected wave produced as a result of reflection from the spherically curved surface of the earth.}$$

r = magnitude of reflection coefficient as calculated by Eq. (27).

ρ = phase angle of reflection coefficient as calculated by Eq. (27).

$\theta = 2\pi\dfrac{l_2 - l_1}{\lambda}$ = phase difference between direct and ground-reflected wave due to differences in path lengths.

λ = wave length.

$l_1 = \sqrt{d^2 + (h_1' - h_2')^2}$ = path length of direct wave.

$l_2 = \sqrt{d^2 + (h_1' + h_2')^2}$ = path length of ground-reflected wave.

ψ_1', ψ_2', d, h_1', h_2', l_1, and l_2 are as shown in Fig. 13, and as used above.

The + sign is used for vertically polarized waves (where $\rho \sim 180°$), while the − sign applies to horizontal polarization (where $\rho \sim 0°$). The same unit of length must be used throughout. Equation (22) applies whenever the receiving point is above the line of sight, and takes into account the earth's conductivity, dielectric constant, and curvature.

When the height becomes small enough in proportion to distance so that

$$\frac{(h_1' + h_2')}{d} < 0.1,$$

Eq. (22) can be written

$$\text{Space wave} = \frac{E_0}{d}\left[1 \pm Dr\underline{/\rho - \frac{4\pi h_1' h_2'}{\lambda d}}\right]^{1/2}$$

$$= \frac{E_0}{d}\left[1 + (Dr)^2 \pm 2Dr\cos\left(\rho - \frac{4\pi h_1' h_2'}{\lambda d}\right)\right]^{1/2} \tag{23}$$

The + sign is used with vertical polarization, the − sign with horizontal. This equation holds for values of ψ_2' less than 0.1, and greater than $(\lambda/2\pi ka)^{1/3}$ as plotted in Fig. 3. When the angle ψ_2' with which the ground-reflected wave strikes the earth is less than $(\lambda/2\pi ka)^{1/3}$, the laws of geometrical optics fail to hold, and Eq. (22) or (23) can no longer be used to calculate the strength of the space wave.

The conductivity and dielectric constant of the earth affect the intensity of the space wave by causing the reflection coefficient to have a magnitude and phase angle different from that existing for a perfect reflector (earth with infinite conductivity). The magnitude and phase of the reflection coefficient also depends upon angle of incidence and upon the frequency, as discussed in Par. 5.

The curvature of the earth produces two effects on the space wave when the receiving point is above the line of sight. First, it makes the effective height h_1' and h_2' of the transmitting and receiving antennas that must be used in Eq. (22) or (23) less than the actual heights of these antennas, as is illustrated in Fig. 13. This effect tends to make the actual received field less than the field that would be calculated for the same antenna height on the assumption of a plane earth. Second, the wave reflected from the slightly spherical ground diverges more than would be the case if the reflection took place from a plane surface. This effect is taken into account by the factor D that appears in Eqs. (22) and (23). This divergence of the wave upon reflection from the curved earth reduces the intensity of the ground-reflected wave that arrives at the receiving point and is equivalent to reducing the magnitude of the reflection coefficient. This reduces the extent to which the ground reflected wave can cancel the direct wave and therefore tends to increase the strength of the received field.

The two effects produced by earth curvature hence tend to oppose each other, with the result that the assumption of a plane earth will give valid results even when considerable curvature is present. The exact limit where calculations based on a plane earth fail to hold depends both on the absolute and relative antenna heights, and with antennas of moderate heights approaches the condition where the receiver is barely above the line of sight.

When the angle ψ_2' with which the ground-reflected wave strikes a plane earth is nearly glancing, the reflection coefficient can be considered as having the value $\underline{1/0°}$ and $\underline{1/180°}$ for horizontally and vertically polarized waves, respectively, and Eq. (23) can then be written as

$$\text{Space wave} = E_0\frac{4\pi h_1 h_2}{\lambda d^2} \tag{24a}$$

Here h_1 and h_2 are the actual antenna heights as in Fig. 13. This special case of the general Eq. (22) is valid when the following conditions are simultaneously met:

1. Distance sufficient for numerical distance $p > 20$.
2. Distance great enough in proportion to height so that $p > 100\,(q_1 + q_2)$.
3. Antenna heights great enough so that $(q_1 + q_2) < q_1 q_2/2$

4. Earth curvature can be neglected.
5. $2\pi h_1 h_2/\lambda d << 1$.

Limitation (5) can be removed by using the following slightly more complicated relation in place of Eq. (24*a*).

$$\text{Space wave} = \frac{2E_0}{d} \sin\left(2\pi \frac{h_1 h_2}{\lambda d}\right) \tag{24b}$$

Equation (24*a*) or Eq. (24*b*) can be used in practical cases up to distances of about $32.5/f_{mc}^{1/3}$ miles, and can be used up to antenna heights of about $2{,}000/f_{mc}^{2/3}$ feet subject to limitations (1) and (2) above. Note that since the curvature of the earth introduces two effects that act in opposite directions, Eqs. (24*a*) and (24*b*) can be used where there is a moderate amount of earth curvature.

The simple relation represented by Eq. (24*a*) or (24*b*) is very widely used in calculating field strength at ultra-high frequencies, since a large proportion of transmission conditions come within its region of validity. This is particularly so in the case of horizontally polarized waves where the formula gives the total field unless the antenna height is almost infinitesimally small, and holds up to points quite close to the transmitter. With horizontally polarized waves requirement (4) is usually the limiting factor.

Discussion of Effect of Antenna Height.—The full variety of effects that can be produced by elevating the antennas appears only in ultra-high-frequency cases. This is because it is only at such frequencies that it is practicable to obtain heights of many wave lengths and also heights great enough so that the space wave can be stronger than the ground wave with vertical polarization.

Consider ultra-high-frequency horizontally polarized waves. When the heights of both of the antennas involved are not negligible compared with the distance between them, as might occur, for example, in communication between airplanes, the received field consists of a relatively strong direct (or free-space) ray upon which is superposed a ground-reflected wave that is much smaller in amplitude because of the greater distance traveled by the ground-reflected wave (*i.e.*, $\psi_2' >> \psi_1'$ in Eq. (22)). The phase with which these two waves combine will depend upon the transmission distance as compared with the antenna heights, and will alternately be additive and subtractive. The amplitude of the received field accordingly oscillates about the free-space field with increasing distance as shown in Fig. 20. As this distance of transmission increases, the length of the direct and indirect paths becomes more nearly equal, so that the minima in the oscillations tend to become progressively deeper. At the same time, however, the increase in distance required to produce a complete oscillation becomes greater because with the changed geometry that goes with greater distance the relative phase of the direct and ground-reflected waves is less sensitive to increments in distance. Finally, at large distances such that the ground-reflected wave strikes the earth with almost grazing incidence, the two waves are almost in phase opposition at the receiving point. The resultant field is then much less than the free-space field, and drops off rapidly with further increase in distance, and without showing any further tendency to oscillate.

The situation with vertically polarized waves is very similar to that for horizontally polarized waves except that, since with such waves the magnitude and phase of the reflection coefficient is much more sensitive to angle to incidence with the earth than is the case with horizontally polarized waves, the amplitude of the oscillations in field strength is affected by an additional factor. Also, when the distance is so related to the height that the angle of incidence with which the ground-reflected wave strikes the earth approximates the pseudo-Brewster angle, the amplitude of oscillation is quite small because the ground-reflected wave is largely destroyed. Under such conditions the field tends to approach the free-space value.

If one antenna is at a fixed moderate height, and the height of the other is then varied, the effects produced at very high frequencies, assuming that the receiving point is well below the line of sight, are as shown in Fig. 21*a*. With vertically

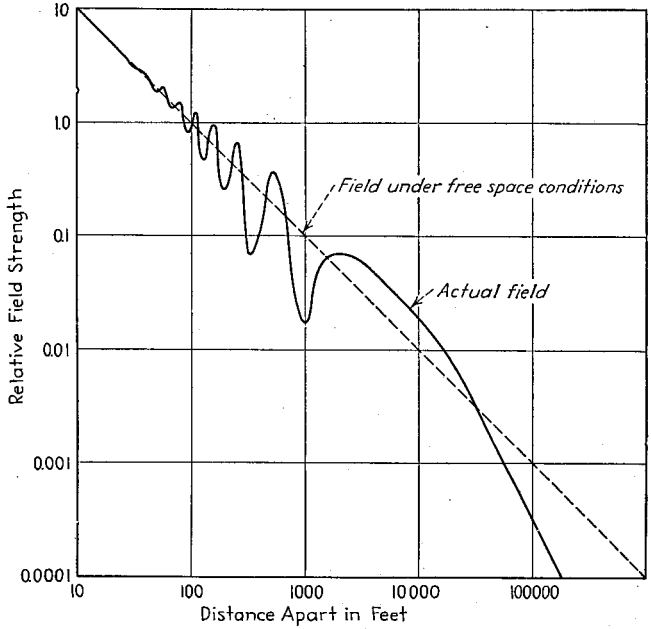


FIG. 20.—Curve of field strength as a function of distance in a typical case involving equally elevated antennas at ultra-high frequencies.

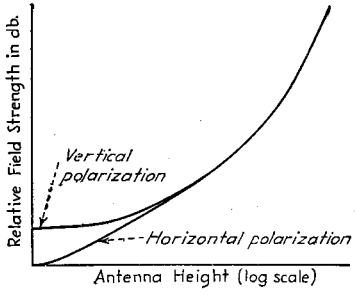


FIG. 21*a*.

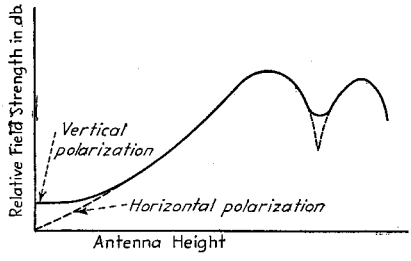


FIG. 21*b*.

FIG. 21*a*.—Variation of field strength with antenna height for vertically and horizontally polarized waves when the antenna is well below the line of sight.

FIG. 21*b*.—Variation of field strength with antenna height for vertically and horizontally polarized waves when the antenna heights are not negligible compared with the transmission distance.

polarized waves one finds that as the antenna is raised starting first from the ground, the field strength is first substantially constant, and then increases in strength, first slowly, then in proportion to height, and then exponentially with height. In contrast with this, the field obtained with horizontally polarized waves is almost negligible even very close to the earth, and increases in direct proportion to height

until it merges at a moderate height with the curve for vertical polarization. The reason for this difference with polarization is that the surface wave is much weaker with horizontal polarization than it is with vertical polarization. Thus, in the case of Fig. 21*a*, the space wave is not appreciably greater than the surface wave in the case of vertical polarization until the antenna height is about $\lambda/2$. When the transmission distance is not too great, then though the behavior is the same as in Fig 21*a* for small and moderate heights, it takes on an oscillatory character for large heights, as shown in Fig. 21*b*. The oscillations are the result of the ground-reflected wave alternately adding to and subtracting from the direct ray as tne height is varied, in accordance with Eq. (24*b*). Further discussion of ground-wave characteristics at ultra-high frequencies is given in Par. 17.

At the lower frequencies, such as 1,000 kc, the situation is much simpler. With vertically polarized waves the antenna heights will never be large when expressed in terms of numerical height. Under most conditions one need then consider only the surface wave, although if the height is unusually great, as might be the case in communication with an airplane, one can apply the height factors derived from Fig. 15 or 16.

In contrast with this, it is found that with horizontally polarized waves of low and moderate frequency, the intensity of the surface wave is negligible even very close to the surface of the ground, and one need consider only the space wave. Under these conditions one can calculate the intensity of the field as a function of height and of distance by the use of Eq. (22), or some of its simpler modifications, such as Eqs. (23) and (24). In dealing with horizontally polarized waves at the lower frequencies, the ground-reflected wave tends to cancel the direct wave in all practical cases; so the resultant field intensity is appreciably less than the free-space field. Also, there are no oscillations in the curve of field intensity as a function of distance, or in the curve of field intensity as a function of height (unless heights such as obtainable only by aircraft are involved).

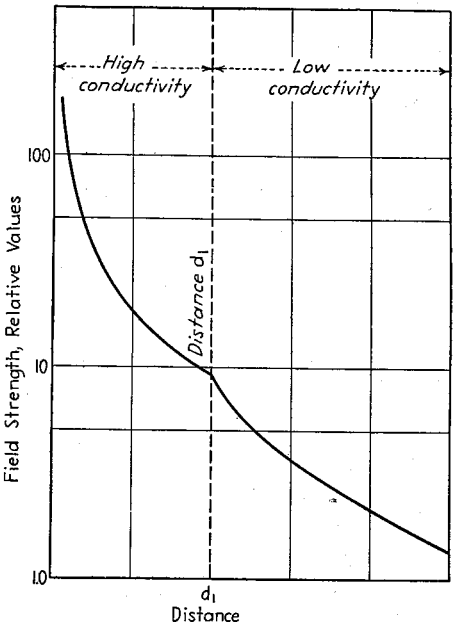


FIG. 22. Attenuation curves for ground waves over a composite path obtained by piecing together curves calculated for two earth conductivities.

4. Miscellaneous Considerations in Ground-wave Propagation. *Composite Paths.* The electrical constants of the earth encountered by the surface wave in traveling away from the transmitter may vary greatly at different places. No accurate method has been devised for determining the surface-wave field intensity under these conditions. Practical methods that have so far been used to handle this situation are of two types: (1) an average of earth constants can be used as determined by propagation tests; (2) attenuation curves for different earth constants are joined. This second method is illustrated in Fig. 22 for the simple case of two earth conditions. Up to the distance d_1 the intensity is calculated in the usual manner. Beyond d_1 the field is assumed to die away at a rate corresponding to the new earth constants, calculated on the basis of the total distance to the transmitter. The absolute magnitude of this

second portion of the curve is adjusted so that it will give a field at the distance d_1 corresponding to the field delivered to this point by the better conducting earth close to the transmitter. The first method of taking into account a composite path is recommended where the transmission path is over earth that varies continuously in its characteristics with no large homogeneous stretches. The second method is recommended where there are large stretches of earth along the transmission path of substantially uniform characteristics, as, for example, where part of the transmission path may be over a flat valley or marshy land with a subsequent portion of considerable length over a foothill region of poorer conductivity.

An important case of composite paths occurs in ship-to-shore communication over moderate distances in which the surface wave is depended upon. Experiments have shown that the strength of the surface wave on the shore immediately adjacent to the water is of the order of 8 or 12 db greater than the strength of the surface wave a mile or more inland.[1] These results are explainable theoretically, at least in part, on the basis that a portion (6 db) of the excessive attenuation that occurs within the first mile is the result of the land losses being sufficient to destroy the optical image of the surface wave that existed over the highly conducting water ground.

Effects Produced by Variations of Dielectric Constant of the Atmosphere with Increase in Height.[2]—The dielectric constant of the atmosphere is slightly greater than the value unity obtained in a vacuum. This is a result of the presence of the gas molecules, and the value depends on the number and kind of molecules per unit volume, and is particularly sensitive to water molecules because of the high dielectric constant of water vapor. The dielectric constant of air can, accordingly, be expected to decrease with height above the earth because of decrease in atmospheric pressure. The actual value, and also the rate of decrease with height, can be expected to vary from time to time as a result of changes in the amount and distribution of water vapor in the atmosphere.

This variation in the dielectric constant of the atmosphere causes radio waves passing through the air to be refracted downward toward the earth in accordance with Eq. (41). This effect can be taken into account by using an effective radius of the earth that is k times the actual radius, and then neglecting the variation in dielectric constant with height. On the assumption that the dielectric constant ϵ of the air changes at a uniform rate $d\epsilon/dh$, then

$$k = \frac{1}{1 + \frac{a}{2}\frac{d\epsilon}{dh}} \tag{25}$$

where a is the actual radius of the earth expressed in the same units as the height h.

Under average conditions the value of k is of the order of 1.33, and this value has been used in all the curves in Pars. 2 and 3. The actual value of k will vary from hour to hour, day to day, and from season to season. Under very unusual conditions it may conceivably become infinity, in which case the refraction is sufficient to enable the wave to follow the earth's curvature, giving propagation conditions that are essentially the same as for a plane earth up to very great distances.

[1] R. A. Heising, Effect of Shore Station Location upon Signals, *Proc. I.R.E.*, Vol. 20, p. 77, January, 1932.

[2] For further information see J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, Ultra-short Wave Propagation, *Proc. I.R.E.*, Vol. 21, p. 427, March, 1933; E. R. Englund, A. B. Crawford, and W. W. Mumford, Further Results of a Study of Ultra-short-wave Transmission Phenomena, *Bell System Tech. Jour.*, Vol. 14, p. 369, July, 1935.

This action was first pointed out by R. Jouaust, Some Details Relative to Propagation of Very Short Waves, *Proc. I.R.E.*, Vol. 19, p. 479, March, 1931.

Wave Tilt and Penetration into the Earth of the Surface Wave.[1]—The electric vector of the wave produced at the surface of the earth by a vertical antenna possesses a slight forward tilt. As a result the electric vector has, in addition to the normal vertical component, a slight horizontal component parallel to the surface of the earth and lying in the direction of propagation. These two components generally have a phase difference, so that the resultant electric vector is elliptically polarized. The magnitude of the forward tilt of the wave is independent of the height of the transmitting antenna.

The magnitude of the wave tilt and the eccentricity of the elliptical polarization depend upon the earth constants, and the frequency, and can be calculated with the aid of Fig. 23.[2] This shows that the maximum tilt that can be encountered in practice is less than 22°, with large values requiring a combination of a very high-frequency

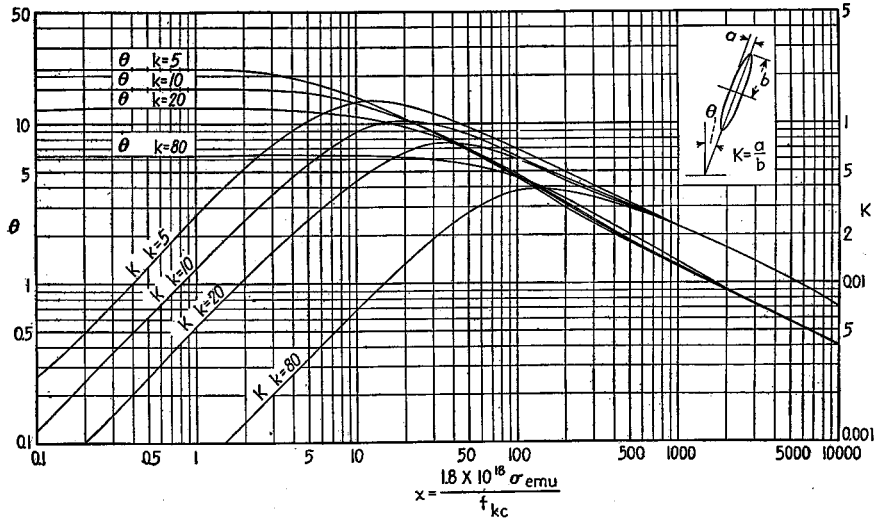


FIG. 23.—Characteristics of wave tilt for different values of dielectric constant k and ground-conductivity factor x.

and a low-conductivity soil. At broadcast frequencies the maximum tilt will always be less than 15°, and it will be still less at lower frequencies. The ratio of horizontal to vertical components of the surface wave is greatest at an intermediate value of frequency as shown in Fig. 23.

The component of the electric vector lying parallel to the earth's surface that is present because of wave tilt represents a voltage applied across the earth's surface. This voltage causes a current to penetrate down into the earth. The character of this current and its penetration can be calculated by assuming that the electric vector parallel to the earth's surface is an applied voltage acting at the sending end of a trans-

[1] For further discussion, including methods of measuring wave tilt, see C. B. Feldman, The Optical Behavior of the Ground for Short Radio Waves, *Proc. I.R.E.*, Vol. 21, p. 764, June, 1933; R. H. Barfield, Some Measurements of the Electrical Constants of the Ground by the Wave-tilt Method, *Jour. I.E.E.*, Vol. 75, p. 214, 1934; K. A. Norton, The Physical Reality of Space and Surface Waves in the Radiation Field of Radio Antennas, *Proc. I.R.E.*, Vol. 25, p. 1192, September, 1937; W. H. Wise, The Physical Reality of Zennech's Surface Wave, *Bell System Tech. Jour.*, Vol. 16, p. 35, January, 1937.

Excellent elementary analysis of the wave tilt and earth penetration from the physical point of view is given by G. W. O. Howe, Wireless Waves at the Earth's Surface, *Wireless Eng.*, Vol. 17, p. 385, September, 1940.

[2] From K. A. Norton, The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, *Proc. I.R.E.*, Vol. 25, p. 1203, September, 1937.

mission line having distributed constants represented by the conductivity and dielectric constant of the earth. As these currents flow into the earth, they are attenuated with distance at a rate determined by the earth's constants and the frequency. The depth of penetration is greater the lower the frequency, the smaller the dielectric constant of the earth, and the lower the earth's conductivity. The rate at which the current density decreases with depth is given by the relation[1]

$$\frac{\text{Current density at depth } d}{\text{Current density at surface}} = \epsilon^{-pd} \qquad (26)$$

where p =

$$\left[\frac{XB}{2}\left(\sqrt{1+\frac{(\sigma\times 10^9)^2}{B^2}}-1\right)\right]^{1/2}$$

d = depth, cm.
$X = 0.008\pi^2 f_{mc}$
$B = 0.556\times 10^{-6} k f_{mc}$
k = dielectric constant of the earth.
f_{mc} = frequency, mc.
σ = earth conductivity, emu.

The current penetrates to a considerable depth. Thus with $k = 5$, and

$$\sigma = 2.8\times 10^{-14} \text{ emu},$$

the depth at which the current is one-fourth of its value at the earth's surface is 6.5 meters (20 feet) at 10 mc, and 14 meters (45 feet) at 1 mc.

The considerable depth to which currents penetrate causes the conductivity and dielectric constant that is effective to radio waves to be determined by the average conditions existing for a considerable distance below the surface of the earth.

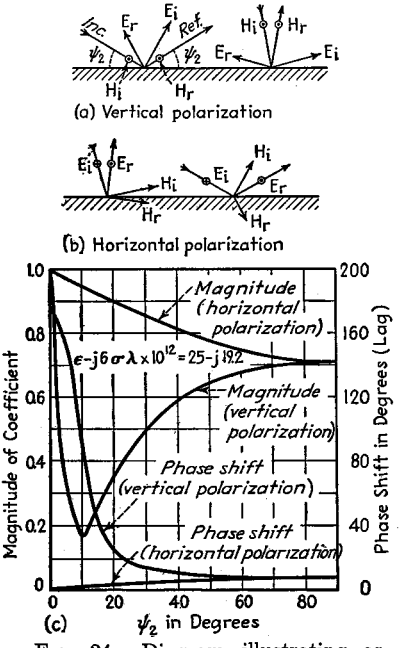


FIG. 24.—Diagram illustrating assumed positive polarities for Eq. (27) for the case of a perfect earth, together with curves showing the reflection coefficient as a function of angle of incidence for an imperfect earth. (E and H denote electric and magnetic flux, respectively, and subscripts i and r denote incident and reflected components.)

5. Reflection of Radio Waves.[2]—When a *plane* wave strikes a surface such as the earth, it is reflected with an angle of reflection equal to the angle of incidence. When the reflecting surface has infinite conductivity, the amplitude of the reflected wave equals the amplitude of the incident wave, and the reflected wave has either exactly the same phase or exactly the opposite phase of the incident wave, depending upon the polarization. With an actual earth the reflected wave tends to be smaller in amplitude than the incident wave, and the phase relations are modified from those existing with an infinitely conducting reflector.

The vector ratio of reflected to incident wave is termed the *reflection coefficient*, and designated as $R = r/\rho$, where r is the ratio of magnitude of reflected and incident waves and ρ is the modification of the phase produced by the imperfections of the earth. The exact value of the reflection coefficient depends upon the dielectric con-

[1] From Howe, *loc. cit.*

[2] For further information see C. B. Feldman, The Optical Behavior of the Ground for Short Radio Waves, *Proc. I.R.E.*, Vol. 21, p. 764, June, 1933; L. C. Verman, Reflection of Radio Waves from the Surface of the Earth, *Proc. I.R.E.*, Vol. 18, p. 1396, August, 1930; G. W. O. Howe, Reflection of Waves at Earth's Surface, *Exp. Wireless and Wireless Eng.*, Vol. 11, p. 59, February, 1934. This last reference presents an especially fine elementary quantitative analysis and discussion of the subject.

stant and conductivity of the earth, the frequency, and the angle of incidence with which the wave strikes the surface of the earth, according to the relations[1]

Vertical polarization (Fig. 24):

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = r/\underline{\rho} = \frac{\epsilon' \sin \psi_2 - \sqrt{\epsilon' - \cos^2 \psi_2}}{\epsilon' \sin \psi_2 + \sqrt{\epsilon' - \cos^2 \psi_2}} \tag{27a}$$

Horizontal polarization (Fig. 24):

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = r/\underline{\rho} = \frac{\sqrt{\epsilon' - \cos^2 \psi_2} - \sin \psi_2}{\sqrt{\epsilon' - \cos^2 \psi_2} + \sin \psi_2} \tag{27b}$$

where ψ_2 = angle of incidence as shown in Fig. 24.

$\epsilon' = \epsilon - j60\sigma\lambda \times 10^{12}$

ϵ = dielectric constant of earth (air taken as unity).

σ = earth conductivity (emu).

λ = wave length, meters.

$j = \sqrt{-1}$.

The angle ρ obtained in Eqs. (27) is the angle by which the phase of the reflected wave leads the phase that it would have in the case of perfect earth ($\sigma = \infty$). The phase relations existing in this ideal case are as shown in Fig. 24. As a result of the fact that the earth's impedance is capacitive the phase angle of reflection ρ is always negative, corresponding to a lag in the reflected wave.

Curves from which the magnitude and phase angle of the reflection coefficient can be readily obtained with a minimum of calculation are given in Fig. 26.[2]

The reflection coefficient varies with the angle of incidence ψ_2 in the manner illustrated in Fig. 24, which is for earth of high conductivity at a wave length of 15 meters. With horizontally polarized waves, the reflection coefficient has a magnitude of unity and a phase shift of zero at glancing angles of incidence, but as the wave approaches vertical incidence, the reflection coefficient drops off gradually in magnitude, and the phase shift ρ gradually becomes more lagging but always has a small value.

In contrast with this, a vertically polarized wave undergoes a reflection in which the magnitude of the reflection coefficient, starting from unity at very small angles of

[1] A more general case of reflection is when a wave traveling in one medium having $\epsilon' = \epsilon_1'$, strike a boundary with another medium for which $\epsilon' = \epsilon_2'$. Under these conditions the reflection coefficient is obtained by substituting ϵ_2'/ϵ_1' for ϵ' in Eqs. (27), with the dielectric constants and conductivities of the corresponding mediums being used in evaluating ϵ_1' and ϵ_2'.

The relations existing in this general case are shown in Fig. 25. The angle of reflection is equal to the angle of incidence ψ, while the angle of refraction ψ' of the transmitted wave is given by Snell's law

$$\frac{\cos \psi'}{\cos \psi} = \frac{\mu_1}{\mu_2} \tag{28}$$

where μ_2 and μ_1 are the indexes of refraction of the respective media. The index of refraction is the real part of $\sqrt{\epsilon'}$.

The magnitude and phase of the transmitted wave must be such that the vector sum of transmitted and reflected waves equals the incident wave.

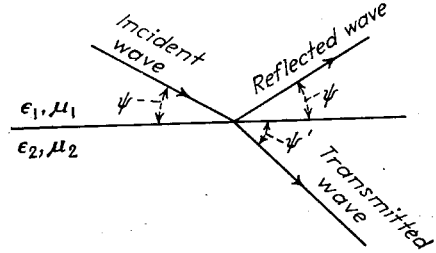
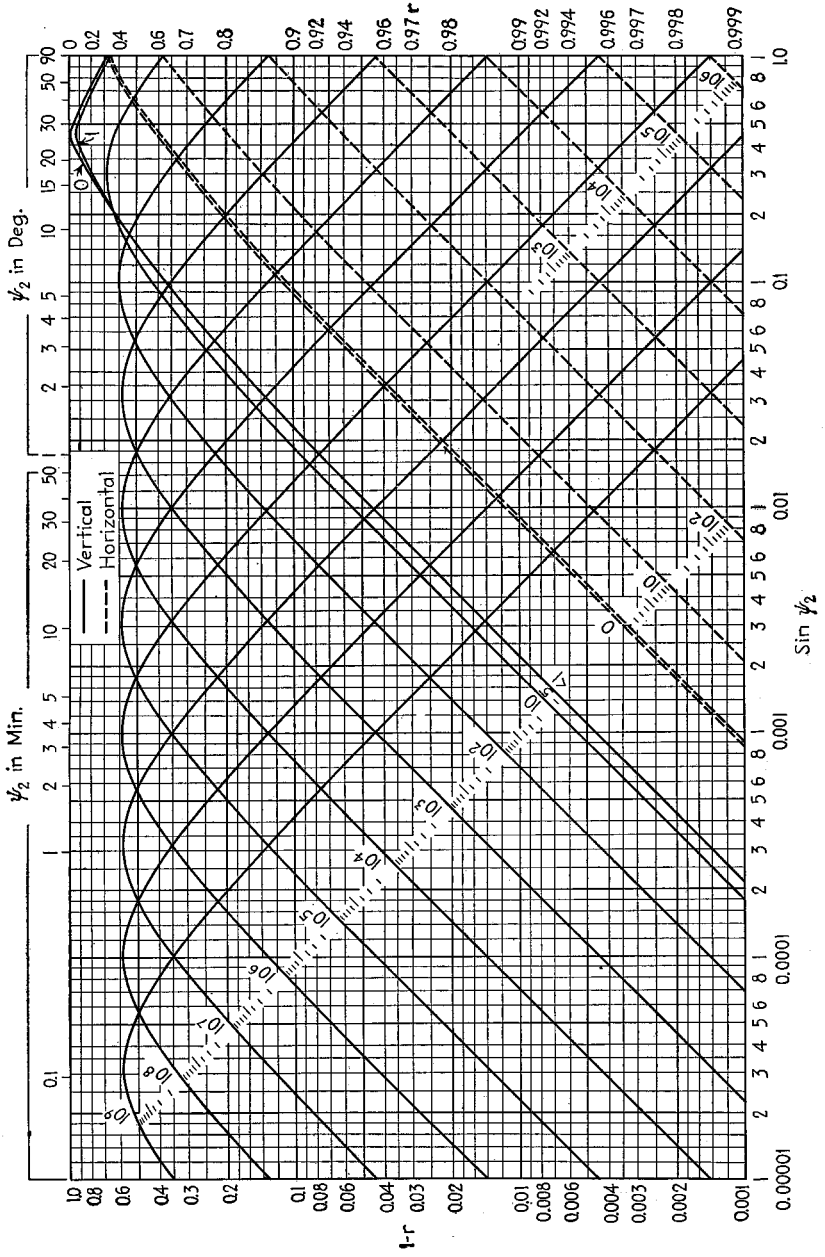


FIG. 25.—Diagram illustrating general case of reflection occurring at the boundary of two mediums having different refractive indexes.

[2] From C. R. Burrows, Radio Propagation over Plane Earth—Field Strength Curves, *Bell System Tech. Jour.*, Vol. 16, p. 45, January, 1937. Equivalent curves are given by J. S. McPetrie, The Reflection Coefficient of the Earth's Surface for Radio Waves, *Jour. I.E.E.*, Vol. 82, p. 214, 1938; Reflection Curves and Propagation Characteristics of Radio Waves along the Earth's Surface, *Jour. I.E.E.*, Vol. 87, p. 135, 1940.


ψ2 in Min.
ψ2 in Deg.
Sin ψ2
1-r
r
Vertical
Horizontal
0
0.2
0.4
0.6
0.7
0.8
0.9
0.92
0.94
0.96
0.97
0.98
0.99
0.992
0.994
0.996
0.997
0.998
0.999
10
15
20
30
50
90
1
2
3
5
0.1
0.01
0.001
0.0001
0.00001
0.8
0.6
0.4
0.3
0.2
0.08
0.06
0.04
0.03
0.02
0.008
0.006
0.004
0.003
0.002
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹

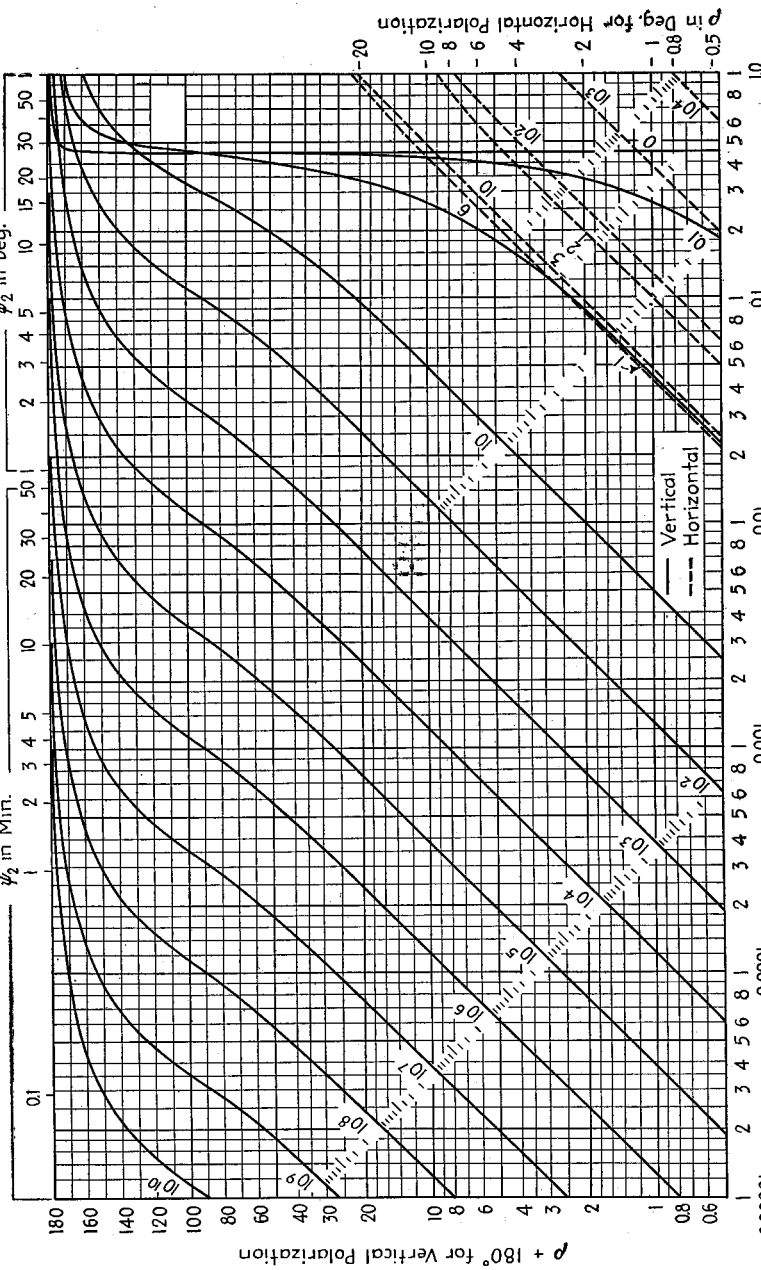
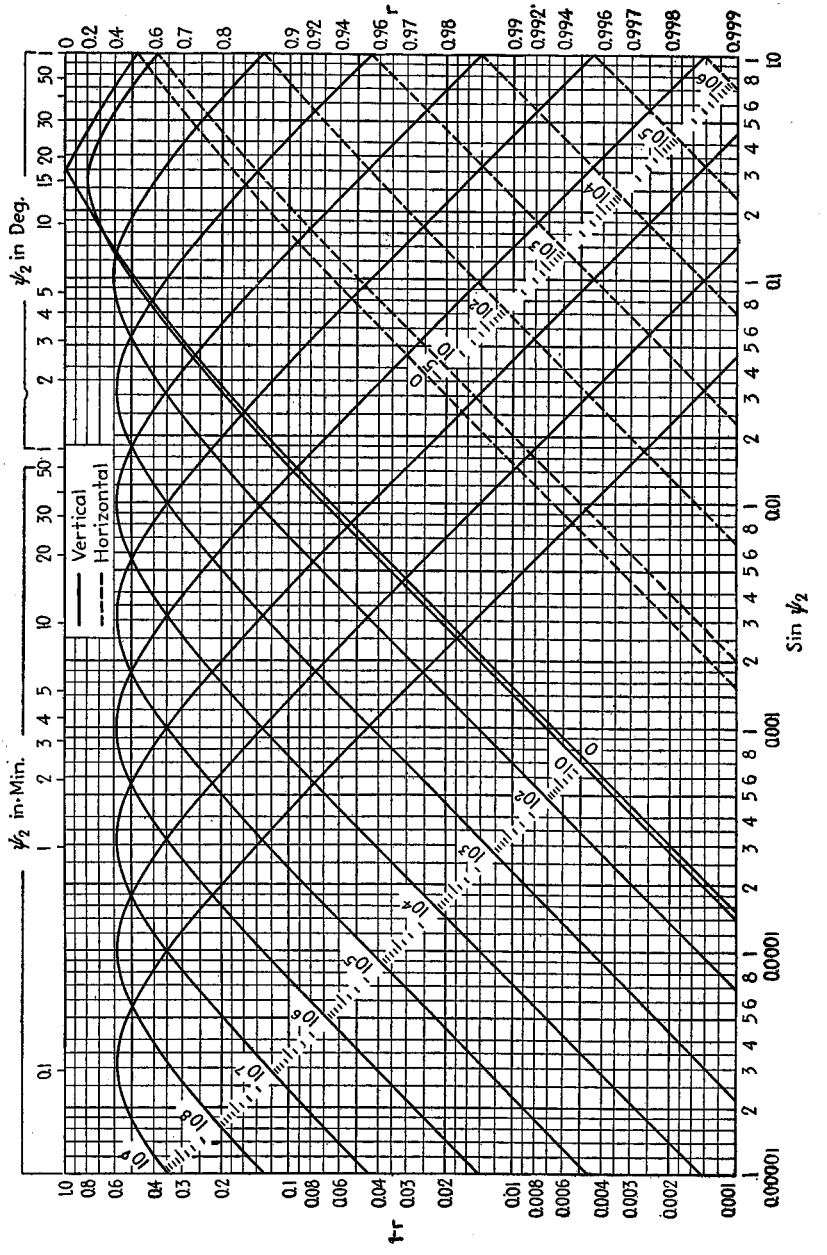


FIG. 26a.—Magnitude and phase of reflection coefficient for $\epsilon = 4$. The number on each curve gives the value of the quantity $x = (\sigma_{emu}/f_{mc})18 \times 10^{14}$ to which it applies, and ψ_2 is the angle of incidence (see Fig. 24).


ψ2 in Min.
ψ2 in Deg.
Vertical
Horizontal
r
1-r
Sin ψ2

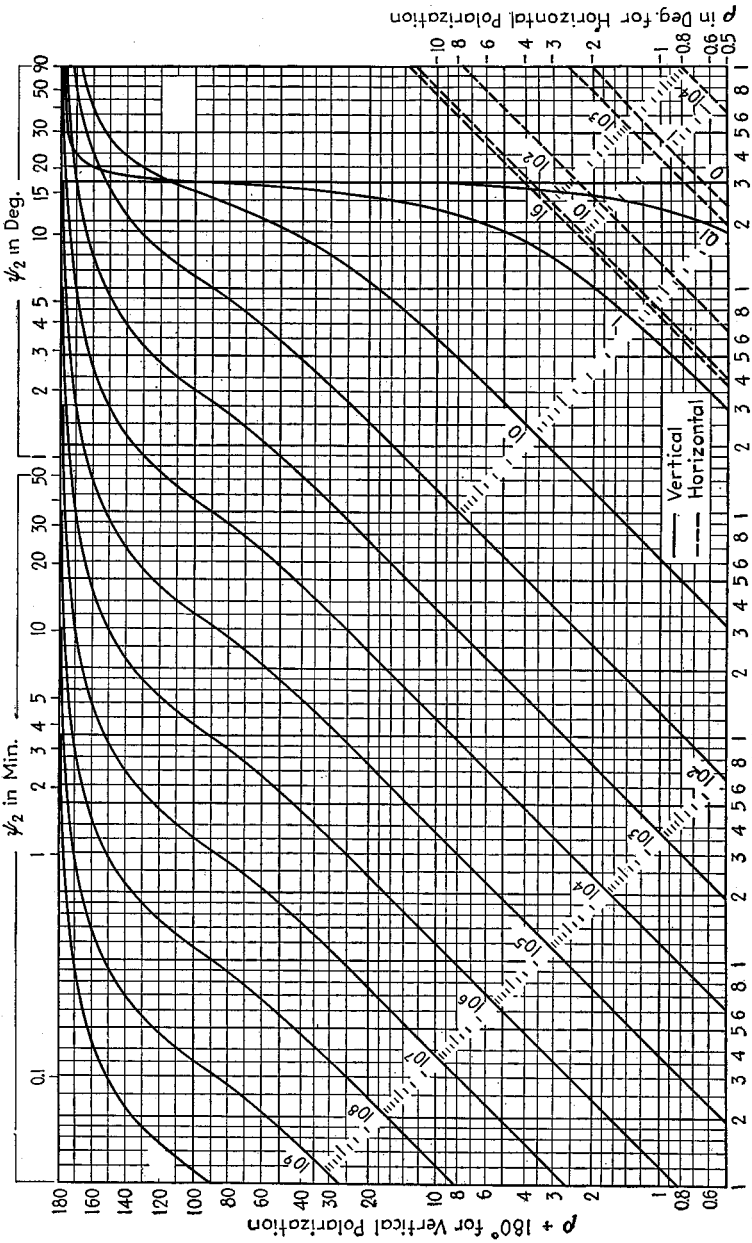
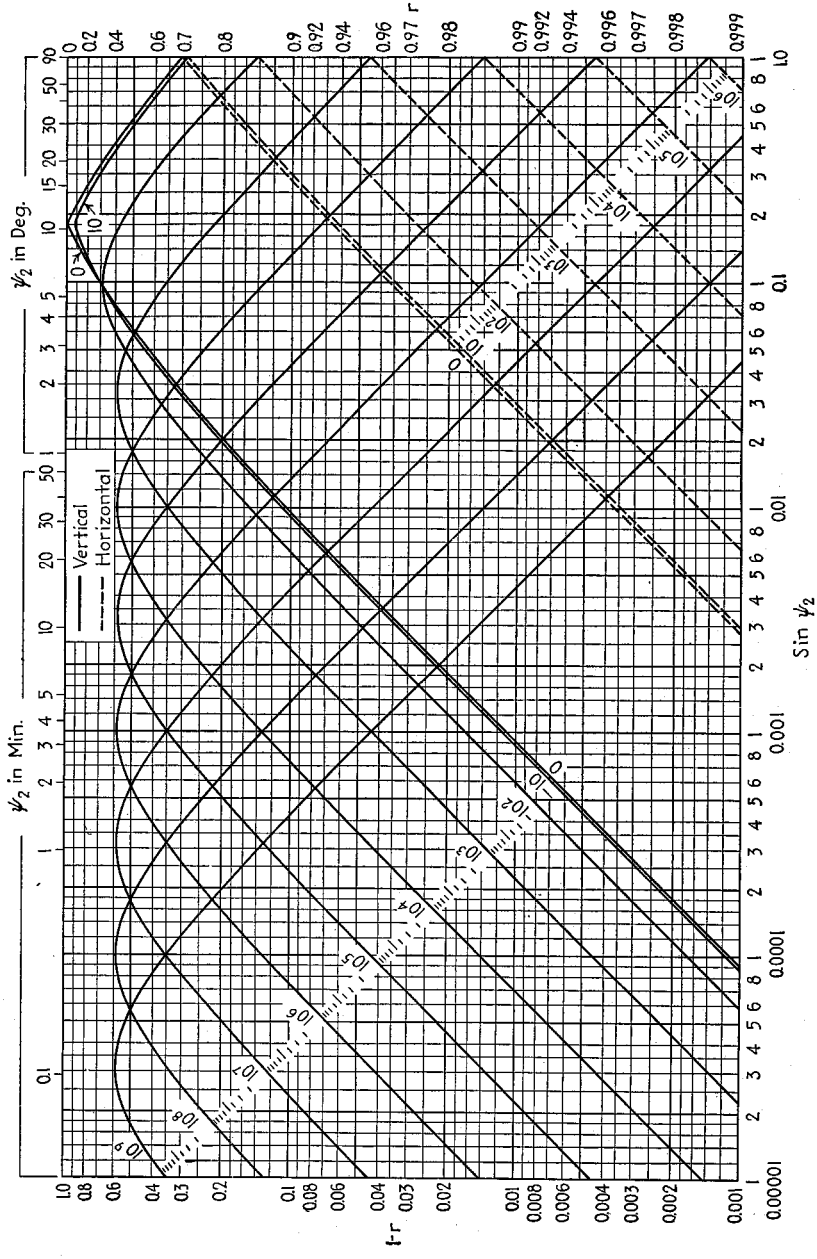


FIG. 266.—Magnitude and phase of reflection coefficient for $\epsilon = 10$. The number on each curve gives the value of the quantity $x = (\sigma_{emu}/f_{mc})18 \times 10^{14}$ to which it applies, and ψ_2 is the angle of incidence (see Fig. 24).


ψ2 in Min.
ψ2 in Deg.
Vertical
Horizontal
1-r
r
Sin ψ2

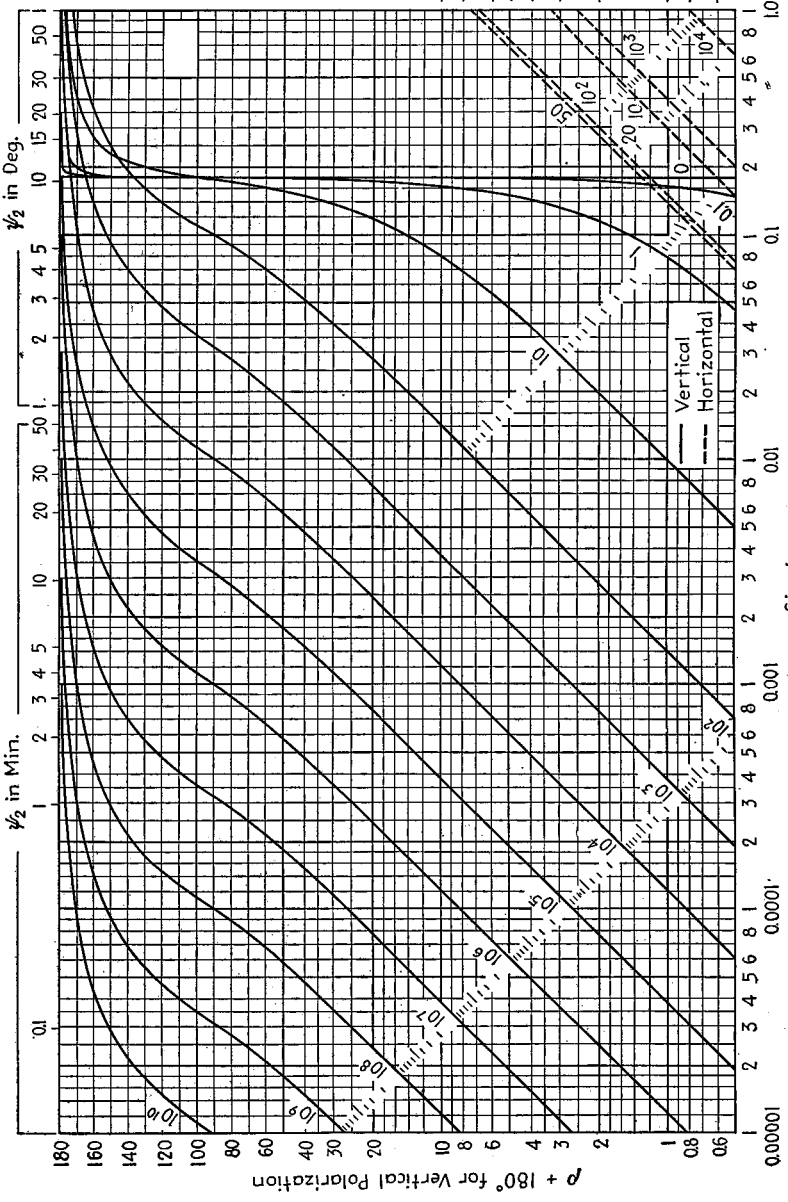
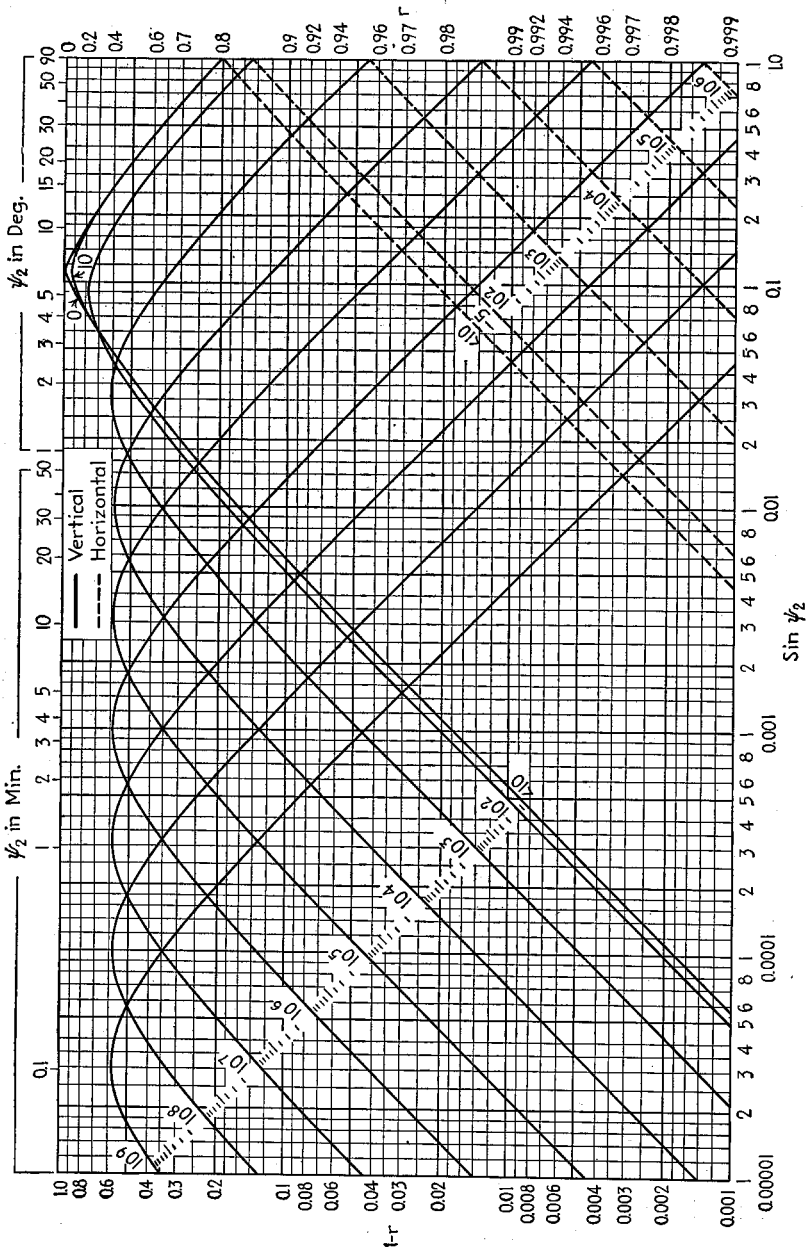


FIG. 26c.—Magnitude and phase of reflection coefficient for $\epsilon = 30$. The number on each curve gives the value of the quantity $x = (\sigma_{emu}/f_{mc})18 \times 10^{14}$ to which it applies, and ψ_2 is the angle of incidence (see Fig. 24).


ψ2 in Min.
ψ2 in Deg.
Sin ψ2
1-r
r
Vertical
Horizontal
10
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹

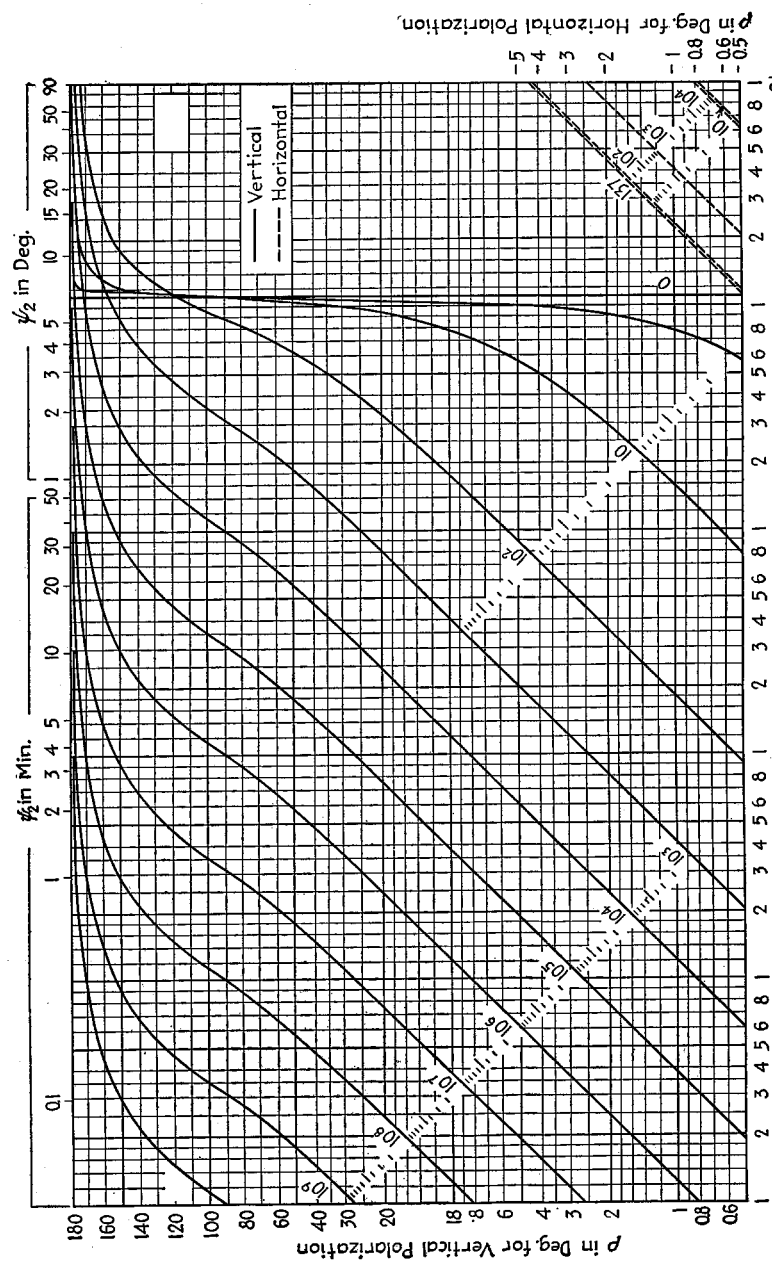


FIG. 26d.—Magnitude and phase of reflection coefficient for $\epsilon = 80$. The number on each curve gives the value of the quantity $(\sigma_{emu}/f_{mc})18 \times 10^{14}$ to which it applies, and ψ_2 is the angle of incidence (see Fig. 24).

incidence, quickly drops to a relatively small value at a small angle of incidence, and then rises until at normal incidence it assumes the same value as with horizontal polarization. The phase angle of the reflection coefficient for vertically polarized waves undergoes correspondingly wide variations. At glancing incidence the phase lag is 180°, corresponding to a reversal in polarity such as occurs in the reflection of horizontally polarized waves from a perfect earth. This angle of lag then drops with increased angles of incidence, becoming 90° when the magnitude of the reflection coefficient goes through a minimum, and finally at vertical incidence assuming the same small value as is obtained with horizontal polarization.

The angle at which the magnitude of the reflection coefficient goes through a minimum with vertically polarized waves corresponds to the Brewster angle in optics, and can be termed the pseudo-Brewster angle. It is given with an accuracy sufficient for most purposes by the relation[1]

Pseudo-Brewster angle ψ_B

$$= \sin^{-1} \sqrt{\frac{\epsilon - 1 + \sqrt{(x^2 + \epsilon^2)^2(\epsilon - 1)^2 + x^2[(x^2 + \epsilon^2)^2 - 1]}}{(x^2 + \epsilon^2)^2 - 1}} \qquad (29a)$$

The notation is the same as in Eq. (27).

When $x >> \epsilon$ (low frequencies) this becomes

$$\sin^2 \psi_B = \frac{1}{x} \qquad (29b)$$

Similarly, when $x << \epsilon$ (high frequencies)

$$\sin^2 \psi_B = \frac{1}{\epsilon + 1} \qquad (29c)$$

The value of reflection coefficient r_B obtained when the angle of incidence equals ψ_B is

$$r_B = \tan\left(\frac{\pi}{8} + \frac{b'}{4} - \frac{b''}{2}\right) \qquad (29d)$$

where $\tan b' = \dfrac{\epsilon - \cos^2 \psi_B}{x}$.

$\tan b'' = \dfrac{\epsilon}{x}$.

6. Earth Constants.—The conductivity and dielectric constant of the earth vary greatly according to conditions. Typical values are given in Table 1. It is found that high values of dielectric constant tend to go with large conductivities, and vice versa, and that the highest conductivities are obtained with wet loam, while the poor conductivities and low dielectric constant are generally associated with dry, rocky, and sandy soil. Earth having a conductivity of the order of 5×10^{-14} to 10×10^{-14} emu is considered as average to better than average, values above 10×10^{-14} emu are very high, and conductivities of the order of 1.0×10^{-14} emu are considered as very low. Sea water has a conductivity many times that of earth, and also possesses a very high dielectric constant.

The effective conductivity of the earth tends to be reduced if the earth's surface is not level, as in the case of mountainous or hilly regions. The effective conductivity also tends to be low in wooded areas, and in regions containing many buildings, particularly cities with large office buildings.

[1] See K. A. Norton, Ground Wave Propagation, *F.C.C. Mimeo. Rept.* 47475, presented before Broadcast Engineering Conference, February, 1941.

TABLE 1.—SOME TYPICAL GROUND CONSTANTS*

Type of terrain	Dielectric constant	Conductivity, emu
Fresh water	80	1×10^{-14}
Sea water, minimum attenuation	81	4.64×10^{-11}
Pastoral, low hills, rich soil, typical of Dallas, Tex., Lincoln, Neb., areas	20	3×10^{-13}
Pastoral, low hills, rich soil, typical of Ohio and Illinois	14	10^{-13}
Flat country, marshy, densely wooded, typical of Louisiana near Mississippi River	12	7.5×10^{-14}
Pastoral, medium hills and forestation, typical of Maryland, Pennsylvania, New York, exclusive of mountainous territory and seacoasts	13	6×10^{-14}
Pastoral, medium hills and forestation, heavy clay soil, typical of central Virginia	13	4×10^{-14}
Rocky soil, steep hills, typical of New England	14	2×10^{-14}
Sandy, dry, flat, typical of coastal country	10	2×10^{-14}
City, industrial areas, average attenuation	5	10^{-14}
City, industrial areas, maximum attenuation	3	10^{-15}

* From Standards of Good Engineering Practice Concerning Standard Broadcast Stations, *Federal Register*, p. 2862, July 8, 1939.

The value of conductivity and dielectric constant that is effective for radio waves represents the average value for a distance below the surface of the earth determined by the depth to which ground currents of appreciable amplitude exist. This depth of penetration depends upon the frequency, dielectric constant, and conductivity, and is commonly of the order of 5 to 10 feet at the frequencies used in short-wave communication, and 50 or more feet at broadcast and lower frequencies (see Par. 4). As a result, the earth constants are not particularly sensitive to conditions existing at the very surface of the earth, as, for example, recent rainfall. The effective value of the earth constants tends to be substantially independent of frequency over a relatively wide frequency range.

The constants of the earth may be measured in a variety of ways.[1] The most widely used method consists in determining a curve giving the surface-wave intensity as a function of distance from the transmitter. A smooth curve is drawn to represent the experimental data as accurately as possible, and the earth constants are then chosen by trial and error so that the best agreement with Eq. (1) and Fig. 2 will be obtained. A second method involves measuring the tilt and polarization of the surface wave. A third method that has been used consists in taking samples of the earth and measuring their effective conductivities and dielectric constants. Finally, the earth constants may be deduced from a measurement of the reflection coefficient. It has been found that these various methods will give results that are in substantial agreement, provided that they are carefully followed out.

THE IONOSPHERE AND ITS EFFECT ON THE PROPAGATION OF RADIO WAVES[2]

7. The Ionosphere.—The term *ionosphere* refers to the ionized region that exists in the upper atmosphere, and that has an important influence on the propagation of radio

[1] An excellent discussion of various measuring methods is given by Feldman, *loc. cit.* See also R. H. Barfield, Some Measurements of the Electrical Constants of the Ground at Short Wavelengths by the Wave-tilt Method, *Jour. I.E.E.*, Vol. 75, p. 214, 1934; also, Wireless Section, *I.E.E.*, Vol. 9, p. 286, September, 1934; R. L. Smith-Rose, Electrical Measurements on Soil with Alternating Currents, *Jour. I.E.E.*, Vol. 75, p. 221, 1934; also, Wireless Section, *I.E.E.*, Vol. 9, p. 293, September, 1934.

[2] For summaries and bibliographies on wave propagation, the reader is referred to J. H. Dellinger,

waves. This ionized region extends from a lower limit of 50 to 80 km, upward as far as there is atmosphere to be ionized. The ionization results principally from the action of ultraviolet radiation from the sun, but the processes by which it is produced and maintained are obscured by the lack of precise knowledge of the composition, temperature, etc., of the atmosphere at the levels involved. It can be said, however, that the atmospheric pressure in the ionosphere is extremely small, being little if any greater than that often found in a vacuum tube. Because of this, collisions between electrons and ions are relatively infrequent even in the lower part of the ionosphere; so recombination takes place only very slowly. The ionization produced in the daytime is also carried over into the night by some means, the details of which are at present uncertain.

Because of the variation in chemical composition of the air with height, and because different gases differ in their ability to absorb solar radiation of different frequencies, there is a tendency for the ionization in the ionosphere to become stratified, so that the curve of electron density as a function of height commonly has several maxima, as shown in Fig. 27. There are two semipermanent "layers" of this character, the E and the F or F_2 layers, with a third designated as F_1 usually present in the daytime. The

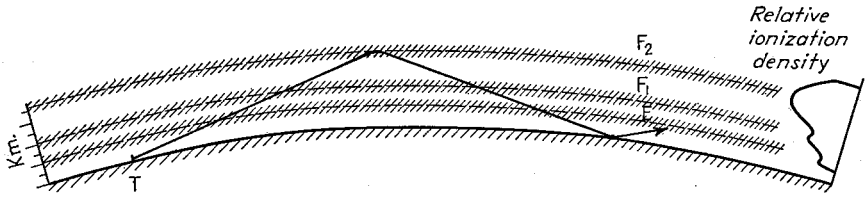


FIG. 27.—Schematic diagram drawn approximately to scale showing location of various ionosphere layers, approximate distribution of ionization, and typical path of a wave refracted from the F_2 layer.

height of a particular layer and the maximum electron density in the layer will vary at different times of the day and of the year as the result of variations in the composition and temperature of the air at different heights, and of the radiation received from the sun. The distribution of ionization with height in a typical case is shown schematically in Fig. 27, in which three distinct layers are shown corresponding to typical daytime conditions. It will be noted that the ionization does not drop to zero between the layers, but merely has a value less than the maxima on either side.

A radio wave entering the ionosphere from the earth has a tendency to be bent earthward, and if the conditions are favorable, the bending will be sufficient to cause the wave to return to earth as shown in Fig. 27. Upon striking the earth it will then be again reflected upward as shown. This action makes it possible to carry on radio communication over great distances in spite of the earth's curvature. The wave that reaches these distant points as a result of the action of the ionosphere is termed the *sky wave.*

8. Propagation of a Radio Wave in an Ionized Medium. *Fundamental Mechanisms Involved.*—When a radio wave enters the ionosphere the electric field of the wave exerts a force upon the electrons, setting them in vibration at the frequency of the

The Role of the Ionosphere in Radio-wave Propagation, *Trans. A.I.E.E.*, Supplement, Vol. 58, p. 803, 1939; S. S. Kirby, L. V. Berkner, and D. M. Stuart, Studies of the Ionosphere and Their Application to Radio Transmission, *Proc. I.R.E.*, Vol. 22, p. 481, April, 1934; Karl K. Darrow, Analysis of the Ionosphere, *Bell System Tech. Jour.*, Vol. 19, p. 455, July, 1940; G. W. Kenrick and G. W. Pickard, Summary of Progress in the Study of Radio Wave Propagation Phenomena, *Proc. I.R.E.*, Vol. 18, p. 649, April, 1930; Bibliography on Radio Wave Phenomena and Measurement of Radio Field Intensity, *Proc. I.R.E.*, Vol. 19, p. 1034, June, 1931; S. K. Mitra, "Report on the Present State of our Knowledge of the Ionosphere" (booklet), University of Calcutta, India, 1935; H. R. Mimno, The Physics of the Ionosphere, *Rev. Mod. Phys.*, Vol. 9, p. 1, January, 1937.

wave.[1] Under conditions where the earth's magnetic field has negligible effect, this vibration is along a line parallel to the electric field of the wave. The velocity of vibration lags 90° behind the electric field, and is inversely proportional to the frequency. Since a moving electron represents a current, each electron set in vibration by the radio wave acts as a miniature parasitic antenna that absorbs energy from the wave and then reradiates this energy in a different phase. The net effect, after the phase difference between the original and reradiated fields is taken into account, is to bend the wave path away from the regions of high electron density toward regions of lower density. The magnitude of this effect varies with the amplitude of the electron vibration, and therefore becomes increasingly great as the wave frequency is lowered.

The electrons in the ionosphere exist in the presence of the earth's magnetic field. Such a magnetic field exerts a force on a moving electron that is proportional to the instantaneous velocity of the electron, and to the component of the magnetic field at right angles to the direction of motion. The direction of this force is at right angles to the direction of motion of the electron, and to the component of the magnetic field producing the deflecting force. The effect of the earth's magnetic field at the higher radio frequencies is to cause each electron to vibrate in an elliptical path, as shown at *a* or *b* in Fig. 28, with the major axis of the ellipse lying in the direction of the electric field of the wave. The ratio of minor axis to major axis increases as the velocity with which the electron vibrates becomes greater, *i.e.*, as the frequency is reduced. This trend continues until at a frequency termed the gyro frequency and having a value of approximately 1.4 mc, the electron vibrates in a spiral path as shown at *c*, in which the velocity becomes increasingly great. At still lower frequencies the electrons vibrate in loops as shown at *d* and *e*, commonly making several loops during each half cycle of the radio wave.

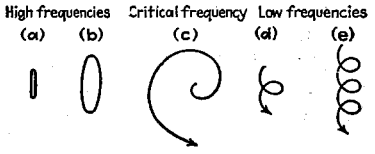


FIG. 28.—Paths followed by an electron when vibrating under the influence of a radio wave, in the presence of the earth's magnetic field. To avoid confusion the paths for the low-frequency cases represent only a half cycle of the vibration that proceeds cyclically.

It will be noted that in all cases where the magnetic field of the earth has an influence, the effect is to cause the vibrating electrons to have some motion at right angles to the direction of vibration that would exist in the absence of a magnetic field. The polarization of the fields reradiated by the vibrating electrons will hence differ from that of the passing wave, causing the polarization of the resultant wave to be affected by the presence of the earth's magnetic field. This effect is maximum at the gyro frequency, but is present to some extent even for waves of much higher and much lower frequencies.

The magnetic field, in addition to affecting the polarization, also causes a wave to be split into two components, which follow different paths, which have different phase velocities, differ in polarization effects, and suffer different attenuations. This is double refraction.

An electron set in vibration by a passing radio wave will from time to time collide with a gas molecule. In such a collision, the kinetic energy that the electron has acquired from the radio wave is partly transferred to the gas molecule, and partly reradiated in the form of a disordered radio wave that contributes nothing to the transmission. The net result is therefore an absorption of energy from the passing wave. The rate at which energy is thus absorbed depends upon the gas pressure (*i.e.*,

[1] The electric field of the wave also acts upon any ions that may be present, but because of their much heavier mass, the effects produced by the ions are relatively unimportant as compared with electrons.

upon the likelihood of the electron colliding with a gas molecule) and upon the average velocity of the vibrating electron (*i.e.*, upon the average energy lost per collision).

Refractive Index and Dielectric Constant of an Ionized Medium.—The path in the ionosphere followed by an electromagnetic wave depends upon the refractive index μ of the medium defined by[1]

$$\text{Refractive index} = \mu = \frac{\text{velocity of light}}{\text{phase velocity of wave}} \tag{30}$$

The refractive index of an ionized medium when not influenced by the earth's magnetic field, and under conditions where the loss of energy is negligible, is[2]

$$\text{Refractive index, } \mu = \sqrt{k} = \sqrt{1 - \frac{4\pi e^2 N}{m\omega^2}} \tag{32a}$$

where k = dielectric constant of ionized medium.
e = electron charge, esu.
m = electron mass, grams.
$\omega/2\pi$ = frequency.
N = number of electrons per cu cm.

Substituting values of e and m appropriate for electrons gives

$$\text{Refractive index} = \mu = \sqrt{1 - \frac{81N}{f_{kc}^2}} \tag{33}$$

where f_{kc} is the frequency in kilocycles.

It will be noted that the presence of electrons causes the dielectric constant and the refractive index of the space to be less than unity. This results from the fact that the current represented by the vibrating electrons lags 90° behind the electric field of the wave, and so subtracts from the capacitive displacement current due to the capacity of space.

[1] The phase velocity is determined by the rapidity with which the phase changes along the path of a wave, and is not to be confused with group or signal velocity. In the case of waves traveling in an ionized medium the phase velocity is greater than the velocity of light, and the corresponding refractive index is less than unity. The signal or group velocity represents the velocity with which the energy of the wave travels, rather than the rate at which the phase changes. The group velocity can be obtained by sending a short pulse of energy along a path of known distance and determining the time it takes the pulse to arrive at the receiving point. The group velocity cannot exceed the velocity of light, and may even approach zero in certain limiting cases. The phase and group velocities are related to each other by the equation

$$\frac{\text{Group velocity}}{\text{Phase velocity}} = \frac{1}{1 - \frac{\omega}{v}\frac{dv}{d\omega}} \tag{31}$$

where $\omega = 2\pi f$ and v is phase velocity. A good discussion of the relationship group and phase velocity is given by Karl K. Darrow, "Analysis of the Ionosphere," *Bell System Tech. Jour.*, Vol. 19, p. 455, July, 1940. See also Par. 52, Sec. 3.

[2] This equation is derived by many authors. Thus see William G. Baker and Chester W. Rice, Refraction of Short Waves in the Upper Atmosphere, *Trans. A.I.E.E.*, Vol. 45, p. 302, 1926; P. O. Pederson, Propagation of Radio Waves, *G.E.C. Gad*, Copenhagen, Denmark (in English).

The refractive index is also sometimes written as

$$\mu = \sqrt{1 - \frac{4\pi e^2 N}{\omega^2 m + a(4\pi e^2 N)}} \tag{32b}$$

where a is a constant that is assumed to have the value of zero in Eq. (32*a*), but is assigned the value 0.33 by some writers. There is uncertainty as to which is the proper formula. For further discussion see Newbern Smith, Oblique-incidence Radio Transmission and the Lorentz Polarization Term, *Jour. Research, Nat. Bur. Standards*, Vol. 26, p. 105, February, 1941; H. G. Booker and L. V. Berkner, An Ionospheric Investigation Concerning the Lorentz Polarization Correction, *Terr. Mag. and Atmos. Elec.*, Vol. 43, p. 427, December, 1938

When the action of the earth's magnetic field is taken into account, the refractive index, assuming negligible loss, is given by the relation[1]

$$\mu = \sqrt{1 - \frac{2}{2\alpha - \left(\frac{\gamma_T^2}{\alpha - 1}\right) \mp \sqrt{\frac{\gamma_T^4}{(\alpha - 1)^2} + 4\gamma_L^2}}} \tag{34}$$

where $\alpha = \dfrac{m\omega^2}{4\pi Ne^2}$.

$\gamma_L = \dfrac{m\omega \dfrac{H_L e}{mc}}{4\pi Ne^2}$.

$\gamma_T = \dfrac{m\omega \dfrac{H_T e}{mc}}{4\pi Ne^2}$.

H_L = magnetic field in direction of propagation in gauss.

H_T = magnetic field at right angles to direction of propagation.

$c = 3 \times 10^{10}$ = velocity of light, cm per sec.

The remaining notation is the same as in Eq. (32*a*). The double sign in Eq. (34) means that there are now two values of the refractive index. The index corresponding to the upper sign is nearer the value given by Eq. (32*a*) in the absence of a magnetic field and gives rise to the *ordinary ray*, discussed below, while the other sign represents the refractive index associated with the *extraordinary ray*.

Energy absorption in the ionosphere reduces the effect of the electrons since the collisions that give rise to the loss interfere with the motion of the electrons. As a consequence, absorption causes the refractive index and the dielectric constant to have values closer to unity than would otherwise be the case. When the magnetic field of the earth is neglected, one has[2]

$$\text{Dielectric constant of ionosphere} = k' = 1 - \Delta k \left(\frac{1}{1 - \left(\frac{v}{\omega}\right)^2}\right) \tag{35}$$

$$\text{Refractive index} = \sqrt{\frac{k'}{2} + \sqrt{\frac{k^2}{4} + \left(\frac{2\pi\sigma c^2}{\omega}\right)^2}} \tag{36}$$

where $\Delta k = 81N/f_{kc}^2$ is the amount by which the refractive index is less than unity in the absence of losses, v is the number of collisions per second that an electron makes with the gas molecules, σ is the conductivity of the ionosphere in emu, and the remaining notation is as used above. The quantities v and σ are discussed in detail below. It will be noted that energy loss has little effect if $v << \omega$; *i.e.*, if there are appreciably less than 2π collisions per cycle between electrons and gas molecules, while if $v >> \omega$, then the dielectric constant and refractive index both approach unity.

When the effect of the earth's magnetic field is included, the formulas for refractive indexes and dielectric constant become quite involved when losses are taken into account.[3] However, the qualitative effect is similar to the case where the magnetic effects are ignored; *i.e.*, the losses make the refractive indexes and the dielectric constants have a value closer to unity than in the absence of losses, and this tendency is large or small according to whether v/ω is appreciably more or less than unity.

[1] E. V. Appleton, Wireless Studies of the Ionosphere, *Jour. I.E.E.*, Vol. 71, p. 642, 1932; also, Wireless Section *I.E.E.*, Vol. 7, p. 257, September 1932; S. S. Kirby, L. V. Berkner, and D. M. Stuart, Studies of the Ionosphere and their Application to Radio Transmission, *Proc. I.R.E.*, Vol. 22, p. 481, April, 1934.

[2] For derivations, see Baker and Rice, *loc. cit.*, or Pederson, *loc. cit.*

[3] The complete formula is given by Appleton, *loc. cit.*

Energy Absorption in the Ionosphere.—The ionosphere absorbs energy from waves passing through it as a result of collisions between the vibrating electrons and gas molecules. This action causes the ionosphere to act with respect to the radio wave as a medium having a certain conductivity σ. This conductivity will have a value proportional to the average number of collisions between electrons and gas molecules per second, and will be greater the higher the average velocity possessed by the vibrating electrons. To the extent that the effects of the earth's magnetic field can be neglected one has[1]

$$\sigma = N \frac{e^2}{m} \frac{v}{\omega^2 + v^2} \frac{1}{9 \times 10^{20}} \tag{37}$$

where σ = conductivity, emu.
e = electronic charge, esu.
m = electron mass, grams.
v = number of collisions per second that an electron makes with the gas molecules.
$\omega/2\pi$ = frequency.
For electrons $e^2/m = 2.81 \times 10^{-13}$.

The contribution made by a single electron to the ionosphere conductivity is maximum when ω equals v, *i.e.*, when there are 2π collisions per cycle. The number of collisions per second is determined by the thermal-agitation velocity of the electrons, and by the gas pressure. It is independent of the frequency or amplitude of the radio wave. The value of v is a function of height. Assuming atmospheric pressures given by Chapman and Milne,[2] and an electron "temperature" of 6000°K, v is given in Table 2. When the distribution of electrons with height is taken into account, it is found that because the electron density drops off very rapidly below 80 km, the maximum conductivity occurs in the ionosphere at heights of the order of 60 to 80 km for all ordinary frequencies. This is at the lower edge of the ionosphere. Losses well up in the ionosphere are smaller because even though the electron density is high, the gas pressure is enormously low.

TABLE 2

Height, km	Pressure, dynes per sq cm	Number of molecules per cu cm	Molecular mean free path, cm	Electron collision frequency per sec
0	1.01×10^{6}	2.7×10^{19}	$9. \times 10^{-6}$	9.5×10^{11}
12	1.92×10^{5}	6.5×10^{18}	$4. \times 10^{-5}$	2.1×10^{11}
20	5.53×10^{4}	1.9×10^{18}	$1. \times 10^{-4}$	8.5×10^{10}
40	2.55×10^{3}	8.6×10^{16}	$3. \times 10^{-3}$	2.8×10^{9}
60	1.24×10^{2}	4.2×10^{15}	$6. \times 10^{-2}$	1.4×10^{8}
80	6.27	2.1×10^{14}	1.0	8.5×10^{6}
100	0.363	1.2×10^{13}	20.	4.3×10^{5}
150	1.49×10^{-2}	5.0×10^{11}	500.	1.7×10^{4}
200	5.62×10^{-3}	1.8×10^{11}	1,000.	8.5×10^{3}
300	6.99×10^{-4}	2.4×10^{10}	1×10^{4}	8.5×10^{2}
400	1.05×10^{-4}	3.6×10^{9}	7×10^{4}	1.2×10^{2}
600	2.59×10^{-6}	8.8×10^{7}	3×10^{6}	2.8
800	7.97×10^{-8}	2.7×10^{6}	9×10^{7}	0.95
1,000	2.92×10^{-9}	9.9×10^{4}	3×10^{9}	2.8×10^{-3}

[1] For derivation, see Baker and Rice, *loc. cit.*, or Pederson, *loc. cit.*
[2] See S. Chapman and E. A. Milne, *Quart. Jour. Roy. Met. Soc.*, Vol. 46, p. 357, 1920.

Hence practically all the absorption that a wave suffers as a result of the ionosphere occurs where the wave enters and leaves the ionosphere.

The conductivity of the ionosphere is affected by the presence of the earth's magnetic field, since the magnetic field modifies the average velocity with which the electrons vibrate under the influence of the radio wave. At frequencies considerably higher than the gyro frequency, the mean velocity of the vibrating electron is only very slightly affected by the earth's magnetic field, with resulting negligible influence on the conductivity. However, at frequencies in the vicinity of the gyro frequency, the average velocity with which an electron vibrates is increased as a result of the presence of the earth's magnetic field, and the conductivity becomes high. Such frequencies accordingly tend to be more strongly absorbed by the ionosphere than either higher or lower frequencies. At frequencies much lower than the gyro frequency, the action of the earth's magnetic field is to reduce the average velocity of the vibrating electrons to a value less than that which would be obtained in the absence of a magnetic field, and this trend becomes more pronounced the lower the frequency. The result is that at the lower frequencies the earth's magnetic field reduces the conductivity to values less than given by Eq. (37). The exact expression for conductivity is too cumbersome to be given here.[1]

A wave traveling a distance d through a portion of the ionosphere having a conductivity σ will be attenuated according to the law

$$\frac{E_2}{E_1} = \epsilon^{-\alpha l} \tag{38}$$

where E_1 = initial amplitude of the wave.
E_2 = amplitude of wave after traveling a distance l.
α = attenuation constant.
l = distance in same units of length as α.

The attenuation constant is[2]

$$\alpha = \frac{\omega}{c\sqrt{2}} \sqrt{\sqrt{k'^2 + \left[(1 - k')\frac{\nu}{\omega}\right]^2} - k'} \tag{39}$$

where k' is given by Eq. (35) and c is the velocity of light in the same units as l. The total attenuation $\int \alpha\, dl$ that a wave suffers in the ionosphere can be calculated from Eq. (39) if ν and the electron density N are known as a function of height h, remembering that dl/dh depends on the angle of the wave path with the horizontal. In the important special case of a section of the ionosphere where the frequency is much less than the critical frequency and $\omega >> \nu$, Eq. (39) reduces to

$$\alpha = \frac{2\pi\sigma}{c\sqrt{k'}} = \frac{2\pi e^2 N\nu}{c\sqrt{k'}m\omega^2} \tag{40}$$

The notation is the same as in Eqs. (35), (37), and (39). Equation (40) indicates that the energy loss of a high-frequency wave in traversing the lower part of the ionosphere (where most of the loss tends to occur) is inversely proportional to the square of the frequency.

Wave Path in the Ionosphere.[3]—The actual path that a radio wave follows in the ionosphere is controlled by the refractive index μ. Variation in the index along the

[1] It is given by Appleton, *loc. cit.*
[2] For derivation, see Baker and Rice, *loc. cit.*
[3] This discussion assumes that the change in refractive index in a distance corresponding to one wave length is negligibly small. If this is not true, reflections take place as discussed below.

path causes the wave to follow a curved path such that

$$\frac{1}{\text{Radius of curvature}} = \frac{1}{\mu}\frac{d\mu}{ds} \tag{41}$$

Here s is distance in the direction normal to that which the wave is following, and the curvature is away from the region of high refractive index toward the region of lower index. It will be noted that the rate of curvature of the path is greatest when the refractive index varies most rapidly, and that the path becomes a straight line when the refractive index is constant. The differential equation of the path can be derived from Eq. (41), and is[1]

For short distances such that the earth curvature can be neglected:

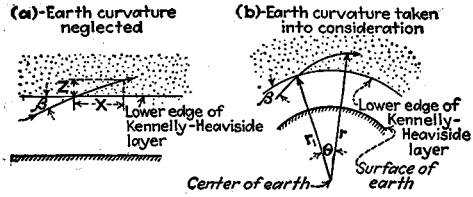


FIG. 29.—Diagrams illustrating notation of Eqs. (42*a*) and (42*b*).

$$dx = \frac{\cos\beta\, dz}{\sqrt{\mu^2 - \cos^2\beta}} \tag{42a}$$

For distances great enough that the earth's curvature must be taken into account:

$$d\theta = \frac{r_1 \cos\beta\, dr}{r\sqrt{r^2\mu^2 - r_1^2\cos^2\beta}} \tag{42b}$$

where β = angle of incidence of wave entering ionosphere.
μ = refractive index of ionosphere, and is a function only of z.
z = height of ray above lower edge of ionosphere.
x = distance parallel to earth's surface, measured from the point at which wave enters ionized region.
θ = polar angle (in radians) intercepted by wave path.
r = distance from wave to center of earth.
r_1 = distance from lower edge of ionosphere to center of earth.
This notation is shown in Fig. 29.

In making use of Eqs. (42*a*) and (42*b*) it is to be noted that with small angles of incidence, the angle δ between the wave path and the surface of the earth differs from the angle of incidence β at the ionosphere because of the curvature of the earth. The relationship is

$$\cos\beta = \frac{\cos\delta}{1 + \frac{h}{r_0}} \tag{43}$$

where β and δ = angles illustrated in Fig. 30.
h = height of the lower edge of ionosphere above earth.
r_0 = radius of earth in same units of length as used to express height.
The relationship expressed by Eq. (43) is shown in Fig. 31 for small and moderate values of δ. At large values of δ one can normally assume $\beta \cong \delta$.

The total angle through which a wave that has entered the ionosphere is deflected can be obtained according to Snell's law by the relation

$$\cos\beta = \mu\cos\alpha \tag{44}$$

[1] Derivations of wave path equations are given by many authors. Thus see Baker and Rice, *loc. cit.*, Pedersen, *loc. cit.* (p. 171), or G. W. Kenrick and C. K. Jen, Measurements of the Height of the Kennelly-Heaviside Layer, *Proc. I.R.E.*, Vol. 17, p. 74, April, 1929.

where α and β are the angles of refraction and incidence, respectively, as shown in Fig. 32, and μ is the refractive index at the point where α is observed. This relationship

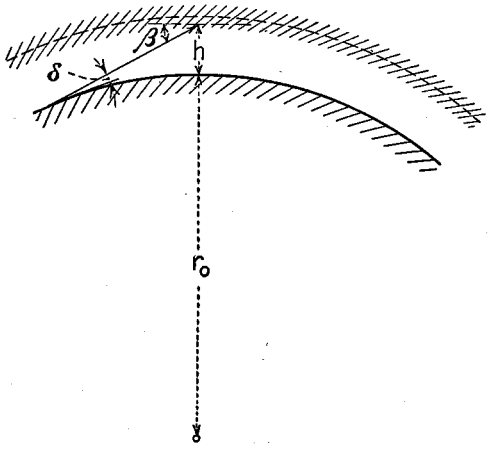


FIG. 30.—Diagram illustrating notation of Eq. (43).

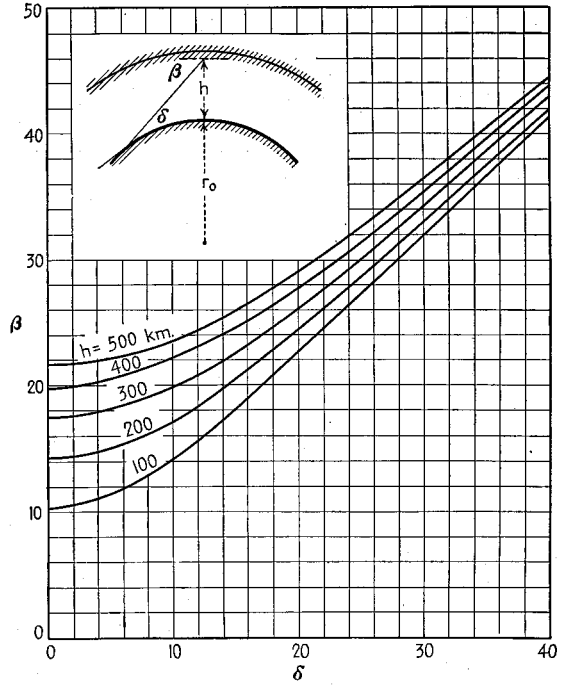


FIG. 31.—Relation of angle of incidence β at the ionosphere to the angle of incidence δ at which the wave leaves the earth, for different ionosphere heights.

is independent of the way in which the refractive index varies between the lower edge of the ionosphere and the point of observation.

The special case of Eq. (44), where $\alpha = 0$, is of particular importance since this corresponds to the condition that must exist at the highest point in the path of a wave that is bent back to earth. The refractive index required to bend back to earth a wave entering the ionosphere with an angle β is

$$\mu = \cos \beta \tag{45}$$

With vertical incidence $\beta = 90°$, and the wave will penetrate the ionized region until it reaches a point where the refractive index is zero.

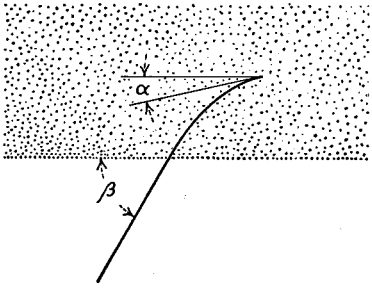


FIG. 32.—Diagram illustrating notation of Eq. (44).

When the action of the earth's magnetic field causes the ionosphere to have two refractive indexes, the wave that enters the ionosphere splits. One part follows the path called for by one index, the other the path corresponding to the second index. These two waves are propagated with different phase velocities, different polarization changes, and different absorptions. The one behaving most nearly as though the magnetic field were absent is termed the *ordinary ray*, since its characteristics are not much affected by the presence of the magnetic field. The other ray is termed the *extraordinary ray*, and in general suffers the greater attenuation. The paths followed by the two rays in the ionosphere are not greatly different except where one or the other of the rays is just barely returned to earth (or just fails to be returned) by the maximum electron density of one of the ionosphere layers.[1]

Virtual Height.—The *virtual height* is the height h shown in Fig. 33 that a wave would reach if it traveled in a straight line through the ionosphere and was then reflected as though from a mirrorlike surface. The virtual height is the height calculated on the basis of the angles at which the wave leaves and returns to the earth. It is also the height obtained from pulse measurements, since the length of time it requires a wave traveling with the velocity of light to cover the triangular path TAR in Fig. 33 is the same time taken by the wave to travel the actual path TBR. This is so because although the path TBR is shorter, the group velocity in the ionosphere is just enough less than the velocity of light to make up for the shorter distance.

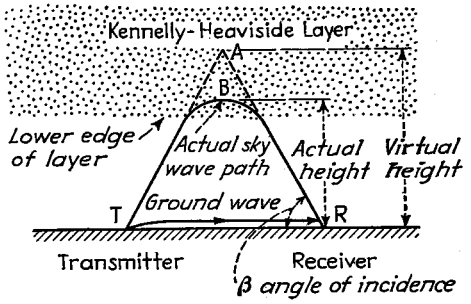


FIG. 33.—Diagram showing actual path TBR followed by a wave and the equivalent triangular path TAR used to define the virtual height.

The actual height of a layer, and also the electron distribution within a layer, can be determined with reasonably good accuracy by assuming that the electron distribution is parabolic with height. If f_c is the critical frequency of the layer, then at a frequency 0.834 f_c the virtual height is the actual height of the layer. Also, the dif-

[1] A comprehensive discussion of the effects produced by the earth's magnetic field is given by H. W. Nichols and J. C. Schelleng, Propagation of Electric Waves over the Earth, *Bell System Tech. Jour.*, Vol. 4, p. 215, April, 1925.

ference between this height and the virtual height at $0.648 f_c$ (or at $0.925 f_c$) is half the thickness of the layer.[1]

Critical Frequencies.—The maximum electron density of an ionosphere layer is commonly expressed in terms of the *critical frequency.* This is the highest frequency that the layer can return to earth when the ray enters the ionosphere with vertical incidence. The critical frequency corresponds to a refractive index of zero at the point of maximum electron density in the layer [see Eq. (45)], so that this density can be determined from measurements of critical frequency.

When the earth's magnetic field can be neglected, then for negligible attenuation at the center of the layer, one has[2]

$$\text{Critical frequency, kilocycles} = f_c = 9\sqrt{N} \tag{46}$$

When the magnetic field of the earth is taken into account there are two critical frequencies, f_c^0 and f_c^x, corresponding to the ordinary and extraordinary rays. The critical frequency f_c^0 for the ordinary ray is exactly the same as given by Eq. (46) when the magnetic field is neglected. The critical frequency, f_c^x, is related to the critical frequency f_c^0 and the gyrofrequency $f_H = He/2\pi mc$, by the equation

$$f_c^{x2} = f_c^0(f_c^0 + f_H) \tag{47}$$

The value of f_c^x is always more than f_c^0, and at high frequencies the difference, $f_c^x - f_c^0$ approaches $f_H/2$, or about 730 kc for $H = 0.52$ gauss.

Reflection of Radio Waves by the Ionosphere.—When the change in refractive index in a distance corresponding to a wave length is not negligibly small, then reflection as well as refraction takes place. This situation occurs at the lower radio frequencies where the distance represented by a wave length is large, and it also can occur even at relatively high radio frequencies as a result of electron clouds having very sharply defined edges. Such clouds are frequently present in the *E* layer, and account for the sporadic *E*-layer reflections.

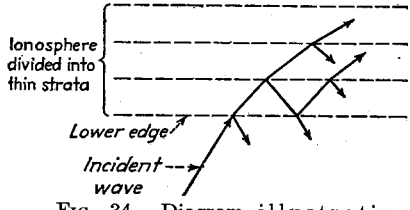


FIG. 34.—Diagram illustrating method of determining reflection in the ionosphere by assuming that the ionosphere is divided into a succession of thin layers.

Analysis of the ionosphere behavior under these conditions is carried out by dividing the ionosphere into a number of horizontally stratified layers, lying one above another, shown in Fig. 34.[3] Each layer is assumed to have a thickness appreciably smaller than a wave length, and is assigned a conductivity and a dielectric constant corresponding to the values that exist at the center of the layer. A wave reaching the interface between two such hypothetical layers experiences an abrupt change in the refractive index that produces a partial reflection. The angle of the reflected wave equals the angle of the incident wave, and the relative magnitude of the reflected wave can be expressed in terms of the reflection coefficient calculated with the aid of Eq. (27), provided that values of dielectric constant and conductivity appropriate to the layers involved are used.[4]

[1] These relations, as well as others of a similar character, together with discussions on how to check the validity of the parabolic assumption, and how to correct for the effect of lower layers when these introduce errors, are given by H. G. Booker and S. L. Seaton, Relation between Actual and Virtual Ionosphere Height, *Phys. Rev.*, Vol. 57, p. 87, Jan. 15, 1940.

[2] When the constant a appearing in Eq. (32b) is taken as 0.33, the effect is to make the electron density N corresponding to a given critical frequency about ⅔ times as great as otherwise.

[3] Shogo Namba, General Theory on the Propagation of Radio Waves in the Ionized Layer of the Upper Atmosphere, *Proc. I.R.E.*, Vol. 21, p. 238, February, 1933.

[4] A general formula for reflection covering the case where the conductivity can be neglected has been developed by Lord Rayleigh, On the Propagation of Waves through a Stratified Medium with Special

9. Normal Characteristics of the Ionosphere. *Normal Structure of the Ionosphere and Its Specification.*[1]—The electron distribution in the ionosphere has the general character illustrated in Fig. 27, and is normally characterized by maxima, termed "layers," which are designated by letters such as E, F, etc. There are two such layers having a permanent existence. The lower is termed the E layer; the upper is designated by F at night and F_2 in the daytime. In the day there is also usually an intermediate layer, the F_1 layer, that appears in early morning, persists throughout the day, and then fades out at night. Other layers have been reported by some investigators, of which the most important is the D layer, which lies below the E layer at a height of about 60 km and probably has an influence on daytime broadcast signals at considerable distances.[2] Very little is known about these other layers, beyond the fact that they are present only part of the time and are much less important than the E and F layers in returning the radio waves to earth.

These layers in the ionosphere are described in terms of their critical frequencies and virtual heights. The critical frequency is a direct measure of the maximum electron density of the layer. The virtual height is an indication of the height at which maximum electron density occurs. The virtual height is greater than the true height, but the distribution of ionization is such that the virtual height at frequencies considerably less than the critical frequency of the layer is very close to the true height.

Regular Variations of the Ionosphere with Time of Day and Time of Year.—The critical frequencies and virtual heights of the various ionosphere layers have diurnal and seasonal characteristics as shown in Fig. 35.[3] These curves represent monthly averages, and individual days can accordingly be expected to deviate from them to some extent. The E layer maintains a substantially constant virtual height of between 110 and 120 km throughout the day, and from season to season. The critical frequency of the E layer has, however, a regular diurnal and seasonal variation. During the day it varies synchronously with the altitude of the sun, with the daily maximum occurring at noon. The diurnal maximum is also greater in summer than in winter, *i.e.*, is greater when the zenith angle of the sun at noon increases. This critical frequency f_E follows very closely the law

---

Reference to the Question of Reflection, "Collected Works," Vol. VI, p. 80 (Eq. 5). The result unfortunately involves integrations that can be carried out only in special cases, however, Several assumed laws of variation of the refractive index that can be handled analytically are discussed by T. R. Gilliland, G. W. Kenrick, and K. A. Norton, Investigations of Kennelly-Heaviside Layer Heights for Frequencies between 1600 and 8650 Kilocycles per Second, *Proc. I.R.E.*, Vol. 20, p. 286, February, 1932.

[1] The literature giving experimentally determined characteristics of the ionosphere is almost limitless. For those who wish to follow this subject further, attention is called to the Reports of High-frequency Radio Transmission Conditions that appear monthly in *Proc. I.R.E.* Reference is also to be made to the following articles, all of a summary character, which are recommended for one wishing to start on the literature that is available: J. H. Dellinger, The Role of the Ionosphere in Radio-wave Propagation, *Trans. A.I.E.E.*, Supplement, Vol. 58, p. 803, 1939; Karl K. Darrow, Analysis of the Ionosphere, *Bell System Tech. Jour.*, Vol. 19, p. 455, July, 1940; L. V. Berkner and H. W. Wells, Report of Ionosphere Investigations at the Huancayo Magnetic Observatory (Peru) during 1933, *Proc. I.R.E.*, Vol. 22, p. 1102, September, 1934; S. S. Kirby and E. B. Judson, Recent Studies of the Ionosphere, *Proc. I.R.E.*, Vol. 23, p. 733, July, 1935; S. S. Kirby, L. V. Berkner, and D. M. Stuart, Studies of the Ionosphere and Their Application to Radio Transmission, *Proc. I.R.E.*, Vol. 22, p. 481, April, 1934; J. P. Schafer and W. M. Goodall, Diurnal and Seasonal Variations in the Ionosphere during the Years 1933 and 1934, *Proc. I.R.E.*, Vol. 23, p. 670, June, 1935; T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, Characteristics of the Ionosphere and Their Application to Radio Transmission, *Proc. I.R.E.*, Vol. 25, p. 823, July, 1937; N. Smith. T. R. Gilliland, and S. S. Kirby, Trends of Characteristics of the Ionosphere for Half a Sunspot Cycle, *Jour. Research Nat. Bur. Standards*, Vol. 21, p. 835, December, 1938.

[2] See S. K. Mitra and P. Syam, *Nature*, Vol. 135, p. 953, 1935; R. C. Colwell and A. W. Friend, *Nature*, Vol. 137, p. 782, 1936; J. H. Piddington, The Scattering of Radio Waves in the Lower and Middle Atmosphere, *Proc. I.R.E.*, Vol. 27, p. 753, December, 1939.

[3] The E layer ordinarily returns only the ordinary ray, but the F, F_1, and F_2 layers commonly return the extraordinary ray as well as the ordinary ray, and so have two slightly different critical frequencies, the higher being for the extraordinary ray.

The critical frequencies under these conditions are designated as f_E^o, $f_{F_2}^x$, etc., where the subscript denotes the layer involved and the superscript indicates the type of ray.

$$\left.\begin{matrix}\text{Critical frequency } f_E \\ \text{of } E \text{ layer}\end{matrix}\right\} = K\sqrt[4]{\cos\psi} \tag{48}$$

where ψ is the zenith angle of the sun and depends upon the time of day and the season of the year and K is a factor determined by the intensity of solar radiation, which varies in a random way but has no relationship to time of day or season of the year. The relation expressed by Eq. (48) holds within fairly close limits during the daylight

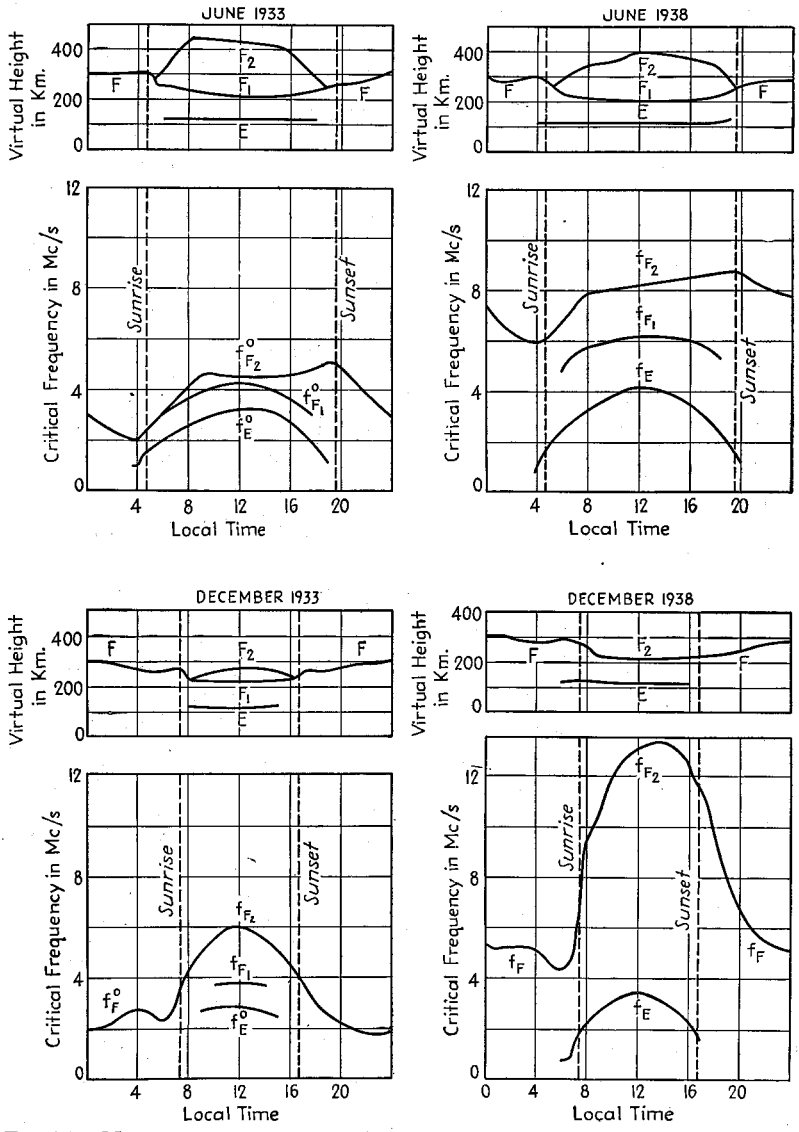


FIG. 35.—Monthly averages of critical frequencies and virtual heights as a function of time of day for winter and summer and for two parts of the 11-year sunspot cycle. These results were obtained at Washington, D.C.

hours, provided that ψ is not too close to 90°. The relationship expressed by Eq. (48) indicates that the daytime ionization of the E layer is produced by ultraviolet or similar radiation from the sun, and that the recombination rate is relatively rapid. At night the critical frequency of the E layer is below one megacycle, and scarcely any experimental data are available giving exact values to be expected.

The F_1 layer exists only in the daytime and has a virtual height of the order of 225 km during the middle of the day, a value that is rather consistently maintained from season to season. The critical frequency of the F_1 layer has a diurnal variation similar in character to that of the E layer, with the maximum value occurring at noon and with winter values usually being less than summer values. The close relationship between critical frequency and zenith angle represented by Eq. (48) for the E layer does not exist for the F_1 layer, however. The F_1 layer is always present in the summer, but does not necessarily exist during the winter days, particularly during the more active part of the 11-year solar cycle.

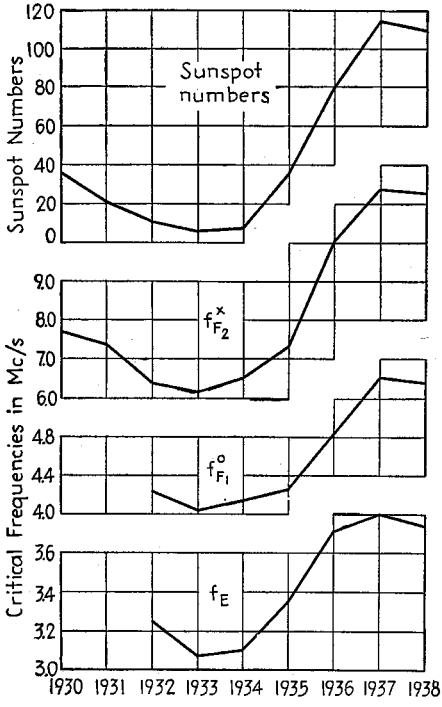


FIG. 36.—Comparison of annual averages of critical frequencies and sunspot numbers.

The F_2 layer varies greatly in both virtual height and critical frequency during the day, and from season to season. The virtual height in the summer ranges from 300 to 400 km, while in winter the virtual height may be as low as 225 km. The critical frequency of the F_2 layer is much higher in winter than in summer.[1] In winter the maximum critical frequency tends to occur around noon, while the maximum is in the late afternoon in summer.

Around sunset the F_1 layer disappears, but the F_2 layer continues through the night, when it is designated as the F layer. During the period of darkness the height of this layer averages about 300 km. After sunset, the critical frequency drops off rapidly from the daytime F_2 value, and reaches a minimum between midnight and sunrise. The minimum value of the night critical frequency is somewhat lower in the winter than in summer.

Year-to-year Variations in the Ionosphere.—The virtual heights of the various layers of the ionosphere do not vary appreciably from year to year (see Fig. 35). However, the critical frequencies show long-time variations that follow closely the 11-year cycle of the sun ultraviolet radiations (sometimes termed the 11-year sunspot cycle). There is a progressive increase in the critical frequencies of the various layers as the sunspot cycle advances, and yearly averages of critical frequencies and sunspot numbers show very close correlation (Fig. 36). There is also a definite relation between

[1] Monthly averages of critical frequency of the F_2 layer show definite correlation with the character figure for central zone calcium flocculi of the sun, and some but less pronounced relation with sunspot numbers. See W. M. Goodall, The Solar Cycle and the F_2 Region of the Ionosphere, *Proc. I.R.E.*, Vol. 27, p. 701, November, 1939.

monthly averages of sunspot numbers, and of critical frequencies for the E layer (Fig. 37), provided that the E-layer critical frequencies are corrected by Eq. (48) to take into account the seasonal variations in the zenith angle of the sun.

Another year-to-year effect of importance is that as the solar cycle advances, there is a tendency for the daytime F_1 layer to fail to appear in the winter.

Ionosphere Variations with Latitude and Longitude.—The discussion of normal ionosphere characteristics given above is based mainly on results obtained in the vicinity of Washington, D.C., where the latitude is 39 degrees north. Scattering results obtained at other locations in the Northern Hemisphere indicate that longitude has relatively little effect, and that curves of ionosphere characteristics plotted as a function of local sun time are very much the same, irrespective of longitude.

Latitude, however, is quite important, since the varying angle of incidence of solar radiation with latitude leads to greater ionization in equatorial regions as compared with polar regions. Measurements taken near the equator indicate that the average critical frequencies are somewhat higher and the average virtual height somewhat lower than at Washington, D.C. There is also very little seasonal variation near the equator. In contrast with this, measurements taken at latitude 70 degrees north

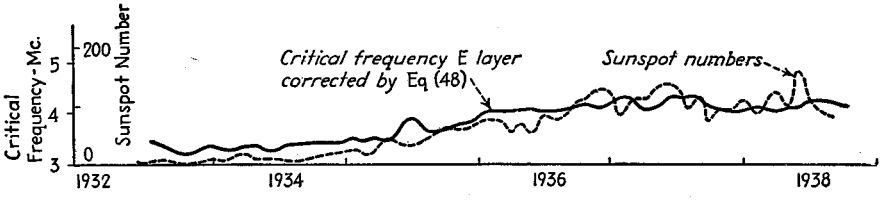


FIG. 37.—Correlation between monthly averages of sunspot numbers and of the E-layer critical frequency for one-half of a solar rotation.

show lower average critical frequencies and greater variations with season than are present at Washington, D.C. Results obtained in the Southern Hemisphere, except for the F_2 layer, are in general similar to those for the Northern Hemisphere when the reversal of seasons is taken into account. With the F_2 layer, however, the action is the same in both hemispheres (*i.e.*, there is no reversal of the seasons).

10. Abnormal Ionosphere Behavior.—While the characteristics discussed in Par. 9 are those normally observed, the ionosphere under certain conditions exhibits irregularities of various types, as discussed below.

Sporadic E-layer Reflections.—Waves are sometimes returned to earth by the E layer at frequencies greater than the critical frequency of the layer.[1] These are termed *sporadic* E-layer reflections and are thought to be the result of partial reflections from a sharp boundary of stratified ionization, according to the mechanism discussed in connection with Fig. 34. These reflections are apparently the result of electron (or ion) clouds or patches in the E layer, which have very sharp boundaries. These clouds drift through space, with the result that the reflections come and go; hence the term "sporadic." The clouds may range from one kilometer across to several hundred kilometers in extent. Sporadic E-layer reflections occur most commonly in the summer, particularly at night, but are found to some extent at any time of day or night and occasionally at all seasons. Sporadic E-layer reflections tend to be more frequent during the active part of the sunspot cycle, but the correlation is not very close, and neither is the occurrence of these reflections correlated with other types

[1] The critical frequency is defined here as the frequency at which waves are just barely returned from the layer, with a large virtual height as the frequency is increased. This frequency is quite sharply defined, whereas the sporadic E reflections are not critical with frequency.

of ionosphere irregularities. There is some evidence that sporadic reflections are more common at high latitudes than at the equator.[1]

Sporadic *E*-layer reflections tend to be less intense the higher the frequency, but under special circumstances it is possible to observe sporadic *E*-layer reflections with vertical incidence at frequencies up to about 12 megacycles. Under these conditions, and because of the low height of the *E* layer, it is possible to obtain long-distance sky-wave communication at frequencies as high as 60 megacycles. This frequency is about twice as high as can be returned from any of the layers by normal refraction processes, even when the angle of incidence at the ionosphere has the lowest value possible in view of the curvature of the earth and the layer height.

Sudden Ionosphere Disturbances—Fade-outs.[2]—Sky-wave signals sometimes suddenly disappear. This phenomenon, often termed the *Dellinger effect,* is the result of a burst of ionizing radiation from a solar eruption on the sun that causes a sudden abnormal increase in the ionization of the portion of the ionosphere below the *E* layer. This results in a great increase in the absorption that waves undergo in entering and leaving the ionized region. Such a radio fade-out is usually complete within less than a minute, and lasts from ten minutes to several hours. The effect occurs simultaneously throughout the portion of the earth illuminated by the sun, and does not occur at night. The action of a fade-out is more intense, and the duration longer, the lower the frequency, at least for frequencies above 1,500 kc. The intensity of the disturbance tends to be greatest in the region where the sun's radiation is perpendicular, *i.e.*, is greater at noon than at other times of day, and greater at the equator than at higher latitudes.

There is no seasonal variation in the occurrence of radio fade-outs, but there is apparently a tendency for such fade-outs to be more numerous in those years characterized by high sunspot numbers.

Radio fade-outs occur only when there is a solar eruption as observed with the aid of a telescope. The converse does not necessarily follow, however, since most solar eruptions fail to produce even a mild radio fade-out.

Prolonged Periods of Low-layer Absorption.[3]—This phenomenon is similar in all respects to the sudden ionosphere disturbances associated with radio fade-outs except that the beginning as well as the recovery is gradual, and the duration is much longer, commonly several hours. Also the diminution in intensity of the sky wave is generally less severe than in the more intense fade-outs, although sometimes the waves are completely absorbed.

It appears that low-layer absorption arises from the same sort of ionizing radiation from the sun that accounts for the radio fade-outs, but is the result of a great outpouring of such radiation extending over a period of time rather than from a sudden burst.

Periods of low-layer absorption occur only in daytime, and are most intense for the radio waves of medium high frequency. They frequently occur during periods when radio fade-outs are numerous.

Ionosphere Storms.[4]—An ionosphere storm is characterized by poor radio transmission at frequencies above 500 kc, lasts for one or more days, and is usually accom-

[1] L. V. Berkner and H. W. Wells, Abnormal Ionization of the *E*-region of the Ionosphere, *Terr. Mag. and Atmos. Elec.*, March, 1937.

[2] J. H. Dellinger, Sudden Disturbances of the Ionosphere, *Proc. I.R.E.*, Vol. 25, p. 1253, October, 1937; L. V. Berkner and H. W. Wells, Study of Radio Fade-outs, *Terr. Mag. and Atmos. Elec.*, June, 1937; L. V. Berkner, Concerning the Nature of Radio Fade-out, *Phys. Rev.*, Vol. 55, p. 536, Mar. 15, 1939.

[3] T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, Characteristics of the Ionosphere at Washington, D.C., December, 1937, *Proc. I.R.E.*, Vol. 26, p. 236, February, 1938.

[4] T. R. Gilliland, S. S. Kirby, and N. Smith, Characteristics of the Ionosphere at Washington, D.C., June, 1938, *Proc. I.R.E.*, Vol. 26, p. 1033, August, 1938.

panied by a magnetic storm.[1] An ionosphere storm begins with a turbulent phase, which consists of a violent turbulence of the entire ionosphere in the auroral zone. This causes the normal stratification of the ionosphere to be destroyed, with the production of small clouds of ionization that move in an irregular manner. The turbulent period is followed by a moderate phase, in which the effects initiated in the auroral zone gradually extend to much lower latitudes by expansion and diffusion of the higher F regions. The ionosphere then gradually returns to normal after a period that in severe cases may be several days.

During the turbulent phase of an ionosphere storm, layers with definite virtual heights and critical frequencies tend to disappear. During the moderate phase that follows, the virtual heights are abnormally great, and the critical frequencies unusually low, particularly for the F and F_2 layers. There is also a tendency for the absorption to be greater than normal.

The effects of ionosphere storms are greater near the polar regions and become negligible at the equator. The turbulent phase is usually confined to within about 20° of the magnetic poles, and even in the case of very severe storms does not extend beyond about 50° from the magnetic pole. The moderate phase is the one normally observed in temperate latitudes, and lags some hours behind the initial severe disturbances observed in the polar areas. Unlike radio fade-outs and periods of low-layer absorption, ionosphere storms may occur at night as well as in the daytime. The frequency with which ionosphere storms occur correlates reasonably closely with solar activity, and so follows the 11-year solar cycle as do sunspots, the index for magnetic character of the days, etc.

11. Methods of Ionosphere Investigation. *Pulse Method.*[2]—In this method short wave trains lasting perhaps 10^{-4} seconds are transmitted. A receiver located in the immediate vicinity of the transmitter (commonly only a few feet away) is then used to pick up the transmitted waves and any wave trains returned from the ionosphere. The output of the receiver is applied to some form of oscillograph, usually a cathode-ray tube but in some cases a magnetic oscillograph, and the oscillograph is provided with a time axis that is synchronized with the pulses sent out by the transmitter, usually by controlling both the time axis and the transmitted pulses from the same 60-cycle power source.

When the time axis is horizontal, and the received signals produce vertical deflection, the oscillogram obtained has the character shown in Fig. 38*a*. The first received pulse represents the transmitted wave train, while the second is a wave train that has traveled up to the ionosphere and been refracted back to earth. The time interval between these two pulses is a measure of the path length and determines the virtual height of the layer. Additional pulses, when present, represent either waves returned from higher layers or waves that have made more than one round trip between earth and the first layer. When the received record is to be photographed continuously it is desirable to modify the preceding arrangement by biasing the control electrode of the cathode-ray oscillograph tube just enough negatively to turn off the light spot, and then superimposing the output of the receiver upon this bias.[3] By proper arrangement of polarities, an incoming pulse will turn the spot on, giving a pattern such as is illustrated in Fig. 38*b*. In such an arrangement it is very convenient to

[1] At frequencies below 500 kc an ionosphere storm may improve radio transmission.

It is frequently said that radio reception of short waves is poor when there is a magnetic storm. This is a result of the close connection between ionosphere and magnetic storms.

[2] This method was first proposed by G. Breit and M. Tuve, A Test of the Existence of the Conducting Layer, *Phys. Rev.*, Vol. 28, p. 554, September, 1926.

[3] The first use of this general method was by L. C. Verman, S. T. Char, and Aijaz Mohammed, Continuous Recording of Retardation and Intensity of Echoes from the Ionosphere, *Proc. I.R.E.*, Vol. 22, p. 906, July, 1934.

superimpose upon the control electrode bias, very short positive pulses having a frequency of 3,000 cycles, synchronized with the timing wave of the oscillograph. This will cause spots to appear at intervals along the time (or height) axis representing 50 km of virtual height, and so automatically will calibrate the height scale.

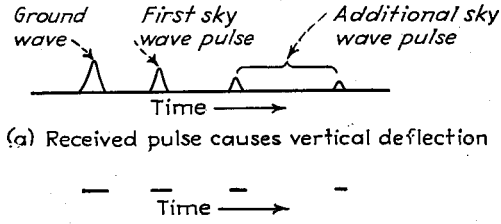


FIG. 38.—Example of oscillograms received when a transmitter is within ground-wave range, and the transmitted signal consists of a short wave train. The additional sky-wave pulses may be either multiple reflections of the first pulse or waves returned from higher layers. In many cases the additional sky-wave pulses are absent.

Equipment has been developed that sends 10 to 60 pulses per second, with the frequency of the pulses gradually changing so that a frequency range of perhaps 1,000 to 10,000 kc is swept through in a period of perhaps 15 minutes.[1] In receiving such multifrequency records, the vertical deflection is obtained from a timing wave synchronized with the transmitted pulses in such a way that deflections along the vertical

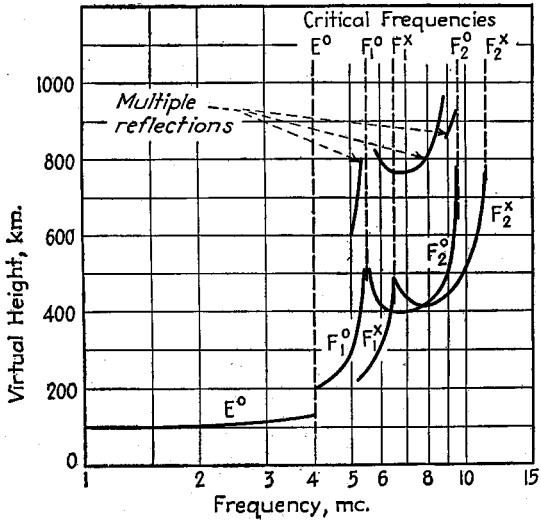


FIG. 39.—Record of virtual height as a function of frequency as obtained with a multifrequency recorder. The superscripts 0 and x denote ordinary and extraordinary rays, respectively.

are a measure of virtual height. Deflections along the horizontal are obtained from a voltage controlled by the same shaft that varies the transmitter frequency, so that

[1] T. R. Gilliland and A. S. Taylor, Field Equipment for Ionosphere Measurements, *Jour. Research Nat. Bur. Standards*, Vol. 26, May, 1941; T. R. Gilliland, Note on a Multifrequency Automatic Recorder of Ionosphere Heights, *Proc. I.R.E.*, Vol. 22, p. 236, February, 1934.

distance along the horizontal represents transmitted frequency. A typical record obtained in this way is illustrated in Fig. 39,[1] and has the advantage of making it possible to evaluate the critical frequencies of the various layers. These critical frequencies are marked in Fig. 39, and are indicated either by a sudden jump in the virtual height corresponding to refraction that shifts from one layer to a higher layer, as in the case of the E critical frequency in Fig. 39, or by a crinkle in the curve, as with the F_1 and F_2 layers. The large virtual height associated with these crinkles arises from the very low group velocity of a wave train traveling in an ionized medium when the frequency is close to the critical frequency. Multifrequency records make it a simple matter to separate the ordinary and extraordinary rays, and they also make it possible to distinguish multiple reflections from reflections produced by a

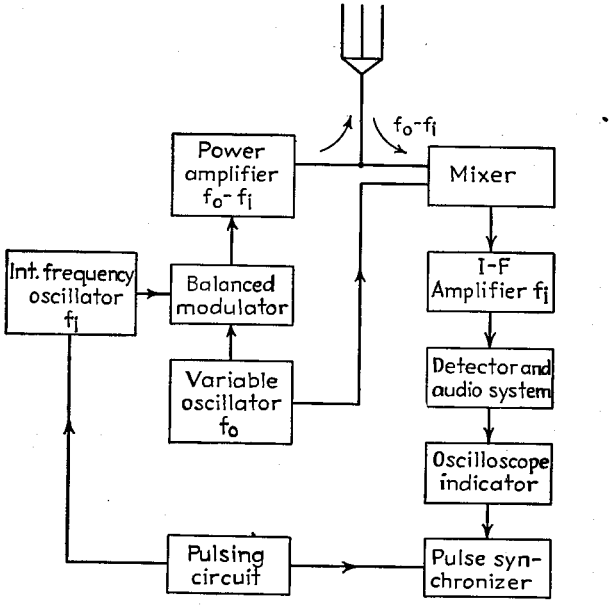


FIG. 40.—Interlocking system for producing and receiving pulses for ionosphere investigations.

higher layer. For example, in Fig. 39 the reflections corresponding to a virtual height of the order of 800 km are obviously multiples of the F_1 and F_2 layers since the critical frequencies are the same as for the F_1 and F_2 layers. This would be the case only by pure chance if the reflection was produced by an independent layer.

In making continuous records of the ionosphere by the pulse method it is necessary that the receiver stay accurately tuned to the transmitter frequency over long periods of time. This problem is particularly difficult in the case of multifrequency records, where the receiver must always be in tune with the frequency being transmitted at the particular moment, and yet at the same time the transmitter frequency is continually varying. This situation can be handled by the interlocking system of producing and receiving pulses shown schematically in Fig. 40.[2] Here a fixed-frequency oscillator operating at the intermediate frequency of the receiver (about 475 kc) is pulsed by

[1] From Darrow, *op. cit.*

[2] This is due to T. R. Gilliland. See Gilliland and Taylor, *loc. cit.*, for further discussion of such equipment, or T. R. Gilliland, Ionospheric Investigation, *Nature*, Vol. 134, p. 379, September, 1934.

means of a thyratron or other convenient arrangement. These pulses are then applied to a mixer tube, to which there is also applied voltage from a continuously operating oscillator that generates a frequency f_0 that is variable in the case of the multifrequency equipment, or fixed if operation is to be at a fixed frequency. The difference frequency produced by the mixer is selected by a tuned circuit in the output of the mixer, applied to a power amplifier, and then delivered to the transmitting antenna. At the same time output from the continuously operating oscillator f_0 is applied to the first detector of the receiver. The detector hence always operates to give the desired intermediate frequency irrespective of the transmitted frequency, provided only that the intermediate-frequency oscillator maintains the assigned frequency to within a few kilocycles. It will be noted that there is no oscillator in the system operating at the frequency that is to be received, and furthermore the intermediate-frequency oscillator operates only when pulses are being transmitted, and so does not interfere with the reception of the relatively weak pulses returned with a time delay by the ionosphere.

The receiver in a pulse system is subjected to very high overloads by the transmitted pulse, and must be so designed that these overloads will not cause damage. It is also necessary that in the interval between the end of the transmitted pulse and the time the first ionosphere reflection returns, the receiver be able to recuperate from any overloading and be restored to normal sensitivity. The principal requirement for achieving this result is so to proportion all resistance-capacity combinations in the receiver that they have a small value of RC. Abnormal charges received by condensers as a result of the action of the transmitted pulse can then be discharged very quickly.

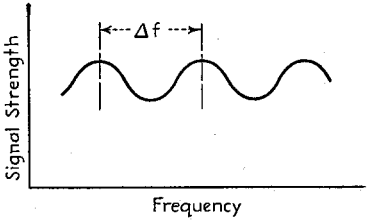


FIG. 41.—Interference fringes produced at receiving point when the transmitter frequency is varied under conditions where there is only one sky wave of somewhat smaller intensity than the ground wave.

Miscellaneous Methods of Studying the Ionosphere.—While the pulse method of Breit and Tuve is the most widely used method of investigating the ionosphere, other techniques have been employed and have their uses.

In one of these the frequency of an unmodulated carrier is varied slowly, and the variations in the signal strength noted at a receiving point where the strength of the sky and ground waves are of the same order of intensity. These variations are characterized by regular maxima and minima as shown in Fig. 41. From a knowledge of the frequency increment Δf required to go from one maximum to the next, together with a knowledge of the distance to the transmitter, one can determine the length of the path followed by the sky wave (on the basis that the sky wave travels with the velocity of light), and accordingly can deduce the virtual height from the triangle TAR of Fig. 33. This method was used by Appleton and Barnett in obtaining the first experimental proof of the existence of the ionosphere.[1] It has a very serious practical limitation, however, in that if there is more than one sky-wave reflection, as is very commonly the case, the interference pattern produced at the receiving point with change in frequency becomes so complicated that it is difficult to interpret. The method is accordingly not generally suitable for high-frequency investigations, but it is probably the best of all methods for studying the ionosphere at frequencies so low that pulses of the shortness necessary for resolving an ionosphere layer cannot be produced.[2]

[1] E. V. Appleton and M. A. S. Barnett, *Nature*, Vol. 115, p. 333, 1925.

[2] An example of an application to a moderately low frequency is given by C. R. Smith, Indirect Ray Measurements on the Droitwich Transmitter, *Wireless Eng.*, Vol. 14, p. 537, October, 1937.

Another method of determining the virtual height of the ionosphere is to observe the interference patterns produced at the ground between the ground wave and the sky wave as the distance to the transmitter is varied. As this distance increases the average field intensity will diminish, but superposed upon this major trend will be oscillations as shown in Fig. 42. The maxima of these correspond to points at which the ground and sky waves are in the same phase, with the minima corresponding to points where they are of opposite phase. From this pattern one can readily determine the virtual height of the ionosphere, and likewise the reflection coefficient of the wave at the ionosphere.[1] This method is suitable only at the lower radio frequencies where the wave paths are stable (*i.e.*, no fading).

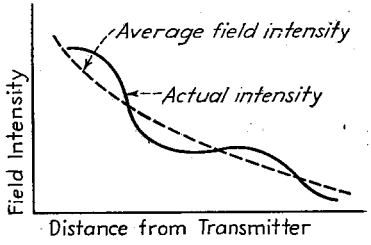


FIG. 42.—A schematic diagram illustrating the way in which field intensity can be expected to vary with distance at moderate distances from a long-wave transmitter, showing how the actual intensity deviates from the average field intensity as a result of interference between a down-coming sky wave and a somewhat stronger ground wave.

Still another method of determining virtual height involves measuring the angle with which the down-coming sky wave strikes the earth at a receiving point a moderate distance from the transmitter. From this angle, and the distance to the transmitter, one can calculate the height of the equivalent triangle that gives the virtual height, as shown in Fig. 33.

Theoretical studies show that the virtual heights measured in these three ways are all exactly the same virtual height obtained by the pulse method. The various methods are therefore equivalent, and the choice between them is a matter of convenience.[2]

Polarization Characteristics of Down-coming Waves—Elliptical Polarization.—The polarization characteristics of down-coming sky waves can be determined by observing the interference fringes obtained from three spaced-loop receivers as the transmitter frequency is varied. This gives sufficient information to determine the vertically and horizontally polarized components, the principal axes of an elliptically polarized wave,[3] and the sense of the polarization.

[1] Examples of such studies are given by J. Hollingworth, Propagation of Radio Waves, *Jour. I.E.E.*, Vol. 64, p. 579, May, 1926; J. D. Best, J. A. Ratcliffe, and M. V. Wilkes, Experimental Investigations with Very Long Waves Reflected from the Ionosphere, *Proc. Roy. Soc. (London)*, Vol. 156, p. 614, September, 1936.

[2] See J. C. Schelleng, Note on the Determination of the Ionization of the Upper Atmosphere, *Proc. I.R.E.*, Vol. 16, p. 1471, November, 1928; also, *Proc. I.R.E.*, Vol. 17, p. 1313, August, 1929; E. V. Appleton, Some Notes on Wireless Methods of Investigating the Electrical Structure of the Upper Atmosphere, *Proc. Phys. Soc. London*, Vol. 41, Pt. II, p. 43, December, 1928.

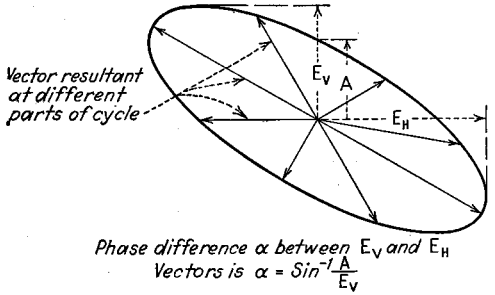


FIG. 43.—Diagram showing electric field of an elliptically polarized wave when the wave front is in the plane of the paper.

[3] Elliptical polarization results whenever there are vertical and horizontal components having different phase. Elliptical polarization is characterized by the fact that the resultant magnetic and electrostatic fields of the wave never at any instant pass through zero; rather, the resultant fields rotate in the plane of the wave front at a rate corresponding to the frequency of the wave, while at the same time pulsating in amplitude. The resultant field produced by elliptical polarization can therefore be represented by a rotating vector of varying length as illustrated in Fig. 43. The field can never be zero because the vertical and horizontal components do not become zero at the same instant.

This technique has been used to verify the theoretical conclusion that the behavior of a wave in the ionosphere is influenced by the presence of the earth's magnetic field. In particular, theory indicates that the direction of rotation of polarization should be left-handed in the Southern Hemisphere and right-handed in the Northern Hemisphere as a result of the different orientation of the earth's magnetic field in the two hemispheres, and this has been verified experimentally.[1]

Field-strength Curves as a Means of Studying Ionosphere Conditions.[2]—A study of curves giving a continuous record of received field intensity from a particular transmitter, as shown in Fig. 44, will give much information concerning the ionosphere, particularly when these curves are correlated with results obtained at the same time on critical frequencies and virtual heights by means of pulse measurements. For example, the sudden increase in intensity of the received signal at 0700 in Fig. 44 indicates a shift from the *F* layer to the *E* layer, and the sudden drop in intensity just before 1800 marks a return to the *F* layer in the evening. From the virtual

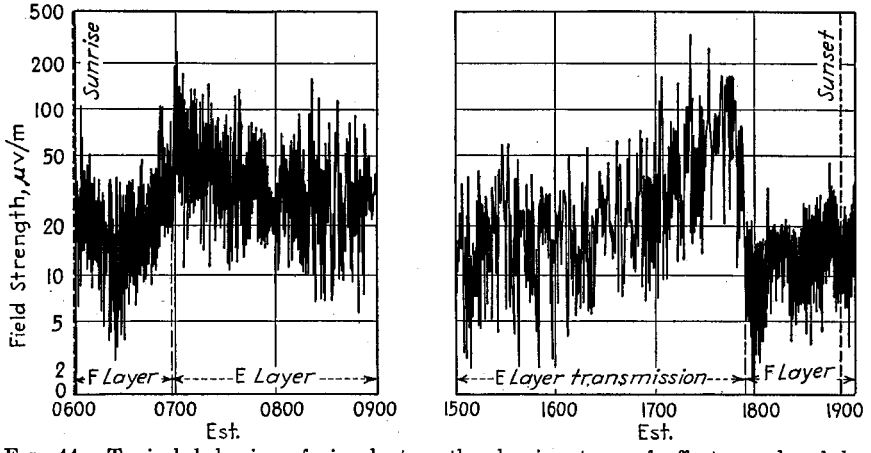


FIG. 44.—Typical behavior of signal strength, showing type of effects produced by ionosphere.

height of the *E* layer (110 km), the transmission distance, and the transmitter frequency (648 km and 6,060 mc, respectively, in Fig. 44) one can deduce that the critical frequency of the *E* layer for waves with vertical incidence must have been approximately 2,450 kc at 0700 and 1800.

It is possible in this manner to correlate many observations on field-strength curves with ionosphere data, or vice versa. When comprehensive information is simultaneously available on critical frequencies and virtual heights of the various layers, one can determine the causes for various changes that appear in the field-strength records, as occur, for example, when the transmission path shifts from one to two hops, when sporadic reflections are present, etc.

12. Ionosphere—Miscellaneous. *Eclipse Effects.*[3]—Numerous studies have been made to determine the effect of a solar eclipse on the ionosphere. An eclipse is found

[1] See E. V. Appleton and J. A. Ratcliffe, A Method of Determining the State of Polarization of Down-coming Waves, *Proc. Roy. Soc. (London)*, Vol. 117*A*, p. 576, March, 1928; A. L. Green, The Polarization of Sky Waves in the Southern Hemisphere, *Proc. I.R.E.*, Vol. 22, p. 324, March, 1934; W. G. Baker and A. L. Green, The Limiting Polarization of Down-coming Radio Waves Travelling Obliquely to the Earth's Magnetic Field, *Proc. I.R.E.*, Vol. 21, p. 1103, August, 1933.

[2] For further discussion see J. H. Dellinger, The Role of the Ionosphere in Radio-wave Propagation, *Trans. A.I.E.E.*, Supplement, Vol. 58, p. 803, 1939.

[3] S. S. Kirby, L. V. Berkner, T. R. Gilliland, and K. A. Norton, Radio Observations of the Bureau of Standards during the Solar Eclipse of August 31, 1932, *Proc. I.R.E.*, Vol. 22, p. 247, February, 1934; E.

to cause a decrease in the critical frequency of E and F_1 layers, with the variation being approximately in phase with the optical eclipse. This indicates that the ionizing agency in these regions is of solar origin and has a speed closely agreeing with that of light.

The effect of the eclipse upon the F_2 region is less marked than upon the lower regions, and it is not yet certain whether the effects produced by an eclipse on this region travel from the sun with the speed of light or at a very much lower velocity, as would be the case if corpuscles were involved.

Effect of Meteors on the Ionosphere.—It has been established that meteors can produce sufficient ionization in the upper atmosphere to effect wave propagation, and the actual effect of individual meteors on the E region has been observed.[1]

It has been suggested that meteors might account for sporadic E-layer effects, and also that the ionization that persists in the ionosphere at night might be attributable, at least in part, to the ionizing effect of the minute meteors that are continually being received by the earth.[2]

Scattering of Radio Waves and Related Deviations from the Great-circle Path.[3]—Scattered reflections are frequently produced in the ionosphere as a result of irregularities in the electron distribution. These reflections are usually weak, irregular or sporadic in character, and diffused. They may be produced in any of the layers and in some cases become quite complex as a result of interaction between layers. The virtual heights depend upon the type of scattering involved and may range from 90 km to values in excess of 1,500 km.

Scattered reflections are also produced when a sky wave strikes the earth after refraction by the ionosphere. Such reflected waves then propagate away from the scattering point in the same way as would waves produced by a transmitting antenna. The signals produced in this way are commonly of the order of 40 db lower in intensity than the main wave.

Scattering can cause a number of effects of importance. A scattered signal may, for example, be returned inside the skip region. Also, if the transmitting antenna is sharply directional, an observer considerably to one side of the path of the main beam may be able to receive the signal, although perhaps weakly, as a result of scattered waves that travel at right angles to the main beam. When bearings are taken upon the received signal under such conditions it is found that the apparent direction of the transmitter tends to be at right angles to the actual direction of the main beam, as illustrated in Fig. 45.

It is sometimes found possible to deliver better signals to a given receiving point by making use of scattering than can be obtained by direct transmission. Thus when short-wave signals are transmitted across the North Atlantic it is found that during severe magnetic storms (with corresponding severe ionosphere storms) it is impossible to produce an audible signal at the receiving point by aiming the transmitted energy along the great-circle path. This is because the great-circle path passes near the polar regions, and the absorption during a magnetic storm is practically complete.

V. Appleton and S. Chapman, Report on Ionization Changes during a Solar Eclipse, *Proc. I.R.E.*, Vol. 23, p. 658, June, 1935; S. S. Kirby, T. R. Gilliland, and E. B. Judson, Ionosphere Studies during Partial Solar Eclipse of February 3, 1935, *Proc. I.R.E.*, Vol. 24, p. 1027, July, 1936.

[1] A. M. Skellett, The Ionizing Effects of Meteors, *Proc. I.R.E.*, Vol. 23, p. 132, February, 1935.

[2] J. A. Pierce, Abnormal Ionization in the E Region of the Ionosphere, *Proc. I.R.E.*, Vol. 26, p. 892, July, 1938.

[3] For further information on scattering see C. F. Edwards and K. G. Jansky, Measurements of the Delay and Direction of Arrival of Echoes from Nearby Short-wave Transmitters, *Proc. I.R.E.*, Vol. 29, p. 322, June, 1941; T. L. Eckersley, Analysis of the Effect of Scattering in Radio Transmission, *Jour. I.E.E.*, Vol. 86, p. 548, 1940; also Wireless Section, *I.E.E.*, Vol. 15, p. 74, June, 1940; J. B. Hoag and Victor J. Andrew, A Study of Short-time Multiple Signals, *Proc. I.R.E.*, Vol. 16, p. 1368, October, 1928; E. Quäck and H. Mögel, Short-range Echoes with Short Waves, *Proc. I.R.E.*, Vol. 17, p. 824, May, 1929.

However, if the transmitted signals under such conditions are aimed in a southerly direction, it is found that the path of the main beam is sufficiently far from the polar regions so that the absorption is only nominal, and some of this energy will reach moderately high latitudes through the action of scattering.[1] Bearings on such signals received at New York indicate a southeasterly direction of arrival. These effects in transatlantic communication ordinarily occur at night, and are most pronounced during ionosphere storms, but may be present to some extent at all times. These scattered signals are the cause of flutter fading.

The Luxemburg Effect.[2]—The action of the ionosphere upon a radio wave is not linear, *i.e.*, the effect is not proportional to the amplitude of the wave. As a result there is a possibility that two strong radio waves passing through the ionosphere simultaneously could interact to produce cross-modulation products. Theoretical work has indicated that the magnitude of this cross-modulation could under favorable conditions be sufficient to produce appreciable effect, and such interference has been reported by observers. It is termed the "Luxemburg effect" because the first such case reported involved Radio Luxemburg. However, since the Luxemburg effect has

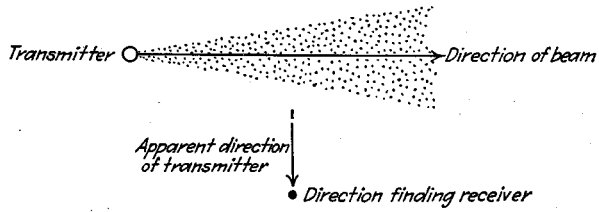


FIG. 45.—Scattering of beam signals in sidewise directions.

not been generally observed with most powerful transmitters, there is a possibility that external cross-modulation (see Par. 5, Sec. 9) has been the cause of interference ascribed to cross-modulation in the ionosphere.

Echo Signals with Very Long Time Delay.—Echo signals having a time delay of several seconds have been reported a number of times, and there are several authentic cases of echoes having a time lag of several minutes. The cause of such echoes has not definitely been established, but theoretical work indicates that retardation up to ten seconds could be accounted for by low group-velocity propagation in the upper part of the ionosphere. Signals of greater retardation could be accounted for only by waves that have traveled great distances in the empty space outside the earth's atmosphere, and that then by a fortuitous combination of circumstances are finally reflected back to earth by ionized regions either within the influence of the earth's magnetic field or in the vicinity of the sun.[3]

Correlation of Ionosphere and Signal Strength with Weather.—Various investigators have attempted to correlate ionosphere and radio propagation data with weather.[4] The results have in general, however, been disappointing, for although at times some

[1] C. B. Feldman, Deviations of Short Radio Waves from the London-New York Great-circle Path, *Proc. I.R.E.*, Vol. 27, p. 635, October, 1939.

[2] G. W. O. Howe, Accurate Measurements of the Luxemburg Effect, *Wireless Eng.*, Vol. 15, p. 187, April, 1938; V. A. Bailey and D. F. Martyn, The Interaction of Radio Waves, *Exp. Wireless and Wireless Eng.*, Vol. 12, p. 122, March, 1935.

[3] P. O. Pedersen, Wireless Echoes of Long Delay, *Proc. I.R.E.*, Vol. 17, p. 1750, October, 1929; G. Breit, Group-velocity and Long Retardations of Radio Echoes, *Proc. I.R.E.*, Vol. 17, p. 1508, September, 1929.

[4] For typical examples of such studies see I. J. Wymore-Shield, A Correlation of Long-wave Radio Field Intensity with Passage of Storms, *Proc. I.R.E.*, Vol. 19, p. 1675, September, 1931; Robert C. Colwell, Cyclones, and the Kennelly-Heaviside Layer, *Proc. I.R.E.*, Vol. 21, p. 721, May, 1933; Greenleaf W. Pickard, Some Correlations of Radio Reception with Atmospheric Temperature and Pressure, *Proc. I.R.E.*, Vol. 16, p. 765, June, 1928.

correlation seems to exist, this does not always persist indefinitely, and it is usually impossible for one investigator to duplicate the correlations observed by another investigator at a different location.

PROPAGATION OF RADIO WAVES OF DIFFERENT FREQUENCIES

13. Propagation Characteristics of Very Low-frequency Radio Waves (below 100 Kc).[1]—The propagation characteristics of radio waves of very low frequency are controlled by the following facts: (1) the ground-wave attenuation is very small, so that ground-wave signals of appreciable strength are produced up to distances of 1,000 km, particularly at the lowest frequencies; (2) the sky wave is able to penetrate only a very slight distance into the ionosphere, which acts almost like a reflecting mirror having a height of the order of 60 to 90 km; (3) the energy absorption at the ionosphere is relatively small and is less the lower the frequency, is less in winter than in summer, and during the night as compared with the day, and varies somewhat from year to year.

Low-frequency waves that have traveled considerable distance act as though they were propagated in the space between the two concentric reflecting spherical shells representing the earth and the lower edge of the ionosphere. The attenuation under such conditions is that caused by spreading, plus the energy lost at the earth's surface and at the edge of the ionosphere. The loss in the ionosphere has diurnal, seasonal, and year-to-year variations. The daylight signal strength of low-frequency radio waves over water is given with fair accuracy by the Austin-Cohen formula[2]

$$e = 377 \frac{hI}{\lambda d} \sqrt{\frac{\theta}{\sin \theta}}\, e^{\frac{-0.0015d}{\sqrt{\lambda}}} \tag{49}$$

where e = base of Napierian logarithms.
E = field strength, mv/m.
h = effective height of transmitting antenna, km.
λ = wave length, km.
d = distance, km, from transmitter.
θ = angle at center of earth intercepted by transmission path, in radians.
I = current flowing in the vertical part of the transmitting antenna, amp.

Curves showing typical diurnal, seasonal, and year-to-year variations in the strength of low-frequency radio signals from distant transmitters are given in Figs. 46 and 47. It will be noted that the variation between day and night strength increases the higher the frequency, and that in all cases the diurnal curves are char-

[1] The most comprehensive discussion of propagation at low frequencies available is that of L. Espenschied, C. N. Anderson, and A. Bailey, Trans-atlantic Radio Telephone Transmission, *Proc. I.R.E.*, Vol. 14, p. 7, February, 1926.

Attention is also called to the long series of field-strength measurements of transatlantic low-frequency signals reported in *Proc. I.R.E.* up to 1933.

[2] This formula was derived empirically but it has been found that an attenuation factor of the form $-\alpha d/\lambda^n$ has theoretical justification. Thus see G. N. Watson, The Transmission of Electric Waves around the Earth, *Proc. Roy. Soc. (London)*, A, Vol. 95, p. 546, July 15, 1919. The values used for α and n in Eq. (48) are those originally proposed by L. W. Austin, Some Quantitative Experiments in Long Distance Radio Telegraphy, *Bur. Standards Bull.*, Vol. 7, p. 315, 1911. Other investigators have found that other values sometimes fit data over a limited range of conditions with better accuracy, but on the whole, the original formula is found to be as good as or better than the modifications. A discussion of some of these variations is given by E. Yokoyama and T. Nakai, East-west and North-south Attenuations of Long Radio Waves on the Pacific, *Proc. I.R.E.*, Vol. 17, p. 1240, July, 1929. See also L. W. Austin, Preliminary Note on Proposed Changes in the Constants of the Austin-Cohen Transmission Formula, *Proc. I.R.E.*, Vol. 14, p. 377, June, 1926.

There is evidence that for overland transmission the proper value of α is 0.003 to 0.01. See H. Fassbender, F. Eisner, and G. Kurlbaum, Investigation of the Attenuation of Electromagnetic Waves and the Distances Reached by Radio Stations in the Wave Bands from 200 to 2000 Meters, *Proc. I.R.E.*, Vol. 19, p. 1446, August, 1931.

acterized by a period of minimum field intensity around sunset, with in some instances also a smaller dip in intensity at sunrise.[1]

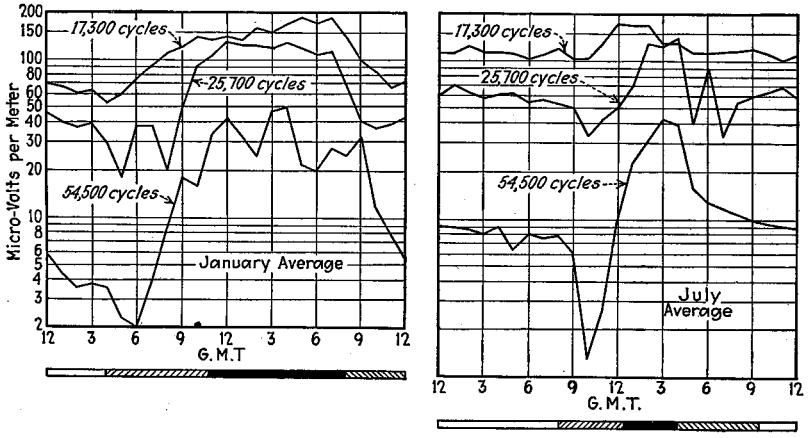


FIG. 46.—Curves showing average diurnal variation in strength of long-wave signals of different frequencies propagated across the North Atlantic during midwinter and midsummer months. (Note that signal strengths at different frequencies cannot be compared because the radiated power was not the same at all frequencies.) The solid and clear strips at the bottom of the figure indicate periods when the entire transmission path was in darkness and light, respectively, while the shaded strips indicate that part of the path was in darkness and part in light.

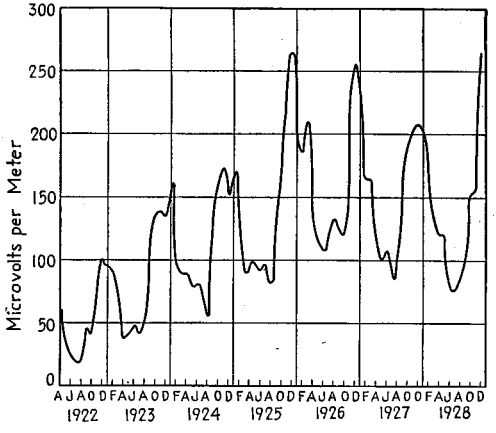


FIG. 47.—Variation in the monthly average of signal strength at Washington, D.C., of signals from 15.9-kc transmitter at Bordeaux, France.

Fading is never observed at the low radio frequencies. Any changes of signal strength that occur take place gradually.

[1] The characteristics of this sunset dip in signal strength have been studied in considerable detail by A. Bailey and A. E. Harper, Long-wave Radio Transmission Phenomena Associated with a Cessation of the Sun's Rays, *Bell System Tech. Jour.*, Vol. 15, p. 1, January, 1936.

It has been suggested that the sunset dip can be explained on the basis that the wave suffers reflection rather than refraction at the lower edge of the ionosphere, and that during the sunset period the conductivity and dielectric constant of the ionosphere undergo changes such that the pseudo-Brewster angle momentarily coincides with the angle of incidence and causes high absorption of vertically polar-

There is evidence to indicate that low-frequency radio signals propagate differently in an east-west direction than when traveling between north and south.[1] Factors that could cause this difference are the direction of the earth's magnetic field in relation to the transmission path, and the fact that in north-south transmission the sun time is substantially the same along the entire transmission path, a situation not realized with east-west transmission. Also in north-south operation the transmission path is always well away from the auroral zone.

At distances of a few hundred miles or less, the field strength is the resultant of a ground wave and a wave reflected from the ionosphere. Depending upon the distance, height of the reflecting layer, and the frequency, these two waves may either add or subtract, and their phase relationship will vary with the season and time of day. Thus the field strength of very low-frequency radio waves at moderate distances may be either stronger or weaker during daylight than at night, according to the way the relative phases of ground and sky waves happen to be.[2]

Effect of Solar Activity.[3]—Data on the field strength of transatlantic long-wave signals accumulated over many years show that the received field is definitely related to the eleven-year cycle of terrestrial magnetic activity, sunspots, solar limb prominences, and ultraviolet radiation. The correlation between yearly averages is particularly good, as shown in Fig. 48,[4] but with short periods, irregular effects tend to obscure the long-time trend, and there is very little day-by-day correspondence.

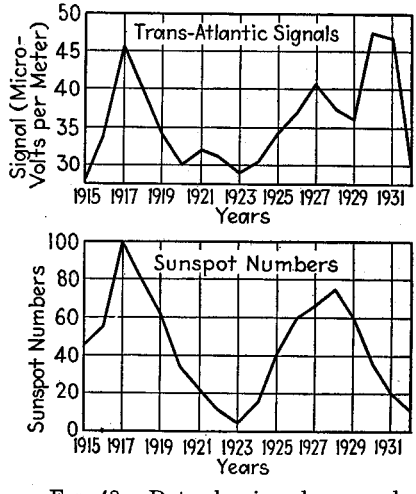


FIG. 48.—Data showing close correlation between sunspot numbers and yearly average of signal strength of long-wave signals received at Washington, D.C.

Magnetic storms have a very marked effect upon the propagation of low-frequency radio waves. A magnetic storm is characterized by a rapid and erratic fluctuation of the earth's magnetic field, which begins almost simultaneously over the earth with full intensity, and then gradually subsides in three or four days. During a magnetic storm the daytime field strength of low-frequency radio waves is increased somewhat above normal, the sunset drop in signal intensity disappears, and the night field is

ized waves. See S. Namba, General Theory of the Propagation of Radio Waves in the Ionized Layer of the Upper Atmosphere, *Proc. I.R.E.*, Vol. 21, p. 238, February, 1933. Such partial extinction of the vertically polarized component would account for the very weak signals observed at the earth's surface, since at these low frequencies the horizontally polarized field is of negligible strength at any reasonable height above the earth.

[1] Thus see Yokoyama and Nakai, *loc. cit.*

[2] For example, see P. A. de Mars, G. W. Kenrick, and G. W. Pickard, Low-frequency Radio Transmission, *Proc. I.R.E.*, Vol. 18, p. 1488, September, 1930.

[3] L. W. Austin, Solar Activity and Radiotelegraphy, *Proc. I.R.E.*, Vol. 20, p. 280, February, 1932; A. Bailey and H. M. Thomson, Transatlantic Long-wave Radio Telephone Transmission and Related Phenomena from 1923 to 1933, *Bell System Tech. Jour.*, Vol. 14, p. 680, October, 1935; C. N. Anderson, Correlation of Long Wave Transatlantic Radio Transmission with Other Factors Affected by Solar Activity, *Proc. I.R.E.*, Vol. 16, p. 297, March, 1928; K. Sreenivasan, On the Relation between Long-wave Reception and Certain Terrestrial and Solar Phenomena, *Proc. I.R.E.*, Vol. 17, p. 1793, October, 1929; E. Yokoyama and T. Nakai, Effects of Sun Spots and Terrestrial Magnetism on Long-distance Reception of Low-frequency Waves, *Proc. I.R.E.*, Vol. 19, p. 882, May, 1931. See also Espenschied, Anderson, and Bailey, *loc. cit.*

[4] From E. B. Judson, Low-frequency Radio Receiving Measurements at the Bureau of Standards in 1931 and 1932, *Proc. I.R.E.*, Vol. 21, p. 1354, September, 1933.

subnormal, approaching more nearly the day value. These effects are illustrated in Fig. 49, and are more pronounced as the frequency of transmission is increased.

14. Propagation Characteristics of Radio Waves of Moderately Low Frequency (100 to 500 kc).[1]—Radio waves in the frequency range 100 to 500 kc show the same general propagation characteristics as lower frequency radio waves, but with certain features accentuated. Thus the ground wave dies away somewhat more rapidly, with the result that it is necessary to depend almost solely upon the sky wave for propagation over considerable distances. The sky-wave absorption, while quite small at night, is relatively large in the daytime. The daytime absorption increases rapidly with frequency, and at the higher frequencies in the range 100 to 500 kc is sufficient to make long-distance daytime communication impossible, at least in the summer, when the absorption is much greater than in the winter. Daytime signals over water are given with at least fair accuracy by Eq. (49).[2]

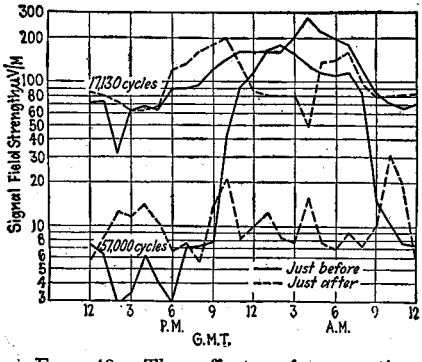


FIG. 49.—The effects of magnetic storms on low-frequency waves, showing how the day field strength is increased and the night field strength reduced by the magnetic storm.

Solar activity and magnetic storms presumably produce the same general effects at these frequencies as at lower frequencies, but the data on this point are very meager.

15. Propagation Characteristics of Broadcast Waves (500 to 1,500 Kc). *Field Strength Required for Broadcast Service.*—Practical propagation problems involving broadcast waves are dominated by the fact that the primary objective of a broadcast station is to deliver to the receiver a signal strong enough to override all ordinary interference, and that is as free as possible from fading, distortion, etc. This is in contrast with most circumstances where the primary objective is to deliver an understandable but not necessarily interference-free and entirely distortionless signal to the receiver.

The highest grade broadcast service (primary service) can be rendered only when the ground wave is relatively strong compared with interference present, and also has a field strength of at least several times the strength of the sky wave at the receiving point.

A lower grade of service (secondary coverage) can be provided relatively distant receivers at night through the medium of the sky wave. Secondary coverage obtained from the sky wave is of distinctly lower quality than primary coverage, since the sky wave always fades in intensity, and is accompanied by at least some selective fading that produces distortion.

The field strengths required for satisfactory primary and secondary coverage depend upon the interference level, which, in turn, is dependent upon the local conditions at the receiving point. This interference level is in general greatest in the business and industrial areas of large cities, is less in residential sections of large cities, still less in small cities, and least of all in rural areas. Values of field strength considered necessary for primary coverage under various types of average conditions are given in Table 3. In areas where local conditions are more favorable than the average, primary coverage will be obtained with weaker fields than specified, and in all cases some coverage of an intermittent character that varies with the hour-to-hour intensity of local interference can always be obtained with very much weaker fields than those

[1] Curves of field strength as a function of distance at these frequencies are given by Fassbender, Eisner, and Kurlbaum, *loc. cit.*

[2] Thus see C. N. Anderson, Notes on Radio Transmission, *Proc. I.R.E.*, Vol. 19, p. 1150, July, 1931.

in the table. Secondary coverage via the sky wave requires field intensities of 0.5 mv/m or more under average conditions, except that in rural areas fields of as low as 50 μv/m will often give signals that are considered of value by the listener.

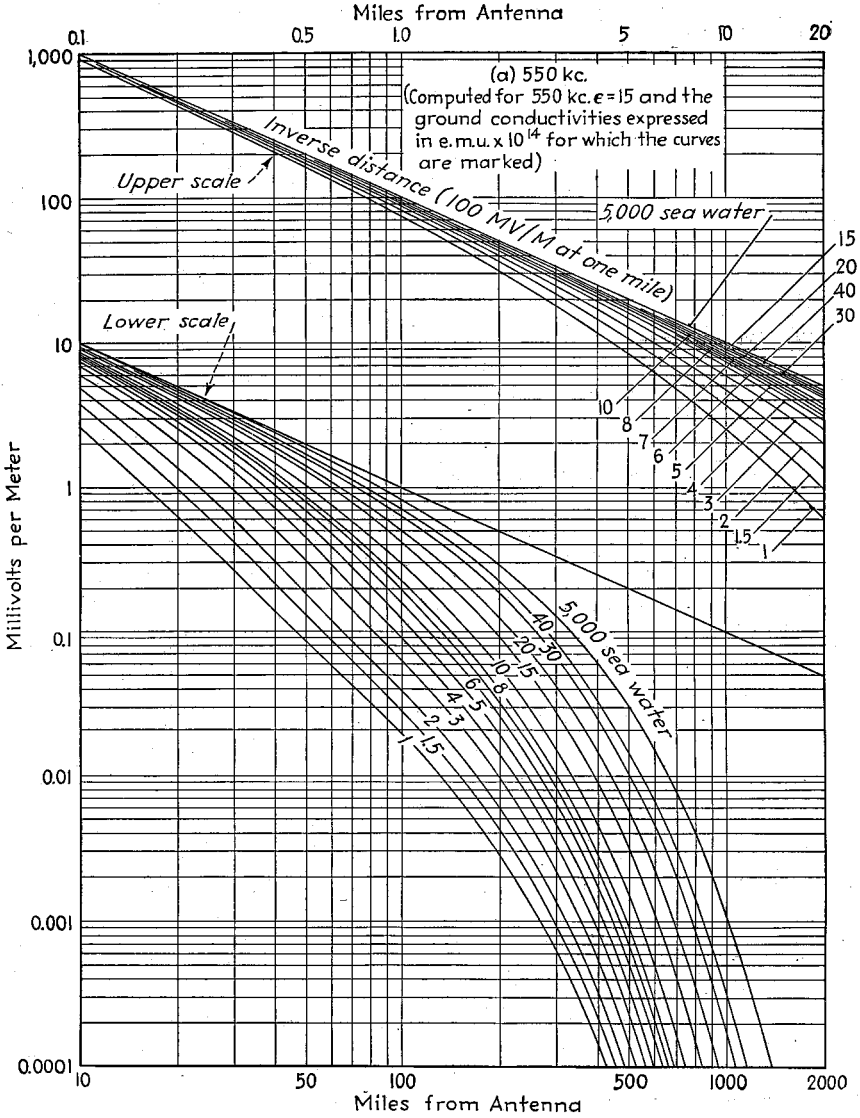


FIG. 50.—Curves of ground-wave strength at broadcast frequencies. These curves are for the case where the field strength at one mile is 100 mv/m.

Ground Wave at Broadcast Frequencies.—The primary coverage of a broadcast station results from the ground wave at the surface of the earth, and so is determined by the surface wave discussed in Pars. 1 and 2. The curve of ground-wave intensity as a function of distance for a given transmitter power depends upon the earth conductivity, the frequency, and the directivity of the transmitting antenna. The

dielectric constant of the earth is not important at broadcast frequencies since at these frequencies the earth acts substantially as a conductor.

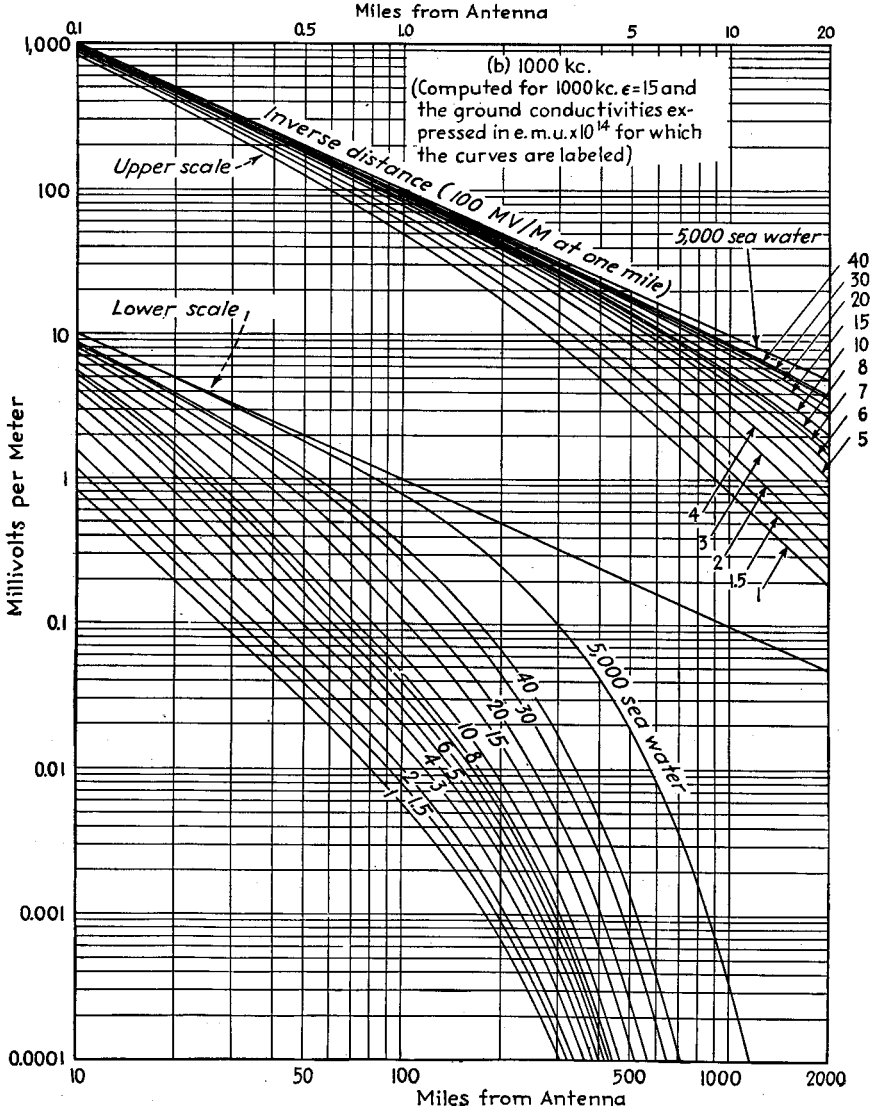


FIG. 50.—(*Continued*)

Curves showing ground-wave intensity as a function of distance for various frequencies are given in Figs. 8 to 10, and additional curves, especially for broadcast frequencies, are given in Fig. 50.[1]

[1] These latter curves are similar to those published by the Federal Communications Commission, and differ somewhat at large distances from those of Figs. 8 to 10 because the computations were made in a slightly different manner.

Examination of these curves shows that the distance at which the ground wave can provide primary coverage is very sensitive to frequency, and is much less at the high-frequency end of the broadcast band than at the low-frequency end. Earth conductivity is also important, and makes it desirable that the site of the transmitting

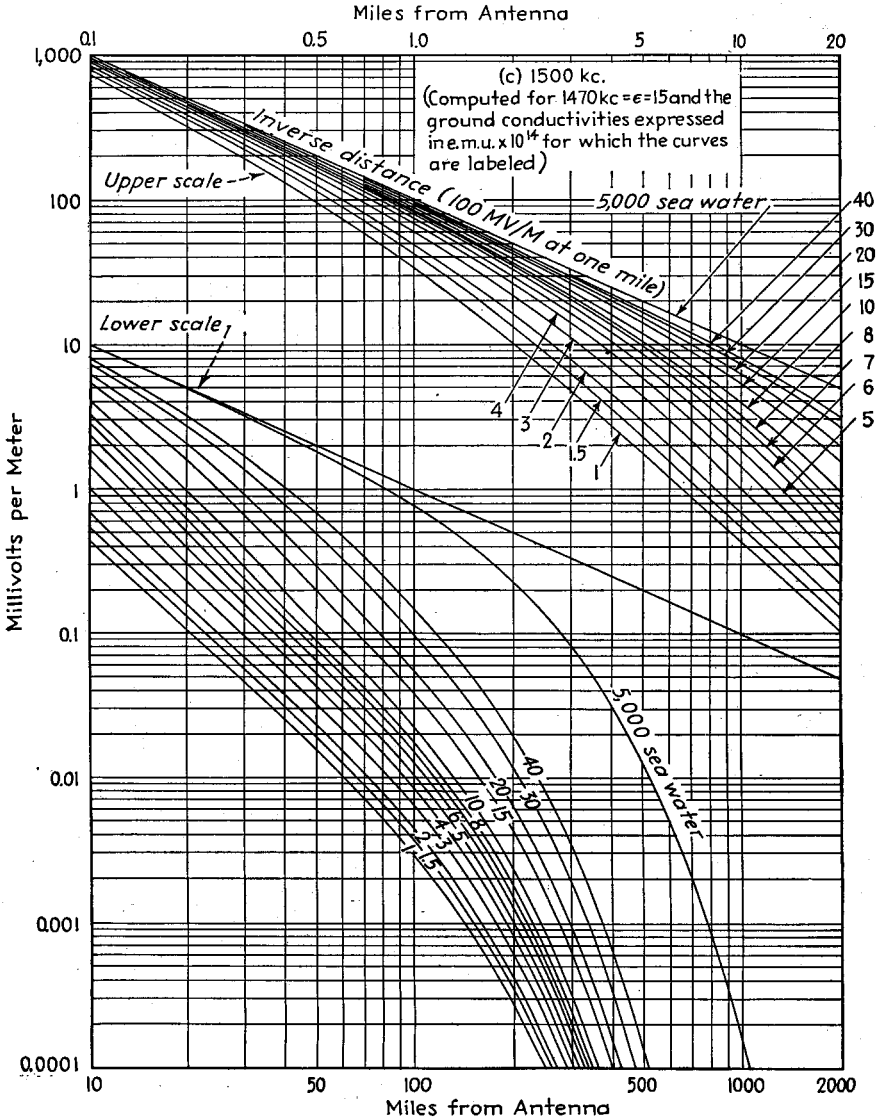


FIG. 50.—(*Continued*)

station be carefully chosen to take maximum advantage of regions of high conductivity, as, for example, water, marshlands, etc.[1] Directional transmitting antennas,

[1] The selection of a transmitter location is discussed by W. B. Lodge, The Selection of a Radio-broadcast-transmitter Location, *Proc. I.R.E.*, Vol. 27, p. 621, October, 1939; W. A. Fitch and W. S.

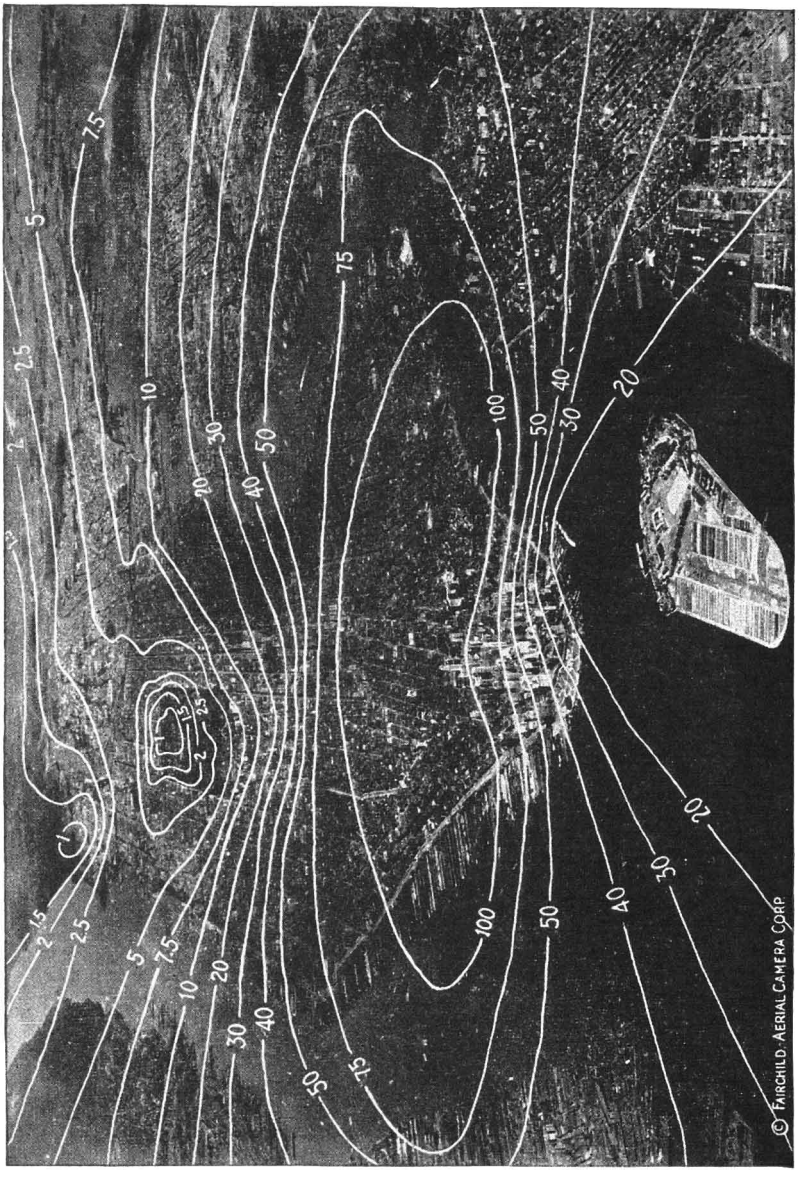


FIG. 51.—Radio field intensity in millivolts per meter for a transmitter located on the top of a building in Lower Manhattan, New York City. (*After R. Bown and G. D. Gillett. Courtesy of Fairchild Aerial Camera Corporation and the Institute of Radio Engineers.*)

which concentrate a large portion of the radiated energy along the horizontal, give an effect on the ground wave that is equivalent to increasing the transmitter power, and so are helpful in improving ground-wave coverage.

Duttera, Measurement of Broadcast Coverage and Antenna Performance, *R.C.A. Rev.*, Vol. 2, p. 396, April, 1938; R. F. Guy, Engineering Factors Involved in Relocating WEAF, *R.C.A. Rev.*, Vol. 5, p. 455, April, 1941.

TABLE 3.—REQUIRED FIELD STRENGTHS

	Field Intensity
Primary coverage:	
City business or factory areas	10–50 mv/m
City residential areas	2–10 mv/m
Towns (2,500–10,000 population)	2 mv/m
Towns (under 2,500 population)	0.5 mv/m
Rural—all areas during winter or northern areas during summer	0.1–0.5 mv/m
Rural—southern areas during summer	0.25–1.0 mv/m
Sky-wave secondary coverage:	
Average conditions	0.5 mv/m
Low noise areas (rural, etc.)	0.1 mv/m

In cities and built-up areas the effective conductivity that must be used in determining ground-wave propagation is considerably less than the actual conductivity of the earth because of additional energy losses occasioned by man-made structures. This effect is particularly pronounced where there are many buildings with steel framework, or numerous reinforced concrete structures, as in the business, industrial, and large apartment-house areas of big cities. Some of these effects are illustrated in Fig. 51,[1] where the "shadow" cast by the skyscraper district of lower Manhattan is very apparent and contrasts strikingly with the low attenuation propagation over the water area.

The Sky Wave.—In the daytime the sky wave at broadcast frequencies is so completely absorbed as to be of negligible importance.[2] With the approach of sunset, however, the sky-wave absorption decreases rapidly, as shown in Fig. 52,[3] and within a few hours after sunset the absorption by the ionosphere has become relatively small. This condition persists throughout the night, but with the approach of sunrise the loss is increased, and the sky wave disappears.

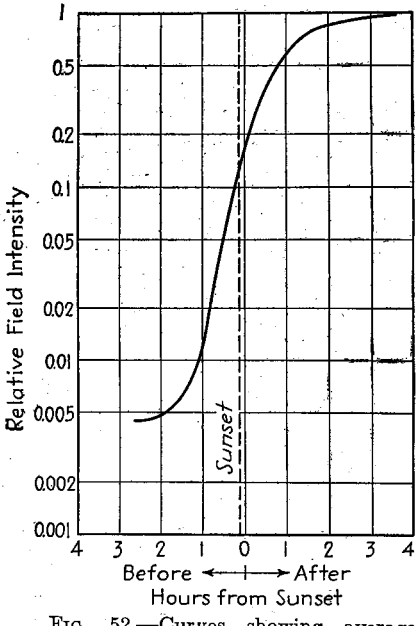


FIG. 52.—Curves showing average variation of strength of sky wave at the receiving point during sunset periods. These results were obtained at a frequency of 800 kc over a distance of 560 miles during the spring.

The absorption of the sky wave by the ionosphere occurs at the lower edge of the *E* layer, or in a lower absorbing layer. The very large loss during the daytime is presumably a result of greater ionization at the lower edge, combined with a slight lowering of the whole ionized region so that it extends into regions of higher atmospheric pressure. The absorption is on the average less during winter nights than in summer nights, and at least to a first approximation is independent of the frequency. The absorption is also commonly assumed to be independent of the angle of incidence with

[1] From R. Bown and G. D. Gillett, Distribution of Radio Waves from Broadcasting Stations over City Areas, *Proc. I.R.E.*, Vol. 12, p. 395, August, 1924.

[2] At considerable distances from the transmitter a sky wave may be detectable, particularly in the winter. The strength of the sky wave, even under favorable conditions, is, however, usually too small to give satisfactory broadcast service.

[3] From Allocation Survey, *F.C.C. Rept.* 18108, Sept. 1, 1936.

which the waves strike the ionosphere, at least when the angle is not too close to grazing. There is evidence to indicate that the absorption increases as the angle of incidence becomes small (*i.e.*, approaches grazing). The loss in the ionosphere varies greatly from day to day and even from hour to hour. However, it appears that under typical night conditions the effective reflection coefficient of the ionosphere for broadcast waves is of the order of 0.25 for angles of incidence exceeding 25°.[1] This means that the wave returning to earth from the ionized region has a field intensity that is approximately one-fourth the intensity (and hence one-sixteenth the energy) that would be obtained if there were no losses whatsoever.

There is also evidence that the sky-wave intensity is related to solar activity, although the data bearing on this point are far from complete. It has, for example, been found that the average value of night signal strength correlates relatively closely with sunspot numbers and the character of the earth's magnetic field, when moving

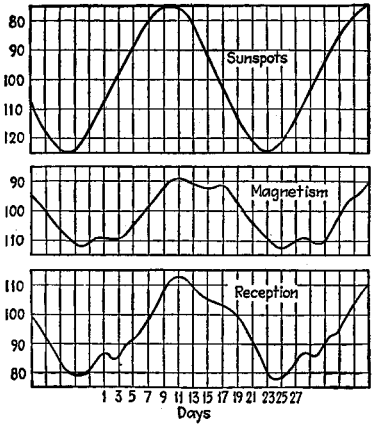


FIG. 53.—Average of sunspot numbers, magnetic character of days, and radio reception on 1,330 kc for eight solar rotations. These curves have been smoothed by the use of a 13-day moving mean.

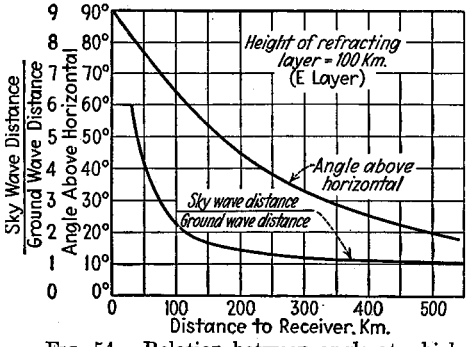


FIG. 54.—Relation between angle at which wave leaves the earth and the distance from transmitter at which return takes place, and also the factor by which the ground-wave distance must be multiplied to give the distance the sky wave travels. These curves take into account the curvature of the earth but assume only slight penetration of the ionized *E* layer.

averages are taken and the results of a number of 27.3 day periods of solar rotation are averaged, as shown in Fig. 53.[2] This curve indicates that strongest night sky waves are received during quiet periods of the solar cycle.

The intensity of the sky wave can be estimated by calculations that assume: (1) that the night sky wave suffers a mirrorlike reflection at a height of 100 km (*E* layer); (2) that the angle of reflection at the layer is equal to the angle of incidence; (3) that the equivalent reflection coefficient at the ionosphere is of the order of 0.25 and is independent of the frequency and of the angle of incidence; (4) that except for the ionosphere losses, the wave field strength is inversely proportional to distance traveled by the wave in going from transmitter to ionosphere and back to earth.

Calculation of sky-wave intensity is facilitated by the use of Figs. 54 and 55, which also give a general idea of how the sky-wave intensity can be expected to vary with dis-

[1] W. Ross, Ground and Ionospheric Rays, *Wireless Eng.*, Vol. 14, p. 306, June, 1937.

[2] G. W. Pickard, Correlation of Radio Reception with Solar Activity and Terrestrial Magnetism II, *Proc. I.R.E.*, Vol. 15, p. 749, September. 1927.

tance. The sky wave is weak very close to the transmitter because broadcast transmitting antennas radiate negligible energy at very high vertical angles. However, as this vertical angle decreases as the distance to the receiver is increased, the field intensity first increases with distance. It then goes through a maximum, and then dies off with still greater distances since a reduction in vertical angle with increased distance is not accompanied by a sufficiently great increase in radiated field at the lower angle to compensate for the increased distance that the wave must travel to reach the receiver. At great distances from the transmitter the sky-wave intensity drops off more rapidly than inversely with distance (see Fig. 56), presumably because the reflection coefficient becomes less when the angle of incidence of the sky wave at the ionosphere approaches glancing, and also because at considerable distances the sky wave may have made two or three round trips between the ionosphere and the earth

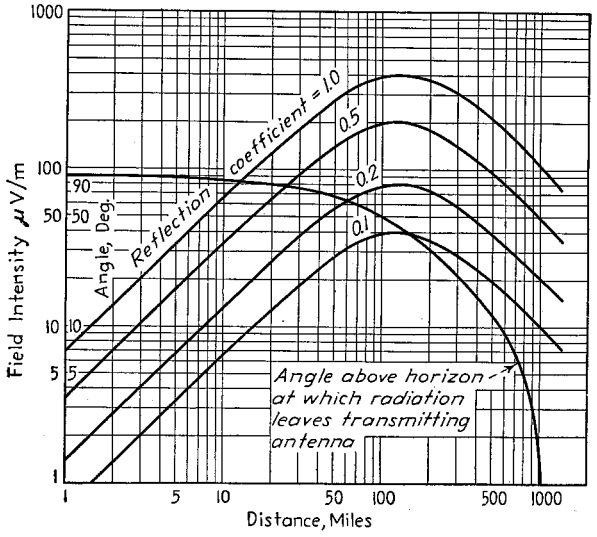


FIG. 55.—Curves useful in the calculation of sky-wave intensity at broadcast frequencies. These curves assume that the wave is reflected from the *E* layer at a height of 100 km and that the transmitting antenna radiates a field 100 mv/m at one mile along the horizontal, with the strength at angles above the horizontal proportional to the cosine of the angle of elevation. The curves take into account the curvature of the earth.

before reaching the receiving point. Experimental results of sky-wave field intensity obtained by averaging a large quantity of data are given in Fig. 56.[1] These curves take into account the fading that occurs at considerable distance by giving the field strength that is exceeded for various percentages of the time. The curve giving the field strength that is exceeded for 5 per cent of the time is often termed the "quasi-maximum."

The curves of Fig. 56 represent general trends. They do not take into account hour-by-hour and night-to-night variations in the ionosphere absorption or seasonal effects, sunspots, etc. Also, it has been found that sky waves at broadcast frequencies will propagate with less attenuation in a north-south direction than between east and west, particularly when the transmission distances are great.[2]

[1] These data were taken during a period that had more than the normal number of ionospheric storms, and there is reason to believe that the fields shown are somewhat lower than for more typical conditions, the difference being perhaps 15 db at the greater distances.

[2] For example see J. H. Dellinger and A. T. Cosentino, A Radio Transmission Anomaly; Co-operative Observations between the United States and Argentina, *Proc. I.R.E.*, Vol. 28, p. 431, October, 1940.

Sky waves from a distant transmitter always exhibit fading. The fading is usually relatively slow, commonly taking some minutes to go from a maximum to minimum signal strength, although in some instances the variation in signal strength is more rapid. This fading is the result of interference between waves that have traveled slightly different paths in the ionosphere. The fading is sensitive to frequency, with the result that with a modulated wave the carrier and the various side-band frequencies do not necessarily fade in and out together. This is termed *selective fading* and introduces distortion in the received wave, which is particularly great when the carrier wave fades to a small amplitude in proportion to the side bands.

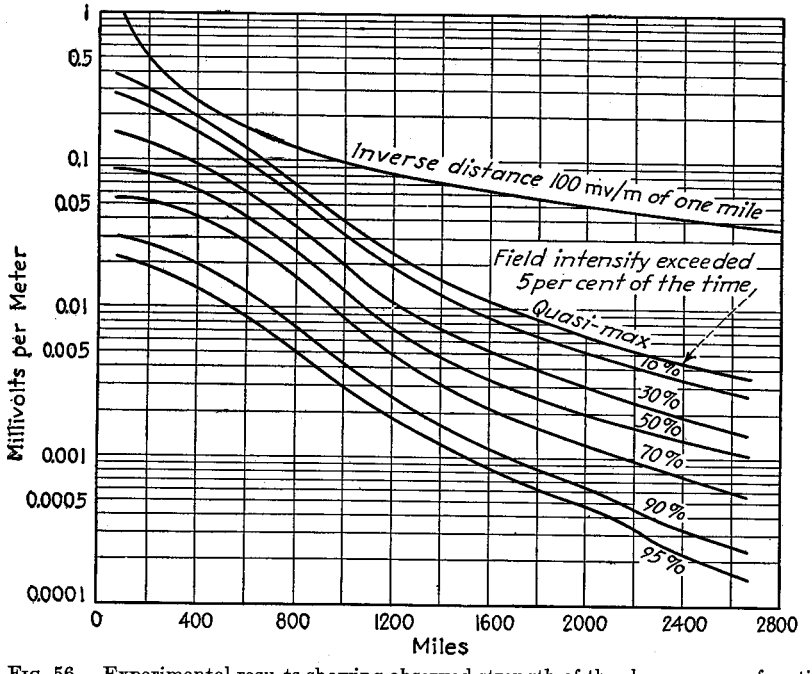


FIG. 56.—Experimental results showing observed strength of the sky wave as a function of distance for the frequency range 640 to 1,190 kc. These particular curves correspond to the second hour after sunset at the recording station and assume that the transmitting antenna radiates a field strength of 100 mv/m at one mile along the horizontal with the radiation in other directions proportional to the cosine of the angle of elevation. (*From F.C.C. survey made February through May*, 1935).

Daytime Coverage.—In the daytime only the ground wave need be considered at broadcast frequencies. The signals that result under these conditions were discussed above, where it was pointed out that they are particularly sensitive to frequency and to the conductivity of the earth. The intensity of the daytime signals can be accurately calculated when the conductivity of the earth is known, and this conductivity can be determined experimentally by methods given in Par. 6.

Nighttime Coverage.—At night the received field is the vector sum of the ground and sky waves, and the situation that exists is as shown in Fig. 57. Close to the transmitter the ground wave dominates, while at large distances the sky wave is the strongest. At some moderate distance the two waves are of approximately equal amplitude, and the resultant field obtained will depend primarily upon the phase relations of the ground and sky waves. This relative phase is dependent upon frequency, and may be

substantially different for the various frequency components contained in a modulated wave. Furthermore, the ionosphere conditions will vary from moment to moment sufficiently to cause the relative phase to vary continually. As a consequence, the resultant signal in the region where ground and sky waves are of substantially equal intensity can be expected to exhibit severe *selective fading*, with accompanying distortion.[1]

The distance from the transmitter where this region of bad distortion occurs will be greater the higher the earth conductivity and the lower the frequency, because these factors increase the strength of the ground wave. It will also be greater the more the transmitting antenna concentrates its radiation along the horizontal, since such antenna directivity causes the ground wave to be stronger, and at the same time reduces the strength of the high-angle radiation producing the part of the sky wave

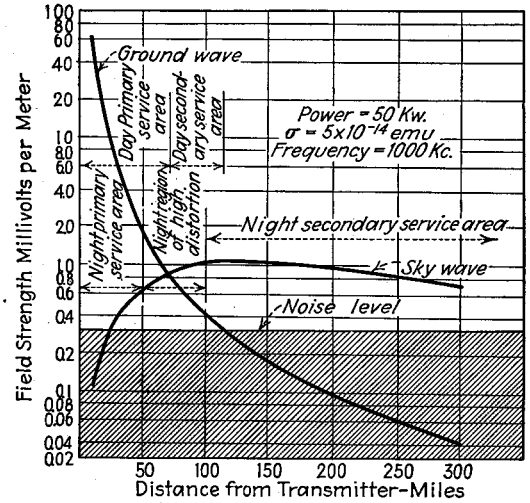


FIG. 57.—Diagram illustrating different types of coverage obtained from a high-power broadcast station during the day and night. The top of the shaded area represents the lowest field strength that completely overrides the noise level under the assumed conditions.

that returns to earth close to the transmitter. The effect of antenna height on the location of the region of high distortion is indicated by Fig. 58.

The position of the region of high distortion is independent of transmitter power, but will vary with conditions in the ionosphere, and so changes from hour to hour and day to day.

With high-power broadcast transmitters, the field strength in the high-distortion region is adequate to give good service if it were not for the distortion. In such cases, the area of primary coverage is greater in the daytime than at night, when some of the daytime coverage is lost because of the distortion region.

16. Propagation Characteristics of Short Waves (1.5 to 30 mc).—Radio communication at high frequencies depends upon ground waves only in special applications such as police radio, where the distances are small and frequencies of the order of 1.5 to 5 mc are used, and also in the case of moderate-distance transmission over water, as in ship-to-shore work.

[1] A thorough study of selective fading at broadcast frequencies is reported in the classic paper by R. Bown, De L. K. Martin, and R. K. Potter, Some Studies in Radio Broadcast Transmission, *Proc. I.R.E.*, Vol. 14, p. 57, February, 1926.

The principal use of frequencies in the range 1.5 to 30 mc is in moderate and long-distance communication, in which the ionosphere is used to refract the sky wave back to earth at the receiving point. The propagation characteristics under such conditions depend upon the transmitted frequency and the conditions in the ionosphere.

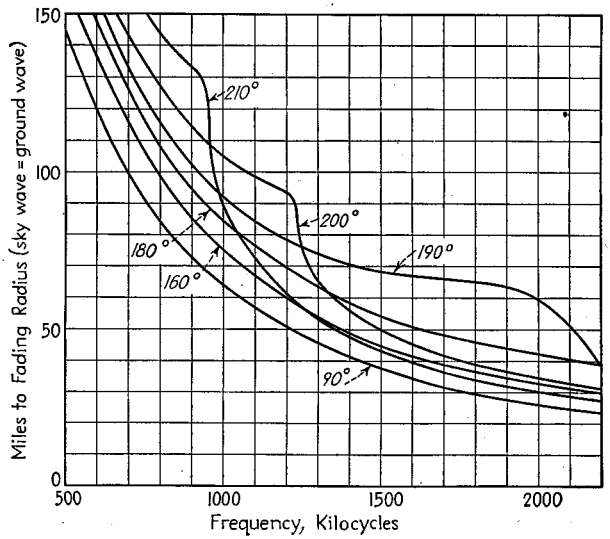


FIG. 58.—Approximate location of high distortion regions for broadcast antennas of different heights. The number of degrees marked for each curve represents the antenna height expressed in electrical degrees (180° corresponds to height of $\lambda/2$).

Skip Distance and Maximum Usable Frequency.—Radio waves of frequency less than the critical frequency of the F_2 (or F) layer of the ionosphere will be returned to earth irrespective of the angle of incidence at the ionosphere. For such frequencies it is therefore possible to obtain sky-wave transmission to a receiving point quite close to the transmitter. However, when the frequency of the transmitted wave is more than the critical frequency of the F_2 (or F) layer, then the only waves that will return to earth will be those that strike the ionosphere with an angle of incidence β such that $\cos \beta > \mu_0$, where μ_0 is the refractive index at the point of maximum electron density for the frequency involved. Waves striking the ionosphere with an appreciably greater angle of incidence will pass on through as illustrated in Fig. 59 (path *a*), while waves with a lesser angle will return to earth beyond the skip distance (path *c*). As a result no sky-wave energy will return to earth closer to the transmitter than some distance *d*, which is termed the *skip distance.*

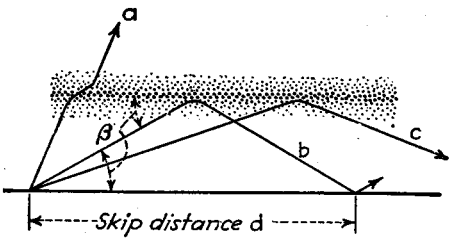


FIG. 59.—Hypothetical wave paths illustrating skip-distance phenomenon. For the sake of simplicity the illustration is drawn for ionosphere conditions involving only a single layer.

As a result of this skip phenomenon, and the fact that the ground wave dies out very rapidly with distance, there is commonly a region beginning a few miles from a short-wave transmitter, and extending out to the skip distance, where no signals at

all are received even though strong signals are returned to earth beyond the skip distance.[1]

The skip distance will depend upon the relationship of the frequency of transmission to the critical frequency of the ionosphere layer involved, upon the height of the layer, and to a slight extent upon the distribution of ionization within the layer. In general it may be said that the higher the frequency, the less the ionization in the layer, and the higher the layer, the greater will be the skip distance. Each particular layer of the ionosphere has associated with it at each frequency a skip distance determined mainly by the critical frequency of that layer and its virtual height. The skip distances of the various layers at a particular frequency will in general be different. The particular layer having the smallest skip distance is sometimes the E and sometimes the F_2 layer, depending upon relative heights and critical frequencies. The F_1 layer is only very seldom the controlling factor.

With given ionosphere conditions the skip distance increases with frequency. Accordingly, for any given distance at a particular time there is a maximum frequency that may be used without causing the sky wave to skip over the receiving point. This frequency is termed the *maximum usable frequency*, and is of great importance because it is also the frequency that ordinarily gives the strongest sky-wave signal at the receiving point as well as being the highest frequency that can be used to obtain sky-wave transmission to the receiver. The maximum usable frequency for the extraordinary ray is always greater than for the ordinary ray. However, except for short distances the difference is not great and decreases with increasing frequency and distance.

An approximate calculation of the maximum usable frequency for a given distance of transmission (or what is equivalent, the calculation of skip distance for a given frequency) can be made by assuming that the wave undergoes a mirrorlike reflection at the point of maximum electron density in the layer, and that the height of this reflecting point is the virtual height of the layer. This gives[2]

$$\mu_0^2 = \frac{\sin^2\left(\frac{s}{2r}\right)}{\sin^2\left(\frac{s}{2r}\right) + \left[1 + \frac{h}{r} - \cos\left(\frac{s}{2r}\right)\right]^2} \tag{50}$$

where r = radius of earth in same units as h.
h = virtual height of ionosphere layer.
s = skip distance in same units as h.
μ_0 = refractive index that must exist at point of maximum electron density for frequency corresponding to a skip distance s.

In applying Eq. (50) to determine the maximum usable frequency, one first calculates the refractive index μ_0 required for the given virtual height and skip distance. The actual frequency of transmission required to give this refractive index can then be determined from a knowledge of the critical frequency of the layer.

At distances short enough so that the earth's curvature may be neglected, the relationship expressed in Eq. (50) takes the simplified form[3]

$$f_m = f_c \sqrt{\left(\frac{s}{2h}\right)^2 + 1} \tag{51}$$

[1] Actually, scattered reflections and sporadic E-layer reflections sometimes result in signals being observed within what is normally this skip distance, but these circumstances are the exception rather than the rule.

[2] For derivation see A. H. Taylor and E. O. Hulburt, The Propagation of Radio Waves over the Earth, *Phys. Rev.*, Vol. 27, p. 189, February. 1926.

[3] For example, see J. H. Dellinger, The Role of the Ionosphere in Radio-wave Propagation, *Trans. A.I.E.E.*, Supplement, Vol. 58, p. 803, 1939.

where f_m is the maximum usable frequency, f_c the critical frequency, and s and h have the same significance as previously.[1]

Results obtained with the aid of Eq. (50) are approximate in that the virtual height h that should be used in this equation will depend somewhat upon frequency. Methods have been devised by which one can start with a curve giving virtual height as a function of frequency as observed for vertical incidence pulse measurements, and determine exactly the maximum usable frequency for a given distance.[2] The analytical basis of such a determination is moderately involved, but the actual result can be readily obtained by graphical means. The maximum usable frequency under typical conditions behaves as shown in Figs. 60 and 61.[3]

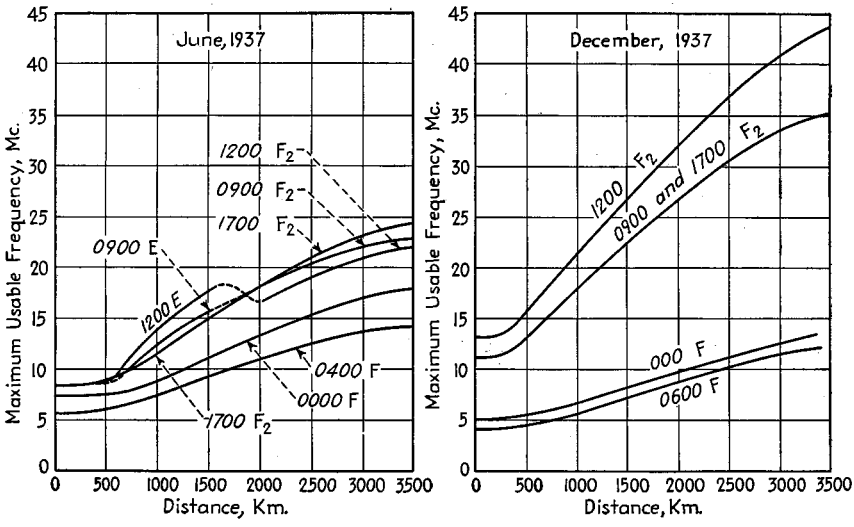


FIG. 60.—Typical curves showing maximum usable frequency as a function of distance for various times of day. Each curve is marked with the appropriate local time and with the layer determining the maximum usable frequency. The results are for Washington, D.C.

The same methods used to determine the maximum usable frequency from a knowledge of the virtual height as a function of frequency for vertical incidence can also be used to obtain the virtual height as a function of frequency for oblique incidence. A curve of virtual height for oblique incidence is shown in Fig. 62. For this particular case the maximum usable frequency is determined by the F_2 layer and is 20 mc.

[1] Another form of Eq. (51) is

$$f_m = f_c \sec (90° - \beta) = \frac{f_c}{\sin \beta} \tag{52}$$

where β is the angle of incidence at the ionosphere measured with respect to the horizontal, as shown in Fig. 59, from which it is also seen that $\tan \beta = h/(s/2)$, where s and b mean the same as in Eq. (50).

[2] Newbern Smith, The Relation of Radio Sky-wave Transmission to Ionosphere Measurements, *Proc. I.R.E.*, Vol. 27, p. 332, May, 1399; Application of Vertical Incidence Ionosphere Measurements to Oblique-incidence Radio Transmission, *Jour. Research Nat. Bur. Standards*, Vol. 20, p. 683, May, 1938; Application of Graphs of Maximum Usable Frequency to Communication Problems, *Jour. Research Nat. Bur. Standards*, Vol. 22, p. 81, January, 1939.

[3] Such curves for a variety of conditions are given by T. R. Gilliland, S. S. Kirby, and N. Smith, Maximum Usable Frequencies for Radio Skywave Transmission, 1933 to 1937, *Jour. Research Nat. Bur. Standards*, Vol. 20, p. 627, May, 1938; also, *Proc. I.R.E.*, Vol. 11, p. 1347, November, 1938.

Transmission Paths—Angle of Departure and Arrival.—If it is assumed that 3½° is the minimum practical angle above the horizon at which rays can depart from the transmitting antenna, the maximum skip distance theoretically possible is about 1,700 km for *E*-layer transmission and 3,000 to 3,500 km for F_2-layer transmission. When communication is to be carried on over greater distances the transmission path

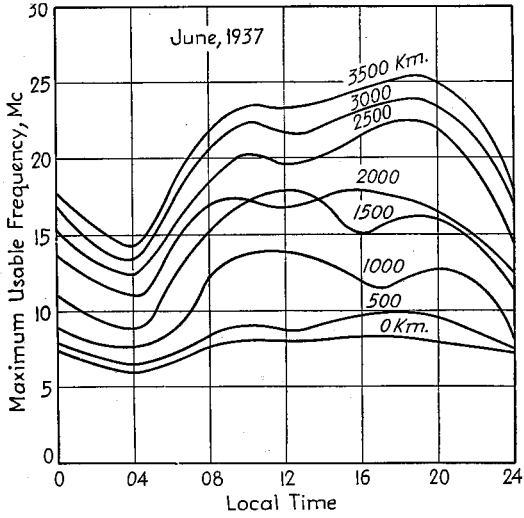


FIG. 61.—Curves showing maximum usable frequency as a function of time of day for different distances as observed at Washington, D.C.

must include two or more hops, as illustrated in Fig. 63. When more than one hop is involved it is to be noted that the virtual height, length of each individual hop, etc., are determined by the ionosphere conditions existing at each apex of the path, and that the conditions may be different for the different apexes. This is particularly so in the case of east-west transmission, where large differences in the local sun time may exist, and will give rise to differences in ionosphere behavior. As a consequence,

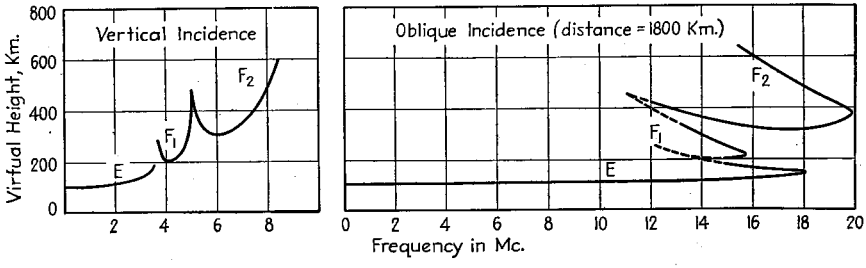


FIG. 62.—Curve of virtual height as a function of frequency for oblique incidence, together with corresponding curve for vertical incidence.

one may expect that the length of different hops will not, in general, be the same, and that in some cases the wave may even be refracted from one layer during one hop and from another layer on another hop.

When the frequency of transmission is appreciably less than the maximum usable frequency, as is commonly the case, then there are normally two or more transmission

paths by which energy may reach a distant receiving point in appreciable amplitude. Examples of such multiple transmission paths are illustrated in Fig. 64. At *a* the multiple paths result from the fact that when the skip distance is appreciably less than the distance to the receiver then the energy may travel by routes involving different numbers of hops. The conditions illustrated at *b* represent multiple-path transmission as a result of different depths of penetration into the ionosphere, and can occur at frequencies differing not very greatly from the maximum usable frequency. Finally, multiple-path transmission is possible as illustrated at *c* as a result of energy being refracted from different layers.

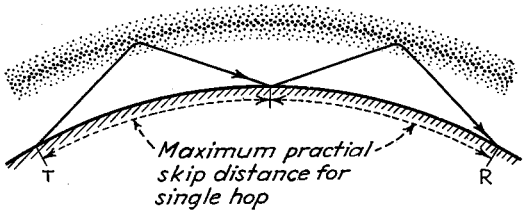


FIG. 63.—Transmission to a distant receiver via a sky wave taking two hops.

Experiments indicate that short-wave transatlantic communication ordinarily involves two to four transmission paths that contribute appreciable energy to the receiver. Each corresponds to energy arriving at the receiver at a particular angle of arrival above the horizon. The angle for a particular path is relatively stable with time, and usually lies in the range 10 to 25°.[1] It has further been observed that the higher the angle at which the energy arrives at the receiving point the greater the length of time required for transmission. This is consistent with Fig. 64, which indicates that the waves associated with the paths of larger angles above the horizon have traveled a greater total distance in order to reach the receiver.

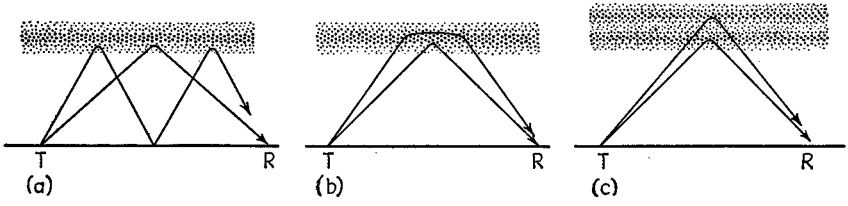


FIG. 64.—Common ways in which multiple transmission paths may arise. For the sake of simplicity the curvature of the earth is neglected.

Optimum Frequency.—In order that a signal usable for practical communication may reach a particular receiving point, it is necessary not only that the frequency not exceed the maximum usable frequency for the path involved but also that the absorption not be excessive. The general rule that applies here is that the highest frequency that will deliver a signal to the receiver will normally give the greatest field strength, and that as the frequency is reduced the field strength becomes less, particularly during the daytime. When frequencies are selected on the basis of curves of monthly averages of maximum usable frequency, such as are given by Fig. 61, it is necessary to choose a frequency at least 15 per cent below the monthly average for maximum

[1] See H. T. Friis, C. B. Feldman, and W. M. Sharpless, The Determination of the Direction of Arrival of Short Radio Waves, *Proc. I.R.E.*, Vol. 22, p. 47, January, 1934.

usable frequency to take care of day-to-day and hour-to-hour deviations.[1] Communication with fair efficiency will usually be obtained in daytime for frequencies somewhat greater than half the maximum usable frequency; at night, down to somewhat less than half the maximum usable frequency.

Since the maximum usable frequency over any given distance varies greatly with time of day, and also somewhat from season to season and year to year, the optimum transmission frequency will vary likewise. In particular, where continuous communication is desired it is necessary to employ at least two frequencies, one for day operation and the other for night conditions. A third frequency is also sometimes desirable for transition conditions.

When ionosphere storms are encountered the maximum usable frequency is reduced, and the absorption is greatly increased, particularly when the transmission path approaches high altitudes. The absorption associated with a severe ionosphere storm is then usually sufficient to extinguish the signal completely.

In long-distance communication, particularly over east-west paths, transmission conditions at different places along the path may be greatly different. The highest frequency that will deliver a signal to the receiver may then be much less than the optimum frequency for one of the hops, so that the absorption will be excessive. This condition tends to occur when one end of the path is in darkness and the other in full sunlight.

Where sporadic *E*-layer reflection is obtained, communication is possible at frequencies much greater than the maximum usable frequency calculated on the basis of the ordinary layers.[2] The signals received under such conditions are often of great intensity as a result of the low absorption encountered by the very high frequency. However, because of the transient character of sporadic *E*-layer reflections, they cannot be depended upon for commercial communication, but rather represent freak conditions of interest and value to amateurs, experimenters, etc., which sometimes also give rise to unexpected interference.

Field Strength.—A large amount of information has been published concerning field strengths that can be expected under different conditions from short-wave transmitting stations. Some of the more important articles summarizing information of this sort are given below.[3] These data indicate among other things that communication over the North Atlantic is much worse than from North America to South America. Thus the difference between the New York-Buenos Aires circuit has been reported as approximately 25 db.[4]

The signal strengths required for carrying on short-wave communication depend upon conditions. Values of the order of 10 microvolts per meter are sometimes required, while 1 microvolt per meter will commonly be sufficient to give intelligible

[1] A discussion covering the use of graphs of maximum usable frequency for the practical problem of selecting the optimum frequency is given by N. Smith, S. S. Kirby, and T. R. Gilliland, Application of Graphs of Maximum Usable Frequency to Communication Problems, *Bur. Standards Research Paper* RP1167, Vol. 22, p. 81, January, 1939.

[2] Further discussion of transmission by sporadic *E*-layer reflections is given by E. H. Conklin, 56-megacycle Reception via Sporadic *E*-layer Reflections, *Proc. I.R.E.*, Vol. 27, p. 36, January, 1939; D. R. Goddard, Transatlantic Reception of London Television Signals, *Proc. I.R.E.*, Vol. 27, p. 692, November, 1939.

[3] C. R. Burrows, The Propagation of Short Waves over the North Atlantic, *Proc. I.R.E.*, Vol. 19, p. 1634, September, 1931; M. L. Prescott, The Diurnal and Seasonal Performance of High-frequency Radio Transmission over Various Long-distance Circuits, *Proc. I.R.E.*, Vol. 18, p. 1797, November, 1930; Clifford N. Anderson, North Atlantic Ship-shore Radio-telephone Transmission during 1932–1933, *Proc. I.R.E.*, Vol. 22, p. 1215, October, 1934; Report of Committee on Radio-wave Propagation, *Proc. I.R.E.*, Vol. 26, p. 1193, October, 1938; C. R. Burrows and E. J. Howard, Short-wave Transmission to South America, *Proc. I.R.E.*, Vol. 21, p. 102, January, 1933; R. A. Heising, J. C. Schelleng, and G. C. Southworth, Some Measurements of Short-wave Transmission, *Proc. I.R.E.*, Vol. 14, p. 613, October, 1926.

[4] R. K. Potter and A. C. Peterson, Jr., The Reliability of Short-wave Radio Telephone Circuits, *Bell System Tech. Jour.*, Vol. 15, p. 181, April, 1936.

radio-telephone signals and permit commercial radio-telegraph communication, provided that the receiving location is relatively free from local noise, and a well-designed receiving antenna is employed.

The strength of short-wave signals received from a distant transmitter (at least for North Atlantic paths) has been found to show a correlation with the character figures indicating terrestrial magnetic activity and earth currents. These character figures show a 27-day recurrence tendency, so that it is possible to predict with fair accuracy the strength of signals to be expected in the near future from the conditions that existed in the immediate past.[1]

Selective Fading with Amplitude- and Frequency-modulated Signals.—When short-wave signals that have traveled to a distant receiver along a particular path fade, the carrier and side-band frequencies (up to 5 or 10 kc) normally fade in and out together. Thus the amount of *selective* fading associated with an individual transmission path tends to be small. However, under the rather common conditions where there are two or more transmission paths delivering signals of comparable strength to the receiver, then the resultant received signal will be the vector sum of the signals coming in over the various paths. This vector sum will, in general, be quite sensitive to the exact frequency of the wave, with frequencies differing by as little as 200 cycles behaving differently. Thus when an ordinary amplitude-modulated wave is received under such conditions, the various side-band frequencies, and also the carrier, will fade in and out in a more or less random way. This causes *selective fading* of the resultant signal, with accompanying distortion.[2]

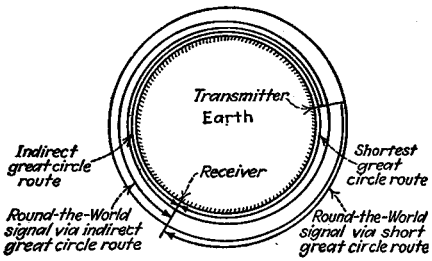


FIG. 65.—Paths followed by multiple and round-the-world signals.

Frequency-modulated signals transmitted over great distances under conditions where there is more than one transmission path suffer much greater distortion than do amplitude-modulated signals. The distortion is greatest at low modulation frequencies and high depths of modulation. This distortion results from the fact that the instantaneous frequency of a frequency-modulated wave is continually varying, so that when two waves arrive at a receiving point after having traveled different distances, they have different instantaneous frequencies. The resultant wave then contains a new modulation involving both amplitude and frequency modulation at a frequency that is not harmonically related to the modulation produced at the transmitter, but that rather depends upon the difference in transit time for the different paths.[3]

Round-the-world Signals.[4]—Signals from a transmitter to a distant receiver may reach the receiver either by traveling directly toward the receiver or by traveling

[1] A. M. Skellett, On the Correlation of Radio Transmission with Solar Phenomena, *Proc. I.R.E.*, Vol. 23, p. 1361, November, 1935; Henry E. Hallborg, Terrestrial Magnetism and Its Relation to World-wide Short-wave Communications, *Proc. I.R.E.*, Vol. 24, p. 455, March, 1936; Short-wave Radio Transmission and Geomagnetism, *R.C.A. Rev.*, Vol. 5, p. 395, April, 1941; R. M. Morris, and W. A. R. Brown, Transoceanic Reception of High-frequency Telephone Signals, *Proc. I.R.E.*, Vol. 21, p. 63, Janary, 1933.

[2] Results of a classic study of selective fading at high frequencies are reported by R. K. Potter, Transmission Characteristics of a Short-wave Telephone Circuit, *Proc. I.R.E.*, Vol. 18, p. 581, April, 1930.

[3] Murray G. Crosby, Observations of Frequency-modulation Propagation on 26 Megacycles, *Proc. I.R.E.*, Vol. 29, p. 398, July, 1941; Frequency Modulation Propagation Characteristics, *Proc. I.R.E.*, Vol. 24, p. 898, June, 1936.

[4] For further discussion see E. Quäck and H. Mögel, Double and Multiple Signals with Short Waves, *Proc. I.R.E.*, Vol. 17, p. 791, May, 1929.

indirectly the opposite way around the earth, as shown in Fig. 65, thus leading to long and short great-circle paths. It is also possible under certain conditions for signals to travel from the transmitter past the receiver, on around the earth, and then to the receiver a second time. There even are cases on record where the same signal has been heard five times at a distant receiving point, while signals repeated once or twice

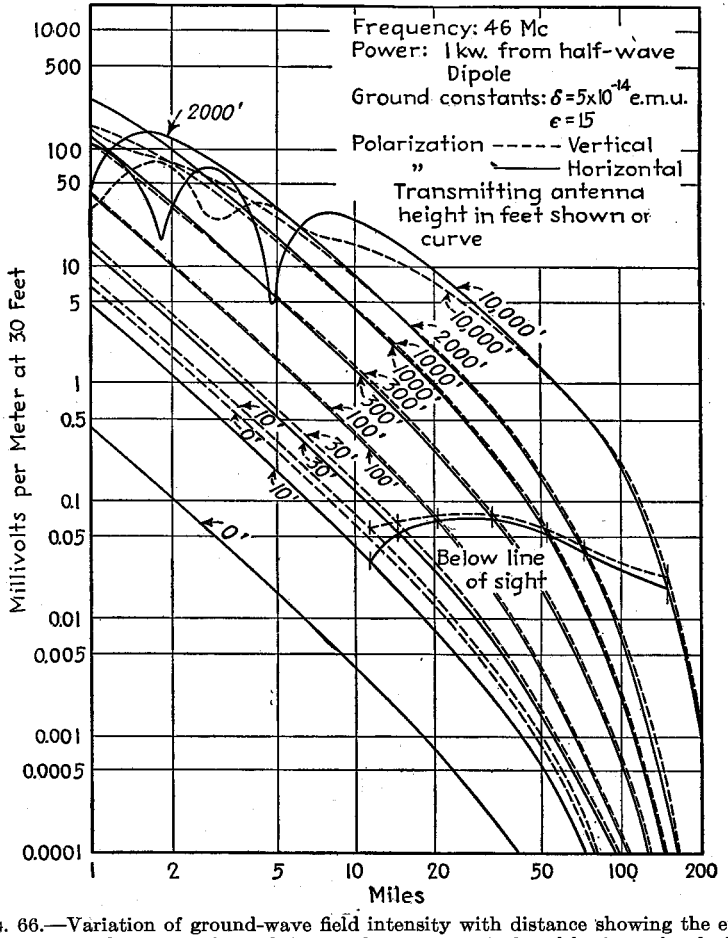


FIG. 66.—Variation of ground-wave field intensity with distance showing the effect of different antenna heights and the difference between vertical and horizontal polarization. The point on each curve corresponding to line-of-sight conditions is indicated by a short vertical line.

are not infrequent. In order that double signals may reach the receiver by propagation along both long and short great-circle paths, it is necessary that the conditions in the ionosphere be roughly the same along both routes, and the same requirement must also be fulfilled in the case of round-the-world signals. As a result, multiple and round-the-world signals are observed only when the part of the great-circle path that lies in darkness is experiencing summer, or when the great-circle path coincides very closely with the twilight zone. The time delay of round-the-world signals is of the order of $\frac{1}{7}$ second, and is so great that when these signals occur it is necessary to

reduce the speed of telegraph transmission to an extremely low value. With radiotelephone signals the fading and selective distortion is increased.

17. Propagation of Ultra-high Frequency Waves (Frequencies above 30 mc).—Frequencies above 60 mc are never refracted or reflected back to earth by the ionosphere, and it is only under special circumstances that frequencies in the range 30 to 60 mc are so returned. As a consequence, communication at ultra-high frequencies is obtained by means of the ground wave discussed in Pars. 1 to 4, with the added complication that at receiving points well below the line of sight, tropospheric reflections may be of importance.

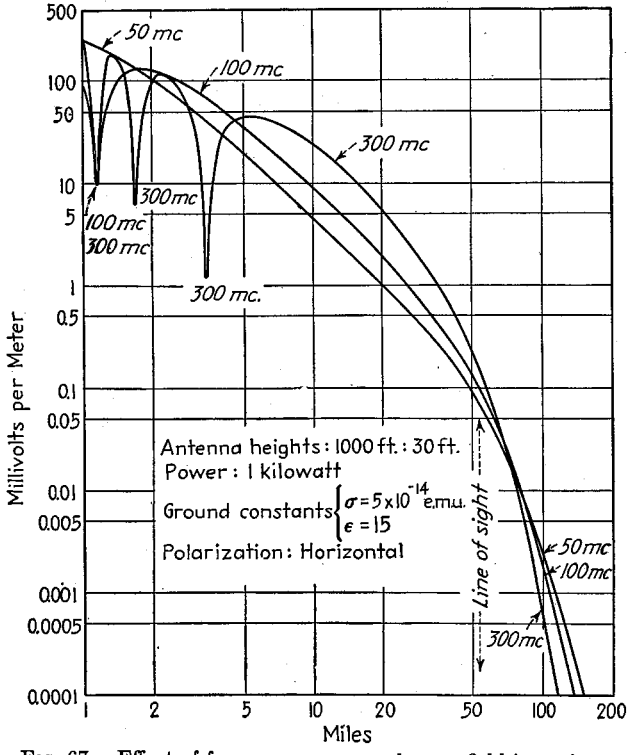


FIG. 67.—Effect of frequency on ground-wave field intensity.

Field-strength Curves.—The factors determining the strength of the ground wave are discussed at length in Pars. 1 to 6, and the essential characteristics are illustrated in Figs. 19 to 21.

The field strength will vary with distance in the manner illustrated in Fig. 66 (also see Fig. 20). At receiving points appreciably above the line of sight the field strength is almost exactly inversely proportional to the square of distance d, and is also proportional to the product h_1h_2 of the antenna heights, provided that the transmission distance is large compared with the antenna height. If the transmission distance is not large compared with antenna height,[1] then the curve of field strength as a function of distance shows an oscillatory character, as illustrated in Fig. 20 and by the horizontally polarized curve for 10,000 feet in Fig. 66. At points appreciably

[1] Whenever $(2\pi h_1h_2/\lambda d) > 0.5$, the transmission distance can be considered too small in relation to antenna heights for the field to be proportional to h_1h_2/d^2.

beyond the optical horizon, the field strength falls off faster than the square of the distance, while as the antennas are raised (though still well below the horizon) it increases more rapidly than the product of antenna heights.

The effect of frequency is illustrated in Fig. 67. Under conditions where the field strength oscillates with increasing distance, the effect of an increase in frequency is to increase the number of oscillations occurring in a given distance without affecting the maximum amplitude of the oscillations. Increasing the frequency also causes the oscillatory distribution of field strength to continue until closer to the optical horizon. At distances less than the optical horizon, but great enough so that the oscillatory distribution of field strength no longer exists, increasing the frequency increases the field strength, whereas at distances appreciably beyond the optical horizon the field strength is less the higher the frequency.[1]

The field strength is very sensitive to antenna height. At receiving points well beyond the optical horizon, and below the line of sight, the field intensity will depend upon height in the manner shown in Fig. 21*a*. As the antenna is raised above the earth's surface with vertically polarized waves the field strength is first constant, and then becomes higher at an increasing rate until at several wave lengths the field is almost linearly proportional to height. This linear relation is then maintained until the height is very large, after which the field strength increases faster than the height. Horizontally polarized waves are seen to behave in substantially the same manner as vertically polarized waves except that the height below which the field intensity levels out and becomes independent of height is a very small fraction of a wave length above the earth.

At receiving points within the line of sight, the action at small and moderate heights is as in Fig. 21*a*, but now if the height becomes sufficient in proportion to transmission distance so that the curve of field strength as a function of distance displays an oscillatory character, then the curve of field strength with height likewise shows oscillations, as illustrated schematically in Fig. 21*b*.

Curves giving the approximate field strength as a function of distance for various transmitter antenna heights are presented in Fig. 68. These are for the case of a small vertical doublet transmitting antenna radiating a power of one kilowatt, and assume that the receiving antenna is located at the surface of the earth. The increase in signal strength that results by raising the receiving antenna above the earth for any particular case is given by the curve to the right of the corresponding field-strength curve.[2]

The application of the curves of Fig. 68 to a practical communication problem is illustrated by the following example:

Example.—It is desired to determine the field intensity expected at a distance of 100 km over land when the transmitter is at a height of 200 meters, the receiving antenna is at the height of 100 meters, and the frequency is 150 mc (2 meters). The transmitter power is 250 watts.

If the transmitter power were 1 kilowatt and the receiving antenna were at the surface of the ground, then the curves show that the received field would be 0.8 $\mu v/m$. However, the fact that the receiving antenna is at a height of 100 meters results in the field strength being 42 db (= 126 times) greater, or $126 \times 0.8 = 101$ $\mu v/m$ for one kilowatt, or 55.5 $\mu v/m$ for 250 watts.

While the curves of Fig. 68 were derived on the basis of vertically polarized waves, they can also be applied to horizontally polarized waves, provided that the antenna

[1] Further discussion of the effect of frequency is given by K. A. Norton, The Effect of Frequency on the Signal Range of an Ultra-high-frequency Radio Station, *F.C.C. Mimeo. Rept.* 48466, Mar. 20, 1941.

[2] These curves are due to T. L. Eckersley, Ultra-short-wave Refraction and Diffraction, *Jour. I.E.E.*, Vol. 80, p. 286, 1937; also, Wireless Section, *I.E.E.*, Vol. 12, p. 42, March, 1937, and Report of Committee on Radio Wave Propagation, *Proc. I.R.E.*, Vol. 26, p. 1193, October, 1938. They were calculated by a somewhat less exact method than that outlined in Pars. 1 to 4, but are sufficiently accurate for most purposes, and also indicate the influence of various factors on field strength. These curves neglect the refraction in the atmosphere, and so give field strengths slightly low, particularly at considerable distances.

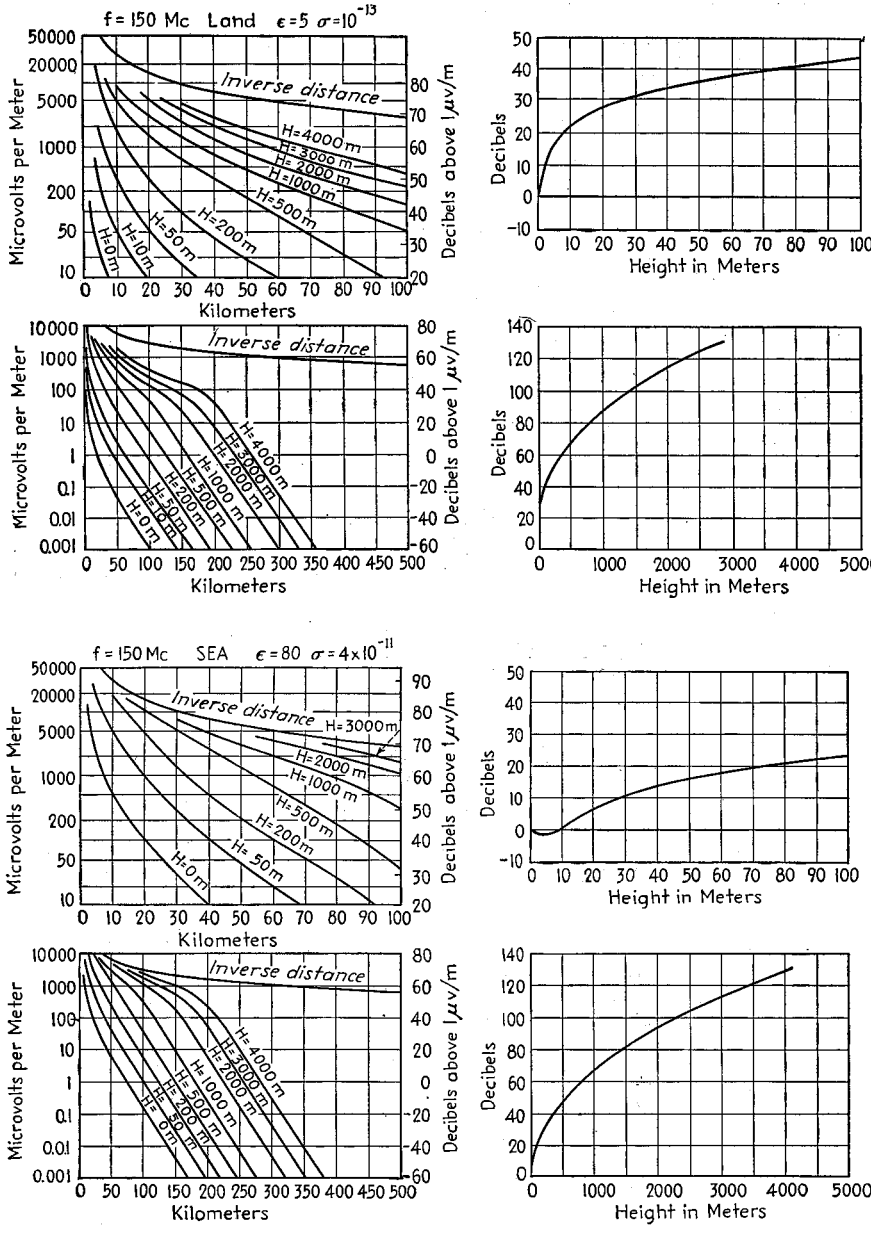


FIG. 68a.—Field-strength curves for 150 mc.

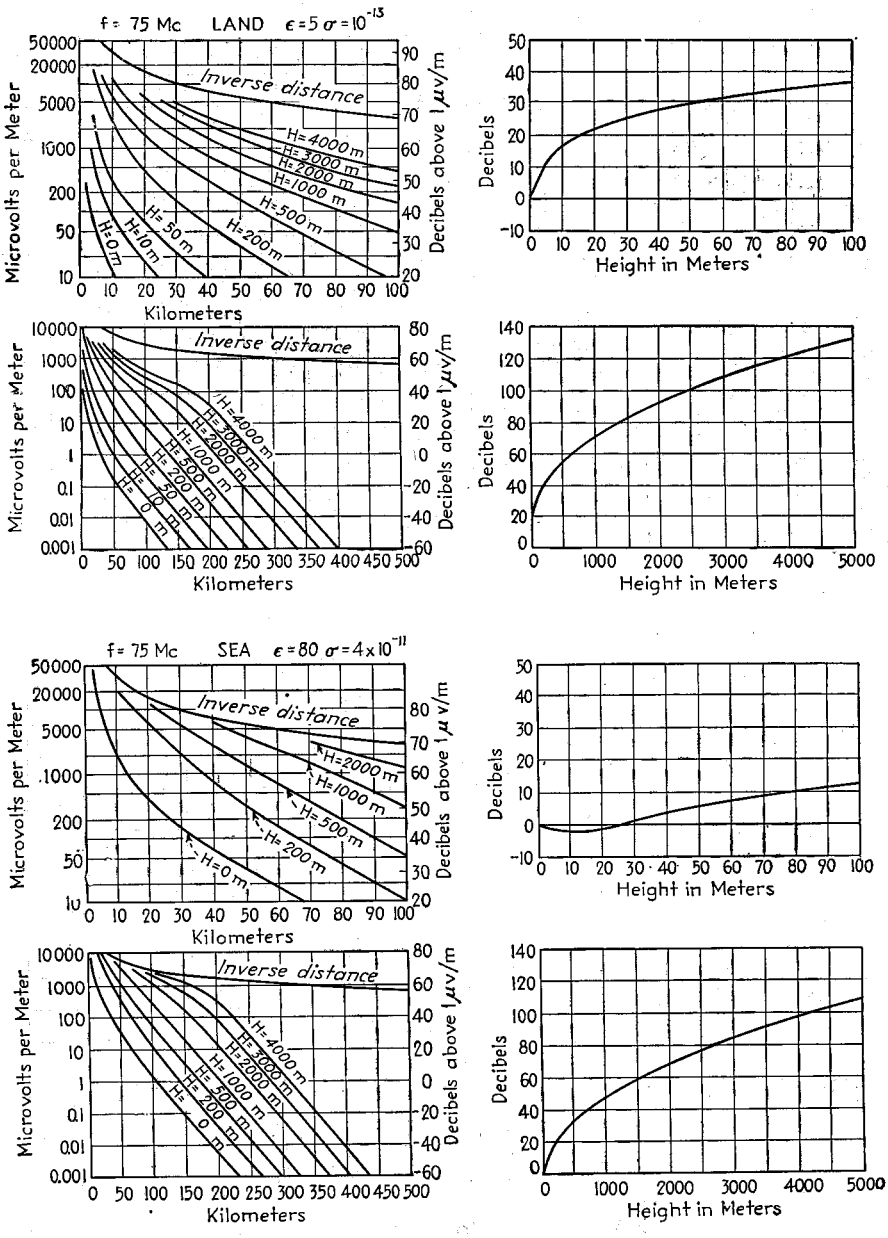


FIG. 68b.—Field-strength curves for 75 mc.

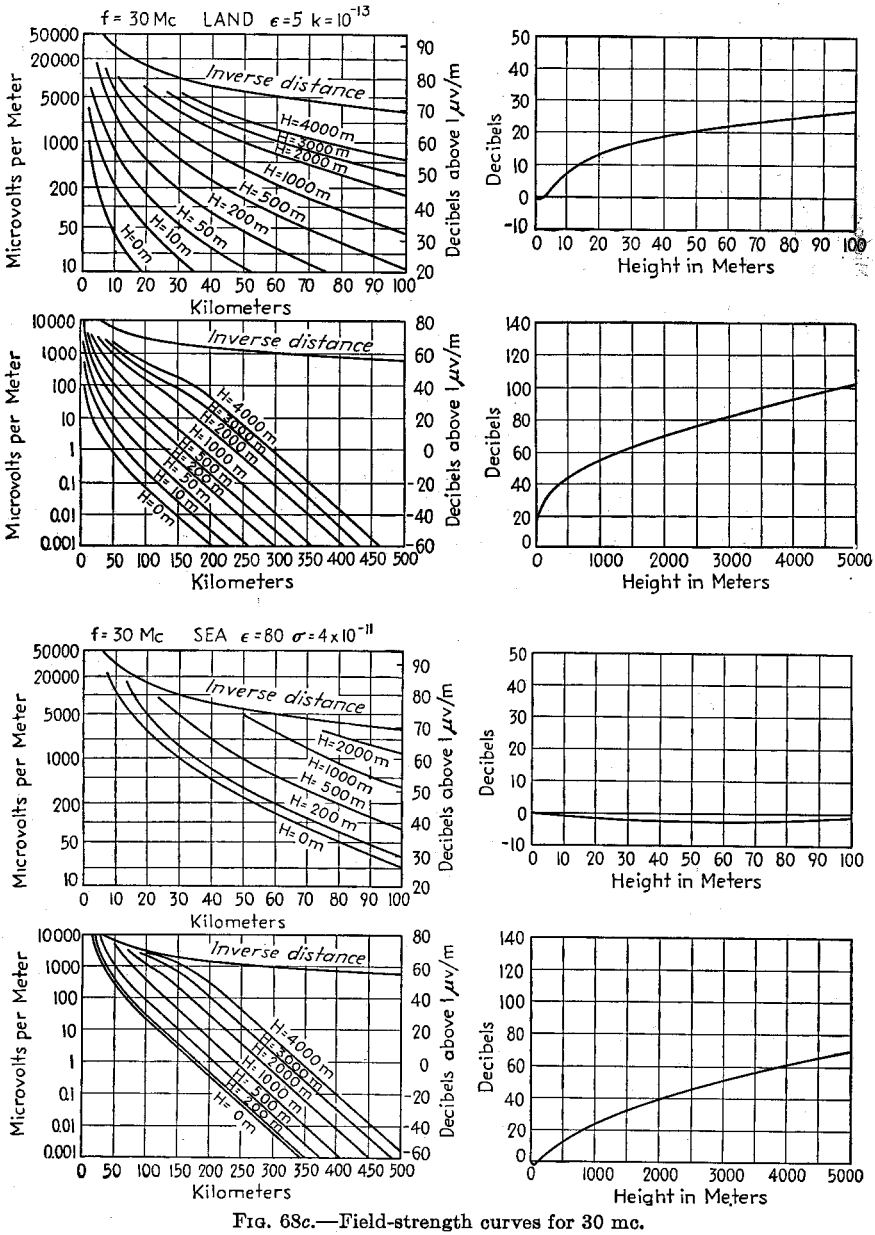


FIG. 68c.—Field-strength curves for 30 mc.

heights are at least several wave lengths. Under this condition vertically and horizontally polarized waves give almost identical received fields, provided that the transmission distance is great compared with the antenna heights (see Fig. 66).

Practical Limitations to Theoretical Field-strength Curves.—Theoretical field-strength curves determined in the manner outlined and discussed above will give accurate results under idealized conditions. Practical conditions, however, nearly always differ to greater or less degree from the ideal case, so that while field strength curves actually observed follow the general trend of the theoretical curves, they can be expected to differ in detail. The theoretical curves cannot, for example, take into account the irregularities introduced by cities, by large buildings, by irregular terrain, slopes, etc., or by brush and trees. Theoretical calculation of ground-wave field intensity likewise ignores the presence of tropospheric reflections.

The effect of cities on ultra-high-frequency waves is obviously too complicated to be taken into account in a formula. However, experimental results obtained on the strength of a received signal in built-up areas show that the signal is less than that received under idealized conditions by approximately 10 db.[1] This loss accounts for the scattering of the wave required in order to reach street level.

In cities where there are large buildings it is found that there are commonly several wave paths along which energy travels from the transmitter to the receiver, these multiple paths arising as a result of reflections at the surfaces of buildings.[2] Such reflections frequently produce a change in the plane of polarization, so that signals that were radiated with horizontal polarization possess a vertical component at the receiving point, and vice versa. These multiple-path transmissions are very disturbing in television work because the differences in path length when measured in wave lengths will vary with frequency. In the wide frequency band required by television signals the result is that the energy arriving over the multiple paths will add at some frequencies, and subtract at other frequencies within the transmitted band. Multiple-path transmission is also serious in connection with ultra-high-frequency landing beams, etc., since an auxiliary wave path created by reflection from a near-by building, for example, will produce a spurious course that may lead to serious consequences. Trouble from multiple-path transmission tends to be greater the higher the frequency because the minimum size of a surface that can act as an effective reflector is proportional to the wave length.

Brush, trees, etc., will have an important effect upon the field strength at ultra-high frequencies. Thus, tests made at 250 to 500 mc with antenna heights of about 3 meters have indicated an attenuation through 500 feet of wood and underbrush that is of the order of 10 to 20 db greater than for propagation over level, cleared ground.[3]

In the case of transmission to a receiving point above the line of sight, the character of the earth's surface in the region where the ground-reflected wave is reflected from the earth is of importance. Such factors as irregularities, trees, buildings, etc., can have an important influence on the reflection coefficient.

Irregularities of terrain will also influence the received field strength. The slope of the earth in the immediate vicinity of the transmitter and the receiver is of particular importance.[4] Hills, large buildings, and other obstacles will cast shadows; *i.e.*,

[1] Charles R. Burrows, Loyd E. Hunt, and Alfred Decino, Ultra-short Waves in Urban Territory, *Elec. Eng.*, Vol. 54, p. 115, January, 1935.

[2] P. S. Carter and G. S. Wickizer, Ultra-high-frequency Transmission between the R.C.A. Building and the Empire State Building in New York City, *Proc. I.R.E.*, Vol. 24, p. 1082, August, 1936; R. W. George, A Study of Ultra-high-frequency Wide-band Propagation Characteristics, *Proc. I.R.E.*, Vol. 27, p. 28, January, 1939.

[3] B. Trevor, Ultra-high-frequency Propagation through Woods and Underbush, *R.C.A. Rev.*, Vol. 5, p. 97, July, 1940.

[4] C. R. Englund, A. B. Crawford, and W. W. Mumford, Some Results of a Study of Ultra-short-wave Transmission Phenomena, *Proc. I.R.E.*, Vol. 21, p. 464, March, 1933.

the field strength will be less immediately behind the obstacle than it is some distance beyond.[1]

Various investigators have reported the results of field-strength surveys.[2] Typical data taken along a radial away from a transmitter are shown in Fig. 69, where the extremities of each vertical line represent the maximum and minimum field strength observed in that section of the chart. These particular results were taken in reasonably open surroundings, yet even then it will be noted that variations of the order of 15 db are present, which is typical.

Fading of Ultra-high-frequency Signals.—Ultra-high-frequency signals received at a considerable distance from the transmitter will normally have a tendency to fade. There are two causes of such fading: *first*, changes in the rate with which the refractive index of the earth's atmosphere varies with height; *second*, variations in troposphere reflections.

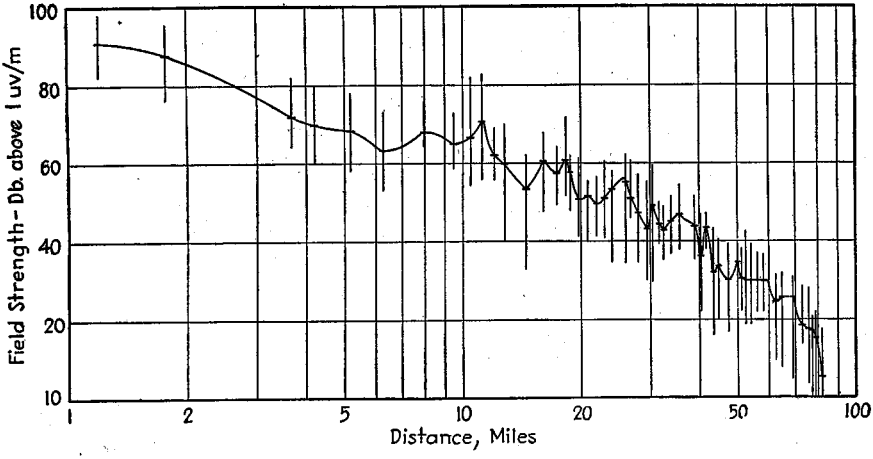


FIG. 69.—Observed variation of field strength with distance from an ultra-high-frequency transmitter. The vertical line represents the range of field strength observed in each region.

It is indicated in Par. 4 that the dielectric constant of the earth's atmosphere varies slightly with height in such a way as to cause a bending of the wave path. On the average, this bending is equivalent to increasing the effective radius of the earth by a factor of approximately 1.33. However, the conditions in the atmosphere vary from time to time, and the degree of bending of the wave path (and therefore the corresponding equivalent radius of the earth) will likewise vary. This will affect the field strength of the received signal, and can cause slow fading of moderate intensity at points an appreciable distance from the transmitter.

Troposphere reflections (see below) are produced by discontinuities in the troposphere, and since these discontinuities may vary in magnitude, orientation, and in height from moment to moment, one can expect severe fading whenever the tropo-

[1] An analysis of this effect showing the magnitude of the shadow is given by T. L. Eckersley, Direct-ray Broadcast Transmission, *Proc. I.R.E.*, Vol. 20, p. 1555, October, 1932.

[2] G. W. Wickizer, Field-strength Survey, 52.75 Megacycles from Empire State Building, *Proc. I.R.E.* Vol. 28, p. 291, July, 1940; Mobile Field-strength Recordings of 49.5, 83.5, and 142 Mc from Empire State Building, New York—Horizontal and Vertical Polarization, *R.C.A.*, Vol. 4, p. 387, April, 1940; Field Intensity Survey of Ultra-high-frequency Broadcasting Stations, *F.C.C. Mimeo. Rept.* 40004; R. S. Holmes and A. H. Turner, An Urban Field Strength Survey at Thirty and One Hundred Megacycles, *Proc. I.R.E.*, Vol. 24, p. 755, May, 1936.

spheric wave is dominant at the receiving point, or when its intensity is of the same order of magnitude as that of the ground wave.

Experimental studies of fading have been made by many investigators,[1] and some typical fading records are shown in Fig. 70. When the optical horizon is at a distance of the order of 40 miles, it is found that the signal strength at the optical horizon can fade through a range of at least 15 db, while at appreciably greater distances than this, which are well below the line of sight, the variation in signal strength may reach as much as 50 db. The extent of the fading, and also its character, vary greatly, as is apparent from the records in Fig. 70.

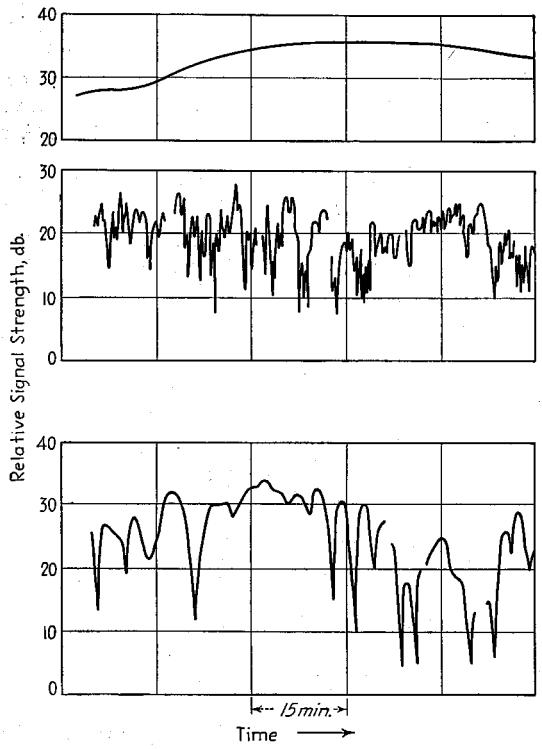


FIG. 70.—Typical types of fading curves observed at ultra-high frequencies. These are for a wave length of 4.7 meters and a distance of 70 miles.

A relationship exists between meteorological conditions and fading. Effects observed include a difference in the tendency for fading to occur in the day and night, and seasonal actions in which the severity of fading will differ in winter and summer. Such behavior can be expected to depend upon local weather conditions, however. Hence any behavior, however regularly observed at one location, may not be observed at some other locality.

[1] For example, see C. R. Burrows, A. Decino, and Loyd E. Hunt, Stability of Two-meter Waves, *Proc. I.R.E.*, Vol. 26, p. 516, May, 1938; C. R. Englund, A. B. Crawford, W. W. Mumford, Ultra-short-wave Transmission over a 39-mile "Optical" Path, *Proc. I.R.E.*, Vol. 28, p. 360, August, 1940; Ultra-short-wave Transmission and Atmospheric Irregularities, *Bell System Tech. Jour.*, Vol. 17, p. 489, October, 1938; A. H. Waynick, Experiments on the Propagation of Ultra-short Radio Waves, *Proc. I.R.E.*, Vol. 28, p. 468, October 1940; K. G. MacLean and G. S. Wickizer, Notes on the Random Fading of 50-megacycle Signals over Nonoptical Paths, *Proc. I.R.E.*, Vol. 27, p. 501, August, 1939.

Fading at ultra-high frequencies is influenced somewhat by frequency, the higher frequencies generally fading the most. Some observers have reported that horizontally polarized waves fade more severely than do vertically polarized waves, but others have not found any polarization effects.

The fading observed at considerable distances from an ultra-high-frequency transmitter is sufficiently selective with respect to frequency to produce distortion in television signals. This distortion can be severe enough to reduce or even entirely destroy the usefulness of the reproduced picture for entertainment purposes.

Tropospheric Reflections.—The troposphere is that part of the earth's atmosphere adjacent to the earth's surface, where clouds form and convection is appreciable.

It has been found that reflections of appreciable importance in ultra-high-frequency wave propagation can be produced in the troposphere.[1] These reflections may occur at heights above the earth from 10 km down to 1 km or less, the lower heights being more common. The magnitude of the reflected waves is small, the reflection coefficient being commonly of the order of 10^{-3} to 10^{-4}, although larger values are possible under certain circumstances.

Tropospheric reflections are caused by small discontinuities in the dielectric constant of the air at boundaries of air masses of different characteristics. Ionization does not play any part in tropospheric reflections, since observations made with meteorological sounding balloons indicate that the necessary ionization is not present in the troposphere.[2] The discontinuities in the dielectric constant required to explain observed results are of the order of 10^{-5}, and these can be readily accounted for on the basis of air masses of different temperature and moisture content in proximity to each other.

The reflection coefficient resulting from a discontinuity in dielectric constant can be calculated as explained in the footnote associated with Eq. (27). When the change $\Delta\epsilon$ in dielectric constant is very small, as in the case of troposphere reflections, these formulas reduce to the following simplified forms:

For vertical polarization:

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = r = \frac{\Delta\epsilon}{2} - \frac{\Delta\epsilon}{4\cos^2\phi} \tag{53a}$$

For horizontal polarization:

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = r = \frac{\Delta\epsilon}{4\cos^2\phi} \tag{53b}$$

Here ϕ is the angle of incidence at the troposphere layer measured with respect to the vertical, as shown in Fig. 71. Under conditions involved in tropospheric reflections, ϕ is given with sufficient accuracy by the following formula:

$$\cos^2\phi = \frac{\left(\frac{d^2}{8a^2} + \frac{h_t}{a}\right)^2}{\frac{d^2}{4a^2}\left(1 + \frac{h_t}{a}\right) + \left(\frac{h_t}{a}\right)^2} \tag{54}$$

The notation is illustrated in Fig. 71. This equation assumes that the reflecting sur-

[1] Experimental evidence of such reflections is given by A. W. Friend and R. C. Colwell, The Heights of the Reflecting Regions in the Troposphere, *Proc. I.R.E.*, Vol. 27, p. 626, October, 1939; J. H. Piddington, The Scattering of Radio Waves in the Lower and Middle Atmosphere, *Proc. I.R.E.*, Vol. 27, p. 753, December, 1939; C. R. Englund, A. B. Crawford, and W. W. Mumford, Ultra-short-wave Transmission and Atmospheric Irregularities, *Bell System Tech. Jour.*, Vol. 17, p. 489, October, 1938.

[2] O. H. Gish and H. G. Booker, Nonexistence of Continuous Intense Ionization in the Troposphere and Lower Stratosphere, *Proc. I.R.E.*, Vol. 27, p. 117, February, 1939.

face is parallel to the earth, and that the antenna heights are small compared with the height of the reflecting boundary.

It is to be noted that the reflection coefficient tends to increase as the wave strikes the reflecting layer with a more glancing angle, and will approach unity when the angle of incidence approaches 90°. Such conditions may occur if the transmitting antenna is very high, or if the plane of the reflecting boundary is tilted with respect to the surface of the earth. With vertically polarized waves the reflection coefficient will be zero when $\phi = 45°$. This is the Brewster angle, and with angles appreciably more glancing, the reflection coefficient is approximately the same for both vertically and horizontally polarized waves. The reflection coefficient is not affected by frequency.

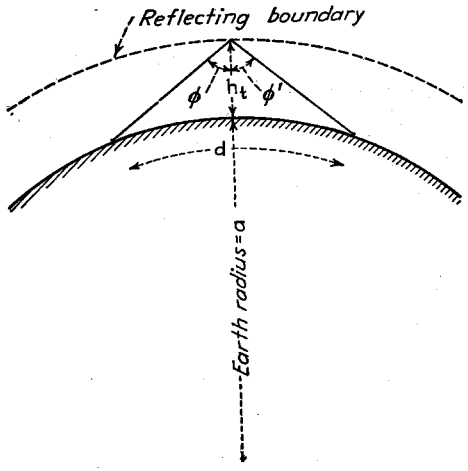


FIG. 71.—Diagram illustrating notation of Eq. (54).

The maximum possible range for a tropospheric wave when the transmitting and receiving antennas are elevated is

$$\left.\begin{matrix}\text{Range of tropospheric}\\ \text{wave, miles}\end{matrix}\right\} = \sqrt{2h_1} + 2\sqrt{2h_t} + \sqrt{2h_2} \tag{55}$$

The heights are in feet, the range in miles. The notation and also the general situation

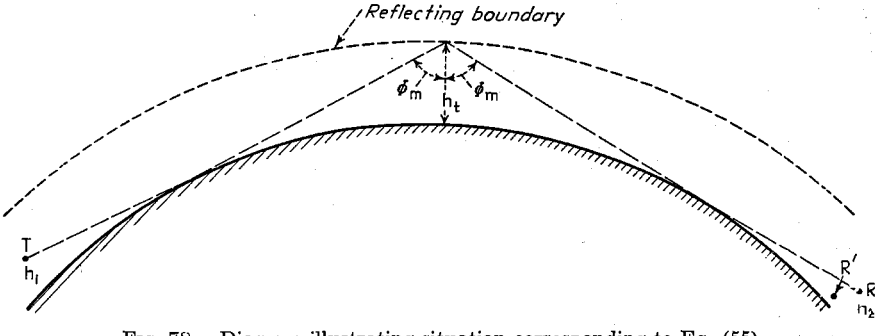


FIG. 72.—Diagram illustrating situation corresponding to Eq. (55).

involved are illustrated in Fig. 72. With the conditions shown in this figure the angle of incidence ϕ has the value ϕ_m such that

$$\cos^2 \phi_m = \frac{\dfrac{2h_t}{a}}{1 + \dfrac{3h_t}{2a}} \tag{56}$$

When the receiving point is at a greater distance than that given by Eq. (55), as, for example, if the receiver should be at R' in Fig. 72, the energy of the wave reflected

from the troposphere must diffract around the earth's curvature in order to reach the receiver. The situation then is very much as though there were in the troposphere a source of radio waves having an antenna height corresponding to the height of the reflecting boundary and transmitting to the receiving point. Inasmuch as propagation under such conditions is affected by the frequency, the strength of the tropospheric waves can be expected to be affected by frequency if the distance is greater than given by Eq. (55).

The waves reflected from the troposphere can be expected to vary widely in intensity under different conditions, and to be correlated with the weather. At times the discontinuities in dielectric constant may reach 10^{-4}. At other times they may be either negligible or entirely nonexistent, in which case no tropospheric waves are produced. It is also possible for two or more discontinuities to exist simultaneously. Furthermore, since the troposphere is the region in which "weather" occurs, a correlation between tropospheric effects on radio waves and weather is to be expected. This has been found, although the data are too few to permit any general conclusions being drawn as to the nature of the correlation.

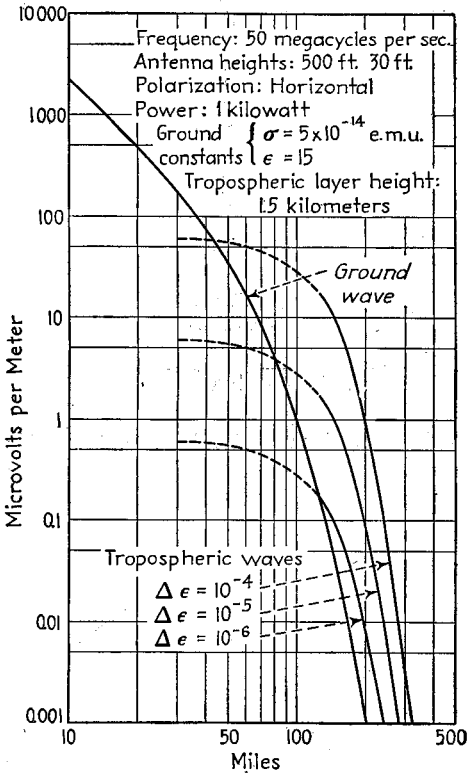


FIG. 73.—Effect of magnitude of discontinuity in dielectric constant $\Delta\epsilon$ at the reflecting boundary.

Tropospheric reflections are of especial importance when the receiving point is well below the sight of the transmitter. The waves received via the troposphere are then frequently of much greater intensity than those arriving at the receiving point by diffraction of the ground wave around the curvature of the earth. The behavior that can then be expected under various conditions is given in Figs. 73 to 76, which are theoretical curves based upon reasonable assumptions.[1]

Vertical vs. Horizontal Polarization. The effect of polarization on field strength is shown in Fig. 66. There is very little difference except at very low antenna heights and when the height is not negligible compared with the distance. At low heights the field strength is greater with vertical polarization, as is to be expected from Figs. 21*a* and *b*. Also, with the combination of large heights and small distance the angle of incidence at the ground approaches the pseudo-Brewster angle, so that the reflection coefficient with vertical polarization becomes small. The amplitude of the ground-reflected wave is then much less with vertical than with horizontal polarization, causing the oscillations in the field strength to be less pronounced.

[1] These are from K. A. Norton, A Theory of Tropospheric Wave Propagation, *F.C.C. Mimeo. Rept.* 40,003, Mar., 18, 1940.

Interference at ultra-high frequencies caused by man-made sources such as automobiles normally produces stronger vertically polarized waves than horizontally polarized waves.[1] This results from the fact that such noise sources are ordinarily located close to the surface of the earth, and it has been shown above that when the antenna is close to the earth, vertical polarization gives a stronger field than does horizontal polarization (see Figs. 21*a* and *b*).

When all considerations are taken into account, it is found that the ratio of signal to noise with horizontal polarization will, in general, be superior to that obtained with vertical polarization, particularly when the height of the receiving antenna is not

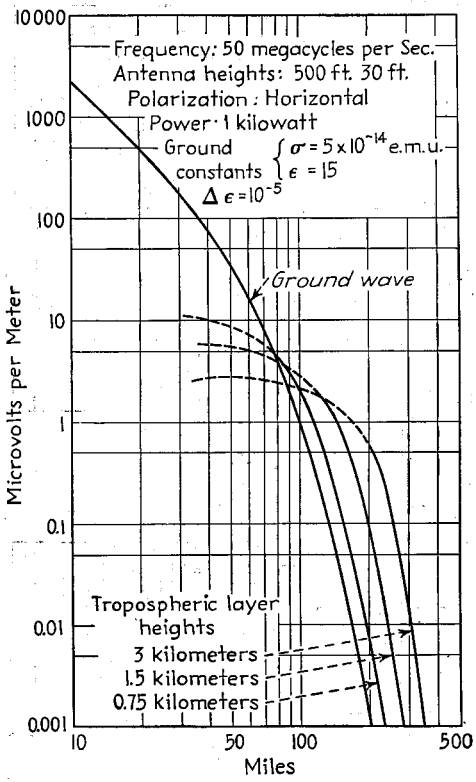


FIG. 74.—Effect of layer height on tropospheric waves in a typical case.

particularly great. In the case of transmission between mobile units, where both transmitting and receiving antennas are located close to the earth, vertical polarization is usually superior because of the greater field strength that is produced, though the signal-to-noise ratio may be no better than with horizontal polarization. The difference between the two types of polarization tends to disappear at moderate to large heights and distances.[2]

18. Static and Noise.—Radio disturbances caused by natural causes such as lightning, electrical storms, etc., are designated variously as static, atmospherics, strays. The effects produced by such disturbances on radio receivers are also some-

[1] A discussion of ignition noise is given by R. W. George, Field Strength of Motorcar Ignition Between 40 and 450 Megacycles, *Proc. I.R.E.*, Vol. 28, p. 409, September, 1940.

[2] George H. Brown, Vertical vs. Horizontal Polarization, *Electronics*, Vol. 13, p. 20, October, 1940.

times termed noise, although this word is sometimes used to indicate electrical disturbances that have a man-made origin, as, for example, disturbances produced by electric razors, power lines, etc.

Types and Sources of Static.—Static disturbances as received in a receiver can be divided into the following types:[1] (1) impulses of very high intensity, occurring only intermittently; such static is caused by local thunderstorms. (2) A steady rattling or crackling that is due to electrical disturbances produced by distant thunderstorms. (3) A steady hiss type of static observed at high frequencies, apparently having an interstellar origin. This type of static produces a sound similar to that caused by

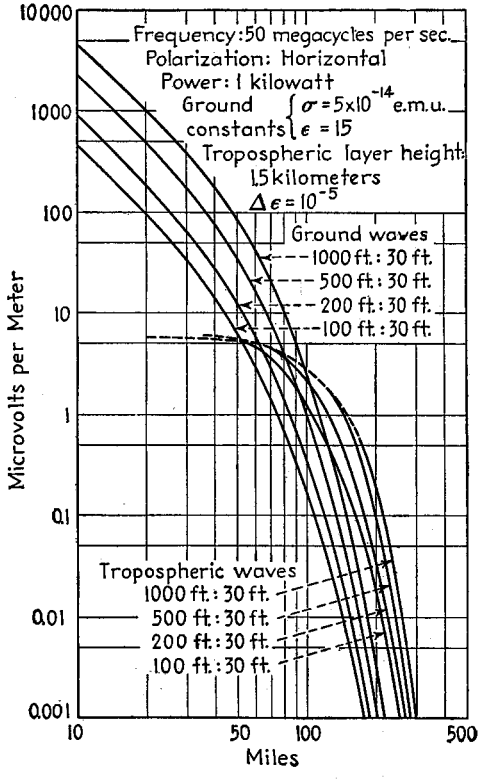


FIG. 75.—Effect of antenna height on tropospheric waves in a typical case. Receiving antenna 30 ft. high, transmitting antenna 100 to 1,000 ft. as indicated.

thermal agitation, and its direction of arrival indicates the possibility of a source in the direction of the Milky Way system.[2] (4) Precipitation static (see below).

Mountains, desert areas, and the tropics are important sources of static interference. Storm centers, particularly of tropical hurricanes, are common causes of static, and it has been found possible to trace the course of a tropical hurricane by taking bearings upon the static received at a moderately distant point.[3]

[1] Karl G. Jansky, Directional Studies of Atmospherics at High Frequencies, *Proc. I.R.E.*, Vol. 20, p. 1920, December, 1932.

[2] Karl G. Jansky, A Note on the Source of Interstellar Interference, *Proc. I.R.E.*, Vol. 23, p. 1158, October, 1935.

[3] S. P. Sashoff and J. Weil, Static Emanating from Six Tropical Storms and Its Use in Locating the Position of the Disturbance, *Proc. I.R.E.*, Vol. 27, p. 696, November, 1939; S. W. Dean, Correlation of Directional Observations of Atmospherics with Weather Phenomena, *Proc. I.R.E.*, Vol. 17, p. 1185, July, 1929.

When most of the static arriving at a receiving location is of distant origin it is commonly found that these sources tend to be in certain directions. Thus observations of long-wave static received in Maine indicate that, barring local thunderstorms, most of the disturbances have a southwesterly origin apparently in the Gulf of Mexico or Texas, while similar observations in Europe indicate sources in Africa.[1] The same is also true of static received at high frequencies.[2]

When the static is of distant origin it is found that the same impulse will simultaneously produce interference at widely distant points, as, for example, Maine and

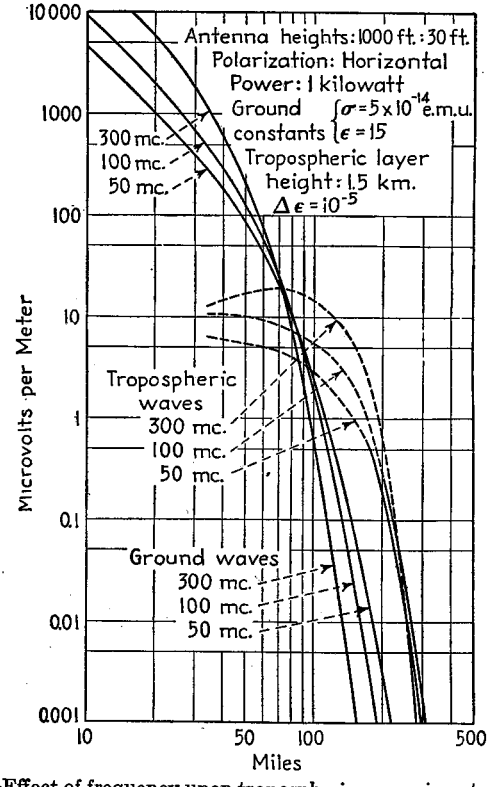


FIG. 76.—Effect of frequency upon tropospheric waves in a typical case.

England.[3] It is also possible at times to correlate impulses received at low frequencies with corresponding high-frequency static.

Effect of Frequency on Propagation Characteristics of Static.—The field strength of static appears on the average to be approximately inversely proportional to frequency.[4] This indicates that the electrical discharge that generates the radio wave representing

[1] A. E. Harper, Some Measurements on the Directional Distribution of Static, *Proc. I.R.E.*, Vol. 17, p. 1214, July, 1929.

[2] See Karl G. Jansky, Directional Studies of Atmospherics at High Frequencies, *Proc. I.R.E.*, Vol. 20, p. 1920, December, 1932.

[3] S. W. Dean, Long-distance Transmission of Static Impulses, *Proc. I.R.E.*, Vol. 19, p. 1660, September, 1931.

[4] R. K. Potter, An Estimate of the Frequency Distribution of Atmospheric Noise, *Proc. I.R.E.*, Vol. 20, p. 1512, September, 1932; L. W. Austin, The Present Status of Radio Atmospheric Disturbances, *Proc. I.R.E.*, Vol. 14, p. 133, February, 1926.

the static interference is a pulse having relatively long duration, and there is experimental evidence to bear this out.[1]

When the static is of local origin the intensity accordingly is less the higher the frequency and, in general, becomes quite small at ultra-high and micro-wave frequencies.[2] In the case of distant sources, there is superposed upon this effect, the propagation characteristics over the path from the source to the receiving point. Thus at ultra-high and micro-wave frequencies sky-wave transmission is not possible, and distant sources cannot produce static interference. Similarly, at high frequencies, the range at which a static source can produce interference on a particular frequency will depend upon the time of day, the distance, etc. As a consequence, high-frequency static tends to be greatest when conditions are most favorable for the propagation of long-distance signals, since then distant static sources become increasingly effective.[3]

At broadcast frequencies, the daytime static is largely of local origin, whereas at night one has not only the local static but also that generated by distant sources. The static level on the average is hence appreciably higher at night, when propagation conditions are good.

Static at the lowest radio frequencies has a very high intensity. This is in part because the natural sources of static generate greater intensities at very low frequencies and in part because low-frequency radio waves propagate very efficiently over great distances under practically all conditions.

Means of Overcoming Static.—The static energy absorbed by a receiver is proportional to the width of the response band of the receiving system. Accordingly, the first and most obvious means of minimizing static interference is to design the receiving system so that the response band is no wider than that necessary to accommodate the side bands of the desired signal.[4]

The interference produced by static in a receiver can be reduced by limiting, as discussed in Par. 6, Sec. 9. Such limiting is particularly effective in cases where the static consists of occasional crashes of very large amplitude, but tends to become ineffective when the static is of the continuous crackling or rattling sort, since then the peak amplitudes are smaller in proportion to the average amplitude of the static energy.

Frequency modulation is an effective means of minimizing the effect of static, since when the signal is appreciably stronger than the static the latter is almost completely suppressed, and does not produce a background of disturbance as with amplitude modulation. Furthermore, even when the static is stronger than the signal and so suppresses the latter, the usual frequency-modulation receiving system possesses a limiting action that minimizes the intensity of the static disturbance produced even during these moments. Unfortunately, however, frequency modulation, because of its wide-band requirements, can be used only at the ultra-high frequencies, where the static is least severe, and is not available at lower frequencies, where static tends to give more trouble.

[1] J. E. I. Cairns, Atmospherics at Watheroo, Western Australia, *Proc. I.R.E.*, Vol. 15, p. 985, December, 1927.

[2] For example, it has been found that a lightning flash 5 miles away will produce only a 5 $\mu v/m$ field at 150 mc with a reception band width of 1.5 mc. See J. P. Schafer and W. M. Goodall, Peak Field Strength of Atmospherics Due to Local Thunderstorms at 150 Megacycles, *Proc. I.R.E.*, Vol. 27, p. 202, March, 1939.

[3] A classic investigation of high-frequency static is reported by R. K. Potter, High-frequency Atmospheric Noise, *Proc. I.R.E.*, Vol. 19, p. 1731, October, 1931.

[4] See John R. Carson, Selective Circuits and Static Interference, *Trans. A.I.E.E.*, Vol. 43, p. 789, 1924. This is a classical paper in which it is shown that, if static is a random series of impulses, then the amount of static energy absorbed by a receiver is directly proportional to the frequency range to which the receiver responds.

The received static intensity can be reduced by employing directional receiving antennas except when the principal sources of static happen to lie in exactly the same direction as the transmitting station being received. Directional antennas are particularly effective in reducing static in the case of long-wave transatlantic communication, for nearly all the static except that from local sources has a southerly origin.

When the full possibilities of narrow-band reception, limiting, and directional receiving antennas have been utilized, the only further means of combating static is to provide a stronger signal. This may be done either by greater transmitter power or by the use of directivity at the transmitting antenna to concentrate the available energy more strongly in the direction of the receiving point.

Much effort has been expended in attempting to devise "static eliminators." These have commonly taken the form of attempting to balance static received on one frequency with a signal, against static received free of signal on a slightly different frequency. Investigation of such systems has shown, however, that no appreciable gain results from these arrangements. The best that can be hoped from them is the elimination of static in the absence of a signal, which benefit is counterbalanced by more static energy being delivered to the receiver output when the signal is present than in the case of a simple receiving system that does not attempt to balance out the static.[1]

Precipitation Static.[2]—The term *precipitation static* is used to designate a type of interference commonly experienced in an airplane receiver in flying through rain, snow, hail, ice crystals, and dust clouds. It consists of a continuous interference that may completely blanket reception for appreciable periods. Precipitation static occurs when a plane passes through a region where the cloud particles are electrically charged. Under such conditions the plane will accumulate a charge, which builds up until corona breaks out at some sharp point on the plane. The static resulting from this corona discharge is "precipitation static," and consists of a strong induction noise field in which the electrostatic induction is much stronger than the magnetic component.

Precipitation static can be almost completely eliminated by (1) providing a means of electrically discharging the plane in a way that will create a minimum of radio interference, (2) using shielded loop receiving antennas. A suitable arrangement for discharging the plane consists of a five-foot length of very fine wire connected to the plane by a resistance of approximately 100,000 ohms, five feet long, made of graphite-impregnated rubber-covered rope. With such an arrangement the corona that discharges the plane is concentrated in the trailing wire, where it is located as far as possible from the plane itself, and is in the form of a "soft" corona discharge that creates a minimum of interference. The use of an electrostatically shielded loop antenna then further reduces the static disturbance reaching the receiver, since such an antenna responds only to magnetic fields, while the induction field produced by the corona discharge is primarily electrostatic.

Precipitation static is sometimes observed at ground stations in snowstorms, thunderstorms, and dust storms. The causes here are apparently either the impact of dust particles against the antenna, or the creation of induction fields by near-by corona discharges.

[1] John R. Carson, The Reduction of Atmospheric Disturbances, *Proc. I.R.E.*, Vol. 16, p. 966, July, 1928.

[2] Further discussion of this subject is given by H. M. Hucke, Precipitation-static Interference on Aircraft and at Ground Stations, *Proc. I.R.E.*, Vol. 27, p. 301, May, 1939; Howard K. Morgan, Rain Static, *Proc. I.R.E.*, Vol. 24, p. 959, July, 1936; Tomozo Nakai, Some Notes on Rain Static in Japan, *Proc. I.R.E.*, Vol. 25, p. 1375, November, 1937; E. C. Starr, Aircraft-precipitation-static Radio Interference, *Trans. A.I.E.E., Supplement*, Vol. 60, p. 363, 1941.

SECTION II

ANTENNAS

FUNDAMENTAL RELATIONS

1. Radiation. *Nature of the Radiated Wave.*—A radio wave represents electrical energy that has escaped into free space. A radio wave travels with the velocity of light, and consists of magnetic and electric fields at right angles to each other, and also at right angles to the direction of travel. The intensity of the electric and magnetic fields of the wave are such that one-half of the electrical energy contained in the wave is in the form of electrostatic energy, while the remaining half is in the form of magnetic energy. This leads to

$$\epsilon = 300H \tag{1}$$

where ϵ is the voltage gradient of the electric field in volts per centimeter and H is the magnetic field intensity in gilberts.

The strength of a radio wave is expressed in terms of the voltage stress produced in space by the electric field of the wave, and is usually expressed in either millivolts or microvolts stress per meter. The stress expressed in this way is exactly the same voltage that the magnetic flux of the wave induces in a conductor one meter long when the wave sweeps across this conductor with the velocity of light.

The plane parallel to the mutually perpendicular lines of electrostatic and magnetic flux of the wave is termed the *wave front.* The wave travels in a direction at right angles to the wave front, with the direction of travel depending upon the relative direction of the lines of electromagnetic and electrostatic flux. If the direction of either magnetic or electrostatic flux is reversed, the direction of travel is likewise reversed, but reversing both sets of flux has no effect.

The direction of the electrostatic lines of flux is termed the direction of *polarization* of the wave. Thus, when the electrostatic lines are vertical, the wave is said to be vertically polarized.

The *wave length* represents the distance traveled by the wave in a length of time corresponding to one cycle, assuming a sinusoidal force producing the wave. Since a wave travels with the velocity of light, the relationship between wave length λ in meters and frequency f in cycles per second is

$$\lambda = \frac{300{,}000{,}000}{f} \tag{1a}$$

Radiation from an Elementary Doublet.—A doublet consists of a length of conductor (antenna) short compared with a wave length, which is assumed to have such large capacity areas at its ends that the current throughout the length of the conductor can everywhere be considered the same.

The strength of the field radiated from such an elementary doublet having a length δl and carrying a current I (see Fig. 1) is given by the formula

$$\epsilon = \frac{60\pi}{d\lambda}(\delta l)I \cos \omega \left(t - \frac{d}{c}\right) \cos \theta \tag{2}$$

where ϵ = strength of wave, volts per meter, at point P (see Fig. 1).
δl = length of wire from which radiation takes place, measured in the same units as λ.
$I \cos(\omega t + 90°)$ = current flowing in wire, amp.
d = distance from P to antenna, meters.
θ = angle of elevation of point at which field is desired with respect to a plane perpendicular to the conductor δl.
f = frequency of current.
$\omega = 2\pi f$.
t = time.
c = velocity of light $= 3 \times 10^8$ meters per sec.
λ = wave length corresponding to frequency f.

The wave front of the wave radiated from a doublet is the plane perpendicular to a line drawn to the doublet. The waves are polarized in such a manner that a plane can be passed through the antenna and the electrostatic flux lines of the radiated wave.

The Radiation Field Produced by an Actual Antenna System.—The total radiation field produced by an antenna system having any known arbitrary current distribution can be found by adding the separate fields produced by the elementary lengths of such an antenna system, with each elementary length treated as a doublet antenna.[1] In adding the contributions to the total field that are made by the various doublets that can be considered as making up the actual current distribution of an antenna, it is necessary to take into account the phase and the plane of polarization of each contribution.

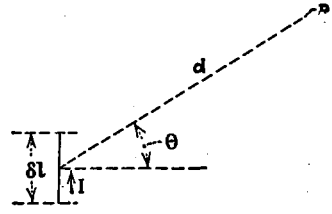


FIG. 1.—Elementary doublet consisting of a length of wire δl carrying a current I.

2. Fields in the Vicinity of an Antenna—Induction Fields.—The electric and magnetic fields in the immediate vicinity of an antenna are greater in magnitude, and differ in phase, from the radiation field as calculated with the aid of Eq. (2). The electric and magnetic fields that must be added to the radiation field in order to give the fields actually present are termed *induction fields*. These induction fields diminish in strength more rapidly than inversely proportionally to distance. Thus the induction magnetic field from a doublet is inversely proportional to the square of the distance, and the induction electric field from a doublet has one component that is inversely proportional to the square of the distance and another that is inversely proportional to the cube of the distance. Inasmuch as the radiation field is inversely proportional to the distance, the induction fields die away much more rapidly with distance than do the radiation fields, and at distances of a few wave lengths become negligible in comparison with the radiation field. However, at distances from the antenna that are small compared with a wave length (or small compared with the antenna dimensions if the antenna is large), the induction electric and magnetic fields will be much greater than the radiation field of the antenna.

The induction field also differs from the radiation field in that, unlike the latter, the magnetic and electrostatic field intensities of the induction wave are not proportional to each other, nor are they in phase.

The electric induction field becomes proportionately stronger than the magnetic induction field as the distance to the antenna becomes less in the case of a doublet

[1] Such summations can be carried out by either graphical or mathematical integration. Examples are to be found in many places in the literature, as, for example, F. E. Terman, "Radio Engineering," 2d ed., p. 658. McGraw-Hill, New York, 1937.

antenna. With a loop antenna, the magnetic induction field is increasingly predominant as the distance becomes less.

The field in the vicinity of an antenna can be determined by the following methods:[1] (1) Considering the antenna system with its current distribution to be replaced by a series of doublets, and then summing up the fields produced by the doublets, taking into account the phase and polarization. This method, although apparently simple, leads to integrations that normally cannot be made mathematically. (2) Derivation of the fields with the aid of the vector and scalar potentials. (3) Derivation of the fields from a Hertzian vector.

The vector and scalar potential method can be applied only to radiating systems that are complete in themselves (*i.e.*, in which no current either enters or leaves the radiating conductor). In order to apply this method of analysis to such cases as a terminated conductor, where the current entering and leaving the antenna system may be relatively large, it is necessary to postulate suitable terminating charges at the ends of the antenna system to take into account the effect of these terminal currents. Otherwise the calculated electrostatic fields will be in error.[2]

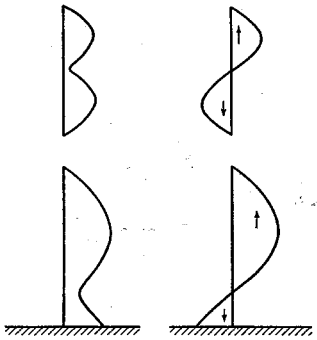


FIG. 2.—Actual and schematic current distribution in resonant antennas. The arrows represent the direction of current flow (*i.e.*, the polarity).

3. Current Distribution in Antenna Systems.—An antenna is a circuit having distributed constants. The current distribution accordingly has the same general character as in a transmission line. Although some deviation from transmission-line behavior may be expected because the inductance and capacity per unit length of an antenna system are not the same for all parts of the antenna, experiments have shown that these effects are quite small under most conditions.

Resonant Antennas.—In the case of a wire having an open end, the current will be distributed as in Fig. 2*a*. It is zero at the open end and passes through minima at distances that are multiples of a half wave length away from the end.[3] The currents in the loops on either side of a current minimum are almost exactly 180° out of phase except in the immediate vicinity of the minimum, which is where substantially all the phase change takes place. Since the current minima are quite small with respect to the current maxima, it is common practice to picture the current distribution as in Fig. 2*b*, where the current at the minima is assumed to be zero instead of merely small, and the currents in adjacent loops are taken as exactly 180° out of phase. The equa-

[1] The fields existing in the immediate vicinity of a doublet antenna are given by C. R. Englund and H. T. Friis, Methods for the Measurement of Radio Field Strength, *Trans. A.I.E.E.*, Vol. 146, p. 492, 1927.

Use of the vector and scalar potential method is discussed by P. S. Carter, Circuit Relations in Radiating Systems and Application to Antenna Problems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932; F. R. Stansel, A Study of the Electromagnetic Field in the Vicinity of a Radiator, *Proc. I.R.E.*, Vol. 24, p. 802, May, 1936.

The Hertzian vector method is described by R. Bechmann, Calculation of Electric and Magnetic Field Intensities of Oscillating Straight Conductors, *Proc. I.R.E.*, Vol. 19, p. 461, March, 1931. Both the Hertzian vector, and vector and scalar potential methods are considered by Andrew Alford, A Discussion of Methods Employed in Calculations of Electromagnetic Fields of Radiating Conductors, *Elec. Comm.*, Vol. 15, p. 70, July, 1936.

[2] See Alford, *loc. cit.;* also, F. M. Colebrook, The Electric and Magnetic Fields of a Linear Radiator Carrying a Progressive Wave, *Jour. I.E.E.*, Vol. 86, p. 169, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 17, 1940.

[3] End effects cause the current loop adjacent to an open end to be a few per cent less than a half wave length on the basis of a radio wave traveling through space.

tion of the resulting current distribution for this idealized case is

$$\text{current} = I_0 \sin\left(\frac{2\pi x}{\lambda}\right) \sin \omega t \tag{3}$$

where I_0 = current at current maximum.
x/λ = distance in wave lengths from open end.
$\omega = 2\pi \times$ frequency.
t = time.

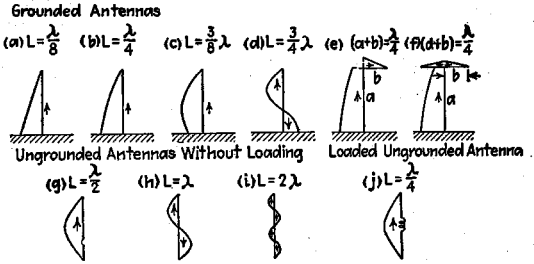
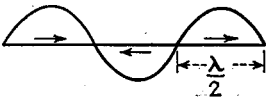
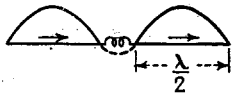


FIG. 3.—Current distribution in typical antennas, assuming sinusoidal distribution of current.

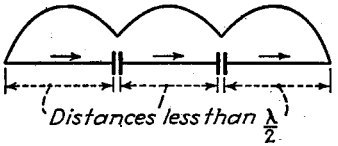
The current distribution represented by Eq. (3) is that which would be obtained if the antenna acted exactly as a transmission line with zero losses. This condition is never realized perfectly in practice, because obviously an antenna will radiate some energy and therefore will consume some energy. Practically, however, radiation field patterns calculated on the assumptions implied by Eq. (3) are sufficiently accurate for nearly all purposes.[1]



(a) Antenna without loading



(b) Antenna with series inductance



(c) Antenna with series condensers

FIG. 4.—Typical current distributions with loaded antennas.

Typical examples of current distribution based on Eq. (3) are shown in Fig. 3. In each case, the current is assumed to follow a sinusoidal distribution and to be zero at the open end. When the lower end of the antenna is grounded, as in *a* to *f*, the current distribution is made up of a section of a sine wave. A change in direction of the antenna, as, for example, in *e*, has no effect upon the distribution, at least to a first approximation.

The current distribution of an antenna can be modified by adding suitably spaced series inductances or series capacities, as illustrated in Fig. 4. The effect of a series inductance is to shorten the physical distance between adjacent current minima, as well as to reduce the maximum current in the section of the antenna containing the inductance. This is illustrated in Fig. 4*b*, and if the inductance is sufficiently large, it is possible to suppress a current loop almost entirely. In contrast with this, a series

[1] A discussion of the actual current distribution in a vertical grounded antenna, as used in broadcast work, and the effect on the field pattern of departure from the sinusoidal assumption is given by W. L. McPherson, Electrical Properties of Aerials for Medium and Long Wave Broadcasting, *Elec. Comm.*, Vol. 17, p. 44, July, 1938; see also R. M. Wilmotte, *Jour. I.E.E.*, Vol. 66, p. 617, 1928, or Wireless Section, *I.E.E.*, Vol. 3, p. 136, 1928.

condenser increases the distance between adjacent minima to a value greater than a half wave length. If a number of spaced condensers are employed somewhat less than a half wave apart, it is possible to eliminate all current minima except those at the ends of the wire, as shown in Fig. 4c, and cause the current to have substantially the same phase throughout the entire length of the antenna.

Nonresonant Antennas.—When an antenna is terminated by an impedance equal to the characteristic impedance of the antenna wire considered as a transmission line,

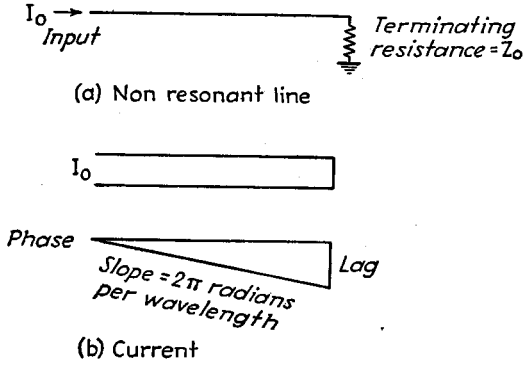


FIG. 5.—Current and phase distribution in a nonresonant antenna.

as in Fig. 5, the current distribution, instead of having the resonant character illustrated in Fig. 2, is nonresonant, as in Fig. 5. Such a current distribution is characterized by progressive uniform distributed phase shift amounting to 2π radians per wave length, and to the extent that the energy losses can be neglected, the current is everywhere of constant amplitude. Further discussion of such current distributions is given in Par. 11, Sec. 3.

The phase shift associated with a nonresonant current distribution can be modified by means of series inductances or series capacities spaced not over a quarter wave length apart. Series inductances introduce an additional phase lag as one travels along the antenna toward the terminated end, making the average phase shift per unit length greater. Series condensers have the opposite effect, causing the average phase lag per unit length to be less.

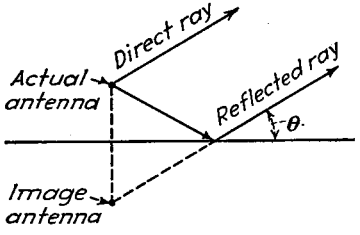


FIG. 6.—Diagram illustrating how the reflected wave from the earth can be considered to be produced by an image antenna.

4. Effect of Ground; Image Antennas.—When an antenna is near the ground, energy radiated toward the earth is reflected as shown in Fig. 6. The total field in any direction then represents the vector sum of a direct wave plus a reflected wave.

For purposes of calculation, it is convenient to consider that the reflected wave is generated not by reflection but rather by a suitable image antenna located below the surface of the ground. This image antenna has a physical configuration that is the mirror image of the actual antenna, as illustrated in Fig. 7. In the case of a perfect earth (infinite conductivity) the reflection coefficient is unity. The currents in corresponding parts of the actual and image antennas (*i.e.*, parts lying on the same vertical line and at the same distance from the earth's surface) are of the same magnitude. The direction of current flow is such that the vertical component of the current in the

image is in the same direction as the vertical component of the current in the corresponding part of the actual antenna, and the horizontal component of the current in the image is directed oppositely from the horizontal component of the current in

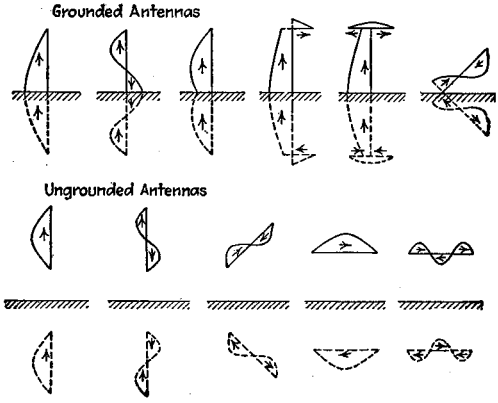


FIG. 7.—Images for common types of antennas.

the corresponding part of the actual antenna. The application of these principles to the determination of the direction of current flow in the image in a doubtful case is illustrated in Fig. 8, and the images for a number of typical cases are given in Fig. 7.

In the case of finite earth conductivity, the magnitude and phase of the image currents are modified in accordance with the reflection coefficient calculated on the basis of the angle of reflection θ in Fig. 6. That is, the vertically polarized component of the reflected wave can be considered as being produced by an image antenna in which the currents, instead of being the same as with a perfect earth, are the currents in this ideal case multiplied by the magnitude of the reflection coefficient for vertically polarized waves as calculated for the angle θ, with the phase of the currents in the image adjusted in accordance with the phase angle of the reflection coefficient. It will be noted that in making such a calculation, the departure of the current in the actual image from that existing in the ideal case will vary with the angle of reflection θ, and will likewise be different for the vertically and horizontally polarized components of the reflected wave.

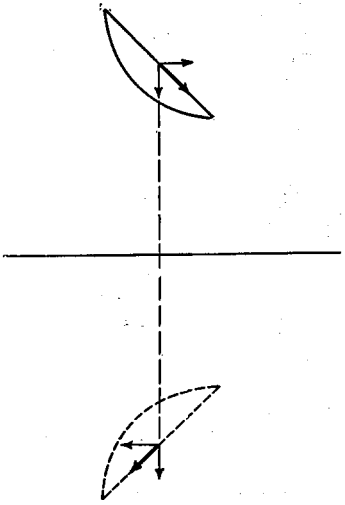


FIG. 8.—Diagram showing horizontal and vertical components of current flow in actual antenna and in image.

In practical antenna calculations, it is usually permissible to assume that the earth is a perfect reflector. This is justified by the fact that the errors thus introduced are of secondary importance for most purposes.

The electric and magnetic fields existing in the space above the earth in the presence of a perfect earth are exactly the same fields that would be produced in this same region by the joint action of the actual antenna and its image, with the earth removed. This is not true in the case of an imperfect earth, however, since although

the use of an image modified in accordance with the reflection coefficient gives the proper reflected ray, the effect that the current in the imperfect earth has on the field just above the earth is not properly taken into account.

5. Impedance and Mutual Impedance of Antennas. *Self-impedance.*—The impedance of an antenna system is the impedance offered to a voltage applied to the two terminals obtained by opening up the antenna system at the point involved, when there are no other antennas near by.

The impedance characteristics of an antenna are similar to those of a transmission line because of the fact that the antenna is a circuit with distributed constants. Thus a vertical antenna with the lower end grounded acts like a single-wire transmission line with ground return. The impedance between the lower end of such an antenna and ground varies with frequency in the manner illustrated in Fig. 9. Similarly, the impedance at the center of an isolated ungrounded antenna is similar to the impedance that would be obtained from a two-wire transmission line having a length equal to the distance from the center of the antenna to the ends (see Fig. 9). In such a case, resonance is obtained whenever the total antenna length approximates a multiple of a half wave length for the frequency involved.[1] At frequencies on one side of a resonance point, the reactance will be inductive; on the other side of the resonance, it will be capacitive.

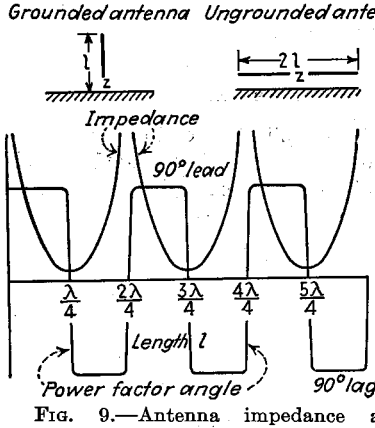


FIG. 9.—Antenna impedance and power-factor angle as a function of antenna length in wave lengths.

The resistance component of the antenna impedance takes into account the energy consumed by the antenna system as a result of radiation, ground losses, antenna wire resistance, dielectric losses, etc. In order for the antenna to be efficient, it is necessary that the various losses be small compared with the energy lost by radiation, *i.e.*, the radiation should account for as much as possible of the total antenna resistance. The radiation resistance is discussed further in Par. 6.

Mutual Impedance.—An antenna excited so that it has current flowing in it will induce a voltage in any other antenna in the vicinity. This action is as though the two antenna systems were coupled, and the behavior that results can be expressed in terms of a mutual impedance Z_{12} existing between the antennas, which is defined by the relation

$$\text{Mutual impedance} = Z_{12} = -\frac{E_2}{I_1} \tag{4}$$

where I_1 is the current flowing at the reference point in antenna 1 and E_2 is the voltage that would have to be applied to the reference point in antenna 2 to produce in this antenna, with antenna 1 removed, the current that actually flows in it; *i.e.*, E_2 is the equivalent voltage induced by antenna 1 in antenna 2 when this induced voltage is referred to the reference point in the second antenna.

[1] The antenna length required for resonance is always slightly less (of the order of 5 per cent) than a multiple of a half wave length based on the wave length in free space. This is because of end effects and couplings between parts of the same wire. See C. R. Englund, The Natural Period of Linear Conductors, *Bell System Tech. Jour.*, Vol. 7, p. 404, July, 1928; P. S. Carter, Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932.

In defining mutual impedance, it is necessary to do so in terms of a reference point in each antenna, *i.e.*, the mutual impedance represents the equivalent coupling between the antennas expressed in terms of the current at some specified point in the first antenna and the resulting effect at some specified point in the second antenna. These specified points can be arbitrarily chosen in whatever manner is most convenient, and must be specified or implied. If not specified, it is taken for granted that the points of reference in both antennas are at current loops.

The mutual impedance between two antennas is calculated by assuming a convenient current flowing in one antenna and then calculating the resulting electric field existing at the position of the second antenna and parallel with this antenna. This electric field then gives the distributed voltages that are induced in the second antenna by the current in the first antenna. The voltage thus induced in each differential length of the second antenna has the same effect, insofar as current in the antenna

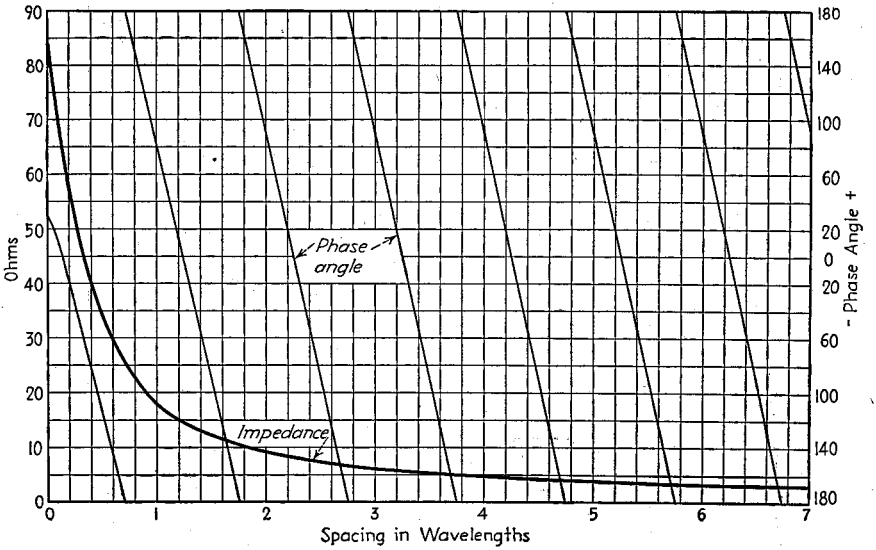


FIG. 10.—Mutual impedance between parallel half-wave radiators.

system is concerned, as some other induced voltage acting at the reference point, and the total voltage E_2 is obtained by integration of these equivalent voltages referred to the reference point.[1]

Values of Mutual Impedance in Typical Cases.—The magnitude and phase angle of the mutual impedance that can be expected between two antennas will depend upon the spacing and upon the geometrical configuration. The magnitude will be smaller the greater the spacing, and will be greater the larger the antenna dimensions in wave lengths. The phase angle will vary greatly, depending upon the spacing and the geometry, and may have any value from 0 to 360°. Phase angles between 90 and 270° signify that the resistive component of the mutual impedance is negative.

Values of mutual impedance in some of the more important cases are given in Figs. 10, 11, and 12. The resistance component of the mutual impedance is given in Figs. 13, 14, 15 and in Table 1 for a number of common cases. Formulas for calculating

[1] An excellent discussion of the actual details of making such calculations, together with a number of examples, is given by P. S. Carter, Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932. See also G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937.

these and other cases are to be found in the literature but are too involved to be given here. In some cases the formula involves an integration that can only be carried out graphically, and even when this is not the case, graphical methods are often the easiest to use.[1]

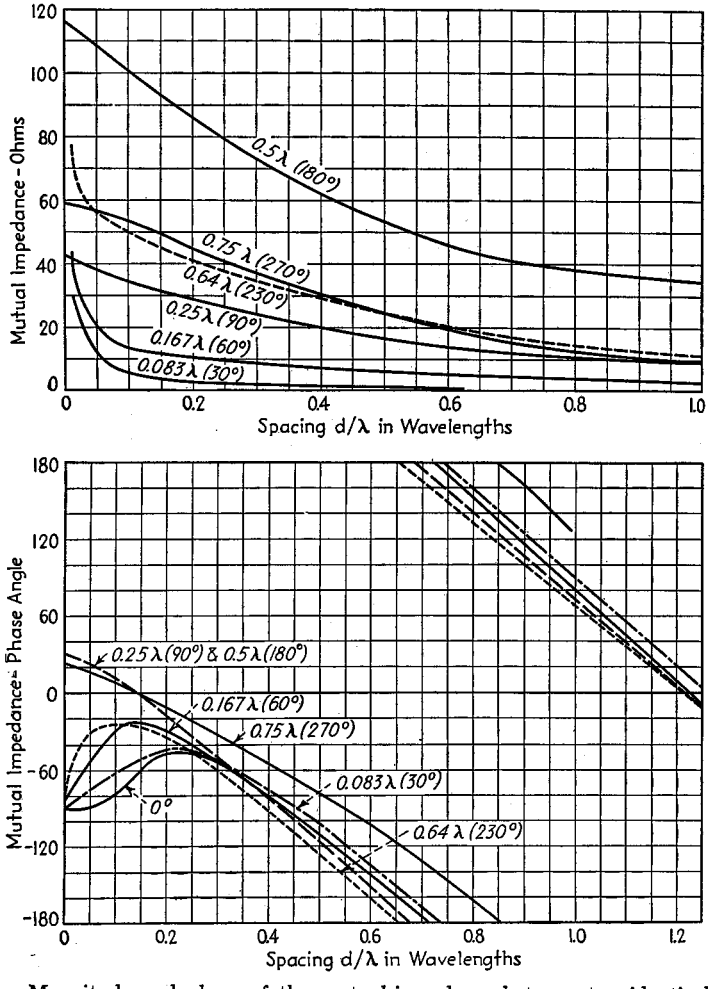


FIG. 11.—Magnitude and phase of the mutual impedance between two identical vertical grounded antennas. The designations of the curves are antenna heights.

Voltage and Current Relations in Systems Involving Coupled Antennas.—The voltage and current relations existing in a system of radiators between which mutual imped-

[1] The results given in Figs. 10 to 15 and Table 1 are from Brown, *loc cit.;* Carter, *loc. cit.;* and A. A. Pistolkors, The Radiation Resistance of Beam Antennas, *Proc. I.R.E.*, Vol. 17, p. 562, March, 1929. The Carter paper also gives formulas for the mutual impedance between parallel wires of equal length (both staggered and not staggered), colinear wires, and for two wires forming at V. The Brown paper gives formulas for vertical grounded antennas such as are used in broadcast work. Mutual impedance between two antennas lying in the same plane, but not necessarily parallel, is given by F. H. Murray, Mutual Impedance of Two Skew Antenna Wires, *Proc. I.R.E.*, Vol. 21, p. 154, January 1933.

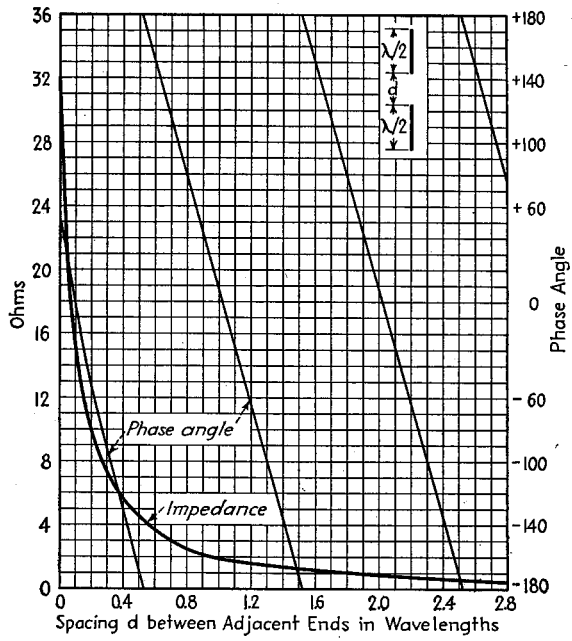


FIG. 12.—Magnitude and phase angle of mutual impedance between two colinear half-wave-length resonant antennas.

TABLE 1.—RESISTANCE COMPONENT OF MUTUAL IMPEDANCE EXISTING BETWEEN HALF-WAVE-LENGTH PARALLEL RADIATORS

Spacing d (see Fig. 16) in λ	Relative height h (see Fig. 16) in wave lengths						
	0.0	0.5	1.0	1.5	2.0	2.5	3.0
0.0	+73.29	+26.40	−4.065	+1.78	−0.96	+0.58	−0.43
0.5	−12.36	−11.80	−0.78	+0.80	−1.00	+0.45	−0.30
1.0	+ 4.08	+ 8.83	+3.56	−2.92	+1.13	−0.42	+0.13
1.5	− 1.77	− 5.75	−6.26	+1.96	+0.56	−0.96	+0.85
2.0	+ 1.18	+ 3.76	+6.05	+0.16	−2.55	+1.59	−0.45
2.5	− 0.75	− 2.79	−5.67	−2.40	+2.74	−0.28	−0.10
3.0	+ 0.42	+ 1.86	+4.51	+3.24	−2.07	−1.59	+1.74
3.5	− 0.33	− 1.54	−3.94	−3.76	+0.74	+2.66	−1.03
4.0	+ 0.21	+ 1.08	+3.08	+3.68	+0.51	−2.49	−0.09
4.5	− 0.18	− 0.85	−2.50	−3.40	−1.30	+2.00	+1.12
5.0	+ 0.15	+ 0.69	+2.10	+3.14	+1.82	−1.35	−1.87
5.5	− 0.12	− 0.57	−1.80	−2.90	−2.24	+0.49	+1.77
6.0	+ 0.12	+ 0.51	+1.56	+2.61	+2.28	−0.06	−2.02
6.5	− 0.10	− 0.45	−1.18	−2.31	−2.29	−0.45	+1.71
7.0	+ 0.06	+ 0.36	+1.14	+2.06	+2.26	+0.85	−1.32
7.5	− 0.03	− 0.30	−1.00	−1.86	−2.14	−1.03	+0.66

ances exist have exactly the same form as the voltage and current relations in a general network (see Par. 20, Sec. 3), and so can be written as

$$\left.\begin{aligned} E_1 &= I_1Z_{11} + I_2Z_{12} + \ldots + I_nZ_{1n} \\ E_2 &= I_1Z_{12} + I_2Z_{22} + \ldots + I_nZ_{2n} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ E_n &= I_1Z_{1n} + I_2Z_{2n} + \ldots + I_nZ_{nn} \end{aligned}\right\} \tag{5}$$

where E_1, E_2, etc. = voltages applied to antennas 1, 2, etc.
I_1, I_2, etc. = currents flowing in antennas 1, 2, etc.
Z_{11}, Z_{22}, etc. = self-impedances of antennas 1, 2, etc.
Z_{12}, Z_{2n}, etc. = mutual impedances between antennas denoted by subscripts.

In applying Eq. (5), it is to be noted that if some of the individual radiators involved are not excited, then the corresponding applied voltages are zero, whereas if

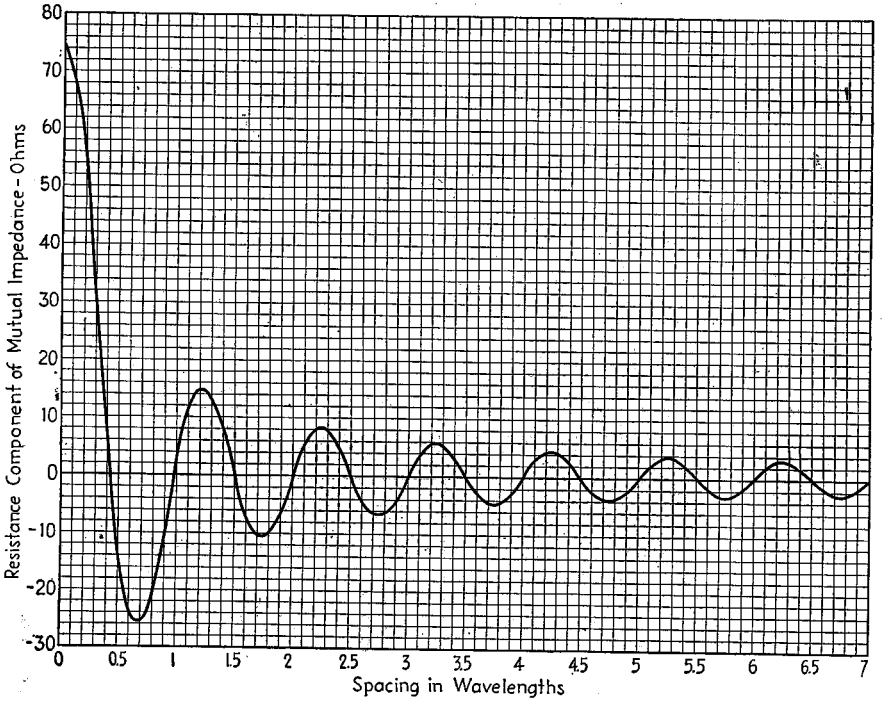


FIG. 13.—Resistance component of mutual impedance between two parallel nonstaggered half-wave-length antennas.

certain of the antennas are open-circuited so that they carry no current, then the currents in these antennas are assigned the value zero in the equations.

The impedance offered to an exciting voltage applied to the kth antenna (the driving-point impedance) is the vector ratio E_k/I_k obtained by dividing the right-hand side of the kth equation of Eq. (5) by I_k. Thus for $k = 1$

$$\left.\begin{matrix}\text{Driving-point} \\ \text{impedance for} \\ \text{first antenna}\end{matrix}\right\} = \frac{E_1}{I_1} = Z_{11} + \frac{I_2}{I_1}Z_{12} + \frac{I_3}{I_1}Z_{13} + \ldots \tag{6}$$

The effect that a perfect earth has on the impedances existing in an antenna system can be taken into account by replacing the earth by a suitable image antenna system. The impedance and voltage and current conditions that exist in the actual antenna

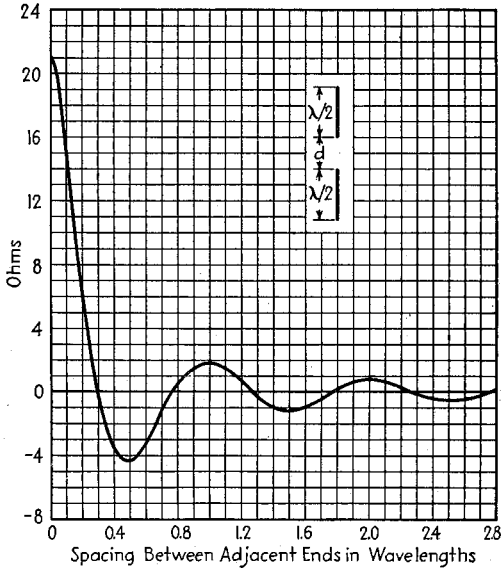


FIG. 14.—Resistance component of mutual impedance between two colinear half-wave antennas.

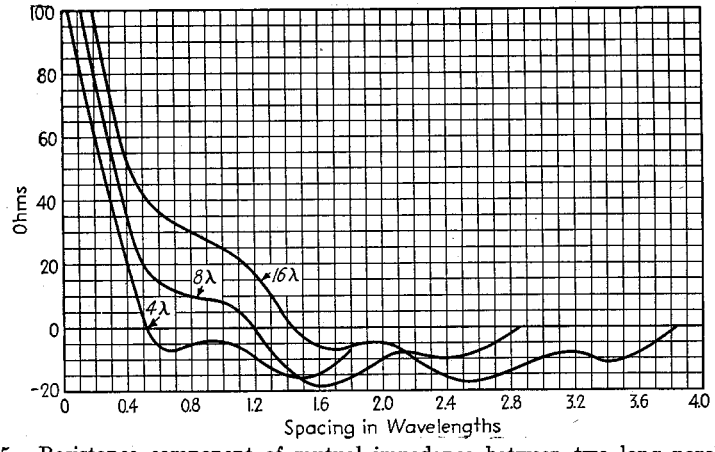


FIG. 15.—Resistance component of mutual impedance between two long parallel non-staggered resonant wires.

system in the presence of such an image antenna system, in which all the image exciting voltages are taken as zero, are then the conditions in the actual antenna in the presence of a perfect earth.

When the action of the earth cannot be approximated by a perfect reflector, the effect of the earth on the antenna impedance will be modified somewhat. The analysis

of the exact behavior then becomes quite involved, and is generally not carried out, because the impedance modifications introduced by the imperfections of the earth are not of great importance under most practical conditions.[1]

Measurement of Self- and Mutual Impedances.—The various self- and mutual impedances involved in a system of antennas can be readily measured by determining the driving-point impedances at the various reference points for two or more conditions. For example, in a system involving three antennas, Eq. (5) takes the form

$$\left.\begin{aligned} E_1 &= I_1Z_{11} + I_2Z_{12} + I_3Z_{13} \\ E_2 &= I_1Z_{12} + I_2Z_{22} + I_3Z_{23} \\ E_3 &= I_1Z_{13} + I_2Z_{32} + I_3Z_{33} \end{aligned}\right\}. \tag{7}$$

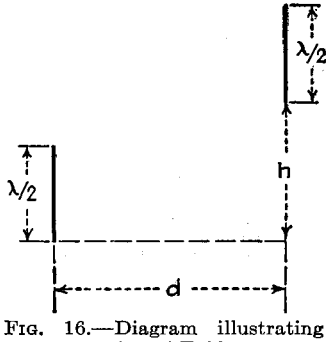


FIG. 16.—Diagram illustrating notation of Table 1.

The self-impedance Z_{11} of the first antenna can then be determined by open-circuiting the second and third antennas, thus making I_2 and I_3 zero.[2] Z_{11} is then the impedance measured at the reference point in antenna 1, *i.e.*, $Z_{11} = E_1/I_1$. The mutual impedance Z_{12} can then be measured by open-circuiting antenna 3, thereby putting it out of commission and measuring the driving-point impedance at the reference point in antenna 1. Under these conditions, one has

$$\left.\begin{aligned} E_1 &= I_1Z_{11} + I_2Z_{12} \\ 0 &= I_1Z_{12} + I_2Z_{22} \end{aligned}\right\} \tag{8a}$$

Simultaneous solution of these equations to obtain the driving-point impedance E_1/I_1 gives

$$\text{Driving-point impedance} = \frac{E_1}{I_1} = Z_{11} - \frac{Z_{12}^2}{Z_{22}} \tag{8b}$$

Thus a measurement of the driving-point impedance E_1/I_1, together with a knowledge of the self-impedances Z_{11} and Z_{22}, which can be determined separately as described above, gives the information needed to determine the mutual impedance Z_{12}. Mutual impedances between other pairs of antennas can be determined in a similar manner from impedance measurements.[3]

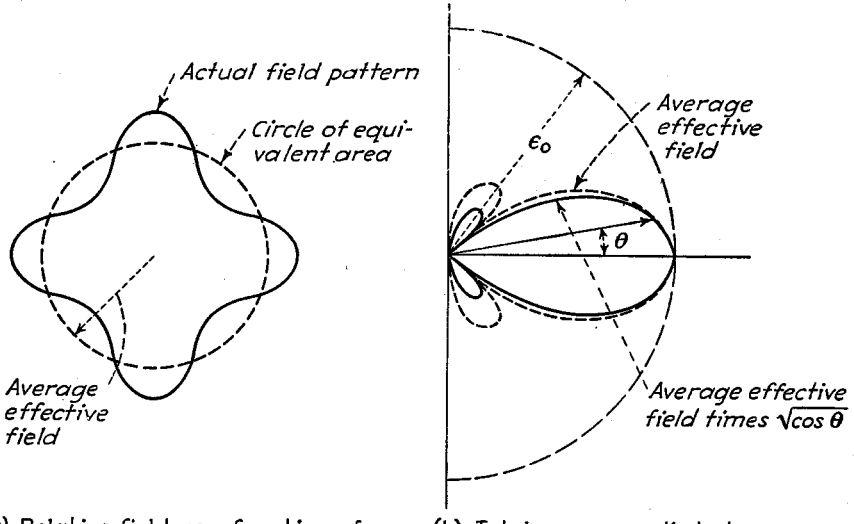
6. Power Relations in Antennas. *Calculation of Radiated Energy.*—There are two principal methods for determining the energy radiated by an antenna system. In the first, known as the *Poynting vector method,* the antenna is assumed to be at the center of a very large sphere, and the resulting fields produced over the spherical surface by the assumed current distribution in the antenna calculated. The power passing through each square meter of such a spherical surface is $0.00265\epsilon^2$ watts, where ϵ is the field strength in rms volts per meter. The total energy radiated from the antenna is then found by a process of summation or integration over the spherical surface. This summation can be carried out in many cases by mathematical or graphical integration.

[1] Thus W. L. Barrow, On the Impedance of a Vertical Half-wave Antenna above an Earth of Finite Conductivity, *Proc. I.R.E.*, Vol. 23, p. 150, February, 1935, shows that the modifications due to imperfection of the ground are negligible for wave lengths above 10 meters except for very dry soil.

[2] In cases where antennas 2 and 3 are in partial resonance, even when open-circuited at the reference points, then they must be further sectionalized or otherwise modified in such a manner as to reduce the induced current in all parts to a negligible value.

[3] Further discussion on the measurement of mutual impedance is given by G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937.

Graphical integration can be readily carried out as follows:[1] *First*, the *average effective* field strength at the spherical surface is calculated as a function of the angle of elevation. This average effective field strength for any given angle of elevation is defined as the strength of the field that, if constant as one rotates about the vertical axis, would represent the same total energy as the actual field distribution for that vertical angle. This average effective field strength can be obtained by determining the radius of the circle that would inclose the same area as does the actual distribution of field strength when plotted in polar form, as shown in Fig. 17*a*. *Second*, the values of average effective field strength as calculated in the first step are next multiplied by $\sqrt{\cos\theta}$, where θ is the angle of elevation, and the resulting product plotted in polar form, as shown in Fig. 17*b*.[2] *Finally*, the area under this curve is determined. This area is proportional to the total radiated power, and can be evaluated in terms of



(a) Relative field as a function of azimuth for a particular vertical angle

(b) Total power radiated as a function of vertical angle

FIG. 17.—Diagram illustrating method of graphically calculating total radiated power.

watts from the fact that the area under the semicircle shown in Fig. 17*b* is equal to $0.0522 e_0^2 r_0^2$ watts, where r_0 is the radius in meters of the spherical surface over which the integration is being made, e_0 is the radius of the half circle plotted in Fig. 17*b*, in volts per meter, and the field strengths used in deriving Fig. 17*b* are in volts per meter.

When the antenna is near the earth, the procedure for evaluating the radiated energy by the Poynting vector method must be modified slightly. In this case, the effect of the earth on the radiated field is taken into account in the usual manner by postulating an image antenna system. The summation or integration is then carried out only for the hemispherical surface above the level of the earth. In the case of graphical integration, as illustrated in Fig. 17*b*, this means that the area to be evaluated

[1] A slightly different graphical procedure is described in the appendix of the paper by H. E. Gihring and G. H. Brown, General Consideration of Tower Antennas for Broadcast Use, *Proc. I.R.E.*, Vol. 23, p. 311, April, 1935.

[2] The multiplication by $\sqrt{\cos\theta}$ takes into account the fact that, on approaching the vertical, the circumference of a ring located on the surface of the sphere with its center on the vertical axis is proportional to $\cos\theta$. The square-root sign is necessary because the area under a curve plotted in polar form is proportional to the square of the radius vector.

is that lying in the first quadrant, whereas in the absence of the earth, the area that is effective is the area in the first and fourth quadrants.

Another method of evaluating the energy radiated from a transmitting antenna consists in assuming an appropriate current distribution and then determining the power that must be supplied to the antenna system to sustain these currents.[1] This power is

$$\text{Power} = \int F_x I_x \, dx \tag{9}$$

where I_x is the current in an elementary length dx of the antenna system, F_x is the component of the electric field parallel to dx at its surface and in phase with I_x, and the integration is carried out over all conductors in the antenna system. The field F_x is calculated in accordance with the methods discussed in Par. 2, and must include the effects of the current in each part of each conductor of the antenna system. When the antenna system is in the vicinity of the earth, the effect of the earth on the field component F_x is taken into account by postulating the usual image antenna system. The integration indicated by Eq. (9) is carried out only over the actual antenna system, however.

The integration indicated by Eq. (9) can be carried out either mathematically or graphically. The graphical procedure is relatively simple and straightforward, and except in the simplest cases is usually to be preferred to the direct mathematical approach because of the relatively complicated formulas that the latter leads to.

A third method[2] of determining the energy radiated by an arbitrary current distribution consists in replacing the actual current distribution by a dipole, quadrupole, octopole, etc., system located at the origin and producing the same field as is actually present. These parts (dipoles, quadrupoles, etc.) are so chosen that the energy radiated by each of them is independent of the presence or absence of other parts, so that one adds the radiated energies from each part without the necessity of taking into account interference effects. This method, although relatively new and not yet fully evaluated in comparison with the first two methods, represents a powerful tool for attacking problems involving power radiation.

Radiation Resistance.—The radiation resistance referred to a certain point in an antenna system is the resistance which, inserted at that point with the assumed current I_0 flowing, would dissipate the same energy as is actually radiated from the antenna system. Thus

$$\text{Radiation resistance} = \frac{\text{radiated power}}{I_0^2} \tag{10}$$

Although this radiation resistance is a purely fictitious quantity, the antenna acts as though such a resistance were present, because the loss of energy by radiation is equivalent to a like amount of energy dissipated in a resistance. It is necessary in defining radiation resistance to refer it to some particular point in the antenna system, since the resistance must be such that the square of the current times radiation resistance will equal radiated power, and the current will be different at different points in the antenna. This point of reference is ordinarily taken as a current loop, although in the case of a vertical antenna with the lower end grounded, the grounded end is often used as a reference point.

[1] The use of this method is developed by A. A. Pistolkors, The Radiation Resistance of Beam Antennas, *Proc. I.R.E.*, Vol. 17, p. 562, March, 1929; R. Bechmann, On the Calculation of Radiation Resistance of Antennas and Antenna Combinations, *Proc. I.R.E.*, Vol. 19, 1471, August, 1931; also, Vol. 21. p. 1367, September, 1933; P. S. Carter, Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932; G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937.

[2] W. W. Hansen and J. G. Beckerley, Concerning New Methods of Calculating Radiation Resistance Fither with or without Ground, *Proc. I.R.E.*, Vol. 24, p. 1594, December, 1936.

In an antenna system composed of several component antennas, the radiation resistance of the combination may be expressed in terms of the radiation resistances of the individual component antennas when isolated from the remaining parts of the antenna system, and the mutual impedances existing between the various parts of the antenna system. In particular, the effective radiation resistance encountered by the voltage E_1 applied to antenna 1 of an antenna array is the resistance component of the driving-point impedance offered to E_1 as calculated by Eq. (6). The total power delivered to the antenna by the voltage E_1 is the square of the current at this point multiplied by this effective radiation resistance. The radiation resistance to the voltages applied to other component antennas in the system and the power delivered by these other voltages can be determined in the same way. The total radiated power is then the sum of the powers supplied by the various applied voltages.

Gain of Directional Antenna Systems.—The term *gain*, as applied to an antenna system, is a measure of the directivity of the antenna field pattern as compared with some standard antenna. Quantitatively, the gain is the ratio of power that must be supplied to the comparison antenna to deliver a particular field strength in the desired direction, to the power that must be supplied to the directional antenna system to obtain the same field strength in the same direction. The gain can be expressed either directly as a power ratio or in terms of the equivalent number of decibels.

The procedure for calculating the gain of an antenna system consists in assuming currents in the antenna to be investigated and in the comparison antenna, such that the field strength produced in the desired direction is the same in both cases. The total energy involved is then determined either by the Poynting vector method or in terms of the radiation resistance.

7. Receiving Antennas.—A radio wave sweeping over a receiving antenna induces voltages that are distributed along the length of the antenna. The value of the induced voltage per meter of conductor length is $\epsilon \cos \phi \cos \theta$, where ϵ is the field strength of the wave in volts per meter, ϕ is the angle between the plane of polarization and the wire in which the voltage is induced, and θ is the angle between the wave front and the direction of the antenna wire. The quantity $\epsilon \cos \phi \cos \theta$ is the component of the field that has a wave front parallel to the antenna and is polarized in the same plane as the antenna.

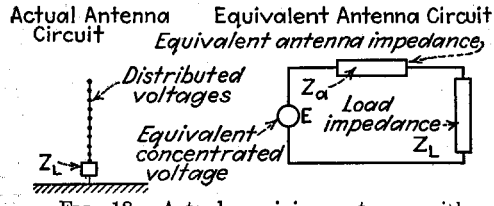


FIG. 18.—Actual receiving antenna with a load impedance Z_L and distributed induced voltages together with equivalent antenna circuit.

The distributed voltages induced in a receiving antenna cause currents to flow, so that if a load impedance is inserted in series with the antenna, energy that has been abstracted from the passing radio wave will be delivered to this load. The antenna, insofar as such a load impedance is concerned, can be considered as a generator of voltage E that has an internal impedance Z_a, as illustrated schematically in Fig. 18. The ratio of the lumped voltage E to the strength of the radio wave is termed the *effective height* of the receiving antenna (*i.e.*, $E = \epsilon h$, where h is the effective height), while the equivalent impedance Z_a is the impedance that the load sees when looking toward the antenna. The current that flows in this equivalent circuit is exactly the same as the current that flows in the antenna at the point where the load is inserted.

The resistance component of the antenna impedance Z_a consists of the radiation resistance R_r of the antenna, plus whatever additional resistance R_l must be postulated

to take into account other losses, such as arise from the presence of earth, the resistance of the antenna wires, dielectrics in the neighborhood, etc.

The total energy that the receiving antenna abstracts from a radio wave is given by the relation

$$\left.\begin{array}{l}\text{Total power in watts}\\ \text{abstracted from}\\ \text{radio wave}\end{array}\right\} = \frac{(\epsilon h)^2}{R_L + R_r + R_l} \tag{11}$$

where ϵ = field strength (rms value) of the radio wave, volts per meter.
h = effective height of the antenna, meters.
R_r = antenna radiation resistance.
R_l = antenna loss resistance.
R_L = load resistance.

The fraction $R_L/(R_r + R_l + R_L)$ of this total energy is delivered to the load impedance; the portion $R_l/(R_r + R_l + R_L)$ represents antenna losses such as occur in the wire and ground resistance, etc.; the remainder, the fraction $R_r/(R_r + R_l + R_L)$, is reradiated. This reradiation of energy results from the fact that when current flows in the antenna, radiation takes place irrespective of whether the voltage producing the current is derived from a passing radio wave or from some other source. The maximum amount of energy that it is theoretically possible for a given antenna to abstract from a passing radio wave occurs when the total loss resistance $R_l + R_L$ equals the radiation resistance, and the load reactance is equal in magnitude but opposite in sign to the reactance component of Z_a. Under these conditions, the abstracted energy is $(\epsilon h)^2/4R_r$ watts.

Calculations show that when the receiving antenna is small compared with a quarter wave length, the maximum amount of energy that can be absorbed from a passing radio wave is roughly the energy contained in a section of the wave front extending about one-fifth of a wave length from the receiving antenna. This maximum energy is independent of the exact size of the antenna, provided that the antenna is small, and is directly proportional to the square of the wave length. When the receiving antenna is not small compared with a quarter wave length, the situation is somewhat more complicated, but, in general, it can be said that the antenna is not capable of abstracting energy passing by at a distance much more than one-quarter of a wave length from the antenna conductors, but will intercept all the energy closer than this.

The electric and magnetic fields in the vicinity of a receiving antenna are the sum of the fields produced by the radio wave and by the current induced in the receiving antenna. The result is that the receiving antenna causes a distortion of the field in its immediate vicinity.[1] Part of this distortion arises from the induction field produced by the antenna current, and part results from the radiation field that combines with the passing radio wave to redistribute the energy of the passing wave in the same manner as a parasitic antenna (see Par. 17).

8. Reciprocal Relations between Receiving and Radiating Properties of Antenna Systems.—The properties of an antenna, when used to abstract energy from a passing radio wave, are similar in nearly all respects to the corresponding properties of the same antenna when acting as a radiator. Thus the relative response of the antenna to waves arriving from different directions is exactly the same as the relative radiation in different directions from the same antenna when excited as a transmitting antenna. Also like the antenna directivity, the effective height, and the impedance of the antenna are the same in reception as in transmission. These reciprocal relations

[1] Thus see Henry C. Forbes, Re-radiation from Tuned Antenna Systems, *Proc. I.R.E.*, Vol. 13, p. 363, June, 1925.

between transmission and reception properties make it possible to deduce the merits of a receiving antenna from transmission tests, and vice versa.

The reciprocal relation between the transmitting and receiving properties of an antenna are incorporated in several theorems, the most important of which was discovered by Rayleigh and extended to include radio communication by John R. Carson.[1] It is to the effect that *if an electromotive force E inserted in antenna* 1 *causes a current I to flow at a certain point in a second antenna* 2, *then the voltage E applied at this point in the second antenna will produce the same current I (both in magnitude and phase) at the point in antenna* 1 *where the voltage E was originally applied.* The Rayleigh-Carson theorem fails to be true only when the propagation of the radio waves is appreciably affected by an ionized medium in the presence of a magnetic field, and so holds for all conditions except short-wave transmission over long distances.

CHARACTERISTICS OF ANTENNA SYSTEMS

9. Doublet Antenna.—The field radiated by a doublet antenna is given by Eq. (1). The radiation resistance of such a doublet is[2]

$$\left.\begin{matrix}\text{Radiation resistance}\\ \text{of doublet antenna}\end{matrix}\right\} = 789\left(\frac{\delta l}{\lambda}\right)^2 \tag{12}$$

where $\delta l/\lambda$ is the doublet length in wave lengths.

When a doublet antenna is employed for reception under conditions where the maximum possible energy is absorbed from the wave (load resistance plus antenna loss resistance equal to radiation resistance), the energy abstracted from the wave corresponds to the energy in a portion of the wave front lying in a circle centered upon the doublet antenna and having a radius of 0.20λ.

10. Resonant Single-wire Antenna.[3] *Radiation Pattern of Single Wire Remote from Earth.*—The field radiated from an isolated wire that is an exact whole number of half wave lengths long is

a. When the wire is an odd number of half wave lengths long

$$\epsilon = \frac{60I}{d}\,\frac{\cos\left(\pi\frac{L}{\lambda}\cos\theta\right)}{\sin\theta} \tag{13a}$$

b. When the wire is an even number of half wave lengths long

$$\epsilon = \frac{60I}{d}\,\frac{\sin\left(\pi\frac{L}{\lambda}\cos\theta\right)}{\sin\theta} \tag{13b}$$

where ϵ = field strength, volts per meter.
d = distance to antenna, meters.

[1] John R. Carson, Reciprocal Theorems in Radio Communication, *Proc. I.R.E.*, Vol. 17, p. 952 June, 1929.

Additional discussion of reciprocity theorems, including some additional relations of more limited scope, are given by John R. Carson, A Generalization of the Reciprocal Theorem, *Bell System Tech. Jour.*, Vol. 3, p. 393, 1924; The Reciprocal Energy Theorem, *Bell System Tech. Jour.*, Vol. 9, p. 325 April, 1930; Stuart Ballantine, Reciprocity in Electromagnetic, Mechanical, Acoustical, and Interconnected Systems, *Proc. I.R.E.*, Vol. 17, p. 929, June, 1929.

[2] See G. W. Pierce, "Electric Oscillations and Electric Waves," p. 434, McGraw-Hill, New York 1920.

[3] The formulas in this section are derived by S. A. Levin and C. J. Young, Field Distribution and Radiation Resistance of a Straight Vertical Unloaded Antenna Radiator at One of Its Harmonics, *Proc. I.R.E.*, Vol. 14, p. 675, October, 1926; P. S. Carter, C. W. Hansell, and N. E. Lindenblad, Development of Directive Transmitting Antennas by R.C.A. Communications, Inc., *Proc. I.R.E.*, Vol. 19, p. 1773, October, 1931.

I = current, amp., at a current loop.
L = length of antenna, meters.
λ = wave length, meters.
θ = angle of elevation measured with respect to wire axis.

The field distribution is a figure of revolution about an axis coinciding with the antenna wire, is symmetrical about a plane perpendicular to the center of the wire, and has the character illustrated in Fig. 19. The relative magnitude and position of the various lobes in the directional pattern can be quickly obtained with the aid of Figs. 19 and 20.

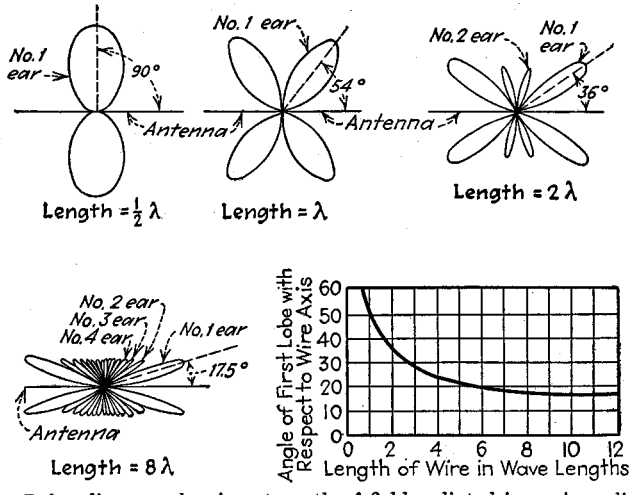


FIG. 19.—Polar diagram showing strength of field radiated in various directions from an antenna consisting of a wire remote from the ground. These diagrams can be considered as cross sections of a figure of revolution in which the axis is the antenna.

Radiation Resistance and Gain.—Radiation resistance of isolated resonant wire is

$$\left.\begin{matrix}\text{Radiation}\\ \text{resistance}\end{matrix}\right\} = 30\left[0.5772 + \log_e\left(4\pi\frac{l}{\lambda}\right) - Ci\left(4\pi\frac{l}{\lambda}\right)\right] = 30S_1\left(\frac{4\pi l}{\lambda}\right) \quad (14)$$

where the resistance is in ohms, l/λ is the antenna length in wave lengths, and $Ci(x)$ is

$$Ci(x) = \int_{\infty}^{x}\frac{\cos x}{x}\,dx \quad (15)$$

Values of $Ci(x)$ can be obtained from the tabulated values of $S_1(x)$ given in Table 13, Sec. 1. When $\frac{l}{\lambda} > 1$, the radiation resistance is approximately

$$\left.\begin{matrix}\text{Radiation}\\ \text{resistance}\end{matrix}\right\} = 17.32 + 30\log_e\left(4\pi\frac{l}{\lambda}\right) \quad (16)$$

The relation between radiation resistance and length is given in Fig. 21, which also gives the gain of an isolated long-wire antenna as compared with an antenna a half wave length long. The power gain of the latter as compared with a doublet is 1.09.

Effect of a Perfect Ground on Characteristics of a Resonant Wire.—The effect of a perfect earth on the directional characteristics of a resonant wire antenna is determined by the method of images discussed in Par. 4. For horizontal antennas, and also

for vertical antennas that are an even number of half wave lengths long, the field in the presence of a perfect ground is

$$\left.\begin{matrix}\text{Actual radiation in}\\ \text{presence of ground}\end{matrix}\right\} = 2\sin\left(2\pi\frac{H}{\lambda}\sin\theta\right)\left\{\begin{matrix}\text{radiation from}\\ \text{antenna when}\\ \text{in free space}\end{matrix}\right. \tag{17}$$

In the case of vertical antennas that are an odd number of half wave lengths long, the field in the presence of a perfect ground is

$$\left.\begin{matrix}\text{Actual radiation in}\\ \text{presence of ground}\end{matrix}\right\} = 2\cos\left(2\pi\frac{H}{\lambda}\sin\theta\right)\left\{\begin{matrix}\text{radiation from}\\ \text{antenna when}\\ \text{in free space}\end{matrix}\right. \tag{18}$$

In these equations, H/λ is the height of the antenna center above the ground, measured in wave lengths, and θ is the angle of elevation above the horizontal. The nature of the factor in Eq. (17) is indicated in Figs. 22 and 23. It is seen that the ground reflection causes cancellation along the ground and at certain vertical angles. Where Eq. (18) applies, the nulls occur when Figs. 22 and 23 show maxima, and vice versa.

The radiation resistance of a resonant-wire antenna is modified by the presence of the earth. This effect can be most conveniently taken into account by adding to the radiation resistance of the wire as given by Eq. (14) the resistance component of the mutual impedance between the antenna and its image. Thus, in the case of a horizontal wire, where the image has currents flowing in the opposite direction from those in the actual antenna, the effective resistance of the latter is its radiation resistance calculated on the basis of an isolated wire, minus the resistance component of the mutual impedance as given by Figs. 13 or 15. Hence, for a wire four wave lengths long and parallel to the earth at a height of three-quarters wave length, the radiation resistance is $132 - (-15) = 147$ ohms, where -15 ohms is the resistance component of mutual impedance for a spacing of 1.5λ ($= 2 \times 0.75\lambda$). In the case of vertical wires, the resistance component of the mutual impedance of the image is calculated on the basis of colinear conductors, and is added to the radiation resistance of the isolated wire when the wire is an odd number of half wave lengths long, and is subtracted when the wire is an even number of half wave lengths long.

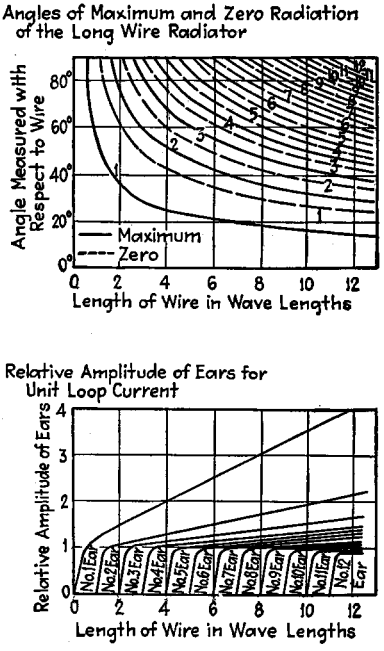


FIG. 20.—Angles at which the radiation from an isolated wire antenna is zero and maximum, together with relative field strength of the ears. The ears are numbered so that ear 1 is the lobe adjacent to the antenna.

Examples showing the effect on radiation resistance of antenna height above earth in typical cases are presented in Fig. 24.

Losses in the earth do not affect the radiation resistance appreciably. The effect of an imperfect earth on the field distribution is taken into account by modifying the

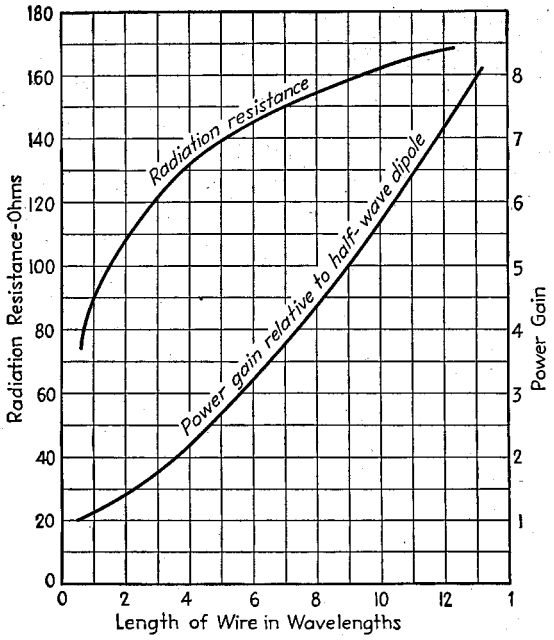


FIG. 21.—Radiation resistance and power gain of an isolated resonant antenna.

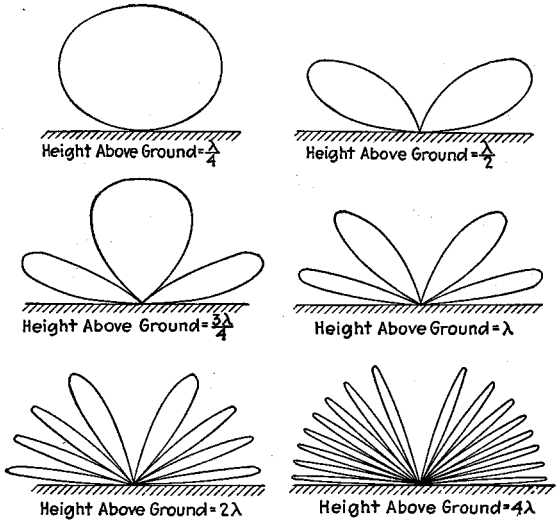


FIG. 22.—Polar diagram of the factor $|2 \sin (2\pi \frac{H}{\lambda} \sin \theta)|$ for various values of H/λ, showing how the height above earth affects the directional characteristics of the antenna.

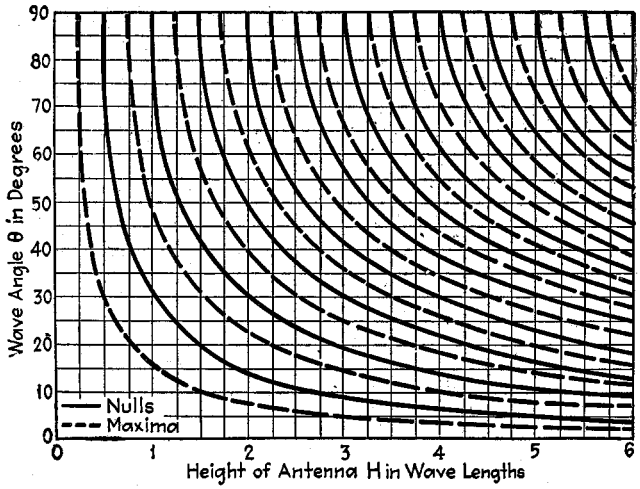


FIG. 23.—Chart showing the vertical angles at which the factor $[2 \sin(2\pi \frac{H}{\lambda} \sin \theta)$ is maximum and zero. These represent the vertical angles at which the reflections from the ground cause complete reinforcements and complete cancellation, respectively, of the radiation.

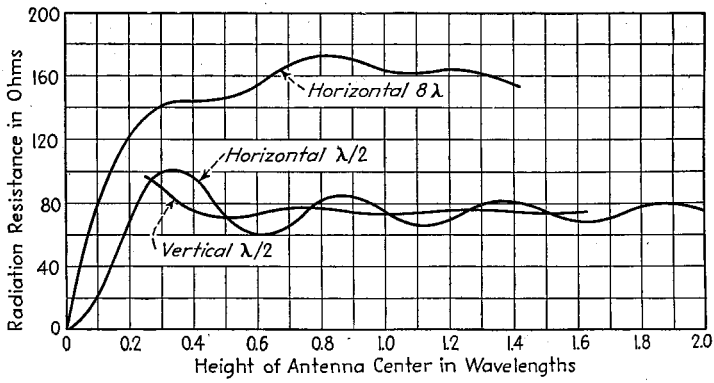


FIG. 24.—Effect of antenna height on radiation resistance in typical cases.

magnitude and phase of the current image in accordance with the reflection coefficient of the earth, as described in Par. 4. Under these conditions

$$\left.\begin{array}{l}\text{Actual radiation}\\ \text{in presence of}\\ \text{ground}\end{array}\right\} = \left[1 + A^2 \pm 2A \cos\left(\phi + 4\pi \frac{H}{\lambda} \sin\theta\right)\right]^{\frac{1}{2}} \left\{\begin{array}{l}\text{radiation}\\ \text{from}\\ \text{antenna}\\ \text{when in}\\ \text{free space}\end{array}\right. \quad (19)$$

Here A is the magnitude of the reflection coefficient, and ϕ is the phase angle of the reflection coefficient as defined in Par. 5, Sec. 10. The negative sign is used for the case corresponding to Eq. (17) when there is a perfect earth, and the positive sign

corresponds to the case of Eq. (18); the remaining notation is the same as previously employed.

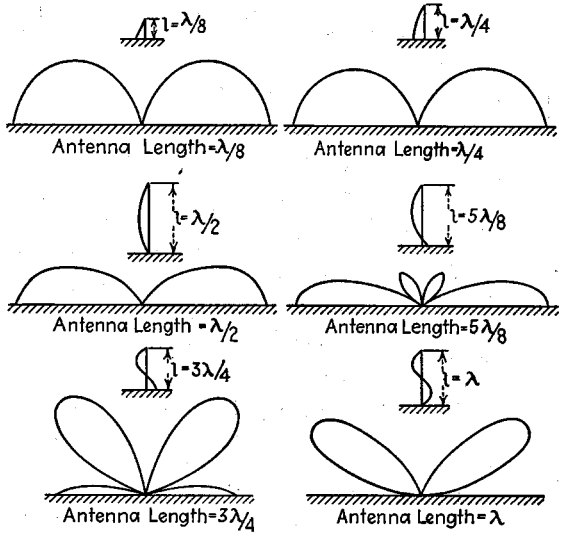


FIG. 25.—Directional characteristics in a vertical plane of field produced by grounded vertical antennas of varying lengths. These polar diagrams can be thought of as cross sections of figures of revolution about the antenna, and assume a perfectly conducting earth.

11. Grounded Vertical Antennas. *Grounded Vertical Wire.*[1]—The field pattern obtained with a grounded vertical wire depends upon the height (length) H/λ of the wire in wave lengths according to the equation

$$\epsilon = \frac{60I}{d}\left[\frac{\cos 2\pi\frac{H}{\lambda} - \cos\left(2\pi\frac{H}{\lambda}\sin\theta\right)}{\cos\theta}\right] \tag{20}$$

where ϵ = field strength, volts per meter.
I = current, amp., at current loop.[2]
d = distance, meters.
H/λ = height (length) in wave lengths.
θ = angle of elevation with respect to the earth.
This equation assumes a perfect earth.

[1] Stuart Ballantine, On the Radiation Resistance of a Simple Vertical Antenna at Wave Lengths below the Fundamental, *Proc. I.R.E.*, Vol. 12, p. 823, December, 1924; E. B. Moullin, The Radiation Resistance of Aerials Whose Length Is Comparable with the Wavelength, *Jour. I.E.E.*, Vol. 78, p. 540, 1936; also, Wireless Section, *I.E.E.*, Vol. 11, p. 93, June, 1936; N. Wells, Aerial Characteristics, *Jour. I.E.E.*, Vol. 89, Part 3, p. 76, June, 1942.

This assumes a sinusoidal distribution of current. Analysis of cases where the current distribution has been modified is discussed by H. E. Gihring and G. H. Brown, General Consideration of Tower Antennas for Broadcast Use, *Proc. I.R.E.*, Vol. 23, p. 311, April, 1935; W. L. McPherson, Electrical Properties of Aerials for Medium and Long Wave Broadcasting, *Elec. Comm.*, Vol. 17, p. 44, July, 1938.

[2] The loop current in a vertical wire is related to the base current by

$$\text{Loop current} = \frac{\text{base current}}{\sin\left(2\pi\frac{H}{\lambda}\right)}. \tag{21}$$

This loop current will not be present anywhere in the actual antenna when $H < \lambda/4$, but it can be calculated, and fields and resistance can then be referred to it, even if it is only hypothetical.

When the length of the grounded antenna is of the order of one-eighth wave length or less, the radiation is almost exactly proportional to the cosine of the angle of elevation. As the height is increased the radiation is concentrated increasingly along the horizontal until the height exceeds a half wave length, beyond which high-angle lobes of increasing amplitude appear until at a height of one wave length none of the energy is radiated along the horizontal. This behavior is illustrated in Fig. 25.

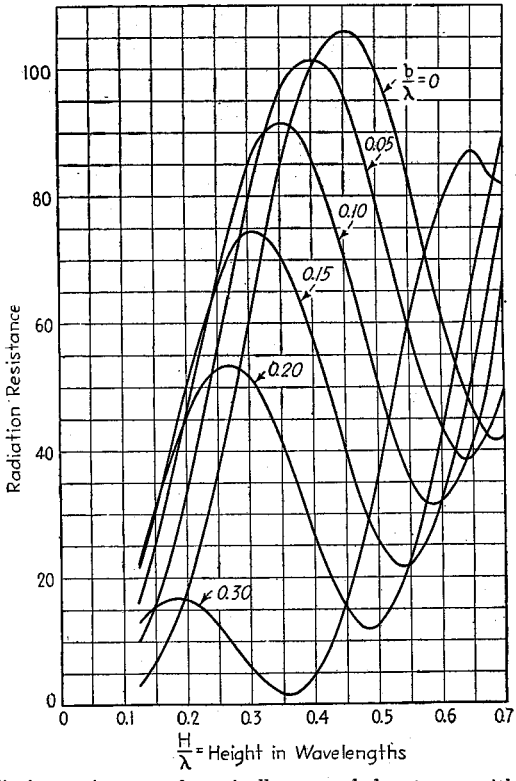


FIG. 26.—Radiation resistance of vertically grounded antenna with varying amounts of capacity top loading (see Fig. 27 for notation). The resistances shown are referred to a current loop. When the resistance is referred to the base current, it is $1/\sin^2\left(2\pi\frac{H+b}{\lambda}\right)$ times as great as the value shown above.

The radiation resistance (referred to a current loop) of a grounded vertical antenna, with sinusoidal current distribution and a perfect earth assumed, is

$$\left.\begin{matrix}\text{Radiation}\\ \text{resistance}\\ \text{in ohms}\end{matrix}\right\} = 60\left\{S_1\left(4\pi\frac{H}{\lambda}\right)\cos^2\left(2\pi\frac{H}{\lambda}\right) - \frac{1}{4}S_1\left(8\pi\frac{H}{\lambda}\right)\cos\left(4\pi\frac{H}{\lambda}\right) - \frac{1}{2}\sin\left(4\pi\frac{H}{\lambda}\right)\left[Si\left(4\pi\frac{H}{\lambda}\right) - \frac{1}{2}Si\left(8\pi\frac{H}{\lambda}\right)\right]\right\} \quad (22)$$

where $Si(x) = \int_0^x \frac{\sin x}{x}\,dx$ (23)

Values of the sine integral $Si(x)$ and of $S_1(x)$ are tabulated in Tables 12 and 13,

respectively, of Sec. 1. The radiation resistance given by Eq. (21) is referred to the loop current. The radiation resistance referred to the base current is the loop resistance divided by $\sin^2\left(2\pi\frac{H}{\lambda}\right)$, the square of the ratio of base to loop current.

Results calculated with the aid of Eq. (22) are plotted in Fig. 26 (curve for $b/\lambda = 0$).

Vertical Grounded Antennas with Capacity Tops—Sectionalized Antennas.—Vertical grounded antennas are frequently provided with a capacity top. This may be in the form of a ring or spider, an L (see Fig. 3*e*), or a T (see Fig. 3*f*). Occasionally the capacity top is connected to the upper end of the vertical portion of the antenna through an inductance. This gives the same effect as increasing the capacity of the top. The current in the vertical portion of an antenna provided with a capacity top is distributed in exactly the same way as would be the case if the capacity top were removed and the antenna height increased somewhat, as illustrated in Fig. 27.

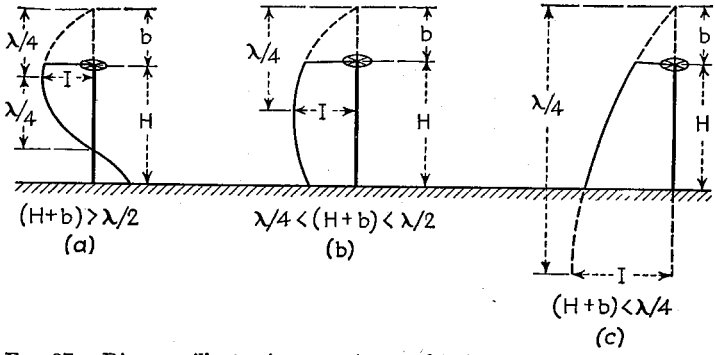


FIG. 27.—Diagram illustrating notation used in Eq. (24) and Fig. 26.

In calculating the behavior of an antenna with a capacity top, it is customary to neglect the radiated field produced by the horizontal currents flowing in the top capacity. This is permissible because this field is of negligible amplitude compared with the radiated field produced by the currents flowing in the vertical portion of the antenna.

The variation of field with vertical angle for an antenna with a capacity top is[1]

$$\epsilon = \frac{60I}{d}\left[\frac{\cos 2\pi\frac{b}{\lambda}\cos\left(2\pi\frac{H}{\lambda}\sin\theta\right) - \sin\theta\sin 2\pi\frac{b}{\lambda}\sin\left(2\pi\frac{H}{\lambda}\sin\theta\right)}{\cos\theta} - \frac{\cos 2\pi\left(\frac{H+b}{\lambda}\right)}{\cos\theta}\right] \quad (24)$$

where ϵ = field strength, volts per meter.

d = distance, meters.

I = current at current loop $= \dfrac{\text{(base current)}}{\sin\left(2\pi\dfrac{H+b}{\lambda}\right)}$.

θ = angle of elevation above the earth.

H/λ = height of vertical portion in wave lengths (see Fig. 27).

b/λ = vertical length that added to antenna would cause same current in H that is present with flat top (see Fig. 27).

[1] G. H. Brown, A Critical Study of the Characteristics of Broadcast Antennas as Affected by Antenna Current Distribution, *Proc. I.R.E.*, Vol. 24, p. 48, January, 1936.

When H/λ is small and $b >> H$, then the current in the vertical part is substantially constant, and Eq. (24) reduces to

$$\epsilon = \frac{120\pi}{d}\frac{H}{\lambda} I_0 \cos\theta \tag{25}$$

where I_0 is the current in the vertical portion.

Top loading has the same effect on the field distribution in a vertical plane as a greater height. Thus an antenna for which $H = 0.45\lambda$ can by suitable top loading be made to have a field distribution in the vertical plane that is substantially the same as for a vertical wire of $H = 0.6\lambda$.

The radiation resistance of a top-loaded antenna referred to the current at a current loop is given in Fig. 26.[1] In the special case of antennas where the total equivalent electrical length $H + b$ does not exceed a tenth of a wave length, as is the case with long-wave antennas, the radiation resistance referred to the base current, is[2]

$$\left.\begin{matrix}\text{Radiation resistance}\\ \text{in ohms}\end{matrix}\right\} = 1{,}578\left(\frac{H}{\lambda}\right)^2\left[1 - \frac{H}{H+b} + \frac{1}{4}\left(\frac{H}{H+b}\right)^2\right] \tag{26}$$

The notation is explained in Fig. 27.

A slightly different form of top loading is shown in Fig. 28, and involves the insertion of an inductance coil in a vertical grounded antenna a little way down from the top. This causes the current distribution to be modified as illustrated, the currents flowing in the portion of the antenna below the coil being the same as those that would be present if the coil was removed and the actual height increased somewhat, as indicated by the dotted line in Fig. 28.[3] This method of top loading gives results equivalent to those obtained with a capacity top.

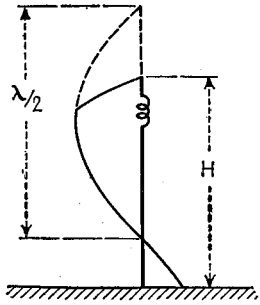


FIG. 28.—Current distribution in vertical antenna having an inductance inserted near the top.

12. Antennas Terminated so that the Current Distribution is Nonresonant.—The current distribution in this case is

$$I = I_0\epsilon^{-\alpha L} \tag{27}$$

where L = distance from end of line to which power is supplied.

I = current at distance L from end of line to which power is supplied.

I_0 = current at $L = 0$.

α = equivalent attenuation constant of line.

The phase of the current I lags $2\pi L/\lambda$ radians behind the phase of I_0.

The radiated field obtained with the current distribution of Eq. (27) is[4]

$$\epsilon = \frac{30 I_0 \sin\theta}{d(1-\cos\theta)\sqrt{1+\left(\frac{\alpha}{p}\right)^2}}\left(1 + \epsilon^{-2\alpha L} - 2\epsilon^{-\alpha L}\cos pL\right)^{1/2} \tag{28}$$

[1] These curves are based on data given by W. L. McPherson, Electrical Properties of Aerials for Medium and Long Wave Broadcasting, *Elec. Comm.*, Vol. 16, p. 306, April, 1938.

[2] Fulton Cutting, A Simple Method of Calculating Radiation Resistance, *Proc. I.R.E.*, Vol. 10, p. 129, April, 1922. This paper also gives formulas for the radiation resistance of other antenna configurations operated at frequencies much lower than that giving quarter-wave resonance.

[3] Further discussion of this case, with formulas for field patterns, radiation resistance, and relative fields radiated along the horizontal, is given by G. H. Brown, A Critical Study of the Characteristics of Broadcast Antennas as Affected by Antenna Current Distribution, *Proc. I.R.E.*, Vol. 24, p. 48, January, 1936; A Note on the Placement of the Coil in a Sectionalized Antenna, *Broadcast News*, Vol. 22, p. 14, October, 1936.

[4] See Terman, *op. cit.*, p. 662.

where ϵ = field strength, volts per meter.
d = distance, meters.
θ = angle measured with respect to wire axis.
$p = \frac{2\pi}{\lambda}(1 - \cos\theta)$.

The remaining notation is as in Eq. (27).

In the common case where the amplitude of the current can be assumed constant along the wire, *i.e.*, zero losses corresponding to $\alpha = 0$, then Eq. (28) becomes[1]

$$\epsilon = \frac{60I_0}{d} \cot\frac{\theta}{2} \sin\frac{pL}{2} \tag{29}$$

Typical field patterns calculated from Eqs. (28) and (29) are shown in Fig. 29. These patterns exhibit a strong unidirectional tendency with the angle between the

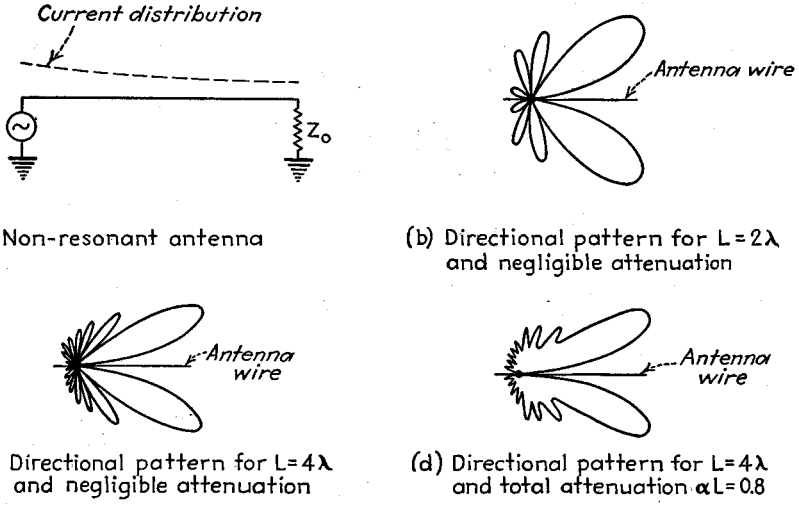


FIG. 29.—Typical directional characteristics for a nonresonant wire in space. These patterns may be considered as cross sections of a figure of revolution in which the wire is the axis.

major lobe and the wire being less the greater the wire length measured in wave lengths. This angle for the assumption of zero losses in the transmission line is $90° - \phi$, where ϕ is given by the solid curve of Fig. 40.

Examination of Fig. 29 shows that even with losses that represent as much as or more attenuation than would be encountered in practice ($\alpha L = 0.08$), the field pattern still has the same general character as with zero losses. The chief difference is that the losses cause the minima to be filled in.

The radiation resistance of a terminated antenna on the assumption of uniform current distribution (negligible attenuation) is given in Fig. 30.[2] This resistance is defined as the quantity that, when multiplied by the square of the current I_0 in the

[1] This relation is also given by Andrew Alford, A Discussion of Methods Employed in Calculations of Electromagnetic Fields of Radiating Conductors, *Elec. Comm.*, Vol. 15, p. 70, July, 1936; F. M. Colebrook, The Electric and Magnetic Fields of a Linear Radiator Carrying a Progressive Wave, *Jour. I.E.E.*, Vol. 86, p. 169, 1940; also, Wireless Section, I.E.E., Vol. 15, p. 17, March, 1940.

[2] From Alford, *loc. cit.* See also F. M. Colebrook for details regarding radiation resistance calculations.

wire, will give the total radiated power. The radiation resistance as given in Fig. 30 involves an approximation, since if there is radiation, there necessarily must be some attenuation. The resulting error is small when the radiation resistance is small in comparison with the characteristic impedance of the terminated wire considered as a one-wire transmission line. However, when the value of radiation resistance differs only slightly from the characteristic impedance, then the results obtained from Fig. 30 must have a correction applied to them that may be fairly large.

13. Antenna Arrays.[1]—An antenna array is an arrangement of antennas so spaced and phased that the fields radiated from the individual antennas composing the array tend to add in some preferred direction, and to cancel in other directions. It is possible in this way to obtain very great directivity and consequent large antenna gain.

Field Pattern for a Rectangular Array of Similar Radiators.—The most important general case of an antenna array consists of a number of identical radiators equally spaced along the x, y, and z axes, carrying equal currents, with the phase difference between adjacent radiators along any one axis constant. Such an arrangement can

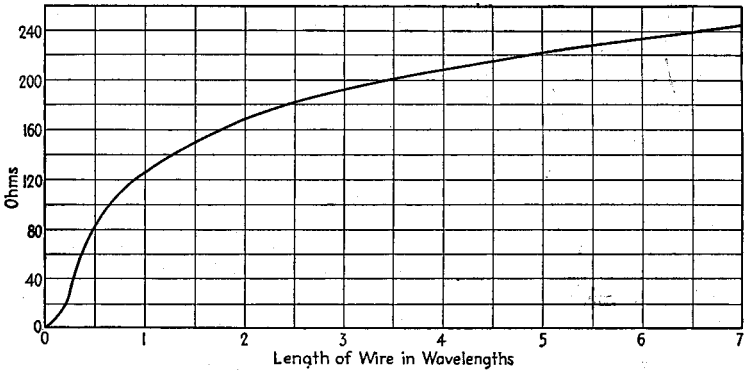


FIG. 30.—Radiation resistance of nonresonant antenna having low attenuation and remote from earth.

be termed a rectangular array, and may be regarded as consisting of $\mathfrak{N}$ parallel planes, each made up of N parallel columns, where each column is made up of n individual radiating elements.

The arrangement of radiators in the form of a rectangular array gives a resultant field pattern that is a modification of the field pattern of the individual array. The field obtained is[2]

$$\text{Relative field strength} = R\left|\frac{\sin n\pi(a\cos\phi\cos\theta + b)}{n\sin\pi(a\cos\phi\cos\theta + b)}\,\frac{\sin N\pi(A\sin\phi\cos\theta + B)}{N\sin\pi(A\sin\phi\cos\theta + B)}\,\frac{(\sin \mathfrak{N}\pi(\mathfrak{A}\sin\theta + \mathfrak{B})}{\mathfrak{N}\sin\pi(\mathfrak{A}\sin\theta + \mathfrak{B})}\right| \tag{30}$$

where n, N, and $\mathfrak{N}$ = number of radiators along x, y, and z axes, respectively.

a, A, and $\mathfrak{A}$ = spacing of adjacent radiators along x, y, and z axes, respectively, measured in fractions of a wave length.

b, B, and $\mathfrak{B}$ = phase displacement between adjacent radiators along x, y, and z axes, respectively, measured in fractions of a cycle.

[1] G. C. Southworth, Certain Factors Affecting the Gain of Directive Antennas, *Proc. I.R.E.*, Vol. 18, p. 1502, September 1930; E. J. Sterba, Theoretical and Practical Aspects of Directional Transmitting Systems, *Proc. I.R.E.*, Vol. 19, p. 1184, July, 1931.

[2] See Southworth, *loc. cit.*

θ = angle with respect to xy-plane (angle of elevation).
ϕ = angle with respect to xz plane (bearing angle).
R = directional pattern of individual radiators composing the array.

The field distribution is determined by four factors. The first is the radiation pattern of the individual radiator; the second, third, and fourth take into account the distribution and phasing of the radiators along the x, y, and z axes, respectively.

Broadside Arrays.—A broadside array consists of a number of identical radiators carrying currents of the same phase and spaced at uniform intervals along a line. It is

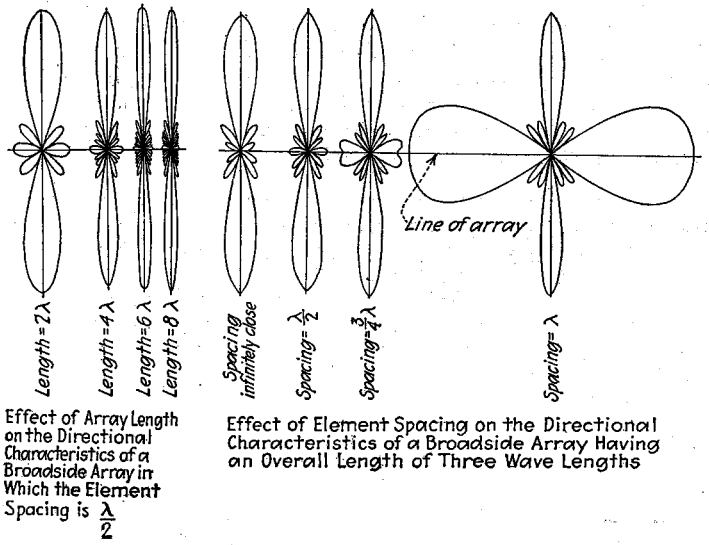


FIG. 31.—Effect of array length and element spacing on directional characteristics of the field radiated from a broadside array under conditions where the individual radiator has a circular pattern in the plane under consideration.

a special case of the general rectangular array of Eq. (30) in which $N = 1$, $\mathfrak{N} = 1$, $b = 0$. The field distribution of a broadside array is given by the equation

$$\left.\begin{matrix}\text{Relative field} \\ \text{strength}\end{matrix}\right\} = R\left|\frac{\sin\left[n\pi(a\cos\phi\cos\theta)\right]}{n\sin\left[\pi(a\cos\phi\cos\theta)\right]}\right| \tag{31}$$

where n = number of radiators in the array (assumed spaced along x axis).
a = spacing of adjacent radiators in wave lengths along x axis.
R = field distribution of individual radiator when isolated.
θ = angle of elevation with respect to xy plane.
ϕ = bearing angle measured with respect to a plane containing the line of the array (xz plane).

The principal properties of a broadside array are illustrated in Figs. 31 to 33. The sharpness of the directional pattern in a plane containing the line of the array increases with the length of the array and is substantially independent of the spacing of the radiators until a critical spacing is exceeded. This critical spacing depends upon the directivity of the individual radiators, and is approximately $3\lambda/4$ when the individual antennas radiate uniformly in all directions in a plane containing the line of the array. The minimum permissible spacing becomes greater, however, as the individual antennas possess increasing directivity in a plane containing the line of the array. Thus,

when the individual radiators are half-wave antennas parallel to the line of the array, as in Fig. 32, the critical spacing is something greater than a wave length, and when

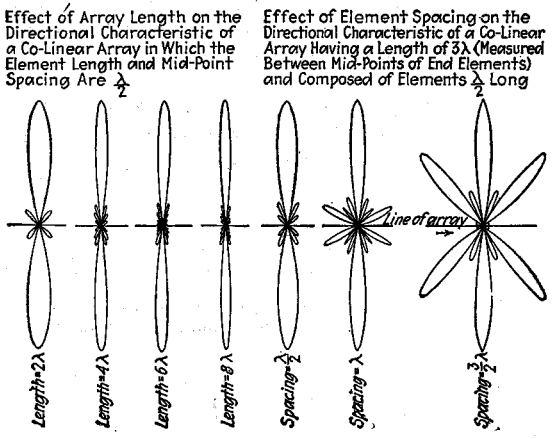


FIG. 32.—Effect of array length and element spacing on directional characteristics of the field radiated from a broadside array under conditions where the individual radiator is a half-wave antenna having its axis parallel with the line of the array (collinear arrangement).

the individual radiators have still greater directivity, as is the case with the V antennas of Fig. 43*b*, still larger spacings between radiators are permissible without loss of the broadside concentration. These effects of radiator spacing come about because increased spacing tends to result in large parasitic lobes of radiation in directions other than at right angles to the line of the array. However, if the directivity of the individual radiators is such that little if any energy is radiated in a side direction corresponding to such a parasitic lobe, as in Fig. 32 for a spacing of one wave length, then these potentially large parasitic lobes will still be small.

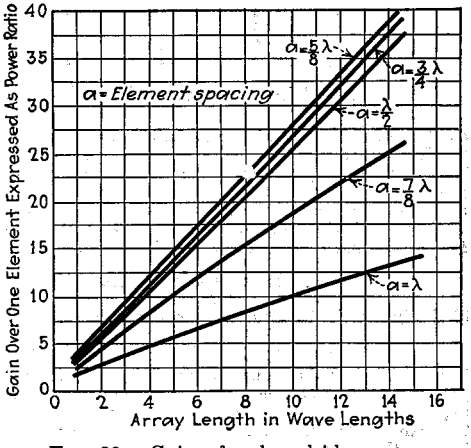


FIG. 33.—Gain of a broadside array as a function of array length for various element spacings. These curves are for the case corresponding to Fig. 31, with the radiation from the individual antenna proportional to the cosine of the angle of elevation, and give the gain over a single element of the array.

The effect of the ground on the field pattern of a broadside array in which the individual antennas are vertically polarized is obtained by multiplying the field pattern as calculated by Eq. (31) by the factor given by Eq. (18), where the height H is the height to the center of the broadside array. When the individual radiators of the broadside array are horizontally polarized, then the field pattern as given by Eq. (31) for an isolated system must be multiplied by the factor calculated from Eq. (17) to take into account the effect that a perfect ground has on the direction pattern of the array.

The power gain of a broadside array is almost exactly proportional to the length of the array, provided that this length exceeds one or two wave lengths. The power gain is also substantially independent of the spacing of the individual radiators, provided that this spacing does not exceed the critical spacing beyond which the directional pattern begins to exhibit large parasitic lobes in other than the broadside direction. This is illustrated in Fig. 33.

The power gain of a broadside array can be readily calculated when tables or curves are available giving the resistance component of the mutual impedances between parallel antennas of the type employed in the array. This is illustrated by the following example:

Example.—Calculate the power gain of the broadside array of Fig. 34, assuming that the earth is a perfect reflector (1) when the power gain is referred to an isolated antenna similar to that composing each element of the array and (2) when the comparison antenna is in the same position with respect to earth as the antenna elements.

If the individual antennas of the array of Fig. 34 are assumed to carry the same current as the comparison antenna, the field radiated along the horizontal at right angles to the line of the array will be

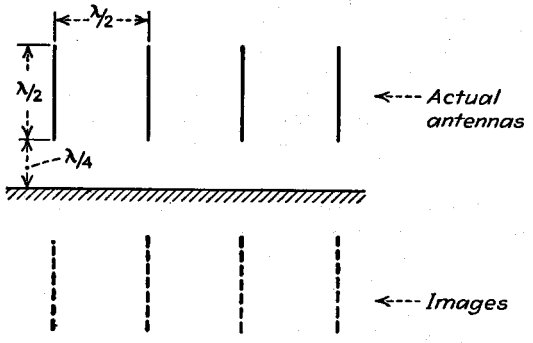


FIG. 34.—Diagram of simple broadside array illustrating notation.

eight times as strong as the field from the single isolated comparison antenna, corresponding to a power 64 times as great, and 4 times the field or 16 times the power when the comparison antenna is not isolated but is near the earth. The amount of power required to produce this field in the array is proportional to the summation of the radiation resistances of the individual wires of the array. The total resistance of any wire of the array is the radiation resistance of this wire when isolated plus the resistance component of the mutual impedances between the wire and every other wire of the array and its image. For the two end wires, with the aid of Table 1, this total resistance is

$$73.3 - 12.4 + 4.1 - 1.8 - 4.1 - 0.8 + 3.6 - 6.3 = 55.6 \text{ ohms.}$$

Similarly, for the two middle wires this resistance totals

$$73.3 - 2(12.4) + 4.1 - 4.1 - 2(0.8) + 3.6 = 50.5 \text{ ohms.}$$

The total power that must be supplied to maintain the assumed current in the array is accordingly proportional to $2(55.6 + 50.5) = 212.2$ ohms. This is $212.2/73.3 = 2.90$ times as great as the power required by the isolated comparison antenna, and produces a field representing 64 times as much power. The gain of the array is hence $64/2.90 = 22.1$ times, as compared with a half-wave radiator remote from earth. The gain with respect to a half-wave vertical radiator with the lower end $\lambda/4$ above earth is similarly $16/[212.2/(73.3 - 4.1)] = 5.22$ times, where $73.3 - 4.1 = 69.2$ ohms is the radiation resistance of the comparison antenna taking into account its image.

In cases where the necessary tables giving the resistance component of the mutual impedance are not available, graphical methods are generally preferred.[1]

End-fire Arrays.—The end-fire array consists of a number of identical antennas arranged along a line, carrying equal currents excited so that there is a progressive phase difference between adjacent antennas equal in cycles to the spacing between

[1] Formulas giving the gain of rectangular arrays in which the elementary radiators are doublet antennas are given by Southworth, *loc. cit.* and Sterba, *loc. cit.*

these antennas in wave lengths.[1] Thus if the adjacent antennas are a quarter of a wave length apart, a phase difference of 90° between adjacent antennas is called for.

The end-fire array can be considered a special case of the rectangular array in which $N = 1$, $\mathfrak{N} = 1$, and $a = |b|$. The field pattern that results is given by the equation

$$\left.\begin{matrix}\text{Relative field} \\ \text{strength}\end{matrix}\right\} = R\left|\frac{\sin\left[n\pi a(\cos\phi\cos\theta \pm 1)\right]}{n\sin\left[\pi a(\cos\phi\cos\theta \pm 1)\right]}\right| \tag{32}$$

The notation is the same as in Eq. (31), with the + sign being used if b is positive and the − sign if b is negative.

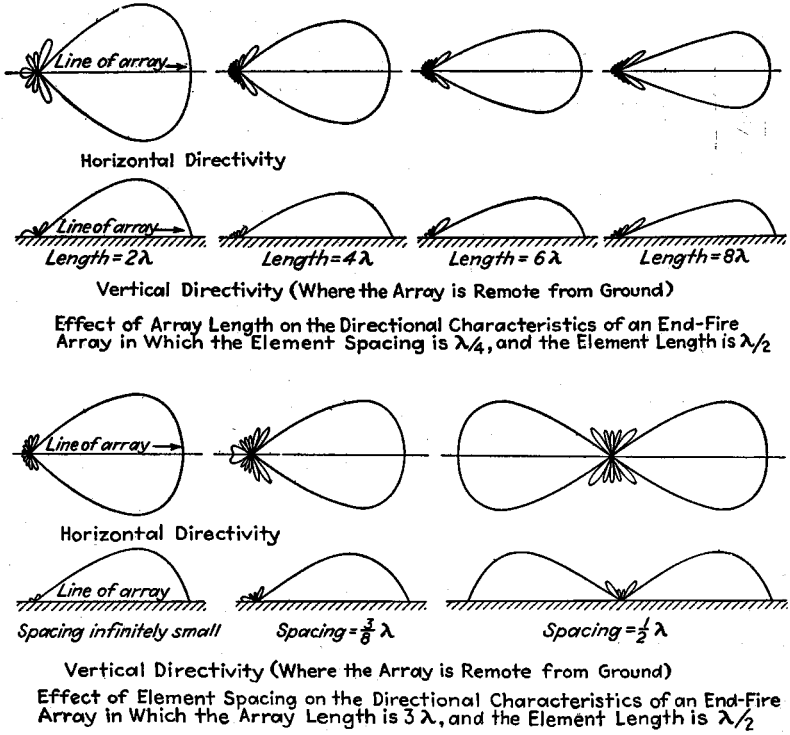


FIG. 35.—Effect of array length and element spacing on the directional characteristics of the field radiated from an end-fire array. The elementary antennas of the array are oriented at right angles to the line of the array and are vertical.

The field pattern of a typical end-fire array under various conditions is illustrated in Fig. 35, and is characterized by concentration of the resultant radiated field in the direction of the end of the array having the most lagging phase. This unidirectional nature of the field pattern makes the end-fire array of considerable importance.

The end-fire arrangement concentrates the radiation in both the vertical and horizontal planes. This concentration is proportional to the length of the array, and is substantially independent of the spacing of the elementary radiators, provided that

[1] It has been found possible to increase the directivity and power gain of an end-fire array by making the phase difference between adjacent radiators about π/n radians greater than in the usual end-fire arrangement, where n is the number of radiators. This increases the power gain by a factor of about 1.8 if the array is at least two wave lengths long. See W. W. Hansen and J. R. Woodyard, A New Principle in Directional Antenna Design, *Proc. I.R.E.*, Vol. 26, p. 333, March, 1938.

this spacing does not exceed a critical value. This critical spacing depends upon the directivity of the individual radiators and has a value about $3\lambda/8$ when the field pattern of the individual radiators of which the array is composed is circular in a plane containing the line of the array. Greater spacing is permissible, however, if the individual radiators have directivity that is of the same general character as that produced by end-fire action.

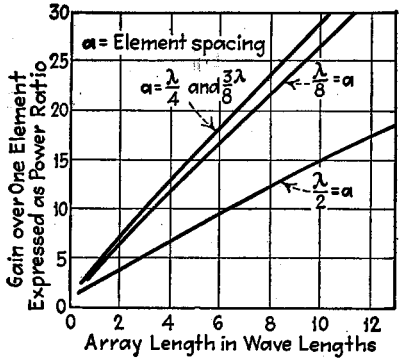
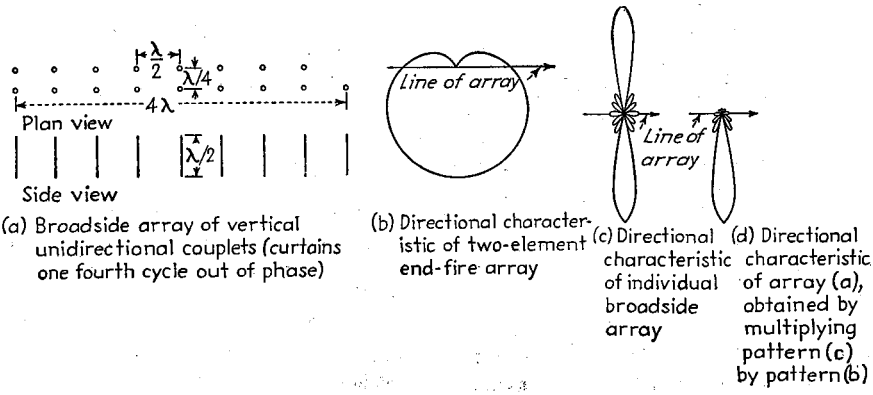


FIG. 36.—Gain of end-fire array as a function of array length for various element spacings. These curves are calculated on the assumption that the individual antennas are vertical doublets and that the array length is in a horizontal direction.

The power gain of an end-fire array expressed with reference to an individual antenna such as those that make up the array, is proportional to the length of the array, and nearly of independent element spacing up to critical spacing, as illustrated in Fig. 36. The gain of an end-fire array having a given length is almost exactly the same as the gain of a broadside array of the same length, if it is assumed that both arrays have the same type of individual radiators oriented in the same direction with respect to the line of the array.

Array of Arrays.—Many antennas used in practice can be considered as an array, each element of which is in itself an array. The directional characteristics of such an array of arrays can be calculated with the aid of Eq. (30) by taking R as the directional pattern of the individual elementary array. An example of such an array of arrays is shown in Fig. 37, which can be considered as two broadside arrays arranged in end-fire fashion, with a spacing of a quarter of a wave length. The directional characteristics of the individual broadside array is given by Fig. 37*c*, corresponding to R in Eq.



(a) Broadside array of vertical unidirectional couplets (curtains one fourth cycle out of phase)

(b) Directional characteristic of two-element end-fire array

(c) Directional characteristic of individual broadside array

(d) Directional characteristic of array (a), obtained by multiplying pattern (c) by pattern (b)

FIG. 37.—An array of arrays in which the elementary array is a broadside arrangement having the characteristics shown at (*c*), with two such broadside arrays arranged in end-fire fashion to form an array of arrays.

(30). This must be multiplied by the directional pattern shown in Fig. 37*b* of a two element end-fire array spaced a quarter of a wave length. The resultant pattern in the horizontal plane is shown in Fig. 37*d*, and is seen to possess the sharpness in the horizontal plane of a broadside array, combined with the unidirectional characteristics of an end-fire array.

Arrays of arrays are used to increase the directivity obtainable from simple arrays. In general, arranging arrays side by side in broadside fashion sharpens the directional pattern in the horizontal plane, while stacking them one above the other in tiers increases the directivity in the vertical plane by giving a broadside action in the vertical direction. The use of a second curtain of arrays, located an odd number of quarter wave lengths behind or ahead of the original array and phased correspondingly to give end-fire action, then results in a unidirectional characteristic. Thus the combination of two curtains of stacked broadside arrays gives vertical and horizontal directivity combined with a unidirectional pattern.

The power gain of an array of arrays must usually be found by graphical methods. Formulas are, however, available in certain special cases, as, for example, in the case of a rectangular array in which the radiating elements are doublets.[1] In certain cases, notably the broadside array, tabulated values of mutual impedance may be available from which array gain can be determined.

In systems involving two arrays or curtains arranged in end-fire manner to give a unidirectional characteristic, the addition of the second array or curtain always exactly doubles the power gain (adds 3 db) over that given by the single array or curtain. This makes it necessary to determine only the gain of one array or curtain, and thereby simplifies the gain calculation in certain cases.

Two-element Antenna Arrays.—The two-element antenna array is widely used in broadcast work, where it consists of two vertical antennas that are so spaced and supplied with currents of such relative phase and magnitude as to give the desired field pattern in the horizontal plane. The field distribution of such an array consisting of identical elements A and B is

$$\text{Radiated field} = \{[1 + k \cos (\gamma + \phi)]^2 + k^2 \sin^2 (\gamma + \phi)\}^{\frac{1}{2}} E_\delta \qquad (33)$$

where k = ratio of current in antenna B to current in A.

$$\gamma = \frac{2\pi S}{\lambda} \cos \theta \cos \delta.$$

ϕ = phase angle by which current in B lags current in A.

S/λ = spacing of radiators in wave lengths.

θ = angle in horizontal plane (bearing angle), measured with respect to a line joining the antennas.

δ = vertical angle measured with respect to the horizontal.

E_δ = field radiated by antenna A, with antenna B removed but no change of current in antenna A.

In the special case where the currents in the two antennas are of equal magnitude, the field pattern in the horizontal plane is as shown in Fig. 38 for various phases and spacings.[2] The power gain of a two-element array tends toward a maximum when the spacing is 0.3λ or less and the currents in the two antennas are equal and 180° out of phase. This maximum is approximately 2.5 times (= 4 db) over a single radiator.[3]

The gain of two-element antenna arrays can be readily determined by using the self- and mutual impedances involved in the system to evaluate the power required to maintain the desired currents, as explained in Par. 6.

[1] See Southworth, *loc. cit.* and Sterba, *loc. cit.*

[2] From G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937. Similar curves are given by A. J. Ebel, Directional Radiation Patterns, *Electronics*, Vol. 9, p. 29, April, 1936.

Mechanical devices for solving this equation and plotting the results are described by F. Alton Everest and Wilson S. Pritchett, Horizontal Polar Pattern Tracer for Directional Broadcast Antennas, *Proc. I.R.E.*, Vol. 30, p. 227, May, 1942; W. G. Hutton and R. M. Pierce, A Mechanical Calculator for Directional Antenna Patterns, *Proc. I.R.E.*, Vol. 30, p. 233, May, 1942.

[3] Further discussion of the properties of such arrangements, and also of arrays composed of closely spaced pairs of antennas of this type, is given by J. D. Kraus, Antenna Arrays with Closely Spaced Elements, *Proc. I.R.E.*, Vol. 28, p. 76, February, 1940.

14. Rhombic Antenna.[1]—The rhombic antenna consists of four nonresonant wires arranged in the form of a diamond or rhomboid as in Fig. 39. Such an arrangement gives greater directivity than can be obtained from a single long nonresonant wire, and

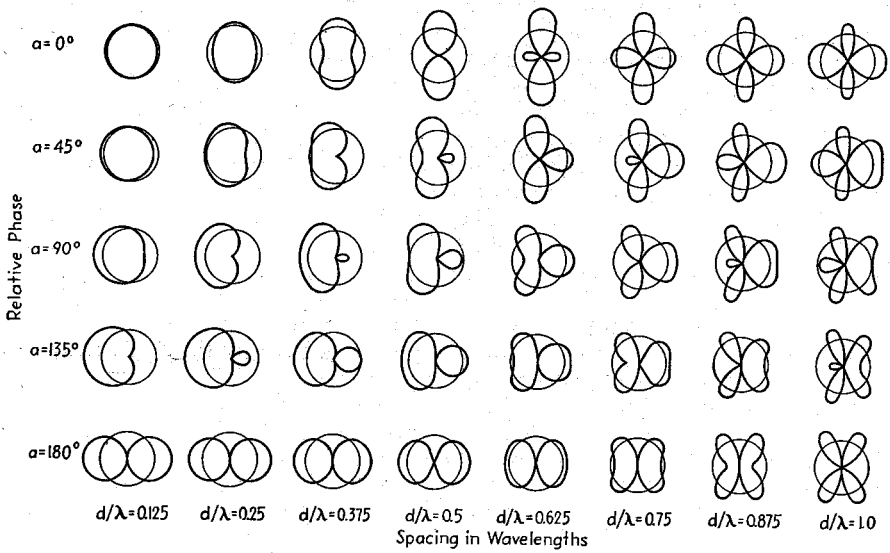


FIG. 38.—Horizontal radiation patterns of arrays of two vertical antennas fed with currents of equal magnitude, showing effect of spacing d and phase a. The circle superimposed on each diagram has the same area as the field pattern actually attained.

furthermore provides a convenient means of terminating the antenna system to obtain the nonresonant condition.

The directional characteristics of the radiation from each leg of the rhomboid is as shown in Fig. 29, and when the tilt-angle ϕ, as defined in Fig. 39, has the optimum value according to Fig. 40, all four legs have major lobes that tend to be pointed in the

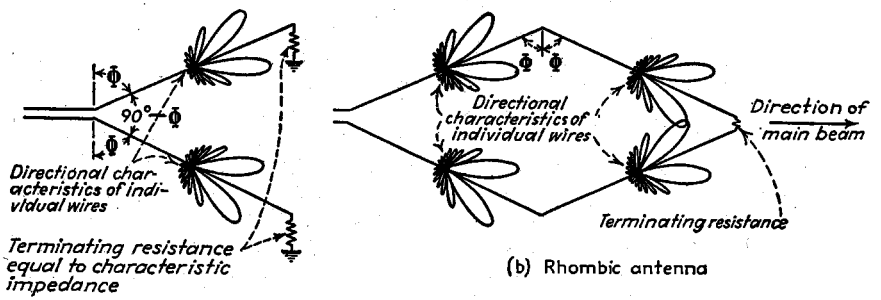


FIG. 39.—Rhombic antenna showing how the radiation from the individual legs combines to produce a maximum in the direction of the terminating resistance.

direction of a line drawn through the apexes, and at the same time tend to add in phase in this direction.

[1] For further information, see E. Bruce, Developments in Short-wave Directive Antennas, *Proc. I.R.E.*, Vol. 19, p. 1406, August, 1931; E. Bruce, A. C. Beck, and L. R. Lowry, Horizontal Rhombic Antennas, *Proc. I.R.E.*, Vol. 23, p. 24, January, 1935.

Radiation from a rhombic (diamond) antenna in space, with characteristic impedance termination, when the attenuation along the wires is neglected, is[1]

$$\left.\begin{matrix}\text{Relative field}\\ \text{strength}\end{matrix}\right\} = \left(\frac{\cos(\phi-\beta)}{1-\cos\theta\sin(\phi-\beta)}+\frac{\cos(\phi+\beta)}{1-\cos\theta\sin(\phi+\beta)}\right) \times \sin\left\{\frac{\pi l}{\lambda}[1-\cos\theta\sin(\phi+\beta)]\right\}\sin\left\{\frac{\pi l}{\lambda}[1-\cos\theta\sin(\phi-\beta)]\right\} \tag{34}$$

where ϕ = tilt angle of antenna (see Fig. 39*b*).
β = bearing angle with respect to line passing through the apex having the terminating resistance and the diagonally opposite apex.
θ = angle of elevation with respect to the plane of the antenna.
l = length of each side of the rhomboid.
λ = wave length.

In the usual case, the rhombic antenna has its plane parallel to the earth and is supported from wooden telephone poles at the corners. The modification in directional pattern obtained under such conditions as a result of the presence of a perfect earth is taken into account by multiplying the relative field strength as calculated from Eq. (34) by the factor given in the right-hand side of Eq. (17).

The optimum tilt angle of a rhomboid antenna is not especially critical with respect to the length of the legs, provided that this length is not less than two wave lengths (see Fig. 40). As a result, the rhomboid can be used over a considerable frequency range without adjustment. The effect of a change in frequency is mainly to alter the vertical angle of the main lobe and the sharpness of the directional pattern somewhat without changing its fundamental character. The frequency range obtainable under practical conditions is approximately 2 to 1, with the higher frequencies having the sharpest patterns, and the main lobe directed closer to the horizontal than the lower frequencies, as shown in Fig. 41.

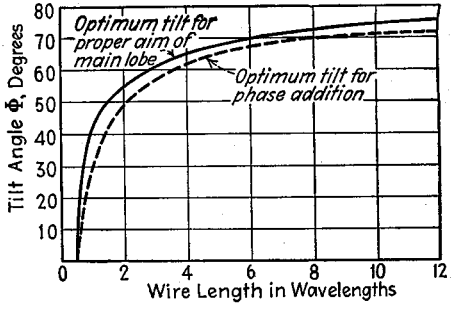


FIG. 40.—Optimum tilt angle for nonresonant antennas in space, assuming zero attenuation.

The designer of a rhombic antenna has as variables at his disposal the height above ground, the length of the legs, and the tilt angle ϕ. It is possible to control the angle above the horizontal for maximum response by varying the tilt angle while leaving the length of the legs constant. Similarly, the effect that insufficient height has on the directional characteristics can in large measure be compensated for by suitably increasing the leg length and suitably modifying the tilt angle.[2]

[1] See Bruce, Beck, and Lowry, *loc. cit.*
This equation involves two approximations. First, the assumption that the attenuation along the wires is zero implies that the antenna radiates no energy, an obvious error. The amount of attenuation actually encountered in practical designs, however, is not sufficient to cause a very large modification. Second, Eq. (34) applies only to the portion of the field that is polarized parallel to the plane of the rhombic antenna. Actually, the antenna radiates some energy that is also polarized at right angles to this and, likewise, when acting as a receiving antenna, is responsive in some degree to waves polarized in this manner. A discussion of the field pattern in the general case, when the various polarizations involved are taken into account, is given by Donald Foster, Radiation from Rhombic Antennas, *Proc. I.R.E.*, Vol. 25, p. 1327, October, 1937. See also the discussion by L. Lewin, *Proc. I.R.E.*, Vol. 29, p. 523, September, 1941.

[2] A discussion of the various possibilities of this sort that can be utilized to advantage in designing rhombic antennas is given by Bruce, Beck, and Lowry, *loc. cit.*

The power gain of a rhombic antenna array depends upon the length of the legs measured in wave lengths, and upon the other proportions of antenna. With leg lengths ranging from two to four wave lengths, the power gain is commonly of the order of 20 to 40 in typical cases. The higher gains tend to go with the longer lengths, since then the concentration of energy in the desired direction is greater, and furthermore the amount of energy radiated is greater in proportion to that dissipated in the terminating resistance.

The radiation resistance of a rhombic can be defined as that quantity which when multiplied by the square of the average current in the wire will equal the radiated power. When the length and breadth are both considerably greater than λ, this resistance R in ohms is[1]

$$R = 240\left[\log_e\left(4\pi\frac{l}{\lambda}\cos^2\phi\right) + 0.577\right] \tag{35}$$

In considering its effect on current distribution, this resistance can be considered as being uniformly distributed along the wire.

The terminating resistance of a rhombic antenna must dissipate a considerable amount of power when the antenna is used for transmitting. This may in typical

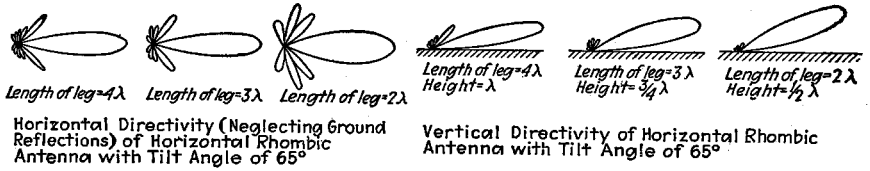


FIG. 41.—Polar diagram showing directional characteristics of the same horizontal rhombic antenna for three different frequencies.

cases be of the order of a quarter to a half of the total power supplied to the antenna, with the exact value depending upon the antenna design. When high-powered transmitters are involved, a convenient way of obtaining a terminating resistance of the required power-handling capacity is to use a two-wire transmission line having a characteristic impedance equal to the desired terminating resistance and employing iron wire to give high loss. This transmission line can be run back from the terminating apex toward the input apex, and after being made sufficiently long to dissipate all except a negligible proportion of the power, can be terminated in a low-wattage resistance, or even left unterminated.

It is possible to modify the minor lobes in the rear of the directional pattern, and in particular to obtain a null in any desired backward direction, merely by modifying the magnitude or phase angle, or both, of the terminating resistance.

There is an advantage in making each conductor of a rhombic antenna consist of two or more spaced wires connected in parallel. This lowers the characteristic impedance of the rhombic antenna, thereby making the terminating impedance less critical and also causing a greater proportion of the total energy supplied to the rhombic to be radiated. There is a further advantage to be gained by arranging such a spaced-wire conductor so that the effective conductor diameter is greater at the two corners of the rhombic that are between the apexes than at the corners of the apexes. It is possible in this way to compensate for the fact that the varying spacing between the sides of the rhombic tends to cause the characteristic impedance of the antenna to be different at different places.

[1] Lewin, *loc. cit.*, or Rhombic Transmitting Aerial Efficiency, *Wireless Eng.*, Vol. 18, p. 180, May, 1941. This latter article also contains additional useful information on the performance of rhombic antennas.

15. The Resonant V Antenna.[1]—This antenna consists of two long resonant wires arranged to form a V and excited so as to carry equal currents that are in phase opposition. The apex angle of the V is made twice the angle that the first lobe in the field pattern of a long resonant wire makes with the wire (see Fig. 19).[2] This gives a strong concentration of radiation in the plane of the V, with the major lobe of the directional pattern in the direction of the line bisecting the V as shown in Fig. 42.

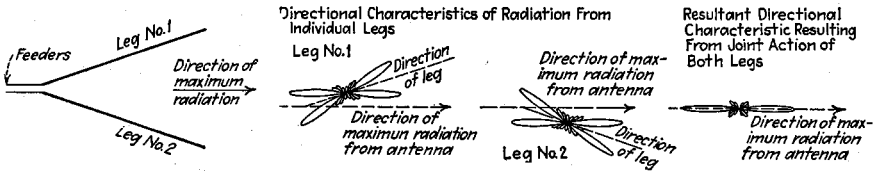


FIG. 42.—V antenna, showing how the radiation from the two legs combine to give a well-defined beam.

The radiation pattern from a V antenna, if it is assumed that the antenna is remote from earth and that each leg is an even multiple of a half wave length long, is[3]

$$\text{Field strength in plane of V} = \sqrt{E_a^2 + E_b^2 - 2E_aE_b \cos\left(2\pi \frac{l}{\lambda} \sin\alpha \sin\phi\right)} \qquad (36a)$$

Radiation in vertical plane passing through bisector of apex angle

$$\epsilon = \frac{120I}{d}\left[\frac{\sin\left(\frac{n\pi}{2}\cos\alpha\cos\theta\right)\sin\alpha}{1-\cos^2\theta\cos^2\alpha}\right] \qquad (36b)$$

where E_a and E_b = radiation in desired direction from individual legs of antenna as given by Eq. (13).
l = length of leg.
λ = length corresponding to one wave length.
α = half of angle at apex.
ϕ = bearing angle with respect to bisector of apex.
ϵ = field strength, volts per meter.
n = number of half wave lengths in each leg of antenna.
θ = angle of elevation with respect to plane of antenna.
I = current at current loop.
d = distance to antenna, meters.

Increased directivity can be obtained by means of an array, each element of which is a V antenna. Thus the directivity in a vertical plane can be improved by stacking two or more V's one above the other, as illustrated in Fig 43*a*. Similarly, a unidirectional pattern can be developed by the use of a second system of V antennas spaced an odd number of quarter wave lengths behind the original system and excited

[1] For further information see P. S. Carter, C. W. Hansell, and N. E. Lindenbald, Development of Directive Transmitting Antennas by RCA Communications, Inc., *Proc. I.R.E.*, Vol. 19, p. 1773, October, 1931; P. S. Carter, Circuit Relations in Radiating Systems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932.

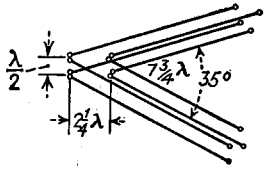
[2] When the sides of the V are short, for example, one wave length or less, the apex angle at which the power gain of the antenna is maximum is less. Thus, for legs one wave length long, maximum gain is obtained with an apex angle of 90° rather than the 105° corresponding to exact superposition of the major lobes. When the antenna is near the earth, the optimum angle is also slightly less than when the V is isolated.

[3] See Carter, Hansell, and Lindenblad, *loc. cit.*

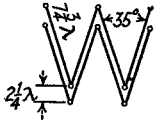
90° out of phase to give an end-fire action (see Figs. 43*a*, 43*b*, and 43*c*).[1] This end-fire action doubles the power gain otherwise obtained. The directional pattern can be sharpened in a horizontal plane by placing two or more sections of V antennas in broadside, as illustrated in Fig. 43*b*. The power gain from this broadsiding is approximately equal to the number of sections; *i.e.*, if three sections are used the power gain will be roughly tripled, etc.

A single section of the type illustrated in Fig. 43*a*, with the sides of the V about eight wave lengths long and with a vertical spacing of $\lambda/2$ between V's, has a power gain of approximately 39 referred to a half-wave antenna.

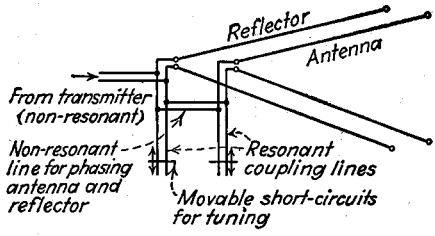
16. Fishbone Antennas.—The fishbone antenna is illustrated in Fig. 44, and is a special form of end-fire array in which the necessary phase relations are maintained by a nonresonant line. The outstanding characteristics of a fishbone antenna are its



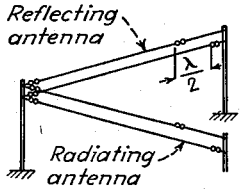
(a) Commercial Design of V Antenna



(b) Plan View of Two Arrays Such as Shown at (a) Placed Broadside to Give Increased Horizontal Directivity



(c) Schematic Design of Feeder System



(d) Alternative Arrangement for Reflecting Antenna

FIG. 43.—Commercial design of V resonant antenna arrays, together with feeder circuits.

ability to operate over a considerable band of frequencies without any readjustment whatsoever and the smallness of the minor lobes in its directional pattern.

The fishbone antenna consists of a series of collectors arranged in collinear pairs, loosely coupled to the transmission line by small capacities supplied by an insulator, as indicated in Fig. 44. The collectors are usually, although not necessarily, horizontal. The phasing is obtained by taking advantage of the fact that when an unloaded transmission line is terminated in its characteristic impedance, the phase shift along the line is proportional to length, and is 360° per wave length.

In the usual design, the collectors are each about 0.3 wave lengths long at the optimum frequency, and are spaced at a distance not to exceed $\lambda/12$ at this frequency. The length of the array is commonly 3 to 5 wave lengths, and is limited by the reactive loading that the coupled antennas place upon the transmission line. In order that

[1] An alternative arrangement for eliminating back-end radiation is shown in Fig. 43*d*, and consists in placing a second V a short distance above or below the first. This second V is a half wave length shorter than the first V, and is excited in phase quadrature. Since its center is almost exactly a quarter wave length closer to the apex than the center of the other V, the two acting together are essentially equivalent to the arrangement of Fig. 43*c*, but only half as many supporting poles are required.

this reactive loading may not be too great, the coupling capacitators are so designed that their loading under the most unfavorable conditions within the range frequencies to be received will not reduce the velocity of phase propagation along the line to less than 90 per cent of the velocity of light. An antenna designed in this way is effective for frequencies ranging from 1.2 to 0.5 times the optimum frequency. This frequency range is limited at low frequencies by the fact that very little energy is coupled into the line, because the collectors are so far out of resonance, and because at low frequencies the antenna begins to lose its directivity. At high frequencies, the limit is fixed by the fact that, as the collectors approach resonance, the resulting reactive loading of the line is sufficient to reduce the phase velocity excessively and so spoils the phasing.

The terminating resistance used in a fishbone antenna should approximate the characteristic impedance. The field pattern in the backward direction can be controlled to some extent by varying either the magnitude or phase, or both, of the terminating impedance. It is possible in this way to make a null occur in almost any backward direction that may be arbitrarily chosen.

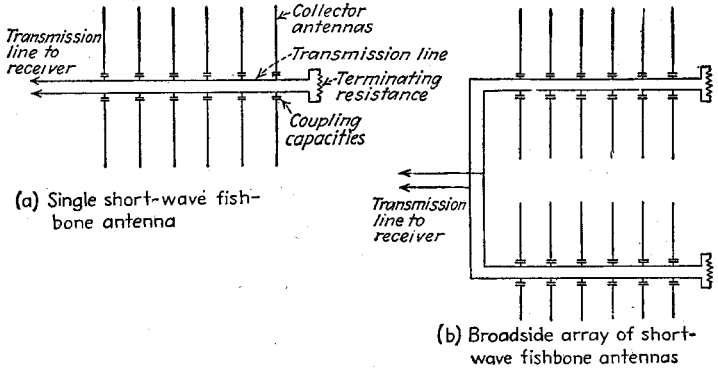


FIG. 44.—Plan view of single fishbone antenna and of an array consisting of two such antennas in broadside.

In practical applications, two fishbone antenna arrays are commonly placed broadside, as illustrated in Fig. 44. This materially improves the directivity in the horizontal plane, and corresponds to an improvement in the power gain of approximately two.

17. Antenna Systems Involving Parasitic Antennas.—In the case of transmitting systems, a parasitic antenna is defined as an antenna that influences the directional pattern but that is excited only by the voltage induced in it by the action of power delivered to a neighboring antenna that is excited. In receiving systems, a parasitic antenna is an antenna associated with the antenna to which the receiver is connected, but reacting on the receiving system only through mutual impedance.

Analysis of Parasitic Antenna Action.[1]—The voltage and current relations that exist in a system involving parasitic antennas can be calculated by the system of equations given in Eq. (5). In the transmitting case, the applied voltages in all the parasitic antennas are zero, while in the receiving case, these applied voltages are the voltages induced in the individual antennas by the radio wave.

The method of using Eq. (5) can be made clear by considering a simple example, such as a driven antenna (designated by subscript 1) associated with a single parasitic

[1] See P. S. Carter, *loc. cit.*; G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937; L. S. Palmer, W. Abson, and R. H. Barker, Multiple Reflections between Two Tuned Receiving Antennae, *Jour. I.E.E.*, Vol. 83, p. 424, 1938; also, Wireless Section *I.E.E.*, Vol. 13, p. 294, September, 1938.

antenna (designated by subscript 2). One then has

$$\begin{aligned} E_1 &= I_1 Z_{11} + I_2 Z_{12} \\ 0 &= I_1 Z_{12} + I_2 Z_{22} \end{aligned} \tag{37}$$

Simultaneous solution of these two equations yields

$$\begin{aligned} I_1 &= \frac{E_1}{Z_{11} - \left(\frac{Z_{12}^2}{Z_{22}}\right)} \\ I_2 &= -I_1 \frac{Z_{12}}{Z_{22}} = -\frac{E_1}{Z_{12} - \frac{Z_{11}Z_{22}}{Z_{12}}} \end{aligned} \tag{38}$$

The field distribution of an antenna array involving parasitic antennas is obtained by assuming a convenient current in the driven antenna and then calculating the

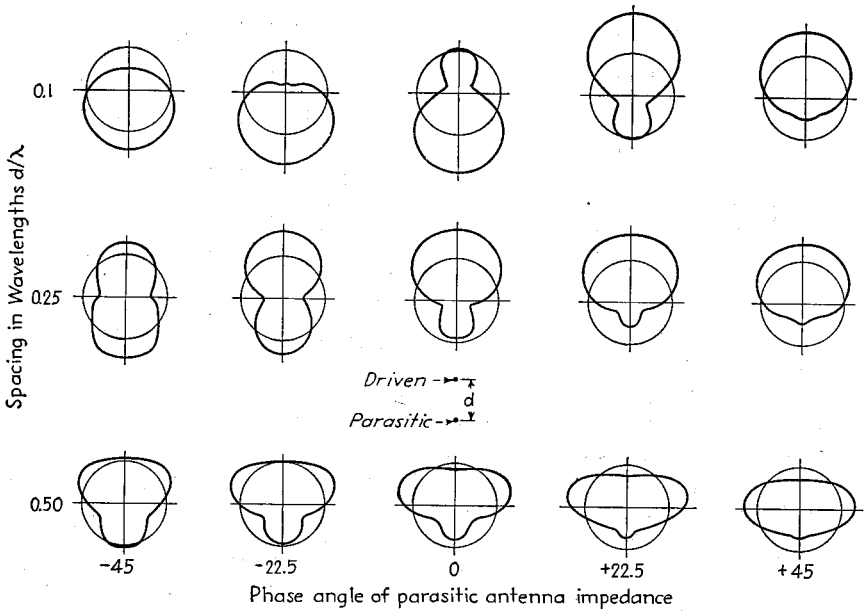


FIG. 45.—Directional patterns obtainable in a horizontal plane with a combination of two vertical antennas, one driven and one parasitic, for different spacings between antennas and different tuning of the parasitic antenna. The phase angles shown are the angles of the self-impedance of the parasitic antenna corresponding to the particular tuning condition. (*From G. H. Brown.*)

magnitude and phase of the currents in the remaining antennas, as illustrated above. The directional pattern obtained with these resulting currents in the various antennas is then determined in the usual manner.

The power gain of an antenna array involving parasitic antennas can be determined from the radiated field produced in the desired direction for an assumed current in the driven antenna, and the power that must be supplied to the driven antenna to produce the currents that give the field. This power can be readily determined from the resistance component of the driving-point impedance of the driven antenna and the current that was assumed in the driven antenna in calculating the field distribution.

Typical Antenna System Involving Parasitic Antennas.[1]—The simplest possible parasitic antenna arrangement consists of a single driven antenna associated with a single parasitic antenna. Even with such a simple system, a wide variety of directional patterns can be obtained, since two variables are involved, namely, the spacing between antennas and the tuning of the parasitic antenna. This is illustrated by Fig. 45, which shows a series of directional patterns obtainable in the horizontal plane from a simple array of two vertical antennas, one driven and one parasitic. A relatively marked unidirectional effect is obtained when the spacing is close and the parasitic antenna is tuned so that its resonant frequency differs slightly from the exciting frequency. This is also shown in Fig. 46. With close-spaced arrangements, a parasitic antenna that is resonant at a lower frequency than that being transmitted will act as a reflector and reduce the field strength in the direction of the parasitic antenna, while if resonant at a higher frequency than that being transmitted, then the parasitic

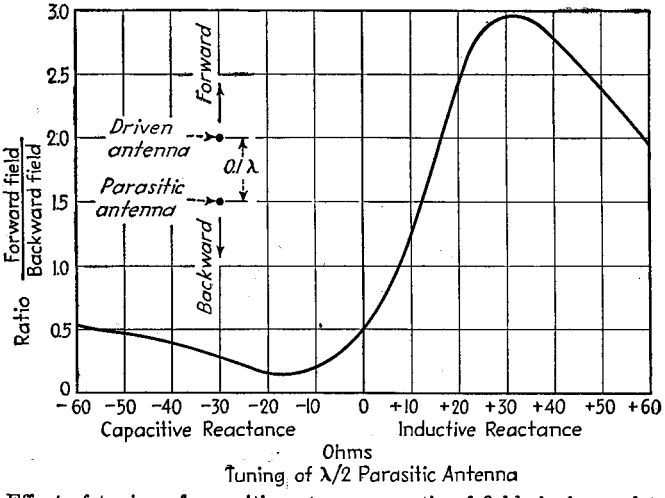


FIG. 46.—Effect of tuning of parasitic antenna on ratio of fields in forward to backward direction in a typical case.

antenna acts as a "director" that tends to concentrate the radiated field in its direction.

Close-spaced arrays involving a driven antenna, and one or more parasitic antennas are frequently used to obtain directivity with a minimum of size. Typical examples are illustrated in Figs. 47*b* and 47*c*, and consist of a radiating antenna, a reflector, and one or more directors with spacings as indicated. The desired directivity is obtained by experimental adjustment of the directors and reflector, it being kept in mind that the reflector is always tuned to a slightly lower frequency than that being radiated, and the directors to a slightly higher frequency. A properly adjusted three-element array such as illustrated in Fig. 47*b*, composed of radiators approximately a half wave length long, will give a gain of 5 to 7 db over a half-wave antenna. Still greater gains are obtainable by the use of additional directors.

Arrangements such as shown in Fig. 47*c*, involving a number of parasitic antennas arranged along a line, are often termed *Yagi arrays.*

[1] For further information and examples, see Brown, *loc. cit.;* Hidetsugu Yagi, Beam Transmission of Ultra-short Waves, *Proc. I.R.E.*, Vol. 16, p. 715, June, 1928; A. Wheeler Nagy, An Experimental Study of Parasitic Wire Reflectors on 2.5 Meters, *Proc. I.R.E.*, Vol. 24, p. 233, February, 1936.

Examples of other arrays involving parasitic antennas are also illustrated in Fig. 47. A variety of directional patterns can be obtained with these arrangements, according to the spacings involved and the tuning of the parasitic antennas. It is

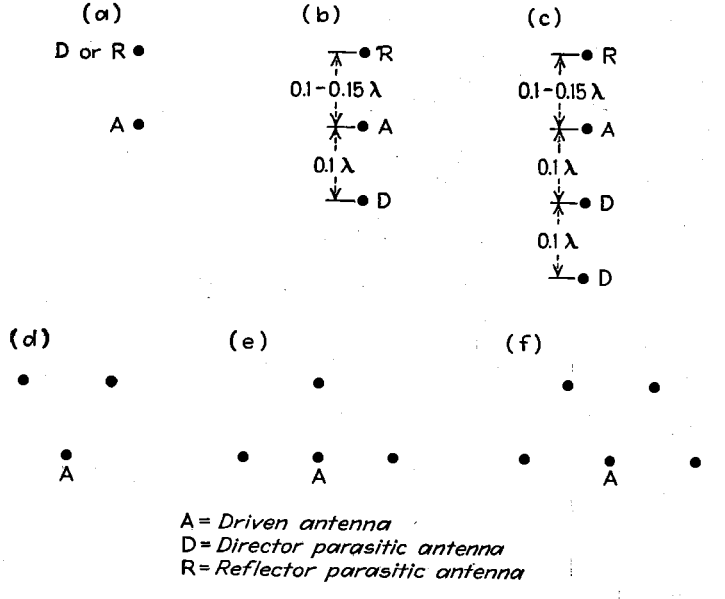


FIG. 47.—Examples of antenna arrays involving parasitic antennas.

possible with arrangements of this sort to obtain very marked directivity, as illustrated by Fig. 48.[1]

In antenna arrangements such as are illustrated in Fig. 37, where two antenna arrays are spaced an odd number of quarter wave lengths apart in order to obtain an end-fire effect that will provide a unidirectional pattern, the reflecting curtain or array is frequently excited parasitically instead of being driven. When this is the case, the magnitude and phase of the induced currents are controlled by the tuning of the parasitic array, which must be carefully adjusted to give the most desirable characteristics.

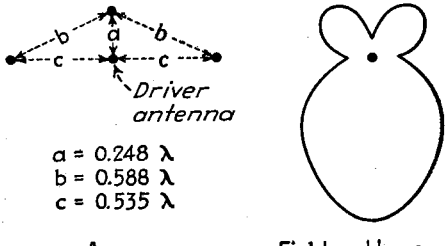


FIG. 48.—Field distribution in horizontal plane obtained from an array of vertical antennas, one of which is driven and three of which are parasitic, arranged as shown.

Adjustment of Antenna Systems Involving Parasitic Radiators.—In transmitting systems, the most convenient way to adjust the parasitic radiators is to excite the driven antenna, place a receiver at a convenient distance in the desired direction, and then vary the tuning of the parasitic radiator by cut and try until the best results are obtained. In the case of an array intended for receiver use, one can adjust by placing a small transmitter at some distance in the undesired direction and then adjusting the parasitic antenna by

[1] From G. H. Brown, Directional Antennas, *Proc. I.R.E.*, Vol. 25, p. 78, January, 1937.

cut and try until a receiver associated with the receiving antenna indicates minimum response.

18. Loop Antennas.—The loop antenna is essentially a coil of any convenient cross section (see Fig. 49 for examples), carrying a radio-frequency current. The ordinary loop is so designed that its dimensions are small compared with the wave length, in which case the currents are of the same magnitude and phase throughout the loop.[1]

Characteristics of Ordinary Loops.—The field radiated in a plane perpendicular to the plane of the loop by an ordinary loop in which the dimensions are so small compared with a wave length that the current can be considered as constant throughout the loop is[2]

$$\epsilon = \frac{120\pi^2}{d} N \left(\frac{A}{\lambda^2}\right) I_s \cos\theta \tag{39}$$

where ϵ = field strength, volts per meter.
d = distance, meters.
N = number of turns in loop.
A = area of loop, square meters.
λ = wave length, meters.
A/λ^2 = area in wave lengths squared.
I_s = loop current, amp.
θ = angle with respect to plane of loop.

This relation holds irrespective of the shape of the loop, provided that the loop is small compared with a wave length.

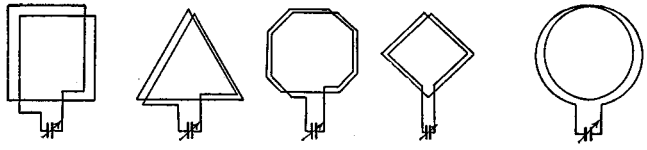


FIG. 49.—Typical loop antennas.

When a loop is used as a receiving antenna, the voltage induced in the loop by a wave polarized in the plane of the loop is

$$\left.\begin{matrix}\text{Voltage induced} \\ \text{in loop}\end{matrix}\right\} = 2\pi\epsilon N \frac{A}{\lambda} \cos\theta \tag{40}$$

The notation is the same as in Eq. (39), except that ϵ is the strength of the passing radio wave in volts per meter.

[1] When the loop is small compared with a wave length, but not extremely small, the distributed capacity and leakage inductance between turns of the loop may cause the current to vary in magnitude throughout the loop, although the phases will be substantially the same. Under these conditions, the directional pattern may be slightly different from that of an idealized loop. See E. M. Williams, Radiating Characteristics of Short-wave Loop Aerials, *Proc. I.R.E.*, Vol. 28, p. 480, October, 1940. Also, the voltage step-up ratio of the loop will be affected, so that if the loop is used in measuring field strength, a correction must be made for its distributed constants (see Par. 40, Sec. 13).

Further information on loop behavior under conditions where the distributed nature of the loop inductance and capacity must be taken into account is given by Paul B. Taylor, Theory of Loop Antenna with Leakage between Turns, *Proc. I.R.E.*, Vol. 25, p. 1574, December, 1937; F. M. Colebrook, The Application of Transmission Line Theory to Closed Aerials, *Jour. I.E.E.*, Vol. 83, p. 403, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 273, September, 1938; V. I. Bashenoff and N. A. Mjasoedoff, The Effective Height of Closed Aerials, *Proc. I.R.E.*, Vol. 19, p. 984, June, 1931; L. S. Palmer and D. Taylor, Rectangular Short-wave Frame Aerials for Reception and Transmission, *Proc. I.R.E.*, Vol. 22, p. 93, January, 1934.

[2] J. H. Dellinger, Principles of Radio Transmission and Reception with Antenna and Coil Aerials, *Bur. Standards Sci. Paper*, 354, December, 1919; also, *Trans. A.I.E.E.*, Vol. 38, p. 1348, 1919.

The field pattern represented by Eqs. (39) and (40) is shown in Fig. 50, and when the loop is remote from ground can be considered as a cross section of a figure of revolution about an axis that is perpendicular to the plane of the loop. It will be noted that the field distribution produced by a loop antenna when remote from ground is exactly the field distribution from a doublet antenna located at the center of the loop, having its axis perpendicular to the plane of the loop. The two patterns differ, however, in that the field radiated from the loop is everywhere polarized in a direction parallel to the plane of the loop.

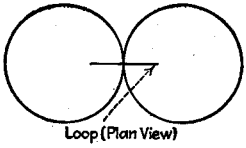


FIG. 50.—Directional characteristic of loop antenna. This applies to loops of all shapes, provided that the loop is small.

Equations (39) and (40) assume that the loop is remote from earth. If this is not the case the method of images may be used to determine the influence that the earth has on the directional pattern. When the loop is vertical, the effect of the ground is to modify the distribution in the vertical plane, but the relative distribution in a horizontal plane is still as shown in Fig. 50.

The radiation resistance of a loop antenna that is small compared with a wave length and is remote from earth is[1]

$$\left.\begin{matrix}\text{Radiation resistance} \\ \text{in ohms}\end{matrix}\right\} = 31{,}200 \left(N \frac{A}{\lambda^2}\right)^2 \tag{41}$$

where N, A, and λ have the same meaning as in Eq. (39).

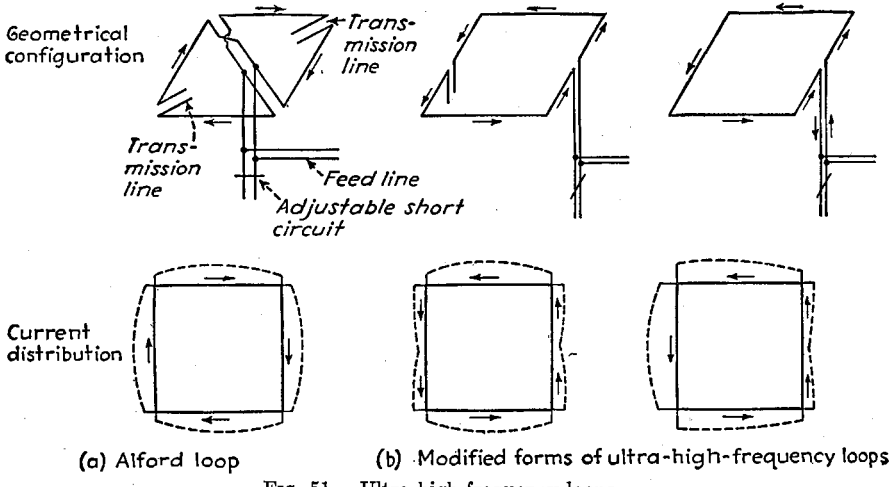


FIG. 51.—Ultra-high-frequency loops.

The radiation resistance of the ordinary loop is very small. As a result, it is difficult to make the loss resistance low enough to obtain high efficiency when a loop antenna is used as a radiation or to obtain appreciable energy pickup when the loop is used in reception.

Ultra-high-frequency Loop Antennas.—The low efficiency obtained with an ordinary loop transmitting antenna can be overcome at ultra-high frequencies by an ingenious arrangement due to Alford,[2] which is shown in Fig. 51*a*. The sides of the Alford loop

[1] Andrew Alford and A. G. Kandoian, Ultra-high-frequency Loop Antennas, *Trans. A.I.E.E.*, Vol. 59, p. 843, 1940.

[2] See Alford and Kandoian, *loc. cit.* A modified form of high-frequency loop involving three voltage fed half-wave antennas in a triangular configuration is described by N. E. Lindenblad, Antennas at Empire State, *Communications*, Vol. 21, p. 9, May, 1941.

are sections of a resonant transmission line so arranged that there is a current loop at the center of each side, with the current in the various sides all in phase.

The Alford loop has radiation characteristics that are nearly the same as those of an ordinary loop, but at the same time permits the use of loop dimensions that are not negligible compared with a wave length.[1] This results in a relatively high radiation resistance, enabling the Alford loop to act as an efficient transmitting or receiving antenna.

The radiation resistance of an Alford loop such as shown in Fig. 51*a*, referred to the current I_0 in the center of the leg, is given by the semiempirical equation

$$\left.\begin{matrix}\text{Radiation resistance}\\ \text{in ohms}\end{matrix}\right\} = 320\left[\sin\left(\pi\frac{l}{\lambda}\right)\right]^4 \tag{42}$$

where l/λ is the length of the sides in wave lengths.

Modified forms of the Alford loop are illustrated in Fig. 51*b*. In these, the radiation pattern in the plane of the loop is no longer always nearly circular, because the currents in the various sides of the loop may show marked asymmetries.

Combination of Loop and Vertical Antenna.—When the output of a vertical antenna is coupled into the circuit of a vertical loop in such a way that the voltage induced

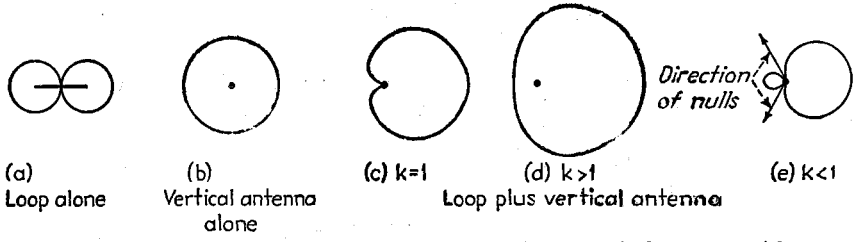


FIG. 52.—Directional patterns obtained by combining a vertical antenna with a vertical loop. The quantity k represents the contribution of the vertical antenna to the output, as compared with the loop contribution.

in the loop by the coupling is 90° out of phase with the voltage the passing wave induces in the vertical antenna, and if at the same time the vertical antenna is located near the loop, then the directional characteristic in the horizontal plane is

$$\text{Relative response} = k + \cos\theta \tag{43}$$

where θ is the bearing angle with respect to the plane of the loop and k is the relative contribution of the loop and antenna actions ($k = 1$ for equal effects). The resulting directional patterns for different amounts of vertical antenna pickup are illustrated in Fig. 52. When the loop and antenna contribute equally to the output, the result is a cardioid, as in Fig. 52*c*, whereas if they are unequal and k is less than unity, the directional pattern exhibits two minima that are not exactly 180° with respect to each other as is the case with a simple loop.

19. Polyphase Antenna Systems. *Turnstile Antenna.*[2]—The turnstile antenna consists of two half-wave radiators crossing each other at right angles, as shown in Fig. 53*a*, and excited 90° out of phase. Such an arrangement has a field distribution

[1] The field pattern of an Alford loop will differ from that of an ordinary small loop in two respects. First and most important, there will be departures from a circular pattern in the plane of the loop as a result of the fact that the current is not constant around the perimeter of the loop. Second, the radiation drops off a little more rapidly with angle θ measured with respect to the plane of the loop than $\cos\theta$. This is because the length of the side is not negligible compared with a wave length. These effects are negligible when the sides do not exceed $\lambda/8$ in length, and are not large even for sides $\lambda/4$ long.

[2] George H. Brown, The Turnstile Antenna, *Electronics*, Vol. 9, p. 15, April, 1936.

in the plane of the loop that is approximately circular, as shown in Fig. 53*b*. This is obtained from the field patterns of the individual antennas by noting that the individual radiations in any particular direction are in time quadrature, and so combine to give a resultant that is the square root of the sum of the squares of the individual radiations in that direction.

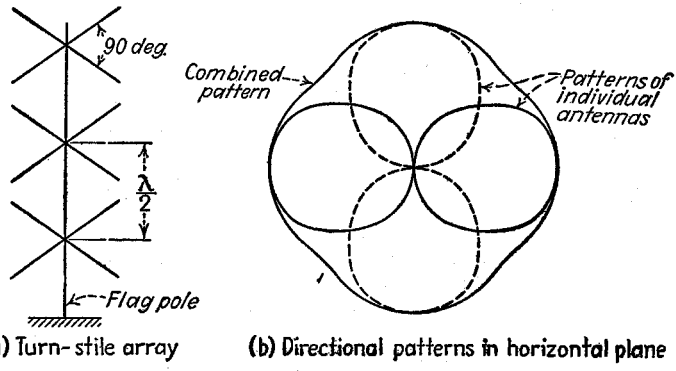


FIG. 53.—Stacked array of turnstile antenna systems, together with directional patterns in horizontal plane.

Turnstile antennas are sometimes arranged in an array in which a number of individual turnstiles in a horizontal plane are arranged one above the other in a broadside array, in which the line of the array is vertical, as shown in Fig. 53*a*. This gives approximately uniform radiation in all horizontal directions, but provides directivity in a vertical plane.

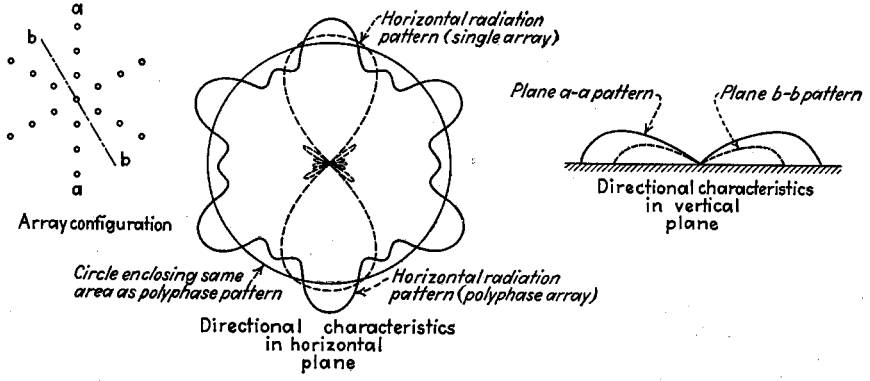


FIG. 54.—Polyphase antenna array that concentrates radiation along the horizontal without requiring a high pole structure. Varying the angle between the lines of antennas or modifying the arrangement of antennas within a line, etc., makes it possible to control the horizontal and vertical patterns.

Polyphase Antenna Arrays.[1]—Marked directivity in a vertical plane can be obtained at broadcast and lower frequencies by using a large number of short vertical radiators in a polyphase array, as illustrated in Fig. 54. Here short vertical radiators are arranged at regular intervals in linear arrays, a number of such linear arrays are arranged as shown, and then the various linear arrays are excited in a polyphase

[1] Such arrangements have been proposed by Terman, *op. cit.*, p. 678.

manner. It is possible in this way to realize an approximately nondirectional characteristic in the horizontal plane, while marked directivity in the vertical is maintained. Directional characteristics for a typical case are shown in Fig. 54. Such polyphase arrangements are very flexible, since by simple changes in the individual array it is possible to vary the directional characteristics in the horizontal plane in almost any manner desired.

20. Ring Antenna Systems.[1]—Marked directivity in the vertical plane combined with a substantially circular pattern in a horizontal plane can be obtained by arranging short vertical radiators in concentric rings and providing a uniform progressive phase shift between adjacent antennas in the individual rings such that the total phase shift around the circumference of each ring is the same and is a whole number of cycles.

Two types of such ring antenna systems have especially desirable properties. The first, termed the J_0 type, consists of a multiplicity of concentric rings, with all antennas in the same ring having the same phase. The other, termed the J_n type, has only a single ring, but with the total phase shift around the ring being n cycles, where n is an integer not less than 1.

The design of a J_0 antenna is carried out as follows: First, an integer m is selected that will give the desired power gain when substituted in the approximate relation

$$\text{Power gain} \cong 1.17\sqrt{m - 0.5} \tag{44}$$

This power gain is expressed with respect to a short vertical grounded antenna, and assumes that the radiators of which the ring array is composed are likewise short vertical radiators. The value of m obtained in this way then determines a value of the quantity $k'\rho_0$ according to the relation

$$k'\rho_0 = \pi(m - 0.25) \tag{45}$$

The quantity k' is defined by the relation

$$k' = \frac{2\pi}{\lambda} S \tag{46}$$

where

$$S = 1 + \frac{1.8}{k'\rho_0 - \frac{\pi}{4}} \tag{47}$$

The radius ρ of a ring is then made such that $J_0(k'\rho)$ is a maximum, corresponding to

$$\rho \cong 0.62\frac{\lambda}{S} + \frac{p\lambda}{2S} \text{ wave lengths} \tag{48}$$

where p has the value 0 for the first ring, 1 for the second ring, 2 for the third ring, etc., and the outer ring in the array is the one for which the radius ρ is a little less than $k'\rho_0$. There is also a central antenna. The number of antennas in each ring should be not less than $1 + k'\rho_0/S$. The total current in each ring, when the current in the center antenna is taken as unity, will then be 2.00 for the first ring, 2.9 for the second

[1] W. W. Hansen, and J. R. Woodyard, A New Principle in Directional Antenna Design, *Proc. I.R.E.*, Vol. 26, p. 333, March, 1938; W. W. Hansen and L. M. Hollingsworth, Design of "Flat-shooting" Antenna Arrays, *Proc. I.R.E.*, Vol. 27, p. 137, February, 1939; H. Chiereix, Antennas a Rayonnement Zenithal Réduit, *L'Onde electrique*, Vol. 15, p. 440, July, 1936. See also U. S. Patent 2,218,487, Oct. 15 1940, issued to W. W. Hansen and F. E. Terman.

ring, 3.6 for the third, then 4.2, 4.7, etc.[1] This total current in a particular ring is equally divided among the antennas in the ring, and the phases of successive rings differ by exactly 180°. An example of a J_0 antenna system is given in Fig. 55.

The J_n antenna is designed by first selecting the total number of cycles n of phase shift around the ring to give the desired gain, according to the approximate equation

$$\text{Power gain} = 0.752\sqrt{n + 1.5} \qquad (49)$$

This assumes that the array is composed of short vertical radiators, and gives the power gain with respect to a single such radiator. The minimum number of radiators required is $2n + 3$. With fewer than this, the radiation pattern in the horizontal plane will not be circular. The radius ρ of the ring must be small enough to satisfy the relation $(2\pi\rho/\lambda)^2 << 4$. The exact value of the radius is otherwise immaterial except for the fact that the smaller the radius the lower will be the radiation resistance, and hence the more difficult it becomes to reduce the other losses sufficiently to provide an efficient radiating system. The J_n antenna is characterized by a directivity in a vertical plane that has no minor lobes, as shown in Fig. 56.

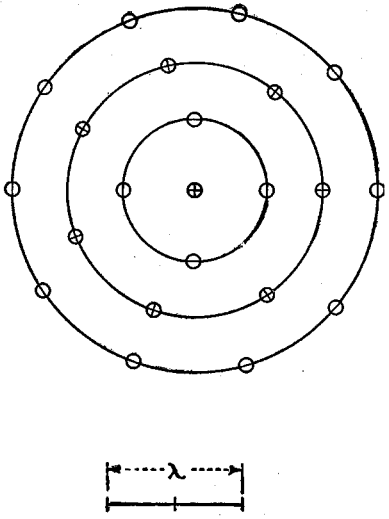


FIG. 55.—Plan view of J_0 antenna array with $m = 4$, having a gain of 2.3. The circles represent array elements, those with a (−) inside having the current flowing 180° out of phase with the currents in the antennas marked (+). The radii of the successive rings are 0.52λ, 0.95λ and 1.39λ, and the relative currents in the *individual* antennas are 1, 0.50, 0.41, and 0.36 for the successive rings.

Comparison of the characteristics of the J_0 and J_n arrays indicates that the latter is preferable when high gain is desired, whereas the former becomes more economical with moderate gains. The transition point at which the multiple-ring arrangement has approximately the same desirability as the single ring occurs when $(k'\rho_0)/S$ is in the range 6.65 to 10.

21. Antenna Systems Involving Plane Reflectors. *Systems Employing a Single Flat Reflector.*—In ultra-high-frequency work, a flat conducting sheet, usually of copper, is often placed near an antenna system to modify the field pattern and increase the gain. Such a reflector acts in the same as a perfectly conducting earth, and its effect can be analyzed in the same way.

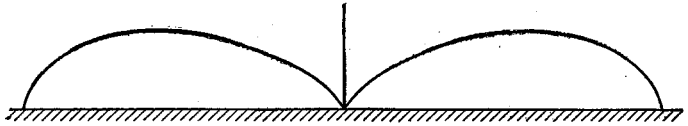


FIG. 56.—Field produced in a vertical plane by a single-ring antenna for $n = 5$, corresponding to a gain of 2.1.

The simplest arrangement involving a flat reflector of this type consists of a half-wave antenna adjacent to and parallel with a copper sheet or screen. In such an arrangement, the spacing between the antenna and the reflector is preferably of the

[1] These total relative currents in each ring are proportional to the areas under the successive loops of the function $\rho J_0(k'\rho)$, with each ring of antennas being identified with the corresponding loop of the Bessel function.

order of **0.1 to 0.3 wave lengths.** The maximum radiation is then in a direction at right angles to the plane of the reflector, and the power gain referred to an isolated half-wave antenna is of the order of 6 db (4 times). The effect of spacing on the gain for zero loss resistance of the antenna, and for the case where a loss resistance of 1 ohm is assumed, is shown in Fig. 59 by the curves marked 180°. The variation of radiation resistance with spacing is shown by the curve marked 180° in Fig. 60. The low resistance with very small spacings makes such close spacings undesirable, because then even a small loss resistance will reduce the antenna efficiency substantially.

An antenna spaced close to a reflector is equivalent to a close-spaced two-element array such as is discussed in Par. 13. The two arrangements differ, however, in that the antenna with reflector gives a unidirectional pattern and so has somewhat more power gain. The arrangement with a reflector is accordingly to be preferred to a close-spaced array where it can be employed.

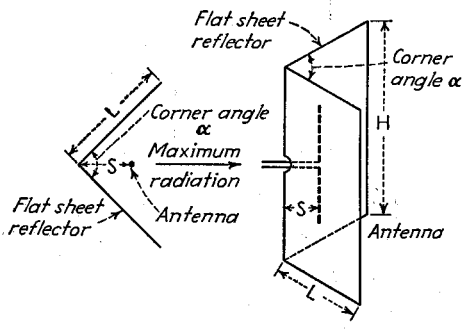


FIG. 57.—Corner reflectors, in cross section and in perspective.

Flat reflectors are sometimes employed in association with broadside antenna arrays at ultra-high frequencies as a substitute for a reflecting curtain, to give a unidirectional characteristic. Such an arrangement has been termed a "billboard" array, and has an advantage over the equivalent arrangement employing a reflecting curtain in that the adjustments associated with the latter are eliminated. The only adjustment in the billboard arrangement is the spacing between the arrays and the reflector. This spacing can be an odd number of quarter wave lengths, or it can be

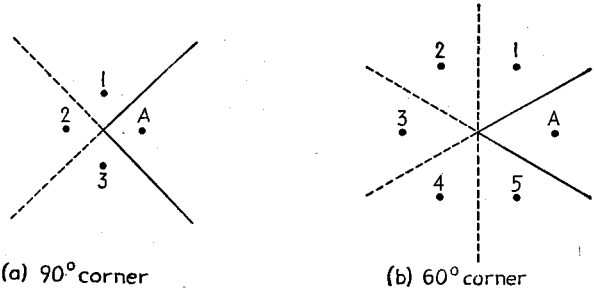


FIG. 58.—Image system for corner antennas. Here *A* represents the exciting antenna, and the numbered antennas are images. The antenna *A*, in conjunction with its associated images (and in the absence of all reflectors), produces the same field in the direction of the reflector mouth as does the system consisting of *A* and the corner reflector.

less than a quarter of a wave length. In the latter case, the spacing is not at all critical, but the radiation resistance of the antennas is lowered appreciably by the presence of the reflector, and the directivity will usually not be quite so great as that obtainable with larger spacing, as, for example, $3\lambda/4$ or $5\lambda/4$.

Antenna with Corner Reflector.[1]—The corner reflector antenna consists of a driven radiator, normally a half-wave antenna, associated with a reflector constructed of two

[1] John D. Kraus, The Corner-reflector Antenna, *Proc. I.R.E.*, Vol. 28, p. 513, November, 1940

flat conducting sheets (or their electrical equivalent), which meet at an angle to form a corner. An arrangement of this sort is shown in Fig. 57.

When the corner angle α is $180/n$, where n is an integer, the effect of the corner reflector can be taken into account by the method of images. For a 90° corner, it is

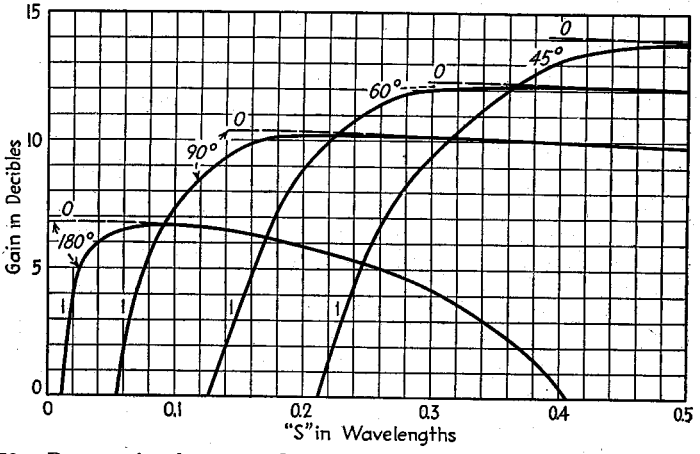


FIG. 59.—Power gain of corner-reflector antenna systems compared with a half-wave antenna in free space for different values of the corner angle α. The curves marked "0" are for zero loss resistance; those marked "1" are for an assumed loss resistance of the exciting antenna of 1 ohm.

necessary to postulate three images as shown in Fig. 58*a*, a 60° corner requires five images, as in Fig. 58*b*, and a 45° corner, seven images, etc.

The principal effect of a corner reflector is to concentrate the radiation in the direction of the bisector of the corner, with the maximum concentration taking place

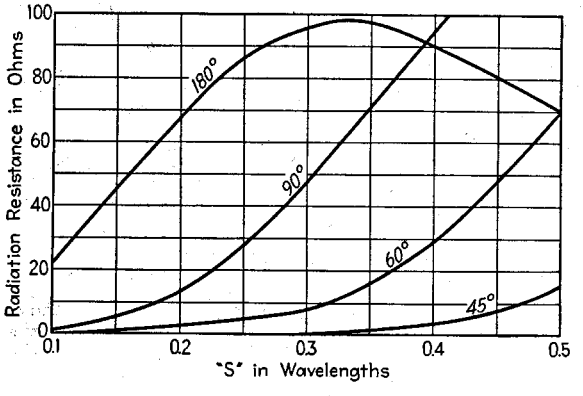


FIG. 60.—Radiation resistance of driven half-wave antenna in corner-reflector systems as a function of the antenna-to-corner spacing S for various values of the angle α.

in the horizontal plane, but with incidental concentration in the vertical plane. For best results there is an optimum spacing S between the antenna and the corner. If the spacing is too large, the directional pattern begins to show multiple lobes, the gain drops off, and the arrangement also becomes physically large. On the other hand, as the spacing becomes small, the radiation resistance is reduced, and incidental

loss resistances then consume an increasingly large fraction of the energy. The relationship between spacing and gain is shown in Fig. 59 for various corner angles. The radiation resistance is shown in Fig. 60. It will be noted that power gains of the order of 10 to 13 db (10 to 20 times), as compared with an isolated half-wave antenna, are readily obtainable in a relatively simple, compact arrangement having a reasonable radiation resistance.

In order to realize the theoretical possibilities, the corner reflector should have a height H such that the difference between the antenna length and the height is preferably not less than S. Likewise, the length of L of the reflector sides (see Fig. 57) should be at least $3S$ to $4S$.

The conducting reflector can be approximated satisfactorily by a grid-type arrangement involving spaced parallel wires orientated in the same plane as the antenna, as illustrated in Fig. 61. The spacings between the wires of such a reflector must be a relatively small fraction of the antenna spacing S, if the action is to be the full equivalent of a solid reflector. It is to be noted that the reflector wires are not parasitic antennas in the ordinary sense, in that their length is not critical, but rather must merely be sufficient to give an effect approximately the same as though one had an infinite length. This arrangement has the advantage that it has lower resistance to wind, and is also much lighter than a solid reflector.

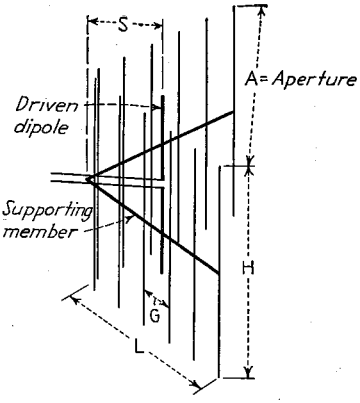


FIG. 61.—Grid-type corner reflector having spaced parallel wires in place of the usual conducting sheet.

22. The Wave Antenna (Beverage Antenna).[1]—This is a form of nonresonant antenna used in the reception of long-wave signals. It consists of a wire ranging from one-half to several wave-lengths long, pointed in the direction of the desired transmitting station and mounted at a convenient height (usually 10 to 20 feet)

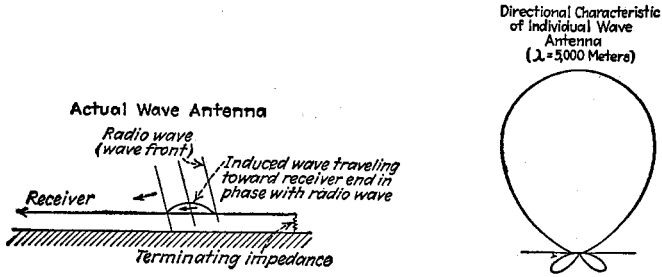


FIG. 62.—Details of action taking place in wave antenna when the incident wave arrives from the direction in which the response is maximum, together with directional diagram when the wire is 0.9 wave lengths long.

above earth. The end toward the transmitting station is grounded through an impedance approximating the characteristic impedance of the antenna when considered as a one-wire transmission line with ground return, while the energy abstracted from the radio wave is delivered to a radio receiver at the end of the antenna farthest from the transmitter.

[1] Harold H. Beverage, Chester W. Rice, Edward W. Kellogg, The Wave Antenna—A New Type of Highly Directive Antenna, *Trans. A.I.E.E.*, Vol. 42, p. 215, 1923.

The wave antenna abstracts energy from passing waves as a result of the wave tilt that the earth losses produce in vertically polarized low-frequency waves traveling along the surface of the ground. When the wave is traveling in the direction of the wire toward the receiver, as in Fig. 62, the currents induced in different parts of the wire all add up in phase at the receiver because the currents induced in the wire travel with the same velocity as the radio wave and keep in step with each other. When the wave travels in the opposite direction, the energy builds up as the terminated end is approached, but the terminating resistance absorbs this energy, and there is little or no effect on the receiver. Waves arriving from the side have little resultant effect at the receiver, because, although the various induced currents are more or less in phase at their point of origin, they are initiated at widely different distances from the receiver.

The wave antenna has marked directivity in the horizontal plane, as shown in Fig. 62, and this can be still further improved by arranging several wave antennas in a simple array.[1] Since the wave antenna is nonresonant, it can be used to receive simultaneously signals of different wave lengths, provided that the transmitters all lie in the same general direction.

The wave antenna is suitable only for reception. In transmitting, the earth losses are so great compared with the radiated energy that the low efficiency more than counterbalances the directivity obtained.

23. Antenna Systems with Steerable Directivity.[2]—Antenna systems of this type are arranged to have a very sharp directivity in a vertical plane, and are provided with means whereby the angle of maximum effectiveness in the vertical plane can be adjusted so as to coincide with the angle of arrival of the radio waves, as explained in Par. 7, Sec. 9.[3] Antennas of this type have been referred to as *Musa systems*, from the first letters of the words in the phrase "multiple unit steerable antenna."

The Musa antenna system, as employed in commercial practice, consists of a number of rhombic antennas, uniformly spaced along a line. The outputs of the antennas are then so combined as to provide a uniform constant progressive phase difference between adjacent antennas. The vertical angle for maximum reception can be controlled by adjusting this phase difference according to the equation

$$\psi = 2\pi \frac{d}{\lambda} \cos \theta \tag{50}$$

where d/λ = spacing between centers of adjacent antennas in wave lengths.
θ = angle of wave arrival above horizontal for which reception is to be best.
ψ = phase difference required between adjacent antennas to make reception of wave at angle θ a maximum.

The phasing is accomplished by means of adjustable phase shifters that are interconnected so that they can be operated from a single control. In the Musa system these phase shifters operate at the intermediate frequency of the receivers. This is accomplished by having the output of each antenna system delivered to a separate mixer, with all such mixers supplied by voltage from a common beating oscillator. This arrangement takes advantage of the fact that when two radio-frequency voltages are separately heterodyned with the same beating oscillator, the difference frequency outputs obtained have the same relative phases as do the original radio-frequency voltages before the frequency change was accomplished.[4]

[1] Austin Bailey, S. W. Dean, and W. T. Wintringham, The Receiving System for Long-wave Transatlantic Radio Telephony, *Proc. I.R.E.*, Vol. 16, p. 1645, December, 1928.

[2] H. T. Friis and C. B. Feldman, A Multiple Unit Steerable Antenna for Short Wave Reception, *Proc. I.R.E.*, Vol. 25, p. 841, July, 1937.

[3] In the case of a transmitting antenna, the object would be to adjust the directivity so that it coincided with one of the preferred angles at which energy is effectively transmitted to the receiving point.

[4] H. T. Friis, A New Directional Receiving System, *Proc. I.R.E.*, Vol. 13, p. 685, December, 1925.

A simpler system of directivity steering, but one that has much less directivity in a vertical plane, can be obtained by employing a rhombic antenna in which the tilt angle is varied.[1] A structural arrangement is employed in which the apex of the rhombic opposite the terminating resistance is fixed, while the apex at the terminating resistance end is kept taut by a constant pull obtained from a rope passing over a pulley and loaded with weights. The directivity is adjusted with the aid of ropes connected from the other two corners of the rhombic, passing over pulleys at the supporting towers and leading to a winch. With this arrangement, winding up these ropes on the winch will open up the sides of the rhombic and reduce the tilt angle, which, in turn, changes the vertical directivity.

24. Antenna Systems to Produce a Specified Directional Characteristic.[2]—Given a field pattern that has mirror image symmetry about an axis, it is then always possible to determine a system of radiators, which, when distributed along this axis, will produce a field pattern approximating the given pattern to any desired degree of accuracy.

The current distribution required to do this can be built up from pairs of radiators, with the currents in the two antennas of a pair being of equal magnitude (but not necessarily equal phase). The pairs are arranged symmetrically along the axis about a common center point, and the spacing of each pair is an integral multiple n of the spacing of the pair of radiators closest to the center point. The arrangement is shown in Fig. 63, when $d_n = nd_1$.

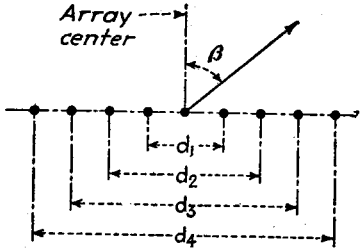


FIG. 63.—Antenna system composed of symmetrical pairs.

The specification of the basic distance d_1 of the first pair of radiators, and of the relative magnitudes and phases of the currents in the various pairs of radiators is accomplished as follows: First, it is to be noted that the nth pair of radiators makes a contribution to the total field in the horizontal plane that is given by the equation

$$E_n = A_n \cos\left(\pi n \frac{d_1}{\lambda} \sin\beta + \frac{\phi_n}{2}\right) \tag{51}$$

where E_n = field of nth antenna pair.
A_n = proportionality factor.
d_1/λ = spacing of first pair of antennas in wave lengths.
ϕ_n = phase difference between the currents in the two antennas of the nth pair.
β = bearing angle measured with respect to perpendicular of line of array (see Fig. 63).

The total field E radiated by a system of n such antenna pairs, including the radiation A_0 from a single antenna at the array center, is accordingly

$$E = A_0 + \sum_{n=1}^{n=n} A_n \cos\left(nw + \frac{\phi_n}{2}\right) \tag{52}$$

where

$$w = \pi \frac{d_1}{\lambda} \sin\beta \tag{53}$$

Equation (52) assumes that the mean phase of the currents in each pair of antennas is the same as the phase of the current in A_0 *i.e.*, that the current in one antenna of

[1] E. Bruce and A. C. Beck, Experiments with Directivity Steering for Fading Reduction, *Proc. I.R.E.*, Vol. 23, p. 357, April, 1935.

[2] Irving Wolff, Determination of the Radiating System Which Will Produce a Specified Directional Characteristic, *Proc. I.R.E.*, Vol. 25, p. 630, May, 1937.

the pair has the phase $+\phi_n/2$ and in the other has $-\phi_n/2$. Equation (52) is a Fourier series, each term of which determines the magnitude and phase of the currents in one of the pairs of radiators of which the antenna system is composed. It is obvious that by suitable choice of the coefficients A_n and ϕ_n, it is possible to realize any desired field distribution as a function of the variable w, from $w = \pi$ through $w = 0$ to $w = -\pi$.

In applying this principle, the desired field distribution is expressed as a function of w from $\beta = +\pi$ to $\beta = -\pi$. Now the transformation from the variable β to the variable w depends upon the choice of d_1/λ, *i.e.*, upon the spacing between the first pair of antennas. Since any value of d_1/λ less than 1 is permissible, there are in general an infinite number of solutions to the problem. In the special case[1] where the desired field intensity at $\beta = \pi/2$ is the same as the desired intensity at $\beta = -\pi/2$, it is permissible to assume that $d_1 = \lambda$. If this is done, the value of w is specified over the range extending on either side of $w = 0$, to $w = +\pi d_1/\lambda$ and to $w = -\pi d_1/\lambda$. For the more general case, where it is necessary that $d_1 < \lambda$, specification of the field from $\beta = +\pi$ to $\beta = -\pi$ leaves regions from $w = \pi d_1/\lambda$ to $w = \pi$ and also from $w = -\pi d_1/\lambda$ to $w = -\pi$ that are not specified by the assumed field distribution. The curve of w may be completed to π and to $-\pi$ by assuming any convenient amplitude of field in these intervals, provided only that the assumption satisfies Dirichlet's conditions in these regions. In general, the curve of field strength should be continued through these gaps to $w = +\pi$ and to $w = -\pi$ in such a way that the complete resulting curve, when plotted as a function of w, can be represented by the simplest possible Fourier series.

The final step in the specification of the antenna system is to determine the coefficients in the Fourier series of Eq. (52) that will give the desired field-strength pattern, expressed as a function of w with the arbitrary extensions as explained above. This specifies the relative amplitudes A_n of the currents required in the successive pairs of antennas and phase differences ϕ_n required between currents of the two antennas of each pair, for the case where the basic spacing d_1/λ of the pairs of antennas has the value used in defining w. The mean phase of each pair of antennas should all be the same, and should be the phase of the central antenna, as indicated above.

The procedure outlined above is applicable directly to the case where the individual radiators have a circular pattern in the plane in which the directivity of the array is specified. The procedure can also be readily extended to cover the case where the individual antennas have directivity in this plane. This is done by dividing the directional pattern desired, by the directional characteristic of the individual radiator, to obtain a desired directional characteristic for the assumption that the radiators have a circular pattern. The relative amplitudes and phases of the currents in the antenna pairs used to build up this desired characteristic are then determined on the basis that each antenna has circular directivity.

ANTENNAS WITH HORN AND PARABOLIC RADIATORS

25. Horn Radiators. *General Properties.*—In the horn radiator, directivity is obtained by using a horn to guide and concentrate the radiated wave into a sharply defined pattern. Examples of horn radiators are shown in Figs. 64, 74, 75, and 77. Such radiators are capable of giving highly directional patterns when the mouth has a dimension large compared with the wave length. Horns become practical at very high frequencies, since then a dimension large compared with the wave length does not involve excessive physical size. Since horn radiators do not involve resonant elements, they can be operated over a wide frequency band.

[1] There are other special cases in which it is permissible to make d_1 greater than λ, although d_1 must always be less than 2λ. See Wolf, *loc. cit.*, for an example.

The field pattern obtained by exciting a horn radiator is determined by the shape and dimensions of the mouth measured in wave lengths and the variation in the magnitude and phase of the field distribution produced across the mouth of the horn. The action of a horn is, in fact, exactly as though the mouth of the horn was replaced by a distribution of elementary dipoles, as illustrated in Fig. 64, in which each dipole radiates a field[1] proportional to $E(1 + \cos \theta)$, where θ is the angle measured with respect to the direction in which the horn points, and E is the intensity of the electric field where the dipole is assumed to exist. The dipoles are polarized in the same direction as the electric field at the mouth of the horn. The actual field pattern of the radiation from the horn is the summation of the radiated fields produced by all these elementary dipoles as obtained by integration.[2]

The nature of the field distribution across the mouth opening is determined by the horn's shape and by the mode of excitation. The relative phases of the field at differ-

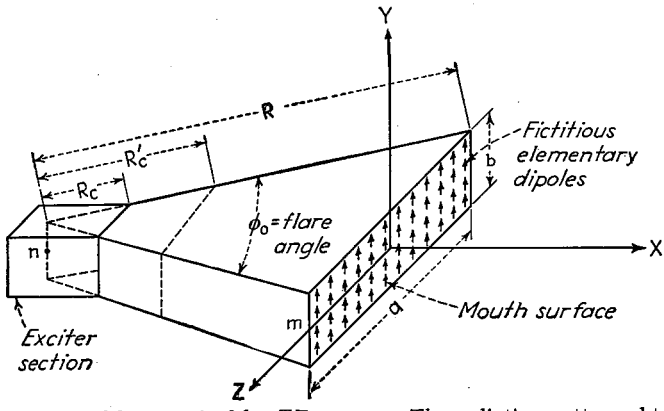


FIG. 64.—Sectoral horn excited by $TE_{0,1}$ wave. The radiation pattern obtained from the horn can be considered as arising from an array of fictitious elementary dipoles arranged across the mouth of the horn as illustrated and having a strength proportional to $E(1 + \cos \theta)$, where E is the intensity of the electric field at the mouth of the horn and θ is the angle with respect to XY plane.

ent parts of the mouth are determined by the relative distances of different parts of the opening from the throat of the horn and are also influenced by the mode of excitation when higher order modes are involved.

When a horn radiator is used as a receiving antenna, the mouth gathers in a section of the passing wave. If this wave is traveling directly toward the mouth of the horn, then the horn in effect cuts out a slice of the wave having a cross section approximating that of the horn mouth, and causes the energy in this slice to travel down the horn toward the throat. If the oncoming wave approaches from an angle, the effective cross-section of mouth that this wave sees as it approaches the horn is reduced, and the amount of energy that can be abstracted from the wave is likewise less. The end result is a directional pattern that is exactly the same as that obtained when the horn is used for transmitting. A horn antenna will, in general, abstract considerably more energy from a radio wave approaching from the favored direction than can be abstracted from the same radio wave by a dipole. In the case where the

[1] This corresponds to a cardioid pattern, with maximum radiation forward and zero radiation backward into the horn. Such a result comes about when the effect of both the electric and magnetic fields across the horn mouth are taken into account.

[2] The theory of horns, with particular reference to sectoral horns, is presented in detail by W. L. Barrow and L. J. Chu, Theory of the Electromagnetic Horn, *Proc. I.R.E.*, Vol. 27, p. 51, January, 1939.

mouth of the horn is in a plane, as in Fig. 64, then if the horn is oriented in the direction for interception of maximum energy from the passing wave, one has approximately

$$\frac{\text{Energy abstracted by horn}}{\text{Energy abstracted by doublet dipole}} = 8.4\,\frac{A}{\lambda^2} \tag{54}$$

where A/λ^2 is the mouth area in square wave lengths. This equation becomes inaccurate when the smaller dimension involved in defining the mouth opening (the width b in Fig. 64) becomes comparable with or less than a quarter wave length.

Sectoral Horn.[1]—The sectoral horn is illustrated in Fig. 64. It has a rectangular cross section, and is flared linearly in one dimension only. Such a horn when used for transmitting produces a fan-shaped beam that is very sharp in the plane in which the horn is flared (the XZ plane of Fig. 64) and a less sharp beam in the corresponding plane at right angles (the XY plane in Fig. 64).

Coupling to a sectoral horn may be either by means of a wave guide or from an antenna. Wave-guide coupling is obtained by terminating the throat in a rectangular

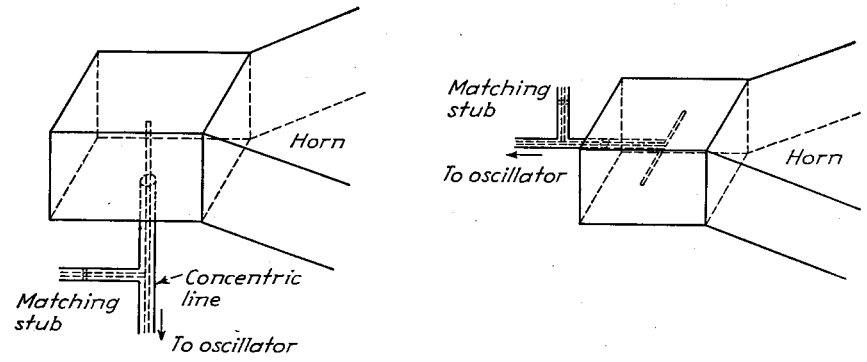


FIG. 65.—Antenna excitation of sectoral horn, together with impedance-matching arrangement to provide a nonresonant concentric transmission-line input.

wave guide, as illustrated in Fig. 64. The horn under these conditions can be thought of as a rectangular wave guide flared in one dimension. Antenna excitation is obtained by terminating the throat of the horn in a short section of wave guide that forms a box, and then exciting this box by means of a probe or a doublet antenna, as illustrated in Fig. 65. In such arrangements, the impedance offered by the antenna system will depend upon the physical size of the antenna and upon the distance of the antenna from the rear of the exciting box.

A sectoral horn operated as a transmitting antenna has the possibility of being excited in several modes. The most important of these and the only ones used in practical operation of sectoral horn radiators correspond to $TE_{0,1}$ and $TE_{1,0}$ waves traveling down the wave guide used for coupling. The field distributions inside the horn corresponding to these two basic modes of excitation are illustrated in Fig. 66. It will be noted that in the $TE_{0,1}$ mode, the electric field is constant in the vertical plane (plane XY of Fig. 64), while in the horizontal plane (plane XZ of Fig. 64), the electric field is maximum at the center of the horn and gradually diminishes approxi-

[1] Further information on sectoral horns is given by Barrow and Chu, *op. cit.;* W. L. Barrow and F. D. Lewis, The Sectoral Electromagnetic Horn, *Proc. I.R.E.*, Vol. 27, p. 41, January, 1939; L. J. Chu and W. L. Barrow, Electromagnetic-horn Design, *Trans. A.I.E.E.*, Vol. 58, p. 333, July, 1939; W. L. Barrow and Carl Shulman, Multiunit Electromagnetic Horns, *Proc. I.R.E.*, Vol. 28, p. 130, March, 1940.

mately sinusoidally to zero at the sides. With excitation by the $TE_{1,0}$ wave, the situation is interchanged, the electric field being constant in the horizontal or XZ plane and tapering off sinusoidally in the vertical or XY plane from a maximum in the middle of the horn to zero at the upper and lower sides.

Excitation by the $TE_{0,1}$ mode, free from the distorting effect produced on the pattern by higher order modes, can be obtained by suitable design of the throat. Waves of the $TE_{0,n}$ type suffer attenuation near the apex of a sectoral horn, and the

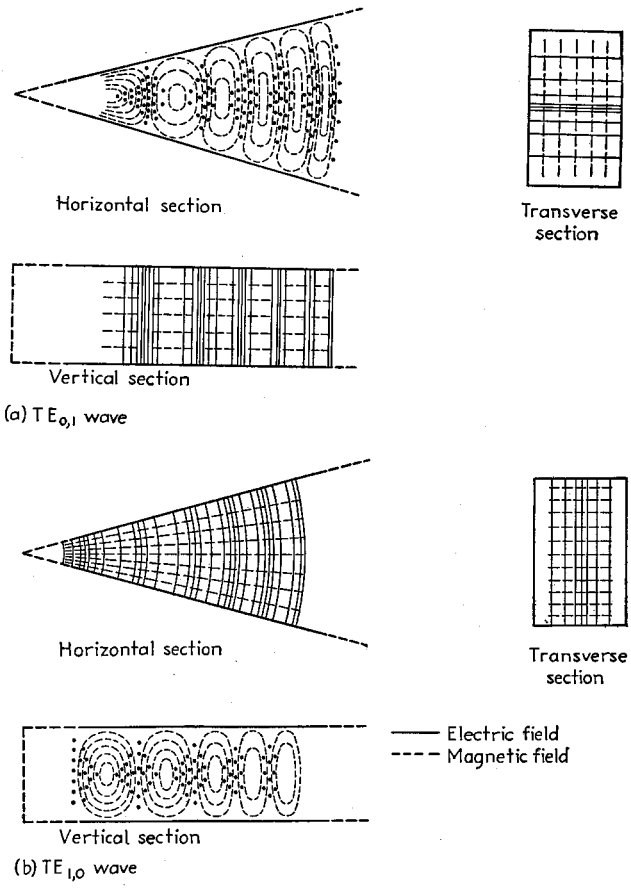


FIG. 66.—Field distributions corresponding to $TE_{0,1}$ and $TE_{1,0}$ waves in the sectoral horn of Fig. 64.

radial distance over which the region of high attenuation extends is greater the higher the order of the wave. Consequently, by designing the throat so that the sectoral horn is cut off a suitable distance R_c from the apex in forming the terminating wave guide or box (see Fig. 64), it is possible to permit the fundamental mode $TE_{0,1}$ to reach the mouth substantially unattenuated, while at the same time providing very high attenuation or filtration of high order modes, such as the $TE_{0,3}$, etc. This optimum cutoff length R_c for $TE_{0,1}$ excitation, and also the extent of the region of high attenuation for the $TE_{0,3}$ mode of excitation, are shown in Fig. 67. Higher

modes of even order can be prevented from appearing by providing symmetrical excitation systems, as well as by proper throat design.

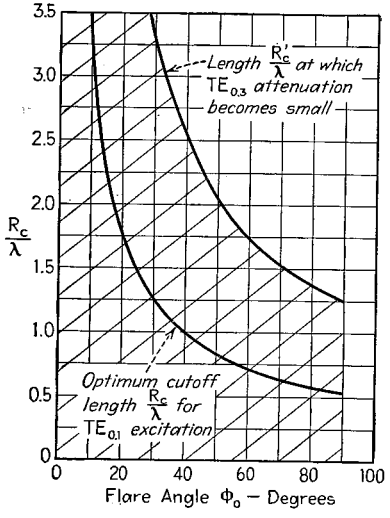


FIG. 67.—Extent of attenuating region near apex of sectoral horn for $TE_{0,1}$ and $TE_{0,3}$ waves.

When radiation by means of the $TE_{1,0}$ wave is desired, higher order modes of the $TE_{n,0}$ type may be eliminated by suitably adjusting the vertical dimension b in Fig. 64. The $TE_{1,0}$ wave is able to travel freely in the radial direction when the dimension b is greater than a half wave length, whereas unattenuated propagation of the third-order mode $TE_{3,0}$ requires that the vertical dimension b exceed $3\lambda/2$, etc. Discrimination against even-order modes, such as $TE_{2,0}$, can be obtained by providing an exciting system that is symmetrical in the horizontal plane. Hence one should have

$$\lambda/2 < b < 3\lambda/2.$$

Discrimination against higher order modes of excitation is particularly important in the case of antenna excitation. This is because antenna arrangements such as those illustrated in Fig. 65 simultaneously generate fields corresponding to several modes, and if the higher order components are not suppressed, the resulting field pattern will be irregular in outline and generally undesirable.

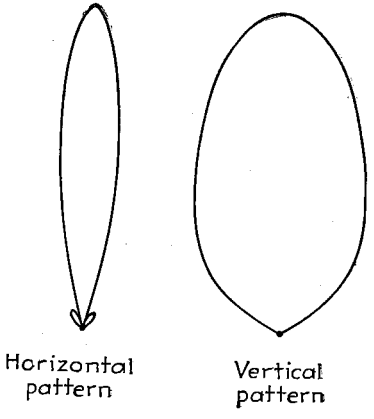


FIG. 68.—Typical field pattern of sectoral horn excited by a $TE_{0,1}$ wave. For the case shown, the flare angle is 30°, the horn length R is 11λ, and the vertical aperture is one wave length.

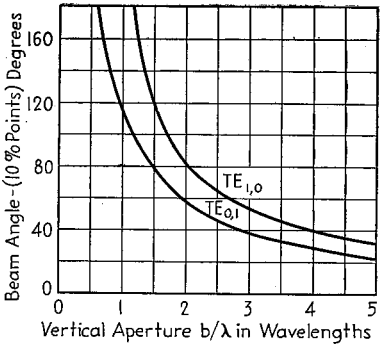


FIG. 69.—Width of beam in vertical plane for sectoral horns of various heights excited by $TE_{0,1}$ and $TE_{1,0}$ waves. The term beam angle is here used to indicate the angle included between the points of the major lobe, where the field intensity is 10 per cent of the maximum.

The directional pattern of a properly designed sectoral horn excited in the $TE_{0,1}$ mode has the general character shown in Fig. 68. The pattern in the vertical plane

is determined only by the vertical aperture (dimension b in Fig. 64), and is independent of the width of the mouth or the flare angle. The vertical pattern has only a single lobe if the vertical aperture does not exceed a wave length; while with greater apertures one or more small minor lobes are present. In the case of vertical apertures of one wave length or greater, the width β_v of the beam in the vertical plane between the two nulls is given by the equation

$$\sin \frac{\beta_v}{2} = \frac{1}{\left(\frac{b}{\lambda}\right)} \tag{55}$$

where b is the vertical aperture measured in wave lengths. This leads to the result shown in Fig. 69.

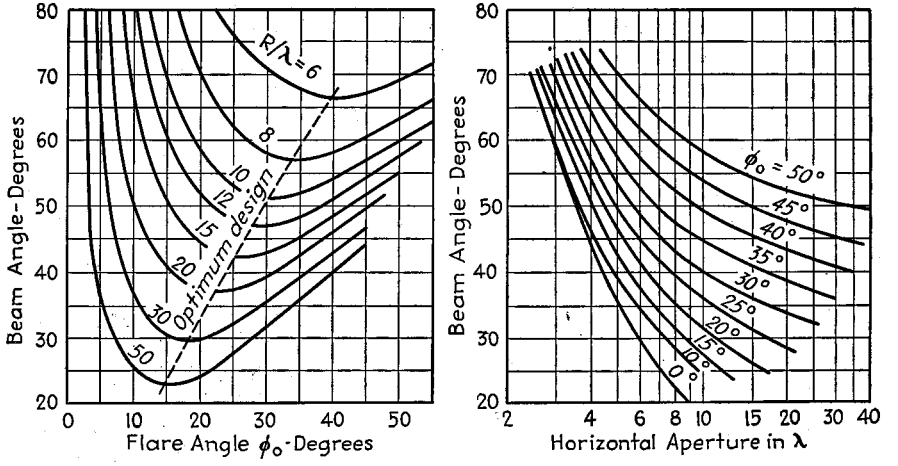


FIG. 70.—Beam angle in horizontal plane for sectoral horn of Fig. 64, excited by $TE_{0,1}$ wave, as a function of flare angle ϕ_0 and horizontal aperture, for various horn lengths R. The beam angle is here defined as the angle between points at which the field strength has dropped to 10 per cent of that at the maximum of the beam. The dotted curve marked "optimum design" gives the proportions for which a beam of given width is obtained with a horn of minimum length.

The wave pattern in the horizontal plane of the sectoral horn of Fig. 64 excited by a $TE_{0,1}$ wave is affected both by the width of the aperture and by the flare angle, but is independent of the height of the aperture in the vertical plane. With constant flare angle, the beam tends to become sharper the greater the width of the aperture; while with an aperture of constant width, the beam becomes wider the greater the flare angle (*i.e.*, the shorter the horn). The quantitative relations are shown in Fig. 70. When the flare angle is very small, the beam in the horizontal plane has its maximum sharpness for an aperture of given width, but the horn is very long. The beam width β_H between nulls when the flare angle approaches zero is

$$\sin \frac{\beta_H}{2} = \frac{1.5}{\left(\frac{a}{\lambda}\right)} \tag{56}$$

where a/λ is the horizontal aperture in wave lengths (see Fig. 64). As the flare angle is increased, the front of the wave traveling down the horn diverges as is apparent from Fig. 66*a*. This causes a broadening of the beam, and if the length of the horn

is very large when measured in wave lengths, the width of the main beam approaches the flare angle, irrespective of the aperture width.[1]

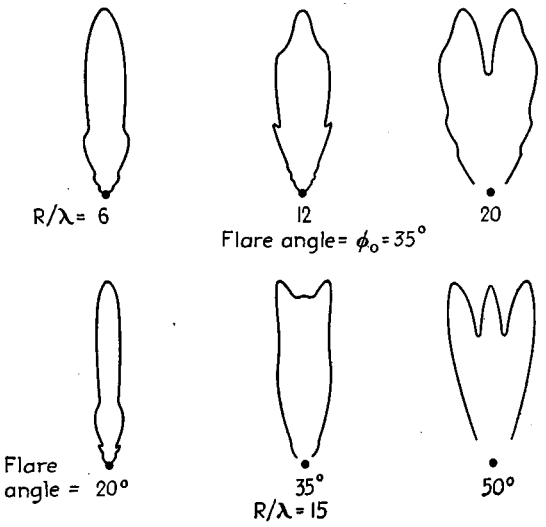


FIG. 71.—Typical series of horizontal radiation patterns for $TE_{1,0}$ wave in sectoral horn of Fig. 64.

For design purposes, one is usually interested in obtaining a beam of a given sharpness using a horn having the shortest possible length R. This condition is shown by the dotted line in Fig. 70 marked "optimum design." It is to be noted that this optimum design corresponds to horn proportions such that the difference in distance from the apex of the horn to the middle of the mouth (to line Y in Fig. 64) and the distance to the edge of the mouth (distance R in Fig. 64) is approximately 0.375λ.

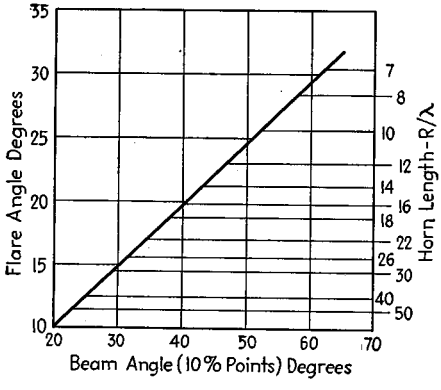


FIG. 72.—Optimum proportions for sectoral horns excited by $TE_{1,0}$ wave, corresponding to the shortest horn giving the indicated beam width. The beam angle in this curve corresponds to the angle between points at which the field strength of the main lobe is 10 per cent of the maximum.

The field pattern of a sectoral horn excited by a $TE_{1,0}$ wave differs in certain details from the pattern obtained with the $TE_{0,1}$ excitation. The width β_v of the beam in the vertical plane is now approximately 50 per cent broader than given by Eq. (55) (see Fig. 69). At the same time, the horizontal pattern is characterized by a major lobe that for very small flare angles is smooth and approximately two-thirds as wide as with $TE_{0,1}$ excitation but that becomes jagged with large flare angles, as illustrated in

[1] The statement that the diverging character of the field within the horn causes the beam to become broader is equivalent to stating that the field at different parts of the mouth of the horn is not in the same phase because different parts of the mouth of the horn are at different distances from the horn apex.

Fig. 71. The proportions corresponding to the shortest horn that will have a specified beam width in the horizontal plane, when excited in the $TE_{1,0}$ mode, are given in Fig. 72.[1] Upon comparison of these proportions with those

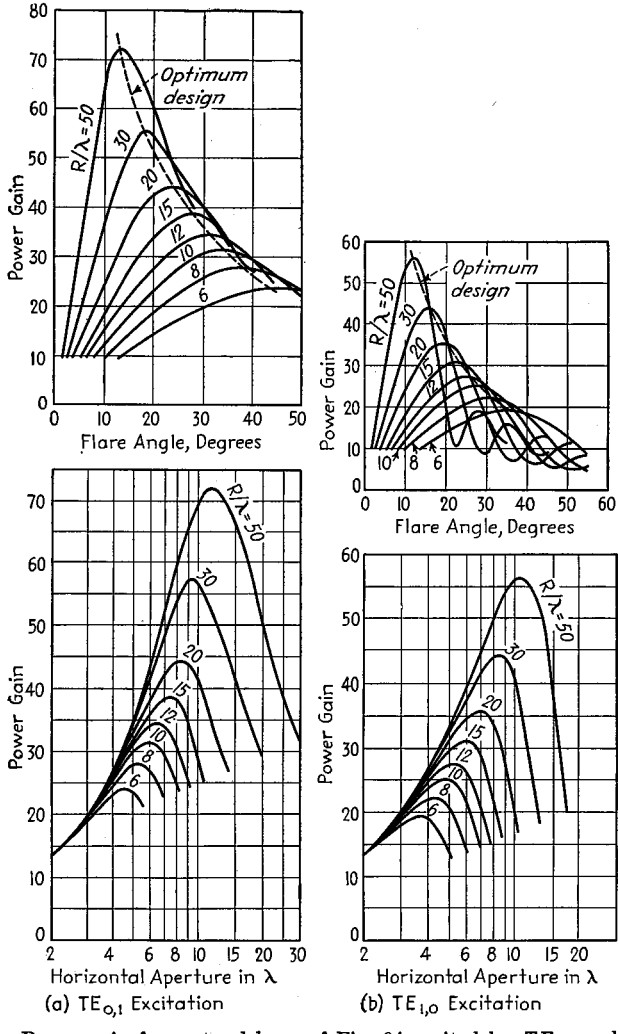


FIG. 73.—Power gain for sectoral horn of Fig. 64 excited by $TE_{0,1}$ and $TE_{1,0}$ waves, showing the effect of flare angle and horizontal aperture. These curves are for a vertical aperture of one wave length and give the gain compared with a doublet antenna. For other vertical apertures, the power gain given by the curves should be multiplied by the vertical aperture measured in wave lengths.

represented by the dotted line marked "optimum design" in Fig. 70, it is noted that

[1] These optimum proportions correspond to a condition such that the distance from the point n on the middle of the apex (see Fig. 64) to the point m on the side of the mouth is approximately a quarter of a wave length greater than the distance from n to the center of the mouth where the X, Y, and Z planes intersect in Fig. 64. For this optimum condition, the irregularity in the directional pattern is only moderate, as can be found from an examination of Fig. 71.

$TE_{1,0}$ excitation in a given horn will result in a slightly sharper beam in the horizontal plane than $TE_{0,1}$ excitation. At the same time, the pattern in the horizontal plane tends to be irregular with the $TE_{1,0}$ mode, and the pattern in the vertical plane is broader, so that in general, $TE_{0,1}$ excitation is preferable.

The power gain of a sectoral horn is high if the horn aperture is at all large. Curves giving the power gain using a doublet antenna as the standard of comparison are given in Fig. 73 for the case where the vertical dimension b of the aperture is one wave length. For other vertical apertures, the power gain as given by the curves should be multiplied by the vertical aperture in wave lengths. The dotted curves through the maxima in Figs. 73*a* and 73*b* represent proportions giving the maximum possible power gain for a given horn length, *i.e.*, optimum proportions from the point of view of power gain. These proportions are substantially identical with those giving maximum sharpness of beam in the horizontal plane (see Figs. 70 and 72). This confirms the expected fact that conditions giving the narrowest beam will also give the maximum power gain.

Pyramidal Horns.—A pyramidal horn is illustrated in Fig. 74 and can be thought of as a modified form of sectoral horn that is flared in two dimensions instead of only one. This provides a mouth that can be large in both dimensions, combined with a throat small enough to be coupled to an ordinary wave guide and to attenuate undesired modes of excitation. The factors controlling the two flares of a pyramidal horn are the same as in the simple sectoral horn. Thus when excitation is by a $TE_{0,1}$ wave, the relationship between beam width in the horizontal plane, the horizontal flare angle ϕ_H (flare in XZ plane of Fig. 74), and the length of horn are exactly the same as in the case of a sectoral horn having a type $TE_{0,1}$ excitation. Similarly, the factors determining the width and character of beam and flare angle ϕ_v in the vertical plane (flare in XY plane of Fig. 74) are the same as the factors that would determine width of beam and flare angle in the horizontal plane of a sectoral horn excited by a $TE_{1,0}$ wave; *i.e.*, flare in the vertical plane with a $TE_{0,1}$ wave corresponds to flare in the horizontal plane and a $TE_{1,0}$ wave.

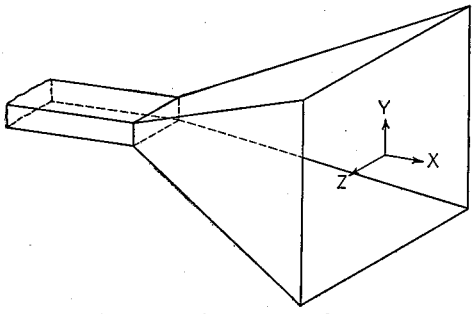


FIG. 74.—A pyramidal horn.

In designing pyramidal horns for $TE_{0,1}$ excitation,[1] it will be noted that the flare angle in the vertical plane with a given length of horn should not exceed the value corresponding to the proportions indicated in Fig. 72. Otherwise the major lobe in the vertical directional pattern becomes ragged.

The power gain of pyramidal horns excited by the $TE_{0,1}$ mode and having small flare in the vertical plane is given accurately by Fig. 73*a*. However, with larger flare angles in the vertical plane, the power gain is reduced by a factor that depends on the horizontal flare angle and the radial length of the horn.

Conical Horns.[2]—A conical horn is a section of a cylindrical cone, the apex of which is terminated in a cylindrical wave guide or cylindrical coupling section, as shown in Fig. 75*a*.

[1] This also covers the $TE_{1,0}$ case, since a pyramidal horn with $TE_{1,0}$ excitation becomes the $TE_{0,1}$ case simply by rotating the horn 90° so that XY is the horizontal plane.

[2] G. C. Southworth and A. P. King, Metal Horns as Directive Receivers of Ultra-short Waves, *Proc. I.R.E.*, Vol. 27, p. 95, February, 1939; A. P. King, Metal Horns as Radiators of Electric Waves, *Bell Lab. Rec.*, Vol. 18, p. 247, April, 1940.

The conical horn is usually excited by a $TE_{1,1}$ wave, although higher modes are possible. The field configurations of the $TE_{1,1}$ wave in a circular horn is similar to that of a $TE_{0,1}$ wave in a rectangular horn. A cross-sectional view of the field configuration is shown in Fig. 75*b*.

The character of the radiation pattern obtained with a conical horn is determined by the diameter of the mouth opening and by the flare angle, or, what is equivalent, by the diameter and length of the horn. The effect of variation in the flare angle and in the length of the horn is given in Fig. 76. The field patterns in the vertical and horizontal planes differ slightly as a result of the different orientation of the electric and magnetic fields in the horn with respect to these planes. Examination of Fig. 76 shows that for a given horn length L there is an optimum flare angle, and likewise for a given flare angle, there is an optimum length. If the flare angle is too great or the horn length excessive, the pattern becomes jagged in its outline. On the other hand, if the flare angle is small or the horn is short, the pattern becomes quite broad. The optimum proportions correspond to

$$L = \frac{0.3\lambda \cos \frac{\psi}{2}}{\left(1 - \cos \frac{\psi}{2}\right)} \qquad (57a)$$

or the equivalent

$$\cos \frac{\psi}{2} = \frac{1}{1 + 0.3 \frac{\lambda}{L}} \qquad (57b)$$

From the geometry for Fig. 76, one also has

$$D = 2L \sin \frac{\psi}{2} \qquad (58)$$

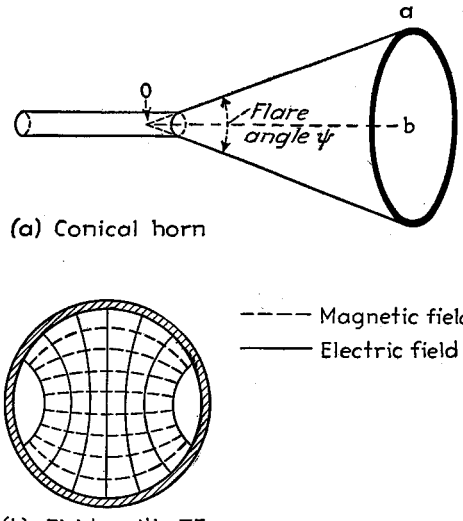


FIG. 75.—Conical horn and field distribution within horn for $TE_{1,1}$ wave.

The proportions represented by Eqs. (57) correspond to the conditions where the difference in distances *oa* and *ob* ($=L$) in Fig. 75*a* from horn apex to horn edge and center, respectively, is 0.3λ. Conical horns designed in accordance with Eqs. (57) will give the sharpest beam that can be obtained for a horn of given length.

Biconical Horns.[1]—A biconical horn consists of a figure of revolution having a cross section such as is shown in Fig. 77*a*. This can be thought of as the sectoral horn of Fig. 64 rotated in the XY plane about point n. When excited in a fundamental mode, such a horn is characterized by a pancake type of radiation pattern in which the radiation is concentrated in a sharp beam in the vertical plane, as shown in Fig. 77*d*, but is uniformly distributed in the horizontal plane.

The fundamental modes that may be used to excite a biconical horn are the TEM and $TE_{0,1}$ waves. These involve field distributions as shown in Figs. 77*b* and 77*c*, respectively. The TEM wave is analogous to the $TE_{1,0}$ mode of the sectoral horn of Fig. 64, and the same rules regarding optimum flare angle apply in the two cases.

[1] W. L. Barrow, L. J. Chu, and J. J. Jansen, Biconical Electromagnetic Horns, *Proc. I.R.E.*, Vol. 27, p. 769, December, 1939.

The $TE_{0,1}$ mode in the biconical horn corresponds to the wave of the same designation in the sectoral horn of Fig. 64, and again the relationship between sharpness of beam, flare angle, horn length, etc., given for the sectoral horn applies to the corresponding biconical case. When the axis of revolution of the biconical horn is vertical, it will be noted that the *TEM* mode results in the radiation of vertically polarized waves, whereas the $TE_{0,1}$ mode radiates horizontally polarized waves.

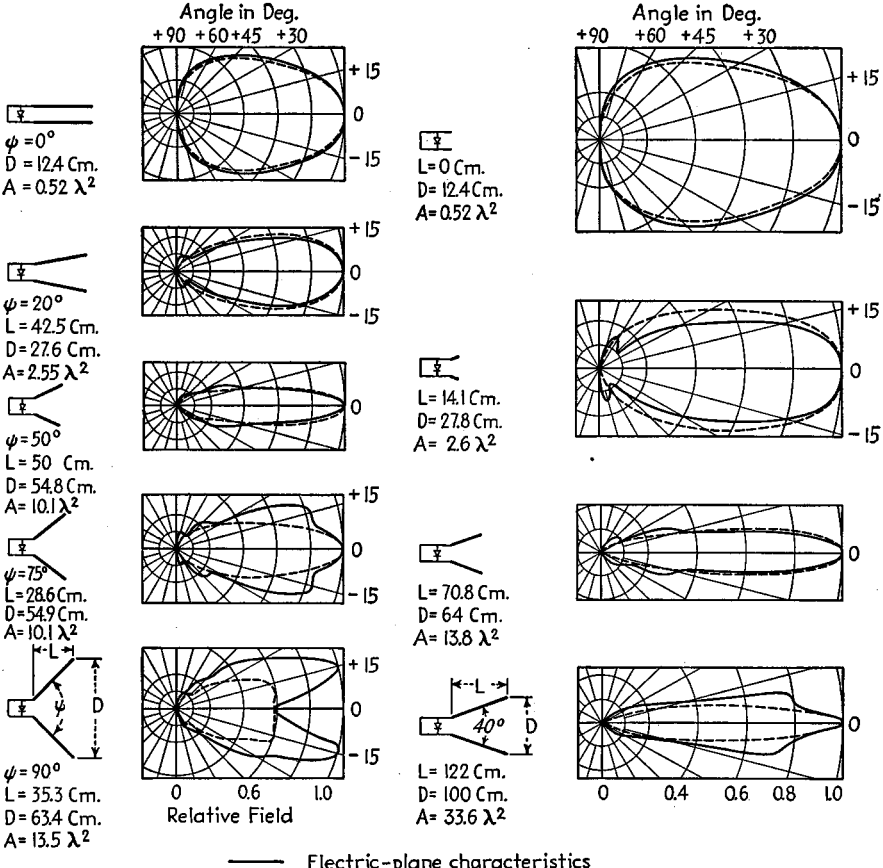


FIG. 76.—Directional properties of conical horns of roughly the same length but of various angular openings and of the same angular openings but of different lengths. The wave length for all cases is 15.3 cm.

The relationship between power gain, flare angle, and radial length of horn in the case of biconical horns follows the same general rules that apply to the sectoral horn. The quantitative relations existing are given in Fig. 78. The dotted curve marked "optimum design" corresponds to the horn giving the maximum power gain (*i.e.*, sharpest beam) for a given length, and is seen to correspond very closely to the optimum design for the corresponding sectoral horn excited with the corresponding mode.

The *TEM* mode in a biconical horn is most readily produced by providing a space between the cones, either by separating the apexes slightly, as in Fig. 79*a*, or by cutting

away a section of the apex, as in Fig. 77*a*. A short antenna carrying a uniform current distribution is then placed in the resulting space coincident with the axis of revolution (see Fig. 77*a*). The $TE_{0,1}$ wave may be produced by a small loop fed by a two-wire balanced line located between the cones, as shown in Fig. 79*a*. In this arrangement,

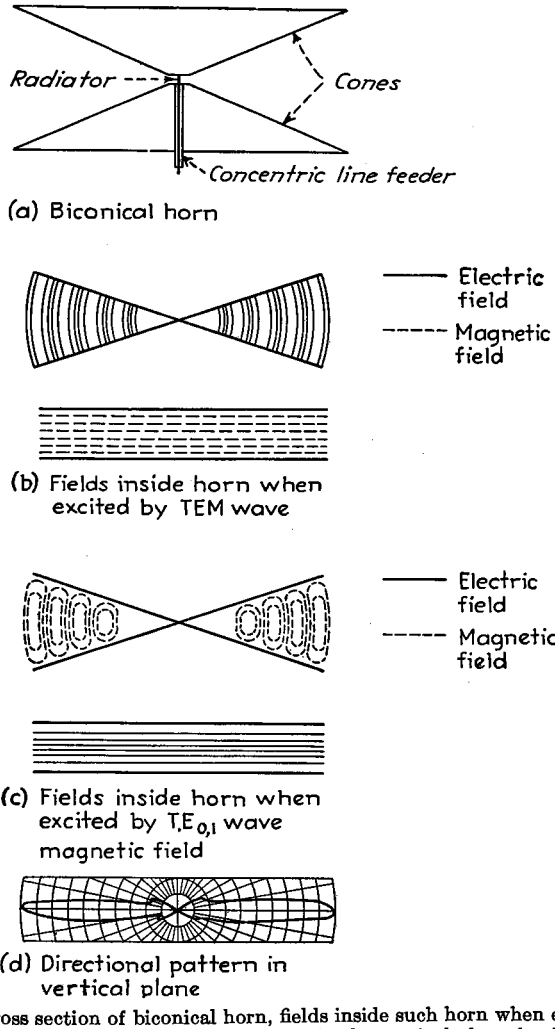


FIG. 77.—Cross section of biconical horn, fields inside such horn when excited by *TEM* and $TE_{0,1}$ waves, and typical directional pattern in the vertical plane for *TEM* excitation. In the representation of fields, the dotted lines indicate magnetic flux and the solid lines electric field, and the upper drawing for each case is a cross section of the horn in a plane containing the axis of revolution, while the lower diagram shows the fields that are seen when one looks toward the apex of the cone.

the length of the loop must be appreciably less than a half wave length in order to maintain substantially uniform current distribution around the loop. As a result, the radiation resistance is quite low. A higher radiation resistance can be obtained, as illustrated in Fig. 79*b*, where a number of antennas are bent in circular segments and

placed about a circular center section. These antennas are fed in such a manner that the currents flow in the same angular direction in all the antennas, and the diameter of the circle formed by the antennas is preferably an odd multiple of a half wave length. The excitation of *TEM* waves can be avoided by taking care to provide a balanced feed arrangement.

One of the critical features of biconical horn design is the spacing between the cone vertexes. There should be enough spacing so that desired mode is propagated freely from the apex outward with negligible attenuation. Also, the greater the spacing the easier it is to obtain a high radiation resistance in the exciting antenna. At the same time, the spacing should be small enough to provide attenuation of higher order

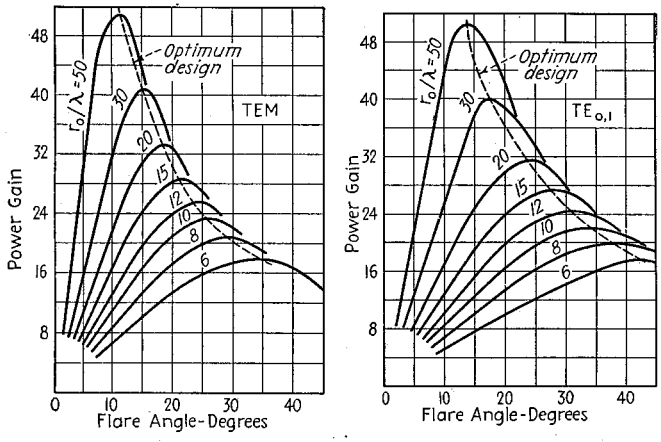


FIG. 78.—Power gain of biconical horns relative to doublet antenna. The dotted curves marked "optimum design" correspond to the condition of maximum power gain for a given radial length. In these curves, r_0/λ is the radial length of the horn, measured in wave lengths.

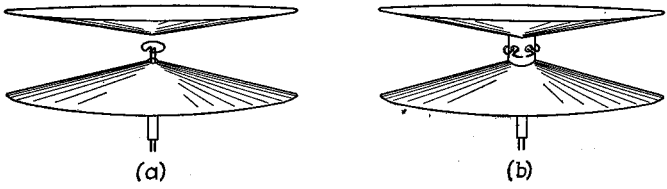


FIG. 79.—Methods of exciting $TE_{0,1}$ waves in a biconical horn.

modes. When these factors are balanced, the result is that it is generally desirable to keep the spacing considerably less than a half wave length.

Pipe Radiators.[1]—A pipe radiator is formed by a wave guide that is allowed to remain open at one end, and therefore radiates energy. The simplest radiators of this type are circular wave guides with a $TE_{1,1}$ wave and rectangular wave guides with either a $TE_{1,0}$ or a $TE_{0,1}$ wave. Both these cases may be considered as special cases of conical and sectoral horns, respectively, in which the flare angles have become zero. Under such conditions, the directivity in any plane depends upon the dimensions of the mouth in that plane, with the width of the major lobe being almost exactly inversely proportional to the corresponding dimension measured in wave lengths.

[1] For further discussion, see W. L. Barrow and F. M. Greene, Rectangular Hollow-pipe Radiators, *Proc. I.R.E.*, Vol. 26, p. 1498, December, 1938; Southworth and King, *op. cit.*

The practical usefulness of pipe radiators is limited by the fact that if the wave guide is large enough to give a reasonably sharp beam, then it is so large that it does not attenuate higher order modes. Under such circumstances, there is considerable practical difficulty in obtaining action corresponding to a single mode of excitation.

26. Antennas with Parabolic Reflectors.[1]—A parabola can be used to concentrate the radiation from an antenna located at the focus, to form a beam in the same way that a searchlight reflector produces a light beam. The action of a parabola is to convert a spherical wave originating at the focus into a plane wave at the mouth of the parabola, as shown in Fig. 80. The end result obtained with a parabolic antenna is accordingly very similar to that obtained by using a horn, the chief difference being in the way in which the field distribution across the mouth is established.

There are, however, several differences in detail between the parabolic and the horn radiator. In the parabola, the wave across the mouth is a plane wave, whereas in the case of a horn, the wave front is divergent because of the flare of the horn. The characteristics of the parabola are also such that the fields established across the mouth by the parabola are everywhere in the same phase, whereas with the horn radiators, there is a variation in the phase because of the differences in distance from the apex to various parts of the horn mouth. The distribution of field intensity across the mouth of the horn will generally be different from that of the parabola because of the different way in which the excitation is obtained. Finally, in the parabolic antenna, the portion of the radiation from the antenna that is directed away from the parabolic surface has a diverging wave front, and is lost as far as the main beam is concerned. This not only wastes power but also broadens the beam by creating additional radiation to the side.

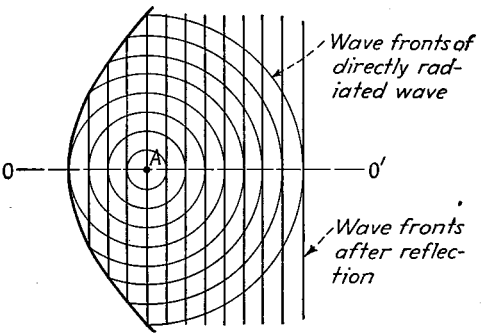


FIG. 80.—Conversion of spherical waves to plane waves by a parabolic reflector. The circles represent the spherical waves radiated from the antenna A; while the vertical lines represent the wave front after reflection from the parabolic surface.

This direct radiation can be eliminated by exciting the parabola with a directional antenna system such that the radiation is all directed toward the parabolic surface. One way that can be used to approximate this result is to employ an array of parasitic antennas such as those illustrated in Fig. 47. Another possibility is to surround the exciting antenna by a hemispherical shield, as shown dotted in Fig. 83. The wave directed toward such a shield is reflected back at the antenna in the proper phase to add to the waves leaving the antenna directly for the parabolic surface. The radius of such a shield should be an exact multiple of a half wave length. The shield should not be too small, for in that case there is appreciable mutual coupling between the

[1] There is almost no information published in English on parabolic antennas. The reader who wishes to follow the subject further is referred to the following German and French articles, which have formed the basis of most of the material in this section: René Darbord, Reflecteurs et Lignes de Transmission pour Ondes Ultra Courtes, *L'Onde électrique*, Vol. 11, p. 54, 1932; Walther Kohler, Die Wirkungsweise von Vollmetal und Gitterreflektoren bei ultra kurzen Wellen, *Hochfrequenz Technik und elektro Akustic*, Vol. 39, p. 207, 1932; R. Brendel, Beitrag zur Berechung von Reflectoren fur elektrische Wellen, *Hochfrequenz Technik und elektro Akustic*, Vol. 48, p. 14, 1936; R. Bromel, The Beam Properties of Small Parabolic Reflectors with Various Excitations, *Hochfrequenz Technik und elektro Akustic*, Vol. 48, p. 81, 121, 1936; C.I.H.A. Staal, Full Parabolic Reflectors for Microwaves, *Philips Transmitting News*, Vol. 3, p. 14, 1937, *Hochfrequenz Technik und elektro Akustic*, Vol. 50, p. 206, 1937; F. Ollendorf, "Die Grundlagen der Hochfrequenztechnik."

dipole and the shield that distorts the action. At the same time, the shield should not be so large as to intercept a considerable portion of the main beam that is reflected from the parabola. The final choice of shield radius will depend largely on the size of the parabola.

Paraboloid of Revolution (Rotational Parabola).—The paraboloid of revolution has a parabolic cross section, with the axis of revolution coinciding with the line oo' shown in Fig. 80. The excitation of such an antenna system is obtained by providing a source A of radio waves at the focus of the parabola. A rotational parabola is approximately equivalent to a conical horn antenna system.

The radiation pattern of a rotational parabola is characterized by a sharply defined main lobe directed along the axis of revolution and surrounded by a succession of minor lobes of relatively small and rapidly decreasing intensity, as shown in Fig. 81. The width of the main lobe between nulls, assuming that the radiation from the exciting antenna is nondirectional, is given approximately by the equation

$$\left.\begin{array}{l}\text{Beam width between}\\ \text{nulls, in degrees}\end{array}\right\} = \frac{137.5}{\left(\frac{D}{\lambda}\right)} \qquad (59)$$

where D/λ is the mouth diameter in wave lengths.

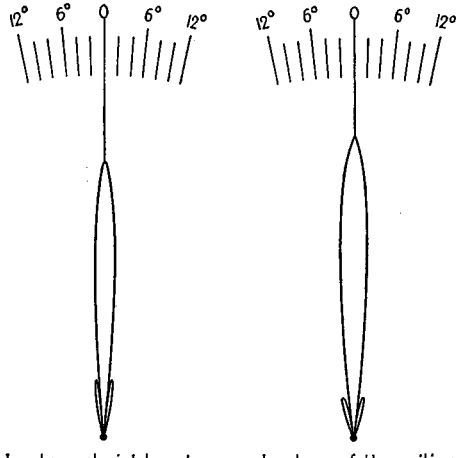


FIG. 81.—Directional pattern of rotational parabola excited by doublet antenna having a mouth diameter of 12λ and a focal length of 0.625 times the mouth radius. The amplitudes of the minor lobes are exaggerated for clarity.

In practice, the main lobe is not circular in cross section, because the source of the exciting radio waves will have some directivity. For example, when the excitation is from a vertical half-wave antenna located at A in Fig. 80, the radiation is not uniform in the vertical plane, even though it is in the horizontal plane. As a consequence, the parabola acts as though it had a mouth diameter in the vertical direction less than the actual diameter. Under such conditions, the beam is broader in the vertical plane than given by Eq. (59), although this equation still gives the horizontal width. In the case of a vertical doublet radiator (field proportional to the cosine of the angle of elevation), the beam width in the vertical is approximately 1.25 times that for the horizontal plane. The "spot" produced by the beam is then elliptical, with the ratio of axes approximately 1.25 to 1.

A parabola with a given mouth diameter may be designed for a large or small focal distance, as illustrated in Fig. 82. The optimum focal distance for any particular case is arrived at after considering several factors. In the first place, the smaller the focal distance the smaller will be the fraction of the total radiation that is not intercepted by the parabolic surface (*i.e.*, the smaller the energy lost by direct radiation). However, a focal distance so short that the focus is located inside the mouth of the parabola (Fig. 82*a*) results in the field reflected from some parts of the parabola being out of phase with the contributions of other parts. For example, when excitation is obtained as in Fig. 83 from a vertical antenna, the rays marked *a* and *b* will have opposite phase after reflection by the parabola and so will tend to cancel. In

the case of vertical antenna excitation, the focus is preferably approximately in the plane of the mouth of the antenna, as illustrated in Fig. 82*b*, corresponding to a focal distance equal to one-half the radius of the mouth. When the antenna A has directivity of such a character as to concentrate a fairly sharp beam along the axis toward the parabolic focus, then the surface is preferably designed so that its focus lies outside

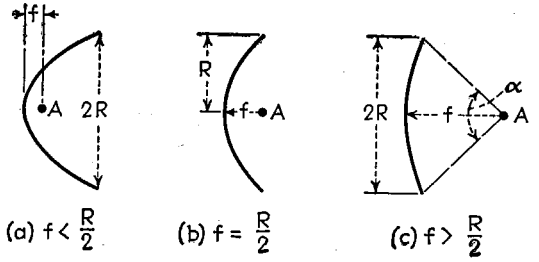


FIG. 82.—Rotational parabolas with same mouth diameter but different focal lengths.

the parabola, as shown in Fig. 82*c*. In this case, the focus is preferably so chosen that the angle α approximates the width between the nulls of the antenna beam.

The power gain of a rotational parabola excited by a doublet antenna (radiated field proportional to the cosine of the angle of elevation) is[1]

$$\left.\begin{matrix}\text{Power gain relative}\\ \text{to doublet antenna}\end{matrix}\right\} = \left(\frac{2\pi}{\lambda}\,\frac{2fR^2}{4f^2+R^2}\right)^2 \tag{60}$$

where f is the focal distance, λ the wave length, and R the radius of the mouth.

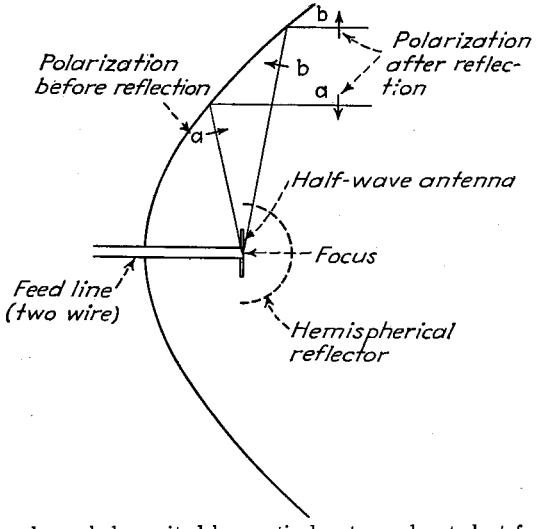


FIG. 83.—Rotational parabola excited by vertical antenna located at focus that is inside the parabola.

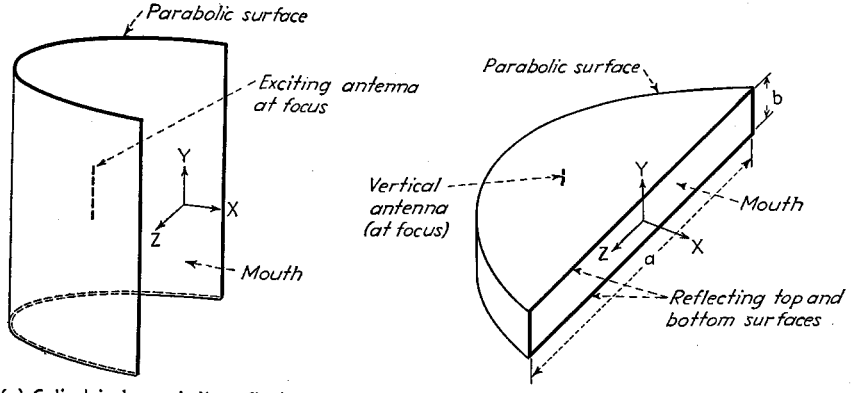
This power gain is maximum when the focal distance is equal to half the radius of the mouth (the case illustrated by Fig. 82*b*), for which case it becomes

[1] See Darbord, *loc. cit.*

$$\left.\begin{array}{l}\text{Maximum power gain}\\ \text{using doublet}\\ \text{radiator, relative}\\ \text{to doublet antenna}\end{array}\right\} = \left(\pi \frac{R}{\lambda}\right)^2 \tag{61}$$

In the event that a spherical shield or other reflector is provided that prevents all loss from direct radiation, then the gain under optimum conditions will be twice that given by Eq. (61).

The paraboloid of revolution may be considered an alternative to a conical horn, in that both produce approximately the same type of directional pattern and both have approximately the same gain for the same mouth opening, provided that the flare angle of the horn is small. When the two are compared, it is found, however, that if the optimum flare angle is used for the conical horn in order to keep the length of the horn to a reasonable value, then the horn has a somewhat broader pattern for a



(a) Cylindrical parabolic reflector with open top and bottom

(b) Cylindrical parabolic reflector with closed top and bottom

FIG. 84.—Cylindrical parabolas with open and closed ends.

given mouth diameter than the rotational parabola, and at the same time the parabola has less depth and is therefore more manageable physically.

Cylindrical Parabolic Reflectors.—Cylindrical parabolic reflectors have a parabolic curvature in one plane, usually called the horizontal plane, and no curvature in any plane perpendicular to this horizontal plane. The top and bottom of the reflector may be open, as in Fig. 84*a*, or closed by plane reflecting surfaces, as in Fig. 84*b*.

Cylindrical parabolas are normally excited by an antenna parallel to the cylindrical surface, located at the axis of the parabola. The parabola is preferably so designed that the focus lies well within the mouth, so that the fraction of the radiated energy that fails to be intercepted by the parabolic surface will be relatively small.

The cylindrical parabola with closed top and bottom is the approximate equivalent of a sectoral horn of corresponding mouth dimensions.[1] The directional pattern in the horizontal plane is the same as that of a sectoral horn of zero flare angle excited in the $TE_{1,0}$ mode. The directional pattern in the vertical plane corresponds to the vertical pattern of the corresponding sectoral horn excited by the $TE_{0,1}$ mode. A cylindrical parabola with closed top and bottom has less axial length for a given mouth opening than the corresponding sectoral horn, and also gives a sharper pattern in

[1] See W. L. Barrow and F. D. Lewis, The Sectoral Electromagnetic Horn, *Proc. I.R.E.*, Vol. 27, p. 41, January, 1939.

either the horizontal or vertical plane while retaining the same width of pattern in the other plane.

A cylindrical parabola with open ends has the same directivity in the horizontal plane as when the top and bottom are closed, but the directivity in the vertical plane is less than with closed top and bottom.

PRACTICAL TRANSMITTING ANTENNAS

27. Transmitting Antennas for Frequencies below 500 Kc.—Transmitting antennas for the lower radio frequencies are always of the grounded vertical-wire type, with a flat top that is usually arranged in the form of an inverted L, or a T, as in Figs. 3*e* and 3*f*, respectively.

The flat top is preferably large enough so that its capacity to ground is considerably greater than the capacity between the vertical wire and ground. In this way, the effective height of the transmitting antenna is only slightly less than the actual tower height.[1]

When steel towers are employed, there should be sufficient spacing between the tower and the vertical down-lead and flat top, so that the capacity from tower to these antenna elements is quite small. Otherwise unnecessary circulating capacity currents will be present that will result in increased loss of energy. There is some advantage in insulating the tower from ground or in using a wooden tower.

Guy wires have much the same effect as the tower, and either should be kept as far as possible from all parts of the antenna or, alternatively, should be broken up into small sections with insulators. The location of these insulators is preferably such that the guy wires with insulators in place produce a minimum of disturbance in the electrostatic field that would exist in the same region if the guy wires were entirely absent.[2]

Within the limitation that the flat-top capacity should be considerably greater than the capacity of the vertical down-lead, the required capacity is, in general, determined by the fact that the current I going into the flat top (*i.e.*, the current in the vertical lead) is related to the voltage E, applied to the flat top, and the capacity C of top to ground, according to the relation

$$E = \frac{I}{\omega C} \tag{62}$$

It is apparent that the flat-top capacity required to avoid excessive voltage and danger of corona is greater the lower the frequency and the higher the transmitter power.

Particular attention must be paid to the design of the antenna insulators when the voltage is high. The common types consist of relatively long rods or tubes of insulating material with properly designed end terminals.[3] In some cases shielding rings are also used to distribute the voltage stress.

In long-wave antenna systems for transoceanic service, where the investment is extremely large and continuous operation is essential, it is necessary in most regions to provide means for removing sleet from the antenna system. This is accomplished by circuit arrangements in which 60-cycle current can be circulated for the purpose of keeping the wires warm without at the same time interfering with the normal antenna action.[4] It is also desirable in such antenna systems to provide means of releasing

[1] The effective height of a grounded vertical-wire antenna is the height that a vertical wire would be required to have to radiate the same field along the horizontal as is actually present if the wire carries a current that is constant along its entire length and of the same value as at the base of the actual antenna. The effective height of a quarter-wave vertical wire is $2/\pi$ times the actual height.

[2] H. P. Miller. Jr., The Insulation of a Guyed Mast, *Proc. I.R.E.*, Vol. 15, p. 225, March, 1927.

[3] The design of such insulators is discussed by W. W. Brown, Radio Frequency Tests on Antenna Insulators, *Proc. I.R.E.*, Vol. 11, p. 495, October, 1923.

[4] J. H. Shannon, Sleet Removal from Antennas, *Proc. I.R.E.*, Vol. 14, p. 181, April, 1926.

any antenna wire that becomes overloaded, either by the use of a weak link or by means of a spring trip. Such arrangements will prevent the loss of towers either from overloading or from unbalanced pull resulting from a failure of only a single part of the antenna system.

Loss Resistance—Ground Systems.[1]—The radiation resistance of grounded vertical-wire antennas with flat top, as given by Eq. (26), tends to be low with reasonable antenna heights, particularly at the lower frequencies. As a result, it is necessary that unusual effort be made to minimize the loss resistance associated with low frequency transmitting antennas. The most important of these losses occur in the ground, in the antenna tuning inductance, in the insulation, and in the conductors.

Losses in the tuning inductance are a coil-design problem, and it is found desirable to employ coils that are physically large, carefully designed, having litz wire. The conductor resistance can be minimized by the use of adequate size and number of antenna wires. In order that insulation losses may be low, it is absolutely essential that there be no corona whatsoever at any place in the system. When high powers are involved, this requires use of suitable corona shields on insulators, the avoidance of sharp points, etc.

The largest power loss in a long-wave antenna usually occurs in the ground. Ground losses arise from the fact that the current charging the capacity between the antenna and ground flows through the capacity from the antenna to the earth and then back through the earth to the grounding point at the transmitter. The earth is a relatively poor conductor, so special provision must be made for returning these currents to the grounding point on the transmitter with a minimum of loss. One way of accomplishing this is to bury wires near the surface of the earth for the purpose of providing a low resistance path through the ground back to the transmitter. In order to be effective, these buried wires must be so arranged that the charging currents entering the earth have only a small or moderate distance to travel through the earth to reach a wire. When the antenna extends over a very large area, as in the case of very low-frequency antennas, it is desirable to modify the ground system so that the ground currents collected by the buried-wire system in given areas are delivered to local central points, and returned from there to the transmitter ground over an aboveground transmission line. In such an arrangement, reactances should be inserted at various places in the system so that the currents will be properly distributed between the various collecting points. It is also helpful in extended systems to employ several antenna grounding points in order to reduce the distance that the current must be transmitted in order to return it to the antenna ground. This is termed *multiple tuning.*

When the soil is very poor or rocky, it may be desirable to employ an artificial ground consisting of a network of wires placed a small distance above the earth and insulated from it. Such an arrangement, termed a *counterpoise,* effectively replaces the earth by a conducting screen provided that the spacing between the wires of the counterpoise is not greater than the height of the counterpoise above the ground. To be fully effective, the counterpoise should extend out from the antenna a distance at least equal to the height of the antenna above ground, and preferably twice as great, so that it will intercept most of the electrostatic field in the vicinity of the antenna. A combination of the counterpoise and buried-wire grounds is also possible. In such arrangements, the counterpoise is connected to the transmitting antenna coil so that if it is assumed that the earth is a neutral plane, the counterpoise has the opposite polarity with respect to ground as does the antenna. The distribu-

[1] An excellent discussion of antenna losses and ground systems is given by N. Lindenblad and W. W. Brown, Main Considerations in Antenna Design, *Proc. I.R.E.*, Vol. 14, p. 291, June, 1926. See also E. F W. Alexanderson, Transatlantic Radio Communication, *Trans. A.I.E.E.*, Vol. 38, p. 1269, 1919.

tion of current between the counterpoise and the buried ground system will depend upon the negativeness of the counterpoise with respect to the ground, and there will be a particular division of current in each case that causes the total ground losses to be a minimum.

28. Transmitting Antennas for Broadcast Frequencies (550–1,500 Kc).—The modern broadcast transmitting antenna consists of a tower[1] that functions as a vertical radiator. The tower is in some cases provided with top loading. This is usually in the form of a ring, spider, or other equivalent arrangement providing a lumped capacity at the top of the tower. Such a capacity "hat" can be connected directly to the top of the tower, or it may be insulated from the tower and connected thereto by a coil. In another form of top loading, the top section of the tower is insulated and connected with the remainder of the tower through a series inductance.

Field Distribution in the Vertical Plane.—The field distribution in a vertical plane of a grounded vertical radiator depends upon the height, as discussed in Par. 11 and

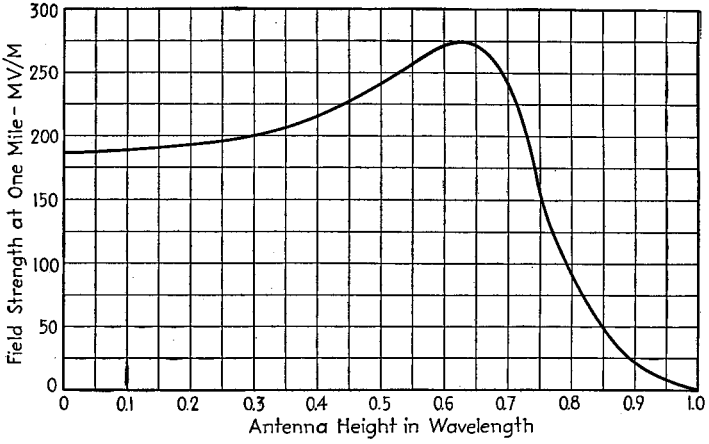


FIG. 85.—Field strength along the horizontal as a function of antenna height for a vertical grounded radiator with one kilowatt radiated power.

shown in Fig. 25 (also see Figs. 85 and 86). With antenna heights of the order of a half wave length or slightly greater, there is marked concentration of energy along the horizontal plane, and a corresponding reduction in the radiation at large vertical angles. This is a desirable result from broadcast purposes, as explained in Par. 15, Sec. 10.[2]

The optimum height of a simple vertical radiator depends upon the transmitter power, transmitter frequency, earth conditions, etc., and will vary with conditions existing in the ionosphere. Experience has shown that an equivalent electrical height of 0.53 wave lengths (190 electrical degrees) is about optimum for ordinary conditions.[3] Although a height of 0.625 wave lengths gives maximum concentration of radiation along the horizontal, it is accompanied by an undesirably large lobe of radiation at high vertical angles.

A tower with top loading gives almost exactly the same field distribution in a vertical plane as is obtained with a simple vertical radiator of somewhat greater vertical

[1] Tower antennas can be analyzed as wide-band antennas of the type discussed in Par. 36, which gives a number of curves and formulas directly applicable to tower antennas.

[2] Stuart Ballantine, On the Optimum Transmitting Wavelength for a Vertical Antenna over a Perfect Earth, *Proc. I.R.E.*, Vol. 12, p. 833, December, 1924.

[3] G. H. Brown, A Critical Study of Broadcast Antennas as Affected by Antenna Current Distribution, *Proc. I.R.E.*, Vol. 24, p. 48, January, 1936.

height. Top loading is used where it is not feasible to construct a simple radiator of the required height, as when commercial air lanes pass near by. Top loading is also used when flexibility is desired, since the equivalent electrical height of a top-loaded antenna can be changed simply by modifying the amount of top loading.

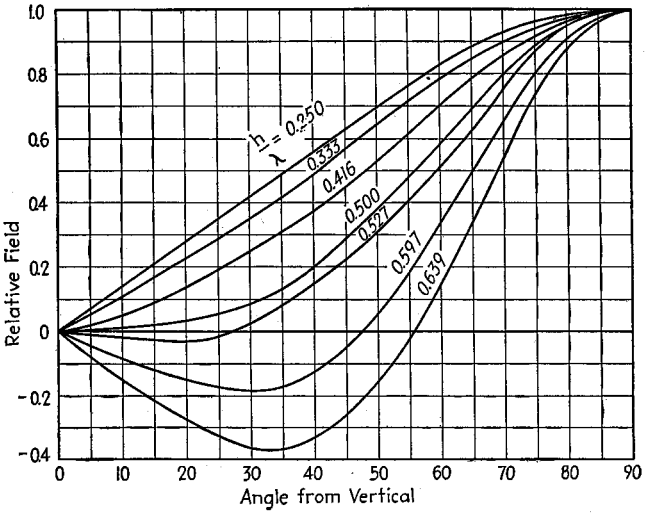


FIG. 86.—Field strength as a function of angle of elevation for vertical radiators of different heights.

Another method of concentrating the radiation along the horizontal is to employ a tower radiator considerably greater than a half wave length high, and then to insert series condensers at intervals of ⅜ wave length or less, thus modifying the current distribution in a manner analogous to that of Fig. 4*c* and maintaining the current everywhere in the same phase, even though the height is greater than a half wave length. Such an arrangement has a lobe of radiation of substantial amplitude at relatively high vertical angles, which can be canceled by combining with an additional radiator consisting of a horizontal circular disk (or equivalent). Such a disk radiates a single high-angle lobe, and if placed at the proper height and excited with current of appropriate magnitude, will neutralize the undesired high-angle lobe that would otherwise be present.[1] The high-angle radiation can also be eliminated by a ring of low vertical radiators arranged about the main antenna and excited in the proper magnitude and phase.[2]

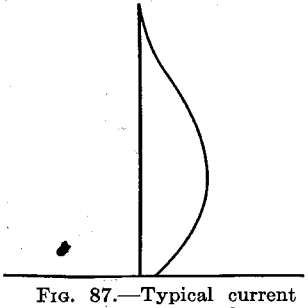


FIG. 87.—Typical current distribution in a tapered tower radiator.

Tower Construction.—The towers used as vertical radiators may be of either uniform or tapered cross section, may be either self-supporting or guyed, and are sometimes insulated from ground and sometimes not.

A tower of uniform cross section has a current distribution that is substantially sinusoidal. With such towers, it is possible to predict with considerable accuracy

[1] J. W. Labus, A Broadcast Antenna for "Low Angle" Radiation, *Proc. I.R.E.*, Vol. 23, p. 935, August, 1935.

[2] Ralph N. Harmon, Some Comments on Broadcast Antennas, *Proc. I.R.E.*, Vol. 24, p. 36, January, 1936.

the equivalent electrical height and the impedance that can be expected between the lower end of the tower and ground. Towers with tapered cross section (*i.e.*, wide base and narrow top) have a current distribution of the character shown in Fig. 87. The departure of this distribution from the sinusoidal distribution results in the effective electrical height being somewhat less than the actual physical height. The quantitative relation between current distribution and taper has not been worked out; so with tapered towers it is not possible to predict the exact electrical height that will be obtained. If top loading is employed with tapered towers, it is possible to adjust the top loading experimentally to get any desired electrical height within wide limits, and thereby obtain in a tapered tower performance that is identical with that which can be obtained with a tower of uniform cross section.

When a guyed mast is used, the supporting guy wires must be broken up at several points by insulators. One such insulator is normally placed at each end of the guy wire, with additional intermediate insulators also being used spaced at intervals, usually of the order of 0.1 wave length. These guy-wire insulators can be of the "egg" type, or equivalent arrangement, such that a mechanical failure of the insulating material does not interfere with the functioning of the guy as a mechanical support

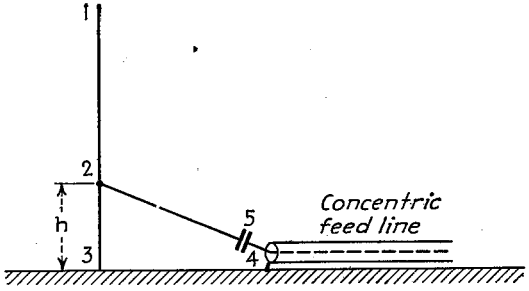


FIG. 88.—Shunt-excited grounded vertical radiator.

for the tower. The voltage duty of the guy-wire insulators depends upon the number and location of the insulators, their electrostatic capacity, the power of the transmitter, and the antenna height. In the case of a tower that is in the order of a half wave length high, a potential of roughly 30 volts per wave length is induced along the wire for one watt of power in the antenna. The voltage for other powers is proportional to the square root of power. The voltage duty on the guy-wire insulators is accordingly quite small, being about 300 volts when the insulators are spaced 0.1 wave length and the transmitter power is 10 kw. The exact value can be determined only by rather elaborate calculations, but this crude approximation gives the order of magnitude.[1]

Methods of Exciting Tower Antennas.—A tower antenna may be insulated from ground and excited by applying a voltage between the lower end of the tower and ground (series excitation), or alternatively the tower may be grounded and the excitation applied across a section h of the tower, as shown in Fig. 88 (shunt excitation).

In series excitation, the important considerations involved are the impedance between the base of the antenna and ground, and the voltage that must be applied to this impedance across the base insulator in order to deliver the full transmitter power to the antenna at modulation peaks. The impedance between the base and ground of a tower radiator acts very much like the impedance at the sending end of a transmis-

[1] A thorough study of the voltage requirements imposed on the guy-wire insulators, in which such calculations are gone into in some detail, is given by G. H. Brown, A Consideration of the Radio-frequency Voltages Encountered by the Insulating Material of Broadcast Tower Antennas, *Proc. I.R.E.*, Vol. 27, p. 566, September, 1939.

sion line open-circuited at the receiving end and having a length equal to the tower height.[1] The correspondence is not exact, however, since it is generally found that the frequencies at which the impedance has unity power factor correspond to antenna

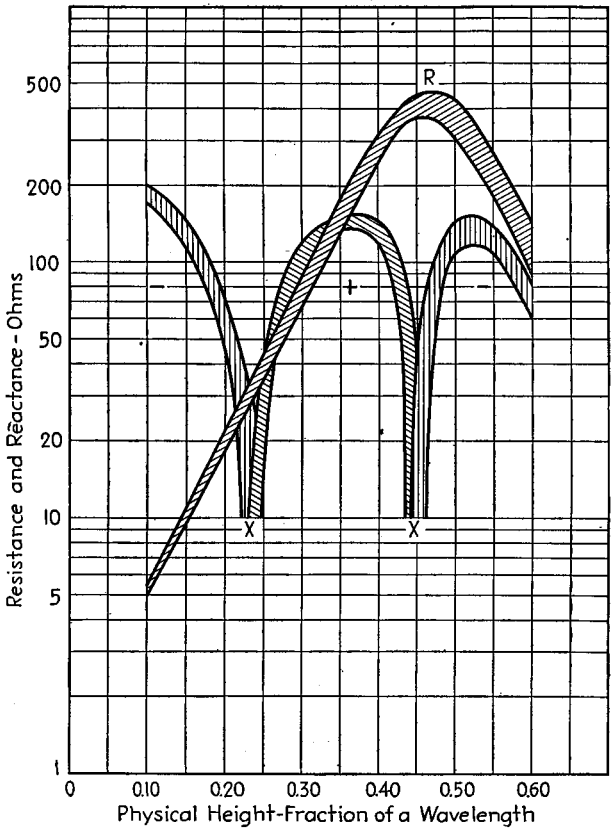


FIG. 89.—Resistance and reactance components of impedance between tower base and ground of vertical radiators as a function of height (average for five guyed towers, as given by Chamberlain and Lodge).

heights somewhat less than would be expected by the simple transmission-line theory, as illustrated in Fig. 89. The situation is further complicated by the fact that there is a lumped capacity between the bottom of the tower and ground that is in shunt with

[1] The characteristic impedance Z_0 to be used in making such a calculation is

$$Z_0 = 138 \log_{10} \frac{h}{\rho} - \left(60 + 69 \log_{10} \frac{2h}{\lambda}\right)$$

where h = tower height above earth.
ρ = radius of antenna conductor.
λ = wave length in free space.
This relation is given by J. F. Morrison and P. H. Smith, The Shunt-excited Antenna, *Proc. I.R.E.*, Vol. 25, p. 673, June, 1937. An alternative relation

$$Z_0 = 60 \left\{\log_{10} \left(\frac{h}{\rho}\right) - 1\right\}$$

is given by G. H. Brown, A Consideration of the Radio-frequency Voltages Encountered by the Insulating Material of Broadcast Tower Antennas, *Proc. I.R.E.*, Vol. 27, p. 566, September, 1939. Still another alternative is given by Eq. (65).

the impedance of the antenna itself. In view of these uncertainties, it is necessary to obtain experimentally the impedance between the tower and ground before it is possible to design the coupling networks and to predict the voltage across the base insulator, unless previous experience has been obtained from a very similar installation.[1]

The voltage across the base insulator depends upon the magnitude and power factor of the impedance at the base of the tower and the amount of power delivered to the antenna, and will vary greatly in accordance with the impedance properties. With antennas that are in quarter-wave resonance, the base impedance is a resistance of approximately 37 ohms, and the base insulator voltage is quite small, being approximately 545 peak volts for one kilowatt of carrier power modulated 100 per cent. The base voltage increases rapidly as the antenna height becomes less than that for quarter-wave resonance. The voltage across the base insulator is a maximum when the equivalent electrical height is a half wave length. The base impedance is then a resistance of the order of 200 to 400 ohms, corresponding to 1,265 to 1,800 volts peak across the insulator during modulation peaks for one kilowatt of carrier. The lower values tend to go with towers that are tapered and that have wide bases. For powers other than one kilowatt, the voltage across the insulator is proportional to the square root of power.

The shunt-excited antenna[2] shown in Fig. 88 can be considered as consisting of an antenna 1-2, in which the exciting voltage between the lower end 2 and ground is developed across the section 2-3 of a one-turn inductance 2-3-4-5, formed by the inclined wire 2-5-4, the antenna section 2-3, and the ground return 3-4. It is possible to calculate roughly the proper location and slope of the inclined wire by determining the impedance of the antenna 1-2 between 2 and ground, by making use of transmission-line theory, and then calculating the performance of the one-turn loop. In practice, however, there are enough unknown factors involved so that it is necessary to employ a certain amount of cut and try to obtain desired results. The shunt-fed arrangement lends itself particularly well to concentric-conductor feed arrangements. In this case, the inclined wire is connected to the transmission-line tower at a point such that the resistance component of the system impedance, as viewed from point 4, looking toward 2-3 (Fig. 88), has a resistance equal to the characteristic impedance of the concentric feed line. The reactive component of the impedance at point 4, which is always inductive, is then resonated out by means of series capacity 5. Point 2, at which the inclined wire should be connected with the tower, will depend upon the height and taper of the tower, etc., but is of the order of one-fifth the tower height in a typical case. The end 4 of the wire is then located at a distance from the tower base such that the slope of the wire is roughly 45°.

The performance that can be obtained from a tower radiator is substantially the same with shunt feed as with series feed, provided that the former is carefully arranged to minimize unnecessary losses. In particular, inasmuch as the current circulating in the loop 2-3-4-5 involves a section 3-4 of ground, it is necessary that considerable care be taken to keep the resistance in this section 3-4 small. This can be done by providing buried or surface wires to carry the current.

[1] Good discussions concerning the measurement of the impedance of broadcast antennas are given by D. B. Sinclair, Impedance Measurements on Broadcast Antenna, *Communications*, Vol. 19, p. 5, June–July, 1939; W. A. Fitch and W. S. Duttera, Measurement of Broadcast Coverage and Antenna Performance, *R.C.A. Rev.*, Vol. 3, p. 340, January, 1939; A. W. Ladner, discussion, Wireless Section *I.E.E.*, Vol. 11, p. 118, June, 1936. In the last reference it is pointed out that in making impedance measurements the exciting voltage must normally be applied at the point in the antenna where the impedance is being measured.

Some results of impedance measurements are given by A. B. Chamberlain and W. B. Lodge, The Broadcast Antenna, *Proc I.R.E.*, Vol. 24, p. 11, January, 1936.

[2] Further information on shunt-excited antennas is given by J. F. Morrison and P. H. Smith, The Shunt-excited Antenna, *Proc. I.R.E.*, Vol. 25, p. 673, June, 1937; Pierre Baudoux, Current Distribution and Radiation Properties of a Shunt-excited Antenna, *Proc. I.R.E.*, Vol. 28, p. 271, June, 1940.

Shunt excitation is not practical, however, when the antenna height is much less than 0.2 wave length. With very short antennas, it is impossible to obtain a reasonable resistance component at the terminal of the inclined wire unless this wire is connected at an excessive height above ground.

Transmission Lines.—The energy can be carried from the transmitter to the antenna by means of either a coaxial line or a two-wire line. The coaxial line is more expensive, but is capable of giving a superior performance. A coaxial line may be buried, or mounted either on or just above the earth's surface, and is preferably arranged so that it is gastight and is then filled with dry nitrogen that is maintained at a slight positive pressure by means of a tank and regulating valve to prevent entrance of moisture through breathing. Considerable care must be paid to the mechanical design of such a line in order to obtain airtight joints, to provide for expansion and contraction with temperature changes, etc.[1]

Two-wire transmission lines are constructed following customary practice, and are widely used. They are somewhat more subject to troubles from weather, such as sleet, etc., than the coaxial line, but are relatively inexpensive.

Transmission lines for broadcast antennas are always made to function as nonresonant lines. This is done by means of an impedance matching network proportioned to convert the actual impedance offered by the antenna system to a resistance equal to the characteristic impedance of the transmission line. In the case of concentric lines, and also with open wires in which one side of the system is grounded, this is accomplished with a simple T or π network of the type discussed in Par. 25, Sec. 3. When a balanced two-wire transmission line is employed, it is necessary to provide a transformer involving inductive coupling to go from the balanced transmission-line system to the unbalanced antenna that has one side grounded.

Ground Systems for Tower Antennas.[2]—The chief source of energy loss in a tower radiator occurs in the ground. Currents flow from tower to ground through the capacity from tower to ground, and then must return through the earth to the base of the tower. Appreciable energy losses occur in the earth as a result of its resistance, because even though the cross section of earth available to carry current may seem large, skin effect confines the current to the region near the earth's surface, and at the same time the earth is at best a poor conductor.

Losses in the ground can be minimized by employing a system of buried wires radiating out from the base of the tower to provide low resistance paths by which current can return to the tower. Study has shown that these radials should extend at least a quarter of a wave length from the base of the tower and, preferably, a half wave length, and that there should be at least 90 such radials. Such an arrangement makes it possible for nearly all the charging current returning through the earth to find its way to a wire without having to travel an excessive distance through the earth. The exact depth to which the ground wires are buried, the earth conductivity, and the size of the wires are relatively unimportant. The principal thing is that they be of adequate length and be not too far apart, so that the distance from any point on the ground in the vicinity of the tower to the nearest radial will not be great.

Investigation of the ground currents in the vicinity of an antenna shows that the height of an antenna has little effect on the length of ground wire that should be used.[3]

[1] For discussions of coaxial line design, see J. B. Epperson, Installation of Coaxial Transmission Lines, *Electronics*, Vol. 12, p. 30, July, and p. 31, August, 1939; W. S. Duttera, New Coaxial Transmission Line at WTAM, *Electronics*, Vol. 12, p. 30, March, 1939.

[2] A thorough study of ground systems for broadcast antennas, including much additional information, is given by G. H. Brown, R. F. Lewis, and J. Epstein, Ground Systems as a Factor in Antenna Efficiency, *Proc. I.R.E.*, Vol. 25, p. 753, June, 1937.

[3] It has been found possible to calculate the distribution of currents in the ground by assuming that a moderate amount of resistivity does not cause the distribution to differ appreciably from that existing in a perfect earth. One then calculates the currents that enter the earth and return to the tower, on

In tower radiators that are of the order of a half wave length high and that are insulated from ground, there is a considerable voltage between the base of the tower and ground. This causes a large current to flow through the lumped capacity between the tower base to ground, and it is therefore important that care be taken to avoid power loss from these capacity currents. A wire screen connected to the buried ground system is sometimes placed under the tower for this purpose,[1] although this is not necessary if an adequate buried-wire system is employed.

With a properly designed ground system, the losses in the earth are relatively small compared with the energy radiated, provided that the tower has a height of $\frac{1}{8}$ wave length or more.

Antenna Arrays.—Extensive use is made of simple antenna arrays for broadcast purposes. These arrays consist of two or occasionally three to four spaced tower radiators. The directional pattern obtained depends upon the spacing and upon the relative magnitude and phase of the currents in the various towers.[2]

An antenna array can serve two principal purposes in broadcast work. First, when the population is not uniformly distributed in all directions about the transmitter, it is possible by means of a directional array to concentrate the radiated energy in the areas where the population is greatest, thereby giving better signals where improved service will be of greatest value. Second, it is possible by means of a directional array to reduce greatly the field strength radiated in a particular direction, and in this way to reduce the interference that the transmitter will cause with another station some distance away operating on the same channel.

In order to obtain the desired directional pattern, it is necessary that the relative magnitudes and phases of the currents in the various antennas be accurately controlled. The best way of ensuring that the desired conditions are being obtained is to measure these quantities directly. This can be done by mounting a one-turn loop rigidly on the tower, so that the loop is coupled to the magnetic flux produced by the current flowing in the tower, and then bringing the loop output to the transmitter building through a concentric line. By using a cathode-ray tube to compare the outputs obtained in this way from the various antennas involved, one can readily obtain the relative magnitude and phase.[3]

It is to be noted that one cannot determine the relative magnitudes and phases of the antenna currents from the relative magnitudes and phases of the transmission-line currents delivered to the coupling network associated with the individual antennas. This is a result of the mutual impedance between the antennas, which causes the line and antenna currents to differ in both magnitude and phase in different ways for the individual antennas.

29. Short-wave Transmitting Antennas. *Half-wave Antenna Arrangements.*—Half-wave antennas are used in short-wave communication when directivity is not desired or when simplicity is of more importance than directivity. Such half-wave antennas may be placed vertically or horizontally, the latter placement being the most common.

---

the basis of the electric field existing in the vicinity of the tower, and then assumes that these currents are confined to the region near the surface of the earth as a result of skin effect. When this is done, it is interesting to note that the loss in the earth is greatest at a distance of approximately $\frac{1}{3}$ of a wave length from the base of the tower. See George H. Brown, The Phase and Magnitude of Earth Currents Near Radio Transmitting Antennas, *Proc. I.R.E.*, Vol. 23, p. 168, February, 1935.

[1] An increase of 10 per cent in field strength from the use of a ground screen 50 by 50 feet is reported in a particular case by R. F. Guy, Notes on Broadcast Antenna Developments, *R.C.A. Rev.*, Vol. 1. p. 39, April, 1937.

[2] A discussion of simple arrays such as are used in broadcast work is given in Par. 13. See also W. S. Duttera, Some Factors in the Design of Directive Broadcast Antenna Systems, *R.C.A. Rev.* Vol. 2, p. 81, July, 1937.

[3] For further details, see John F. Morrison, Simple Method for Observing Current Amplitude and Phase Relations in Antenna Arrays, *Proc. I.R.E.*, Vol. 25, p. 1310, October, 1937; Simplifying the Adjustment of Antenna Arrays, *Bell Lab. Rec.*, Vol. 17, p. 390, August, 1939.

With the horizontal configuration, the height above earth has an important effect upon the distribution of radiated energy in the vertical plane, and so should be carefully chosen.

Methods by which power may be supplied to a half-wave antenna from a balanced two-wire transmission line are illustrated in Fig. 90. Method (*a*) does not require that the antenna length even approximate a half wave length, since changing the antenna length simply affects the resonances appearing on the transmission feed line. In method (*b*), the termination of the transmission line is completed through the capacity from the open wire of the transmission line to the antenna, as shown dotted. In this arrangement, the antenna length must approximate a half wave length, and the transmission line adjustment must be such as to provide a voltage loop at the antenna termination of the line. While arrangements (*a*) or (*b*) in Fig. 90 fail to provide characteristic impedance termination for the transmission line, nonresonant conditions can

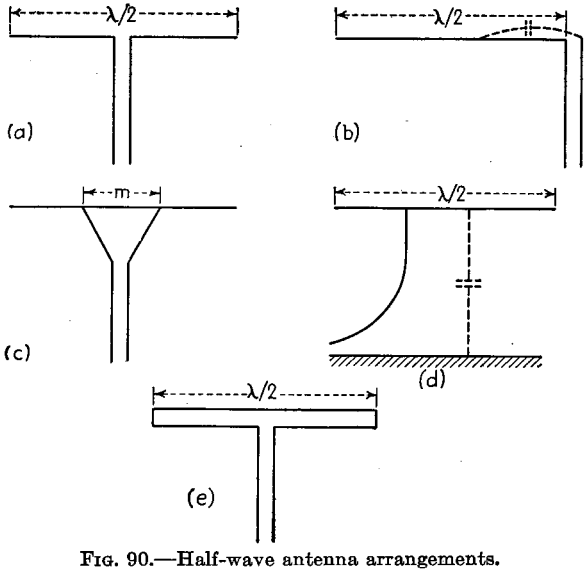


FIG. 90.—Half-wave antenna arrangements.

be obtained over the main part of the transmission line, if desired, by the introduction of an impedance-matching arrangement in the transmission line near the antenna.

An impedance match that permits nonresonant operation can be obtained with the arrangements (*c*), (*d*), and (*e*) in Fig. 90. At (*c*), the antenna length is chosen to give half wave resonance, and the spacing m is adjusted to provide an impedance match. In the case of a horizontal antenna above a perfect earth, this spacing depends upon height in the manner illustrated in Fig. 91.[1] The tapered section of the transmission line must have a taper so gradual as to introduce negligible discontinuity. An apex angle of about 75° is typical. The single-wire transmission-line arrangement shown at (*d*) is a modification of arrangement (*c*). Here the return circuit is completed through the capacity of the antenna to ground, and the ground. The proper adjustment of such an arrangement can be made by observation of the current distribution of the line and antenna.[2] Resonances on the single-wire transmission line that are

[1] P. S. Carter, Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. I.R.E.*, Vol. 20, p. 1004, June, 1932.

[2] See also W. L. Everitt and J. F. Byrne, Single-wire Transmission Lines for Short-wave Radio Antennas, *Proc. I.R.E.*, Vol. 17, p. 1840, October, 1929.

characterized by minima (or maxima) that are exact multiples of a half wave length distant from the antenna indicate that the line is not attached to the correct point on the antenna, while resonances with minima (or maxima) at other distances indicate that the antenna is not in resonance. The arrangement of Fig. 90*e* is basically similar to that at (*a*), except that an impedance-transforming arrangement is involved. Here two closely spaced half-wave antennas are connected together at their ends, and one of the dipoles is fed at the center from a balanced transmission line. The two antennas operate in parallel, each carrying half the total current, so that the impedance offered to the transmission line is four times the loop radiation resistance of a half-wave antenna, or approximately 290 ohms. A two-wire transmission line having this value of characteristic impedance can be readily designed. Other transformation ratios can

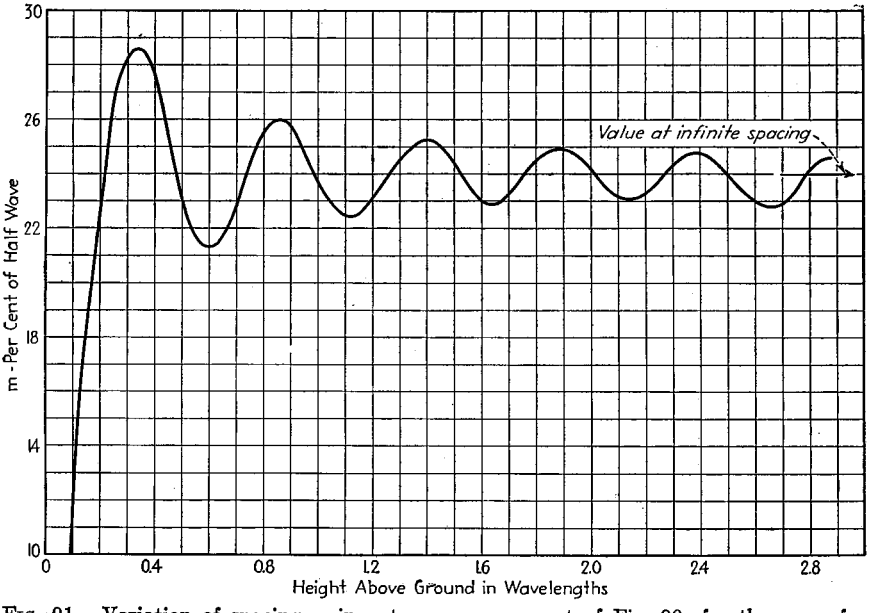


FIG. 91.—Variation of spacing *m* in antenna arrangement of Fig. 90*c* for the case of a horizontal half-wave antenna with variation in height aboveground.

be obtained by making the diameters of the two radiators unequal or by using more than two radiators.[1]

Directional Short-wave Transmitting Antennas.—Directional transmitting antennas are used to concentrate the available transmitter power in the direction in which it is desired to communicate. A moderate amount of directivity can be obtained by simple antenna arrays involving several radiators, such as the close-spaced arrays of Fig. 47, the two-element array of Fig. 92*a*, the simple multielement array of Fig. 92*b*, etc. In the last arrangement, the individual horizontal radiators are made slightly less than a half wave length long, and the loops *aa* are slightly greater than a quarter wave length long, so as to insert capacitive reactance in series with the horizontal wires. The result is to give a current distribution similar to that of Fig. 4*c*, in which the currents in the horizontal elements are all in phase.

When a large amount of directivity is desired, as in commercial point-to-point communication, the usual practice is to employ either a rhombic antenna, or an array of V's

[1] P. S. Carter, Simple Television Antennas, *R.C.A. Rev.*, Vol. 4, p. 168, October, 1939.

as illustrated in Fig. 43. Such arrangements are inexpensive to construct and give a satisfactory amount of directivity.

Highly directional arrays involving many individual radiators in broadside, end-fire, or some other equivalent arrangement, although once very common, no longer find extensive use. This is because their complexity results in expensive construction and difficulty in tuning up. Examples of such multielement arrays are shown in Fig.

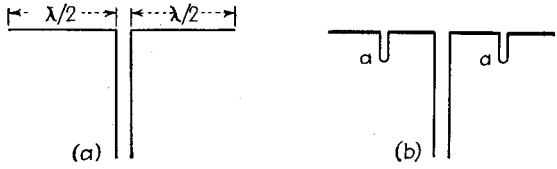


FIG. 92.—Simple directional arrays.

93. At (*a*), the antenna is formed by folding a wire in such a manner that the vertical sections all carry current in the same phase, while the horizontal sections serve as one side of half-wave resonant lines that produce a phase reversal. At (*b*), the antenna wires serve simultaneously as radiators and transmission lines by being so arranged that the currents along any diagonal are in the same phase. The diagonals at 90° then combine to give the effect of a vertically polarized array. At (*c*) and (*d*), the radiators are connected to a resonant line at voltage loops, are spaced a half wave length apart,

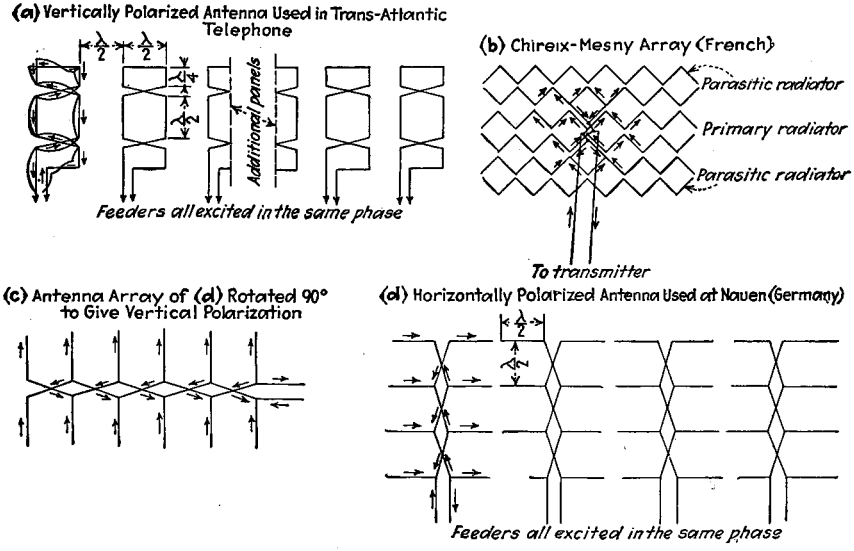


FIG. 93.—Commercial designs of broadside antenna arrays involving combinations of many half-wave radiators. The reflector antenna system that is always used to give a unidirectional pattern is omitted in this figure for the sake of clearness.

and have successive radiators connected to alternate sides of the line. This causes all radiators to be excited in the same phase. Antenna systems such as these, involving numerous radiators, are tuned up by exciting from a convenient source of power and observing the current distribution over the system. The lengths of the various radiators and line sections are then adjusted by trial and error until the current loops are at the desired places.

Aircraft Radio Antennas.[1]—Short-wave aircraft antennas can be classified as fixed-wire, trailing-wire, and shunt-fed-wing types.

The fixed-wire type of antenna is strung from one part of the plane to another, sometimes but not always, with the use of a short mast. The plane acts as the ground, and the effective height is necessarily low, with the result that the radiation resistance is correspondingly small at the lower frequencies, generally being of the order of 1 to 2 ohms at 2 to 4 mc. The fundamental frequency for quarter-wave resonance of such an arrangement depends upon the size of the antenna system, and with the largest planes is about 3 mc, while in small planes, it may be as great as 15 mc. Fixed-wire antennas are accordingly normally operated at frequencies lower than their resonant frequency, and so have a relatively large capacitive component to their impedance, which must be tuned out at the transmitter by the use of a suitable antenna tuning inductance.

With fixed-wire antennas, there is danger that the antenna wire may pick up heavy ice loads. The tendency can be minimized by arranging the antenna so the angle between the antenna and the direction in which the plane travels does not exceed 15 to 20°.

The trailing-wire antenna consists of a wire reeled out from the tail or underside of the plane, of such length as to give quarter-wave resonance when the plane is used as the ground. The trailing wire is normally terminated with either a weight or some sort of drag device. The trailing-wire antenna is relatively efficient, since it has a high radiation resistance, and furthermore has the advantage that by varying the length of wire reeled out, one can adjust the antenna for operation at any desired frequency, thus eliminating much tuning and loading apparatus.

The shunt-fed-wing antenna arrangement makes use of the wings of the airplane as a radiator, and excites them by shunt feeding. To accomplish this, a feeder is taken through the skin of the fuselage and grounded a short distance out on the wing, and the other transmitter terminal is grounded to the airplane body. With this arrangement, the wing becomes excited in a manner similar to that of a shunt-fed broadcast antenna. Experience indicates that such an arrangement may be used if the wing spread of the airplane is 40 per cent or more of the working wave length. Such an arrangement gives results that compare favorably with fixed-wire antennas.

Aircraft antennas are frequently used over a wide frequency range. As a result, the impedance as a function of frequency is of considerable importance, since this determines the design of the antenna coupling system.[2]

The field radiated from an aircraft antenna is usually a mixture of horizontally and vertically polarized waves. The directional pattern for these two components will generally differ greatly, and the ratio of their strengths will vary with the direction with respect to the plane.

Mobile Radiators.[3]—Short-wave transmitting antennas for automobiles, tanks, etc., present a relatively difficult problem, particularly when the maximum length of antenna that can be used is only a small fraction of a wave length.

The usual antenna used for such service consists of some form of whip mounted vertically, with the vehicle body used as the ground. The antenna is brought into resonance by means of an inductance coil connected between its lower end and the body. For best results, the diameter of the antenna should be as large as possible.

[1] For further discussion see George L. Haller, Aircraft Antennas, *Proc. I.R.E.*, Vol. 30, p. 357, August, 1942; R. McGuire and J. Delmonte, Development of Aircraft Radio Antennas, *Communications*, Vol. 20, p. 5, March, 1940; Howard K. Morgan, "Aircraft Radio and Electrical Equipment," Pitman, New York, 1939; P. C. Sandretto, Principles of Aeronautical Engineering, McGraw-Hill, New York, 1942.

[2] Examples of impedance characteristics of typical aircraft antennas are given by Haller, *loc. cit.*

[3] For further discussion, see Millett G. Morgan, Improved Low-frequency Mobile Radiator, *Electronics*, Vol. 13, p. 33, July, 1940; William R. Wilson, A Solenoid-whip Aerial, *Electronics*, Vol. 14, p. 56, January, 1941.

This reduces the reactive component of the antenna impedance, and therefore reduces the size of tuning inductance required, giving lowered losses in the coil. There is also an advantage in top-loading the antenna, either by some form of outrigger arrangement or by inserting an inductance in series with the antenna near the top, as illustrated in Fig. 94. The advantage of such top loading is that it forces the upper parts of the antenna to carry a substantial current, and thereby makes the effective height of the antenna approach more closely the actual physical height.

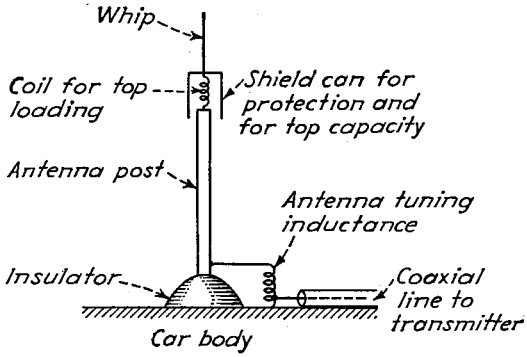


FIG. 94.—Top-loaded antenna for mobile service.

Multifrequency-tuned Antenna Systems.—Circumstances sometimes arise where it is desired to operate the same antenna at a number of specified fixed frequencies at different times. This result is achieved with the arrangement shown in Fig. 95.[1] Here the length a_1 is made equal to a half wave length at the highest frequency to be transmitted. Parallel resonant circuits L_1C_1 and $L_1'C_1'$ are then added at the ends of this section, these circuits being resonant at the frequency for which a_1 is a half wave length. Operation at the next lower frequency is then obtained by adding sections b and b', so that the equivalent electrical length of the section a_2, when the reactance introduced by the resonant circuits is taken into account, corresponds to half-wave-length resonance for the next lower frequency of transmission. These added sections are inoperative for the frequency for which a_1 is in half-wave resonance, since they are isolated by the high impedance of the parallel resonance circuits L_1C_1 and $L_1'C_1'$. Thus one obtains two resonant frequencies corresponding to the two desired frequencies of transmission. Additional resonant frequencies can be introduced as required by adding further parallel resonant circuits tuned to the appropriate frequencies, as L_2C_2 and $L_2'C_2'$, together with additional segments to the antenna length, such as c and c'.

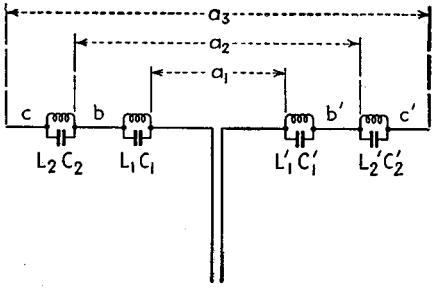


FIG. 95.—Multifrequency antenna.

Transmission Lines for Short-wave Transmitters.[2]—Transmission lines used to transmit power from a transmitter to a short-wave antenna system are usually either conventional concentric or two-wire transmission lines.

[1] Howard K. Morgan, A Multifrequency Tuned Antenna System, *Electronics*, Vol. 13, p. 42, August, 1940.

[2] See also Par. 14, Sec. 3.

Two-wire transmission lines are normally arranged so that they are balanced with respect to ground, and this requires either that the antenna system be likewise balanced or that a suitable transformer be employed. A concentric line is unbalanced with respect to ground, and if it is to be used to deliver power to a symmetrical antenna system, some provision must be made for transforming from an unbalanced to a balanced system.[1] Transformation from a coaxial system to a balanced system may

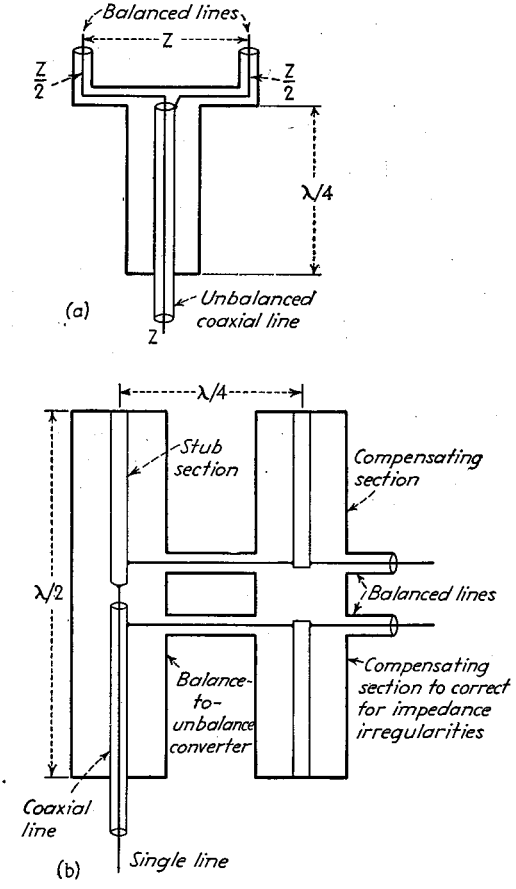


FIG. 96.—Transformers for converting from an unbalanced coaxial line to a balanced line system.

be obtained by inclosing the final quarter-wave end section of the coaxial line within a coaxial cylindrical container that is conductively connected with the outer conductor of the coaxial cylinder, as shown in Fig. 96*a*. This inclosing quarter-wave-length section should have a diameter appreciably greater than that of the coaxial line in order to ensure a good balance.

A somewhat more complicated arrangement, but one that is likewise capable of giving a higher degree of balance, is illustrated in Fig. 96*b*. Here the length of the sec-

[1] The transformer arrangements of Fig. 96 are described by N. E. Lindenblad, Antennas and Transmission Lines at the Empire State Television Station, *Communications*, Vol. 20, p. 13, May, 1940.

tion inclosing the coaxial line (balance to unbalance converter) has been extended to a half wave length, and symmetry is obtained by the use of a stub section physically similar to the outer conductor of the coaxial line to which the inner conductor is connected as shown. The similar section, one quarter of a wave length from the converter, is to give a compensating irregularity for the purpose of canceling any reflections that may be introduced by the converter. This eliminates all possibility of any impedance irregularity being introduced.

Resonant transmission lines are often used with low-power transmitters when the transmission distance is small. Resonant lines do not require an impedance-matching system, and in some cases (as in Figs. 90*a* and 92*a*), it is possible to tune the whole antenna system at the transmitter end of the line and thus avoid the necessity of exact adjustment of the antenna length.

Nonresonant lines must be used if the transmission distance is large, since otherwise the power loss will be excessive. They are also necessary with high power even for short-distance transmission, in order to avoid undesirably high voltage stresses.

30. Transmitting Antennas for Ultra-high Frequencies (30 to 300 Mc).—All kinds of antennas suitable for short-wave operation are also suitable for ultra-high-frequency use. In addition, since a high frequency corresponds to a relatively short wave length, it becomes practical to employ antenna types that would have prohibitive dimensions at lower frequencies. In particular, the corner-reflector antenna is especially suitable for ultra-high frequencies, since it gives good gain, is compact, and is portable. Close-spaced arrays involving parasitic antennas, and the billboard type of antenna also find considerable use at ultra-high frequencies.

A simple antenna arrangement commonly used at ultra-high frequencies is the "flagpole" type, in which the radiating element is in the form of a flagpole and is fed from a concentric transmission line. In such an arrangement, it is desirable to provide an impedance-matching arrangement to prevent resonances on the concentric transmission line. The usual method of doing this is illustrated in Fig. 97*a*.[1] Here the central conductor of the concentric line is extended for three-quarters of a wave length beyond the termination of the outer conductor, and the upper half wave length of this extension functions as a half-wave radiator. The quarter-wave-length section nearest the termination of the concentric line, acting in conjunction with a wire *a* connected to the outer conductor, serves as a quarter-wave-length transforming section. The half-wave antenna acts as a voltage-fed radiator similar to that of Fig. 90*b*, and by using proper spacing between the wire of the quarter-wave-length impedance-matching section, a characteristic impedance termination for the line can be obtained. In practice, it is preferable to interchange the wire *a* and the radiator, as in Fig. 97*b*, to provide a direct contact to ground if lightning should strike the antenna.

Another type of flagpole antenna is illustrated in Fig. 97*c*.[2] Here the antenna *c* is an extension of the inner conductor of a quarter-wave-length mounting section *d*. The purpose of this quarter-wave-length mounting section is to provide means by which rigid mounting can be obtained without affecting the electrical properties, since the impedance at point *e* of section *d* is practically infinite. The antenna is excited from a concentric transmission line *b* provided with a quarter-wave-length matching section *a* to convert the 37.2-ohm radiation resistance of the antenna *c* to the characteristic impedance of the concentric line *b*, which is usually higher than 37.2 ohms. The necessity of this quarter-wave-length matching section can be eliminated by shortening the antenna *c* slightly and likewise by shortening the mounting section *d*, so that the imped-

[1] W. C. Tinus, Ultra-high-frequency Antenna Terminations, *Electronics*, Vol. 8, p. 239, August, 1935.

[2] G. H. Brown and J. Epstein, An Ultra-high-frequency Antenna of Simple Construction, *Communications*, Vol. 20, p. 3, July, 1940.

ance at point e is resistive, but has a value equal to the characteristic impedance of the transmission line. An arrangement of the type of Fig. 97*c* can be mounted at the top of a flagpole by rearrangement as in Fig. 97*d*. Here the ground plane is supplied by the crossed wires *gg* extending a quarter wave length from the radiator. The parts labeled *a*, *b*, *c*, and *d* have the same significance as in Fig. 97*c*.

Another antenna of the flagpole type is illustrated in Fig. 97*e*. Here the outer conductor of a concentric line is extended and folded back on itself by means of a cylinder of relatively large diameter a quarter wave length long, as shown. This outer cylinder, together with the extension of the central conductor, then functions as a half-

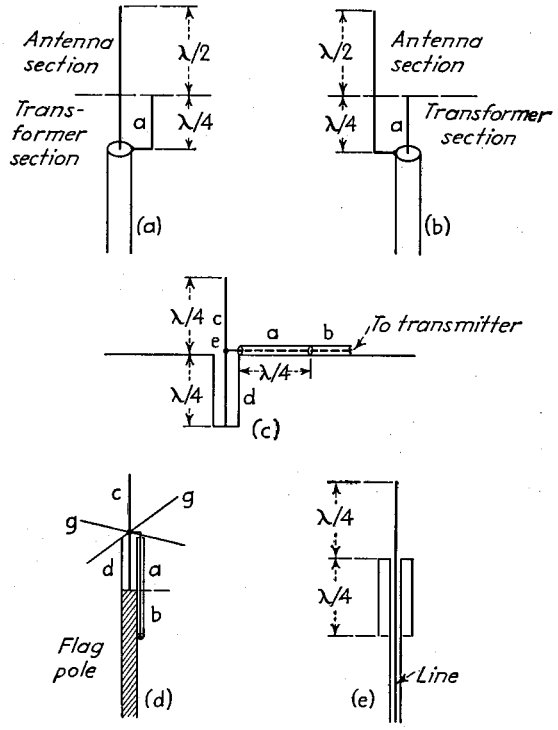


FIG. 97.—Antenna arrangements of the "flagpole" type.

wave-length radiator, and when in resonance, offers an impedance of approximately 74 ohms as a terminating load for the concentric line.

Broadcasting Antennas for Ultra-high Frequencies.—In ultra-high-frequency broadcasting (including frequency modulation, television, etc.), it is desired to concentrate the radiation to a high degree along the horizontal, and at the same time to have a relatively circular pattern in a horizontal plane. This result can be achieved by a stacked array of turnstile antennas or of Alford loops. Such arrangements radiate horizontally polarized waves. A corresponding directional pattern with vertically polarized waves can be obtained by employing a broadside array of vertical half-wave radiators, with the line of the array placed vertically (*i.e.*, a number of vertical half-wave radiators stacked vertically and excited in the same phase).

An ingenious arrangement that gives both vertically and horizontally polarized waves (actually, circular polarization), with a circular pattern in a horizontal plane and

directivity in the vertical plane, has been developed.[1] This consists of eight half-wave radiators tilted at 45° with respect to the vertical and lying uniformly spaced in the surface of a cylinder having its axis vertical, and a diameter slightly over a half wave length. These radiators are all excited in the same phase. The resultant action is then to produce a substantially circular pattern in the horizontal plane, with the radiation being circularly polarized. Several such arrangements can be stacked one above the other to give increased directivity in a vertical plane.

31. Transmitting Antennas for Microwaves.—At wave lengths less than one meter, it is possible to use types of antennas that would be physically too bulky at longer wave lengths. Thus, when a large amount of directivity is desired, parabolic and horn antennas are generally employed. When a moderate amount of directivity is needed, a corner reflector antenna, or a close-spaced array involving a radiator associated with parasitic antennas, can be used; likewise, a rhombic antenna.

When directivity is not important, half-wave radiators are used, excited from a concentric line through some form of transformer, or in a flagpole arrangement. With a half-wave radiator, it is common practice to provide an artificial ground in the form of a copper sheet or disk several wave lengths in extent.

RECEIVING ANTENNAS

32. Long-wave Receiving Antennas.—The usual long-wave receiving antenna is either a wire run diagonally to a tower or other point of high elevation, or a vertical grounded wire with a single-wire flat top in the form of a T or an inverted L.

Directivity in the horizontal plane can be obtained by the use of a loop or wave antenna employed either singly or in an array.[2] The amount of directivity that it is practical to obtain at the longer wave lengths is quite limited, because transmission lines of excessive length are required if an antenna array has dimensions of the order of several wave lengths. Hence the directivity of a simple loop, or of a single wave antenna, is all that is employed except in a few special instances. Even then, a loop, unless physically large, abstracts very little energy from low-frequency radio waves. Likewise, a wave antenna has the disadvantage that the direction from which it receives signals cannot be changed without completely reconstructing the antenna.

33. Broadcast Receiving Antennas.—The ordinary broadcast receiver has enough sensitivity so that an antenna with an effective height of only a few meters is adequate to provide a signal that will override the ordinary set noise. Marked directivity should be avoided in broadcast receiving antennas, since signals are normally to be received from all directions.

The usual broadcast receiving antenna is a single wire having a length of a few feet up to about 100 feet, or consists of a wire leading to some metal object insulated from ground, such as a bed spring, etc. A metal plate placed in the receiver cabinet has also been used. The ground connection for the antenna system is usually supplied by the power cord.

Many present-day broadcast receivers are provided with a built-in loop antenna that can be used as an alternate to the more conventional antenna. Although the effective height of the loop is lower than that of the usual wire antenna, it is normally sufficient to permit adequate reception under average conditions, and the loop has the advantage of making the set completely portable. Built-in loop antennas have a certain amount of directivity, with the result that the reception from a particular station may depend upon the exact position of the receiving set. This directivity is not so great as might be expected, however, since there is considerable residual vertical

[1] N. E. Lindenblad, Antennas and Transmission Lines at the Empire State Television Station, *Communications*, Vol. 21, p. 10, April, 1941.

[2] An array of six loops for long-wave reception is described by A. G. Lee, Radio Communication Services of the British Post Office, *Proc. I.R.E.*, Vol. 18, p. 1690, October, 1930.

antenna effect, because no attention is paid to balancing and shielding the loop. Inclosing a loop in an electrostatic shield will reduce the interference produced by near-by noise sources[1] although such a shield results in a sacrifice in compactness.

Automobile Antennas.[2]—A great variety of antenna arrangements are used in automobile receivers. Perhaps the most effective is a vertical whip, connected to the receiver by means of a transmission line. Other arrangements that are employed include plates mounted below the running board and insulated from the body of the car, and a plate connected in the top. This latter arrangement is suitable only for cars in which the top is not all metal.

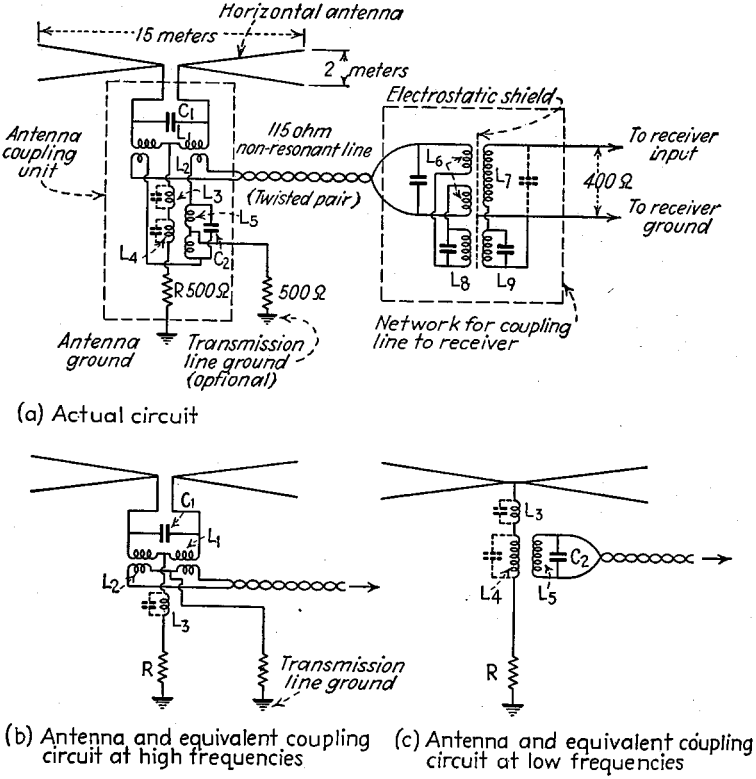


FIG. 98.—All-wave antenna system, together with equivalent circuits at high and low frequencies.

All-wave and Noise-reducing Antenna Systems.—All-wave broadcast receivers operate most effectively when used in conjunction with special antenna systems designed to give adequate pickup over a wide frequency range with a minimum of directivity, which at the same time minimize noise pickup.

An example of a wide-band antenna system is shown in Fig. 98.[3] This employs a horizontal doublet antenna having an over-all length slightly less than a half wave

[1] Stanford Goldman, A Shielded Loop for Noise Reduction in Broadcast Reception, *Electronics*, Vol. 11, p. 20, October, 1938.

[2] Information of the effectiveness of various automobile antennas is given by D. E. Foster and G. Mountjoy, Measurement of Effective Height of Automobile Antennas, *R.C.A. Rev.*, Vol. 3, p. 369, January, 1939.

[3] Harold A. Wheeler and V. E. Whitman, The Design of Doublet Antenna Systems, *Proc. I.R.E.*, Vol. 24, p. 1257, October, 1936.

length at the highest frequency to be received. Each half of the antenna consists of two wires arranged in a V as shown. This lowers the reactive component of the antenna impedance and so simplifies the design of the associated coupling networks. The antenna has its own ground, and is coupled to a balanced nonresonant line by an antenna coupling unit. The receiving terminal of the balanced line is coupled to the unbalanced input of the receiver through a "line-to-receiver filter" or transformer, which has a balanced primary with an electrostatic shield between primary and secondary. The antenna coupling unit is so designed that at the high frequencies the antenna functions as a horizontal doublet coupled to the transmission line through the transformers L_1 and L_2, as shown in Fig. 98*b*, with the coupling network so designed that the impedance match between the line and antenna is sufficiently good to permit the energy transfer to approach the maximum possible value. At low frequencies, the

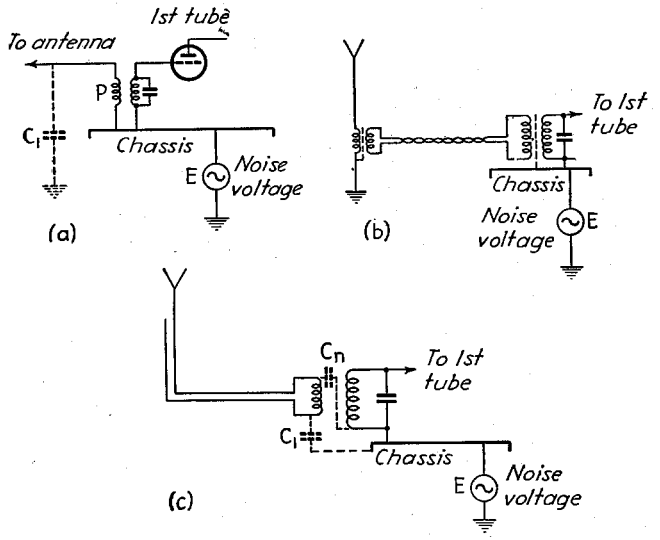


FIG. 99.—Systems for preventing a noise voltage E existing between chassis and ground, as a result of the power-cord connection, from causing a noise voltage from being applied to the grid of the first tube of the radio receiver.

arrangement functions as a flat-top antenna that is coupled to the line through the transformer L_4L_5, as shown in Fig. 98*c*. Transition between doublet and flat-top action occurs in the vicinity of 5 mc, and comes about as a result of the fact that at high frequencies the capacity C_2 is a virtual short circuit, and so prevents action as a flat-top antenna, while at low frequencies the transformer L_1L_2 is so inefficient as to prevent action as a doublet antenna.

An antenna system of the general type illustrated in Fig. 98 commonly results in material improvement in the signal-to-noise ratio. The use of a transmission line makes it possible to place the antenna at a location relatively free from noise, which is often of help. At the higher frequencies, where the system functions as a horizontal antenna, further reduction in noise results from the fact that most man-made sources of noise are located close to the earth, and so generate relatively weak horizontally polarized waves.

Experience has shown that a considerable amount of noise at broadcast and higher frequencies is the result of noise currents that reach the receiver via the power-

line cord. This places a radio-frequency noise voltage E between the chassis and ground, as shown in Fig. 99*a*. The usual input connection to the antenna with one terminal of the primary coil P grounded to the chassis then permits noise currents to flow through P to ground via antenna capacity C. The result is a noise voltage applied to the grid of the first tube. Noise from this source can be largely eliminated by any arrangement that provides the receiver with a balanced transmission-line input, which is mechanically and electrically symmetrical and electrostatically shielded. A simple noise-reducing antenna arrangement based on this principle is illustrated schematically in Fig. 99*b*.[1] It is found that such an arrangement will give marked improvement in signal-to-noise ratio, particularly if advantage is taken of the fact that the transmission line permits the antenna to be placed some distance away from the receiver, where there is less noise.

A modification of the arrangement of Fig. 99*b*, in which the electrostatic shielding can be dispensed with, is shown in Fig. 99*c*.[2] Here the antenna is formed by the extension of one wire of a two-wire transmission line that terminates in a coil coupled to the first tuned circuit of the receiver. An adjustable neutralizing condenser C_n is employed, and is so adjusted as to compensate for the capacity C_1 from the lower end of the primary winding to the chassis. With proper adjustment, the noise voltage E applied between the chassis and ground will produce no potential difference across the input tuned circuit connected to the first tube. This balance is substantially independent of the frequency.

A relatively simple system for getting a response over a wide range of frequencies makes use of a number of half-wave antennas resonant at different frequencies and connected in parallel across the terminals of the transmission line. Each antenna in such an arrangement is made approximately a half wave length for the mean frequency of the band that it is intended to cover, so that by using a number of frequencies and properly spacing them, a wide frequency range can be covered with fairly satisfactory results.

Radio-frequency Distributing Systems.[3]—In apartment houses, it is sometimes desirable to provide a master antenna system that will operate a number of radio receivers simultaneously.

The most satisfactory arrangement of this type makes use of a nonresonant transmission line, across which the various receivers are connected through an attenuation unit that introduces about 30 db loss. The loss in the attenuator is compensated for either by employing an individual buffer amplifier between each receiver and the line or, preferably, by means of a wide-band amplifier connected between the antenna and the transmission line. Such an arrangement will give satisfactory reception with receivers having ordinary sensitivity, and interaction between receivers is practically eliminated because effects originating within one receiver are attenuated 60 db before reaching the input of another receiver.

In simple installations, the attenuating network and amplifier are sometimes omitted. In such cases, the receivers are provided with high-impedance input circuits, and are connected directly across the transmission line.

34. Short-wave Receiving Antennas.—A wire fifteen to twenty feet long is commonly used as a general-purpose receiving antenna for short-wave signals. Somewhat better results, particularly, improved signal-to-noise ratio, can be obtained by using an antenna system similar to systems for the short-wave ranges of all-wave broadcast receivers (see above).

[1] W. L. Carlson and V. D. Landon, A New Antenna Kit Design, *R.C.A. Rev.*, Vol. 2, p. 60, July, 1937.
[2] V. D. Landon and J. D. Reid, A New Antenna System for Noise Reduction, *Proc. I.R.E.*, Vol. 27, p. 188, March, 1939.
[3] F. X. Rettenmeyer, Radio-frequency Distributing Systems, *Proc. I.R.E.*, Vol. 23, p. 1286, November, 1935; D. J. Fruin, A Central Antenna System, *Electronics*, Vol. 12, p. 37, November, 1939.

Where it is possible to employ directional receiving antennas, as, for example, in point-to-point communication or in a case in which the signals to be received originate in a certain portion of the world, there is considerable advantage in employing an antenna system having high directivity. Noise and interfering signals arriving from directions other than that of the desired signal are then reduced, thereby improving the signal-to-noise ratio and reducing interference. When the noise or interference is distributed in a random way with direction, the improvement in signal-to-noise ratio obtained with a directional receiving antenna is the same as the power gain of the antenna when used for transmitting. When most of the noise arrives from some other direction than does the desired signal, the improvement is even greater than the power gain, while in the unfortunate case in which the noise and signal come from the same direction, directivity gives no benefit whatsoever.

It is very desirable that the directional pattern of a receiving antenna have small minor lobes. If there is a minor lobe of considerable intensity, noise or interfering signals coming from this particular direction will be received with almost full intensity, even though this minor lobe may be quite narrow in width and so cause very little power to be wasted if the same antenna were to be used for transmitting purposes. It is also important in the case of receiving antenna systems having high directivity that the transmission line from antenna to receiver, and the circuits associated with the receiver input, be very carefully designed to avoid direct pickup of energy from the passing waves. In view of the fact that the discrimination against signals and noise from undesired directions may be of the order of 20 to 60 db in a directional antenna, only a very small spurious response to such an undesired signal introduced by unbalanced transmission lines, etc., will greatly reduce the benefits obtained from the directivity.

The types of directional receiving antennas most commonly used for commercial work are the rhombic and fishbone antennas. Both of these are characterized by excellent directivity and ability to respond almost equally well over a frequency range of about 2 to 1. The fishbone antenna has the further advantage of possessing a directional pattern characterized by unusually small minor lobes. In both the rhombic and fishbone antennas, it is possible to make the response approach zero in almost any desired backward or side direction by proper adjustment of the magnitude and phase angle of the terminating impedance.

It is, of course, possible to use other types of directional antennas for short-wave reception. V's, broadside arrays, systems involving parasitic radiators, etc., are employed to some extent.

35. Ultra-high-frequency and Microwave Receiving Antennas.—Reception at ultra-high and microwave frequencies is generally carried out with the same types of antennas that are used for transmission at these frequencies.

The principal problem involved with ultra-high-frequency and microwave receiving antennas is in obtaining an adequate pickup sensitivity, so that the amount of energy abstracted will be sufficient to override the set noise. This difficulty comes about from the fact, as explained in Par. 7, that an antenna is capable of abstracting energy from a section of the wave front that extends only about a quarter of a wave length away from the antenna. At very short wave lengths, it is accordingly a difficult matter to abstract much energy. Thus a horn or parabolic reflector will intercept an amount of energy corresponding to that contained in a section of the wave front having about the same area as the mouth of the horn or reflector, and even then a portion of this energy may be reradiated. Thus, even though a horn of reasonable physical dimensions may give an enormous power gain when operated as a transmitting antenna at centimeter wave lengths, the amount of energy that it abstracts is small compared to that of an ordinary broadcast receiving antenna 25 feet long operating at broadcast frequencies.

Because of considerations involving the amount of energy abstracted, it is found that the half-wave antenna becomes increasingly unsatisfactory as the frequency is increased above 30 mc. At these higher frequencies, it is usually necessary to employ some form of antenna array having greater ability to abstract energy.

WIDE-BAND ANTENNAS—TELEVISION ANTENNAS[1]

36. Wide-band Antenna Systems. *General Considerations.*—All antenna systems of the resonant type capable of operating over a wide frequency range are characterized by having a diameter or other crosswise dimension that is relatively large. The use of a large diameter does not affect the radiation resistance greatly if at all, but it reduces the inductance and increases the capacity of the antenna elements. The resultant effect is to cause the equivalent resonant circuit that the antenna represents to have a lowered Q, and hence broader resonance.

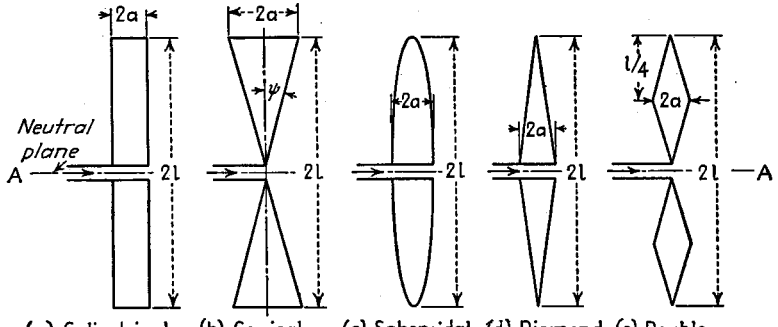


FIG. 100.—Typical wide-band antenna arrangement. All the antennas shown can be considered cross sections of a figure of revolutions. Dash line AA represents a neutral plane that can be replaced by a conductor (earth) with the omission of the lower half of the radiating system.

The large cross-sectional area needed to obtain wide-band characteristic can be obtained in a variety of ways. A number of these are illustrated in Fig. 100 and involve such expedients as a half-wave antenna of large diameter (cylindrical antenna), the use of double cones, etc. The Empire State television antenna shown in Fig. 110 is a special application of the general axiom that a large diameter is favorable for wide-band operation.

Cylindrical Antennas.—The cylindrical antenna in Fig. 100*a* can be considered as the generalized case of a half-wave antenna in which the diameter of the antenna is not necessarily very small compared with the length. The characteristics of such an antenna are given in Figs. 101, 102, and 103,[2] expressed in terms of the parameter K_a, which represents the average characteristic impedance and is given in Fig. 104 and

[1] It is to be noted that while this section considers only antennas of the resonant type, nonresonant antennas, such as described in Pars. 12, 14, and 16, are also wide-band antennas, and would be used where high directivity was desired.

[2] These curves are from S. A. Schelkunoff, Theory of Antennas of Arbitrary Size and Shape, *Proc. I.R.E.*, Vol. 29, p. 493, September, 1941. For further information see E. B. Moullin, The Radiation Resistance of Aerials Whose Length Is Comparable with the Wavelength, *Jour. I.E.E.*, Vol. 78, p. 540, 1936; also, Wireless Section, *I.E.E.*, Vol. 11, p. 93, June, 1936; E. B. Moullin, The Radiation Resistance of Surfaces of Revolution, Such as Cylinders, Spheres and Cones, *Jour. I.E.E.*, Vol. 88, p. 50, March, 1941, Part III; and E. B. Moullin and H. Page, discussion on The Radiation Resistance of Surfaces of Revolution, Such as Cylinders, Spheres and Cones, *Jour. I.E.E.*, Vol. 88, p. 171, June, 1941, Part III; Ronald King and F. G. Blake, Jr., The Self-Impedance of a Symmetrical Antenna, *Proc. I.R.E.*, Vol. 30, p. 335, July, 1942.

by the relation

$$K_a = 276 \log_{10} \frac{2l}{a} - 120 \tag{65}$$

The notation of Eq. (65) and Fig. 104 is given in Fig. 100*a*.

Examination of Figs. 101 to 103 brings out the following points: (1) The resistance that the antenna system offers at its input terminal under conditions where the total

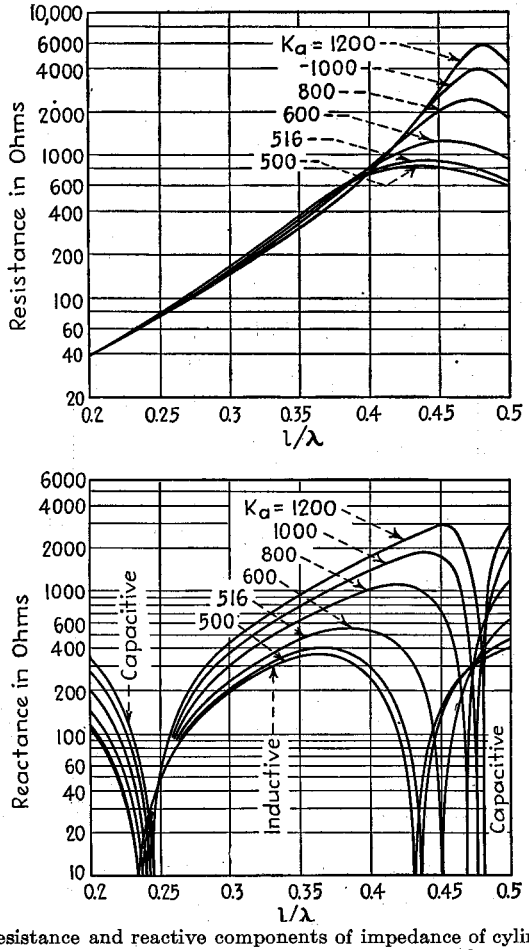


FIG. 101.—Resistance and reactive components of impedance of cylindrical antenna of Fig. 100*a* as calculated for different values of characteristic impedance K_a. (*After Schelkunoff.*) These curves assume that the resistance and reactive components are in series with each other and neglect the effect of any lumped shunting capacity that may be present across the input terminals.

length of the antenna does not exceed 0.8 wave length ($l \leqq 0.4\lambda$) is almost independent of the characteristic impedance K_a; however, when the total length approximates a wave length ($l \sim 0.5\lambda$) the resistance component of the antenna impedance depends very greatly upon the characteristic impedance K_a, being comparatively low when the characteristic impedance is low (*i.e.*, when the diameter is large); (2) the reactive component of the antenna impedance is less the lower the characteristic

impedance, this applying both for conditions close to resonance and for those far from resonance; (3) increasing the diameter of the antenna reduces the length required to obtain half-wave and full-wave resonance.

Cone Antennas.[1]—The cone antenna is illustrated in Fig. 100*b*. It will be noted that when the angle of revolution ψ is large this antenna de-

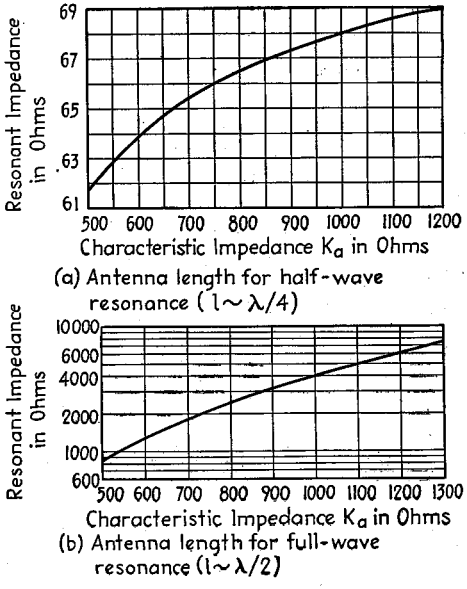


FIG. 102.

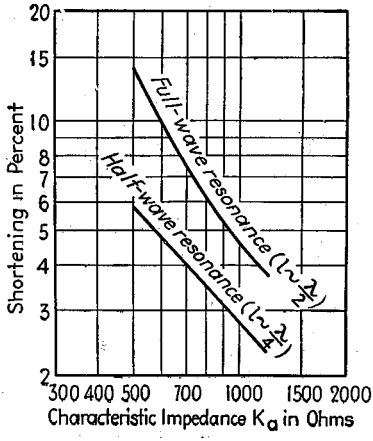


FIG. 103.

FIG. 102.—Input resistance for cylindrical antennas in half-wave and full-wave resonance.

FIG. 103.—Shortening of cylindrical antenna as a function of characteristic impedance, expressed as per cent reduction below a half wave length for half-wave resonance and a full wave length for full-wave resonance.

generates into a biconical horn operating in the *TEM* mode, as illustrated in Fig. 77 (provided that the antenna is a number of wave lengths long).

The characteristics of conical antennas depend upon the angle of revolution ψ (see Fig. 100*b*) and the length of the antenna, expressed in wave lengths. The char-

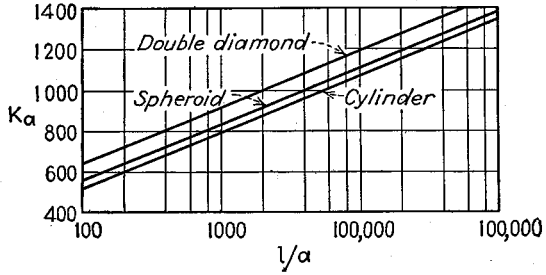


FIG. 104.—Average characteristic impedance of various types of antennas illustrated in FIG. 100.

acteristics that can be expected and the information necessary for antenna design are given in Figs. 105, 106, and 107,[2] expressed in terms of the characteristic impedance K

[1] Further information is given by Schelkunoff, *loc. cit.*; P. S. Carter, Simple Television Antennas, *R C.A. Rev.*, Vol. 4, p. 168, October, 1939; and Moullin, *loc. cit.*

[2] These are from Schelkunoff, *loc. cit.*, and Carter, *loc. cit.*

given in Fig. 108 and by the equation

$$K = 276 \log_{10} \cot \frac{\psi}{2} \tag{66}$$

where ψ is the angle of revolution, as in Fig. 100*b*. The principal conclusions to be derived from these results are: (1) the total antenna length required for half-wave resonance is appreciably less than a half-wave length; (2) the antenna length required for full-wave resonance, and for still higher order harmonic operation, is also less than the corresponding length in wave lengths measured in free space; (3) the resistance component of the input impedance of the cones under conditions of half-wave-length resonance is substantially independent of the angle of revolution, while the input resistance in the region of full-wave-length resonance varies greatly with the angle of revolution; (4) the reactive component of the impedance offered by the cones as one departs from half-wave-length or full-wave-length resonance depends upon the angle of revolution, and decreases rapidly as the angle of revolution is increased.

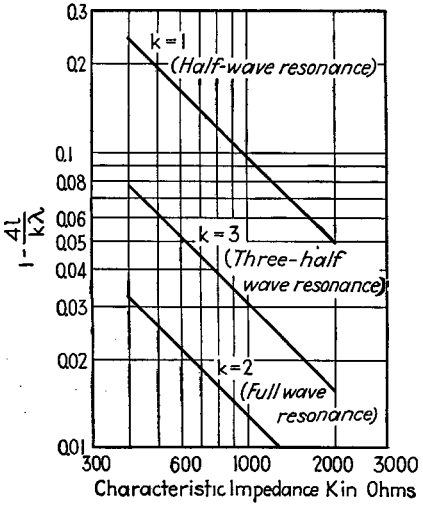


FIG. 105.—Shortening effect in conical antennas, expressed as the deviation of the resonant length from $2l = k\lambda/2$.

Miscellaneous Wide-band Antenna Structures.—The cylinder and cone are by no means the only antenna arrangements with large cross-sectional areas. Typical of the possibilities that remain are the spheroidal, diamond, and double-diamond antennas of Fig. 100 and the wide-band television antenna of Fig. 110. A folded dipole, such as is shown in Fig. 90*e*, is also a first step toward a wide-band arrangement, in that dividing the antenna into two adjacent radiators of small diameter produces an effect equivalent to a substantial increase in diameter.

The characteristics of spheroidal, diamond, and double-diamond antennas have been studied mathematically.[2] The results obtained are analogous to those for the cylindrical and conical radiators; *i.e.*, the behavior depends upon the average characteristic impedance, and this is less the greater the diameter in proportion to length (see Fig. 104). As this characteristic impedance becomes smaller (*i.e.*, the larger the cross-sectional area in proportion to wave length), the length of antenna required for resonance becomes less, and the reactive component of the antenna impedance is greatly reduced. The resistance component of the antenna impedance is not greatly affected by the average characteristic impedance (*i.e.*, by the cross-sectional dimensions) if the antenna is in the vicinity of half-wave resonance or is shorter, but with full-wave resonance ($2l = \lambda$) the resistance component of the antenna impedance is reduced very greatly as the average characteristic impedance is made smaller.

The formulas giving the impedance relations and also the actual current distributions that can be expected are to be found in the literature[3] but are too complicated and specialized to be reproduced here. The effect of cross-sectional area upon antenna

[1] Thus, Carter, *loc. cit.*, gives $2l \sim 0.73\lambda$ as the value determined experimentally for full-wave resonance for values of ψ such as would be used in practice.

[2] Schelkunoff, *loc. cit.*

[3] See Schelkunoff, *loc. cit.*

length required to produce half-wave resonance is, however, given by the following relatively simple expressions:[1]

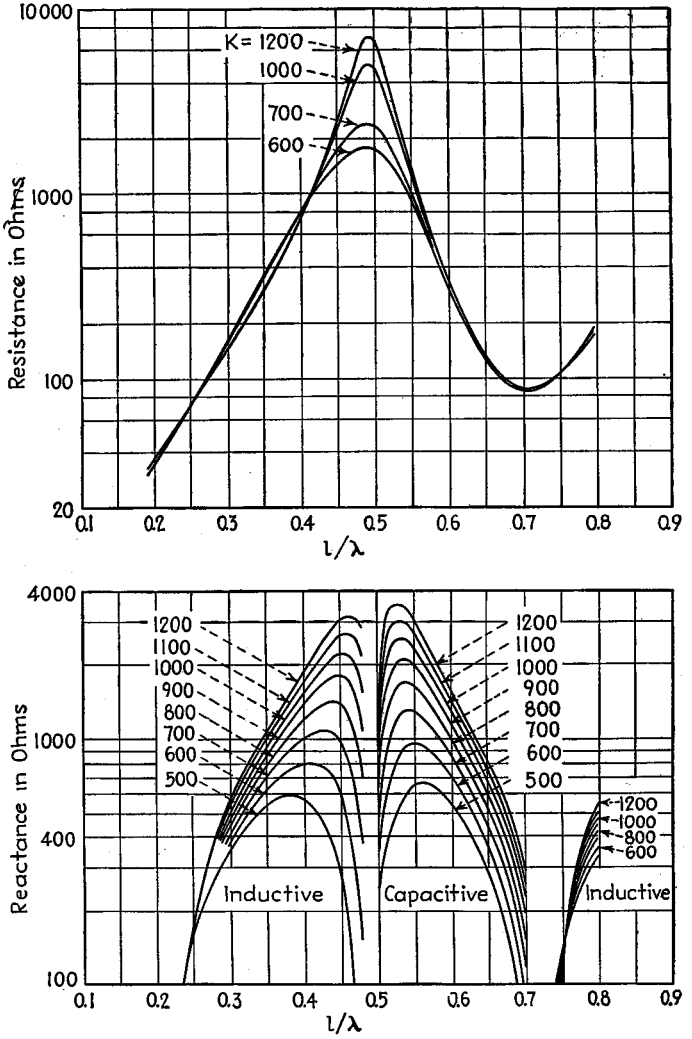


FIG. 106.—Resistance and reactance components of input impedance of conical antenna of Fig. 100*b* (*after Schelkunoff*), for small angles of revolution.

For spheroidal antenna with $a << l$:

$$2l = \frac{\lambda}{2}\left(1 - \frac{5{,}040}{(K_a + 83)^2}\right) \tag{67}$$

For double-diamond antenna with $a << l$:

$$2l = \frac{\lambda}{2}\left(1 - \frac{8.62}{K_a}\right) \tag{68}$$

[1] From Schelkunoff, *loc. cit.*

It will be noted that the arrangements of Figs. 100*a*, 100*d*, and 100*e* represent types of tower radiators in common use for broadcast transmitters (see also Fig. 86), assuming the neutral plane is replaced by the ground. It thus becomes apparent that the ordinary tower radiator used in broadcast work is a form of wide-band antenna, and the characteristics of such antennas can be analyzed in terms of the methods and formulas referred to here.

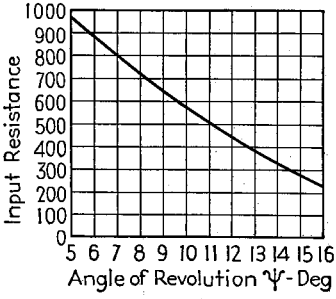


FIG. 107.—Experimental results obtained by Carter giving relationship between input resistance of a pair of cones in full-wave resonance, and the angle of revolution of the cone for large angles of revolution. Under these conditions the overall length of the pair of cones should be approximately 0.73 wave length at the mid-frequency, regardless of the angle of revolution.

37. Requirements of Antennas Used for the Transmission and Reception of Television Signals.[1] In television transmitting systems, it is necessary that the impedance match between the antenna and the transmission line leading from the transmitter be good enough to keep the reflection coefficient, as defined by Eq. (55), Sec. 3, to within a very small value throughout the frequency band covered by the television signal. This band width is of the order of 5 megacycles in typical systems, and with the assigned television channels is of the order of 10 to 15 per cent of the carrier frequency. The maximum allowable reflection coefficient that can be tolerated is of the order of 5 per cent, which means that the vector impedance must not vary more than ± 10 per cent from the mean value. Hence when the magnitude of the impedance is substantially constant over the band, as is typically the case, the reactive component of the impedance must never exceed 10 per cent of the resistance component.

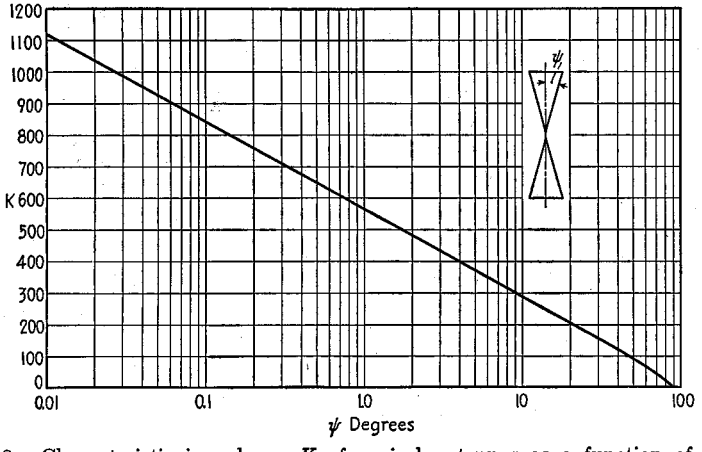


FIG. 108.—Characteristic impedance K of conical antennas as a function of angle of revolution ψ (see Fig. 100*b*).

Reflections at the antenna end of the transmission line of a transmitting system result in the formation of "ghost" phenomena, in which weak displaced images are produced as a result of energy that has been reflected back from the antenna to the

[1] See Carter, *loc. cit.*; S. W. Seeley, Effect of the Receiving Antenna on Television Reception Fidelity, *R.C.A. Rev.*, Vol. 2, p. 433, April, 1938.

transmitter, and then is reflected back to the antenna, to be reradiated with a delay time corresponding to twice the length of time required to travel the transmission line. Failure to match the impedance of the antenna and transmission line to each other will also result in distortion of the image, in that the energy transfer at some frequencies is different from that at others. This distortion is, however, of much less importance than the production of ghosts.

The receiving antenna of a television system must meet requirements that are much simpler than those applying to the case of transmission. Here it is not particularly important that a good impedance match be maintained between the antenna and the transmission line, provided that there is a satisfactory impedance match between the transmission line and the input to the receiver. Mismatching between antenna and line produces a loss of energy transfer and so distorts certain parts of the picture, but a considerable amount of mismatching can be tolerated before a serious degradation of picture quality occurs.

Reflections produced at the receiver input terminals can produce ghost images, as described above, as the result of waves that travel back to the antenna and then are returned to the receiver with a delay time corresponding to twice the line length. However, the impedance match between the transmission line and the receiver input can be made very excellent by designing the receiver input to offer a substantially resistance impedance over a wide frequency range. This imposes some limitations on the input circuit of the receiver, particularly lowering the voltage gain obtainable between the line and the grid of the first tube, but by paying this penalty a good impedance match may be readily achieved over any desired band width.

In the reception of television signals a special problem is encountered in connection with orientation of the antenna to avoid reflected waves that reach the antenna by indirect paths.[1] Waves that reach the receiving antenna from such a path as reflection from a large near-by building will have traveled a somewhat greater distance than the main part of the wave, and so will produce on the television output tube a ghost image displaced by a time delay corresponding to the extra transmission distance. The most practical means of handling this situation under ordinary conditions is to experiment with receiving locations and the orientation of the receiving antenna until a combination is obtained whereby the indirect waves received as a result of reflection are minimized to the point of causing no trouble. It is helpful in such experiments to calculate the time delay involved, from the displacement of the ghost picture, and obtain the corresponding distance in feet on the basis of a wave traveling with the velocity of light. With this information one can frequently determine the source of a disturbing reflection.

Wide-band Antennas Suitable for Television.—An ordinary quarter-wave or half-wave antenna having a diameter of the order of one and a half inches is suitable for use as a receiving antenna in the television band around 50 mc. The impedance match that can be obtained between such an antenna and a two-wire transmission line, with the use of a quarter-wave matching section, likewise of large diameter, is only fair, but is entirely adequate for receiving purposes. In contrast, a small-diameter antenna is inadequate even for receiving purposes (see Fig. 109).

The higher degree of impedance matching required in the case of television transmitting antennas is such that dipoles of even large diameter are unsuitable. A folded dipole of the type shown in Fig. 90*e* is considerably better than a single dipole, as apparent from Fig. 109, but is not quite adequate to meet the rigid requirements of a high-grade television transmitter with a 50-mc carrier. By combining two folded dipoles in a turnstile array, still further improvement is possible, as discussed below, and a passably acceptable result obtained.

[1] See Seeley, *loc. cit.*

Cone antennas or double-cone systems, as discussed in connection with Fig. 100*b*, have characteristics that more than meet the wide-band requirements of television transmitters, as is obvious from Fig. 109. The only objection to such antennas is their rather inconvenient bulk, but this can in many cases be met by replacing the cone by a number of wires. It has been found that twelve wires distributed uniformly

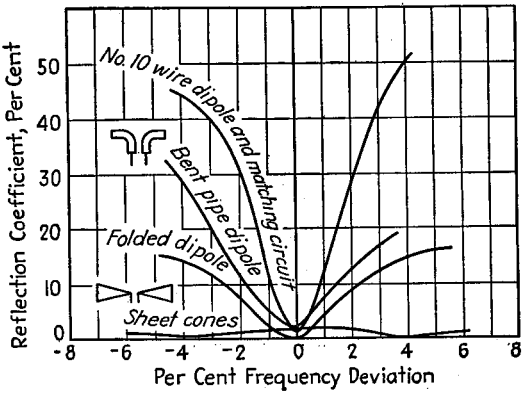


FIG. 109.—Experimentally observed values of reflection coefficient obtained between various types of antennas and a two-wire transmission line. (*After Carter.*)

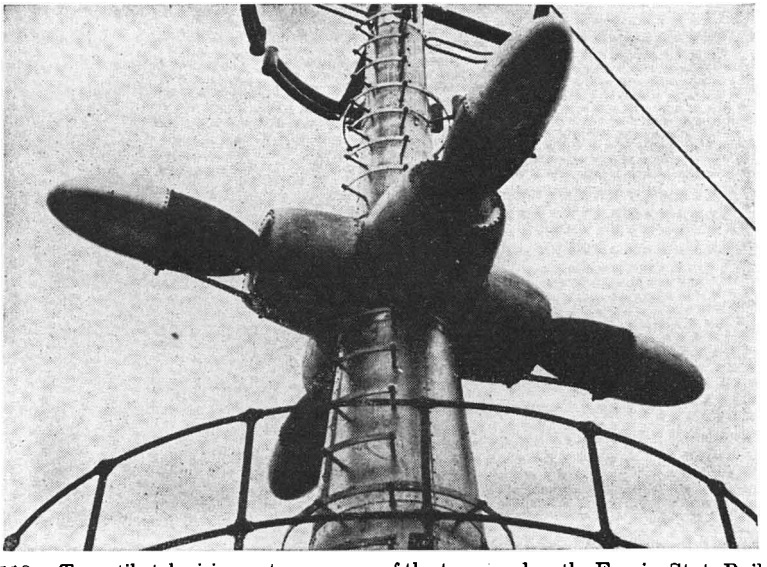


FIG. 110.—Turnstile television antenna array of the type used on the Empire State Building. Note the bracket supporting the ellipsoidal element.

around the conical surface give a result substantially equivalent to that of the solid structure, and have much less wind resistance and weight.

An ingenious wide-band television transmitting antenna has been used on the Empire State Building, and is illustrated in Fig. 110.[1] As shown in Fig. 111, this is

[1] A detailed description is given by N. E. Lindenblad, Television Transmitting Antenna for Empire State Building, *R.C.A. Rev.*, Vol. 3, p. 387, April, 1939.

an evolution of a grounded quarter-wave antenna in which the exciting voltage is applied approximately midway in the antenna structure by means of a concentric line. By using large-diameter components, and particularly by properly proportioning the different parts of the antenna system, it is possible to obtain an extremely good impedance match over a wide frequency range, as apparent from Fig. 111. In the antenna as built, the central ellipsoidal conductor is supported by a bracket, as shown in Fig. 110. This bracket represents an inductive shunt between the ellipsoid and the base portion, and has little effect upon the electrical behavior, other than to

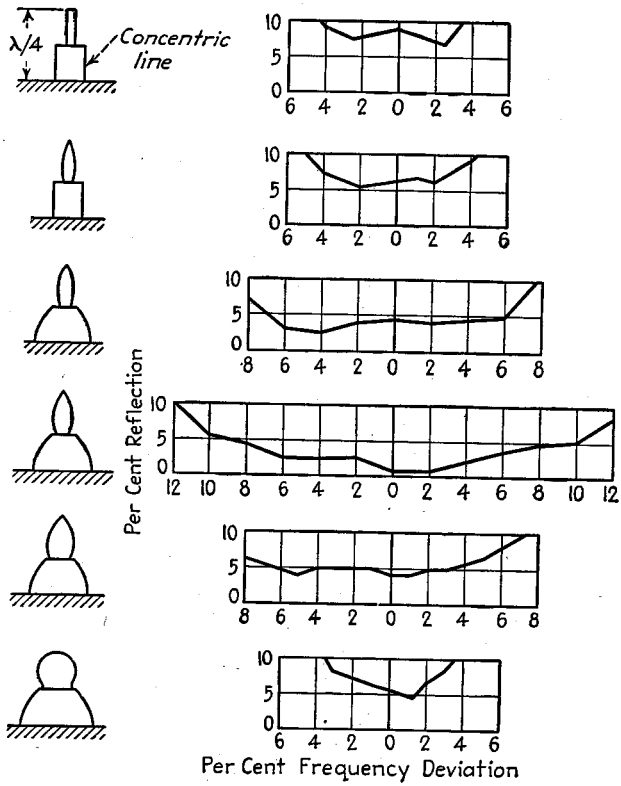


FIG. 111.—Evolution of the Empire State type of antenna structure from a quarter-wave grounded antenna excited by a concentric line at its mid-point.

modify the optimum proportions. In the Empire State installation, four antennas of this character are arranged around a central tower to form a turnstile array, as shown in Fig. 110.

Benefits from Turnstile Arrangements.—Antenna systems of the types discussed above give better performance when connected in a turnstile arrangement than when used alone. This is because of the fact that the two parts of the turnstile array must be excited 90° out of phase, and so are fed through transmission lines of different electrical lengths. When these lines are combined, compensating action is obtained whereby the impedance mismatches of one part of the system tend to become compensated for by corresponding impedance mismatches of opposite character from the other (*i.e.*, quadrature) portion of the turnstile system.

SECTION 12

RADIO AIDS TO NAVIGATION

DIRECTION FINDING[1]

1. General Considerations in Direction Finding.—Radio direction finding is accomplished by taking advantage of the fact that radio waves normally propagate between transmitter and receiver along a great-circle route. Hence a ship or airplane can obtain its location by determining the direction of radio waves sent out by transmitters at known locations. Similarly it is possible to determine the position of a radio transmitter by taking bearings on the radio waves at two receiving points.

When radio direction finding is used to determine the location of a ship, airplane, etc., it makes no difference whether the ship observes the bearings of the waves radiated from known locations or whether receivers at known locations are used to determine the bearings of waves sent out from the ship. This is because the Rayleigh-Carson reciprocity theorem (Par. 8, Sec. 11) shows that under the conditions for which accurate bearings are possible (*i.e.*, where the action of the earth's magnetic field in the ionosphere is unimportant), the observed bearings are unaffected by interchanging the transmitter and direction-finding receiver.

Direction-finding Errors Caused by Lateral Deviation and Scattering of Radio Waves. The accuracy of all radio direction-finding work is based upon the assumption that the radio waves follow a great-circle route. This is usually, but not always, the case.

Lateral deviation of the ground-wave can occur when the great-circle route between transmitter and receiver is roughly parallel to a coast line. Under such conditions the higher ground-wave attenuation over the land results in a shifting of the wave front, and the wave arriving at the receiving point tends to be bent in such a way as to cause the transmitter bearing to be displaced oceanward from the true bearing of the transmitter. Bearings that are within 10 to 15° of a coastline are therefore to be distrusted, particularly if the distance to the transmitter is large. Similar behavior is often found in mountainous regions, since the attenuation of a ground wave traveling up a valley is less than that of a wave traveling crosswise from ridge to ridge. There is hence a tendency for the energy that reaches a receiving point to arrive via a valley, even when the valley does not run in the exact direction of the transmitter.

At the higher radio frequencies, where propagation over even moderately short distances depends upon the ionosphere, the direction from which waves arrive at a receiving point will on occasion deviate appreciably from the true bearing as a result of conditions of the ionosphere. The chief source of such errors is from scattered reflections, which, as indicated in Par. 12, Sec. 10, can cause waves to arrive at a receiving point from other than the great-circle direction. Errors due to scattering are most pronounced when the normal transmission path is seriously disturbed by an ionosphere storm, or when the receiving point is not in the direction toward which the main part of the transmitted energy is directed.

There is also evidence to indicate that lateral deviation of signals away from the great-circle route can be caused by a tilt in the ionosphere in the region where the

[1] An encyclopedic treatment of direction finding is given by R. Keen, "Wireless Direction Finding," Iliffe and Sons, Ltd., London, 1938.

wave is refracted earthward. The amount of tilt likely to occur, though not great, can be sufficient to introduce appreciable error in bearing. Ionosphere tilt may be caused by ionosphere storms, or by the normal variation in the ionosphere that occurs during the day, particularly around sunrise and sunset.[1]

Errors Due to Reradiation from Neighboring Objects.—The true direction in which radio waves are traveling will be obtained only when the point of observation is remote from metal objects. Passing waves striking wires and other metal objects induce currents that produce induction and radiation fields, which, when combined with the fields of the passing waves, result in a distortion of the wave front in the vicinity of the object. Fields from induced currents produce two effects upon direction-finding equipment: (1) a deviation in the apparent direction from which the passing wave seems to arrive and (2) a quadrature effect corresponding to fields 90° out of time phase with those of the passing wave, and giving an effect equivalent to elliptical polarization. In the case of direction finders such as loops, which depend

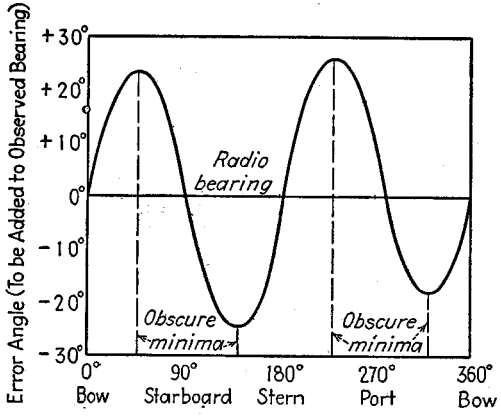


FIG. 1.—Normal quadrantal error characteristic of a ship. When the error is large, there is a tendency for the minimum to be indistinct.

upon obtaining a null for accurate indication, this quadrature component destroys the null, thereby blurring the minimum and reducing the accuracy obtainable.

These errors from induced currents are particularly important in connection with loop or similar direction finders used on shipboard.[2] When a ship-borne direction finder is located along the center line of the ship, the error introduced by induced currents usually has the general character shown in Fig. 1. The error is maximum for waves arriving 45° from bow or stern, and either small or zero when the wave bearing is directly forward, astern, or to either side. This type of error in ships is commonly termed *quadrantal error*, and is normally of such a character as to make the observed bearing angles nearer to 0 or 180°, as the case may be, than the true bearing. The quadrantal error depends on frequency. It is also affected by the loading of the ship, tending to be greater when the ship is lightly loaded, and also will, of course, be influenced by changes in rigging. When the quadrantal error for a particular bearing is large it is commonly found that the minimum of a loop direction finder is blurred.

[1] R. H. Barfield, The Measurement of the Lateral Deviation of Radio Waves by Means of a Spaced-loop Direction Finder, *Jour. I.E.E.*, Vol. 83, p. 98, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 314, September, 1938.

[2] A detailed discussion of such errors, together with compensating means, is given by C. T. Solt, The Development and Application of Marine Radio Direction Finding Equipment by the United States Coast Guard, *Proc. I.R.E.*, Vol. 20, p. 228, February, 1932.

When the error from induced currents is excessive, and particularly when there is sufficient quadrature effect to give a distinctly blurred minimum, it is necessary to minimize the effect of induced currents. This can be done by relocating the direction-finding equipment so that it is more nearly in the clear, or one may in some cases be able to minimize the effect of induced currents by insertion of insulation at points that interfere with the paths of the induced current. It is also usually possible to compensate at least partially for the effects of metal objects by introducing additional objects, as, for example, a short-circuited loop ("current loop"), strategically located, of such size and orientation as to induce in the direction-finding antenna an approxi-

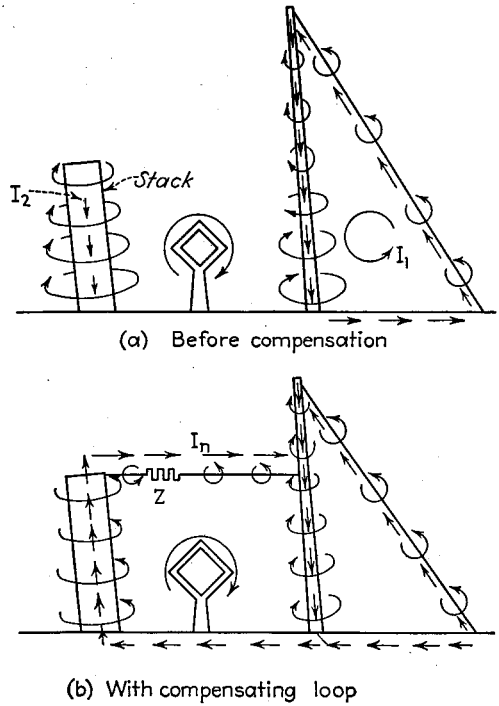


FIG. 2.—Example showing quadrantal error caused by currents induced in stack and in closed circuit formed by mast and stay, and improvement of null by means of a compensating loop. The loop is formed by connecting a wire between stack and mast. The magnitude and phase of the current I_n flowing in this loop, and hence of the compensating effect produced, can be controlled by adjustment of the magnitude and phase of the impedance Z.

mately equal but opposite effect to that produced by other objects. Compensation of this character is frequently necessary in the case of ship direction finders. A typical example is shown in Fig. 2.

Errors from currents induced in "open radiators," *i.e.*, objects such as poles, stacks, etc., that do not form closed loops, are greatly reduced by providing the direction-finding equipment, particularly antenna, with electrostatic shielding. This removes the effect of induction electrostatic fields, and so in the case of near-by "open radiators" eliminates most of the error otherwise present, because near such radiators the induction electrostatic field is much stronger than the remaining fields caused by induced current.

The possibility of errors from induced fields makes it desirable always to calibrate direction-finding equipment after it has been installed, with the use of waves from transmitters at known locations.[1] One can then either use a correction curve, or may compensate automatically for the error by some mechanical arrangement involving an adjustable cam, so that the pointer on the direction finder gives the correct direction even when this differs from the apparent direction indicated by the waves.

Errors from "Mountain Effect."[2]—In mountainous regions errors can occur in airborne direction finders as a result of waves reflected from the rough terrain combining with the direct wave in a manner that causes the apparent direction of arrival to differ from the true bearing. These errors vary with distance to the transmitter, are very sensitive to frequency, and depend upon the character of the terrain, because these factors determine the point of reflection, the magnitude of the reflected wave, and the phase relation of the direct and reflected waves.

When "mountain effect" is present, loop direction finders frequently give a blurred minimum, particularly when the signal used is modulated and so contains several frequencies. This is in addition to the fact that the bearing obtained may be in error.

The presence of mountain-effect error is indicated by irregular fluctuations in apparent bearing about the true course as the plane travels. The absence of such fluctuations is a certain sign that the effect is absent.

2. The Loop Direction Finder.—In the loop direction finder, a receiver is excited by a loop antenna that is rotated until a zero response of the receiver indicates that the plane of the loop is perpendicular to the direction in which the wave travels (see Par. 18, Sec. 11). The loop is adjusted for minimum rather than maximum response, because the percentage change in loop response with a small change in loop position is much greater in the vicinity of the loop minimum.

A simple loop antenna employed in this way gives the bearing angle of the passing radio wave, but leaves a 180° uncertainty in the actual direction of the transmitting station.[3] This ambiguity in the sense of the bearing can be removed by making use of a vertical antenna in conjunction with the loop. The addition of vertical antenna pickup to the output of a loop antenna results in one lobe of the loop pattern being enlarged, while the other is diminished, as discussed in Par. 18, Sec. 11, and also as shown in Fig. 3.

The procedure for direction finding is hence to obtain the bearing by adjusting the loop for zero response with the vertical antenna disconnected. The sense of the bearing is then determined by rotating the loop 90° in a specified direction and coupling the vertical antenna to the loop. If the addition of the vertical antenna increases the signal the sense is one way, while the diminution of the signal indicates the opposite sense.

In combining the vertical antenna pickup with the loop output to obtain a sense determination, it is necessary that the voltage developed by the vertical antenna be not greater than the resultant voltage acting in the loop circuit. It is also necessary that the phase of the vertical antenna output that is coupled into the loop circuit be such that the antenna effect induces a voltage in the loop circuit that is approximately

[1] In the case of ship direction finders the calibration is often done by having a second ship steam around the ship with the loop, and transmit continuously. The distance in such cases must not be too small, or errors may be introduced from the action of the horizontal flat top of the transmitting antenna. See J. F. Coales, Errors in Direction-finding Calibrations in Steel Ships Due to the Shape and Orientation of the Transmitting Antenna, *Jour. I.E.E.*, Vol. 73, p. 280, 1933; also, Wireless Section, *I.E.E.*, Vol. 8, p. 127, September, 1933.

[2] A comprehensive discussion of the subject is given by H. Busignies, Mountain Effects and the Use of Radio Compasses and Radio Beacons for Piloting Aircraft, *Elec. Comm.*, Vol. 19, No. 3, p. 44, 1941.

[3] A direction-finding system in which freedom from sense ambiguity is obtained by using two crossed loops in combination with a vertical antenna is described by C. E. Horton and C. Crampton, A Radio Compass Developed in H. M. Signal School, *Jour. I.E.E.*, Vol. 73, p. 284, 1933; also, Wireless Section, *I.E.E.*, Vol. 8, p. 131, September, 1933.

in phase (or phase opposition) with the resultant voltage acting around the loop. Since the resultant voltage acting around the loop is 90° out of phase with the voltage induced in the vertical antenna, this means that a 90° phase shift is required in the coupling system, which can be conveniently obtained by the use of mutual-inductance

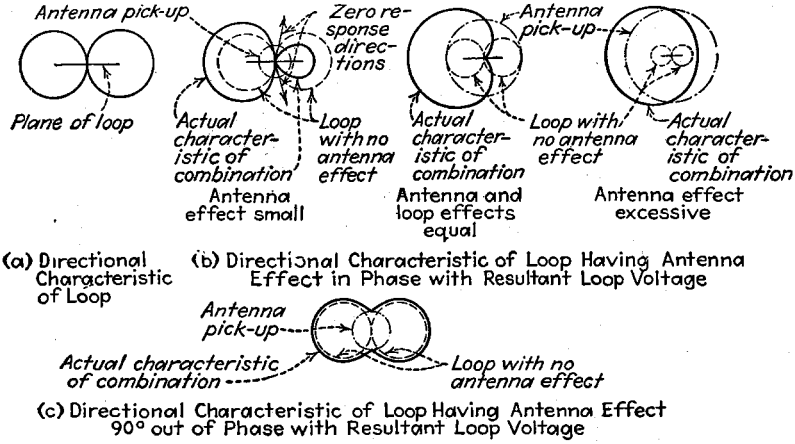


FIG. 3.—Directional characteristics of vertical loop for vertically polarized waves with and without antenna effect. Note that a small antenna effect in phase with the resultant loop voltage causes the null direction to differ from 90° with respect to the plane of the loop, while a small antenna effect in quadrature with the resultant loop voltage results in a blurred minimum.

coupling. The effect produced by adding various amounts of antenna pickup to the loop output, and also the effect of the phase of this pickup, is illustrated in Fig. 3.

Balancing and Shielding of Loop Antennas.—The bearing obtained with a loop antenna will be erroneous unless the loop is carefully balanced electrostatically with respect to ground. This is because if the loop is not so balanced, there will be a residual antenna effect that distorts the directional pattern of the loop, as shown in

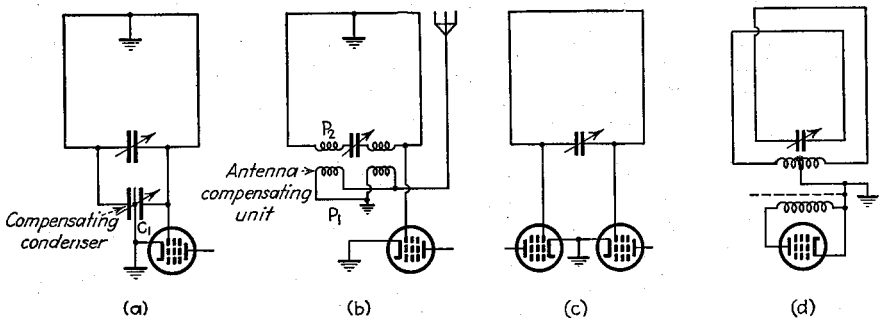


FIG. 4.—Balanced-loop antenna systems, including arrangements involving compensating capacities and mutual inductances for eliminating residual antenna effect.

Fig. 3*b* for the case marked "antenna effect small." This causes the nulls to appear in directions slightly different from 90° with respect to the plane of the loop.

Typical circuit arrangements that can be used in association with a loop to maintain symmetry with respect to ground are shown in Figs. 4*c* and 4*d*. In many practical direction-finding systems it is customary to include also a small auxiliary balancing

condenser or variable mutual inductance to compensate for any residual antenna effect. A condenser used for this purpose normally consists of a three-plate arrangement having two stators, as shown in Fig. 4*a*, while the variable mutual inductance is used to couple a controllable antenna effect into the loop, as in Fig. 4*b*. The adjustment of such a compensating arrangement can be easily made by taking advantage of the fact that in the absence of all antenna effect the two null directions of the loop differ by 180°, but when there is a small antenna effect the two nulls are slightly less than 180° apart, as shown in Fig. 3. The correct adjustment of the compensation is therefore the adjustment for which the loop, after having been rotated to give zero response, will again give zero response when reversed exactly 180°.

The accuracy with which electrostatic balance to ground can be obtained and maintained in a loop antenna is increased by inclosing the loop in an electrostatic shield. Such a shield ensures that all parts of the loop will always have the same capacity to ground irrespective of the loop orientation or of the presence or absence of neighboring objects. A loop shield normally makes use of a metal housing, usually in the form of a pipe, which is provided with an insulated joint or bushing so that the shield does not act as a short-circuited turn. The presence of such an electrostatic shield surrounding a loop antenna has practically no effect upon the performance of the loop, although there are some modifications in the detailed mechanisms of the loop operation.[1]

Shielding of loop antennas also has the advantage of preventing induction electrostatic fields from affecting the loop. This is important in the case of precipitation static (see Par. 18, Sec. 10), and also when there are near-by objects that act as an open antenna (see Par. 1 above).

Polarization Errors in Loop Direction Finders (Night Effect—Airplane Effect).[2]— The bearings obtained from a vertical loop antenna rotated about a vertical axis will be correct if the passing waves are vertically polarized. If, however, there is present a horizontally polarized downcoming component of the waves, the horizontal members of the loop have voltages induced in them that do not give zero resultant loop output voltage when the plane of the loop is perpendicular to the bearing of the radio wave.[3] In fact, if there is present nothing else but horizontally polarized downcoming waves, the null position of the loop occurs when the plane of the loop is parallel, rather than perpendicular, to the bearing of the radio wave. A vertically polarized downcoming component produces the same effect on the bearing determination as any other vertically polarized wave, and so causes no trouble. When the wave possesses both a vertically polarized component and a horizontally polarized downcoming component, the result is either that the loop null indicates an erroneous bearing or that it is impossible to obtain a zero response for any loop position. This latter occurs when the resultant voltages induced in the vertical and horizontal loop members are not in phase, and gives a loop characteristic similar to that of Fig. 3*c*. Troubles introduced by downcoming horizontally polarized waves are termed *polarization errors*.

Horizontally polarized downcoming waves are normally produced by action of the ionosphere. When this is the case, the polarization error is small near the transmitter

[1] The theory of the shielded loop antenna is given by R. E. Burgess, The Screen Loop Aerial, *Wireless Eng.*, Vol. 16, p. 492, October, 1939.

The use of shielding for loop antennas was apparently first suggested by R. H. Barfield, Shielded Loops, *Jour. I.E.E.*, Vol. 62, p. 249, 1924.

[2] An excellent discussion of this subject is given by R. L. Smith-Rose, Radio Direction-finding by Transmission and Reception, *Proc. I.R.E.*, Vol. 17, p. 425, March, 1929.

[3] The magnitudes of the voltages induced in horizontal members is determined by the resultant of the incident downcoming wave and the wave reflected from the surface of the earth. The induced voltages will hence be different for the lower and upper horizontal members and will depend on frequency and on soil conditions. The standard wave error (see Par. 4) of a loop in which the distance from any part to the ground is small compared with a wave length is 35.3° for the case of a perfect ground.

because there the sky wave is much weaker than the ground wave, but becomes of increasing magnitude as the distance is increased. At broadcast and lower frequencies, where loops are most commonly used, the sky wave is much stronger at night than in the daytime, so that the error from downcoming sky waves is commonly termed *night effect.*

Polarization error or night effect reduces the usefulness of the loop as a direction-finding device, limiting the transmission distance for which it can be used, and making this distance depend upon the time of day or night and upon the frequency. In particular, the night-effect or polarization type of error makes it impossible to use a loop at the higher radio frequencies except for very small distances.

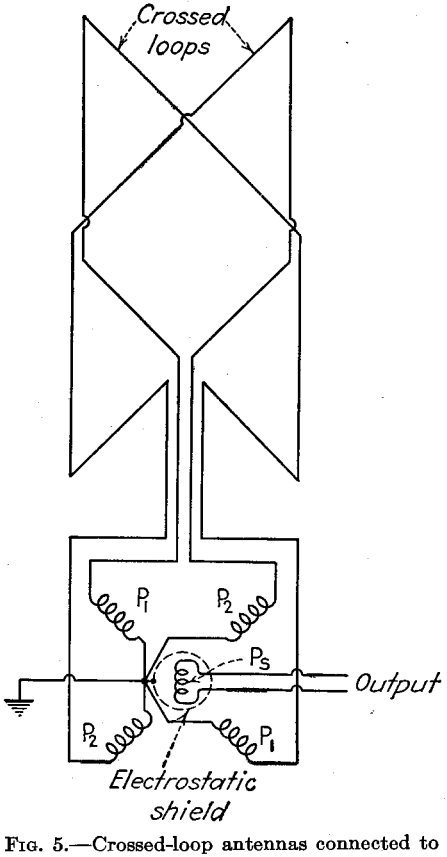


FIG. 5.—Crossed-loop antennas connected to form a goniometer system.

When night-effect error is present, the bearing of the wave from a particular station as observed by a loop often varies continuously as a result of fading effects produced by the ionosphere action. The presence of night-effect error can accordingly be detected by observing the bearing over a period of time and noting whether or not it is steady and sharp. When night effect is present, the true bearing can be obtained, at least approximately, by taking bearings at regular time intervals and averaging them, since the effects of the ionosphere tend to be random and will, over a period of time, average out to zero.

The night-effect error is normally less for bearings taken aboard an aircraft flying at a high elevation than it is for bearings taken at a point on the ground below the same airplane, *provided that the distance of transmission is large.* The reason for this is that the strength of the direct wave reaching the receiver from the transmitter is much greater at a high elevation than it is at the ground level, where this direct wave is a ground wave and suffers high attenuation. Under these conditions the sky wave reflected from the ionosphere is proportionately less important, and so is less able to introduce an error.

When signals are received by a loop antenna from an airplane that is more or less overhead, the loop antenna may be subjected to horizontally polarized downcoming waves just as when the ionosphere produces a sky wave. The result is again the production of polarization errors in the bearing. The action of this case is termed *airplane effect.*

Goniometer Arrangements.—The necessity of rotating a loop antenna to obtain a bearing can be avoided by employing two fixed loop antennas oriented 90° with respect to each other, with their outputs combined in a goniometer, as shown in Fig. 5.

The goniometer consists of two pairs of primary coils P_1P_1 and P_2P_2, one for each loop, arranged at right angles to each other and coupled to a secondary coil P_s as shown in Fig. 5. If the goniometer is built so that the mutual inductance between each pair of primary coils and the secondary is proportional to the cosine of the angle that the axis of the secondary coil makes with the axis of the primary coils, then for any position of the secondary the relation between output of secondary and the direction from which waves arrive will be identical in character to the directional characteristics of a single loop, but the orientation of the pattern depends on the position of the goniometer secondary. Thus rotation of this output coil is equivalent electrically to rotating a loop antenna.

The goniometer arrangement permits use of fixed antennas that can be made much too large to rotate. Furthermore, these fixed antennas may be connected to the goniometer coils by relatively long transmission lines, so that a physical separation is possible between the antenna system and the equipment that must be manipulated.

In the design of goniometer systems it is necessary that the two crossed antennas and their associated transmission lines and goniometer field coils all be symmetrical, both mechanically and electrically. It is also very desirable that the goniometer search coil be electrostatically shielded from the field coils. The field coils should be so designed that the maximum coefficient of coupling obtainable between the search coil and a pair of field coils is large, while at the same time the mutual inductance should be as nearly as possible proportional to the cosine of the angle of rotation. This latter result can be approached by proper distribution of the field-coil winding, and any residual deviations can then be taken into account by means of a correction curve or by a mechanical compensating cam.

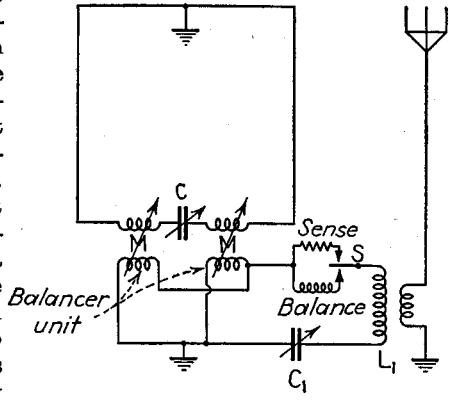


FIG. 6.—Antenna and input circuits of a practical loop direction finder provided with sense determination.

Practical Loop Direction Finders.[1]—The schematic antenna and input circuits of a commercial loop direction-finding system are shown in Fig. 6. In this arrangement the loop antenna is grounded at the center, and antenna effect is balanced inductively by a balancer unit of the type illustrated in Fig. 4*b*. Sense determination is obtained by throwing switch S to "sense," thereby placing a high resistance in series with the antenna input circuit and causing a voltage to be induced in the loop that will cause the normal loop pattern to be distorted to resemble a cardioid. The extent to which a true cardioid is realized can be controlled by adjusting the magnitude of the resistance. The condensers C and C_1 are ganged so that L_1C_1 has approximately the same resonant frequency as the loop circuit. The loop antenna itself is inclosed in an electrostatic shield, and this shield is extended to cover the leads connecting from antenna to receiver, and then terminates in a completely inclosed housing in which the receiver is located. In this way the shielding is fully effective, and the receiver is protected against direct pickup of a signal. Direct pickup is particularly to be avoided because it causes a signal to appear in the receiver output when the loop is in

[1] For further information see Scott, *loc. cit.;* H. B. Martin, Small Vessel Direction Finders, *R.C.A. Rev.*, Vol. 2, p. 69. July, 1937.

the null position and delivers no signal to the receiving set—the equivalent of antenna effect.

A properly balanced and calibrated loop will ordinarily give bearings that are accurate to within one-half to two degrees on near-by radio stations. The accuracy diminishes, however, as the distance to the transmitter increases, is greater during the day than at night (as long as sky-wave propagation is possible), is reduced as the frequency is raised, and is greater when the transmission is over water than when over land. This is because all these factors affect the ratio of ground wave to downward-traveling sky wave. The maximum distance at which satisfactory bearings can be obtained in the daytime is of the order of 50 to 200 miles at frequencies of about 500 kc, and may be as great as several thousand miles at very low radio frequencies. At frequencies greater than 1 to 2 megacycles loop direction finders tend to become increasingly unreliable because of the possibility of sky waves of high intensity being present both day and night up to points quite close to the transmitter.

Loop antennas have been developed for use at ultra-high frequencies, as, for example, 50 megacycles, and when carefully designed can be made to give reasonably accurate results.[1] This is possible because at very high frequencies there is no sky wave, and it is necessary merely that the loop be properly balanced and shielded in order to be usable.

3. Direction Finders Based on Adcock Antennas.—The Adcock antenna in its simplest form consists of two spaced vertical antennas connected as shown in Fig. 7. The action of this arrangement, as far as vertically polarized waves are concerned, is identical with a loop since the resultant current in the output coil of such an Adcock system is proportional to the vector difference of the voltages induced in the two vertical members, exactly as in the case of the loop. Horizontally polarized downcoming waves do not affect the Adcock antenna, however, since the voltages induced in the two horizontal members are of the same magnitude and phase, and so cancel out as the result of the circuit arrangement.

Other forms of Adcock antennas are illustrated in Fig. 9. These are all characterized by the fact that the active members are spaced vertical wires, while the horizontal members are so disposed that either little or no voltage can be induced in them or, alternatively, that voltages induced in them balance out as a result of the circuit arrangement and the fact that the horizontal members are physically close to each other instead of being spaced as in the case of the loop.

The effective height of an Adcock antenna is the same as that of a one-turn loop of corresponding dimensions. Since a loop is not limited to a single turn, the Adcock antenna is definitely inferior to a loop from the point of view of induced signal. This is one of the chief limitations of the Adcock antenna and makes it necessary to use a relatively larger physical structure than necessary with a loop for comparable response. Another disadvantage of the Adcock antenna system is that it represents a source of energy having a relatively high equivalent internal impedance that is capacitive. If the Adcock is to be tuned to resonance, this must be done by means of a relatively high inductance coil that will add appreciable loss to the circuit, while if the antenna system is to be used in a nonresonant manner, then the high source impedance does not lend itself to association in an efficient way with relatively long leads, or with transmission lines. Some help can be obtained by making the vertical members of the Adcock antenna in the form of wire cages of relatively large diameter, but this is only a partial solution of the problem. In contrast with the Adcock antenna system, a loop antenna has an internal impedance consisting of an inductive reactance of moderate value. By the use of a suitable number of turns, the loop can be readily

[1] For example, see H. G. Hopkins, A Loop Direction Finder for Ultra-short Waves, *Wireless Eng.*, Vol. 15, p. 651, December, 1938.

tuned to resonance by means of a variable condenser having negligible loss. If the loop is to be untuned, it can be more readily matched to a transmission line or used with long leads than can the Adcock system.

The value of the Adcock antenna system in direction finding is the greatly reduced susceptibility to polarization errors. Adcock systems are hence especially valuable at short waves, and at ultra-high frequencies, where loops perform least well. Adcock antennas find some use at relatively low frequencies, but here in order to obtain any reasonable amount of energy pickup it is necessary to employ physically large fixed antennas with a goniometer arrangement.

Polarization Errors of Adcock Antenna Systems.—A perfectly symmetrical Adcock antenna system of the type illustrated in Fig. 7 would have zero polarization error if operated in free space remote from all other objects. However, when the antenna system is in the vicinity of the earth, or other equivalent ground plane such as might

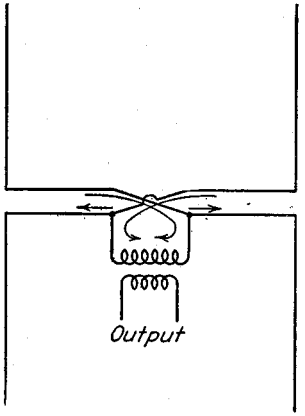


FIG. 7.—Simple Adcock antenna system remote from neighboring objects, particularly the earth.

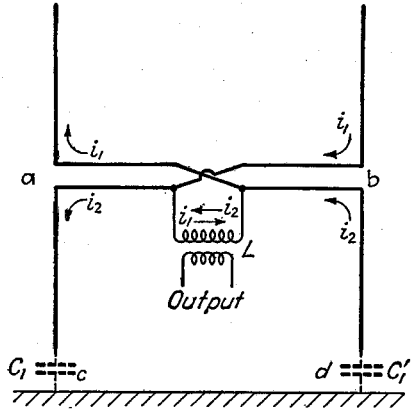


FIG. 8.—Adcock antenna system located close to earth.

be provided by an airplane, polarization errors are present to a greater or less extent. The way in which polarization errors are then introduced in the system can be explained with the aid of Fig. 8. Horizontally polarized downcoming waves induce voltages in the two horizontal members and cause a difference of potential to be produced between points a and b. These voltages cause currents i_1 and i_2 to flow in the vertical members as indicated. When the antenna system is remote from all objects, particularly the earth, these two currents are of equal magnitude, and since they flow through the coil L in opposite directions, no effect is produced in the output. However, when the lower members of the antenna system are close to the ground, as in Fig. 8, the capacities C_1 and C_1' of the lower two legs to ground cause the current i_2 to be greater than i_1 because of the lowered impedance offered to i_2 as a result of the presence of the earth. As a consequence, the currents i_1 and i_2 produced by the downcoming wave no longer balance in the coil L, and there is a residual output. This results in a polarization error that, although substantially smaller than in the case of a loop antenna, is nevertheless by no means negligible.

Types of Adcock Antenna Systems and Their Polarization Errors.[1]—The principles involved in the Adcock antenna system may be applied in a number of ways, of which some of the more practical examples are shown in Fig. 9.

[1] A classic paper on the polarization error of various types of Adcock systems is that of R. H. Bar-

The arrangement of Fig. 9*a* has already been discussed above. The polarization error of this arrangement becomes less the greater the spacing between the earth and the lower limbs, and is also affected by the soil conductivity. The standard wave error will vary accordingly, but in a typical case is of the order of 2 to 6°.

The simple U-type antenna of Fig. 9*b* represents the upper half of the system of Fig. 9*a*. If the earth were a perfect conductor and the horizontal members were extremely close to the earth, the polarization error would be zero. However, the earth is not a perfect conductor and so has a reflection coefficient that is less than unity to downcoming horizontally polarized waves. This results in there being a horizontal electric field at the earth's surface when downcoming waves are present,

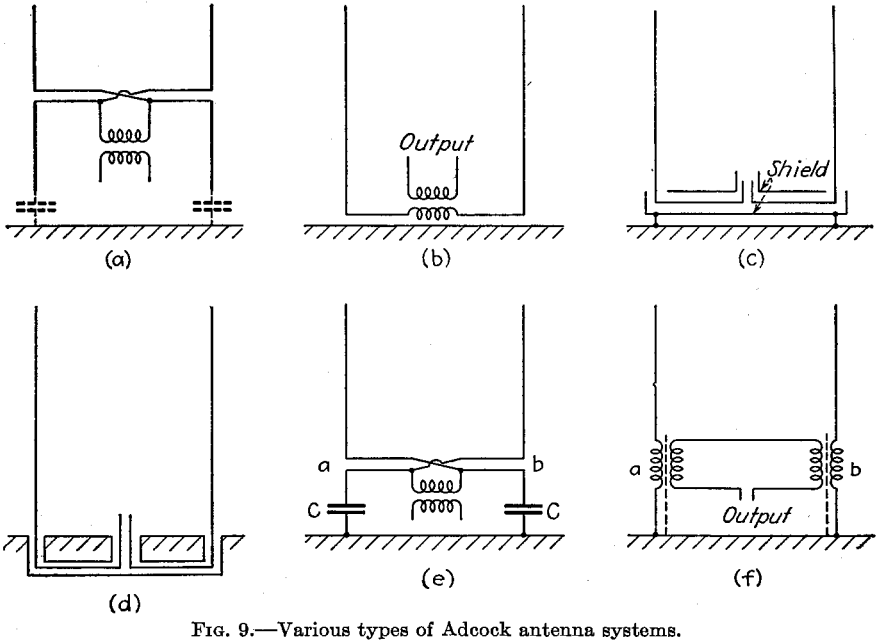


FIG. 9.—Various types of Adcock antenna systems.

which induces a voltage in the horizontal member and gives rise to a polarization error. In a typical case, the standard-wave polarization error to be expected in an arrangement of this sort may easily exceed 10 to 15°.

The polarization error of the simple U-type antenna of Fig. 9*b* can be reduced by placing the horizontal members in a shield grounded at each end as shown in Fig. 9*c*,

field, Some Principles Underlying the Design of Spaced-aerial Direction-finders, *Jour. I.E.E.*, Vol. 76, p. 423, 1935; also, Wireless Section, *I.E.E.*, Vol. 10, p. 55, June, 1935.

Practical examples of Adcock direction-finding systems for various parts of the frequency spectrum, together with data on polarization errors experienced in practice and comparisons between Adcock and loop systems with respect to polarization errors is given by Smith-Rose, *loc. cit.*; R. L. Smith-Rose, Radio Direction-finding on Wavelengths between 2 and 3 Metres (100 and 150 Mc./s), Wireless Section, *I.E.E.*, Vol. 15, p. 161, September, 1940; also, *Jour. I.E.E.*, Vol. 87, p. 154, 1940; R. L. Smith-Rose and H. G. Hopkins, Radio Direction Finding on Wavelengths between 6 and 10 Metres (Frequencies 50 to 30 Mc./Sec.), *Jour. I.E.E.*, Vol. 83, p. 87, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 303, September, 1938; R. H. Barfield and R. A. Fereday, An Improved Medium-wave Adcock Direction-finder, Wireless Section, *I.E.E.*, Vol. 13, p. 33, March, 1938; also, *Jour. I.E.E.*, Vol. 81, p. 676, 1937; R. H. Barfield and W. Ross, A Short-wave Adcock Direction-finder, *Jour. I.E.E.*, Vol. 81, p. 683, 1937; also, Wireless Section, *I.E.E.*, Vol. 13, p. 39, March, 1938; R. H. Barfield, Recent Developments in Direction-finding Apparatus, *Jour. I.E.E.*, Vol. 68, p. 1052, 1930; R. A. Watson Watt, Polarization Errors in Direction Finders, *Wireless Eng.*, Vol. 13, p. 3, January, 1936.

or by burying them as in Fig. 9*d*. The use of a shield will reduce the polarization error to perhaps half of that of the unshielded U type. Burying the horizontal members produces some further reduction in polarization error, with the actual amount of error depending upon the conductivity of the earth, the frequency, and the depth of burial. A wire screen placed over the earth's surface immediately underneath the antenna will help in reducing the error in Fig. 9*d*. The buried arrangement does not eliminate polarization errors completely because the currents the wave induces in the ground penetrate to a depth greater than it is practical to bury the horizontal connections.

The arrangement of Fig. 9*e* is referred to as a balanced system, and can be thought of as the Adcock antenna system of Fig. 9*a* modified by reducing the length of the lower limbs, so that the horizontal members will be closer to ground, and then compensating for the increased impedance of the lower limbs by connecting capacities between them and ground. The condensers *CC* are so selected that the impedance of the upper and lower sections about the dividing points *a* and *b* will be the same. The standard wave error of a typical arrangement of this type is of the order of the corresponding shielded-U system of Fig. 9*c*.

The arrangement in Fig. 9*f* is the same as that of Fig. 9*b* except for the fact that each vertical antenna is separately coupled to the output circuit. In such an arrangement the polarization error is very small because the insertion of mutual inductances between the two vertical sections introduces a very high impedance between points *a* and *b* (see Fig. 9*f*), with the result that the current produced by the voltage induced in the horizontal members becomes extremely small. In fact, if suitable electrostatic shields are used, as shown dotted in Fig. 9*f*, and care is taken to maintain symmetry, it is possible theoretically to reduce the polarization error to zero. Practically, the standard wave error of coupled arrangements of the type of Fig. 9*f* is readily kept to as low as 1° or less.

Miscellaneous Considerations in Adcock Antenna Systems.—The sensitivity of the Adcock antenna system to neighboring objects, and particularly the earth, makes it usually desirable to employ two pairs of Adcock antennas at right angles to each other in a goniometer system, rather than to employ a single rotatable Adcock antenna. The relation of the antenna to these objects then remains fixed.

Adcock antenna systems are subject to vertical antenna pickup just as are loop systems. With Adcock arrangements, antenna effect arises from failure to maintain symmetry between the various vertical antennas and their associated leads and coupling coils. Electrostatic shielding of horizontal members and also the use of electrostatic shields in coupling coils are always helpful in reducing unbalances. It is also sometimes desirable, or even necessary, to inclose the vertical members in electrostatic shields in order to make their capacities to ground independent of neighboring objects, and to minimize the effect of induction electric fields that may be generated by current induced in near-by objects.

Sense determination can be obtained in Adcock systems in the same way as with a loop antenna, *i.e.*, by introducing a certain amount of antenna pickup in a controlled way that changes the normal figure-of-eight pattern to a cardioid pattern.[1]

4. Evaluation of Polarization Error—Standard Wave Error.—The merit of a direction-finding system from the point of view of immunity to polarization error can be conveniently defined in terms of a *standard wave error*. The standard wave

[1] A modified sense-determining arrangement, in which a vertical antenna is substituted for one pair of antennas in a goniometer system, and sense is obtained by comparing the null position for this connection with the null obtained for the usual two pair of antennas, is described by R. A. Fereday, A Sense-finding Device for Use with Spaced-aerial Direction-finders, *Jour. I.E.E.*, Vol. 84, p. 96, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 14, March, 1939. This arrangement is particularly useful at high frequencies.

error is the error in bearing that would be obtained if the incoming wave arrived from an angle of 45° above the horizontal, and had its electric vector inclined at an angle of 45° with respect to the vertical. In calculating the standard wave error, it is always arbitrarily assumed that the resultant voltage induced in the vertical members by the "standard wave" is in the same phase as the resultant voltage induced in the system by the horizontal component of the "standard wave" acting on the horizontal members of the antenna. This actually may not be the case unless the "standard wave" is elliptically polarized, but it greatly simplifies the calculations, and for comparative purposes is perfectly satisfactory.

The standard wave error can be calculated from the geometry of the antenna system, the frequency, and a knowledge of the earth's constants. The intensity of the horizontal electric field produced near the surface of the earth by a downcoming wave—and hence the voltages induced in a horizontal member—depends very greatly upon the reflection coefficient of the earth, and varies greatly with the height measured in wave lengths. The effect of height is especially great when the earth is a good reflector.

The polarization error of a direction-finding antenna system can be determined experimentally by transmitting from a low-power elevated source, and employing a doublet or dipole transmitting antenna that can be rotated to produce waves of any desired polarization. In carrying out such tests it is essential that the distance between transmitter and receiver be not too small. The minimum allowable distance depends upon the antenna dimensions, the wave length, and the inherent accuracy of the direction finder being investigated, and is of the order of 0.6 to 3.5λ.[1] At the higher radio frequencies it is possible to place the transmitter on a high tower and still obtain reasonably large angles of incidence while this requisite distance is maintained.[2] If the frequency is very high, it is even possible to obtain a 45° angle of incidence corresponding to a standard wave. At lower frequencies it is necessary to place the elevated transmitter in an airplane if large angles of incidence are to be obtained with adequate spacing between transmitter and receiver, or alternatively the transmitter must be supported by a balloon or kite.

5. Miscellaneous Direction-finder Systems.[3] *Cathode-ray Direction Finders (Instantaneous Direction Finders).*[4]—A simple form of cathode-ray direction finder giving an instantaneous indication of bearing is shown in Fig. 10*a*. Here the antenna system consists of two loop or Adcock aerials crossed at 90° with respect to each other. The outputs of these aerials are applied to separate radio-frequency amplifiers having identical phase shift and amplification. The outputs of the amplifiers are applied to the horizontal and vertical deflectors, respectively, of a cathode-ray tube. With such an arrangement the effect of the voltages induced in the crossed antennas by a passing radio wave is to produce a line on the cathode-ray tube as shown in Fig. 10*b*, the direction of which indicates the bearing of the signal, while the length depends on the signal strength and so is of no significance.

Sense determination can be obtained in an arrangement of the type shown in Fig. 10*a*, which involves intensity modulating the beam of the cathode-ray tube with a

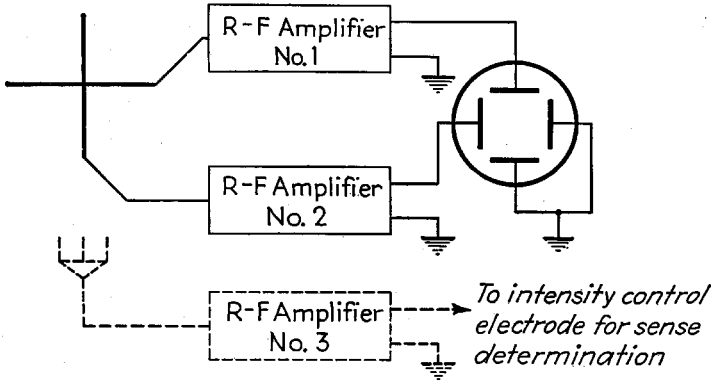
[1] W. Ross, The Calibration of Four-aerial Adcock Direction-finders, *Jour. I.E.E.*, Vol. 85, p. 192, 1939; also, Wireless Section, *I.E.E.*, Vol. 14, p. 229, September, 1939.

[2] R. H. Barfield, An Elevated Transmitter for Testing Direction Finders, *Wireless Eng.*, Vol. 15, p. 495, September, 1938.

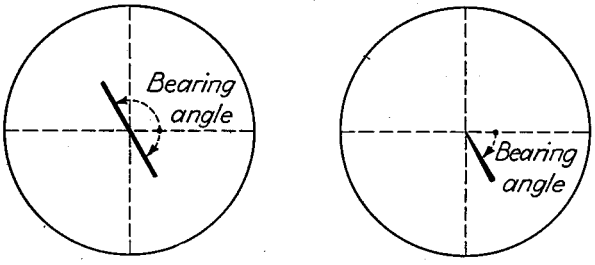
[3] A more complete and also more detailed discussion of various types of direction finders is given by Keen, *loc. cit.*

[4] For further information see R. A. Watson-Watt and J. F. Herd, An Instantaneous Direction Reading Radio Goniometer, *Jour. I.E.E.*, Vol. 64, p. 611, 1926; also, *Exp. Wireless and Wireless Eng.*, Vol. 3, p. 239, April, 1926; G. H. Munro and L. G. H. Huxley, Shipboard Observations with a Cathode-ray Direction-finder between England and Australia, *Jour. I.E.E.*, Vol. 71, p. 488, 1932; also, Wireless Section, *I.E.E.*, Vol. 7, p. 283, September, 1932; The Staff, Radio Research Station, A Short-wave Cathode Ray Direction Finding Receiver, *Wireless Eng.*, Vol. 15, p. 432, August, 1938; H. T. Budenbom, Azimuth Indicator for Flying Fields, *Bell Lab. Rec.*, Vol. 20, p. 58, November, 1941.

radio-frequency voltage derived from a vertical antenna and amplified by a third amplifier, as shown dotted in Fig. 10*a*. This third amplifier must have a phase shift that is identical with that of the first two amplifiers, and its amplification must be such that the amplified signal will be sufficient to produce an appreciable amount of intensity modulation. In addition, a 90° phase shift must be introduced between the voltage induced in the vertical antenna and its amplified output, in order to take into account the fact that the resultant voltage acting around the loop is 90° out of phase with the voltage induced in a vertical antenna.



(a) Schematic circuit



(b) Cathode-ray pattern without sense determination

(c) Cathode-ray pattern with sense determination

FIG. 10.—Instantaneous cathode-ray direction-finder systems.

The effect of intensity modulating the cathode-ray beam is shown in Fig. 10*c*. If the control electrode of the tube is normally biased approximately to cutoff, the effect is to black out one-half of the line of Fig. 10*b* and to cause the intensity of the other half to vary from practically nothing near the origin to maximum intensity near the end of the line, as shown. Since the particular end of the line of Fig. 10*b* that is blacked out by the intensity modulation from the vertical antenna output depends upon the direction from which the waves arrive, a definite sense indication is obtained.

In a system of the type illustrated in Fig. 10*a* it is convenient to employ a superheterodyne system, instead of a simple radio-frequency amplifier. This can be done by employing a common beating oscillator for all three antenna systems, thereby converting the incoming signals to a predetermined intermediate frequency that can

then be amplified and applied to the electrodes of the cathode-ray tube. Such an arrangement takes advantage of the fact that if the frequency of several waves is changed to a new frequency by the heterodyne process, the relative phases of the waves after their frequencies have been changed will not be changed if a common beating oscillator is used for all the waves involved (see Par. 23, Sec. 11).

It is necessary that the various amplifiers used in the system of Fig. 10*a* have identical phase shifts, and it is also necessary that amplifiers 1 and 2 have identical amplification. If amplifiers 1 and 2 have different phase shifts, an ellipse (or a section of an ellipse) will appear on the cathode-ray screen instead of a line (or a section of a line). If the phase shift of amplifier 3 is incorrect, the portion of the line visible on the end of the cathode-ray tube will not have maximum intensity at the point of maximum deflection, and furthermore, the black-out on the end of the line toward the origin will not be clean-cut. If amplifiers 1 and 2 have the same phase shift but unequal amplification, a quadrantal error will be introduced that is maximum for waves arriving at an angle of 45° with respect to the planes of the antennas, but that becomes zero when the arriving wave travels parallel to the plane of either antenna.

A practical method of adjusting amplifiers 1 and 2 to have identical characteristics is to provide in the direction-finding equipment a test oscillator with a loop that will induce in the two crossed antennas voltages of equal magnitude and phase. The phase shift and amplification of one of the amplifiers is then varied until a straight-line pattern is obtained at an angle of 45°. Amplifier 3 can then have its phase adjusted by turning on this amplifier and adjusting phase shift until a satisfactory black-out is obtained. An alternative method of adjusting amplifiers 1 and 2 to equality is to tune in some signal, and then connect the inputs of amplifiers 1 and 2 in parallel across one of the two antennas. When the phase and amplitude characteristics are identical, the bearing indicated will be 45°, and there will be no tendency for the section of line appearing on the cathode-ray screen to open up into an ellipse.

The cathode-ray direction finder giving instantaneous bearing has a number of advantages, and is coming into increasing use. In particular it avoids the necessity of manipulating the direction-finding equipment, and gives an accurate bearing on a signal of brief duration, such as a single dot. Also, the instantaneous direction finder gives a positive indication as to whether or not the signal is present.

Spaced-loop Direction Finders.—In this arrangement the ordinary loop or Adcock antenna system is replaced by two loop antennas spaced some distance apart, as in Fig. 11*a*, and connected in such a manner as to oppose each other. The directional pattern of such an arrangement to vertically polarized waves is the same as that of an ordinary loop, with the null direction in a direction at right angles to the plane of the loops (see Fig. 11*b*). However, unlike the single loop system, horizontally polarized downcoming waves produce no output in such a spaced-loop arrangement when the system is in the null position, because in this position the horizontally polarized downcoming waves induce identical voltages in the two spaced loops and these are balanced out by the connection. The spaced-loop arrangement thus has no night-effect error. This is gained, however, at the expense of reduced energy pickup, and the need of even greater care in design to obtain and maintain symmetry, as compared with an ordinary loop.

A modified form of spaced-loop direction finder has been proposed, in which the loops are as shown in Fig. 11*c*. Such an arrangement has the advantage that the coupling between the antennas and the shield connecting the horizontal members is minimized. The directional pattern of this arrangement for vertically polarized waves is shown by the solid line in Fig. 11*d*, while the dotted line gives the pattern for a horizontally polarized wave coming downward at a sloping angle. It will be noted that the minima in the XX direction are free of polarization errors (night effects), whereas

the minima in the *YY* disappear when night effect is present, which gives a means of identifying the nulls.

Rotating Beacon Direction Finder.—The rotating beacon makes use of a transmitter with which is associated an antenna system that has a well-defined minimum, as, for example, a cardioid pattern. The entire antenna system is then rotated, or a goniometer arrangement is used to cause the polar diagram of the directional pattern

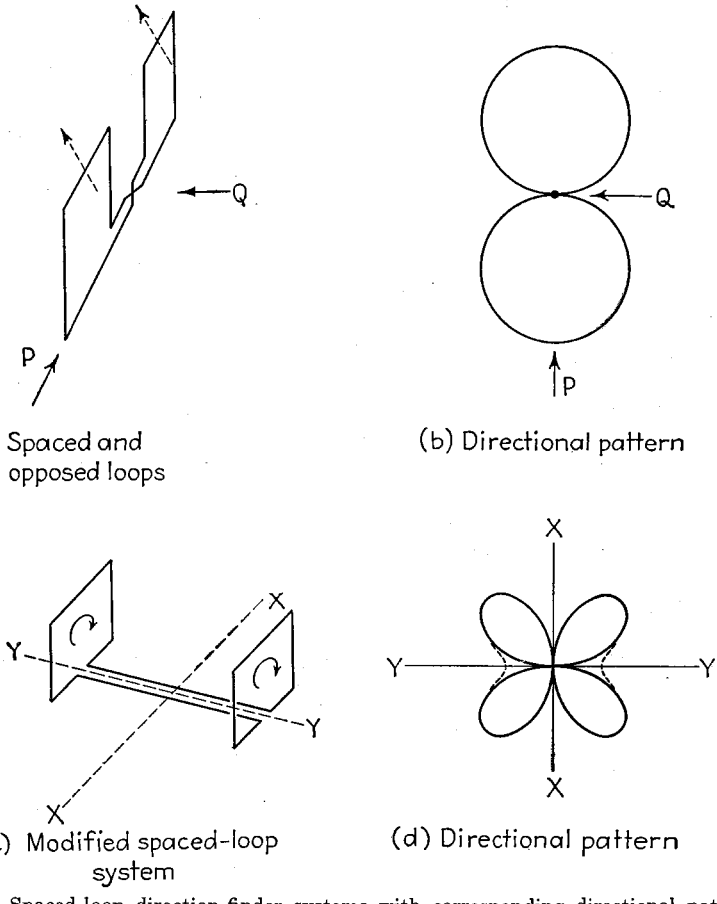


FIG. 11.—Spaced-loop direction-finder systems with corresponding directional patterns.

to rotate, a constant speed being used such that a complete revolution is made in some predetermined time, normally one minute.[1] At the instant that the minimum of the system is directed northward a marker signal in the form of a dash is sent out from a nondirectional antenna.

[1] Aerial systems for rotating beacons and methods of exciting them are described by H. A. Thomas, A Method of Exciting the Aerial System of a Rotating Radio Beacon, *Jour. I.E.E.*, Vol. 77, p. 285, 1935; also, Wireless Section, *I.E.E.*, Vol. 10, p. 256, September, 1935.,

It is also possible to obtain a rotating pattern by setting up two fixed crossed-antenna systems and feeding them with radio-frequency signals modulated at the desired rotation frequency, the modulation of the voltages supplied to the two antenna systems being in phase quadrature. See David G. C. Luck, An Omnidirectional Radio-range System, *R.C.A. Rev.*, Vol. 6, p. 55, July, 1941; p. 344, January, 1942.

Signals from such a rotating beacon can be used to obtain a bearing by starting a stop watch when the marker signal is sent out and stopping the watch when the received signal passes through a minimum. The elapsed time, together with a knowledge of the speed and direction of rotation of the beacon, will then give the bearing.

The rotating-beacon arrangement has the advantages that no special equipment is required at the receiving point in order to obtain a bearing, that there is no possibility of mistakes in bearings, that little skill is required in using the system, and that there are no quadrantal errors and no need of calibration. Night effect is, however, present to the same extent as in the loop arrangements if a loop transmitting aerial is used, and the system requires a very special transmitter.

The principle of the rotating-beacon direction finder can be expanded considerably by utilizing special receiving equipment. Thus by making the rotating beam consist of a limacon (similar to a cardioid, except that there is not sufficient vertical antenna effect to produce a complete null) the receiver output will be an audio wave having a frequency corresponding to the rotational speed of the pattern. By means of a phase splitter it is possible to derive from this single audio wave two waves in quadrature, which can be applied to the vertical and horizontal deflecting electrodes of a cathode-ray tube to give a circular pattern. If now the transmitter winks out its radiation momentarily when the maximum of the pattern is north, then the bearing of the transmitting station as observed from the receiver is indicated by the place on the periphery of the circular pattern on the cathode-ray screen at which the winking out occurs.[1]

Pulse Direction-finding Systems.—The pulse direction finder consists of a direction-finding system modified to operate on transmitted signals that consist of short pulses, such as are used in ionosphere studies. The output indicator of the receiver consists of a cathode-ray tube with a sweep circuit that can be synchronized with the repetition rate of the transmitted pulses.

In such an arrangement the pulses reaching the receiver via the ground wave arrive sooner than the same pulses that have traveled via the ionosphere, and so can be observed separately. The bearing is obtained by manipulating the loop to give a minimum for the ground-wave pulses, while the sky-wave pulses are ignored. In this way night-effect error can be eliminated, since the only downcoming horizontally polarized waves that can be present are contained in the pulse that has traveled via the ionosphere.

Compensated Loop Direction Finder.—The fact that night effect is caused by voltages induced in the horizontal members of the loop by horizontally polarized downcoming waves has led to the suggestion that these voltages causing night effect might be neutralized by coupling into the receiving system the output of a horizontal antenna that rotates with the loop and responds to nothing but horizontally polarized downcoming waves.[2] In order to be effective, such an arrangement must obtain substantially complete neutralization of the effect of horizontally polarized downcoming waves for angles of incidence varying from vertical to very close to the horizontal. This arrangement appears to have some practical possibilities, but the proper evaluation of these will await a more complete investigation than has yet been published.

Homing Devices.—A homing device is a form of direction finder arranged to indicate the orientation of the direction-finding system with respect to the direction

[1] Complete details of a practical system of this type, together with modifications that will make it possible to use the arrangement as a homing device, are described by Luck, *loc. cit.*

[2] F. E. Terman and J. M. Pettit, A Proposal for the Reduction of Polarization Errors in Loop Direction Finders (abstract), *Proc. I.R.E.*, Vol. 28, p. 285, June, 1940; R. H. Barfield, The Performance and Limitations of the Compensated Loop Direction-finder, *Jour. I.E.E.*, Vol. 86, p. 396, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 108, June, 1940.

of travel of the radio waves. The indication may be visual in the form of a zero-center meter, in which no deflection indicates the "on-course" position, while deflections to the left or right indicate deviations to either side of this course. Alternatively, aural indication may be obtained by an interlocked dot-dash (or A-N) arrangement such as is used in radio range systems.

Homing devices find their chief use in guiding aircraft to a particular destination, as, for example, an airfield or a carrier. It is to be noted, however, that an airplane

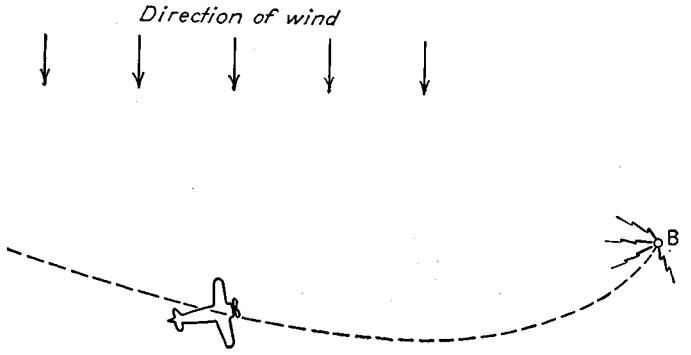


FIG. 12.—Course followed by an airplane homing on a transmitter located at B when there is a strong cross wind. Note that the course followed deviates greatly from the most direct path even though the airplane is always pointing toward B.

guided home by such a device will not fly directly toward its destination if there is a cross wind, but will fly a course such as is illustrated in Fig. 12.

All homing devices that have been developed are based on some form of switched-antenna pattern. The basic principles involved can be explained with the aid of Fig. 13. Here loop and vertical antennas are combined to give a cardioid directional pattern, and the output of the receiver is rectified and passed through a direct-current zero-center galvanometer. The polarity of the loop output that is combined with the

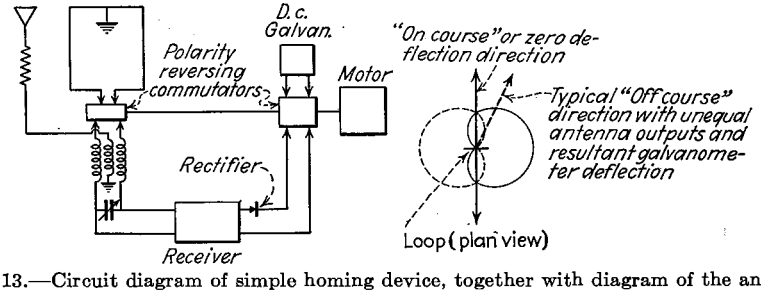


FIG. 13.—Circuit diagram of simple homing device, together with diagram of the antenna directional pattern for the two loop polarities.

vertical-antenna output is continually reversed by means of a commutator, so that the directional pattern of the combined antenna systems is switched back and forth between the solid and dotted cardioids of Fig. 13.[1] A second commutator, synchronous with the first, reverses the terminals of the galvanometer in synchronism

[1] A variation in this antenna system is to use a single loop antenna and switch in such a way as to alter the symmetry of the loop with respect to ground by causing the switching to ground first one side of the loop and then the other. See W. S. Hinman, Jr., A Radio Direction Finder for Use on Aircraft, *Jour. Research Nat. Bur. Standards*, Vol. 11, p. 733, December, 1933.

with the reversals of the loop polarity, so that the galvanometer deflection represents the difference between the receiver output for the two loop polarities. If the plane of the loop is parallel to the direction in which the signals travel, the antenna system output will be the same for both loop polarities. This causes equal d-c currents to pass through the galvanometer alternately in opposite directions, and there will be no deflection. However, if the plane of the loop is not parallel to the bearing of the signal, the output of the antenna system will depend upon the loop polarity, and a resulting d-c current will flow through the galvanometer. This causes a deflection to the right or left, according to which side of the zero direction the signals arrive from. The sense of the bearing obtained in this way can be determined from the fact that, if the loop is deflected to a particular side from the position giving zero output, the direction in which the galvanometer deflects will depend upon the sense of the bearing.

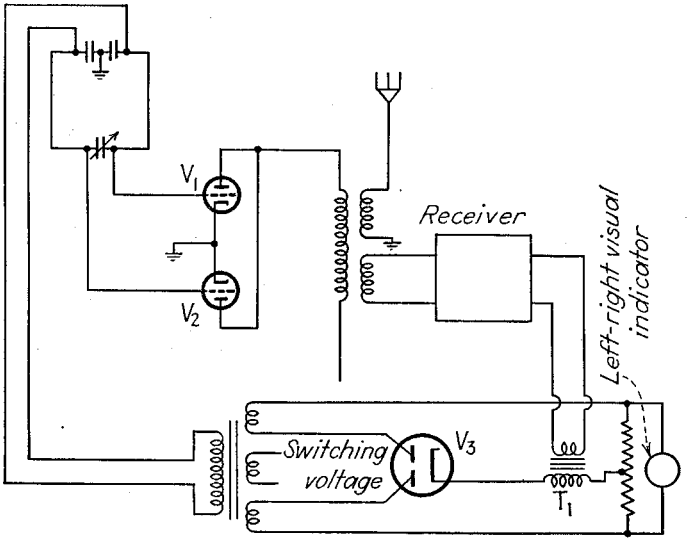


FIG. 14.—Schematic diagram of homing device using electronic switching.

Most homing devices in actual use employ electronic or vacuum-tube switching rather than mechanical switching. A schematic circuit arrangement showing how this is carried out in a typical case is given in Fig. 14. Here tubes V_1 and V_2 are operative on alternate half cycles of the switching voltage. By exciting the inputs in push-pull from the loop, while connecting the outputs of these tubes in parallel and combining with vertical antenna pickup as shown, the end result can be made substantially equivalent to the switched cardioids of Fig. 13. The switching voltage is also applied to the full-wave rectifier V_3, which is so arranged that the output of the radio receiver is inserted in series with the cathode of this rectifier as shown. Visual indication is then obtained by a zero-center meter connected in such a way as to indicate the difference between the rectified currents flowing to the two anodes of tube V_3.

The operation of this output system can be explained briefly as follows. Under "on course" conditions the receiver output is a steady d-c current that, not being transferred by transformer T_1, has no effect on the full-wave rectifier tube V_3. Under these conditions the two anodes of this tube receiver equal d-c currents because

they are equally excited from the switching voltage. However, under "off-course" conditions the output of the receiver pulsates at the switching rate, so that there is a voltage of switching frequency developed in the secondary of transformer T_1. The polarity of this voltage with respect to the switching voltages applied to the anodes of tube V_3 will, however, depend upon whether the course is to the right or left. The d-c currents drawn by the two anodes of V_3 will then be different, and there will be a resultant current through the indicator having a polarity depending upon whether the deviation from the "on course" is to the left or right.

The Automatic Radio Compass.—The term "automatic radio compass" refers to a direction-finding system that sets itself with respect to the passing radio waves without manual manipulation.

The basic principles involved in such an arrangement can be understood by considering the Sperry-RCA radio compass. This makes use of a loop antenna, the output of which is amplified and mixed in a balanced modulator with a locally generated audio frequency, in this case 90 cycles. The output of the balanced modulator contains 90-cycle side bands that will reverse in phase as the antenna system passes through a position of zero response. The modulator output, consisting of 90-cycle side bands minus carrier, is mixed with signal from a vertical antenna, the phase of which has been shifted 90°. This vertical-antenna output thus replaces the carrier of the original signal, but unlike the carrier voltage induced in a loop does not have its polarity reversed as the antenna system is swung through a null. The result is that the 90-cycle modulation envelope of the resulting wave formed by this type of combination of loop and vertical-antenna pickup will have a phase that will reverse as the loop position swings through a null. This modulation envelope is recovered by rectification, and the resulting 90-cycle voltage compared with a 90-cycle component obtained from the modulating source. The comparison is obtained with the aid of an electronic control device operating on a reversible motor, which rotates the loop in such a direction as to bring it to the null position. Whenever the loop is not directly on a null, this action is present and brings the loop to a null, whereupon no 90-cycle component appears in the receiver output and the loop comes to rest. While the loop pattern has two nulls, only one of these provides stable equilibrium. Consequently the 180° ambiguity of the ordinary loop system is avoided.

The automatic radio compass gives a continuous indication of the exact course that one is traveling with respect to a bearing defined by the direction of travel of the radio wave being utilized. In this respect, the automatic compass differs from homing devices, since the latter indicate only whether one is heading along the bearing defined by the wave or is to the right or left of this bearing.

MISCELLANEOUS RADIO AIDS TO AIR NAVIGATION[1]

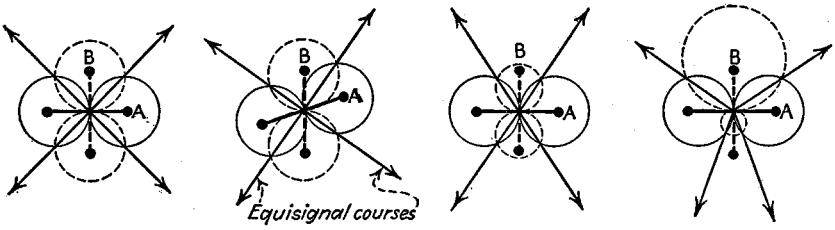
6. The Long-wave Radio Range.—The radio range is a type of radio beacon that lays down a course in a predetermined direction. The characteristics of a typical radio-range transmitter installation are as follows:[2] The antenna system consists of

[1] This division gives merely a brief summary of some of the more important radio aids to air navigation, other than the radio compass, that were in use prior to 1939. This material is presented primarily for the sake of completeness and with the full realization that the many new developments in radio navigation that have been made to meet military needs have rendered much of this material obsolete.

[2] The system in use at the present time is described by W. E. Jackson and D. M. Stuart, Simultaneous Radio Range and Telephone Transmission, *Proc. I.R.E.*, Vol. 25, p. 314, March, 1937. Earlier developments that established the basis of the present system are reported by J. H. Dellinger, H. Diamond, F. W. Dunmore, Development of the Visual Type Airway Radio-beacon System, *Jour. Radio*, Vol. 4, p. 425, March, 1930; *Proc. I.R.E.*, Vol. 18, p. 796, May, 1930; H. J. Walls, The Civil Airways and Their Radio Facilities, *Proc. I.R.E.*, Vol. 17, p. 2141, December, 1929; F. G. Kear and W. E. Jackson, Applying the Radio Range to the Airways, *Proc. I.R.E.*, Vol. 17, p. 2268, December, 1929.

two pairs of crossed Adcock antennas that are 90° in space with respect to each other, together with a central vertical antenna. The transmitter operates in the frequency range 200 to 400 kc and has two independent radio-frequency channels differing in frequency by a convenient audio frequency, commonly 1,020 cycles controlled by matched crystal oscillators having low temperature coefficient plates. The output of one radio-frequency channel is applied to the central tower, while the output of the other channel is switched from one Adcock antenna to the other in a systematic fashion, normally such that the radiation on one Adcock antenna corresponds to the code character for the letter A (– —), while that for the other is the complementary code character, corresponding to the letter N (— –).

When such signals are observed on an ordinary radio receiver, one hears an audio-frequency beat note corresponding to the difference between the frequency of the two r-f channels, normally 1,020 cycles. The intensity of this note is proportional to the intensity of the radiation received from the particular Adcock antenna being



Note: *Dotted curves refer to antenna B*
Solid curves refer to antenna A

(a) Comparison case (b) Angle between A B altered (c) Current in B less than in A (d) Vertical radiation added to B

FIG. 15.—Equisignal courses from an aural radio range under various conditions of adjustment. The solid line represents the directional pattern in the horizontal plane for the Adcock antenna arranged in the direction indicated by *A*, while the dotted line is a pattern in the horizontal plane for the Adcock system oriented in the direction indicated by *B*. The dots at the ends of lines *A* and *B*, and at the center, represent the individual vertical towers comprising the antenna system.

excited at the moment, and will depend upon position, because each Adcock has a marked directional characteristic in the horizontal plane, as shown in Fig. 15*a*. As the second radio-frequency channel is switched from one Adcock antenna to the other, the directional pattern of the antenna system will change from the solid to the dotted pattern as shown. In the particular case where the receiver lies in a direction from the transmitters indicated by one of the arrows in Fig. 15*a*, the intensity of the received beat note is not affected by the switching. These equisignal directions define courses that may be used as an aid to navigation. Any position, either to the right or left of such an equisignal course, will be characterized by a greater intensity of the radiation from one Adcock antenna than from the other. For such an off-course position, the code character corresponding to the Adcock system that produces the strongest signal will predominate, and indicate in which direction the deviation is.

Course Alignment.—The various courses provided by a radio range can be aligned to the actual routes followed by air travel in several ways. First, the whole pattern of Fig. 15*a* may be rotated in azimuth by employing a goniometer arrangement consisting of two primary and two secondary windings. The primary windings are crossed at 90°, as are also the secondary windings. The result is essentially the equivalent of two separate goniometers of the type described in Fig. 5, the combination of one

primary coil with the two secondaries providing one rotatable figure-of-eight pattern, while the other primary coil with the same two secondaries supplies the second figure-of-eight pattern that is crossed with the first.

Second, the individual courses can be shifted by a variety of techniques so that they are no longer symmetrically 90° apart. Thus if the angle between the primary coils of the goniometer differs from 90°, the result is as shown in Fig. 15*b*. Again, if resistances are inserted in the circuit, or other means used so that the power delivered to one of the Adcock antenna systems is less than to the other, the result is as in Fig. 15*c*. Finally, if a phase-shifting network is used to introduce a phase shift between the currents in the two towers comprising one of the Adcock antenna systems, so that these currents are no longer 180° apart, the result is as shown in Fig. 15*d*. A result equivalent to this can likewise be obtained by arranging matters so that the central vertical antenna also receives some excitation from the second radio-frequency channel when this channel is switched to one of the Adcock systems (system *B* in Fig. 15*d*). Still further flexibility in course alignment can be obtained by using a combination of these methods.

Miscellaneous Considerations.—The radio-range installations now employed in the United States provide for broadcasting of weather information without interruption of the radio-range service. This is accomplished by making the power supplied to the central antenna by the first radio-frequency channel somewhat greater than the power supplied to the Adcock antennas and arranging so that this power can be modulated to a reasonable depth, say 70 per cent, by speech. Interference with the speech by the 1,020-cycle beacon signals is eliminated by inserting a band-rejection filter in the speech input circuit to the transmitter modulator, which eliminates speech frequencies in the neighborhood of this frequency. In the aircraft receiver, a band-rejection filter is then employed so that the operator may select only the radio-range signal of 1,020 cycles, or may select another circuit that carries the speech without the radio-range signal, or may obtain at least reasonably satisfactory reception of both speech and the radio-range signal simultaneously.[1]

Night-effect errors[2] appear in radio-range systems, and are due to the same horizontally-polarized downcoming waves that cause night-effect error in radio compass work. The use of crossed-Adcock transmitting antennas in place of crossed-loop antennas that were used earlier greatly reduces the night effect. This is because at the frequencies used in radio-range work the ionosphere produces little or no change in the polarization of the radio waves. Hence the Adcock system gives a performance superior to a crossed-loop system because the Adcock antennas radiate no horizontally polarized waves to begin with. The receiving antenna on the airplane should also be arranged so that it will respond only to vertically polarized waves. This can be accomplished by using a vertical antenna with a symmetrically arranged flat-topped structure. Such a receiving antenna will not respond to horizontally polarized downcoming waves.[3]

The stability with which the courses of a radio range system are maintained in the desired directions is determined by the relative magnitudes and phases of currents in the crossed-Adcock antennas. It is accordingly important that slight changes in antenna reactance caused by weather, ice, etc., have the minimum possible effect.

[1] Further details of this system, which is now in use, are given by Jackson and Stuart, *loc. cit.* The original simultaneous radio-range and radio-telephone system was developed by F. G. Kear and G. H. Wintermute, A Simultaneous Radiotelephone and Visual Range Beacon for the Airways, *Proc. I.R.E.*, Vol. 20, p. 478, March, 1932.

[2] H. Diamond, On the Solution of the Problem of Night Effects with the Radio Range Beacon System, *Proc. I.R.E.*, Vol. 21, p. 808, June, 1933; Haraden Pratt, Apparent Night Variations with Crossed-coil Radio Beacons, *Proc. I.R.E.*, Vol. 16, p. 652, May, 1928.

[3] Further discussion is given by H. Diamond and G. L. Davies, Characteristics of Airplane Antennas for Radio Range Beacon Reception, *Proc. I.R.E.*, Vol. 20, p. 346, February, 1932.

This result can be achieved by employing suitable feeding arrangements.[1] One method is to couple the various vertical radiators to the source of power through low-loss quarter-wave transmission lines. This provides a constant-current transmission system, in which the load currents bear a constant relationship to the sending voltage, irrespective of the load impedance. An alternative arrangement is to connect the two antennas that form an Adcock arrangement in series through a transmission line that is considerably less than a quarter of a wave length long, so that the same current must necessarily flow in each antenna of the pair.

Multiple Courses.—It has been found that in mountainous countries a radio-range system, instead of giving a single equisignal course, sometimes gives several courses. This effect can appear at distances of the order of 30 miles or more from the transmitter, and as many as five or six such courses have been found in some cases. Such multiple courses do not usually point exactly to the transmitter, and may be distributed over an angle of the order of 15°. The tendency for the equisignal course to be accompanied by spurious subsidiary courses is most evident when the main course is more or less parallel with mountain ranges and valleys.

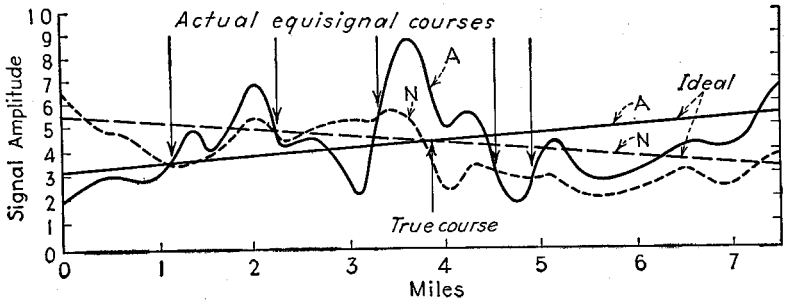


FIG. 16.—Experimentally observed field strength of A and N signals obtained by flying at right angles to the course laid down by a radio-range system under conditions where there are an excessive number of multiple courses.

Multiple courses arise from the fact that over mountainous terrain the resultant signal received at a point in the air is the combination of a direct signal, a sky wave, and a ground-reflected signal, and will vary in an irregular manner according to the circumstances. The details of the combining may furthermore not be exactly the same for the radiations from the two crossed antennas of the radio-range transmitter. The result is that when an airplane flies at right angles to the equisignal course, the field intensity, instead of varying continuously, as shown by the two lines marked "ideal" in Fig. 16, will vary in an irregular manner as shown.[2] The result is then a number of equisignal courses, with resulting confusion to the pilot.

Visual Radio Range.[3]—In the visual radio range the "on-course–off-course" indication is obtained visually rather than by ear. A number of forms of such systems have been devised, but in general all of them are based upon using two modulation frequencies, for example, 65 cycles and 85 cycles. This may be employed in various

[1] For further information see F. G. Kear, Maintaining the Directivity of Antenna Arrays, *Proc. I.R.E.*, Vol. 22, p. 847, July, 1934; Hans Roder, Elimination of Phase Shifts between the Currents in Two Antennas, *Proc. I.R.E.*, Vol. 22, p. 374, March, 1934.

[2] See W. E. Jackson, The Impetus Which Aviation Has Given to the Application of Ultra-high Frequencies, *Proc. I.R.E.*, Vol. 28, p. 49, February, 1940.

[3] For more complete discussions of visual radio-range systems see Kear and Wintermute, *loc. cit.*; J. H. Dellinger, H. Diamond, and F. W. Dunmore, Development of the Visual Type Airway Radio Beacon System, *Proc. I.R.E.*, Vol. 18, p. 796, May, 1930; Haraden Pratt, Field Intensity Characteristics of Double Modulation Type of Directive Radio Beacon, *Proc. I.R.E.*, Vol. 17, p. 873, May, 1929; W. E. Jackson and S. L. Bailey, Development of a Visual Type of Radio Range Having Universal Application to the Airways, *Proc. I.R.E.*, Vol. 18, p. 2059, December, 1930.

ways, one method consisting in using a pair of crossed loops or Adcock antennas, and exciting the loops (or Adcocks) alternately with radio-frequency power, with the carrier power for one loop or Adcock modulated at 65 cycles, and that for the other modulated at 85 cycles. At the receiver the demodulated audio-frequency currents actuate tuned reeds, and the relative intensity of the vibrations of the reeds tuned to 65 and 85 cycles, respectively, indicates the extent to which the equisignal course is being followed.

An alternative visual presentation employs selective circuits to separate the two modulation frequencies, rectifies the outputs separately, and then feeds the rectified outputs differentially to a zero-center meter in such a manner that when an equisignal course is being followed, the two rectified outputs balance and give zero deflection. However, if off the course, one of the currents is stronger than the other, and the meter will deflect either to the left or right, and thereby indicate in which way one is off course.

7. Ultra-high-frequency Radio Range.[1]—Although the established system of radio ranges in this country operates at a relatively low frequency, experimental ultra-high-frequency radio-range systems have been devised, and there are indications these ultra-high-frequency systems will come into practical use.

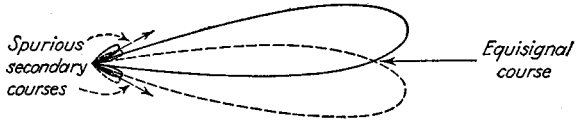


FIG. 17.—Typical antenna patterns that can be used to produce an ultra-high-frequency radio-range signal.

The ultra-high-frequency radio range makes use of the same basic equisignal system as does the low-frequency radio range, but differs in the antenna arrangement used. While it is possible at ultra-high frequency to use crossed figure-of-eight patterns, as illustrated in Fig. 15*a*, it is preferable instead to use antenna arrangements having relatively high directivity in the horizontal plane, such as obtained by employing an antenna array. In this manner it is possible to obtain directional patterns of the general character indicated in Fig. 17. If the transmitter output is switched back and forth between the antenna system having the directional pattern indicated by the solid line and the antenna system having the pattern shown by the dotted line, in the same manner as in the long-wave radio range, then an equisignal course is formed. However, because of the shape of the antenna patterns, the intersection that forms the equisignal course is sharper than for figure-of-eight patterns, causing the course provided by the ultra-high-frequency system to be much more sharply defined. This reduces the width of the course, and also reduces the tendency for spurious courses to be developed.

The most satisfactory ultra-high-frequency systems so far developed employ an antenna array composed of horizontal loop antennas, thus providing pure horizontally polarized waves. Early experimental work on ultra-high-frequency ranges indicated that a frequency of 63 megacycles was superior to a higher frequency, but continued experimentation has resulted in a 125-megacycle system that is superior. It is not altogether unlikely that with further development still higher frequencies will be

[1] The literature on ultra-high-frequency radio-range systems is rather incomplete. The reader interested in the subject will find some help from J. C. Hromada, Development of an Ultra-high-frequency Aural Radio Range (abstract), *Proc. I.R.E.*, Vol. 29, p. 356, June, 1941; E. Kramar and W. Hahnemann, The Ultra-short-wave Guide-ray Beacon and Its Application, *Proc. I.R.E.*, Vol. 26, p. 17, January, 1938; H. A. Chinn, A Radio Range Beacon Free from Night Effects, *Proc. I.R.E.*, Vol. 21, p. 802, June, 1933.

found to be better, particularly since the higher the frequency the easier it is to obtain an antenna array that will concentrate the radiation at a relatively low vertical angle.

One of the chief problems encountered in the development of ultra-high-frequency radio ranges is the avoidance of spurious courses formed by minor lobes of the two antenna patterns intersecting to form secondary equisignal zones. Antenna systems must hence be so designed that the minor lobes have negligible amplitude. It is also important that the angle of elevation of the receiving point with respect to the transmitter have no effect upon the equisignal course, for otherwise a plane flying toward or away from the beacon at constant elevation would have its course progressively shifted because of receiving radiations from steadily varying angles of elevation.

In the practical application of the ultra-high-frequency radio range it is necessary that the directional pattern of the transmitting antennas be stable, and also be such that there be appreciable energy radiated at low vertical angles. Stability can be

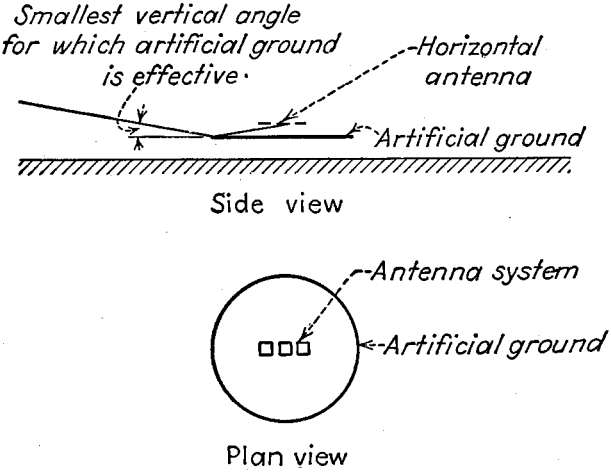


FIG. 18.—Schematic diagram, showing the arrangement of horizontal ultra-high-frequency radio range with artificial ground plane.

achieved by employing a ground plane consisting of a coarse wire mesh mounted some distance above the earth to provide an artificial ground free from influence of snow, vegetation, etc. This plane must be of sufficient extent so that it will be effective for radiation at the lowest angle of elevation that is to be employed (see Fig. 18).

The range of an ultra-high-frequency radio range system is limited by the propagation characteristics of ultra-high frequencies, and cannot be expected to extend greatly beyond the line of sight. The range accordingly will depend upon the height of the airplane in accordance with the principles discussed in Par. 3, Sec. 10. Another factor that influences the effective range is the directional pattern of the transmitting antenna in the vertical plane. This is particularly important in the case of horizontally polarized antennas, because such antennas radiate no energy whatsoever along the horizontal, and have their main lobe at an angle above the horizontal that depends upon the antenna height above ground, and upon the directional characteristics in the vertical plane of the individual antennas comprising the array. For long range it is necessary that the antenna design be such as to concentrate most of the radiated energy at a relatively low angle above the horizontal.

Compared with the low-frequency radio range, the ultra-high-frequency system has the advantage of freedom from polarization errors, freedom from spurious courses,

and a much more sharply defined equisignal zone. Freedom from polarization errors is the result of the fact that the ionosphere plays no part in the propagation of ultra-high-frequency waves. The last two advantages of the ultra-high-frequency system both arise from the fact that antenna patterns much more sharply directional than a figure-of-eight pattern are practical at high frequencies, and this results in a narrower course and less tendency for multiple courses. The chief disadvantage of the ultra-high-frequency system is the fact that it is not effective at great distances for aircraft at low altitude; also, considerable care must be taken in design to avoid the formation of spurious secondary courses as a result of minor lobes in the antenna patterns.

8. Markers.—The radio-range installations in this country are supplemented by markers of various types to increase their usefulness, and to provide additional information for the pilot. Examples are low-frequency marker beacons, "cone-of-silence" markers, and fan markers.

Low-frequency marker beacons are located at intervals along the equisignal course. These stations operate on the same frequency as the radio range with which they are associated, and individual beacons can be distinguished by the coding of their signals. Marker beacons serve to indicate distinctive features of the course, such as the presence of emergency landing fields, changes in topography, cities, etc. They also are used to mark the point where courses from different radio-range systems intersect, and so serve to notify a pilot that he should retune his receiver to the frequency of the radio-range station to be followed during the continuation of his flight. Marker beacons are of relatively low power, and normally have an effective range of only five or ten miles.

The "cone-of-silence" marker beacons are used to indicate the location of a radio-range station. Directly over a low-frequency type of radio-range station there is a small zone of zero signal intensity as a result of the fact that the vertical antenna systems used radiate negligible energy directly upward. While this zone, termed the "cone of silence," can be used to give the pilot the exact location of the radio-range installation, it is very sharp and is difficult to find. The "cone-of-silence" markers consist of a very low-power 75 mc transmitter modulated at 3,000 cycles, exciting a directional antenna array that produces a sharply defined lobe in the vertical plane and has a circular pattern in the horizontal plane. A typical antenna system used to achieve such a result is a simple turnstile mounted one-quarter of a wave length above a large copper screen mounted several feet above the ground to provide a definite ground plane independent of snow, vegetation, etc.[1]

The fan marker is an ultra-high-frequency marker beacon that is characterized by radiating its energy upward in a directional pattern that is fan-shaped, *i.e.*, the beam is relatively narrow in one vertical direction and relatively wide in the vertical plane at right angles to this first vertical plane. Such beacons are used to indicate important divisions in the course provided by a radio range, as, for example, to indicate the region at which the airport control zone begins. The usual fan marker beacon consists of a medium power 75-megacycle transmitter exciting a directional antenna array having the desired pattern. The power required must provide an easily recognized signal up to about 20,000 feet, and the pattern should be such that at 7,000 feet a useful signal some 20 miles wide and a few miles thick is obtained. Such a result can be achieved by a simple broadside array of horizontal half-wave antennas placed a quarter of a wave length above a coarse-mesh screen counterpoise that provides an effective earth plane that is independent of snow and vegetation.

[1] A somewhat improved characteristic can be obtained by an array of turnstiles. See Edmund A. Laport and James B. Knox, Radiating System for 75-megacycle Cone-of-silence Marker, *Proc. I.R.E.*, Vol. 30, p. 26, January, 1942.

9. Absolute Altimeter.—The absolute altimeter, sometimes called a terrain clearance indicator, is a device for enabling an airplane to determine its height above the surrounding terrain independently of the height of this terrain above sea level.

One of the successful devices of this sort is the arrangement shown schematically in Fig. 19.[1] Here the airplane radiates to the earth a wave that is frequency-modulated by a triangular modulating wave. The instantaneous frequency of the transmitted wave accordingly varies as shown schematically by the solid line in Fig. 20. The plane is provided with a receiver and separate receiving antenna so arranged and designed that the energy received directly from the transmitter is not sufficient to block the normal operation of the receiver. The receiver circuits are accordingly subjected to both the direct signal from the transmitter and a reflected signal that has traveled from the transmitter to the ground, and then has been reflected to the plane. This reflected or indirect signal has traveled twice the distance from plane to earth and so has a different instantaneous frequency from the frequency then being radiated by the transmitter. The frequency of the reflected signal reaching the receiving antenna varies according to the dotted line in Fig. 20, which is the solid line displaced horizontally an amount determined by the height of the airplane above ground. As a consequence of this situation the receiver is subjected to two waves of slightly different frequencies, with the difference in frequency dependent upon the height of the airplane above ground. The height is determined by measuring this frequency with a frequency meter that is calibrated in height instead of in cycles. This is done by employing a conventional receiver in which the output of the audio system is fed to a direct-reading frequency meter such as that shown in Par. 26, Sec. 13.

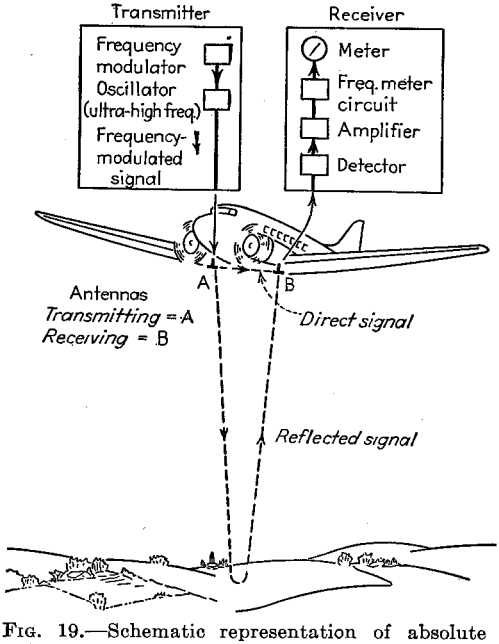


FIG. 19.—Schematic representation of absolute altimeter. (*After Espenschied and Newhouse.*)

The relationship between the height and the frequency f_s fed to the frequency indicating meter, assuming a triangular modulation wave as in Fig. 20, is

$$\left.\begin{matrix}\text{Cycles per second to}\\ \text{frequency meter}\end{matrix}\right\} = f_d = 4F_m\Delta F\left(\frac{\text{height}}{\text{velocity of light}}\right) \qquad (1)$$

where ΔF is the total variation in transmitted frequency and F_m is the frequency of the triangular modulating wave. The minimum height that can be measured is that for which the ratio $f_d/2F_m$ is unity. Any value of height lower than this minimum value gives the same indication as when the height is such as to make $f_d/2F_m$ equal unity.

[1] L. Espenschied and R. C. Newhouse, A Terrain Clearance Indicator, *Bell System Tech. Jour.*, Vol. 18, p. 222, January, 1939; S. Matsuo, A Direct-reading Radio-wave-reflection-type Absolute Altimeter for Aeronautics, *Proc. I.R.E.*, Vol. 26, p. 848, July, 1938.

Absolute altimeters of the type described above must necessarily operate at an extremely high frequency if their minimum height is to be low enough to make the instrument useful in landing operations. For example, if the minimum height that is to be indicated is 10 feet, then Eq. (1) shows that the required frequency sweep Δf must be at least 50 mc. If such a large frequency swing is not to be an excessive percentage of the carrier frequency, it is necessary that the carrier frequency used be 500 mc or greater.

10. Instrument-landing Systems.—An instrument-landing system is a system that makes it possible to land an airplane under conditions of poor or even zero visibility. Such arrangements are often called blind-landing systems.

In order to make an instrument landing or blind landing, an airplane pilot must be provided first with lateral guidance, second with vertical guidance, and third with position "fixes." The lateral guidance enables the plane to be brought down in the middle of the runway. The vertical guidance in association with the position "fixes"

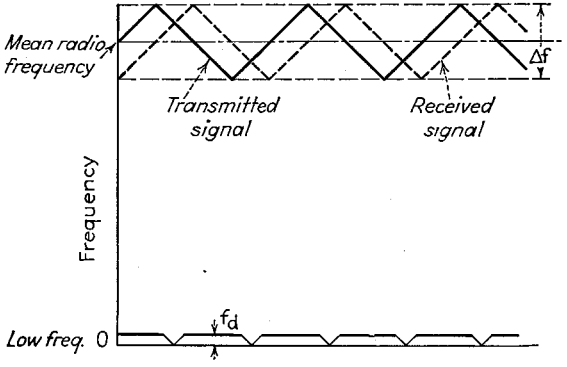


FIG. 20.—Frequency relations in absolute altimeter.

makes it possible to guide the plane at a proper rate of descent to a suitable landing point on the runway.

A variety of blind-landing systems have been devised that incorporate these fundamental elements. The most promising of these that have been described in the literature and tested experimentally are the Lorenz system typified by the installation at Indianapolis, and the MIT-CAA microwave system.[1] These are similar in principle to each other, but differ in the details as to how the basic principles are carried out.

Lorenz System (CAA Installation at Indianapolis).[2]—Lateral guidance is obtained in this installation by an equisignal type ultra-high-frequency beacon, termed a localizer. This is of the visual type, in which field patterns modulated at 90 and 150 cycles are radiated in the manner shown in Fig. 21. This provides an equisignal course where the 90-cycle and 150-cycle modulations are of equal intensity, and this course is aligned accurately with the runway. The power of the localizer in the Indianapolis installation is approximately 300 watts, and the frequency is approximately 110 mc. The antennas consist of an array composed of Alford loops arranged to radiate horizontally polarized waves.

[1] A review of earlier work in the United States is given by W. E. Jackson, Status of Instrument Landing Systems, *Proc. I.R.E.*, Vol. 26, p. 681, June, 1938. A good survey of more recent developments is given in Instrument Landing of Aircraft, *Elec. Eng.*, Vol. 59, p. 495, December, 1940. See also Kramar and Hahneman, *loc. cit.*

[2] W. E. Jackson. P. F. Byrne, A. Alford, H. B. Fischer, The Development of the Civil Aeronautics Authority Instrument Landing System at Indianapolis, *Trans. A.I.E.E.*, Vol. 59, p. 849, 1940; Kramar, *loc. cit.*

Vertical guidance is obtained by means of a second transmitter operating at approximately 94 mc and having a power output of approximately 300 watts. The antenna associated with this transmitter has a directional characteristic such that when an airplane approaches the runway along the path defined by the localizer, it will follow a suitable landing path, provided that the airplane decreases in altitude as the runway is approached in such a manner that the intensity of the glide path signals is kept constant.[1] This result is achieved by placing the transmitting antennas some distance to the side of the localizer path, and so adjusting the directional pattern that the intensity will remain substantially constant along the desired glide path.

In order to utilize this glide path properly, it is necessary that the pilot know when to start a glide. This is accomplished by having the plane approach the glide path along the course defined by the localizer, flying at a predetermined elevation such as 1,500 feet that can be maintained with sufficient accuracy by a sensitive altimeter.

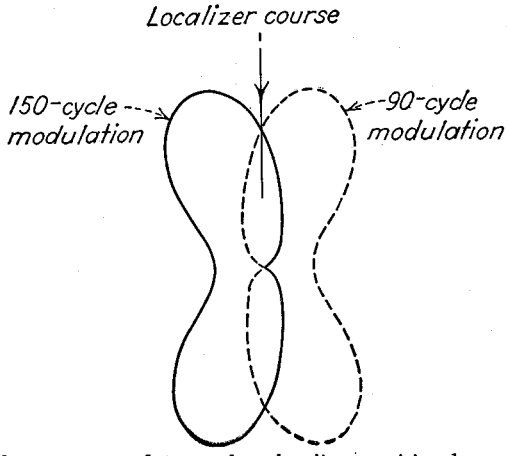


FIG. 21.—Field patterns used to produce localizer equisignal course. The carrier is switched alternately from one antenna system to the other, and the modulation frequency of the carrier depends upon the antenna used to radiate the power at that moment.

When the glide path beam is encountered, the plane then is flown so as to lose altitude at such a rate as to maintain constant intensity of received signal until a landing is accomplished. The marker beacons shown in Fig. 22 serve to indicate the position of the plane at two points along the glide, and provide checks on the accuracy with which the landing is being accomplished. The marker beacons are of the fan type, operating at 75 mc, and each has a power output of 5 watts.

MIT-CAA Microwave System.[2]—In this system the operating frequencies are in the neighborhood of 750 mc. Sectoral horn antenna systems are employed, and because of the high frequency, relatively sharp beams can be obtained. The localizer consists of two such horns arranged to point at a slight angle with respect to each other in the horizontal plane. In the experimental equipment as demonstrated, these two horns were excited from separate transmitters operated at slightly different carrier frequencies, and were modulated at 90 and 150 cycles, respectively, to provide an equisignal localizer path.

[1] The idea of making the desired glide path a path of constant signal intensity is due to H. Diamond and F. W. Dunmore, A Radio Beacon and Receiving System for Blind Landing of Aircraft, *Proc. I.R.E.*, Vol. 19, p. 585, April, 1931.

[2] E. L. Bowles, W. L. Barrow, W. M. Hall, F D. Lewis, D. E. Kerr, The CAA-MIT Microwave Instrument Landing System, *Trans. A.I.E.E.*, Vol. 59, p. 859, 1940.

A glide path is established by means of a second pair of horns arranged in an analogous manner to provide an equisignal zone in the vertical plane. It is possible in this way to obtain a glide path that is essentially a straight line. Furthermore, the glide path is obtained in a positive way by an equisignal indication instead of being dependent upon the ability to obtain and maintain a proper field distribution of intensity in the antenna pattern with respect to both distance above the ground and position along the glide path.

The microwave system employs marker beacons similar to those of the Lorenz system.

Instruments on the Plane.—The indications obtained from an instrument-landing system may be presented to the pilot in a variety of ways. Furthermore, the indications obtained from the instrument-landing system may be combined with information obtained from other sources, as, for example, the bank-and-climb gyroscope and the directional gyroscope, or the radio compass. The possible combinations are very great, and inasmuch as no preferred system has yet evolved, no attempt is made to describe here the various systems that have been used.[1]

Problems Peculiar to Instrument Landing Systems.—In an instrument landing system there are several problems that present particular difficulty. In the first place,

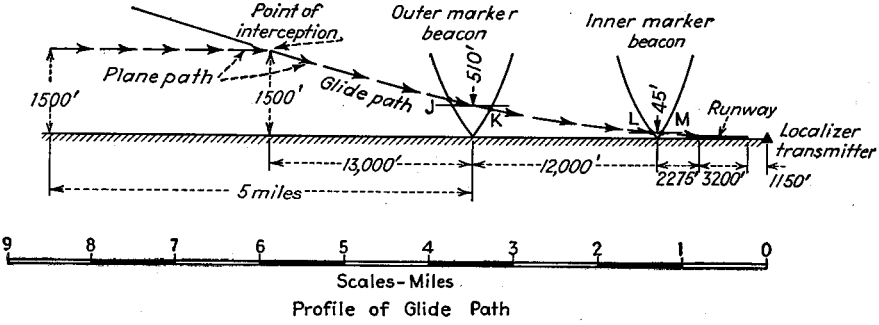


FIG. 22.—Path followed by airplane making a blind landing at the Indianapolis installation.

it is necessary that the directional patterns of the antenna systems used in connection with localizers, and also the antennas used to obtain vertical guidance, be such that no spurious courses arise. This requires that the minor lobes of the antenna patterns be extremely small, or, if possible, entirely absent, and that care be taken so that buildings, wires, etc., in the area do not distort the pattern and create either spurious courses, or "bumps" in the desired course. It is also extremely important that the antenna pattern be stable with respect to weather conditions, etc. This makes the use of horizontally polarized waves desirable.

One of the greatest difficulties in connection with blind-landing systems is that of obtaining a suitably shaped glide path. This path should preferably follow a straight line intersecting the ground at a very small angle. The ideal glide path is not achieved either in the Lorenz system, operating at 100 mc, or in the MIT-CAA system, at 750 mc. The latter achieves a straight-line path, but is at relatively large angle with respect to the earth. The former has a curved path that intersects the runway at a relatively small angle. As a result, these systems and other similar systems that have been devised, from the practical point of view, appear to be useful primarily as blind approach rather than as blind-landing systems.

[1] The interested reader is referred to the following articles that survey some of the more common combinations: Instrument Landing of Aircraft, *Elec. Eng.*, Vol. 59, p. 495, December, 1941; Bowles, Barrow, Hull, Lewis, and Kerr, *loc. cit.*; P. R. Bassett and J. Lyman, Flightray Multiple Instrument Indicator, *Jour. Aero. Sci.*, Vol. 7, p. 199, March, 1940.

SECTION 13

MEASUREMENTS

CIRCUIT CONSTANTS AT LOW FREQUENCIES

1. Direct-current Resistance.—Resistance to direct current can be measured by the voltmeter-ammeter method, on a Wheatstone bridge, or with an ohmmeter.

In the voltmeter-ammeter method, the applied voltage and the resulting currents are measured with ordinary direct-current instruments. In order to minimize errors

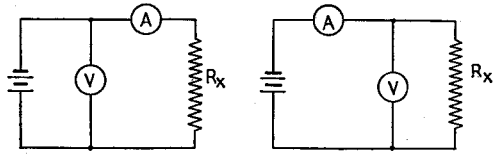


FIG. 1.—Circuit arrangements for measuring resistance by the voltmeter-ammeter method.

resulting from the power consumed by the measuring instruments, the arrangement shown in Fig. 1*a* should be employed when the current drawn by the voltmeter is not negligible as compared with the current in the unknown resistance, while the circuit of Fig. 1*b* is used where the voltage drop in the ammeter is not negligible compared with the applied voltage.

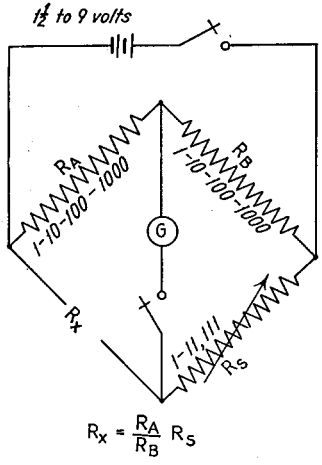


FIG. 2.—Circuit diagram and constants of typical Wheatstone bridge.

Wheatstone Bridge.—The Wheatstone bridge is the standard means used in the accurate determination of resistance. A typical bridge circuit with constants is shown in Fig. 2. In using such a bridge, the ratio R_A/R_B should be so chosen that the unknown resistance is determined to the full number of significant figures available, which is usually four. Also, the magnitudes of resistances used in arms A and B to obtain the desired ratio should be so chosen as to give maximum sensitivity. The proper values depend on the resistance being measured and are usually stated in the instructions accompanying the bridge.

Special difficulties are encountered in bridge measurements of very low and very high resistances. With low resistances the uncertainty introduced by the resistance of leads and contacts can be eliminated by using the Kelvin double bridge, a description of which is found in any book on electrical measurements.

When the resistance being measured is very high, the bridge galvanometer becomes a relatively insensitive indicator of balance. This is because of the high source impedance that the bridge presents to the galvanometer under the conditions existing

in the measurement of very high resistances. This difficulty can be overcome by applying a relatively large d-c voltage to the bridge and using a d-c vacuum-tube voltmeter to indicate balance.

Ohmmeters.—Ohmmeters are suitable for making approximate measurements of resistance, and are widely used in the servicing of communication equipment. Possible circuit arrangements are shown in Fig. 3. To operate, an initial adjustment is made by short-circuiting the terminals XX and adjusting the resistance R until full scale reading is obtained on the milliammeter. When a resistance is then inserted between the terminals XX, the reading will be less than full scale by an amount that depends upon the inserted resistance, and the scale can be calibrated directly in ohms. The arrangement of Fig. 3*a* operates on the assumption that the battery generates a constant voltage during its life, but that the internal resistance of the battery increases with age. The arrangement of Fig. 3*b* operates on the theory that as the battery ages its voltage drops, but the internal resistance remains the same. Although neither assumption is realized completely, ohmmeters built in either way are satisfactory for test work.

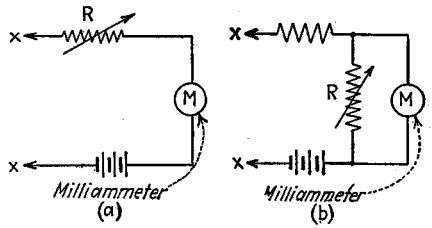


FIG. 3.—Ohmmeters used in servicing communication equipment.

2. Bridge Measurements of Inductance, Capacity, Resistance, and Mutual Inductance at Audio Frequencies.—The alternating-current bridge is the most widely used means of measuring circuit constants at audio frequencies. The schematic diagram of such a bridge is shown in Fig. 4. This is similar to the direct-current bridge of Fig. 2, except that the power source is now an oscillator instead of a battery, and the galvanometer for detecting balance is usually replaced by telephone receivers.[1]

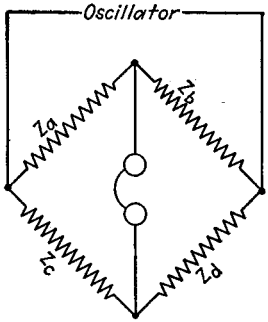


FIG. 4.—Schematic diagram of alternating-current bridge.

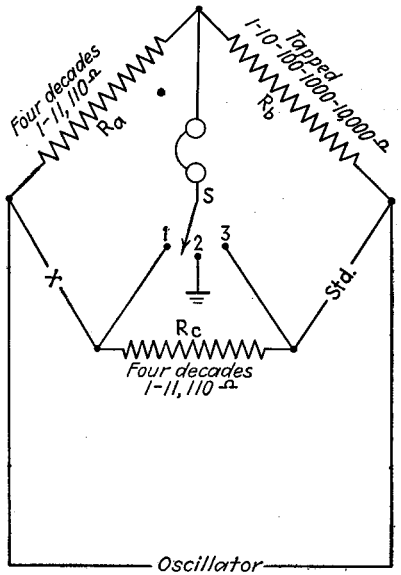


FIG. 5.—Circuit details of a typical general-purpose bridge.

[1] Other means of detecting balance involve the use of amplifier and telephone receivers or amplifier, rectifier, and d-c microammeter; see H. M. Turner, Microammeter Indicator of High-frequency Bridge Balance, *Trans. A.I.E.E.*, Vol. 46, p. 559, 1927. A cathode-ray balance indicator has also been developed that indicates separately the state of the resistance and reactive balances: see H. W. Lamson,

When such a bridge is adjusted so that no current flows through the telephone receivers, then $Z_a/Z_b = Z_c/Z_d$, where Z_a, Z_b, Z_c, and Z_d are now vector quantities.

Common Types of Bridges.—The most widely used bridge is the general-purpose bridge of Fig. 5, in which the two resistance arms R_a and R_b establish an impedance ratio, while the unknown impedance Z_x is measured in terms of this ratio and a standard impedance Z_s of the same kind. An adjustable resistance R_c that can be switched in series with either Z_x or Z_s is provided to balance the power factors of these two arms when reactances are being measured, or to serve as the standard when an unknown resistance is to be determined. The bridge can be operated either with an adjustable standard Z_s, in which case the ratio R_a/R_b is maintained at a constant value, or with a fixed standard, in which case the bridge ratio R_a/R_b is adjusted to be the same as the ratio of impedances Z_x/Z_s.[1]

Other bridge arrangements that find appreciable use in communication work are shown in Fig. 6. The resonance bridge is a special form of the general-purpose bridge, in which the reactances are all concentrated in one arm and adjusted to give series resonance so that this arm offers a resistance impedance. The resonance bridge can be used to measure frequency in terms of inductance and capacity, and is also sometimes used to measure capacity in terms of frequency and a variable inductance, or inductance in terms of frequency and a variable capacity. The Hay bridge compares an inductance with a capacity, and finds its principal use in the measurement of incremental inductance. The Owen bridge also measures inductance in terms of capacity,[2] but in this bridge the arms are so arranged that the balance equations for both resistance and inductance are independent of frequency, provided that the losses in the condenser C_b are negligible. The Maxwell bridge likewise compares an inductance with a capacity. The balance equations do not involve frequency, and the resistance and the reactance balances can be made independent if C_b is continuously variable. Wien's bridge measures capacity in terms of resistance and frequency. This bridge finds use for measuring frequency in the audio range, and is also useful in the precision determination of capacity, since the standards of frequency and resistance are known to very great accuracy.[3] The Schering bridge is of importance because an ordinary general-purpose bridge becomes a Schering bridge when used to measure direct capacity. The Schering bridge is also the standard device for determining dielectric losses in high-voltage cables.

Wagner Ground.—When a bridge is used to measure high impedances, stray capacities to ground will normally cause spurious currents to flow through the telephone receiver, thereby causing an incorrect balance. In the case of the general-purpose and Wien bridges, this difficulty can be eliminated by the use of the Wagner ground, as shown in Fig. 7. This consists of a potentiometer having a total resistance of the order of 500 to 1,000 ohms connected across the oscillator, with the adjustable tap grounded. Where bridges are to be operated at high frequencies, it is also necessary to shunt the potentiometer with a double stator condenser C.

The adjustment procedure for a bridge equipped with a Wagner ground is as follows: (1) First, adjust the bridge as well as possible in the normal manner; (2) switch one terminal of the telephone receivers to ground, and adjust the Wagner earth

---

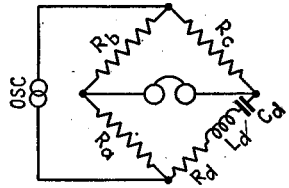
Electronic Null Detector for Impedance Bridges, *Rev. Sci. Instruments*, Vol. 9, p. 272, September, 1938; also, *Gen. Rad. Exp.*, April, 1939.

[1] When the bridge is balanced by adjusting the ratio, the resistance and reactance balances are no longer independent. Under these conditions it can be shown that if the impedance being measured has a high Q, it is difficult to determine the resistance component accurately, whereas if the Q is low, then it becomes difficult to determine the reactive component with accuracy.

[2] J. G. Ferguson, Measurement of Inductance by the Shielded Owen Bridge, *Bell System Tech. Jour.*, Vol. 6, p. 375, July, 1927.

[3] J. G. Ferguson and B. W. Bartlett, The Measurement of Capacitance in Terms of Resistance and Frequency, *Bell System Tech. Jour.*, Vol. 7, p. 420, July, 1928.

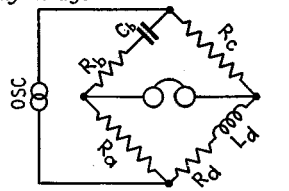
(a) *Resonance bridge*



$$\omega L = \frac{1}{\omega C}$$

$$R_d = \frac{R_a}{R_b} R_c$$

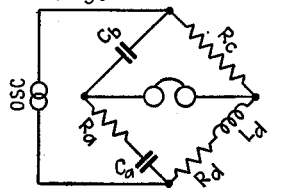
(b) *Hay bridge*



$$L_d = \frac{R_a R_c C_b}{1+(R_b \omega C_b)^2}$$

$$R_d = \frac{R_a R_b R_c (\omega C_b)^2}{1+(R_b \omega C_b)^2}$$

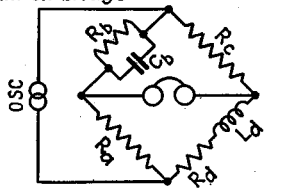
(c) *Owen bridge*



$$L_d = R_a R_c C_b$$

$$R_d = \frac{C_b}{C_a} R_c$$

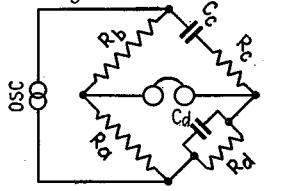
(d) *Maxwell bridge*



$$L_d = R_a R_c C_b$$

$$R_d = \frac{R_a}{R_b} R_c$$

(e) *Wien bridge*

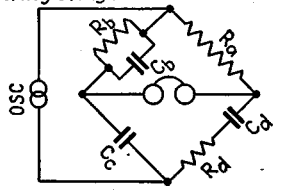


$$\omega^2 = \frac{1}{R_d R_c C_d C_c} \text{ and } \frac{C_d}{C_c} = \frac{R_b}{R_a} - \frac{R_c}{R_d}$$

or

$$C_d^2 = \frac{R_b R_d - R_a R_c}{R_a R_d^2 R_c \omega^2} \text{ and } C_c^2 = \frac{R_a}{(R_b R_d - R_a R_c) R_c \omega^2}$$

(f) *Schering bridge*



$$C_d = \frac{R_b}{R_a} C_c$$

$$R_d = \frac{C_b}{C_c} R_a$$

FIG. 6.—Miscellaneous types of bridge circuits.

for minimum response; (3) switch the telephone receivers back across the bridge output, and check the balance. In difficult cases, it may be necessary to repeat the cycle of adjustment several times

When the Wagner earth is balanced, the impedance ratio Z'_a/Z'_b is the same as the bridge ratio Z_a/Z_b, thus placing the telephone receivers at ground potential. Capacities from corners 2 and 4 of the bridge to ground then have no effect, because there is no potential between these points and ground, while currents flowing to ground through the capacities from corners 1 and 3 of the bridge complete their circuit back to the oscillator through the Wagner ground connection, and so do not affect the telephone receivers.

Special Considerations in Inductance Measurements.—When large inductances are to be measured, the distributed capacity is often sufficient to make the apparent inductance and resistance, as viewed from the terminals of the coil, differ appreciably from the true values. The exact relations existing in any particular case can be calculated by treating the coil and its distributed capacity as a parallel resonant circuit. An approximate relation, sufficiently accurate for all practical purposes, provided that the coil Q is not excessively low and the frequency of measurement does not exceed 80 per cent of the frequency of self-resonance, is

$$\left.\begin{array}{l}\text{Apparent inductance of coil}\\ \text{with distributed capacity}\end{array}\right\} = \frac{L}{1-m^2} \qquad (1a)$$

$$\left.\begin{array}{l}\text{Apparent resistance of coil}\\ \text{with distributed capacity}\end{array}\right\} = \frac{R}{(1-m^2)^2} \qquad (1b)$$

where L and R are the true inductance and resistance, respectively, and m is the ratio of actual frequency to the frequency at which the distributed capacity is in resonance with the coil inductance. Substitution in Eq. (1) shows that, if the error in inductance is to be kept below 1 per cent, measurements must be made at less than one-tenth of the natural resonant frequency of the coil ($m = 0.1$), or corrections calculated by Eq. (1) must be applied.

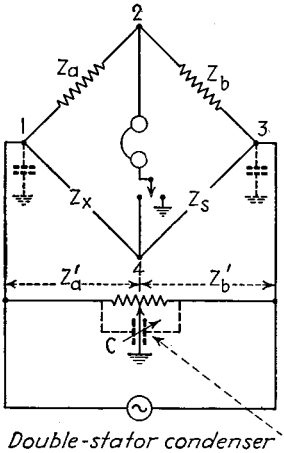


FIG. 7.—General-purpose bridge with Wagner ground.

The accurate measurement of small inductances such as the short-wave coils of a receiver is difficult, because the residual inductance of the bridge wiring, bridge resistances, and leads is not negligible compared with the inductance being measured. The best accuracy is obtained by a substitution method, using as standards a variable inductance and an inductively compensated variable resistance, as illustrated in Fig. 8. The bridge is first balanced with the unknown short-circuited, after which the unknown is placed in the circuit and the variable inductance and inductively compensated resistance readjusted to give balance. The constants of the unknown are then given by the changes in L_s and R_s. It is possible in this way to measure inductance to 0.1 μh or to 0.1 per cent, whichever is the larger.

In measuring very large inductances, it is usually most convenient to employ a type of bridge that evaluates inductance in terms of resistance and capacity. In this way the necessity of an expensive standard inductance of large size is avoided.

Incremental inductance measurements are discussed below.

Mutual-inductance Measurements.—The mutual inductance existing between two coils is commonly determined by connecting the coils in series, measuring the total inductance that results, reversing the terminals of one of the coils, and then measuring

the inductance again. The mutual inductance is one-fourth of the difference of the two measured inductances. In the case of very small mutual inductances, this procedure does not yield accurate results, because it then involves a small difference between two nearly equal quantities. A more accurate determination in this case can be obtained by using a calibrated mutual inductance in the Felici mutual-inductance balance, or by employing the Carey Foster bridge, as shown in Fig. 9.

Special Problems in Capacity Measurements.—The measurement of small capacities, particularly values less than 1,000 $\mu\mu f$, presents a special problem, because the stray capacities associated with the bridge are not negligible in comparison. The most common method of measuring small capacities is the substitution method, using a variable capacity standard, as illustrated in Fig. 10. Here an ordinary general-purpose bridge is used for capacity measurement, with any convenient fixed capacity used for C_c and with variable standard condenser C_s in parallel with the unknown C_x. A balance is first obtained with C_x connected, using a bridge ratio such that the required standard capacity C_s is small, after which C_x is disconnected and C_s increased until balance is regained. The unknown capacity C_x is then obviously the increment in the standard capacity. *The results obtained in this way are independent of bridge errors, since the bridge conditions do not vary.* The accuracy depends only upon the calibration of the standard condenser. The equivalent series resistance of the

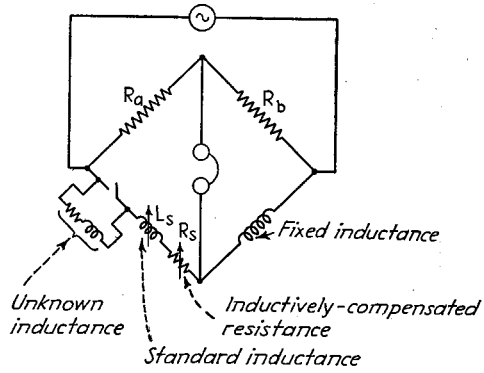


FIG. 8.—Substitution measurement of small inductances using a general-purpose bridge with a variable inductance and an inductively compensated resistance as standards.

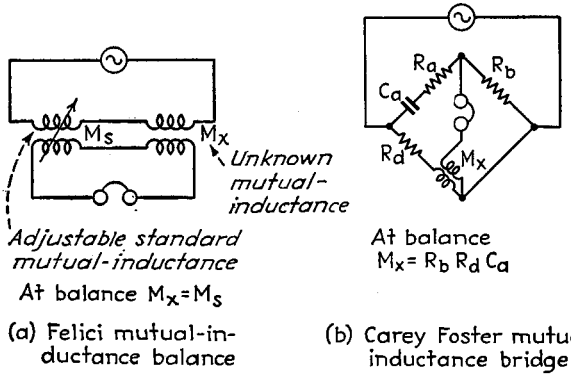


FIG. 9.—Circuits for measuring mutual inductance.

unknown capacity can be determined by using a standard condenser that has audio-frequency losses that are independent of the condenser adjustment.[1] The auxiliary resistance R_s in Fig. 10 is first set to zero in a preliminary balance when both C_s and C_x are present, after which R_s is then adjusted to compensate for the losses removed

[1] Air dielectric condensers satisfy this requirement at audio and moderate radio frequencies (see Par. 32, Sec. 2).

when C_x is disconnected. Then

$$\left.\begin{matrix}\text{Equivalent series resistance}\\ \text{of unknown capacity } C_x\end{matrix}\right\} = \left(\frac{C_s}{C_x}\right)^2 R_s \qquad (2)$$

where C_s is the capacity of the standard *after* disconnecting the capacity C_x, and R_s is the resistance that must be added in series with C_s to compensate for the reduction in losses that results when C_x is disconnected.

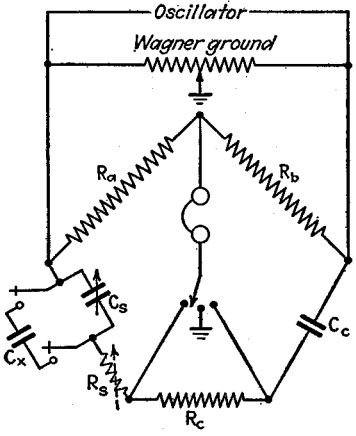


FIG. 10.—Capacity measurements with general-purpose bridge, using substitution method.

Accurate measurements of small capacities can be made with C_x in one arm of the bridge and an adjustable standard C_s in another arm by using a shielded bridge such as is described in connection with Fig. 22.[1] This has the advantage of requiring only a single balance, but requires greater care and tends to give lower accuracy than the substitution method.

The *direct capacity*[2] existing between two electrodes must frequently be measured when there are other capacities associated with the same electrodes. An example is the grid-plate capacity of a vacuum tube, which is associated with capacities from grid to ground, plate to ground, etc. The direct capacity between two electrodes can be measured by the substitution method, using an ordinary general-purpose bridge provided with a Wagner ground. It is merely necessary to return all the remaining electrodes to the earth connection of the Wagner ground, as illustrated in Fig. 11*a*. These additional capacities then become associated with the bridge in such a way as to have no effect upon the measurements. An alternative arrangement is to return all the

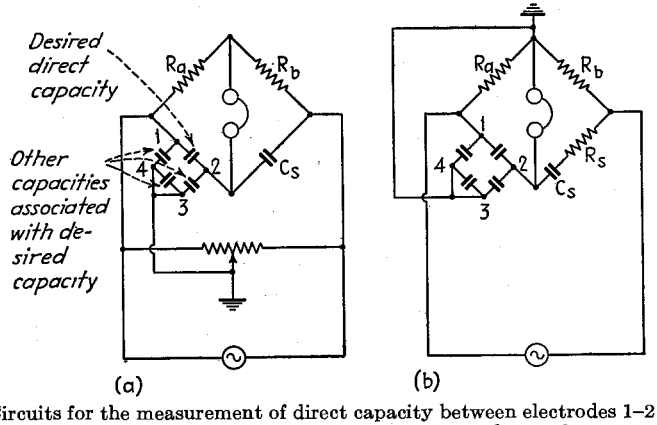


FIG. 11.—Circuits for the measurement of direct capacity between electrodes 1–2 when this capacity is associated with capacity to other electrodes, such as 3–4.

unused electrodes to the junction of the ratio arms, which is then grounded as

[1] Further discussion of the measurement of extremely small capacities is given in Par. 10.

[2] A discussion of the concept of direct capacity is given by George A. Campbell, Direct Capacity Measurement, *Bell System Tech. Jour.*, Vol. 1, p. 18, July, 1922.

in Fig. 11*b*.[1] This leads to a Schering bridge, and the desired direct capacity can be determined either by substitution or in terms of the capacity C_s. The capacity C_{14} in Fig. 11*b* that is shunted across the resistance arm R_a when the unused electrodes are grounded affects only the setting of the resistance R_s, which can no longer be used to give the losses.

Input and Output Impedances of Bridges.—The input and output impedances of a bridge determine the impedance with which the input and output transformers, respectively, must be designed to operate. It is necessary to consider these impedances only for the condition of balance.

The input impedance of the bridge of Fig. 4, when balance is assumed, is the impedance formed by $Z_a + Z_b$ in parallel with the impedance $Z_c + Z_d$. This follows from the fact that since at balance there is no current through the output diagonal, this can be considered as an open circuit.

In a similar manner, the output impedance of the bridge, *i.e.*, the equivalent impedance of the source supplying energy to the telephone receivers or other indicating device, is the impedance $Z_a + Z_c$ in parallel with the impedance $Z_b + Z_d$.

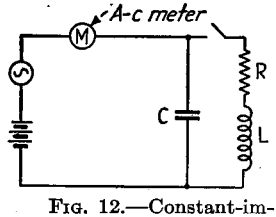


FIG. 12.—Constant-impedance method of measuring large inductances.

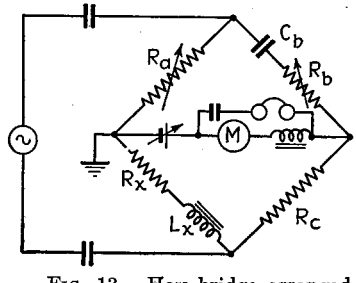


FIG. 13.—Hay bridge arranged for the measurement of incremental inductance.

3. Incremental Inductance.—A simple method of measuring incremental inductance with a minimum of equipment is shown in Fig. 12.[2] Here an adjustable capacity is placed in parallel with the unknown inductance, which is supplied with the desired d-c current and alternating current by sources as shown. The condenser C is then adjusted until the alternating current registered by the meter M is the same with switch S open as with it closed. It can then be shown that

$$\omega L = \frac{1}{2}\left(\frac{1}{\omega C}\right) \tag{3}$$

where $\omega/2\pi$ is the frequency of the supply. This relation holds irrespective of the resistance of the coil.

Incremental inductance can be measured on a bridge, the most satisfactory arrangement being the Hay bridge shown in Fig. 13.[3] Balance can be obtained with the aid of resistances R_a and R_b, and without affecting the direct current in the inductance. The alternating magnetization can be calculated from a knowledge

[1] See also Lincoln Walsh, Direct-capacity Bridge for Vacuum-tube Measurements, *Proc. I.R.E.*, Vol. 16, p. 482, April, 1928; E. T. Hoch, A Bridge Method for the Measurement of Inter-electrode Admittance in Vacuum Tubes, *Proc. I.R.E.*, Vol. 16, p. 487, April, 1928.

[2] H. M. Turner, The Constant Impedance Method for Measuring Inductance of Choke Coils, *Proc. I.R.E.*, Vol. 15, p. 1559, November, 1928.

[3] Constructional details of such a bridge are given by F. E. Terman, "Measurements in Radio Engineering," 1st ed., pp. 55–75, McGraw-Hill, New York, 1935; see also V. D. Landon, A Bridge Circuit for Measuring the Inductance of Coils While Passing Direct Current, *Proc. I.R.E.*, Vol. 16, p. 1771, December, 1928.

of the inductance, and the alternating voltage developed across it as observed by vacuum-tube voltmeter. At balance

$$L_x = \frac{R_a R_c C_b}{1 + (R_b \omega C_b)^2} \tag{4a}$$

$$= \frac{R_a R_c C_b}{1 + \frac{I}{Q^2}}$$

$$R_x = \frac{R_a R_b R_c (\omega C_b)^2}{1 + (R_b \omega C_b)^2} \tag{4b}$$

where Q is the ratio $(1/\omega C_b)/R_b = \omega L_x/R_x$. The fact that the Hay bridge measures inductance in terms of resistance and capacity makes it possible to cover a large range of inductances with a minimum of difficulty with respect to standards.

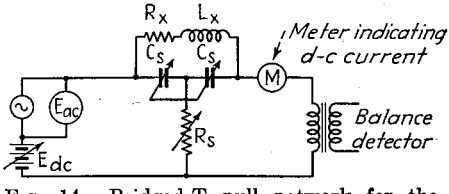


FIG. 14.—Bridged-T null network for the measurement of incremental inductance.

When the losses in the inductance are reasonably low, as is nearly always the case, Eq. (4a) can be simplified to the following

$$L_x = R_a R_c C_b \tag{4c}$$

The extent of the approximation involved in Eq. (4c) decreases as the Q of the coil becomes larger. With $Q = 5$, the error is 4 per cent. It will be noted that to the extent that $1/Q$ is negligible, the balance is independent of frequency.

Another null method of measuring incremental inductance is to employ a bridged-T arrangement, as shown in Fig. 14.[1] Balance is obtained by adjusting the resistance R_s, and simultaneous variation of the two identical condensers C_s. At balance

$$\omega L_x = 2\left(\frac{1}{\omega C_s}\right)$$

$$R_s = \frac{1}{R_x(\omega C)^2} \tag{5a}$$

where $\omega/2\pi$ is the applied frequency. It will be noted that the d-c current is not affected by the process of balancing the bridge, and can be conveniently measured either on the input or the output sides. Likewise, the entire alternating voltage applied to the bridge appears across the unknown inductance, because at balance there is no voltage across the output terminals.

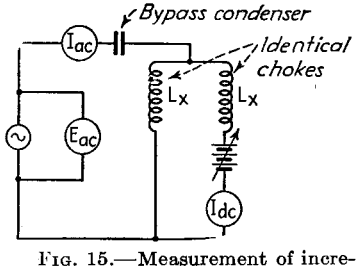


FIG. 15.—Measurement of incremental inductance by voltmeter-ammeter method, using two identical unknown inductances.

Incremental impedance is occasionally determined by the voltmeter-ammeter method by applying a known alternating voltage to the inductance and observing the resulting current. In such methods, it is sometimes convenient to employ two identical chokes, with the direct current introduced as shown in Fig. 15, in order to avoid passing this current through the source of alternating voltage.[2]

A simple method of obtaining incremental inductance of filter chokes with an accuracy sufficient for most purposes is illustrated in Fig. 16.[3] Here the direct cur-

[1] See W. N. Tuttle, Bridged-T and Parallel-T Null Circuits for Measurements at Radio Frequencies, *Proc. I.R.E.*, Vol. 28, p. 23, January, 1940.

[2] H. D. Short, Measurement of Incremental Inductance, *Electronics*, Vol. 13, p. 32, January, 1940.

[3] See Terman, *op. cit.*, p. 58.

rent through the choke is controlled by adjusting the resistance R. The reactance of the choke to the ripple frequency (twice the supply frequency) can then be calculated on the assumption that the higher frequency components of the ripple produce negligible current through the choke. To a fair degree of accuracy, the reactance of the input choke to the ripple frequency is then

$$\left.\begin{matrix}\text{Reactance of inductance}\\ \text{at ripple frequency}\end{matrix}\right\} = \frac{0.424E_{ac}}{I_c} \quad (6)$$

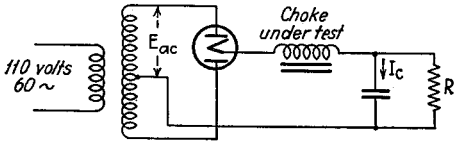


FIG. 16.—Circuit for approximate measurement of incremental inductance at high alternating flux densities.

where E_{ac} is the rms voltage across half of the transformer secondary and I_c is the rms current flowing in the first filter condenser. The alternating flux density can be varied as desired by controlling the primary voltage of the rectifier transformer.

CIRCUIT CONSTANTS AT RADIO FREQUENCY

4. The Substitution Method.—Most measurements of circuit constants at radio frequency make use of some form of substitution. The simplest way in which the substitution method can be carried out is to associate the unknown impedance with a resonant circuit by placing it in parallel with the circuit if the unknown impedance is high or in series if this impedance is low, as in Fig. 17. The resonant circuit is then excited by being loosely coupled to a suitable oscillator, and is adjusted to resonance with the variable standard condenser C_s.[1] The reactance component of the unknown impedance is determined by the change in tuning capacity required to maintain

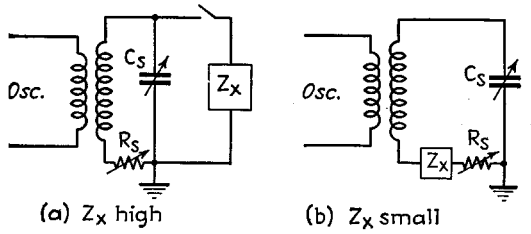


FIG. 17.—Simple arrangements for determining circuit constants by the substitution method.

resonance when the unknown is removed from the circuit. The resistance component is determined either from the change in equivalent series or parallel resistance of the circuit, or from the known resistance that must be added to the circuit in order to maintain the voltage developed across the circuit unchanged when the unknown impedance is removed.

The substitution method is also used in bridge and null-network measurements at radio frequencies. A typical procedure here is first to balance the system with the unknown as well as the standard impedances in the X arm of the bridge, after which the unknown impedance is removed and the standard impedance readjusted to restore balance. The change in the resistance and reactance components of the standard then gives the corresponding components of the unknown. In such measurements,

[1] Resonance between the tuned circuit and the oscillator can be determined by means of a thermocouple milliammeter, a vacuum-tube voltmeter, reaction of the tuned circuit on the oscillator, etc. Where only the reactance component of an impedance is to be measured, it is possible to make the tuned measuring circuit act as the tuned circuit of an oscillating detector. Resonance can then be obtained by zero beating the oscillating detector with the oscillator.

the reactance standard is always a variable condenser, while the resistance standard is usually a variable resistance in series with the standard condenser. The standards can be placed in series with the unknown impedance to give a resonance bridge, as in Fig. 6*a*; or the two impedances can be in parallel. The latter arrangement has the advantage of permitting both the standard and the unknown impedances to have one terminal grounded.

In making substitution measurements with a variable condenser as the reactance standard, it is commonly assumed either that the standard capacity C_s has zero losses, or at least that the losses can be represented by a conductance that is independent of the condenser setting and is shunted across the capacity. These approximations are permissible except when the impedance being measured has an extremely high Q, in which case it may be necessary to determine the exact equivalent circuit of the standard condenser and allow for its change of loss with setting by calculation (see Par. 32, Sec. 2).

The substitution method is capable of very high accuracy when care is taken to avoid the introduction of spurious reactances and resistances as a result of leads, neighboring objects, etc. The precision with which a reactance can be determined is limited only by the accuracy of the standard condenser, and the resistance can be obtained to within better than one per cent if care is given.

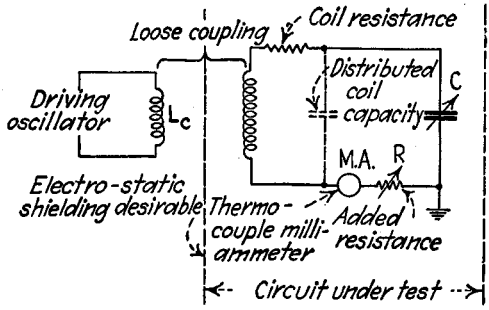


FIG. 18.—Circuit arrangement for measuring radio-frequency resistance by the resistance-variation method.

5. Resistance and Q of Tuned Circuits.—The determination of resistance and Q of resonant circuits is one of the most important radio-frequency measurements. Not only does this give the Q and losses of coils and resonant circuits, but by associating an unknown impedance with a resonant circuit, it is possible to determine the resistance component of this impedance from a knowledge of the effect produced on the circuit resistance.

Many methods have been devised for determining the resistance of tuned circuits. The most important of these are described below.

Resistance-variation Method.—The resistance-variation method of determining the resistance of tuned circuits makes use of the fact that at resonance the current in a circuit is equal to the applied voltage divided by the circuit resistance. If the applied voltage is kept constant, it is then possible to deduce the actual circuit resistance by the current change that results when a known resistance is added to the circuit.

Circuit arrangements suitable for carrying out the necessary measuring operations are shown in Fig. 18. The circuit under test is loosely coupled to a driving oscillator, and has in series with it a thermocouple milliammeter MA and an adjustable resistance R. The circuit is first tuned to resonance with the driver, and the current in the milliammeter is observed when the added resistance R is zero. A known amount of resistance is then added by R, the circuit is retuned to resonance (if this is necessary) without changing the coupling to the driver, and the resulting current noted. The apparent series resistance of the circuit is then given by the following formula:

$$\left.\begin{array}{l}\text{Apparent series resistance}\\ \text{of tuned circuit}\end{array}\right\} = R\left(\frac{I_1}{I_0 - I_1}\right) \qquad (7)$$

where I_0 and I_1 are the currents when the added resistances are, respectively, zero and R. The resistance as measured includes the heater resistance of the thermocouple meter, which must be subtracted.

A modification of the arrangement shown in Fig. 18 consists in using a vacuum-tube voltmeter connected across the capacity C as an indicator rather than the milliammeter. The procedure in using a vacuum-tube voltmeter is the same as outlined above, except that I_1 and I_0 in Eq. (7) are replaced by the corresponding voltages E_1 and E_0. If the input conductance of the vacuum-tube voltmeter is not negligible, it must be determined and allowed for if precise results are desired.

In order to obtain accurate results with the resistance variation method, the current through the coupling coil L_c must be constant and only inductive coupling may exist between the oscillator and the circuit under test. These requirements can be most satisfactorily met by loose coupling between the two circuits or by the use of an electrostatic shield. As a check, it is always desirable to repeat the measurements with several different values of added resistance. In order to avoid errors from capacities to ground, it is necessary to ground one side of the condenser and place the milliammeter and added resistance on the grounded side of the circuit, as shown in Fig. 18. The added resistance must have a negligible skin effect and a reasonably good phase-angle characteristic. It can be either a high-grade decade resistance box or a short link of resistance wire.

The resistance as determined by the resistance-variation method is the apparent series resistance referred to the point R in the circuit of Fig. 18. If the coil has distributed capacity as indicated by the dotted condenser in Fig. 18, this apparent resistance is more than the resistance that should be considered in series with the inductance and total capacity. The relation between the true and apparent series resistance can be obtained by determining the distributed capacity C_0 with the coil in the same situation with respect to surroundings as in the measuring circuit of Fig. 18, and then, using the formula

$$\left.\begin{matrix}\text{True series}\\ \text{resistance}\end{matrix}\right\} = \left(\begin{matrix}\text{apparent series}\\ \text{resistance}\end{matrix}\right)\left(\frac{C}{C_0 + C}\right)^2 \tag{8}$$

where C and C_0 are the external tuning and distributed capacities, respectively. The difference between the true and apparent capacity is usually appreciable, the ratio being 0.83 when $C_0/C = 1/10$.

A modification of the resistance-variation method consists in observing the voltage E_0 developed across the tuned circuit at resonance and then shunting a known resistance R in parallel with the circuit, retuning to resonance, and noting the resulting voltage E_1.[1] Then

$$\left.\begin{matrix}\text{Parallel resonant impedance}\\ \text{of tuned circuit}\end{matrix}\right\} = R\left(\frac{E_0 - E_1}{E_1}\right) \tag{9}$$

The relation between parallel resonant impedance and equivalent series resistance is given by ordinary resonant circuit theory (see Par. 2, Sec. 3).

Reactance-variation (Capacity-variation) Method of Measuring Resistance.[2]—In this method, the circuit under test is loosely coupled to a driving oscillator, as in the resistance-variation method. The tuned circuit is adjusted to resonance with the driving oscillator and the resulting induced current I_0 observed. The tuning capacity

[1] D. B. Sinclair, Parallel-Resonance Methods for Precise Measurements of High Impedances at Radio Frequencies and a Comparison with the Ordinary Series-Resonance Methods, *Proc. I.R.E.*, Vol. 26, p. 1466, December, 1938.

[2] A good discussion of this method, with particular reference to sources of error, is given by Sinclair, *loc. cit.*

is then changed to a value C_1 at which the induced current has dropped to some convenient value I_1. Then

$$\text{Apparent series resistance} \Big\} = \left|\frac{C_1 - C_0}{\omega C_1 C_0}\right| \sqrt{\frac{I_1^2}{I_0^2 - I_1^2}} \tag{10}$$

where $\omega/2\pi$ is the oscillator frequency and C_0 is the tuning capacity at resonance. The resistance includes the heater resistance of the thermocouple used to read the current.

The resistance determined by Eq. (10) is the apparent series resistance referred to the point in the circuit at which the meter is located. If the true series resistance is desired, this can be obtained by correcting for the distributed capacity, as discussed in connection with Eq. (8).

The measuring procedure used with the capacity-variation method is sometimes modified by employing a vacuum-tube voltmeter connected across the circuit capacity as an indicator instead of a thermocouple milliammeter. In this case, Eq. (10) is modified by noting that the voltage across the condenser is $I/\omega C$, leading to the modified equation

$$\text{Apparent series resistance} \Big\} = \left|\frac{C_1 - C_0}{\omega C_1 C_0}\right| \sqrt{\frac{E_1^2}{\left(\frac{C_0}{C_1}\right)^2 E_0^2 - E_1^2}} \tag{11}$$

where E_0 and E_1 are the voltage at resonance and when detuned, respectively. When precise results are required, the input resistance of the vacuum-tube voltmeter must be determined and allowed for.

Frequency-variation Method.—In this method of measurement, the circuit under test is loosely coupled to an oscillator, and the voltage developed across the condenser at resonance is observed using a vacuum-tube voltmeter. The frequency of the oscillator is then changed to a value f_1 at which the voltage developed in the circuit under test has dropped to a convenient value E_1. If, during this process, the *current in the coil of the driving oscillator is kept constant and inductive coupling is used, then*

$$\text{Series resistance} = \left|2\pi \frac{(f_0 + f_1)(f_0 - f_1)L}{f_1}\right| \sqrt{\frac{E_1^2}{E_0^2 - E_1^2}} \tag{12}$$

where f_0 and f_1 are the resonant and detuned frequency, respectively, E_0 and E_1 are the corresponding voltages across the circuit capacity, and L is the circuit inductance.

The resistance obtained in this way after allowing for the input resistance of the vacuum-tube voltmeter is the true series resistance of the circuit, provided that the inductance used in Eq. (12) is the true rather than the apparent circuit inductance at the frequency of measurement.

A rearrangement of Eq. (12) gives

$$Q = \frac{\omega_0 L}{R} = \left|\left(\frac{f_0}{f_0 - f_1}\right)\left(\frac{f_1}{f_0 + f_1}\right)\right| \sqrt{\frac{E_0^2 - E_1^2}{E_1^2}} \tag{13}$$

When it is desirable to use the current induced in the circuit under test as an indication, instead of the voltage developed across the condenser, Eq. (12) can be modified by noting that the voltage developed across the condenser is $I/\omega C$. This leads to

$$\text{Series resistance} = |2\pi(f_0 + f_1)(f_0 - f_1)L| \sqrt{\frac{I_1^2}{\left(\frac{f_1}{f_0}\right)^2 I_0^2 - I_1^2}} \tag{14}$$

The accuracy of the frequency variation method is determined largely by the precision with which the frequency difference $f_0 - f_1$ is known. This can be made very great by measuring the difference directly as an audio beat note, instead of determining f_0 and f_1 individually and then subtracting to obtain their small difference.

Resistance-neutralization Method.[1]—In this method, a negative resistance produced by a dynatron, retarding field, or other negative resistance device is placed in parallel with the resonant circuit being tested, as shown in Fig. 19. The magnitude of the negative resistance is then adjusted by varying the control-grid bias until oscillations are on the verge of being produced. The negative resistance R_n required to accomplish this is then equal to the parallel resonant impedance of the tuned circuit, so that

$$R_s = \frac{(\omega L)^2}{R_n} \qquad (15a)$$

$$Q = \frac{\omega L}{R_s} = \frac{R_n}{\omega L} \qquad (15b)$$

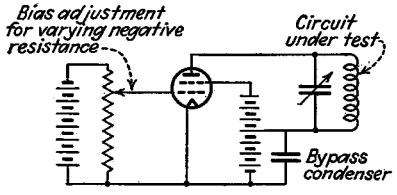
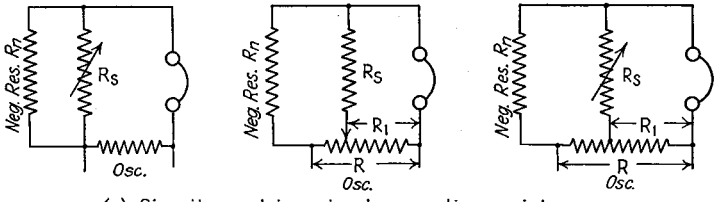


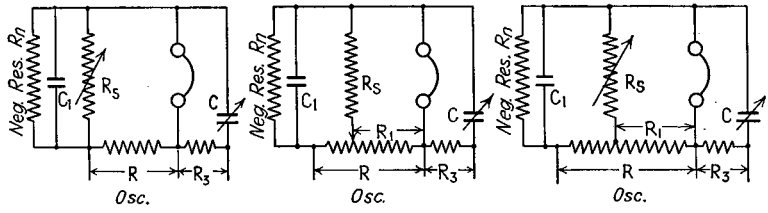
FIG. 19.—A schematic diagram illustrating the resistance-neutralization method of measuring the resistance of a tuned circuit.

where R_s is the true series resistance and ωL is the inductive reactance of the circuit. There is no error from the distributed capacity or the capacities introduced by the measuring tube if the inductance used in the calculation is the true rather than the apparent inductance.



(a) Circuits involving simple negative resistance

$R_n = R_s$ $\qquad R_n = R_s \frac{R}{R_1} + (R - R_1)$ $\qquad R_n = R_s \frac{R}{R_1} + (R - R_1)$



(b) Circuits involving negative resistance shunted by a capacity

FIG. 20.—Bridge circuits for measuring negative resistance.

The magnitude of the negative resistance that neutralizes the positive resistance of the parallel resonant circuit can be determined from the static characteristics of the tube or by switching the tube to a special bridge that will measure the negative

[1] Hajime Iinuma, A Method of Measuring the Radio-frequency Resistance of an Oscillatory Circuit, *Proc. I.R.E.*, Vol. 18, p. 537, March, 1930; Resonant Impedances and Effective Series Resistance of Short-wave Parallel Resonant Circuits, *Proc. I.R.E.*, Vol. 19, p. 467, March, 1931.

resistance by a null balance at audio frequencies. Suitable bridge circuits for obtaining this balance are shown in Fig. 20, together with the equations for balance.[1] In these circuits, the effect of the tube capacity C_1 that shunts the negative resistance is balanced by the adjustable condenser C.

Q Meter.—In the Q meter, a small voltage e is introduced in series with a tuned circuit as a voltage drop across a small resistance R, as shown in Fig. 21. The circuit is then tuned to resonance (or the oscillator frequency is adjusted to the resonant frequency of the circuit), and the voltage E developed across the tuning condenser is observed on a vacuum-tube voltmeter. The circuit Q is then $Q = E/e$.

The results obtained in this way involve certain approximations. In the first place, the internal resistance R of the inserted voltage and also the input resistance of the vacuum-tube voltmeter are charged against the resonant circuit. In the second place, the inserted voltage e is in series with the circuit only if the coil has no distributed capacity. The presence of distributed coil capacity causes the true Q to be higher than that observed by the Q meter. If all the losses are in the coil, as can commonly be assumed, then

$$\text{True } Q = \left(1 + \frac{C_0}{C}\right) (\text{observed } Q) \tag{15}$$

where C_0 is the distributed capacity and C is the added tuning capacity.

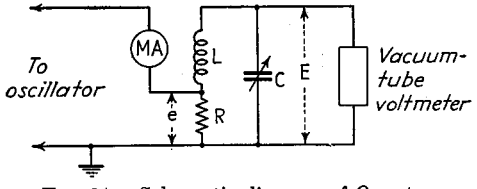


FIG. 21.—Schematic diagram of Q meter.

Q meters are widely used for the measurement of coil Q and coil inductance. The commercial instruments for this purpose are provided with a calibrated tuning condenser so that the apparent coil inductance can be determined from the condenser setting and the frequency. In these instruments, the injected voltage e is held at a predetermined value and the vacuum-tube voltmeter then calibrated directly in Q rather than volts. The instrument is designed so that the resistance R is small compared with the series resistance of ordinary circuits, and the vacuum-tube voltmeter is specifically arranged to have high input resistance, thus minimizing errors. The coil Q is assumed to be the same as the circuit Q; *i.e.*, the condenser and other losses are charged against the coil. In the case of coils with very high Q, this introduces some error, although on a relative basis the results are still correct.

Q meters are frequently used to measure reactance and resistance (or conductance) of choke coils, dielectrics, etc., by the substitution method. The procedure involved consists in making two measurements. First, a convenient coil is used to form a resonant circuit and the circuit Q measured. The unknown impedance is then placed in parallel or series with the coil, as desired, the circuit retuned to resonance, and the Q determined again. The reactance of the unknown is given by the change in tuning capacity required, and the resistance or conductance can be calculated from the effect on the Q and a knowledge of frequency and tuning capacity.

6. Shielded Bridges.—Bridges can be used for impedance measurements at radio frequencies by proper shielding. An example of a completely shielded bridge suitable for radio-frequency measurements is shown in Fig. 22.[2] The effect of the shields is

[1] Details involved in the design of practical bridge circuits of this type are discussed by Terman, *op. cit.*, pp. 76–78. See also F. E. Terman, Improved Circuits for Measuring Negative Resistance, *Electronics*, Vol. 6, p. 340, December, 1933; Edward N. Dingley, Jr., Development of a Circuit for Measuring the Negative Resistance of Pliodynatrons, *Proc. I.R.E.*, Vol. 19, p. 1948, November, 1931.

[2] This particular shielded arrangement should be considered merely as an example. In actual practice, the details can be varied considerably. Extensive discussions, including constructional details of shielded bridges of various types are given by D. B. Sinclair, A Radio-frequency Bridge for Impedance Measurements from 400 Kilocycles to 60 Megacycles, *Proc. I.R.E.*, Vol. 28, p. 497, November, 1940; W. J. Shackelton, Shielded Bridge for Inductive Impedance Measurements at Speech and

to localize the stray capacities so that they act as shown in Fig. 22*b*. The shields surrounding the ratio arms R_1 and R_2 determine the capacities C_1 and C_2. By making the ratio arms and their shields absolutely identical and symmetrical, an equal-ratio bridge is ensured in spite of the capacities. The capacities between shields S_1 and S_2 and ground act as a capacity C_b that shunts the input and so has negligible effect. The capacity of S_4 to ground results in a capacity C_c that is in shunt with the unknown impedance. The ground capacity of shield S_3 supplies C_a.

A shielded bridge may be used to measure radio-frequency impedance either by the substitution method, or directly as a unity-ratio bridge. In the former, the unknown impedance is placed in parallel with C_sR_s, and with $R_s = 0$ and C_s at a convenient setting the bridge is balanced with a convenient capacity and resistance in the X arm. The unknown impedance is then removed and its resistance and reactance components determined by the changes required in C_s and R_s to restore balance. When used as a unity-ratio bridge, a preliminary balance is obtained with the unknown disconnected and R_s short-circuited. This balance is made either by adjusting C_s

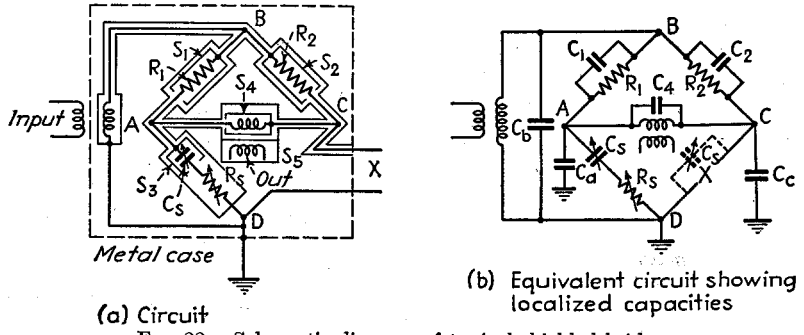


FIG. 22.—Schematic diagram of typical shielded bridge.

or by the use of an auxiliary condenser connected from either terminal A or C to ground, as required. The impedance to be measured is then connected at X and the bridge balanced in the usual way by changing C_s and R_s as required to match the unknown impedance. When the unknown impedance is inductive, it must be changed to capacitive by shunting with a suitable known capacity C_s', as shown dotted in Fig. 22*b*, in order that it may be balanced by a capacity in the "standard" arm.

In using a bridge at radio frequencies, care must be taken with respect to lead inductances, lead capacities, coupling between oscillator and detector, etc. In general, a considerably greater degree of intelligence and skill must be displayed to obtain accurate results at radio frequencies than is necessary in audio-frequency work.

The oscillator exciting the radio-frequency bridge is preferably a modulated oscillator. The output indicator can then be a simple radio receiver. It is necessary that the oscillator be well shielded from the indicator to avoid false indications. A signal generator is frequently used, although a well-shielded modulated oscillator is satisfactory.

Carrier Frequencies, *Trans. A.I.E.E.*, Vol. 45, p. 1266, 1926; also, *Bell System Tech. Jour.*, Vol. 6, p. 142, January, 1927; J. G. Ferguson, Shielding in High-frequency Measurements, *Bell System Tech. Jour.*, Vol. 8, p. 560, July, 1929; *Trans. A.I.E.E.*, Vol. 48, p. 1286, October, 1929; Leo Behr and A. J. Williams, Jr., The Campbell-Shackelton Shielded Ratio Box, *Proc. I.R.E.*, Vol. 20, p. 969, June, 1932; C. L. Fortescue and G. Mole, A Resonance Bridge for Use at Frequencies up to 10 Megacycles per Second, *Jour. I.E.E.*, Vol. 82, p. 687, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 112, June, 1938; C. H. Young, A 5-megacycle Impedance Bridge, *Bell Lab. Rec.*, Vol. 15, p. 261, April, 1937; S. J. Zammataro, An Inductance and Capacitance Bridge, *Bell Lab. Rec.*, Vol. 16, p. 341, June, 1938; D. W. Dye and T. I. Jones, A Radio-frequency Bridge for Impedance and Power-factor Measurements, *Jour. I.E.E.*, Vol. 72, p. 169, 1933 (Wireless Section, *I.E.E.*, Vol. 8, p. 22, March, 1933).

Radio-frequency bridges find their chief use in the measurement of antenna and transmission-line impedances. They also find considerable use in the determination of inductance and capacity, particularly by the substitution method.

7. Bridged-T and Parallel-T Null Networks.—Bridged-T and parallel-T null networks are three-terminal networks so proportioned as to have zero transfer admittance at some particular frequency. Examples of such networks are shown in Fig. 23, together with the circuit proportions required to give zero transmission.[1] These arrangements are equivalent to bridge circuits but have the advantage that the input

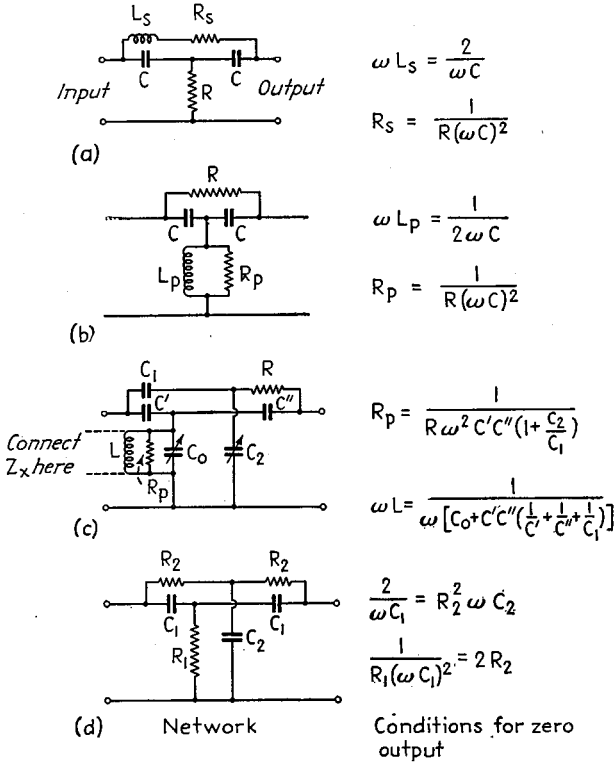


FIG. 23.—Bridged-T and parallel-T null networks useful in measuring circuit constants, together with equations for balance.

and output circuits have one common terminal that can be grounded, as also can one terminal of the unknown impedance in some cases.

The circuit of Fig. 23*a* is suitable for measuring the inductance and Q of radio-frequency coils.[2] In such measurements, the inductance is evaluated in terms of the capacity C, and the parallel resonant impedance formed by the circuit L_sR_sCC is given by the relation[3]

$$\text{Parallel resonant impedance} = 4R \tag{16}$$

[1] These and other null networks are described in detail by W. N. Tuttle, Bridged-T and Parallel-T Circuits for Measurements at Radio Frequencies, *Proc. I.R.E.*, Vol. 28, p. 23, January, 1940.

[2] It is also used for incremental inductance measurements (see Par. 3).

[3] This is an approximate formula in which the assumption is made that $1/Q^2 << 1$. The exact value of parallel impedance is $(1 + 1/Q^2)$ times the value given by Eq. (16).

The only precautions that need to be taken are to provide adequate shielding so that there is no direct coupling between the oscillator and indicator connected to the input and output terminals, respectively, and to avoid excessive capacity shunting the resistance R. The effect of such capacity is equivalent to modifying the value of C that should be used in the balance equations.

The circuit of Fig. 23*b* is similar to *a* but has the disadvantage that stray capacities from the junction of the T to ground modify the balance equations in an inconvenient way. This circuit has been used to compare coils,[1] and is also suitable for the measurement of impedance by the substitution method. In the latter case, a variable condenser and the unknown reactance are connected in parallel with L_p, and the unknown reactance and equivalent shunt resistance determined by the readjustments required in the variable condenser and the resistance R, respectively, to maintain balance when the unknown is removed.

The twin-T circuit of Fig. 23*c* is especially suitable for the general run of impedance measurements at radio frequencies. For this purpose, the substitution method

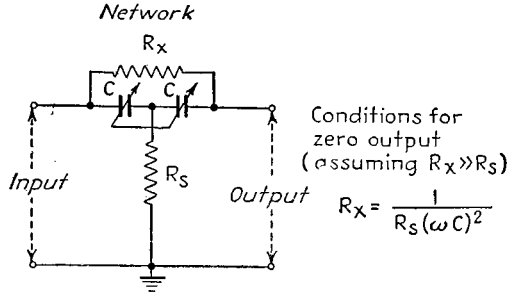


FIG. 24.—Bridged-T network useful in determining a high resistance in terms of a low known resistance and a known capacity.

is employed, with the unknown impedance placed in shunt with the adjustable condenser C_0 and the coil L. A preliminary balance is obtained with the unknown impedance disconnected by adjusting C_0 and C_2 to values C_0' and C_2', as required, after which the corresponding values C_0'' and C_2'' are determined for the case with the unknown impedance in the circuit. The reactance and equivalent parallel resistance of the unknown impedance are then given by the equations

$$\text{Reactance of unknown} = \frac{1}{\omega(C_0' - C_0'')} \tag{17a}$$

$$\text{Parallel resistance of unknown} = \frac{C_1}{R\omega C' C''} \frac{1}{\omega(C_2'' - C_2')} \tag{17b}$$

The results obtained are quite accurate, and by suitably calibrating the variable condensers C_0 and C_2, the arrangement can be direct reading.[2]

The circuit of Fig. 23*d* is equivalent to a Wien bridge. It has been used to measure audio frequencies and also as a feedback network in a degenerative selective amplifier[3]

[1] W. N. Tuttle, A New Instrument and a New Circuit for Coil or Condenser Checking, *Gen. Rad Exp.*, Vol. 12, August–September, 1937.

[2] A detailed discussion of the design features of an instrument of this sort, together with a consideration of the errors resulting from residual inductances and capacities, is given by D. B. Sinclair, The Twin-T, A New Type of Null Instrument for Measuring Impedance at Frequencies up to 30 Megacycles, *Proc. I.R.E.*, Vol. 28, p. 310, July, 1940.

[3] See R. F. Field, a Bridge-type Frequency Meter, *Gen. Rad. Exp.*, Vol. 6, November, 1931; H. H. Scott, A New Type of Selective Circuit and Some Applications, *Proc. I.R.E.*, Vol. 26, p. 226, February, 1938.

Another null network useful in evaluating a high resistance R_x in terms of a low known resistance R_s is shown in Fig. 24.[1] This arrangement does not provide a true null, but if the ratio R_x/R_s is high, the balance is sufficiently sharp for all practical purposes.

8. Use of Transmission Lines in Impedance Measurements.—Transmission lines are frequently used as a tool in making measurements at radio frequencies. Their usefulness arises from the fact that the characteristics of transmission lines can be calculated with considerable accuracy from the mechanical dimensions, combined with the fact that such lines act as an impedance transformer. Thus, one can obtain the ratio of two impedances by connecting these impedances at the two ends of a quarter-wave line, coupling the line to an oscillator, and observing the ratio of currents at the two ends. Under these conditions, one has[2]

$$\left|\frac{I_s}{I_r}\right| = \left|\frac{Z_x}{Z_0}\right| \tag{18}$$

where I_s and I_r are the currents at the two ends of the line, Z_x is the unknown impedance shunted across the receiving end, and Z_0 is the characteristic impedance of the line. The phase angle of the impedance Z_x can be determined by adding a known reactance in shunt and noting the result.

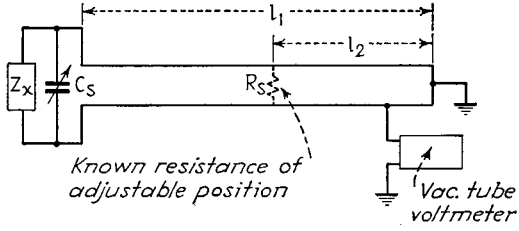


FIG. 25.—Transmission-line arrangement for measuring impedance.

Another example of the use of a transmission line is shown in Fig. 25.[3] Here one end of a two-wire line is short-circuited, and the other end is shunted by a variable condenser and the unknown impedance. A vacuum-tube voltmeter indicates the voltage developed at some convenient point on the line. Measurements are made by connecting the unknown impedance across the line, adjusting the system to resonance with a loosely coupled oscillator, and noting the potential indicated by the vacuum-tube voltmeter. The unknown impedance is then removed, resonance is restored by adjusting the variable condenser, and the vacuum-tube voltmeter made to read the same as previously by shunting a known resistance across the line and adjusting its position as required. The reactance of the unknown is obtained from the change in condenser setting, and the equivalent shunt resistance of the unknown admittance is given by the equation

$$\left.\begin{matrix}\text{Shunt resistance of}\\ \text{unknown impedance}\end{matrix}\right\} = R_s\left[\frac{\sin(2\pi l_1/\lambda)}{\sin(2\pi l_2/\lambda)}\right]^2 \tag{19}$$

[1] P. M. Honnell, Bridged-T Measurement of High Resistances at Radio Frequencies, *Proc. I.R.E.*, Vol. 28, p. 88, February, 1940.

[2] J. W. Labus, Measurement of Resistances and Impedances at High Frequencies, *Proc. I.R.E.*, Vol. 19, p. 452, March, 1931.

A modification consists in observing the voltage across the line at the receiver end and at points one-eighth and one-quarter wave lengths distant from the receiver end. This gives the magnitude and also the phase angle of the unknown impedance. See W. L. Barrow, Measurement of Radio Frequency Impedance with Networks Simulating Lines, *Proc. I.R.E.*, Vol. 23, p. 807, July, 1935.

[3] J. M. Miller and B. Salzberg, Measurements of Admittances at Ultra-high Frequencies, *R.C.A. Rev.*, Vol. 3, p. 486, April, 1939.

where l_1/λ and l_2/λ are the lengths indicated in Fig. 25, measured in wave lengths. In this calculation, it is assumed that the losses in C_s can be represented by a constant resistance independent of capacity setting. If this is not true, a correction must be made.

Impedances of the same order of magnitude as the characteristic impedance of a line can be measured by using the unknown impedance to terminate the line, as in Fig. 26, and deducing the reflection coefficient that results from the observed voltage distribution along the line and the line characteristic impedance. Referring to Fig. 26, the magnitude of the reflection coefficient is

$$\text{Reflection coefficient (magnitude)} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \qquad (20a)$$

$$\text{Reflection coefficient (phase)} = \frac{2l_1}{\lambda} \times 360° \qquad (20b)$$

where E_{max} and E_{min} are the maxima and minima of the voltage distribution and l_1/λ is the distance in wave lengths from the receiving end to the first maximum. With the magnitude and phase of the reflection coefficient determined in this way, and the characteristic impedance known from the line dimensions, the impedance Z_x required to account for this reflection coefficient is obtained from Fig. 42, Sec. 3.[1]

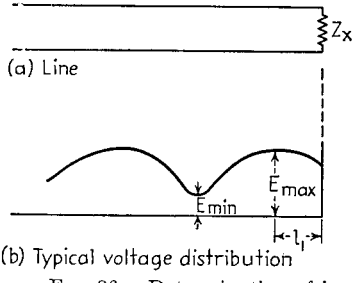


FIG. 26.—Determination of impedance with aid of standing wave patterns on a transmission line.

9. Standards for Radio-frequency Measurements.—The standards used in making measurements at radio frequencies are normally condensers, usually variable, and fixed resistances. Variable condensers are almost ideal reactance standards in that they are continuously variable, have a capacity independent of frequency up to relatively high frequencies, and losses so low as to be negligible for most types of measurements. At the same time, such standards have limitations. There is a residual inductance that, though small, causes the effective capacity to depend on frequency at very high frequencies. Also, although the losses that are present at low and moderate frequencies are primarily dielectric losses and so can be represented by a shunt resistance that is substantially independent of capacity setting and is inversely proportioned to frequency, at high frequencies there are also eddy-current losses of importance that increase with frequency and make the over-all behavior with respect to losses quite complicated.[2]

Certain types of measurements require a capacity in which extremely small accurately known capacity changes must be produced. This can be accomplished by employing a construction in which the change in capacity can be calculated from a change in the mechanical dimensions. An example of such an arrangement is illustrated schematically in Fig. 27. It will be noted that changing the position of the central rod by means of the screw adjustment produces a change in capacity according

[1] Still other means of using transmission lines are described by L. S. Nergaard, A Survey of Ultra-high-frequency Measurements, *R.C.A. Rev.*, Vol. 3, p. 156, October, 1938.

[2] A further discussion of the residual parameters of variable condensers is given in Par. 32, Sec. 2. Methods of evaluating the equivalent inductance, shunt resistance, and series resistance of a variable condenser are described by R. F. Field and D. B. Sinclair, A Method for Determining the Residual Inductance and Resistance of a Variable Air Condenser at Radio Frequencies, *Proc. I.R.E.*, Vol. 24, p. 255, February, 1936; W. Jackson, The Analysis of Air Condenser Loss Resistance, *Proc. I.R.E.*, Vol. 22, p. 957, August, 1934; R. Faraday Proctor, Variable Air Condensers. *Wireless Eng.*, Vol. 17, p. 257, June. 1940.

to the equation

$$\text{Change in capacity} = \frac{0.613}{\log_{10}\frac{b}{a}} \mu\mu\text{f per inch} \qquad (21)$$

End effects and capacities to ground do not affect this result, since they are constant at all times.

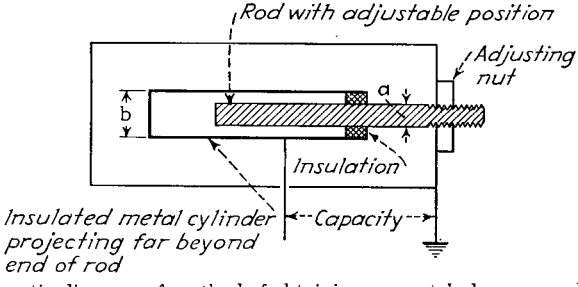


FIG. 27.—Schematic diagram of method of obtaining accurately known small increments of capacity.

The usual resistance standard employed in radio-frequency measurements is a resistance wire of such small diameter that skin effect is negligible at the frequency employed. In this way, the radio-frequency resistance will equal the low-frequency or direct-current resistance. It is not necessary for many purposes that the resistance standard be noninductive, since its reactance can be resonated out without affecting the resistance. It is, however, important that any shunting capacity present have a reactance considerably greater than the resistance. Resistance standards of this type are satisfactory for low and moderate values, *i.e.*, up to the order of one hundred ohms. When standards of higher resistance are desired, it is customary to employ metallized or other similar types of resistors that maintain a substantially constant resistance from direct current up to very high frequencies. When precision is important, it is desirable to check such resistances against low resistance standards composed of fine wire. Resistances of different orders of magnitude can readily be compared by such expedients as (1) measuring a resistance shunted across a resonant circuit in terms of the series of resistance of the circuit;[1] (2) the use of a transmission line; (3) by means of a bridged-T network, as in Fig. 24.

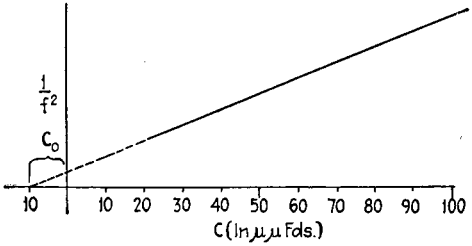


FIG. 28.—Plot of $1/f^2$, where f is the resonant frequency, as a function of external tuning capacity. The value of negative capacity C_0 at which the extrapolated line intercepts the capacity axis is the distributed capacity of the coil, and the slope of the line is a measure of the coil inductance.

Resistance standards are preferably fixed because of the complications introduced if an attempt is made to provide continuous adjustment.

10. Miscellaneous. *Measurement of True Coil Inductance and Distributed Capacity.*—The true inductance and distributed capacity of a coil can be obtained by observing the capacity that must be added to tune the coil to resonance at several

[1] Paul B. Taylor, Method for Measurement of High Resistance at High Frequency, *Proc. I.R.E.*, Vol. 20, p. 1802, November, 1932.

frequencies, and then plotting the added capacity as a function of $1/f^2$, as shown in Fig. 28. This will result in a straight line that will intercept the capacity axis at a negative value C_0 that equals the distributed capacity. The slope of the line gives the true inductance according to the equation

$$\text{Coil inductance in henrys} = 0.0253m \tag{22}$$

where m is the slope of the curve of added capacity plotted against $1/f^2$, where f is in megacycles and capacity is in micromicrofarads.

The distributed capacity can be determined without plotting a curve by adjusting the coil successively to resonance with the fundamental frequency and then the second harmonic of an oscillator.[1] If C_1 and C_2 are the respective tuning capacities, then[2]

$$\text{Distributed capacity} = \frac{C_1 - 4C_2}{3} \tag{23}$$

Measurement of Very Small Capacities.—Very small capacities must sometimes be determined with very high precision. This can be done in several ways. One method that has been used in measuring the direct capacity between the control grid and plate

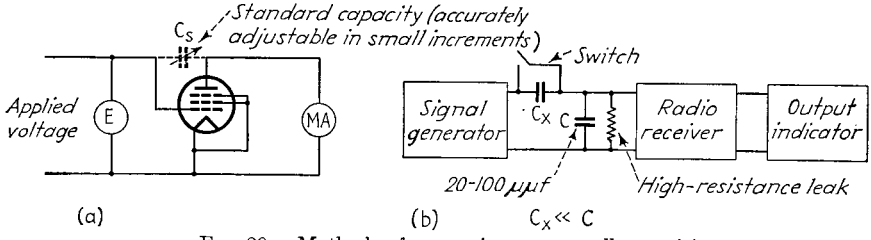


FIG. 29.—Methods of measuring very small capacities.

of a pentode tube consists in applying a known large radio-frequency voltage to the capacity and measuring the resulting current that flows (see Fig. 29*a*).[3] The unknown capacity is then determined either by calculation from known voltage, current, and frequency or by substitution of a known adjustable capacity C_s that is readjusted to maintain the current constant when the plate lead of the tube is disconnected. In the method of Fig. 29*b*,[4] the output of a signal generator is applied to the unknown capacity in series with a very much larger known capacity. The voltage across the latter is applied to the grid of the first tube of a radio receiver, and the signal generator adjusted to give a reasonable receiver output. The unknown capacity is then shorted and the signal generator readjusted to give the same receiver output. The unknown capacity is

$$C_x = C\frac{E_1}{E_2} \tag{24a}$$

where E_1 and E_2 are the signal generator voltages with and without the switch shorted and C is the capacity across the receiver input.

[1] Distributed capacity is sometimes calculated from the frequency at which the coil is in parallel resonance with the distributed capacity. This does not give the distributed capacity effective under practical conditions, however, since it corresponds to a different current distribution within the coil.

[2] Ralph R. Batcher, Rapid Determination of Distributed Capacity of Coils, *Proc. I.R.E.*, Vol. 9, p. 300, August, 1921.

[3] A. V. Loughren and H. W. Parker, The Measurement of Direct Interelectrode Capacitance of Vacuum Tubes, *Proc. I.R.E.*, Vol. 17, p. 957, June, 1929.

[4] Measuring Small Capacities with a Signal Generator, *Hygrade Sylvania News Letter* 56, October, 1939.

Extremely small changes in capacity can be determined by making the capacity involved part of the tuning capacity of an oscillator and then determining the effect upon the oscillator frequency. An auxiliary oscillator can be used as a fixed frequency standard and the frequency variations of the measuring oscillator obtained by observing changes in the beat note between the two oscillators. The relation between the frequency change and capacity is

$$\Delta C_x = C_0 \left[\left(\frac{f_0}{f_1} \right)^2 - 1 \right] \tag{24b}$$

where ΔC_x is the change in capacity, C_0 the original tuning capacity of the oscillator, f_0 the oscillator frequency when tuned only by C_0, and f_1 the oscillator frequency when tuned by C_0 and ΔC_x in parallel. The oscillator tuning capacity C_0 includes tube wire and stray capacities and must be determined experimentally by substituting a known capacity for ΔC_x and noting the resulting frequency change. Since this calibrating capacity can be large enough to be measured accurately by other methods, it is possible to evaluate C_0 very accurately.

Measurement of Dielectrics.[1]—The properties of dielectrics at radio frequencies are normally obtained by the substitution method. The procedure consists in placing a sample of the dielectric between the plates of a condenser that is in parallel with a resonant circuit tuned by a variable condenser. The change in capacity produced by the insertion of the dielectric is obtained by the readjustment of the variable condenser required to maintain resonance, while the shunt resistance resulting from the dielectric losses is determined by evaluating the parallel impedance of the circuit with and without the dielectric present. When solid dielectrics are involved, it is normally permissible to assume that the losses of the standard variable condenser are independent of the setting of this condenser. This approximation is permissible because the losses in solid dielectrics are very much greater than the losses in a good variable condenser.

VOLTAGE, CURRENT, AND POWER MEASUREMENTS

11. Direct-current Voltmeters and Ammeters.—Direct currents and d-c potentials are normally measured with portable instruments of the moving-coil (D'Arsonval) type. Such instruments are available in sensitivities corresponding to full-scale deflection with currents as small as twenty-five microamperes, corresponding to voltmeter sensitivities up to 40,000 ohms per volt. The voltage drop in the moving coil for full-scale deflection is normally under 100 mv.

A given voltmeter or ammeter can be used to cover a wide range of values by the use of suitable multipliers or shunts, respectively. Multirange instruments for measuring current can employ either individual shunts for each range, as in Fig. 30*a*, or a universal shunt, as in Fig. 30*b* and *c*. In the latter arrangement, relative multiplying factor is proportional to R/R_1, irrespective of the meter resistance. When a multirange current instrument is to be protected with fuses, a separate fuse must be provided for each range, with switching arrangements as shown in Fig. 30, in order that the fuse resistance will not affect the calibration.

Multirange voltmeters are obtained by varying the series resistance, using either of the circuits shown in Fig. 31.

[1] For further discussions of dielectric-loss measurements, see J. G. Chaffee, The Determination of Dielectric Properties at Very High Frequencies, *Proc. I.R.E.*, Vol. 22, p. 1009, August, 1934; D. B Sinclair, Impedance Measurements at High Frequencies with Standard Parts, *Gen. Radio Exp.*, Vol. 14, No. 4, September, 1939; E. T. Hoch, Electrode Effects in the Measurement of Power Factor and Dielectric Constant of Sheet Insulating Materials, *Bell System Tech. Jour.*, Vol. 5, p. 555, October, 1926; Miller and Salzberg, *loc. cit.*; L. Hartshorn and W. H. Ward, The Measurement of the Permittivity and Power Factor of Dielectrics at Frequencies from 10^4 to 10^8 Cycles per Second, *Jour. I.E.E.*, Vol. 79, p. 597, 1936; also, Wireless Section, *I.E.E.*, Vol. 12, p. 6, March, 1937.

12. Voltmeters and Ammeters for Power Frequencies.—The iron-vane types of voltmeter and ammeter are widely used at 60 cycles for such purposes as determining line and filament voltages, filament current, etc. Such instruments are inexpensive, and are sufficiently accurate for most applications. The maximum sensitivity of commercial instruments is commonly of the order of 10 milliamperes, corresponding to 100 ohms per volt for voltmeters. The voltage drop in the instrument is normally an appreciable fraction of a volt.

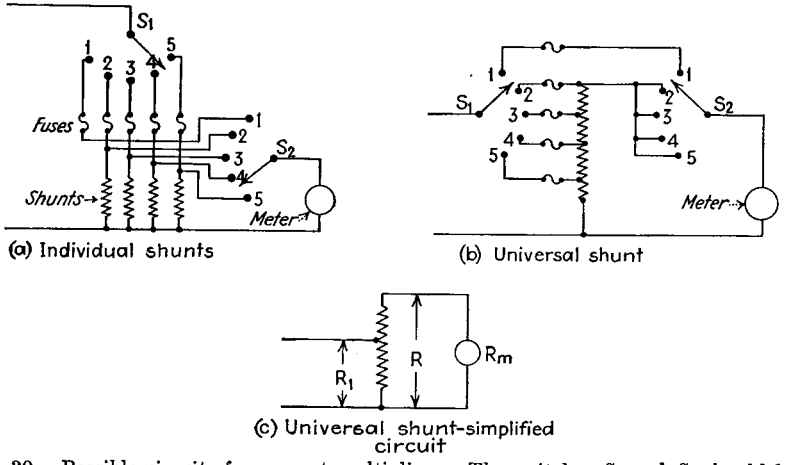


FIG. 30.—Possible circuits for current multipliers. The switches S_1 and S_2 should be of the short-circuiting type operated from a common shaft.

Dynamometer instruments are more accurate than the iron-vane type, but are also considerably more expensive and require more energy to operate.

Both dynamometer and iron-vane instruments give an indication proportional to the effective value of the wave passed through the instrument.

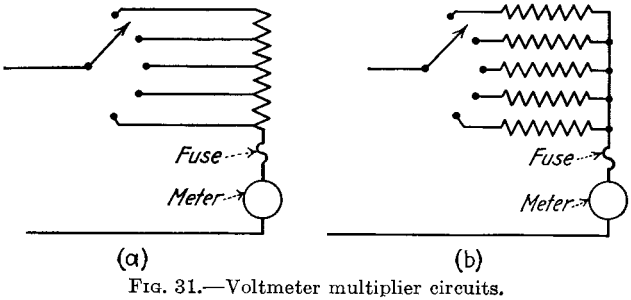


FIG. 31.—Voltmeter multiplier circuits.

13. Rectifier Instruments.[1]—In the rectifier instrument, the current to be measured is passed through a full-wave copper oxide rectifier unit, and the resulting direct current indicated by a moving-coil direct-current instrument, as shown in Fig. 32*a*. Rectifier instruments can be built to give full-scale deflection with alternating currents of less than one milliampere, and so make possible the construction of a-c voltmeters having sensitivities of 1,000 ohms per volt and more. The ruggedness and overload

[1] Further discussion is given by Joseph Sahagen, The Use of Copper-oxide Rectifier for Instrument Purposes, *Proc. I.R.E.*, Vol. 19, p. 233, February, 1931.

capacity compare favorably with moving-coil direct-current instruments. At the same time, rectifier instruments have certain limitations. The best accuracy obtainable is only about 5 per cent because of the variation of rectifier characteristics with temperature. Furthermore, the electrostatic capacity of the rectifier element partially by-passes the higher audio frequencies around the rectifier, causing the instrument to read low at high frequencies by an amount that is of the order of ½ to 1 per cent for each thousand cycles. Another disadvantage of rectifier instruments is their high voltage drop, which is of the order of ½ to 1 volt for full-scale deflection.

The range covered by a given rectifier instrument can be extended by the use of ordinary shunts or multipliers. However, since the resistance that the combination of rectifier and its d-c output meter offers to alternating currents depends upon the current density in the rectifier, it is necessary to provide a separate scale for each range, since the nonlinear resistance of the rectifier causes the law of scale variation to depend upon the resistance associated with the instrument. This difficulty can be avoided by arranging that the multiplier network associated with the rectifier act as a constant-impedance source to the instrument.[1] Thus, with rectifier ammeters, a universal shunt such as is illustrated in Fig. 32*b* will make available various sensitivities,

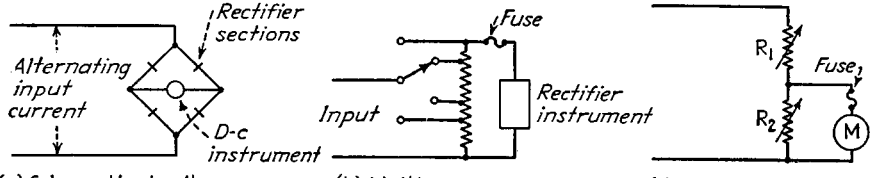


FIG. 32.—Schematic circuit of a rectifier instrument and voltmeter and ammeter multipliers and fusing systems.

with the use of only a single scale. Similarly, multirange voltmeters employing a common scale for all ranges can be obtained by employing the multiplying network of Fig. 32*c* and designing this network so that the equivalent resistance formed by R_1 and R_2 in parallel is constant for all ranges.

Rectifier instruments give an indication that is proportional to the average amplitude of the alternating wave. The phase of the harmonics will accordingly affect the result (see Table 1) when a distorted wave is involved. Commercial rectifier instruments are normally calibrated on a sine wave, and the scale indicates the equivalent effective value of this wave.

14. Thermocouple Instruments.—In a thermocouple instrument, the current to be measured heats a short piece of resistance wire or strip that is associated with a thermocouple. The output of the thermocouple is recorded by a sensitive direct-current microammeter, which accordingly gives an indication of the alternating current passing through the heater. Sensitivities corresponding to full-scale deflection with currents as small as one milliampere can be obtained with ordinary portable d-c instruments, and much greater sensitivity is obtained when a reflecting galvanometer is employed to indicate the thermocouple output. Potential drop produced in the thermocouple heater by full-scale current is commonly of the order of ½ volt.[2]

[1] F. E. Terman, Multirange Rectifier Instruments Having the Same Scale Graduations for All Ranges, *Proc. I.R.E.*, Vol. 23, p. 234, March, 1935.

[2] This does not include the voltage drop due to the inductance of the heater, which at high frequencies may be considerably greater than the resistance drop if a straight heater is employed. The inductance drop can be reduced by folding the heater back on itself to form a noninductive resistance. Low impedance thermocouples of this type are described by H. R. Meahl, P. C. Michel, M. W. Scheldorf, and T. M. Dickinson, Measurements at Radio Frequencies, *Trans. A.I.E.E.*, Vol. 59, p. 654, December, 1940.

The thermocouple is the standard instrument for measuring currents at audio and radio frequencies. It is accurate, stable, and the calibration is independent of frequency up to extremely high frequencies.[1] The only important disadvantage is the low overload capacity, since currents exceeding full load rating by much more than 50 per cent may burn out or at least damage the heater.

The indication of a thermocouple instrument is determined by the effective value of the alternating current in the heater, since the heating is proportional to the square of the effective value. The relationship between the deflection of the d-c microammeter and the heating current follows a square law when the heat generated by the heater is carried away entirely by conduction. This is the case with thermocouples in air and in vacuum thermocouples at low and moderate currents. The resulting nonlinearity of the scale with pronounced bunching at low current values is avoided in some cases by employing especially shaped pole pieces on the d-c instrument.[2]

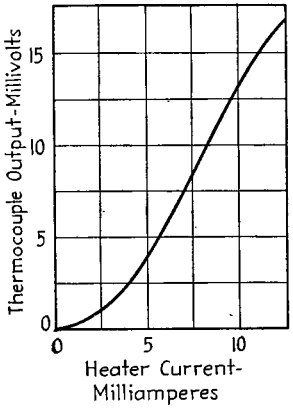


FIG. 33.—Typical calibration curve of a vacuum thermocouple.

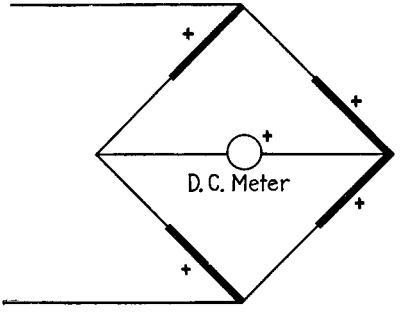


FIG. 34.—Circuit of bridge-type thermocouple instrument.

With vacuum thermocouples, the output varies with heating current as shown in Fig. 33. This starts as a square-law curve with small currents, but as the temperature of the heater rises, the curve flattens off, because an increasing proportion of the total heat is lost by radiation.

Types of Thermocouples.—In most commercial thermocouple instruments, the thermocouple and heater are mounted inside the case of the d-c indicating instrument, and are in air. In instruments for large currents, the heater is commonly a strip of resistance material mounted between copper blocks, with the thermocouple soldered or welded to the center of the strip.[3] Small currents, such as those in the range 100 to 500 ma, are commonly measured using a bridge arrangement as shown in Fig. 34. By combining the heater and thermocouple in this way, the cooling effect of the thermocouple leads is avoided. When still smaller currents are to be measured,

[1] When direct currents are used for calibration, it is necessary to take the average of results obtained with the same d-c current flowing through the heater in opposite directions. This is because the heater may influence the output through the fact that if the thermocouple is in contact with the heater, as is usually the case, some of the d-c current being measured may find its way through the indicating instrument, and either add to or subtract from the true thermocouple output.

[2] See U. S. Patent 1,782,588, issued to F. E. Terman.

[3] Thermocouples of this type are frequently provided with compensation against ambient temperature changes by connecting the thermocouple wires to terminals consisting of conducting strips thermally connected to the mounting block but electrically insulated therefrom. See W. N. Goodwin, Jr., The Compensated Thermocouple Ammeter, *Elec. Eng.*, Vol. 55, p. 23, January, 1936.

the thermocouple and its heater are placed in vacuum to reduce loss of heat through air conduction.

Frequency Characteristics of Thermocouples.—Thermocouple instruments give an indication substantially independent of frequency over wide ranges.[1] However, if the frequency is sufficiently great, the thermocouple instrument will read high. This is primarily because skin effect in the heater increases the effective resistance at high frequencies, and hence the temperature rise resulting from a given current. Other factors, such as stray capacity currents, incipient resonance, etc., also occasionally affect the results.

The correction for frequency is given in Fig. 35 for a number of commercial instruments.[2] It will be noted that the greater the sensitivity of the instrument the better the frequency characteristic. This is because the high-sensitivity instruments employ very small heater wires, which, in turn, have small skin effect.

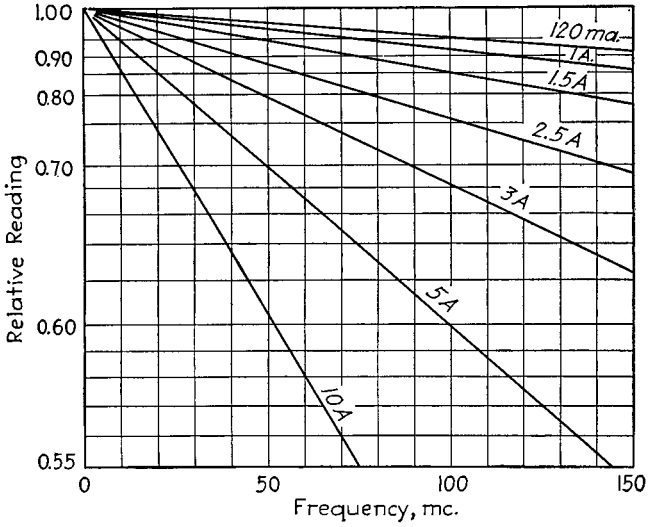


FIG. 35.—Curves giving high-frequency correction for certain commercial thermocouple instruments.

The frequency characteristics shown in Fig. 35 are for thermocouples employing strip heaters for large currents and wire heaters for small currents. A better frequency characteristic may be obtained by employing a tube of resistance material for the heater, since a tube has less skin effect than a wire or strip of equal resistivity (see Par. 4, Sec. 2).

Measurement of Large Currents.—The measurement of large currents with a thermocouple introduces difficulties, because a heater sufficiently large to carry a

[1] This assumes that the thermocouple is at ground potential. If this is not the case, errors are introduced by capacity effects unless suitable shielding is employed. For additional information on this see J. D. Wallace, The Shielding of Radio-frequency Ammeters, *Proc. I.R.E.*, Vol. 29, p. 1, January, 1941.

[2] Accurate calibration of thermocouple instruments can be obtained at high frequencies either by the use of photometric methods involving a straight length of tungsten filament heated by the current to be measured, or by the use of the oscillating-ring electrodynamometer. See John H. Miller, Thermocouple Ammeters for Ultra-high Frequencies, *Proc. I.R.E.*, Vol. 24, p. 1567, December, 1936; J. D. Wallace and A. H. Moore, Frequency Errors in Radio-frequency Ammeters, *Proc. I.R.E.*, Vol. 25, p. 327, March, 1937; H. M. Turner and P. C. Michel, An Electrodynamic Ammeter for Use at Frequencies from One to One Hundred Megacycles, *Proc. I.R.E.*, Vol. 25, p. 1367, November, 1937; Harry R. Meahl; A Bearing-type High Frequency Electrodynamic Ammeter, *Proc. I.R.E.*, Vol. 26, p. 734, June, 1938; Meahl, Michel, Scheldorf, and Dickinson, *loc. cit.*

large current will have considerable skin effect, even at low radio frequencies. Furthermore, ordinary shunts cannot be employed, because the shunting ratio will be affected by the relative inductances of the heater and its shunt, and so will depend upon frequency.

One solution is to employ an array of shunts arranged symmetrically as in Fig. 36*a*. Here each filament possesses the same inductance, so that the current divides in the same way at high frequencies as with direct current. The condenser shunt shown in Fig. 36*b* has also been used successfully in the measurement of large currents.[1] It is merely necessary that the capacity in series with the thermocouple have a much higher impedance than the thermocouple heater.

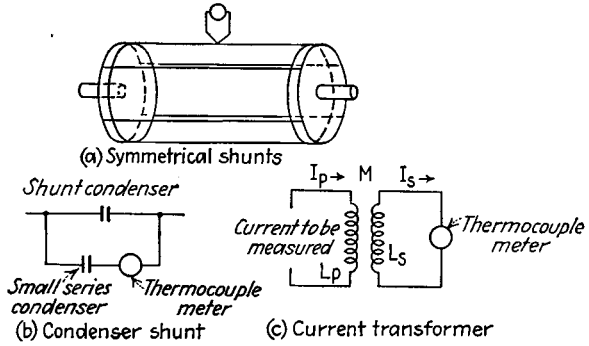


FIG. 36.—Methods that can be employed to measure large radio-frequency currents.

Current transformers, as shown in Fig. 36*c*, are also used at low and moderate radio frequencies. If electrostatic effects are negligible, the transformation ratio is given by the equation

$$\frac{\text{Primary current}}{\text{Secondary current}} = \frac{L_s}{M}\sqrt{1 + \left(\frac{R_s}{\omega L_s}\right)^2} \tag{25}$$

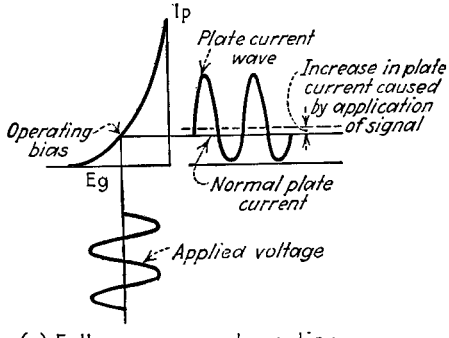
where L_s and M have the meanings indicated in Fig. 36*c* and R_s is the total resistance of the secondary circuit, including the heater. If this total secondary resistance R_s is small compared with the inductive reactance ωL_s of the secondary, the transformation ratio is independent of frequency. The transformer may have either an air or a magnetic core. When the desired transformation ratio is large compared with unity, the secondary can be wound on a toroidal ring, through the center hole of which the wire carrying the primary current is looped once or twice.

15. Vacuum-tube Voltmeters.—The vacuum-tube voltmeter is essentially a vacuum-tube detector in which the rectified d-c current is used as an indication of the applied alternating voltage. The vacuum-tube voltmeter is the standard instrument for the accurate measurement of voltages at audio and radio frequencies. When properly made, it can be calibrated at 60 cycles and used up to extremely high frequencies without any frequency correction. The energy consumption is small in any case, and can be made substantially zero when desired.

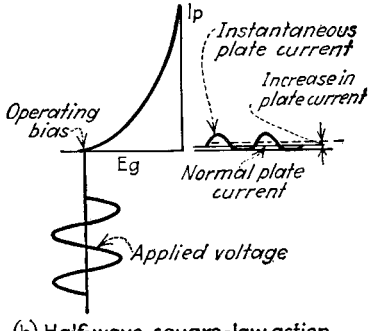
Vacuum-tube Voltmeters Employing Plate Rectification.—The most widely used type of vacuum-tube voltmeter consists of a triode plate (or anode) rectifier. The characteristics obtained with such a device depend upon the adjustments. If the bias is so related to the plate voltage as to allow plate current to flow continuously

[1] Alexander Nyman, Condenser Shunt for Measurement of High Frequency Currents of Large Magnitude, *Proc. I.R.E.*, Vol. 16, p. 208, February, 1928.

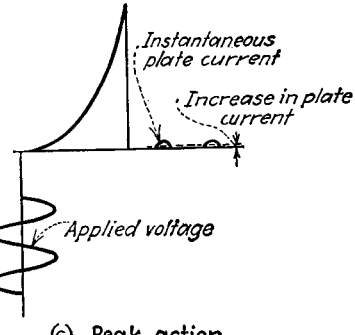
throughout the cycle of applied voltage, as in Fig. 37*a*, the change in plate current produced by the rectifying action is almost exactly proportional to the square of the effective value of the applied voltage. Such operation is termed *full-wave square-law* action. If the grid bias is so chosen that the operating point corresponds to cutoff and the applied signal is appreciable, as in Fig. 37*b*, the negative half cycles of the applied voltage have no effect on the output, and the change in plate current will be very nearly proportional to the square of the effective value of the positive half



(a) Full-wave square-law action



(b) Half-wave square-law action



(c) Peak action

FIG. 37.—Diagrams illustrating action taking place in vacuum-tube voltmeters, with operating points chosen to give full-wave square-law, half-wave square-law, and peak action.

cycles. This is termed *half-wave square-law* action. Finally, if the grid bias is appreciably greater than cutoff bias, as in Fig. 37*c*, the rectified current is determined primarily by the peaks of the positive half cycles, and the instrument tends to become a *peak* voltmeter.

A number of typical vacuum-tube voltmeters circuits are shown in Fig. 38. In the arrangement at *a*, the indicating meter *M* reads the total plate current. This arrangement is satisfactory if the tube is initially given a bias equal to or greater than cutoff, but is otherwise not satisfactory because of the large meter reading with no applied voltage. The plate current that flows with zero applied voltage can be balanced out of the meter by means of an equal and opposite current, as shown at Figs. 38*b* and 38*c*. The circuit of Fig. 38*d*, in which there is a resistance in series with

the meter, gives a more linear calibration between output indication and applied voltage than would be obtained without series resistance. However, this has the disadvantage that the voltage drop in the resistance depends upon the applied signal. With large applied signal voltages, this causes the d-c potential actually applied to the plate to drop considerably, causing a tendency toward peak action. The circuit in Fig. 38*e* is the same as in Fig. 38*c*, except that shunts are shown across the output meter. This permits the use of a very sensitive instrument M for the accurate measurement of small voltages, while still enabling large voltages to be measured by reducing the sensitivity of M.

The calibration of the vacuum-tube voltmeters of Fig. 38 can be made independent of frequency up to where electron transit-time effects become important. To do this, it is necessary merely to make the condensers in the circuit of adequate size, as discussed below.

Vacuum-tube voltmeters of the plate-rectification type normally employ triode tubes. If pentodes are used, they are normally reconnected as triodes. since the

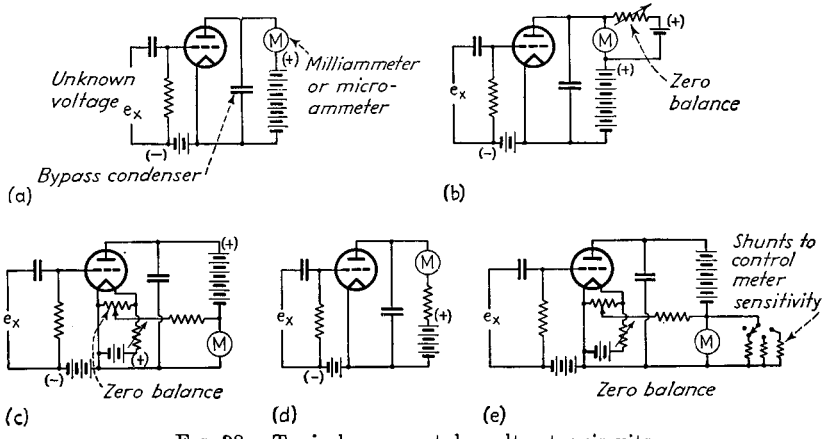


FIG. 38.—Typical vacuum-tube voltmeter circuits.

pentode connection gives no advantages and adds complications. Where the signal voltage to be measured is large, a tube with low amplification factor is desired, since then the grid bias can be sufficient to prevent the large signal voltage from driving the grid positive without at the same time requiring the use of an excessively large plate potential. When high sensitivity to small applied signal is desired, the tube should preferably have a high transconductance in proportion to d-c plate current. This characteristic is most completely realized by tubes designed for video-frequency voltage amplification.

The maximum sensitivity obtainable with a vacuum-tube voltmeter is limited by the stability that can be achieved in balancing out the d-c plate current from the meter. This is because the less the current required for full-scale deflection the more precise must be this balance if the deflection in the absence of a signal is not to drift appreciably from zero. The chief difficulty in maintaining an accurate zero balance comes as a result of variations in the tube voltages, though variations in tube characteristics either through aging or merely as a result of the "warming up" of the tube are also troublesome. When small voltages are to be measured by a vacuum-tube voltmeter, it is very important that the power source, including the filament as well as the anode, screen, and bias voltages, be carefully regulated. It is also helpful to employ circuit arrangements that are inherently stabilized against changes. Thus

it is possible to compensate for the effect of filament voltage changes in filament tubes by obtaining a portion of the grid bias from a resistance in the filament circuit.[1] A more comprehensive balancing system is obtained by using an auxiliary balancing tube to prevent zero drift. By suitable compensating arrangements, this auxiliary tube can be made to have equivalent amplification factor and plate resistance that are the same as for the voltmeter tube, so that when the two tubes are placed in a bridge circuit, the effect of variations in the supply potential on the zero deflection of the output meter can be balanced out.[2]

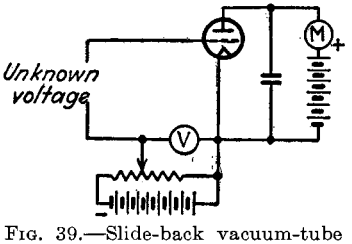


FIG. 39.—Slide-back vacuum-tube voltmeter.

Slide-back Vacuum-tube Voltmeters.—The slide-back vacuum-tube voltmeter consists of a plate-rectifier type of instrument, as shown in Fig. 39, with the bias adjusted so that the plate current is reduced to a few microamperes that are indicated on the meter M. An unknown voltage is measured by applying it to the grid of the tube, after which the negative bias on this grid is increased until the d-c current in the plate circuit is the same as before the application of the signal and the increased bias. The peak value of the positive half cycle of the applied wave is then taken as the increase in bias voltage required.

The results obtained with a slide-back voltmeter are always slightly low by an amount that becomes less the larger the applied signal and the sharper the cutoff characteristic of the tube. An analysis of the relations existing has shown that when the initial plate current is only a few microamperes, the accuracy is independent of the exact value of this current, and is determined only by the quantity KE, where E is the peak value of the applied voltage and K is a constant determined by the kind of cathode and the sharpness of cutoff.[3] The constant K is of the order of 8 with

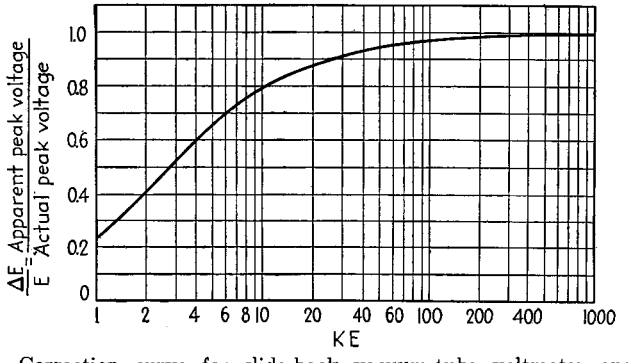


FIG. 40.—Correction curve for slide-back vacuum-tube voltmeter operated with very small initial current, showing ratio of apparent voltage as read to the actual peak voltage E.

sharp cutoff and oxide-coated cathodes, and is less for other tubes. The error to be expected is given in Fig. 40, and is seen to be small if the signal is at least moderately large. It has been found that best results are obtained with a screen-grid or pentode

[1] See H. M. Turner, A Compensated Electron Tube Voltmeter, *Proc. I.R.E.*, Vol. 16, p. 799, September, 1928.

[2] See C. Williamson and J. Nagy, Push-pull Stabilized Triode Voltmeters, *Rev. Sci. Instruments*, Vol. 9, p. 270, September, 1938.

[3] C. B. Aiken and L. C. Birdsall, Sharp Cutoff in Vacuum Tubes—Slide-back Voltmeter, *Trans. A I.E.E.*, Vol. 57, p. 171, April, 1938.

tube connected as a triode, with the screen grid serving as a control electrode and the normal control electrode returned to the cathode. Operation in this way provides a sharper cutoff characteristic than is given by the usual triode.

Diode Voltmeters.—The circuit of a typical diode vacuum-tube voltmeter is shown in Fig. 41*a*. This is essentially a diode detector in which the load resistance R and the microammeter M form a d-c voltmeter that indicates the rectified voltage developed in the output load resistance R. This potential is a measure of the peak value of the applied alternating voltage, and if the load resistance R is high and the applied signal large, this potential will be almost equal to the peak. The exact relation can in any case be determined by suitable calibration.[1] With the diode connected as in Fig. 41*a*, the positive peaks are indicated, while reversing its terminals, as in Fig. 41*b*, will give a measure of the negative peaks.

Unlike the plate-rectifier type of vacuum-tube voltmeter, the diode consumes energy. The equivalent input resistance is $R/2\eta$, where η is the efficiency of rectification and will commonly approach unity (see Par. 9, Sec. 7).

Diode voltmeters can be used in a slide-back circuit, as shown in Fig. 41*c*. Here the signal is applied to the diode in series with a bias that is adjusted so that just a

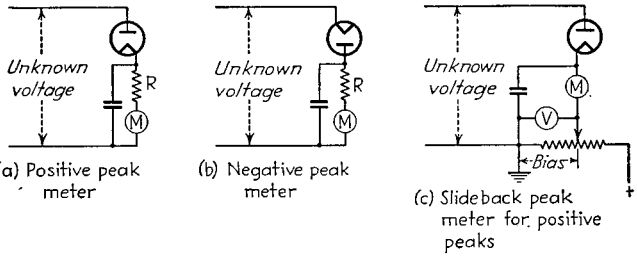


FIG. 41.—Circuits of typical diode vacuum-tube voltmeter.

trace of plate current flows. The amount of bias required is then almost exactly equal to the peak of the applied signal, provided that this signal is reasonably large.

The tubes used in diode voltmeters depend largely upon the applied potentials. When these are small or moderate, it is possible to use the same diodes employed as detectors in receivers. With larger potentials, it is customary to use triode tubes connected as diodes, or rectifier tubes of the type employed in power-supply systems.

Frequency Effects in Vacuum-tube Voltmeters—Behavior at Ultra-high Frequencies.[2]—The indication obtained with vacuum-tube voltmeters of the type shown in Fig. 38 will be substantially independent of frequency for all ordinary frequencies, provided that the condenser C in the plate circuit is an adequate by-pass and provided that any resistance-capacity combination that may be in the grid circuit is so proportioned as to cause negligible loss of voltage. When properly designed, a vacuum-tube voltmeter can be calibrated at 60 cycles and then used at any ordinary audio or radio frequency.

At extremely high frequencies, the calibration of a vacuum-tube voltmeter is modified, however, by incipient series resonance between the inductance of the input lead and the input capacity, and by transit-time effects.[3] Resonance effects in the

[1] An analysis of the diode voltmeter for conditions such that the current through M is only a few microamperes is given by Charles B. Aiken, Theory of the Diode Voltmeter, *Proc. I.R.E.* Vol. 26, p. 859, July, 1938.

[2] See C. L. Fortescue, Thermionic Peak Voltmeters for Use at Very High Frequencies, *Jour, I.E.E.*, Vol. 177, p. 429, 1935 (Wireless Section, *I.E.E.*, Vol. 10, p. 262, September, 1935).

[3] L. S. Nergaard, Electrical Measurements at Wavelengths Less than Two Meters, *Proc. I.R.E.*, Vol. 24, p. 1207, September, 1936; E. C. S. Megaw, Voltage Measurement at Very High Frequencies, *Wireless Eng.*, Vol. 13, p. 65, February, 1936; p. 135, March, 1936; p. 201, April, 1936.

input lead cause the voltage at the input electrode of the tube actually to be greater than the applied voltage, and so make the instrument read high. The amount of error is approximately 1 per cent when the frequency is 10 per cent of the frequency at which the lead inductance is in resonance with the electrode capacity, and varies with the square of the frequency. In the case of the Type 955 "acorn" tube connected as a diode, the lead error with the shortest possible leads external to the tube is approximately 2 per cent at 100 mc.

The transit time of the electrons causes the vacuum-tube voltmeter to read low, since some of the electrons that leave the cathode just before the end of the cycle fail to reach the plate before the cycle reverses. In diode peak voltmeters of the type shown in Fig. 41, the error from transit time is approximately

$$\frac{\text{Error in voltage}}{\text{Applied voltage}} = \frac{kx_a}{\lambda\sqrt{V}} \tag{26}$$

where k = a constant that is approximately 1,055 for cylindrical electrodes and 2,110 with plane electrodes.
x_a = clearance between anode and cathode, cm.
V = peak value of applied signal.
λ = wavelength, cm.

A Type 955 "acorn" tube connected as a diode has a transit time error of approximately 3 per cent, with an applied voltage having a peak value of 10 volts at a frequency of 100 mc.[1]

Wave-form Considerations.—Vacuum-tube voltmeters are normally calibrated on sinusoidal voltages. When they are used to measure a distorted wave, the indication will then be influenced by the type of adjustment employed in the vacuum-tube voltmeter. With a full-wave square-law instrument, the effective value of the wave is read, and the phase of the harmonics has no effect, while with a half-wave square-law voltmeter, the indication will depend not only upon the magnitude of the harmonics but also upon their relative phase, and the same is true even to a greater extent with peak instruments. The effects to be expected are summarized in Table 1.[2]

TABLE 1

	Full-wave square law	Half-wave square law	Full-wave linear[1]	Peak
Turnover possible?	No	Yes	No	Yes
Phase of harmonics affect reading?	No	Yes	Yes	Yes
Effect of harmonic component on reading:				
50 per cent second harmonic.	11 per cent	−6 to +27 per cent	0 to 10 per cent	−25 to +50 per cent
50 per cent third harmonic.	11 per cent	12.5 per cent	−10 to +16 per cent	8 to +50 per cent

[1] This case corresponds to the copper oxide rectifier instrument.

Reversing the input terminals of a vacuum-tube voltmeter will sometimes change the reading of the output meter. This is known as *turnover* and is caused by even harmonics in the applied voltage wave making the positive and negative half cycles

[1] L. S. Nergaard, A Survey of Ultra-high-frequency Measurements, *R.C.A. Rev.*, Vol. 3, p. 156, October, 1938.

[2] This is based on Irving Wolff, Alternating Current Measuring Instruments as Discriminators against Harmonics, *Proc. I.R.E.*, Vol. 19, p. 647, April, 1931.

differ in wave form. Turnover is greatest in peak voltmeters, is present in the half-wave square-law type, but is nonexistent in full-wave square-law and linear instruments. When turnover exists, the average of the direct and reversed polarity readings will be approximately but not necessarily exactly the correct value.

Logarithmic Vacuum-tube Voltmeters.—In a logarithmic voltmeter, the indication is proportional to the logarithm of the applied potential, leading to a linear decibel scale. A logarithmic characteristic can be obtained by applying the voltage to be indicated to an amplifier consisting of several stages employing variable-mu tubes (see Fig. 42*a*).[1] The output of this amplifier is then rectified, and the resulting d-c current used to provide automatic-volume-control action that controls the amplification so that the output tends to be constant with wide variations in input voltage. The d-c current developing the automatic-volume-control voltage is then used as an

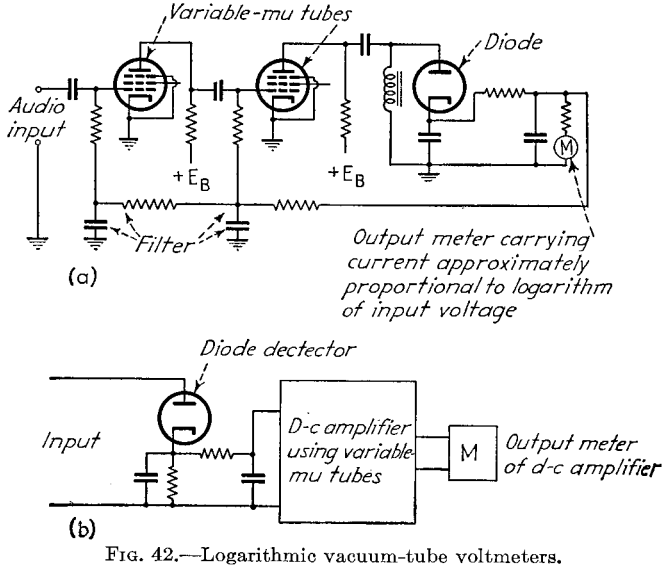


FIG. 42.—Logarithmic vacuum-tube voltmeters.

indication of the amplitude of the applied signal. Such an arrangement has a characteristic that is reasonably close to logarithmic over very wide ranges of amplitude, and is widely used in field-strength recorders, in the testing of loud-speakers, etc.

Another form of the logarithmic voltmeter suitable for large voltages is shown in Fig. 42*b*. Here a diode rectifier is used to develop a d-c voltage proportional to the input voltage. This rectified output, or a fraction thereof, is then amplified by a d-c amplifier employing a variable-mu tube. A logarithmic characteristic over a 20-db range can be obtained in this way, with still greater range possible by employing auxiliary tubes.[2]

Another type of logarithmic meter that has been used for indicating small d-c currents makes use of the fact that the distribution of emission velocities from a

[1] H. A. Wheeler and V. E. Whitman, Acoustic Testing of High-fidelity Receivers, *Proc. I.R.E.*, Vol. 23, p. 610, June, 1935; S. Ballantine, High Quality Radio Broadcast Transmission and Reception, *Proc. I.R.E.*, Vol. 23, p. 618, June, 1935; K. R. Sturley and R. P. Shipway, A Visual Selectivity Meter with Uniform Decibel Scale, *Jour. I.E.E.*, Vol 87, p. 189, 1940; also, Wireless Section, *I.E.E.*, Vol. 15, p. 215, September, 1940.

[2] John P. Taylor, A D-C Amplifier for Logarithmic Recording, *Electronics*, Vol. 10, p. 24, March, 1937; F. V. Hunt, A Vacuum-tube Voltmeter with Logarithmic Response. *Rev. Sci. Instruments*, Vol. 4, p. 672, December, 1933.

cathode follows the Maxwellian law.[1] Accordingly, when the anode potential is slightly negative, the actual difference in potential between anode and cathode resulting from the passage of a current between these electrodes is a logarithmic function of this current.

Feedback Vacuum Tube Voltmeters for Alternating Currents.—A convenient method of measuring small alternating voltages, particularly at audio frequencies, is to apply the voltage to be measured to an amplifier that is designed to have an unusually constant response over the frequency range of interest, and is preferably provided with a large amount of negative feedback. In this way, it is possible to amplify a small potential, such as one millivolt or less, up to a level where the amplified output can be indicated on a thermocouple, diode, or copper oxide rectifier.[2] The use of feedback makes it possible to obtain an amplification substantially independent of tube characteristics, supply-voltage variations, etc. The amplifier gain can accordingly be set to a desired value, with the assurance that it will not change thereafter.

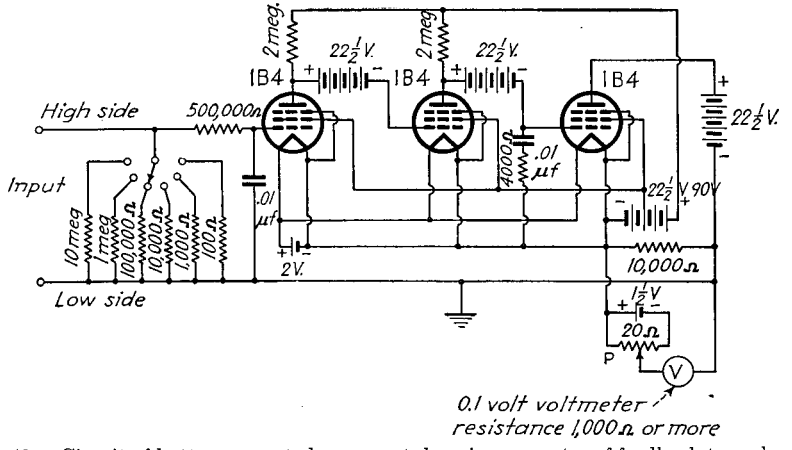


FIG. 43.—Circuit of battery operated vacuum-tube microammeter of feedback type, having full-scale ranges of 10^{-8} to 10^{-3} amperes, suitable for use with ionization gages.

Direct-Current Vacuum-tube Voltmeters and Microammeters.—Vacuum-tube devices are frequently used for the measurement of direct current and voltage. In the case of voltage, they have the advantage of consuming negligible energy, whereas vacuum-tube microammeters can be built that are rugged and at the same time capable of indicating extremely small currents. The simplest vacuum-tube voltmeter for direct current is a d-c vacuum-tube amplifier such as is described in Par. 6, Sec. 5, the output voltage (or the output current) produced by the applied signal being used as an indication of the applied voltage.

A somewhat more elaborate arrangement that is widely used as a d-c voltmeter for radio service and test work makes use of an ordinary push-pull d-c amplifier provided with a large amount of negative feedback to direct current to stabilize the calibration.[3] A high-resistance (16 to 160 megohms) input potentiometer is provided so that the plate, screen, and bias voltages actually existing at the electrodes of a tube

[1] Ralph E. Meagher and E. P. Bentley, Vacuum Tube Circuit to Measure the Logarithm of a Direct Current, *Rev. Sci. Instruments*, Vol. 10, p. 336, November, 1939.

[2] Such instruments are described by F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 10, p. 649, October, 1939; Stuart Ballantine, Electronic Voltmeter Using Feedback, *Electronics*, Vol. 11, p. 33, September, 1938.

[3] This is the Rider Volt-ohmyst.

may be obtained with accuracy, even when extremely high impedances are in series with the electrodes.

D-c voltages can also be measured by a feedback instrument such as is illustrated in Fig. 43.[1] Here a three-stage direct-current amplifier is provided with a large amount of negative feedback that is effective with direct currents. In operation, an initial adjustment is made by short-circuiting the input terminals of the amplifier and adjusting the potentiometer P until voltmeter V reads zero. The magnitude of an unknown d-c voltage applied to the input is then given directly by the reading that this signal produces on voltmeter V. This is because the voltmeter deflection represents the feedback voltage, and with large feedback this is almost equal to the applied voltage. Direct-current instruments of the feedback type are widely used for the measurement of currents of the order of 10^{-3} to 10^{-8} ampere in connection with ionization gages. Here the current to be measured is passed through a resistance such that the resulting voltage drop can be read on a 0.1-volt scale.

16. Cathode-ray Tube Voltmeters and Ammeters.—Cathode-ray tubes are frequently used to measure peak voltage, and are occasionally employed for peak current measurement. A voltage to be determined is applied to the electrostatic deflecting plates, and the length of the resulting line noted. Calibration can be made with a d-c voltage or with an alternating voltage of any convenient frequency. The calibration is independent of frequency until the frequency becomes so high that the time required by the electrons in the beam to travel the length of the deflecting plates is not negligible compared with a cycle.

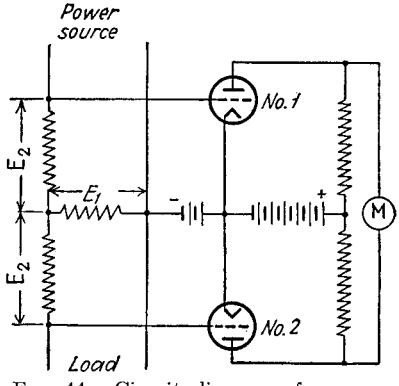


FIG. 44.—Circuit diagram of vacuum-tube wattmeter.

A current can be measured with a cathode-ray tube by passing the current through magnetic deflecting coils and noting the resulting deflection of the spot. Calibration can be obtained by using either direct current or 60-cycle alternating current. This method of measuring current has the disadvantage that the reactance of the deflecting coils produces a high voltage drop unless the frequency is very low.

17. Power Measurements.—The commonest method of determining power at audio and radio frequencies is to measure the resistance R of the circuit, and observe the rms current I that flows. The power is then I^2R.

A modification of this procedure consists in determining the voltage developed across a resistance load and then evaluating power from the fact that the power is E^2/R. This method is frequently used in determining power on nonresonant transmission lines. The common power-level indicator for audio frequencies is also an instrument of this class. Here a copper oxide meter is used to determine the voltage developed across a known impedance, usually the impedance furnished by a 500- or 600-ohm line.

Calorimetric methods of power measurement are sometimes used when large powers are involved or when the frequency is so high that ordinary measuring methods are difficult to employ. A typical example is in the case of water-cooled power tubes, where the power dissipated in the tube (including filament heating power and dissipation at the grid) can be determined with fair accuracy from the temperature rise

[1] A. W. Vance, An Improved Vacuum Tube Microammeter, *Rev. Sci. Instruments*, Vol. 7, p. 489, December, 1936.

of the cooling water. The output is then obtained by subtracting this dissipation from the total input power.

Photometer methods are also employed to some extent by comparing the power to be determined with a known and controllable power measured with direct-current instruments. A typical example consists in using the unknown power to heat the filament of a lamp and then evaluating this power in terms of the d-c heating power required to give the same light emission (or same thermionic emission).

A variation of this is to determine the power dissipated in a filament by a high-frequency current by measuring the increase in resistance of the filament, or the amount of reduction necessary in d-c power supplied the filament in order to keep the filament resistance constant.

Several types of vacuum-tube wattmeters have been proposed.[1] One arrangement consists of two square-law vacuum-tube voltmeters arranged as in Fig. 44. Here one of the applied voltages E_1 or E_2 is proportional to the voltage, and the other is made proportional to current. It can be readily shown that the difference in the d-c output currents as read by the meter M in Fig. 44 is then proportional to $E_1E_2 \cos\theta$, where θ is the phase difference between E_1 and E_2. Other types of vacuum-tube voltmeters employ a single multielectrode tube operated under conditions such that when the voltages E_1 and E_2 are applied to different electrodes, the change in d-c plate current is proportional to $E_1E_2 \cos\theta$, and hence to power.

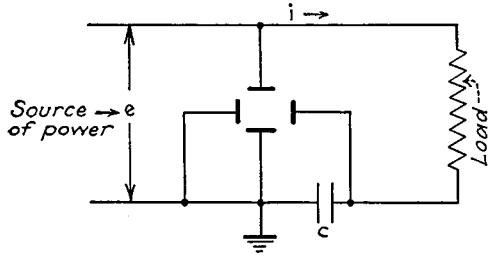


FIG. 45.—Cathode-ray tube connected to measure power dissipated in a load.

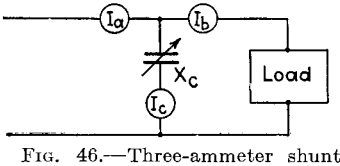


FIG. 46.—Three-ammeter shunt-reactance method of measuring power.

Cathode-ray tubes can be used to measure power by means of the circuit shown in Fig. 45.[2] The deflection along one axis is made proportional to the load voltage, and along the other axis proportional to the instantaneous integral of the current, *i.e.*, to the charge upon the capacity C. The resulting diagram traced by the fluorescent spot is a closed loop, the area of which is proportional to the power loss per cycle. Calibration can be made by introducing known losses or by calculation.

The three-ammeter shunt-reactance method of measuring power shown in Fig. 46 is often used to determine power in radio-frequency transmission lines or the power developed by a radio transmitter.[3] Analysis of this circuit shows that the power in watts flowing through the device is

$$\text{Watts} = 2X_c\sqrt{S(S - I_a)(S - I_b)(S - I_c)} \tag{27}$$

where X_c = reactance of shunt condenser, ohms.

[1] H. M. Turner and F. T. McNamara, An Electron Tube Wattmeter and Voltmeter and a Phase-shifting Bridge, *Proc. I.R.E.*, Vol. 18, p. 1743, October, 1930; H. O. Peterson, U.S. Patent 1,586,553; John R. Pierce, A Proposed Wattmeter Using Multielectrode Tubes, *Proc. I.R.E.*, Vol. 24, p. 577 April, 1936; R. J. Wey, A Thermionic Wattmeter, *Wireless Eng.*, Vol. 14, p. 490, September, 1937.

[2] See Harris J. Ryan, A Power Indicator Diagram for High Tension Circuits, *Trans. A.I.E.E.*, Vol. 30, Pt. II, p. 1089, 1911; A. Hoyt Taylor, The Measurement of Radio-frequency Power, *Proc. I.R.E.*, Vol. 24, p. 1342, October, 1936.

[3] See P. M. Honnell and E. B. Ferrell, The Measurement of Harmonic Power Output of a Radio Transmitter, *Proc. I.R.E.*, Vol. 22, p. 1181, October, 1934.

I_a, I_b, I_c = currents in the three meters (see Fig. 46).

$$S = \frac{I_a + I_b + I_c}{2}.$$

The best accuracy is obtained when the reactance of the shunt condenser is such that I_b and I_c are of the same order of magnitude. The accuracy becomes small when the load is highly reactive.

An approximate determination of a-c power absorbed by a load can be made by using a rectifier to supply the load.[1] If the tube impedance is very small compared with the load resistance, the power dissipated in the tube is small compared with the d-c power dissipated in the load. The latter power accordingly is only slightly less than the total power delivered to the system, and can be measured with d-c instruments, provided an inductance-capacity filter is used to smooth out the current in the load resistance.

18. Volume Indicators.—In the monitoring of program circuits, it is necessary to employ an instrument that will indicate the power level of the audio-frequency currents involved. Inasmuch as these currents fluctuate in intensity in accordance with the amplitude of the speech or music, the ballistic characteristics of such an instrument are fully as important as the characteristics under steady-state conditions. In order to achieve uniformity, the characteristics of a standard volume indicator have been agreed upon, together with a standard reference level to be used in expressing power level in program circuits.[2]

The standard volume indicator consists of a full-wave copper oxide rectifier voltmeter having an input impedance of 7,500 ohms, and is calibrated on the assumption that the instrument will be shunted across a 600-ohm line. The instrument is calibrated in terms of decibels, with zero level (or 100 per cent modulation) corresponding to a milliwatt of power in the 600-ohm line under steady state conditions. The ballistic characteristics are such that the sudden application of a single-frequency voltage that gives a steady state reading of zero level will cause the pointer to overswing by 1 to 1.5 per cent and to reach 99 per cent of the steady state deflection in 0.3 seconds. Because of these ballistic characteristics, the instrument does not read the actual instantaneous power level in db, but rather indicates a mean level based on an average over a short period of time. Because of this, the scale markings are referred to as *volume units*, abbreviated "vu," rather than decibels. The scale reading in volume units will correspond to decibels with reference to one milliwatt only in the case of steady-state sine-wave voltages.

WAVE FORM AND PHASE

19. Harmonic Analysis of Waves.—The wave shape of an alternating voltage or current can be obtained with the aid of an oscillograph, a cathode-ray oscillograph employing a linear sweep being most commonly used.

Periodic electrical waves can always be expressed as the sum of a direct-current component plus a series of alternating components that are all harmonics of the fundamental frequency of the wave; *i.e.*, if E is a periodic function, one can write

$$E = D_0 + C_1 \sin(\omega t + \Phi_1) + C_2 \sin(2\omega t + \Phi_2) + C_3 \sin(3\omega t + \Phi_3) + \ldots \quad (28a)$$

where D_0 = amplitude of direct-current component.
C_1, C_2, etc. = amplitudes of corresponding alternating-current components.

[1] P. M. Honnell, R-F Power Measurements, *Electronics*, Vol. 13, p. 21, January, 1940.

[2] H. A. Affel, H. A. Chinn and R. M. Morris, New Standard Volume Indicator and Reference Level, *Electronics*, Vol. 12, p. 28, February, 1939; H. A. Chinn, D. K. Gannett, and R. M. Morris, A New Standard Volume Indicator and Reference Level, *Proc. I.R.E.*, Vol. 28, p. 1, January, 1940.

ϕ_1, ϕ_2, etc. = phase angle of corresponding alternating-current components.
$\omega = 2\pi$ times the fundamental frequency of the wave.
Equation (28*a*) is often rearranged as

$$E = A_1 \sin \omega t + A_2 \sin 2\omega t + A_3 \sin 3\omega t + \cdots + D_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \cdots \quad (28b)$$

where $\sqrt{A_n^2 + B_n^2} = C_n$.

$$\frac{B_n}{A_n} = \tan \phi_n.$$

Fourier Method of Determining the Coefficients of an Arbitrary Wave.—The coefficients appearing in Eq. (28*b*) can be evaluated with the aid of the following integrals:

$$A_n = \frac{1}{\pi} \int_{\omega t=0}^{\omega t=2\pi} E \sin n\omega t \, d(\omega t) \quad (29a)$$

$$B_n = \frac{1}{\pi} \int_{\omega t=0}^{\omega t=2\pi} E \cos n\omega t \, d(\omega t) \quad (29b)$$

$$D_0 = \frac{1}{2\pi} \int_{\omega t=0}^{\omega t=2\pi} E \, d(\omega t) \quad (29c)$$

where E represents the curve being analyzed.

Examination of Eq. (29) shows that the d-c component D_0 is equal to the average amplitude of the wave when averaged over a full cycle. The B_n component is obtained

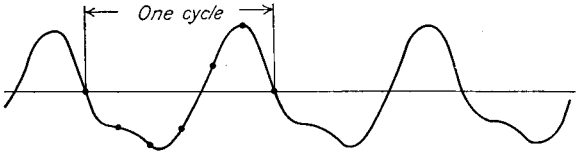


FIG. 47.—Arbitrary curve with six equally spaced ordinates.

by multiplying the curve E by cos $n\omega t$, and dividing the area under the resulting curve for one cycle by π. The A_n coefficient is likewise obtained by multiplying the wave E by sin $n\omega t$, obtaining the area under the resulting product for a full cycle, and dividing by π.

The integrations indicated in Eq. (29) can be carried out graphically even when the wave E cannot be expressed analytically. It is necessary merely to plot the product under the integral sign point by point and obtain the resulting net area (positive minus negative areas) by the use of a planimeter, by Simpson's rule, by counting squares, etc.

The Fourier method gives directly the correct amplitude of any particular component, irrespective of the presence or absence of other components.

Schedule Method.—In the schedule method of wave analysis, one cycle of the wave that is to be analyzed is divided into a number of equally spaced ordinates, as shown in Fig. 47. As many coefficients in Eq. (28*b*) as there are ordinates are then evaluated so that the resulting curve will pass through the selected points. Thus, in Fig. 47, it is possible to obtain an equation that will pass through the selected ordinates by assigning proper values to A_1, A_2, D_0, B_1, B_2, and B_3 while all other coefficients in Eq. (28*b*) are assumed equal to zero. Evaluating the coefficients to accomplish this result involves solving as many simultaneous equations as there are coefficients to be evaluated. By taking advantage of the fact that the ordinates are evenly spaced, the solution of these simultaneous equations can be simplified to the point where it involves performing only a few simple multiplications and carrying out a series of

simple additions and subtractions. These operations are commonly indicated by a form or schedule—hence the name "schedule method."[1] A simple schedule for the case of six ordinates is given in Table 2. Schedules for 12-, 18-, and 36-point analyses are available for the general case where there is a direct-current component and both even and odd harmonics, and also where only odd harmonics are present. These schedules with more ordinates are naturally more complicated than the one shown in the accompanying table, but will likewise evaluate a greater number of coefficients.

TABLE 2.—SIX-POINT SCHEDULE INVOLVING BOTH EVEN AND ODD HARMONICS AND A CONSTANT TERM

Measured ordinates	Sums	Differences		Sine terms	Cosine and constant terms	
				A_1 and A_2	B_1 and B_2	B_0 and B_3
y_0	s_0	d_0				
y_1 y_5	s_1	d_1	sin 30°	.....	$-s_2$ s_1	
y_2 y_4	s_2	d_2	sin 60°	d_1 d_2		
y_3	s_3	d_3	sin 90°	.....	s_0 $-s_3$	$s_0 + s_2$ $s_1 + s_3$
			Sums	S_0' S_e'	S_0'' S_e''	S_0''' S_e'''
				$A_1 = \frac{S_0' + S_e'}{3}$	$B_1 = \frac{S_0'' + S_e''}{3}$	$B_0 = \frac{S_0''' + S_e'''}{6}$
				$A_2 = \frac{S_0' - S_e'}{3}$	$B_2 = \frac{S_0'' - S_e''}{3}$	$B_3 = \frac{S_0''' - S_e'''}{6}$

CHECKS

$$s_0 = (B_0 + B_3) + (B_1 + B_2)$$
$$s_2 = 2(B_0 + B_3) - (B_1 + B_2)$$
$$s_0 + s_2 = 3(B_0 + B_3) \qquad 2s_0 - s_2 = 3(B_1 + B_2)$$
$$s_1 = 2(B_0 - B_3) + (B_1 - B_2)$$
$$s_3 = (B_0 - B_3) - (B_1 - B_2)$$
$$s_1 + s_3 = 3(B_0 + B_3) \qquad s_1 - 2s_3 = 3(B_1 - B_2)$$
$$d_1 = 2(A_1 + A_2) \sin 60°$$
$$d_2 = 2(A_1 - A_2) \sin 60°$$

Procedure.—The measured ordinates are first written down in two columns in the order indicated. In the next two columns appear the sums s_m of the ordinates, found by adding those in the same row, and the differences d_m of the same ordinates. In the fifth column are indicated the trigonometric functions which enter into the calculation. The rest of the schedule indicates in an abbreviated form what products are to be formed, the convention being adopted that each quantity s_m or d_m is to be multiplied by the sine of the angle which appears in the same row at the left. Thus one forms the product $d_1 \sin 60°$ in one case, of $-s_2 \sin 30°$ in another case, and $(s_1 + s_3) \sin 90°$ in still another.

The schedule method evaluates the coefficients so that the resulting curve is correct for the selected ordinates. Between these points, the computed and actual curves will not agree, however, unless the coefficients evaluated are the only coefficients actually present in the wave. Thus, in Fig. 47, the schedule method will not give the correct result if there is some fourth harmonic present. It is therefore necessary

[1] The method of deriving schedules is described by F. W. Grover, Analysis of Alternating-current Waves by the Method of Fourier, with Special Reference to Methods of Facilitating Computations, Reprint 203, *Bur. Standards Bull.*, Vol. 9, 1913. This publication gives a number of schedules.

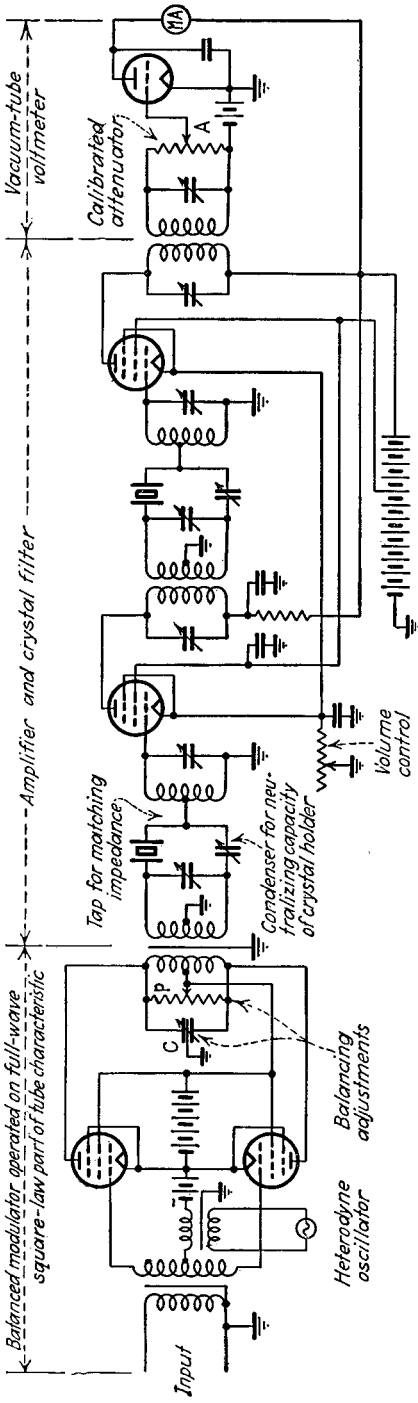


FIG. 48.—Schematic diagram of heterodyne wave analyzer.

to use enough ordinates to permit the evaluation of all important frequency components present in the wave.

20. Wave Analyzers.—A wave analyzer is an instrument for experimentally determining the frequency components of a complex wave. Some of the more important practical types of wave analyzers are described below.

Heterodyne Wave Analyzer.[1]—In the heterodyne method of wave analysis, the complex wave to be investigated is modulated on a carrier by means of a balanced modulator, and the carrier frequency is adjusted so that a side-band component corresponding to the desired component of the complex wave has a predetermined frequency. The output of the balanced modulator then goes to a highly selective amplifier that responds to this particular predetermined frequency, but to no other frequency, and the amplifier output is then measured by a vacuum-tube voltmeter or other device. Calibration is usually accomplished by inserting a known 60-cycle voltage in the system and then adjusting the amplifier gain to give a standard output. A wide range of voltages can be measured by employing an attenuator to the input of the first tube, together with an adjustable gain control in the fixed amplifier. A schematic circuit diagram of a typical heterodyne wave analyzer is shown in Fig. 48.

There are two types of selective fixed amplifiers suitable for use in heterodyne wave analyzers. One arrangement employs a crystal filter having a band-pass characteristic centered at a frequency of about 50 kc. A selectivity curve of an amplifier of this type is shown in Fig. 49, and is characterized by a flat re-

[1] Further discussion of this method is given by L. B. Arguimbau, Wave Analysis, *General Radio Experimenter*, Vol. 7, p. 12, June, 1933; C. R. Moore and A. S. Curtis, An Analyzer for the Voice Frequency Range, *Bell System Tech. Jour.*, Vol. 6, p. 217, April, 1927. The latter reference discusses especially the problem of avoiding false indications produced by high-order modulation products. A modified form to heterodyne wave analyzer capable of operating up to 100 kc is described by A. G. Landeen, Analyzer for Complex Electric Waves, *Bell System Tech. Jour.*, Vol. 6, p. 230, April, 1927.

sponse over a frequency band about four cycles wide, with very sharp discrimination against frequencies outside this band. The other arrangement employs ordinary resonant circuits in which the selectivity can be made high, and also controllable, by the use of negative feedback.[1] The basic circuit arrangement, shown in Fig. 50, makes use of a resonant circuit LC associated with resistors $R_1R_2R_3$ such that at the resonant frequency the negative feedback through R_2R_3 is balanced by positive feedback through R_1LC. At frequencies off resonance, this is no longer the case, and there is a net negative feedback that reduces the output. This reduction in output is in addition to that arising from the selectivity of the tuned circuit, and so increases the effective selectivity or effective Q of the resonant circuit, but does so without changing the response at resonance. The amount of selectivity increase obtained in this way depends upon the amount of negative feedback, and can be varied by the potentiometer P as shown. The selectivity characteristics of a commercial four-stage system of this type operating at approximately 20 kc is shown by the dotted lines in Fig. 49. The maximum selectivity compares favorably with that given by the crystal filter, and there is the added flexibility of ability to control the selectivity.

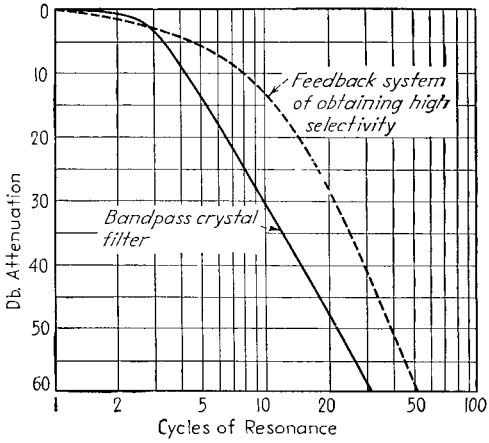


FIG. 49.—Selectivity curves of heterodyne wave analyzers.

The heterodyne wave analyzer is the most widely used device for obtaining the frequency components of a complex audio-frequency wave. It can be built so that the full audio-frequency range is covered by one sweep of the tuning dial controlling the heterodyne frequency, which greatly facilitates the search for frequency components. The chief limiting feature of a heterodyne wave analyzer is that the first

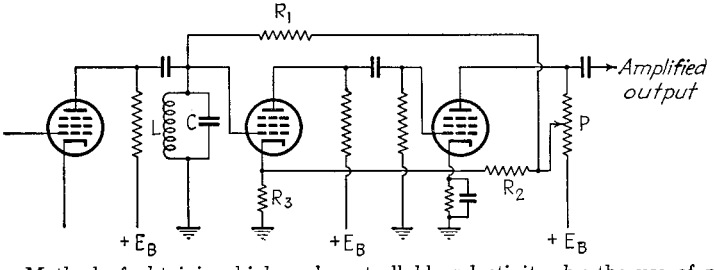


FIG. 50.—Method of obtaining high and controllable selectivity by the use of negative feedback.

detector introduces spurious cross-modulation products that cannot be kept lower than about 60 or 70 db below the desired output. This sets a limit to the smallest frequency component that can be measured in the presence of large voltages of other frequencies.

[1] F. E. Terman, R. R. Buss, W. R. Hewlett, and F. C. Cahill, Some Applications of Negative Feedback with Particular Reference to Laboratory Equipment, *Proc. I.R.E.*, Vol. 10, p. 649, October, 1939.

Fundamental Suppression Methods of Distortion Measurement.[1]—The rms distortion of a wave is defined as the effective value of the harmonics, divided by the fundamental frequency; *i.e.*

$$\text{rms distortion} = \frac{\sqrt{I_2^2 + I_3^2 + I_4^2 + \cdots}}{I_1} \tag{30}$$

where I_2, I_3, etc. are the effective values of the various distortion components and I_1 is the effective value of the fundamental.

The rms distortion can be obtained by suppressing the fundamental frequency and then measuring the part of the wave that remains, with the use of a thermocouple or square-law vacuum-tube voltmeter. Suppression of the fundamental can be accomplished by the use of a high-pass filter so designed that the harmonics lie in the pass band, while the fundamental is severely attenuated. An alternative arrangement is to employ a bridge, or bridged-T network, that is balanced for the fundamental frequency but is decidedly unbalanced for the harmonics. Examples of such networks

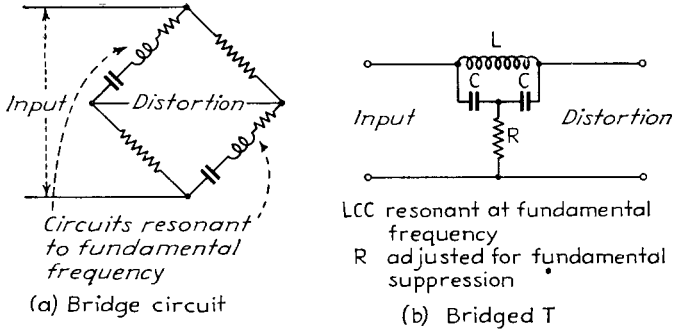


FIG. 51.—Typical networks used in measuring distortion by suppression of the fundamental.

are shown in Fig. 51. By employing resonant circuits in these networks that have a Q of at least 3 to 5, it is possible completely to suppress the fundamental frequency and leave the harmonics practically unaffected.[2] A commercial distortion meter based upon the network of Fig. 51*b* is shown in Fig. 52. Here switch S is at first thrown to A, and the output indication is observed after condenser C and resistance R of the filter network have been adjusted for fundamental suppression (as indicated by minimum output, indication). Switch S is now thrown to B, and the attenuator is adjusted to give the same output indication as before. The attenuator reading is then the rms distortion in db below the fundamental.

Distortion-measuring instruments based on the suppression of the fundamental frequency are much simpler and less expensive than wave analyzers of the heterodyne type. However, they give only the total distortion and not the amplitudes of the individual distortion components; they can be used only in dealing with a fundamental and its harmonics, and not in such cases as two independent waves that have fundamental frequencies unrelated to each other, and they are convenient only when measurements are to be made at a few specified fundamental frequencies.

Miscellaneous Methods of Wave Analysis.—A complex wave can be analyzed by applying it to a tuned amplifier and adjusting the resonant circuits to separate out the

[1] Additional discussion of this subject is given by D. F. Schmit and J. M. Stinchfield, Measuring Harmonic Distortion in Tube Circuits, *Electronics*, Vol. 1, p. 79, May, 1930; H. M. Wagner, A Note on the Fundamental Suppression in Harmonic Measurements, *Proc. I.R.E.*, Vol. 23, p. 85, January, 1935; J. H. Piddington, A Fundamental Suppression Type Harmonic Analyzer, *Proc. I.R.E.*, Vol. 24, p. 591 April, 1936.

[2] When $Q = 3$, the second harmonic suffers only 0.5 db attenuation in the circuit of Fig. 51*b*.

component to be measured so that it can be separately evaluated.[1] This arrangement is not convenient for general audio-frequency work, however, since if a considerable frequency range is to be covered, the manipulation of the tuned circuits is slow, particularly when the exact frequencies contained in the wave being analyzed are unknown. It is also necessary to make a separate calibration for each frequency.

A modified form of tuned analyzer widely used in analysis of sound is illustrated schematically in Fig. 53.[2] This consists of a two-stage amplifier provided with a negative feedback circuit consisting of a parallel-T null network. At the null frequency, there is no feedback, and the full gain of the amplifier is realized. However,

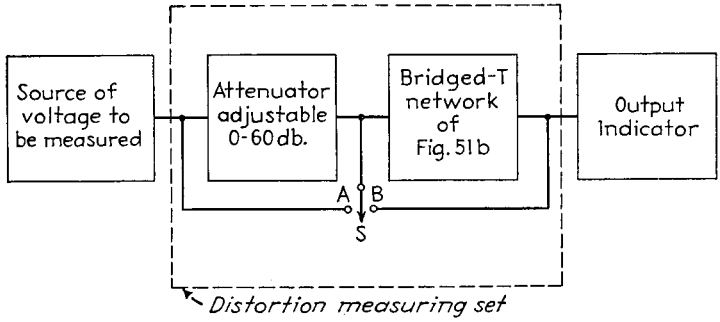


FIG. 52.—A practical distortion meter based on the network of Fig. 51b.

at other frequencies, there is feedback that reduces the gain and results in a response-curve as shown. The null condition is obtained when

$$C_5 = C_6 = \frac{C_7}{2}$$
$$R_5 = R_6 = 2R_7 \tag{31}$$

The null frequency f_0 occurs when

$$f_0 = \frac{1}{2\pi R_5 C_5}$$

This frequency can be readily controlled by simultaneously varying R_5, R_6, and R_7. It is possible in this way to cover a 10 to 1 range, with decimal multiplying factors available by varying C_5, C_6, and C_7 in decade steps. A wave analyzer of the type shown in Fig. 53 provides constant percentage selectivity over a wide frequency range, in contrast with the heterodyne type of analyzer that gives a constant band width in cycles.

The amplitude of a particular frequency component of a complex wave of low or moderate fundamental frequency can be determined by passing the wave through one coil of a dynamometer instrument, while a search current of controllable frequency is passed through the other coil.[3] The operation of the device depends upon the fact that the instrument pointer will not be deflected unless the frequency of the search current is equal to or very close to the frequency of the component being measured. When the difference between the two frequencies is a fraction of a cycle

[1] An instrument of this type for the frequency range 25 to 3,000 cycles is described by R. G. McCurdy and P. W. Blye, Electrical Wave Analyzers for Power and Telephone Systems, *Trans. A.I.E.E.*, Vol. 48, p. 1167, October, 1929.

[2] H. H. Scott, A New Type of Selective Circuit and Some Applications, *Proc. I.R.E.*, Vol. 26, p. 226, February, 1938.

[3] M. G. Nicholson and W. M. Perkins, A Simple Harmonic Analyzer, *Proc. I.R.E.*, Vol. 20, p. 734, April, 1932.

per second, the pointer will pulsate at the difference frequency with an amplitude of pulsation equal to $I_n I_s$, where I_n is the effective amplitude of the unknown component (corresponding to coefficient C_n in Eq. (28*a*), and I_s is the effective amplitude of the search current. Thus, to analyze a wave, one varies the frequency of the search current and notes the frequencies at which beats occur. Each appearance of beats indicates a frequency component in the unknown wave equal to the search frequency, having an amplitude proportional to the amplitude of the beats.

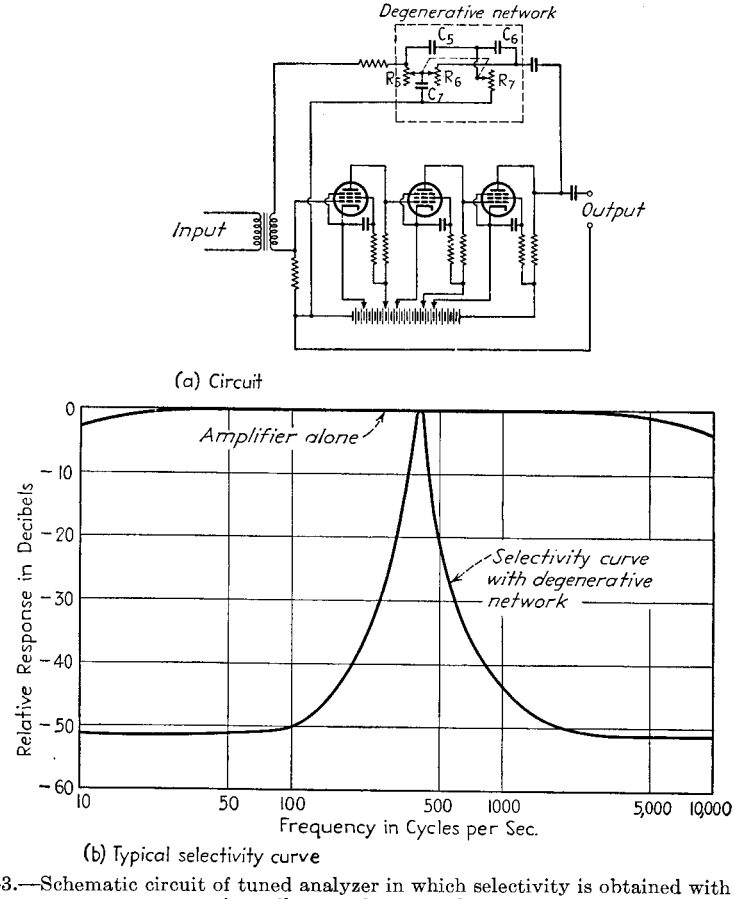


FIG. 53.—Schematic circuit of tuned analyzer in which selectivity is obtained with the aid of a null network in a feedback circuit.

The dynamometer method of analyzing a complex wave has the merits of simplicity and directness. The ability of the method to measure accurately small harmonic components in the presence of large components of other frequencies is determined by the allowable power dissipation in the dynamometer coils, since this limits the current that can be passed through the coils and hence the amplitude of the beats. With practical instruments of the proper sensitivity, there is no difficulty in measuring the distortion commonly encountered in power amplifiers.

Another simple method of harmonic analysis, not limited to the low frequencies at which a dynamometer instrument will operate, involves superimposing the search

voltage upon the wave being analyzed and then applying the combination to a *full-wave square-law vacuum tube* voltmeter, as shown in Fig. 54.[1]

In such an arrangement, the vacuum-tube voltmeter indicating meter will give a steady deflection, depending only upon the effective value of the combined wave. When the search voltage has a frequency within a fraction of a cycle of some frequency component contained in the unknown wave, then beats are superimposed upon this steady deflection. These beats have a crest amplitude proportional to $E_n E_s$, where E_s is the crest amplitude of the search-frequency voltage and E_n is the crest amplitude of the frequency component of the unknown wave that differs from the search frequency by only a fraction of a cycle. One can thus measure amplitude and frequency of each component of the unknown wave by varying the search frequency and noting the frequencies at which beats occur and the amplitude of the beats. This arrangement is simple and direct, but has the disadvantage that the ability to measure small harmonic components in the presence of large components of other frequencies is limited, *first*, by the fact that the total voltage applied to the vacuum-tube voltmeter (sum of unknown and search voltages) must be limited to the square-law part of the tube characteristic and, *second*, by the most sensitive meter that is practicable to use in the output of the vacuum-tube voltmeter.

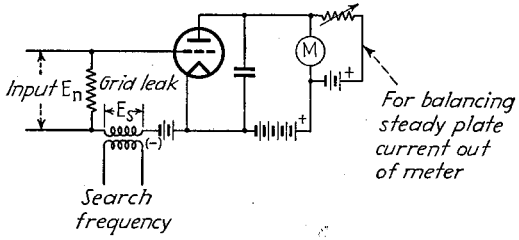


FIG. 54.—Vacuum-tube voltmeter method of analyzing wave forms.

21. Phase Measurement.[2]—The cathode-ray tube is the most widely used instrument for obtaining the phase differences between two voltages. The usual procedure[3] consists in applying one wave to the horizontal deflecting plates and the other wave to the vertical deflectors.[4] This gives an elliptical pattern on the cathode-ray tube, the exact character of which depends upon the relative phase and relative amplitudes of the two voltages involved. Patterns in typical cases are shown in Fig. 55. The phase difference θ between the two waves is given by the formula

$$\sin \theta = \pm \frac{B}{A} \tag{32}$$

where A and B have the significance shown in Fig. 55. The quadrant must be worked out from the orientation of the major axis of the ellipse and the direction in which

[1] C. G. Suits, A Thermionic Voltmeter Method for the Harmonic Analysis of Electrical Waves, *Proc. I.R.E.*, Vol. 18, p. 178, January, 1930.

[2] A method of obtaining the curve of phase shift as a function of frequency of a network, using cathode-ray presentation, is described by B. D. Loughlin, A Phase-curve Tracer for Television, *Proc. I.R.E.*, Vol. 29, p. 107, March, 1941.

[3] Extensive discussions of the subject are given by M. Levy, Methods and Apparatus for Measuring Phase Distortion, *Elec. Comm.*, Vol. 18, p. 206, January, 1940; H. Nyquist and S. Brand, Measurement of Phase Distortion, *Bell System Tech. Jour.*, Vol. 9, p. 522, July, 1930.

[4] An alternative procedure that has been proposed consists in applying one wave to one set of deflectors and using a pulse derived from the reference voltage to increase momentarily the spot intensity at a given phase position by applying this pulse to the control electrode of the cathode-ray tube. For further details, see B. D. Loughlin, Vector Response Indicator, *Trans. A.I.E.E.*, Vol. 59, p. 355, June, 1940.

the spot travels. Uncertainty as to the direction in which the spot travels can always be eliminated by shifting the phase of one of the deflecting voltages in a known direction and noting the effect on the pattern.[1]

Phase shift can also be measured by a variety of other methods. Thus if the amplitudes of two voltages between which the phase difference is desired are measured individually and then their sum is measured, one can construct a vector triangle and obtain the phase difference. Also, it is possible to superimpose the two voltages, read their sum with a vacuum-tube voltmeter, and then shift the phase of one voltage by means of a phase shifter until the indication is either maximum or minimum, indicating the two voltages either in the same phase or phase opposition, respectively.[2]

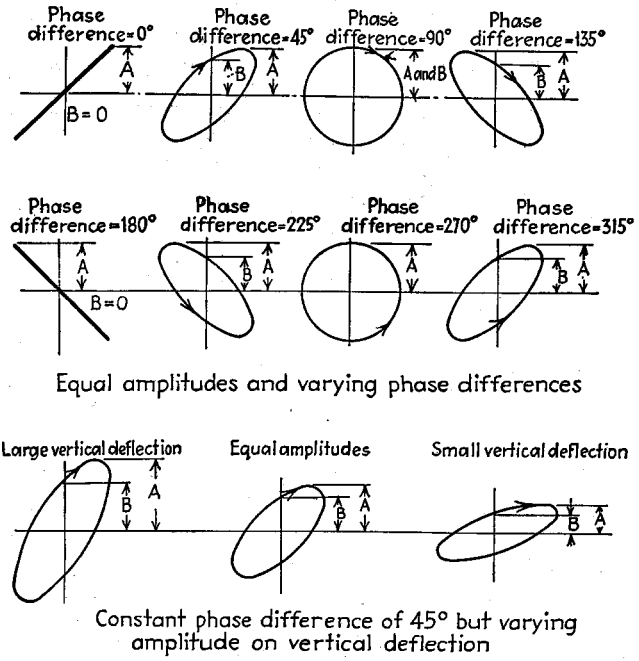


FIG. 55.—Typical patterns produced by cathode-ray tube when sinusoidal voltages of the same frequency but differing in phase and amplitude are applied to the horizontal and vertical deflectors.

Another possibility is to pass one wave through one coil of dynamometer instrument and the other wave through the other coil. The instrument will then read $I_1 I_2 \cos \theta$, where I_1 and I_2 are the currents in the two coils of the meter and θ is the phase difference between these currents.

The phase difference between two waves is not affected by heterodyning both waves simultaneously with the same voltage and rectifying to convert to a new frequency. This permits the actual determination of phase difference to be made

[1] Other methods of determining the direction of travel of the spot are described by E. R. Mann, A Device for Showing the Direction of Motion of the Oscillograph Spot, *Rev. Sci. Instruments*, Vol 5, p. 214, June, 1934; J. R. Haynes, Direction of Motion of Oscilloscope Spot, *Bell Lab. Rec.*, Vol. 14, p. 224, March, 1936.

[2] An alternative arrangement consists in applying the two waves to the two deflectors of a cathode-ray tube and shifting phase until a line is obtained instead of a ellipse. A phasemeter based on this principle is described by J. P. Taylor, Cathode-ray Antenna Phasemeter, *Electronics*, Vol. 12, p. 62, April, 1939.

at any convenient fixed frequency merely by employing a heterodyne oscillator and detector adjusted to give the desired difference frequency.[1]

22. Phase Shifters.—Phase-shifting devices find use in the measurement of phase and also have other applications in measurement work. A great many arrangements for shifting phase have been devised, of which a few of the more common are shown in Fig. 56.[2] At *a*, varying the resistance R causes the phase of the voltage E_{ab} to vary continuously over a 180° range with no change in magnitude.[3] The circuit of Fig. 56*b* employs a capacity load in the plate circuit of a pentode tube, causing the voltage between plate and control grid to differ in phase by 90°. The voltage between the slider on potentiometer P and ground can accordingly be varied continuously over a 90° range by means of this slider. The disadvantage is that the amplitude of the voltage is not constant, although it can be made the same at the two extreme positions

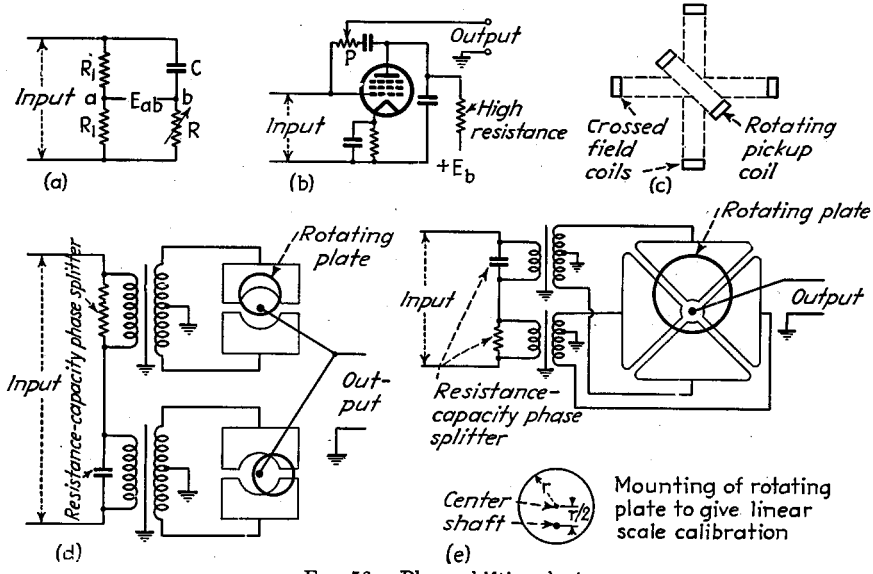


FIG. 56.—Phase-shifting devices.

if the transconductance of the tube is adjusted so that the amplification is exactly unity.

The circuit of Fig. 56*c* employs a rotating magnetic field produced by two crossed coils carrying current differing in phase by 90°. When a pickup coil is placed in the field of such an arrangement, the phase of the induced voltage will depend upon the position and can be varied through 360° by merely rotating the coil.

Electrostatic phase shifters are shown in Figs. 56*d* and 56*e*. The first of these employs two double-stator condensers with specially shaped rotors mounted on a

[1] Phase-measuring equipment making use of this principle is described by Levy, *loc. cit.*; H. T. Friis, Oscillographic Observations on the Direction of Propagation and Fading of Short Waves, *Proc. I.R.E.*, Vol. 16, p. 658, May, 1928; R. R. Law, A New Radiofrequency Phase Meter, *Rev. Sci. Instruments*, Vol. 4, p. 537, October, 1933.

[2] A general survey of bridge type of phase shifters is given by F. A. Everest, Phase Shifting up to 360 Degrees, *Electronics*, Vol. 14, p. 27, November, 1941.

[3] This arrangement can be readily extended to cover other quadrants by means of additional circuits combined with a suitable switching arrangement; see K. Kreielsheimer, Phase Adjuster, *Wireless Eng.*, Vol. 17, p. 439, October, 1940. A slightly different arrangement is described by O. O. Pulley, A Continuously Variable Phase-shifting Device, *Wireless Eng.*, Vol. 13, p. 593, November, 1936.

common shaft.[1] By supplying voltages that are in quadrature to the two pairs of stators as shown, the voltage between the rotor system and ground can be varied through 360° in phase by rotating the shaft. The arrangement at Fig. 56*e* is a modification of this, and involves a single circular rotating plate mounted with the axis off center, as shown, with four stators.[2]

FREQUENCY MEASUREMENTS

23. Standards of Frequency.—The fundamental standard of frequency is the period of rotation of the earth. This can be measured with great accuracy by astronomical methods,[3] and might be thought of as a standard frequency of one cycle per day. All standards of frequency must ultimately be referred to this fundamental source of frequency for calibration sources.

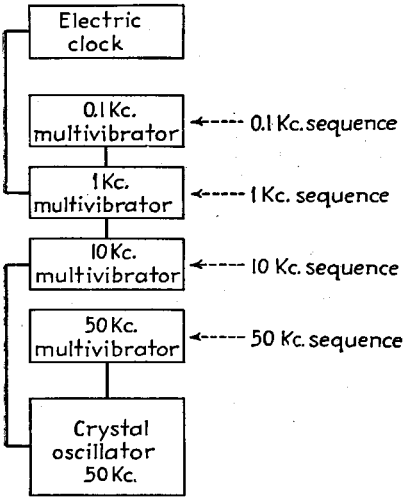


FIG. 57.—Primary standard of frequency with typical multivibrator arrangement.

Primary Standards of Frequency.—A primary standard of frequency is an oscillator that generates a frequency that is very constant over long periods of time and that is checked against the earth's rotation at regular intervals. All commercial primary standards of frequency employ crystal oscillators in which every possible precaution is taken to obtain the maximum possible frequency stability. Thus, the crystal is always a low-temperature coefficient type, and is maintained in a constant temperature oven; the associated electrical circuits are proportioned so that slight changes in their constants will have a minimum of effect on frequency, etc.[4] The frequency of the crystal in a primary standard is usually adjusted to exactly 50 or 100 kilocycles, with means provided to correct the frequency over a range of a few cycles by altering the electrical network associated with the crystal. The crystal of a primary standard is used to control a chain of multivibrators that reduces the crystal frequency to an audio value, commonly 1 kc, which is used to operate an electric clock that will keep exact time provided that the primary standard operates exactly on the proper frequency. Provision is made for accurately comparing the time kept by this clock with time signals sent out by government radio stations and obtained by them from observatories.

Harmonic sequences for measuring purposes are obtained by multivibrators directly or indirectly controlled by the crystal oscillator of the primary standard. A typical multivibrator arrangement is illustrated schematically in Fig. 57. Here the

[1] For details, see J. F. Morrison, A Simple Method of Observing Current Amplitude and Phase Relations in Antenna Arrays, *Proc. I.R.E.*, Vol. 25, p. 1310, October, 1937.

[2] This was developed in the Bell Telephone Laboratories.

[3] Thus see J. F. Hellweg, Time Service of the U.S. Naval Observatory, *Trans. A.I.E.E.*, Vol. 51, p. 538, June, 1932.

[4] Other possible primary standards are precision clocks, that have satisfactory accuracy but develop a frequency too low (*i.e.*, one cycle per second) to be used easily in measuring radio frequencies, and the tuning fork. Electrically driven forks have been used for primary standards, and probably have about the same ultimate possibilities as the crystal oscillator, although the latter has been more fully developed. For further information on precision forks, see J. W. Horton, N. H. Ricker, and W. A. Marrison, Frequency Measurement in Electrical Communication, *Trans. A.I.E.E.*, Vol. 42, p. 730, 1923; E. Norrman, A Precision Tuning Fork Frequency Standard, *Proc. I.R.E.*, Vol. 20. p. 1715, November, 1932.

crystal controls multivibrators with frequencies of 50 and 10 kc. The latter, in turn, controls a 1-kc multivibrator, which operates a 100-cycle multivibrator. Buffer amplifiers are normally provided between the multivibrators and the outputs (not shown in Fig. 57). In these amplifiers, the plate load impedance preferably includes an inductance such that the high-order harmonics will be accentuated in comparison with low-order harmonics. Since a single multivibrator will develop harmonics up to the three hundredth to five hundredth that are detectable on a radio receiver, the arrangement illustrated in Fig. 57 makes available for measurement purposes a great multitude of frequencies, all of which have exactly the same degree of precision as the frequency of the crystal oscillator.

All commercial primary standards of frequency will maintain their frequency constant to within one part in a million over long periods of time. By checking daily against time signals, the degree of precision is considerably higher and can be expected to be nearer one part in ten million. The accuracy that this represents is illustrated by the fact that one ten-millionth of the distance between New York and San Francisco is approximately 16 inches.

Secondary Standards of Frequency.—A secondary standard of frequency is a standard whose frequency is determined by comparison with a primary standard or with some other secondary standard that was originally compared with a primary standard. Secondary frequency standards commonly consist of a good crystal oscillator with one or more multivibrators.

The frequency stability of a secondary standard will depend upon the details of design. In a typical case the frequency will be maintained with an accuracy of about one part in a million if periodical checks are made against a primary standard or the standard-frequency transmissions of the U.S. Bureau of Standards.

Frequency Monitors.—Frequency monitors are secondary standards that are used to check the closeness with which a radio transmitter is maintaining its assigned frequency. Frequency monitors are commonly arranged to provide a continuous indication of the frequency deviation of the transmitters being monitored. This is done either by employing a monitor crystal operating at the assigned frequency, in which case the frequency difference is measured directly, as discussed in Par. 26 below, or by having the frequency of the monitor crystal differ by a predetermined amount, usually 1,000 cycles, from the assigned frequency. In this case a frequency meter is used to indicate the extent that the difference frequency between the transmitter and the monitor departs from this preassigned value.

Radio Signals as Frequency Standards.—Certain classes of radio signals are very useful as frequency standards. Thus the U.S. Bureau of Standards carries on a regular schedule of transmissions that include continuous operation at 5,000 kc, except during special periods when operation is at other frequencies, such as 10,000 and 15,000 kc.[1] The transmissions at different times are unmodulated, modulated at 1,000 cycles, at 440 cycles, and one-second pulses modulated at 1,000 cycles. These signals are accurate to better than one part in ten million, and can be used to check secondary frequency standards and to calibrate heterodyne wave meters, etc.

Broadcast stations are also good sources of standard frequencies, since they are in operation nearly continuously, and ordinarily maintain their assigned frequency to within about 5 to 10 parts in a million, and are required to be within 20 cycles of the assigned value.

24. Miscellaneous Methods of Measuring Radio Frequency.—Where the high accuracy given by primary and secondary frequency standards is not required, the frequency can be determined by the use of a wave meter, by calibrated oscillators,

[1] A discussion of the accuracy of these signals is given by E. G. Lapham, Monitoring the Standard Frequency Emissions, *Proc. I.R.E.*, Vol. 23, p. 719, July, 1935.

Lecher wires, or with the aid of a bridge, as discussed below. Such methods are usually accurate to better than 1 per cent, but less than 0.1 per cent. They are simple, inexpensive, and have a wide field of usefulness.

Wavemeters.—A wavemeter is a resonant circuit tuned by an ordinary variable condenser and provided with a calibration that gives the resonant frequency in terms of the condenser setting. The coil and condenser of the wavemeter should be stable with respect to age and handling, and the temperature coefficient should be as small as possible. It is very desirable that the tuned circuit have low losses, since the accuracy with which a frequency setting can be made is determined primarily by the Q of the tuned circuit.

A variation of the usual wavemeter that is useful at very high frequencies is the variation of both coil and condenser simultaneously, as shown in Fig. 58.[1] This increases the frequency range that can be covered with a single coil and condenser combination.

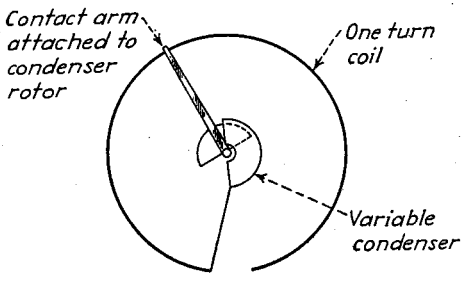


FIG. 58.—Wavemeter for use at ultra-high frequencies, in which both the circuit inductance and capacity are varied.

Resonance between the wavemeter circuit and the oscillations being measured can be determined in a variety of ways. Probably the best arrangement is to employ a diode type of vacuum-tube voltmeter across a portion of the tuned circuit.[2] Such an arrangement will not be damaged by large overloads, will consume negligible power if the diode load resistance is large, and will require only a single flashlight cell to operate the tube. Other methods for resonance indication include a lamp or thermocouple connected in series with the resonant circuit or inductively coupled to the coil, a neon lamp across the tuning condenser, and reaction (absorption) methods. In the latter, a condition of resonance between the wavemeter tuned circuit and the oscillations involved is inferred from the effect that the presence of the wavemeter has on the oscillations, as, for example, dip in the grid current of an oscillator tube.

Heterodyne Frequency Meter.—A heterodyne frequency meter is an oscillator for which a calibration has been made of frequency as a function of the setting of the tuning condenser. The oscillator should be of a type having good frequency stability. A buffer amplifier, preferably a Class C amplifier producing harmonics, is desirable to prevent external circuits from affecting the frequency and to provide a number of frequencies related to the fundamental frequency of the heterodyne frequency meter.

A well-made heterodyne frequency meter will maintain its calibration to 0.1 per cent or better over long periods of time. When provision is made for calibrating one or more points on the dial with the aid of a crystal oscillator, then a precision of the order of 100 parts in a million can be obtained.

Frequency Measurements with Lecher Wires.—At very high frequencies, the wave length may be determined directly by measurements made of the standing wave patterns on a resonant transmission line (often called Lecher wires when used for this purpose).[3] The fundamental idea is illustrated in Fig. 59. As the position of the

[1] E. Karplus, Direct-reading Wavemeter for Ultra-high Frequencies, *Gen. Rad. Exp.*, Vol. 15, p. 1, August, 1940.

[2] At very high frequencies, a crystal can be substituted for the tube to advantage. See E. L. Hall, A Sensitive Frequency Meter for the 30 to 340 Megacycle Range, *Electronics*, Vol. 14. p. 37, May, 1941.

[3] For further information, see Francis W. Dunmore and Francis H. Engel, A Method of Measuring Very Short Radio Wave Lengths and Their Use in Frequency Standardization, *Proc. I.R.E.*, p. 467,

short-circuiting bridge is varied, a series of sharply defined positions will be found for which the transmission line is of the proper length to give resonance, as indicated by a minimum of current through the thermocouple instrument.[1] Minima, rather than maxima, are used, since they are sharper. From the theory of transmission lines, these positions of resonance are almost exactly one-half wave length apart, provided that no dielectric insulation is present in the section of the line involved. The frequency is then

$$f = \frac{300{,}000{,}000}{\lambda} \tag{33}$$

where λ is the wave length in meters and f is in cycles.

In making measurements of frequency with Lecher wires, it is sometimes found that the expected sharp minimum of current is replaced by a broad minimum or even a double minimum. This behavior results from coupling between the parts of the transmission line on the two sides of the short-circuiting bridge. It is accordingly necessary either to associate a shield of considerable diameter with the movable bridge or to place additional short circuits across the unused portion of the wires in order to prevent resonance in them.[2]

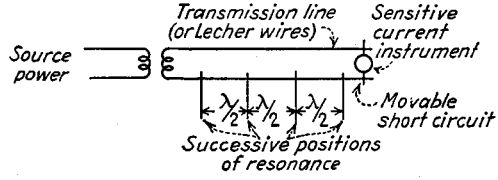


FIG. 59.—Lecher-wire arrangement for measuring wave length at very high frequencies.

Measurements of wave length made with a Lecher-wire system have about the same accuracy as is obtainable with wavemeters, and may reach 0.1 per cent if care is employed. The chief value of this method of measuring frequency is that it can be readily employed at extremely high frequencies, where ordinary measuring methods can be applied only with considerable difficulty.

Bridge and Null Network Methods of Measuring Frequency.—Any radio-frequency bridge that involves a resonant circuit can be used to measure frequency. It is necessary merely to apply the unknown frequency to the input terminals of the bridge or null network, adjust for a balance, and then calculate the frequency from the known circuit constants. In particular, a null network, such as is shown in Fig. 23*d*, can be readily calibrated so that it will be direct reading in terms of frequency.

25. Methods of Comparing Frequency—Interpolation Methods.[3]—When the standard develops a series of fixed frequencies, as is the case with all primary and secondary frequency standards, the measuring problem is one of comparing the

October, 1923. See also August Hund, Correction Factor for the Parallel Wire System Used in Absolute Radio-frequency Standardization, *Proc. I.R.E.*, Vol. 12, p. 817, December, 1924.

[1] A modification of this technique consists in determining when the line is in quarter-wave resonance with the source by using a vacuum-tube voltmeter or a second Lecher-wire system loosely coupled to the active part of the transmission line.

Another modification, in which the experimental results are plotted in a way that tends to increase the accuracy is described by J. Barton Hoag, Measurement of the Frequency of Ultra-radio Waves, *I.R.E.*, Vol. 21, p. 29, January, 1933.

[2] See Eijiro Takagishi, On a Double Hump Phenomenon of Current through a Bridge across Parallel Lines, *Proc. I.R.E.*, Vol. 18, p. 513, March, 1930; A Hikosaburo, Ellipse Diagram of a Lecher Wire System, *Proc. I.R.E.*, Vol. 21, p. 303, February, 1933.

[3] J. K. Clapp, Interpolation Methods for Use with Harmonic Frequency Standards, *Proc. I.R.E.*, Vol. 18, p. 1575, September, 1930; E. G. Lapham, A Harmonic Method of Intercomparing the Oscillators of the National Standard of Radio Frequency, *Proc. I.R.E.*, Vol. 24, p. 1495, November, 1936.

unknown frequency with known frequencies that are slightly different. There are two principal methods by which this can be accomplished: These are, first, the direct measurement of the difference between the unknown and the nearest known frequency, and, second, direct interpolation between the adjacent known frequencies by means of an interpolation oscillator. The direct measurement of the difference frequency is the more accurate but requires somewhat more accessory equipment than does the interpolation method.

A modification is to employ several successive heterodynings between the unknown and the known frequencies to reduce the order of magnitude of the difference to a small residual that can be measured to a small fraction of a cycle.[1] This method requires a rather elaborate equipment, however, and is not so extensively used.

Comparison of Radio Frequencies by Direct Measurement of Difference Frequency.—In this method, the unknown and the nearest known frequencies are combined in a heterodyne detector and the resulting difference frequency measured directly. As normally employed, the known frequencies are the 10-kc sequence provided from a multivibrator controlled by a crystal oscillator. The difference between the unknown and the nearest standard frequency is then less than 5,000 cycles, and can be observed by comparison with a beat-frequency oscillator. The principal problems involved in making such a comparison are the determination of the particular harmonic in the 10-kc sequence that is nearest the unknown, and whether this harmonic is higher or lower than the unknown. The use of an accurately calibrated heterodyne frequency meter is useful in this connection, since by setting to zero beat with the unknown frequency, it will indicate the frequency to within less than 5 kc up to frequencies of at least 3 mc. Also, one can tell if the unknown frequency is greater or less than the known 10-kc harmonic by increasing the frequency of the heterodyne meter slightly. If this increases the audio difference frequency between the nearest harmonic of the 10-kc sequence and the heterodyne frequency meter, the unknown frequency is then greater than the 10-kc harmonic involved, and vice versa.

At frequencies above about several mc, the harmonics of the 10-kc multivibrator are not always of sufficient amplitude to be detected, and it is also difficult to identify the exact harmonic number involved. There is, furthermore, the practical inconvenience of providing the heterodyne frequency meter with sufficient coils to cover all frequency ranges. These limitations are readily overcome by the expedient of using suitable harmonics. Thus, in measuring very high frequencies, one may set a harmonic of the heterodyne frequency meter to zero beat with the signal and then measure the fundamental frequency of the heterodyne frequency meter by the procedure outlined above. In the case of low signal frequencies, on the other hand, one may set the fundamental of the heterodyne frequency meter to zero beat with a harmonic of the unknown frequency, and then measure the fundamental frequency of the heterodyne frequency meter.

Linear Interpolation.—This method makes use of a heterodyne frequency meter whose frequency varies linearly with scale setting. The heterodyne oscillator is successively set to zero beat with the unknown frequency f_x and with the standard frequencies f_1 and f_2 between which f_x lies.[2] This gives three dial settings S_x, S_1, and S_2, and since the frequency intervals are proportional to corresponding dial readings, then

$$\frac{f_x - f_1}{f_2 - f_1} = \frac{S_x - S_1}{S_2 - S_1} \tag{34}$$

[1] See F. A. Polkinghorn and A. A. Roetken, A Device for the Precise Measurement of High Frequencies, *Proc. I.R.E.*, Vol. 19, p. 937, July, 1939.

[2] The zero-beat setting can be done very accurately by allowing the receiver detector to oscillate at a frequency that gives a beat note of about 1,000 cycles with the harmonic in question and observing the waxing and waning of the beats between the heterodyne frequency meter and the unknown frequency.

or

$$f_x = f_1 + \frac{S_x - S_1}{S_2 - S_1}(f_2 - f_1) \tag{35}$$

In case the unknown frequency f_x is too high to be within the range of the heterodyne frequency meter, the latter can be adjusted so that a harmonic is in zero beat with the unknown frequency, and then the fundamental frequency of the heterodyne meter determined by interpolation between the adjacent known frequencies.

26. Measurement of Audio Frequencies.—Several means are in use for measuring audio frequencies, most important of which are the comparison method, the use of bridges, and either mechanical or electronic methods of cycle counting.

Comparison Methods.—An unknown frequency can be compared with a known frequency by means of a cathode-ray tube or by determining the beat or difference frequency between the unknown and the nearest harmonic or subharmonic obtainable from the known frequency.

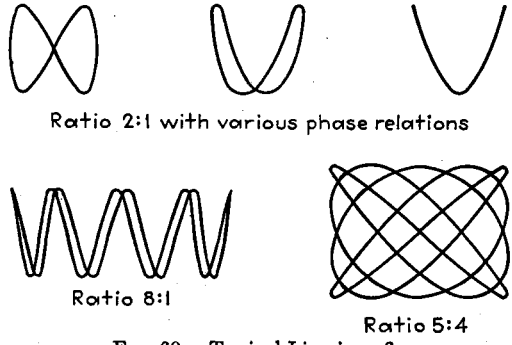


FIG. 60.—Typical Lissajous figures.

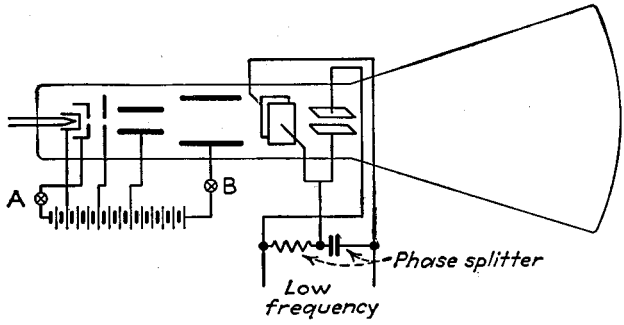
When the standard is adjustable, as in the case of a calibrated oscillator, the cathode-ray tube is the most convenient means of comparison.[1] The usual method of employing a cathode-ray tube is to apply the unknown frequency to one pair of deflecting plates and the known frequency to the other. The resulting pattern will depend upon the frequency ratio and upon the relative phase. If this ratio approaches a value that can be expressed by simple integers, a definite pattern known as a Lissajous figure, is formed. Examples are shown in Fig. 60.

When the ratio of frequencies is an exact integer, the pattern is stationary, and the frequency ratio is the number of times the side of the figure is tangent to a horizontal line divided by the number of times its end is tangent to a vertical line.[2] If the ratio is nearly but not exactly an integer, then the pattern weaves about as though the relative phase of the two deflecting waves was continuously changing, whereas if the ratio of frequencies differs very much from a simple ratio of integers, the pattern is merely a luminous rectangular area. When the frequency ratio is large, the pattern becomes so complicated that it is difficult to determine what the exact ratio is by inspection. In such cases arrangements such as are shown in Fig. 61 can be used. Here the low frequency is caused to produce an elliptical or circular path by use of a resistance-capacity phase splitter. In Fig. 61*b*, the beam intensity is modulated by the high frequency, giving a pattern as shown, with the ratio of high to low frequency

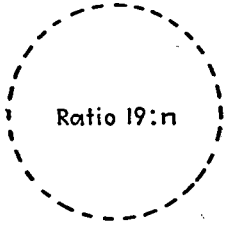
[1] An excellent discussion of this is given by F. J. Rasmussen, Frequency Measurements with the Cathode Ray Oscillograph, *Trans. A.I.E.E.*, Vol. 45, p. 1256, 1936.

[2] This assumes that the forward and return trace do not coincide. If they do coincide, the rule does not hold.

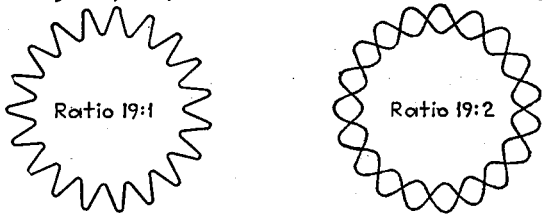
equal to the number of spots divided by some integer n. At Fig. 61c, the high frequency is inserted in series with the anode. This varies the deflection sensitivity and results in a gear-wheel pattern with the ratio of the high to low frequency determined by the number of teeth. In Fig. 61, the wheel or the spots, as the case may be, appear to rotate if the frequency ratio is almost but not exactly expressible as a ratio of simple integers.



(a) Fundamental circuit



(b) Spot-wheel patterns obtained by inserting high frequency at point A to modulate light intensity



(c) Gear-wheel patterns obtained by inserting high frequency at point B to modulate sensitivity

FIG. 61.—Wheel patterns used to compare frequencies.

Instead of using a cathode-ray tube, comparison can be made between an unknown frequency and an adjustable standard frequency by aurally determining the condition for which the beat frequency is zero. It is possible even to compare frequency ratios that are in simple harmonic relation, such as 2 to 1, by aural methods.

When the standard frequency is not adjustable, it is then customary to employ a multivibrator to obtain a series of harmonics related to the standard, as, for example, the 100-cycle sequence in Fig. 57. The unknown frequency is then determined by

measuring directly the difference between the unknown and the nearest known frequency, or by linear interpolation between the known frequencies on either side of the unknown. In the case of the difference-frequency method, if the frequency difference is less than 50 or 100 cycles, it can be determined with considerable accuracy by a counting device, as discussed below.

Bridge Methods of Measuring Audio Frequencies.—An unknown frequency can be determined by applying it to a bridge in which the balance depends upon frequency, and then calculating the frequency from the circuit constants required to give balance. Any type of bridge or null network in which the balance depends on frequency can be used. Typical examples are the Wien, Hay,[1] and resonance bridges and the null network of Fig. 23*d*.

The Wien bridge is particularly satisfactory for measuring audio frequencies, because it contains no inductance to pick up stray magnetic fields, and can be conveniently proportioned to cover a wide frequency range. A typical arrangement is shown in Fig. 62.[2] By proportioning the bridge so that $C_c = C_d$, $R_c = R_d$, and $R_b/R_a = 2$, then the bridge is balanced at a frequency given by

$$f = \frac{1}{2\pi R_c C_c} \qquad (36)$$

Thus, by making R_c and R_d identical slide-wire resistances and mounting them on a common shaft, the dial can be calibrated directly in frequency. Furthermore, convenient multiplying factors, such as decimal values, can be obtained merely by changing the capacities of the condensers C_c and C_d by the appropriate amounts. A frequency range of 10 to 1 can be readily covered with a single condenser value, so that by the use of three pairs of condensers one can cover the complete audio range from 20 to 20,000 cycles. In practical construction, it is impossible to maintain equality between R_c and R_d with the accuracy required to maintain a perfect balance, so that the potentiometer P having a total resistance of perhaps 1 to 2 per cent of R_a is used to sharpen the balance. This has little effect upon the frequency calibration.

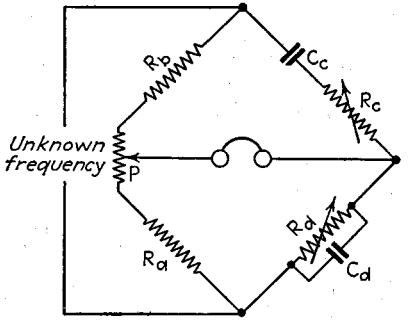


FIG. 62.—Wien bridge arranged for the measurement of frequency.

In the frequency range 300 to 5,000 cycles, balance can be most easily made by the use of telephone receivers. At frequencies outside this range, some indicating arrangement, such as an amplifier and vacuum-tube voltmeter combination, is normally required. The principal difficulty involved in making accurate frequency measurements by bridge methods arises from harmonics of the frequency being measured. The bridge is unbalanced for these harmonics, which are thus very prominent in the bridge output, even though they are only a small percentage in the bridge input current. When telephone receivers are used in the middle audio range, it is usually possible to balance the bridge for the fundamental in spite of the presence of harmonics; but when indicating instruments are employed, appropriate filters must be placed in the output of the neutral arm to prevent spurious voltages from reaching the indicating device.

Measurement of Audio Frequencies by Cycle Counting.—Very low frequencies can be measured by a simple cycle counter operated from a polarized relay.

[1] C. I. Soucy and B. de F. Bayly, A Direct Reading Frequency Bridge for the Audio Range Based on Hay's Bridge Circuit, *Proc. I.R.E.*, Vol. 17, p. 834, May, 1929.

[2] The common network of Fig. 23*d* is equivalent to a Wein bridge, and can be used for frequency measurements by simultaneously varying the three resistances R_2, R_2, and R_1.

When the frequency is still low but is too high for direct counting of cycles, the procedure is to charge a condenser on one-half of each cycle from a source of constant d-c voltage and then discharge the condenser through a resistance on the other half of the cycle. The average or direct current flowing into the condenser is proportional to the number of charges per second, and so to the frequency. A d-c meter placed in the input circuit will accordingly give a deflection proportional to the number of cycles per second, and so can be calibrated to read frequency directly.

At frequencies up to several hundred cycles per second, a polarized relay can be employed to perform the necessary switching operations.[1] At higher frequencies, however, it is necessary to employ an electronic switching device, such as is illustrated in Fig. 63.[2] Here tubes T_1 and T_2 are gas triodes connected in an inverter circuit such that condensers C_m and C'_m are alternately charged from the supply voltage E_b on the positive and negative halves of the cycle. A definite fraction of this charging

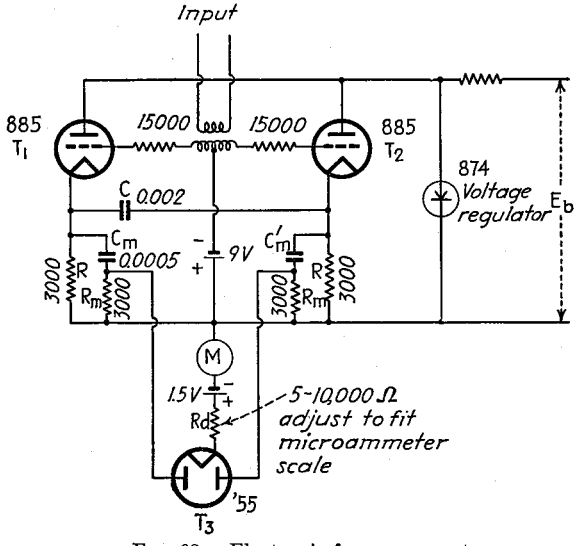


FIG. 63.—Electronic-frequency meter.

current flows through the diode T_3 and the meter M, which accordingly gives a deflection proportional to frequency. The condensers discharge through the resistances R and R_m, the diode preventing any current from returning in the reverse direction through the meter.

The accuracy of electronic cycle counters is of the order of ± 2 per cent, provided that the supply voltage is closely regulated. Any desired frequency range can be covered up to the upper limit of gas triodes, and high vacuum tubes will extend the range still higher.

Determination of Frequency Deviation from a Predetermined Audio Frequency.—It is sometimes desired to obtain an indication of the amount and direction that an unknown frequency deviates from an assigned value. An example is in the case of frequency monitors, where the monitoring crystal differs by a definite amount from

[1] N. P. Case, A Precise and Rapid Method of Measuring Frequencies from Five to Five Hundred Cycles Per Second, *Proc. I.R.E.*, Vol. 18, p. 1586, September, 1930.

[2] F. V. Hunt, A Direct-reading Frequency Meter Suitable for High Speed Recording, *Rev. Sci Instruments*, Vol. 6, p. 43, February, 1935; F. Guarnaschelli and F. Vecchiacchi, Direct Reading Frequency Meter, *Proc. I.R.E.*, Vol. 19, p. 659, April, 1931.

the assigned frequency. An indication of the extent that this difference frequency deviates from its proper value can be obtained with an arrangement such as is illustrated in Fig. 64. Here voltage of the unknown frequency is applied in series to two resonant circuits LC and $L'C'$. One of these circuits is tuned to a frequency slightly higher than the proper frequency, while the other is in resonance slightly below. The two circuits are so proportioned that at the proper difference frequency, the voltages developed across the two tuning condensers are the same. Diode voltmeters are then connected as shown, in such a manner that the meter M reads the difference in rectified outputs. If the voltage applied to the input circuit is then maintained constant, the meter M will be deflected to the right or left of center by an amount indicating

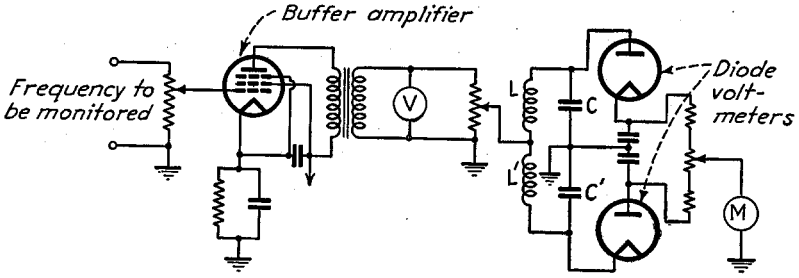


FIG. 64.—Schematic circuit of a frequency-deviation meter.

the extent to which the unknown frequency is higher or lower, respectively, than the "proper" value.

VACUUM-TUBE CHARACTERISTICS

27. Tube Constants.—The principal constants of a vacuum tube are the amplification factor, the plate resistance, and the transconductance. In screen-grid, beam, and pentode tubes, the mu-factor of the screen relative to the control grid with respect to the total space current, and also the dynamic resistance of the screen-grid circuit are of importance. These various constants can be defined as

$$\begin{aligned}
\text{Amplification factor} &= \mu = \left.\frac{\partial E_p}{\partial E_g}\right| I_p \text{ constant} \\
\text{Plate resistance} &= R_p = \frac{\partial E_p}{\partial I_p} \\
\text{Transconductance} &= G_m = \frac{\partial I_p}{\partial E_g} \\
\text{Screen resistance} &= R_{sg} = \frac{\partial E_{sg}}{\partial I_{sg}} \\
\text{Screen mu factor} &= \mu_{sg} = \left.\frac{\partial E_{sg}}{\partial E_g}\right| I_{sp} \text{ constant}
\end{aligned} \qquad (37)$$

where E_g, E_{sg}, and E_p are control-grid, screen, and plate voltages, respectively, and I_p, I_{sg}, and I_{sp} are the plate, screen, and total space current, respectively. Although many other tube constants may be defined, these additional coefficients are usually of little practical use.

Determination of Tube Constants from Static Curves.—The vacuum-tube coefficients defined above can be deduced by evaluating the derivatives involved from the characteristic curves of the tube. Thus the transconductance is given by the slope of the $E_g - I_p$ characteristic, and the plate resistance is the reciprocal of the slope of the $I_p - E_p$ curves, etc.

This method is not particularly accurate but is always available for rough determinations and gives a clear visualization of what each coefficient represents. The method is particularly unsatisfactory in the case of the plate resistance and amplification factor of pentode and similar tubes, since these values are so high that the required ratio cannot be evaluated from ordinary characteristic curves with even fair accuracy.

Experimental Determination of Constants from Increments.—By considering the differentials appearing in Eqs. (37) to be finite increments, the coefficients can be determined experimentally. Thus, to obtain the transconductance, one would increase the control-grid voltage by an increment ΔE_g that could be evaluated from the grid voltmeter, and would observe the resulting change in plate current ΔI_p on a plate milliammeter. Similarly, the amplification factor would be obtained by adding an increment to the grid voltage as above, and then changing the plate voltage by the necessary increment required to leave the plate current unchanged, as indicated by a plate milliammeter.

This method of measuring tube constants is frequently used for approximate determinations. A high degree of accuracy is not easily obtained, however, because the increments used in evaluating the constants must be small, and so are difficult to read accurately.

28. Dynamic Determination of Tube Constants.[1]—The most accurate method of determining tube coefficients involves the use of small alternating current and voltage

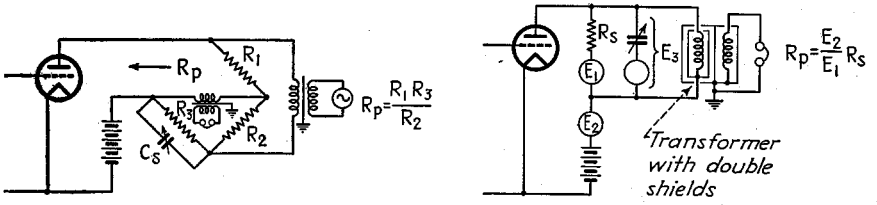


FIG. 65.—Circuits for measurement of electrode resistance.

increments, combined with a circuit arrangement that permits evaluation of the coefficients by a null balance.

Electrode Resistance.—The resistance of any electrode of the tube, such as the plate resistance, can be determined by placing the unknown resistance in the X arm of an a-c bridge. An example of such an arrangement is shown in Fig. 65*a*, where the connection is for the determination of plate resistance. The only special precautions that must be taken are to provide a condenser C_s as shown, to balance the capacity of tube, wiring, and bridge, and to couple the telephone receivers and oscillator to the bridge through transformers, since the bridge is at the d-c plate potential above ground. Electrode resistance can be determined very accurately with such a bridge, and the only disadvantage of the arrangement is that there is a d-c voltage drop in the bridge that varies as the bridge is adjusted to balance.

An alternative means of measuring electrode resistance is shown in Fig. 65*b*.[2] Here voltages E_1 and E_2 are applied as shown, and their ratio adjusted so that a null is obtained in the telephone receiver. The voltage E_3 is to balance out the effect of

[1] For further information, see "Standards on Electronics," Institute of Radio Engineers, New York, 1938; R. W. Hickman and F. V. Hunt, The Exact Measurement of Electron-tube Coefficients, *Rev. Sci. Instruments*, Vol. 6, p. 268, September, 1935; E. L. Chaffee, "Theory of Thermionic Tubes," Chap. IX, McGraw-Hill, New York, 1933; W. N. Tuttle, Dynamic Measurement of Electron Tube Coefficients, *Proc. I.R.E.*, Vol. 21, p. 844, June, 1933.

[2] See Tuttle, *loc. cit.*

any reactive currents in the telephone receivers, and does not influence the balance.

The three voltages E_1, E_2, and E_3 must be in the same phase, and so are conveniently obtained from a transformer with three insulated and shielded secondaries, as shown in Fig. 66. The ratio E_2/E_1 can be controlled by attenuators as indicated, and the polarity and magnitude of the reactive current obtained from E_3 can be controlled with a double stator condenser. The telephone receiver that indicates balance must be connected into the plate circuit with a doubly shielded transformer connected as shown in order to prevent capacity currents from transformer primary to ground from producing a voltage across the secondary terminals.

This voltage-ratio method of measuring plate resistance has the advantage that a very wide range of values can be readily covered and that the d-c voltage drop can be made quite small and constant by using low impedance attenuators of constant impedance. The disadvantage is that a special measuring unit is required to produce the voltages E_1, E_2, and E_3.

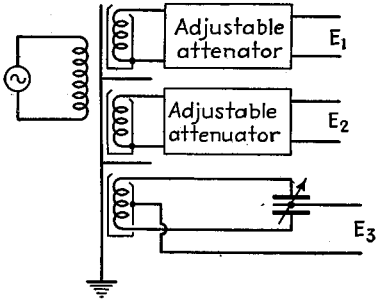


FIG. 66.—Transformer and attenuator arrangements for producing the voltages required in Fig. 65*b*.

Amplification Factor.—The amplification factor, or mu-factor, of a tube is commonly determined by circuit arrangements such as is shown in Fig. 67*a* when the value of the factor is 100 or less. Here an oscillator voltage is applied to resistances R_1 and R_2 in series, with the voltages developed across R_1 and R_2 acting on the grid and plate in the opposite phase. A null balance gives the mu-factor of the electrode associated with R_2 relative to the electrode associated with R_1 with respect to the current passed through the telephone receivers. Balance of stray reactive currents may be obtained by the use of a variable mutual inductance M or by a variable condenser, as shown.

Bridge circuits for measuring amplification factor are satisfactory for moderate values of this constant, although there is the disadvantage that in varying R_2 to obtain balance, the d-c voltage drop in the plate circuit and hence the plate voltage are varied. Also, the plate-supply voltage cannot have its negative lead grounded. The arrangement, however, is quite unsatisfactory in measuring the amplification factor of pentode and similar tubes, because here the required value of R_2 is excessive.

The disadvantages and limitations of the simple mu-factor bridges of Fig. 67*a* can be overcome by the arrangement shown in Fig. 67*b*. This arrangement is analogous to that of Fig. 65*b*, and the voltages E_1, E_2, and E_3 can be obtained as in Fig. 66. This voltage-ratio method of measuring amplification factor has the advantage over the bridge circuits of making it possible to measure very large as well as moderate and small values of amplification factor, and also can be designed to cause a negligible d-c voltage drop in the plate circuit. Furthermore, it permits the operation of bias and plate-voltage sources with one terminal grounded.

Dynamic Measurement of Transconductance.—The simplest and most effective method of determining transconductance is shown in Fig. 68*a*,[1] where R is adjusted to give a null balance in the telephone receiver. In the case of triode tubes where there is direct capacity between grid and plate in shunt with R, a capacity neutralizing arrangement as shown dotted can be used.

Transconductance is frequently measured with the aid of the bridge circuit of Fig. 68*b*. Here it is customary to vary R_2 to obtain balance, and the reactive currents

[1] C. B. Aiken and J. F. Bell, A Mutual Conductance Meter *Communications*, Vol. 18, p. 19, September. 1938.

through the telephone receivers are neutralized with the aid of either a condenser or a variable mutual inductance, as shown in the figure. This bridge arrangement will satisfactorily determine transconductance, but has the disadvantage that the resistance drop in the plate return circuit of the tube varies with the adjustment of the bridge and that neither the grid bias nor anode voltage sources can be returned to ground.

The voltage-ratio method can be employed in the measurement of transconductance as shown in Fig. 68c, where the voltages E_1, E_2, and E_3 are obtained as in

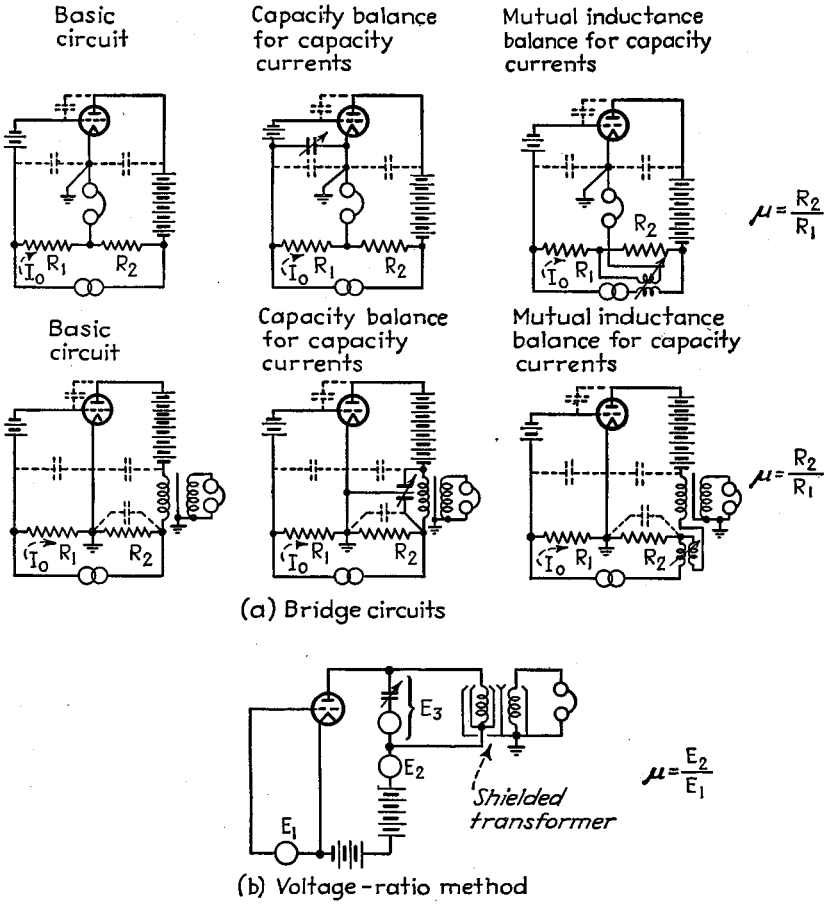


FIG. 67.—Circuit arrangements for measurement of amplification factor or mu factor.

Fig. 66. This method is capable of covering a wide range of values with accuracy, allows the supply voltages to be returned to the grounded cathode, and introduces negligible voltage drop.

Several types of transconductance meters are used for the rough checking of tubes. One arrangement in common use is illustrated in Fig. 69*a*. Here a few volts derived from the secondary of a transformer operated from a 110-volt power line is applied to the grid of the tube, and the resulting alternating current flowing in the plate circuit is read by means of a rectifier instrument that is coupled to the plate

by a transformer. Another arrangement (Fig. 69*b*) makes use of a dynamometer instrument in place of the rectifier meter, with the alternating component of the plate current passed through one coil of the meter and a current derived from the 110-volt

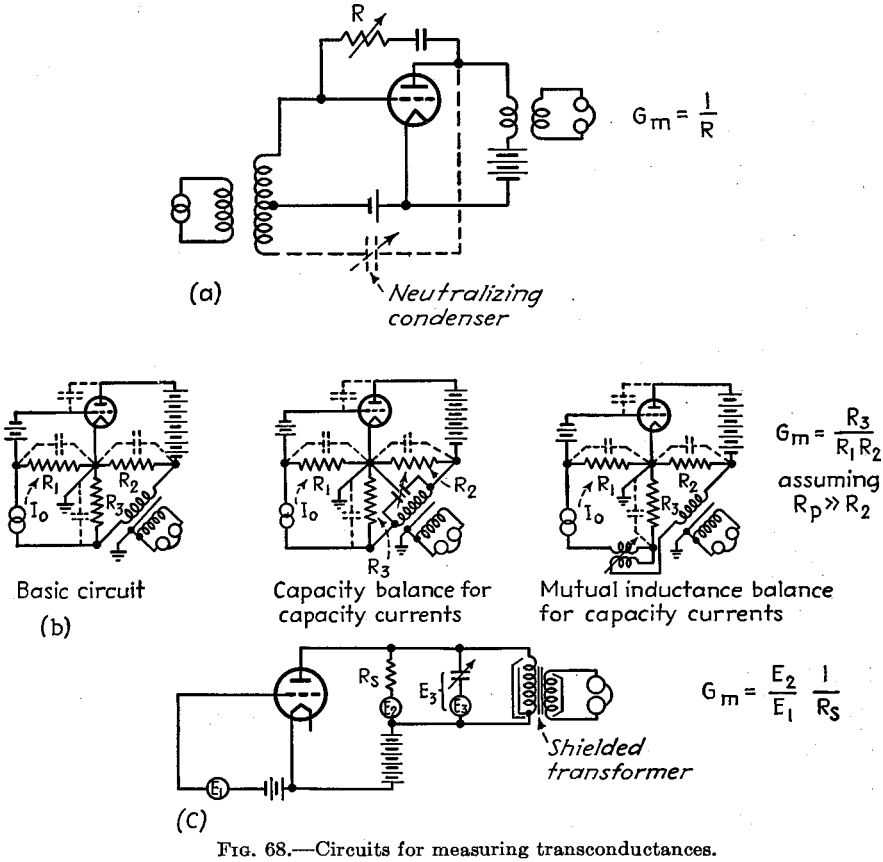


FIG. 68.—Circuits for measuring transconductances.

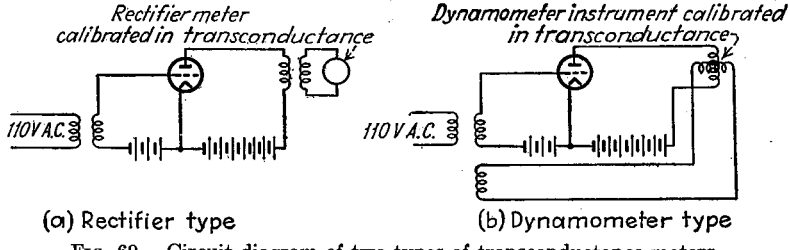


FIG. 69.—Circuit diagram of two types of transconductance meters.

line source through the other. In both these arrangements, the instrument deflection for constant line voltage will be proportional to the transconductance, so that the meter can be calibrated directly in micromhos. The accuracy of both methods depends upon the constancy of the sixty-cycle line voltage, and provision not shown

in Fig. 69 is usually made for adjusting the voltage on the transformer to a standard value by reconnecting the meter so that it serves temporarily as a voltmeter.

29. Determination of Tube Characteristics in the Positive-grid Region.—It is ordinarily not practicable to obtain static characteristics of tubes in the positive-grid region by point-by-point measurement, because the power dissipation will usually exceed the safe value for continuous operation. A typical method[1] devised to overcome this difficulty makes use of two large condensers that are precharged to the desired grid and plate voltages and then are simultaneously connected to the tube with the aid of a thyratron and a tripping circuit. The resulting initial grid and plate currents are then observed with the aid of a cathode-ray oscillograph. This procedure can be modified in a variety of details,[2] and if desired, it is even possible to trace out complete characteristic curves on a cathode-ray oscillograph.[3] Thus, the grid-voltage plate-current characteristic can be obtained by biasing the grid beyond cutoff and then applying to the grid a 60-cycle alternating voltage of suitable amplitude while the plate voltage is maintained constant. A cathode-ray tube connected so that the grid-cathode potential provides the horizontal deflection, while the plate current gives the vertical deflection, will then reproduce the desired characteristic visually.

AMPLIFICATION AT AUDIO AND VIDEO FREQUENCIES

30. Measurement of Voltage Amplification.—The standard method of measuring voltage amplification is to apply a known voltage to the input of the amplifier and

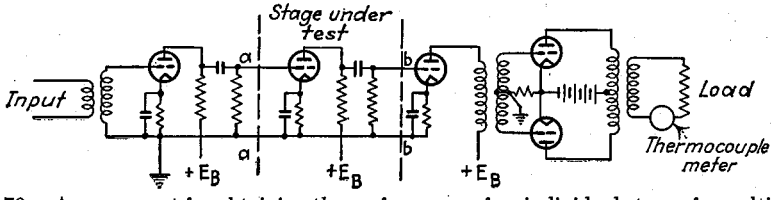


FIG. 70.—Arrangement for obtaining the performance of an individual stage of a multistage amplifier.

observe the resulting output voltage with a vacuum-tube voltmeter. In carrying out these operations, it is usually desirable to maintain the output voltage constant at a value well below the overload limit of the amplifier and then to vary the input voltage as required. It is also important that the output measuring equipment be so arranged as not to affect the behavior of the amplifier. This means that this measuring equipment should, as far as possible, be shunted across points in the amplifier of low impedance level and that, in any case, short leads to vacuum-tube voltmeters must be employed to reduce shunting capacities.

The amplification of individual stages of a multistage amplifier is most satisfactorily obtained by difference. Thus, to obtain the amplification of a stage between points *a* and *b* in Fig. 70, the recommended method is to measure the amplification first from *aa* to the output and then from *bb* to the output. The desired amplification is then the ratio of the two amplifications. By proceeding in this way, the stage under test is operated under normal conditions as far as regeneration is concerned, and the measuring equipment does not in any way modify the amplification characteristic.

[1] O. W. Livingston, Oscillographic Method of Measuring Positive Grid Characteristics, *Proc. I.R.E.*, Vol. 28, p. 267, June, 1940.

[2] H. N. Kozanowski and I. E. Mouromtseff, Vacuum Tube Characteristics in the Positive Grid Region by an Oscillographic Method, *Proc. I.R.E.*, Vol. 21, p. 1082, August, 1933; E. L. Chaffee, Power Tube Characteristics, *Electronics*, Vol. 2, p. 34, June, 1938.

[3] H. F. Mayer, Cathode-ray Tube Applications, *Electronics*, Vol. 11, p. 14, April, 1938; Jacob Millman and S. Moskowitz, Tracing Tube Characteristics on a Cathode Ray Oscilloscope, *Electronics*, Vol. 14, p. 36, March, 1941.

Production of Known Audio-frequency Voltages.—Known audio-frequency voltages from about 1 mv upward can be produced by applying a known voltage to a calibrated voltage divider, as shown in Fig. 71*a*, or by passing a known current through a known resistance, as in Fig. 71*b*. The voltages E indicated in Fig. 71 are those obtained on open circuit, and have an equivalent internal impedance R_{eq}, as shown in the figure. When the voltage is applied to a load, as at Fig. 71*c*, rather than to an open circuit, this internal resistance must be considered as being in series with the load impedance, with the open-circuit voltage applied to the combination as shown.

When audio-frequency voltages of the order of one mv or less are required, or when a wide range of values is to be covered, it is desirable to employ some form of

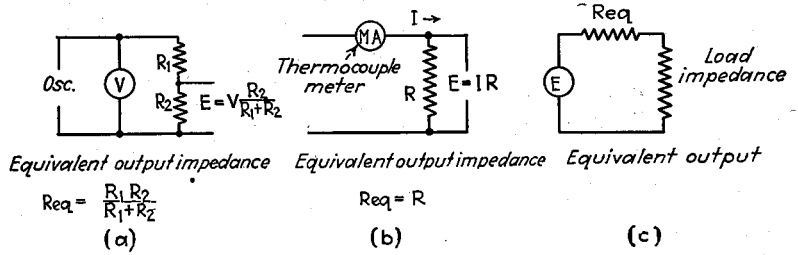


FIG. 71.—Simple methods of producing known voltages, together with equivalent output circuit for calculating reduction in voltage produced by a load connected to the output terminals.

audio-frequency signal generator or microvolter. A schematic diagram of a typical arrangement is illustrated in Fig. 72. It is preferable that the output impedance of such a device be constant, irrespective of the attenuation, since it is then possible to determine, with the aid of Thévenin's theorem, the effect of a finite load impedance placed across the output.

Known voltages at the higher video frequencies can be produced by the simple arrangements in Fig. 71 when these voltages are not too small. If very small, then it is customary to use an ordinary signal generator of the type employed in the testing of radio receivers.

Phase-shift in Amplifiers.—The phase shift in amplifiers can be most easily determined by the use of a cathode-ray oscillograph, as in Fig. 73. Here a relatively

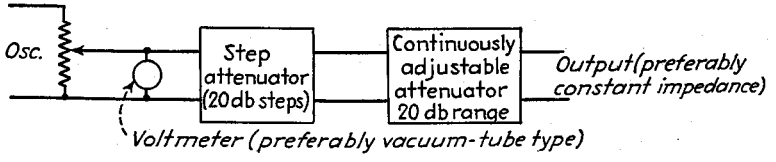


FIG. 72.—Schematic circuit of a microvolter for producing known audio-frequency voltages.

large voltage derived from the oscillator is applied to one set of deflecting plates, while a small fraction of this voltage of the same phase is applied by means of a resistance voltage divider (or attenuator) to the amplifier input. The output of the amplifier is then applied to the other deflecting plates, and the phase difference between the voltages derived from the elliptical pattern that results.[1]

The presence of a time-delay error, particularly at low frequencies, can be readily determined with the aid of a square-wave generator, as discussed below. This

[1] If the amplifiers in the cathode-ray oscilloscope are used in making these measurements, they must be checked to make sure that the phase shift is the same for both vertical and horizontal amplifiers. If not, suitable corrections must be made in the results.

method is very sensitive, and will detect phase shifts that lead to errors in time delay that are too small to be detected by the cathode-ray method.

Curve Tracing Systems for Observing and Recording Amplification Characteristics.—A number of arrangements have been devised for tracing directly the amplification curve as a function of frequency. The simplest arrangement of this type is shown in Fig. 74.[1] Here a beat-frequency oscillator is used that gives a constant output irrespective of frequency, and a fraction of the output voltage is applied to the amplifier input. The output of the amplifier is rectified and the resulting d-c voltage amplified either on a linear or logarithmic vacuum-tube amplifier, and applied to the vertical deflectors of a cathode-ray oscilloscope. A deflecting voltage proportional to the logarithm of frequency is obtained by applying a voltage proportional to the input voltage of the amplifier to the network N. The voltage developed across the resistance R in this network is rectified, applied to a linear d-c amplifier, and used to provide the horizontal deflection. By proportioning the network as shown, the voltage developed across its output is almost exactly proportional to the logarithm of frequency, as indicated in Fig. 74*b*. The result is a horizontal scale proportional to the logarithm of frequency, with a vertical deflection that gives amplification, either on a linear or decibel scale, according to whether a linear or logarithmic vertical deflecting amplifier was used. By making use of a cathode-ray tube having a screen material that gives very high persistence, it is accordingly possible to sweep the beat-frequency oscillator through the frequency range by hand in a period of 20 to 30 seconds, and so trace out the amplification curve. This may be observed visually on the cathode-ray screen and also photographed, if desired.[2]

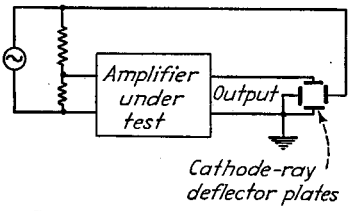


FIG. 73.—Circuit for determining phase shift of an amplifier with the aid of a cathode-ray tube.

Devices that trace out the amplification curve on a piece of paper, with the use of ink or similar recorders, have also been developed.[3] In these, the shafts of the tuning dial of the beat-frequency oscillator and of the recorder drum are normally linked mechanically. Any one of several types of curve-tracing mechanisms may be employed, a typical arrangement being a standard recording meter operated by rectifying and amplifying the amplifier output. A logarithmic frequency scale can be obtained either by using a beat-frequency oscillator with suitably shaped condenser plates or by employing suitable cams in the mechanical link. A logarithmic amplification scale can be obtained by using a logarithmic vacuum-tube amplifier to operate the recording pen.

31. Amplitude Distortion in Amplifiers.—Amplitude distortion in audio- and video-frequency amplifiers can be most readily measured by applying a sine-wave voltage to the amplifier input and investigating the output with a wave analyzer. Unless the amplitude distortion is large, it is necessary that care be taken to ensure the purity of wave form of the input voltage. This requires either a very good oscillator or the use of suitable filter between oscillator and amplifier input.

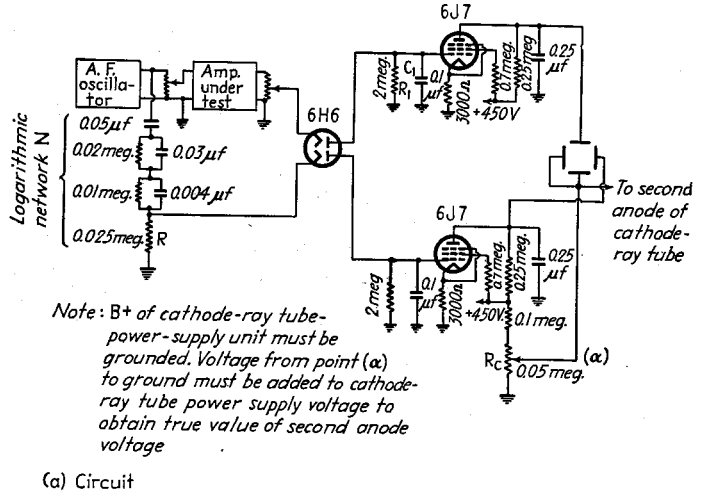
[1] See An Audio-frequency Curve Tracer Using a Cathode Ray Tube, *R.C.A. Mfg. Co. Application Note* 76, June, 1937.

Another arrangement is described by S. F. Carlisle, Jr., and A. B. Mundell, Frequency Response Curve Tracer, *Electronics*, Vol. 14, p. 22, August, 1941.

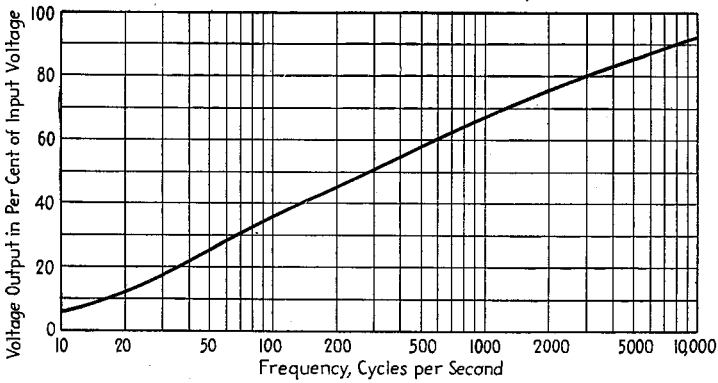
[2] A more elaborate form of curve-tracing device based on this same principle, except that it is made fully automatic by the use of electronic means to sweep the oscillator frequency through the audio range, is described by J. B. Sherman, An Audio-frequency-response-curve Tracer, *Proc. I.R.E.*, Vol. 26, p. 700, June, 1938.

[3] For example, see P. F. Jones, A Recording System for Transmission Measurements, *Bell Lab. Rec.*, Vol. 15, p. 289, April, 1938.

Another method of indicating amplitude distortion, termed the *intermodulation method,* consists in applying to the input of the system two sine waves of different frequencies, for example, 400 cycles and 1,500 cycles, and then observing the magnitude of the difference-frequency component appearing in the output.[1] This system is based on the fact that the same nonlinear distortion that causes the production of harmonics of a pure sine wave also causes intermodulation with resulting difference



(a) Circuit



(b) Characteristic of logarithmic network

FIG. 74.—Circuit arrangement for tracing out the amplification curve as a function of frequency.

and sum frequencies, the magnitudes of which are indicative of the amount of distortion present.[2] This method is particularly popular in sound-recording systems, since

[1] Equipment for such tests is described by J. K. Hilliard, Distortion Tests by the Intermodulation Method, *Proc. I.R.E.*, Vol. 29, p. 614, December, 1941.

[2] A modification of this procedure consists in applying simultaneously a low and a relatively high frequency and observing the extent to which the high-frequency output has its amplitude modulated by the low frequency. This observation is conveniently made by employing a band-pass filter to separate the high-frequency wave in the output, rectifying this, and observing the amount of low-frequency modulation voltage present in the rectifier output.

it can be used to indicate not only the distortion in an amplifier but also that produced in the recording and pickup processes, etc.

32. Square-wave Testing.—Square or rectangular waves are very useful for quickly determining many of the characteristics of amplifiers and circuits.

The usual procedure is to apply the square wave to the input terminals of the circuit or amplifier under investigation and to observe on a cathode-ray oscillograph the character of the output wave that results. The significance of various types of distortion that may occur in the output wave can be understood with the aid of the typical cases discussed below.

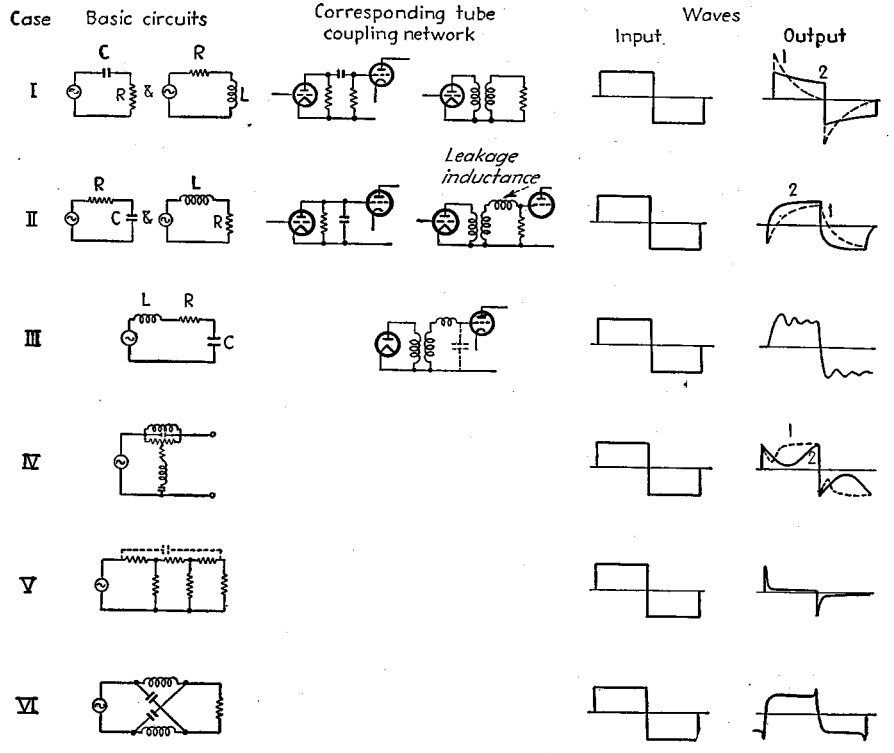


FIG. 75.—Effect of applying voltage from a square-wave generator to typical circuit arrangements.

Distortions of Square Wave Produced by Typical Networks.—The behavior of representative networks under the action of a square-wave input is shown in Fig. 75.[1] The networks of Case I correspond to the low-frequency coupling networks in resistance-coupled and transformer-coupled amplifiers. The output wave marked 1 is for a fundamental frequency of the square-wave generator corresponding to approximately 70 per cent response and 45° phase shift, while the curve marked 2 is for a fundamental frequency approximately 15 times as great. The significant characteristic of these output waves is the slope in the horizontal portion of the wave, which is due primarily to time-delay error (phase distortion) at low frequencies. This

[1] Most of these examples were supplied by W. R. Hewlett. Additional information is given by G. Swift, Amplifier Testing by Means of Square Waves, *Communications*, Vol. 19, p. 22, February, 1939.

slope provides a very sensitive indication, since a 10 per cent slope will be obtained when the phase error is 2° at the fundamental frequency of the square wave, corresponding to a reduction in response to the fundamental frequency of 0.9994.

The second example in Fig. 75 is for networks representing the high-frequency circuit of a resistance-coupled amplifier or a transformer-coupled power amplifier. Here the two curves shown correspond, respectively, to fundamental frequencies of the square wave that are roughly 0.1 and 0.4 of the frequency for 70 per cent response. The resulting output wave is characterized by a gradual rather than an abrupt rise in the vertical portions, with a rounding off of one of the corners. These effects are characteristic of high-frequency deficiencies.

Resonant circuits with less than critical damping act as shown in Case III when the resonant frequency is somewhat higher than the fundamental frequency of the square wave. The vertical part of the square wave has been rounded off somewhat, and a damped oscillation appears. The ratio of the resonant frequency to the square-wave fundamental is equal to the number of half cycles of oscillations that occur in one-half cycle of the square wave. The damping of the resonant circuit is indicated by the rate at which the oscillations die out, and can be calculated from the fact that if the damping is not too close to the critical value, the oscillations will drop to 37 per cent of their initial amplitude in Q/π cycles or to 50 per cent of their initial value in 0.22 Q cycles.

When the resonant circuit is overdamped, the behavior is much the same as in Case II or Case IV, according to the circumstances.

The circuit of Case IV is a dip pad, *i.e.*, a network, having a high attenuation at one frequency, with relatively little effect on either magnitude or phase of transmission at frequencies appreciably higher or lower. Output wave 1 shows the situation when the fundamental frequency of the square wave is much lower than the frequency of high attenuation, while output wave 2 is for the case when the two frequencies are the same. The latter is essentially a square wave, with a certain amount of fundamental frequency subtracted. In the former, the departure from the square wave is in the form of a dip, the duration of which indicates crudely the ratio between the fundamental frequency of the square wave and the frequency of high attenuation.

Example 5 is an attenuator with a high-frequency leak that results from a shunting capacity between input and output. The result is the transmission through the network of an unattenuated pulse at each instant of reversal of the square-wave voltage, with the normal attenuation maintained during the main part of the cycle.

The final example in Fig. 75 is an all-pass network to which a square wave of relatively low fundamental frequency is applied. Each reversal of the applied voltage is characterized by a sharp pulse, followed by a rounding off of the trailing corner of the square wave. This is a result of phase distortion.

Summary of Circuit Behavior with Square Waves.—The low-frequency deficiencies of a circuit can be determined by applying a square wave of low fundamental frequency. A sloping top to the output wave then indicates time-delay error at low frequencies (Case I), while a curved top (Case IV) indicates deficient or excess transmission at the fundamental frequency, according to whether the curve is concave toward or away from the axis.

The high-frequency characteristics of a circuit can be checked by employing a square wave having a fundamental frequency somewhat below the upper limit of the network. Then a gradual rise of the vertical parts, particularly with exponential rounding off of a corner, as in Case II, indicates high-frequency deficiencies, particularly phase distortion, although amplitude distortion may be involved. Oscillations as in Case III or a dip as in curve 1 of Case IV or a short peak as in Cases V and VI indicate amplitude or phase deficiencies, or both, at high frequencies. Oscillations occur when

the circuit is underdamped, while a dip (or peak or gradual rise) will result if resonant circuits have damping greater than the critical value. The approximate frequency at which the deficiency occurs in relation to the fundamental frequency of the square wave is indicated by the ratio of the time occupied by the pulse (or dip or rounding off) compared with the time represented by a full cycle of the square wave.

Square waves provide a means of quickly checking the characteristics of circuits that should have very good amplitude and phase characteristics. This method of testing is more sensitive to time-delay error than any other method that has been devised. It is accordingly widely used in testing television circuits where time-delay error is fully as important as the amplitude characteristics.

Comparison Tests.—A square-wave generator is particularly suitable for determining quickly whether two amplifiers (or circuits) that are supposed to have identical characteristics really are the same. This is because a square wave consists of a large number of frequency components and so provides a means of comparing amplitude and phase for many frequencies in a single test. The procedure is to apply a square wave simultaneously to the amplifier (or network) under test, and an amplifier (or network) that is the standard for comparison. The two outputs are then used to produce vertical and horizontal deflections in a cathode-ray oscilloscope. If the two systems are absolutely identical, a straight-line pattern will be obtained, but if the characteristics differ for any one of the frequency components contained in the square wave, something other than a straight line will result.[1]

33. Square-wave Generators.—Square waves can be generated by successive clipping and amplification of a sine wave or by clipping a wave produced by a multi-

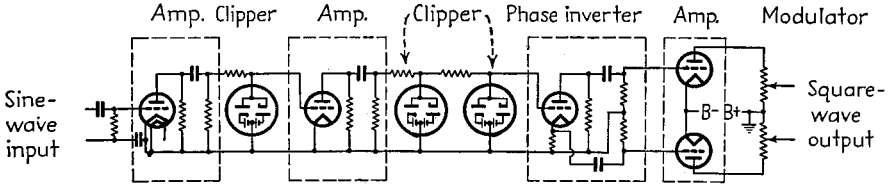


FIG. 76.—Schematic circuit diagram of square-wave generator employing clipping of a sine wave.

vibrator.[2] A square-wave generator of the former type is shown schematically in Fig. 76.[3] Here the sine wave from an oscillator is amplified, and then the positive and negative peaks limited by a biased double diode, as shown.[4] This results in an approximate square wave that is then amplified and subjected to a second clipping process to increase the steepness of the sides. The resulting square wave is then applied to a phase inverter, amplified, and delivered to the output terminals through a volume control.

A multivibrator type of square-wave generator is illustrated in Fig. 77.[5] Here

[1] This method of testing was suggested by W. R. Hewlett.

[2] A very good square wave can be obtained at low and moderate frequencies by clipping the output of a thyratron-inverter circuit. In such an arrangement, the life of the tubes is approximately a certain total number of cycles of operation. As a consequence, although a satisfactory life is obtained at frequencies of the order of 60 to 100 cycles, the tube life becomes impracticably short when the square-wave frequency is of the order of a few thousand cycles per second.

[3] L. B. Arguimbau, Type 769-A Square-wave Generator, *Gen. Rad. Exp.*, No. 7, p. 4, December, 1939.

[4] Another method of clipping consists in biasing a pentode nearly to cutoff and applying sufficient signal to drive the tube well beyond cutoff on the negative peaks. This clips the negative peaks very effectively, and if followed by another such tube, the other peak is limited, and one can then get the same effect as produced by the double diode.

[5] E. H. B. Bartelink, A Wide-band Square-wave Generator, *Trans. A.I.E.E.*, Vol. 60, p. 371, 1941. Another multivibrator type of square-wave generator is described by J. R. Cosby and C. W. Lampson, An Electronic Switch and Square-wave Oscillator, *Rev. Sci. Instruments*, Vol. 12, p. 187, April, 1941.

tubes T_1 and T_2 with their associated circuits represent a conventional multivibrator operated with the grid return brought to a positive bias that can be adjusted to control the frequency. The voltage across the cathode resistor R_c of such an arrangement has the wave shape shown, and can be made square by a single clipping operation that cuts off the positive side of the wave at an amplitude such as is indicated by line *cc*. This is accomplished by connecting the cathode of tube T_3 to the cathode of multivibrator tube T_2. This causes the cathode of T_3 to be driven to cutoff on the positive peaks. The multivibrator type of square-wave generator does not require an external oscillator to operate, although synchronization with an external oscillator can be obtained by injecting a voltage where shown in Fig. 77. The frequency can be continuously controlled over a ten to one range by varying the positive bias on the multivibrator grids; while large step changes in frequency, such as decade values, can be obtained by changing the capacities of the grid condensers C_1 and C_2 with the use of a gang switch. The ratio t_1/t_2 (*i.e.*, the relative pulse lengths) can be controlled by varying the ratio R_2/R_1 of grid-leak resistors.

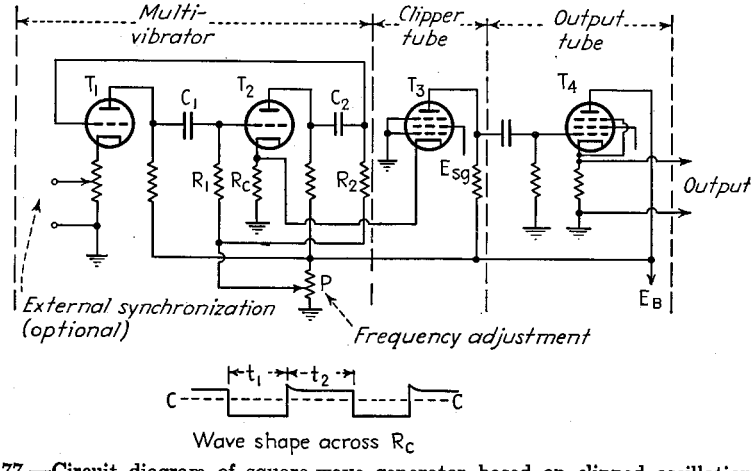


FIG. 77.—Circuit diagram of square-wave generator based on clipped oscillations of a multivibrator.

The circuits in a square-wave generator that handle the square wave after final clipping must be very carefully designed for uniform frequency characteristic and negligible time-delay error. In particular, the phase shift at low frequencies must not exceed a tenth of a degree, and the amplification must be uniform with negligible delay error up to a frequency about thirty times the highest fundamental frequency to be generated.

34. Audio-frequency Transformer Constants.—The most important characteristics of an audio-frequency transformer are the incremental primary inductance for the d-c saturation under which the transformer is expected to operate, the leakage inductance referred to the primary side of the transformer, the ratio of voltage transformation, and in the case of interstage coupling transformers the frequency at which the leakage inductance is in resonance with the distributed capacity of the transformer secondary. Other features of some interest are the losses of the transformer, including particularly the d-c resistances of the windings, and the equivalent eddy-current resistance expressed as a shunt resistance across the primary of the transformer. The hysteresis loss can ordinarily be neglected except possibly at quite low frequencies.

The incremental primary inductance can be measured at the appropriate alternating- and direct-current magnetizations as described in Par. 3. Such a measurement must be made at a frequency well below that at which the transformer is in parallel resonance with its distributed capacity, which normally means at a frequency not to exceed a few hundred cycles.

The leakage inductance reduced to a unity turn ratio can be readily obtained by measuring the inductance between the primary terminals when the secondary is short-circuited. The leakage inductance obtained in this way is independent of frequency, and also of saturation effects in the core since the leakage paths are primarily in air. The measurements accordingly may be made at any convenient frequency and without passing direct current through the core. In the case of interstage coupling transformers, it is not permissible to measure the leakage inductance from the secondary terminals by short-circuiting the primary unless this is done at quite low frequencies, since otherwise errors will be introduced as the result of the secondary distributed capacity.

The effective distributed capacity of the secondary can be obtained in any one of several indirect ways. One method is to determine the frequency $\omega_1/2\pi$ at which parallel resonance occurs across the transformer when the primary terminals are open-circuited. The equivalent secondary distributed capacity C_s, reduced to unity turn ratio with the primary side as reference, is then

$$C_s = \frac{1}{\omega_1^2 L_p} \tag{38}$$

where L_p is the primary inductance under the conditions of measurement. A second method is to determine the frequency at which parallel resonance exists across the secondary terminals when the primary is short-circuited. This frequency corresponds to the condition where C_s and the leakage inductance L_s are in parallel resonance so that

$$C_s = \frac{1}{\omega_2^2 L_s} \tag{39}$$

where $\omega_2/2\pi$ is the resonant frequency, and C_s and L_s are referred to the primary side. A third method of measuring the effective secondary capacity is to determine experimentally the ratio of secondary to primary voltage as a function of frequency for frequencies approaching that at which the leakage reactance is in resonance with the distributed capacity. This ratio will increase with frequency and reach a maximum at resonance. At some frequency $\omega/2\pi$ below resonance, C_s can be deduced from a knowledge of L_s and the voltage ratio according to the relation

$$\frac{\text{Voltage ratio of transformer}}{\text{Turn ratio of transformer}} = -\frac{(1/\omega C_s)}{\omega L_s - (1/\omega C_s)} \tag{40}$$

In measuring the secondary distributed capacity it must be kept in mind that the capacity that is effective in the actual operation of the transformer is the sum of the distributed secondary capacity of the transformer and the input capacity of the tube to which the secondary delivers its voltage. Consequently, C_s is preferably obtained with the output tube actually present and operating with its normal load impedance. Additional measuring equipment, such as vacuum-tube voltmeters, across the secondary, must be arranged to introduce the minimum possible additional shunting capacity, or their shunting capacity should be determined separately and a suitable correction made.

The turn ratio of a transformer can be conveniently measured by applying a known voltage across the primary terminals and measuring the voltage developed across the secondary, using a vacuum-tube voltmeter. The frequency at which such a determination is made must be chosen with reasonable care. If the frequency is very low, the step-up ratio as observed will be less than the actual turn ratio as a result of voltage lost in the d-c resistance of the primary. The error resulting from this source is as follows:

$$\frac{\text{Actual turn ratio}}{\text{Observed voltage ratio}} = \sqrt{1 + \left(\frac{R_p}{\omega L_p}\right)^2} \tag{41}$$

where R_p = d-c resistance of the primary winding.

ωL_p = reactance of the primary inductance at the frequency of measurement.

On the other hand, if the frequency of measurement is too high, the voltage ratio of the transformer will differ from the step-up ratio as a result of partial resonance between the leakage inductance and secondary distributed capacity as given by Eq. (40).

The eddy-current resistance R_e that can be considered as shunting the primary of the transformer in Fig. 12, Sec. 5, is the parallel impedance existing across the primary terminals of the transformer when the transformer is considered as a parallel resonant circuit, in which the inductance is the incremental primary inductance of the transformer and the tuning capacity is the equivalent unity-ratio capacity represented by the secondary distributed capacity. Inasmuch as this resistance does not vary greatly with frequency, it may be conveniently obtained by making a bridge measurement of the primary terminal impedance at some convenient medium low frequency, with the transformer being tuned to resonance at the frequency of measurement by means of a small variable capacity connected across the secondary terminals, that is adjusted to give the transformer a resistance impedance.

MEASUREMENTS ON RADIO RECEIVERS

35. Receiver Characteristics and Their Determination.[1]—Radio receivers are tested by employing an artificial signal from a standard signal generator to provide a voltage corresponding to that induced in the receiving antenna. This voltage is ordinarily applied to the receiver through a network, termed a dummy antenna, having characteristics such that the receiver views substantially the same impedance as it would in normal operation with an actual antenna. The receiver output is then observed by replacing the loud-speaker or telephone receivers by a suitable resistance load, with which is associated a power indicator.

The dummy antenna recommended for use in testing broadcast receivers is given in Fig. 78.[2] The impedance of this network in the frequency range 540 to 1,600 kc approximates that of the typical open-wire antenna resonant at about 2,500 kc, and having a capacity of the order of 200 $\mu\mu$f. At higher frequencies the network approaches a constant impedance of 400 ohms, and so resembles a nonresonant transmission line of corresponding impedance.

[1] Many of the tests and test procedures commonly used with radio receivers, particularly broadcast receivers, have been standardized to ensure uniformity. These standards are described in "Standards on Radio Receivers," Institute of Radio Engineers, New York, 1938.

The standard test procedures used in England are described in the paper R. M. A. Specification for Testing and Expressing the Overall Performance of Radio Receivers, *Jour. I.E.E.*, Vol. 81, p. 104, 1937 (*Wireless Proc.*, Vol. 12, p. 179, September, 1937).

[2] Where tests are to be made only in the standard broadcast frequency range, an alternative network consisting of a capacity of 200 $\mu\mu$f, resistance of 25 ohms, and an inductance of 20 μh, all connected in series, is commonly used. Such a dummy antenna has practically the same impedance as the recommended network in this frequency range.

With loop antennas, the signal-generator voltage can be introduced into the receiver loop by use of a known mutual inductance, as in Fig. 79*a*, or can be inserted directly in series with the loop circuit, as in Fig. 79*b*.[1] The former arrangement is preferable, because it requires no correction for distributed capacity of the loop. In this mutual-inductance method, if the reactance of the primary inductance X_p is at least three times the internal impedance R_p of the signal generator, then the strength

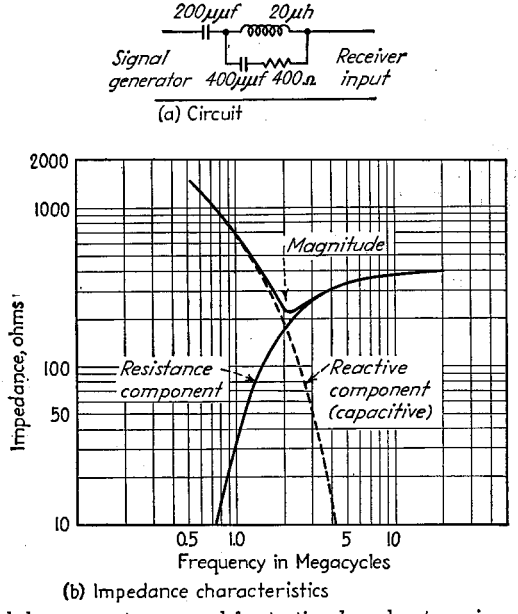


FIG. 78.—Standard dummy antenna used for testing broadcast receivers, together with its impedance characteristic.

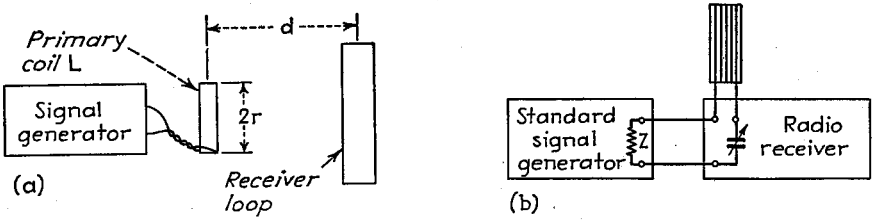


FIG. 79.—Methods of introducing signal-generator voltage into a loop antenna.

of the radio field E required to induce in the loop the same voltage as does the signal generator is[2]

$$\epsilon = \frac{18.85 N_p r_p^2}{d^3} \frac{E}{X_p} \tag{42}$$

where ϵ = strength of radio field, microvolts per meter.

N_p = number of turns in primary coil L_p.

[1] Further details on receiver measurements when loop antennas are used are given by W. O. Swinyard, Measurement of Loop Antenna Receivers, *Proc. I.R.E.*, Vol. 29, p. 382, July, 1941.

[2] When this relation is not satisfied, then one should substitute $\sqrt{X_p^2 + R_p^2}$ in place on X_p in Eq. (42).

r_p = radius of primary coil, cm.
d = distance, meters, between center of primary coil L and center of loop antenna.
E = signal-generator voltage, millivolts.
X_p = reactance of primary coil L.

The distance d should be at least twice the largest dimension of the loop or primary coil, but should be much less than a wave length.

When the signal-generator voltage is inserted directly in series with the loop, as in Fig. 79*b*, then the intensity of the radio field that would produce the same effect in the radio receiver as does the signal generator, is

$$\epsilon = \frac{47{,}750E}{N_2Af} \tag{43}$$

where ϵ and E have the same meaning as in Eq. (42), N_2 is the number of turns in the loop antenna, f the frequency in kilocycles, and A the cross-sectional area of the loop in square meters.

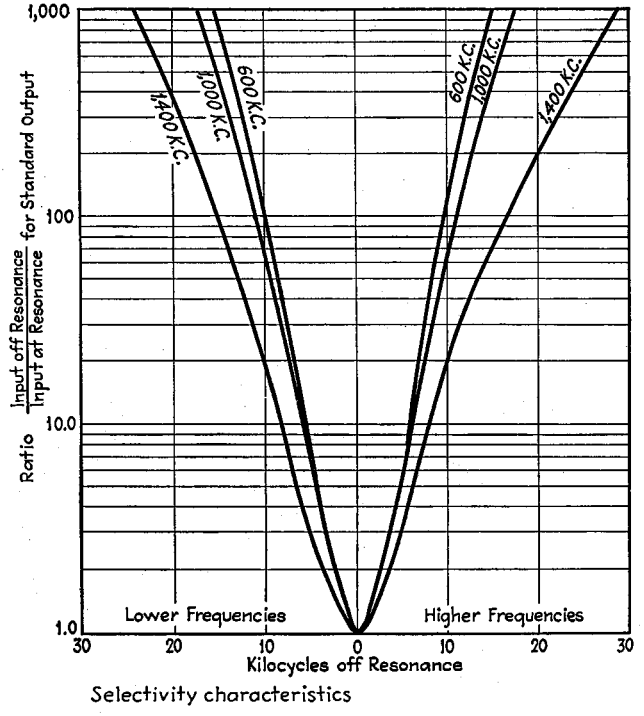
Sensitivity.—The sensitivity of a radio receiver is defined in terms of the voltage that must be applied by a signal generator to the receiver input to produce a specified output. In the case of broadcast receivers, the conditions of the sensitivity test have been standardized on the basis of a signal modulated 30 per cent at 400 cycles, with the standard output to be delivered by the receiver into a dummy load taken as 0.5 watts.[1] The sensitivity is normally expressed either in microvolts, or in decibels below 1 volt. Unless otherwise stated, the sensitivity test is taken with the receiver controls adjusted to give maximum sensitivity. A typical sensitivity curve of a broadcast receiver is shown in Fig. 80.

Selectivity.—The selectivity of a radio receiver is that characteristic which determines the extent to which the receiver is capable of distinguishing between the desired signal and disturbances of other frequencies. Selectivity is expressed in the form of a curve that gives the signal strength required to produce a given receiver output as a function of the cycles off resonance of the signal, with the response at resonance taken as the reference (see Fig. 80).

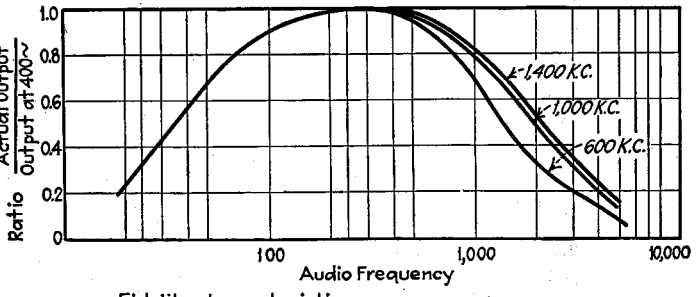
The selectivity curve of a radio receiver is normally obtained by disabling the automatic-volume-control system (or replacing the A.V.C. bias by a fixed bias), setting the signal generator to the desired frequency, tuning the receiver to this frequency, and adjusting the signal to give a convenient output when modulated 30 per cent at 400 cycles. The carrier frequency of the signal generator output is then varied by progressively increasing amounts from the frequency to which the receiver is tuned, and the signal-generator voltage increased as necessary to maintain constant receiver output. Selectivity curves should be carried to 100 kc off resonance or to 80 db above the response of resonance, whichever is encountered first.

Selectivity curves with the automatic-volume-control system operative can be obtained by employing two signal generators, one to represent the desired signal to which the signal is tuned, and the other to represent the interfering signal. In such an arrangement, the "desired" signal is adjusted to the amplitude desired for purposes of the test, and the modulation of this signal is then removed. The "interfering" signal generator is then turned on and modulated 30 per cent at 400 cycles. The selectivity is then expressed as a curve showing the amplitude of the interfering signal required to produce a standard 400-cycle output, as a function of the difference between the frequency of the interfering signal and the frequency of the desired signal. The selectivity curve obtained from a two-signal-generator test of this type

[1] This assumes that the receiver is capable of delivering an output of at least one watt. If the maximum receiver output is between 0.1 and 1 watt, the standard output is 50 milliwatts.



Selectivity characteristics



Fidelity characteristics

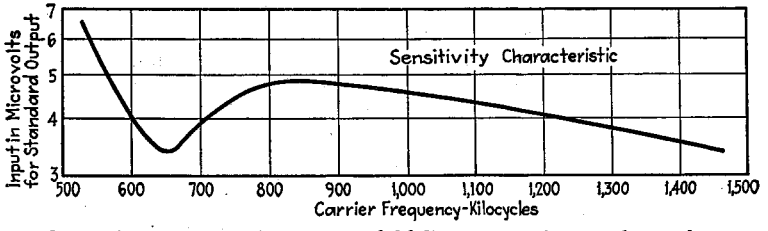


FIG. 80.—Typical sensitivity, selectivity, and fidelity curves of a superheterodyne receiver.

takes into account not only the selectivity of the resonant circuits of the receiver, but also any cross-modulation effects that occur in the receiver. These cross-modulation effects commonly dominate the situation when the desired signal is large and the frequency difference between the two signals is not too small.

Fidelity.—The fidelity of a receiver shows the manner in which the electrical output at a dummy load depends upon the modulation frequency. Fidelity is measured by setting the signal generator to a desired carrier frequency, tuning the receiver to this signal, adjusting the signal generator until a convenient output is obtained, and then observing the variation in receiver output as the modulation frequency of the signal generator is varied, while keeping the degree of modulation constant at 30 per cent. The results of a fidelity test are expressed in the form of a curve, as shown in Fig. 80, with the output at 400 cycles taken as the reference value. In making fidelity tests care must be taken to avoid applying so much input to the receiver as to overload the output. Also, in the event that the noise and hum voltages in the receiver output are appreciable, it is necessary either that the signal be strong enough to override these interfering effects, or that their power be subtracted to give the true output.

Fidelity tests as outlined above using a dummy load do not include the characteristics of the loud-speaker, or the acoustics of the space in which the sound is reproduced. These factors can be taken into account by over-all tests in which the relative sound output of the receiver is measured.[1]

Miscellaneous Receiver Characteristics.—In addition to selectivity, sensitivity, and fidelity, a number of other characteristics are often of interest. These include the maximum undistorted power that the receiver can develop, the hum and noise level, cross-talk, spurious responses, automatic-volume-control characteristics, etc.

The maximum undistorted output that a receiver can develop is arbitrarily defined as the maximum power that can be delivered to the load with an rms distortion not to exceed 10 per cent. Undistorted power is commonly measured with a signal modulated 80 per cent at 400 cycles, although other conditions of degree and frequency of modulation may be specified.

Noise and hum may be objectively measured in several ways.[2] One procedure consists in applying an unmodulated carrier of appropriate amplitude to a receiver from a signal generator and observing the hum and noise output on a square-law indicating device. The signal is then modulated and the degree of modulation adjusted until the square of the rms output, as indicated on the instrument, has been doubled. The hum and noise present under the given conditions can then be expressed in terms of decibels below 100 per cent modulation.

Spurious responses are investigated by the use of one or two signal generators, as the case requires, and are evaluated in terms of the amplitude of the undesired signal required to give a specified output, compared with the amplitude of the desired signal required to give the same output. Thus if an image voltage must be 1,000 times as strong as the desired signal, to produce a given output, then the image discrimination is 60 db.

Operation of Two Signal Generators in Parallel.—Certain receiver tests necessitate the use of two signal generators acting simultaneously on the receiver input. When both generators have one terminal grounded, as is usually the case, this presents a special problem.

[1] Such tests are described by Stuart Ballantine, High Quality Radio Broadcast Transmission and Reception—II, *Proc. I.R.E.*, Vol. 23, p. 618, June, 1935; H. A. Wheeler and V. E. Whitman, Acoustic Testing of High Fidelity Receivers, *Proc. I.R.E.*, Vol. 23, p. 610, June, 1935.

[2] For further information on this subject, see "Standards on Radio Receivers," *loc. cit.;* F. B. Llewellyn, A Rapid Method for Estimating the Signal-to-noise Ratio of a High Gain Receiver, *Proc. I.R.E.*, Vol. 19, p. 416, March, 1931; S. Ballantine, Fluctuation Noise in Radio Receivers, *Proc. I.R.E.*, Vol. 18, p. 1377, August, 1930; H. O. Peterson, A Method of Measuring Noise Levels on Short-wave Radiotelegraph Circuits, *Proc. I.R.E.*, Vol. 23, p. 128, February, 1935.

When the two signal generators have an output impedance that is independent of the attenuator setting, they can be connected directly in parallel. The equivalent output impedance of the combination is then equal to the output impedances of the two generators taken in parallel, while the output voltage of an individual signal generator is less than the output voltage for normal operation by the factor $R_2/(R_1 + R_2)$,

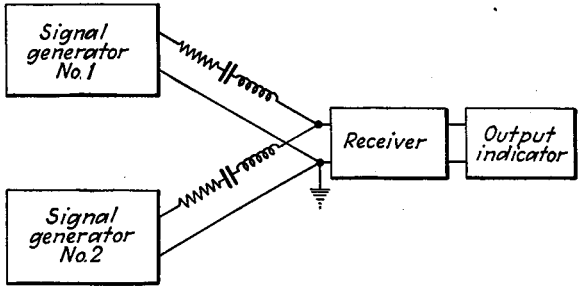


FIG. 81.—Method of superimposing the output voltages of two signal generators upon the input of a radio receiver.

where R_1 is the output resistance of the signal generator in question and R_2 is the output resistance of the other signal generator that is connected in parallel.

An alternative arrangement that must be used when the output impedance of the signal generator is variable, and that is also often more convenient in other cases, is illustrated in Fig. 81 and involves the use of separate artificial antennas for each signal generator, with the outputs of these two antennas connected in parallel. Each individual artificial antenna is arranged to have twice the impedance of the antenna with which the receiver is supposed to operate; *i.e.*, the resistances and inductances are twice as large as in the usual artificial antenna, and the capacity is half as great. Since the two antennas are in parallel as far as the receiver input is concerned, the result is equivalent to the desired artificial antenna. The indicated output voltages

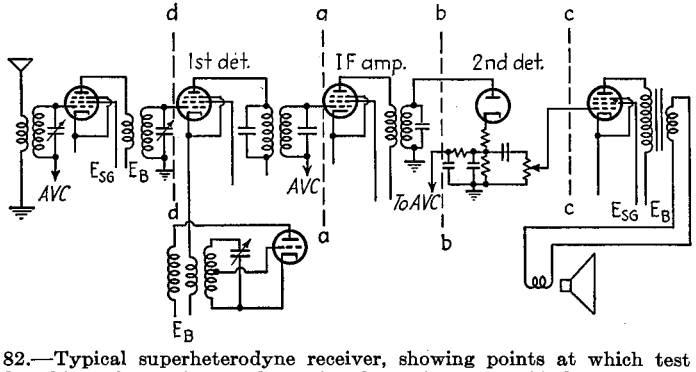


FIG. 82.—Typical superheterodyne receiver, showing points at which test voltages are introduced into the receiver to determine the performance of individual sections of the receiver.

of the signal generator must be divided by two, however, since it takes twice as much voltage in the arrangement of Fig. 81 to produce a given input at the receiver terminals as when a single generator is used with the usual dummy antenna.

Measurement of Performance of Individual Parts of a Receiver.—The performance of the individual portions of a radio receiver, such as the section *ab* in Fig. 82, can be

obtained by determining the signal-generator voltage that must be applied at points *a* and *b*, respectively, to produce the standard receiver output. The amplification in the section under question is then the ratio of these two voltages. It will be noted that if there is a frequency transformation in the section involved, as, for example, in section *da*, the carrier frequency of the signal generator must be changed accordingly, and when one gets into the audio-frequency section, it is necessary to use an audio-frequency signal generator. This procedure by which the performance of a section of a receiver is obtained by difference, instead of by direct measurement of the input and output voltages of the section, must be followed in order that the regeneration, circuit capacities, etc., that would affect the behavior of the section under question be not affected by the measuring procedure. The principal precaution necessary in making tests in this way is not to alter the amplitude of the beating oscillator.

36. Signal Generators.—A signal generator is a device for producing accurately known radio-frequency voltages that can be continuously adjusted from about 1 microvolt in amplitude to approximately 0.5 volts. Signal generators consist of a very completely shielded oscillator that can be modulated, together with attenuator and metering means for producing accurately known small voltages.

Shielding of Signal Generators.[1]—The degree of shielding required in signal generators is the greatest needed in any radio work. The principal considerations required to achieve the desired results are (1) shields having a thickness considerably greater than skin depth of current penetration, constructed with tight-fitting joints making good electrical contact; (2) arrangement of circuits inside the shield so that the shield carries a minimum of current; (3) the return of all ground wires to a common point ground; (4) the use of filters in outgoing leads so arranged that the by-pass condensers of all leads are returned to a common point, and means such as shielding to prevent voltages or currents from being induced in the outgoing leads or chokes on the output side of the by-pass condensers; (5) the use of spaced concentric shields that are connected together electrically at only one point. An example of a suitably arranged shielding system is shown in Fig. 83.

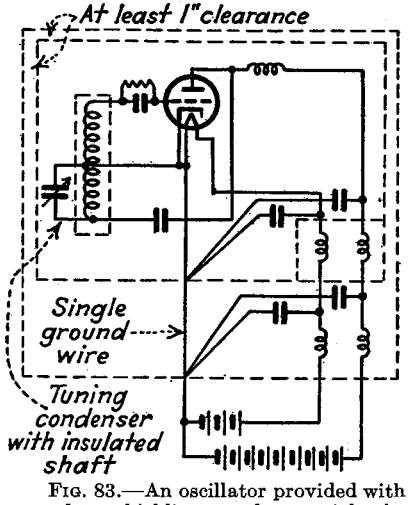


FIG. 83.—An oscillator provided with complete shielding, such as might be used in a signal generator.

Adequate wall thickness of the shield will prevent direct leakage through the shield material itself. Well-made joints providing good electrical conductivity are essential in order that the eddy currents in the shield will not be disturbed. Wherever possible, joints should be soldered, but when this is not feasible, the mechanical fit should be of the very best, since experience shows that a joint that is literally watertight is required if leakage of field is to be avoided. Currents in the shield cause different parts of the shield interior to be at different potentials. This greatly increases the possibility of leakage through outgoing leads, mechanical joints, etc. Currents in the shield can be minimized by using a shielded coil for the oscillator and by insulating everything from the shield except for ground wires all brought to a

[1] See also J. R. Bird, Some Considerations on the Design of Radio-frequency Signal Generators, *Proc. I.R.E.*, Vol. 19, p. 438, March, 1931.

single point. The by-pass condensers in the filters of outgoing leads must all be connected by individual wires to a common point on the shield, preferably the same point at which the internal ground connection within the shield is made. In this way, there will be no current in the shield to produce difference in potential between different parts of the shield and hence a potential difference between different outgoing leads. The use of concentric shields greatly increases the degree of shielding obtainable, but the full possibilities of the concentric construction will not be realized unless there is only a single electrical connection between the inner and outer shields, with by-pass condensers arranged as shown in Fig. 83. It will be noted that insulated couplings must be placed in the shafts of all tuning condensers, etc., located in the inner compartment of a doubly shielded system. The leads for introducing filament and plate voltages do not constitute additional electrical connections between the inner and outer shields, because the filter chokes in these leads make them effectively an open circuit.

External modulating voltage can be introduced into a signal generator by designing the plate filter as a low-pass filter that will transmit the audio modulating frequencies while attenuating the radio carrier frequencies. In some instances, filter sections of the *m*-derived type are used in place of the simple capacity-inductance sections shown in Fig. 83.

37. Signal-generator Attenuators. *Resistance Attenuators.*—The most common method used to develop a known voltage in a signal generator is to employ a resistance attenuator that is supplied by a known current measured by a thermocouple, or by a known voltage measured on a vacuum-tube voltmeter. In such arrangements, the leads from the thermocouple or vacuum tube to the indicating galvanometer are provided with filters so that the indicator can be on the outside panel. A ladder type of resistance attenuator is normally employed to give decade steps, with a slide-wire arrangement used to provide continuous adjustment between decades. The resistance units must be of a low-reactance type, such as described in Par. 6, Sec. 2, and must employ wire small enough so that the skin effect is negligible at the highest frequencies involved. The attenuator is normally constructed to have an impedance level small compared with the impedance of the usual artificial antenna. As a result of the low impedance level, electrostatic effects are not troublesome if reasonable care is taken, but because of the low impedance, every possible precaution must be taken to minimize the inductance in the connecting wires of the attenuator. In particular, these leads should be reduced to the minimum possible length, and, furthermore, the grounded side of the shunt resistors should be returned directly to points closely spaced along a common wire. A good mechanical arrangement of attenuator and switch that carries out these requirements and also provides shielding between parts of the attenuator at widely different power levels is shown at Fig. 84.[1]

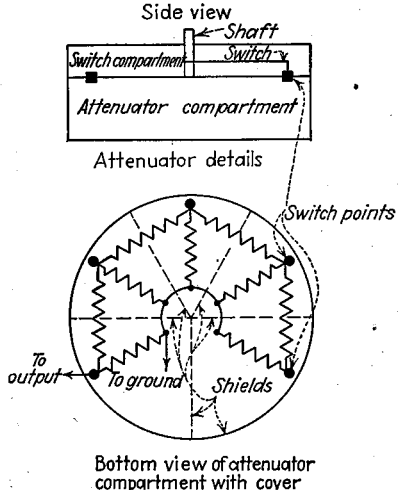


FIG. 84.—Design of attenuator and switch providing short leads, shielding, and compactness.

[1] See also Axel G. Jensen, Potentiometer Arrangement for Measuring Microvoltages at Radio Frequencies, *Phy. Rev.*, Vol. 26, July, 1925.

The ground on the attenuator system should be returned directly to the output ground terminal, and there should be only one ground in the attenuator system.

At radio frequencies above about 20 mc, there is difficulty in transmitting the attenuator output voltage to the input of the receiver without introducing resonance effects. This situation can be handled by employing a concentric line that is terminated with its characteristic impedance at the output (receiver) end.[1]

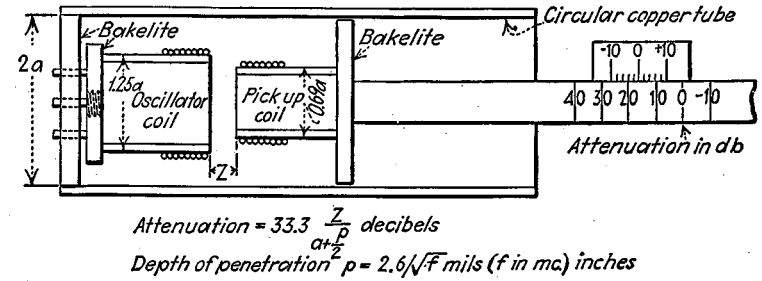


FIG. 85.—Mutual-inductance attenuator involving coaxial coils in a circular tube.

Mutual-inductance Attenuators.[2]—In mutual-inductance attenuators, the output voltage is controlled by varying the mutual inductance between two coils located in a tube. Such attenuators can employ coaxial coils in a circular tube, as in Fig. 85, or coplanar coils, preferably in a square tube, as in Fig. 86. When the spacing between coils in such arrangements is large, the output voltage measured in decibels is a linear function of the spacing Z between coils. In the case of coaxial coils in a circular tube, the attenuation in db is 33.3 times the separation between coils measured in radii of the inside dimension. With a coplanar coil in a square tube,[3] the rate at

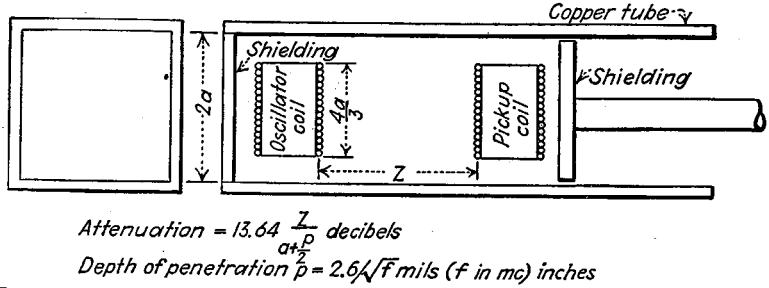


FIG. 86.—Mutual-inductance attenuator involving coplanar coils in a square tube.

which the output varies with large spacings is 13.64 db per radius of the inscribed circle of the square tube. This linear relationship between attenuation and spacing fails to hold if the coils are too close together. The proportions shown in Figs. 85 and 86, however, permit the mutual inductance to be made large enough to give serious reaction between the pickup coil and exciting coil before there is appreciable departure from the linear law.

[1] C. J. Franks, 20–100 Mc.Signal Generator, *Electronics*, Vol. 9, p. 16, August, 1936.

[2] D. E. Harnett and N. P. Case, The Design and Testing of Multirange Receivers, *Proc. I.R.E.*, Vol. 23, p. 578, June, 1935.

[3] When the coplanar coils are used in a round tube, the rate of attenuation is 16.0 db per radius. The radius used in making calculations of attenuation is the inside radius of the tube plus half the distance representing the skin depth of current penetration (see Par. 4, Sec. 2). The latter is very small and can ordinarily be ignored.

In practical mutual-inductance attenuators, it is possible to mark off a relative db scale by calculation. It is then necessary to evaluate the voltage at only one point, which is usually done for a large output, such as one volt, that can be observed on a vacuum-tube voltmeter. The relative attenuation is independent of frequency, although the reference level will usually vary with frequency because the voltage across the exciting coil will in most cases not be constant.

The errors in mutual-inductance attenuators are those arising from capacitive coupling and from mutual-inductance coupling of undesired types. Capacitive coupling can be minimized by proper mechanical design and construction. The coplanar type of attenuator is troubled least by these errors, since with this particular form of coupling, the field that produces the desired output attenuates more slowly with increased spacing than any other form of mutual-inductance or capacitive coupling. Thus, if these spurious couplings are much less than the desired coupling at close spacings, their relative importance is still less with large attenuations. This is not true in the case of coaxial coupling, however, with the result that spurious capacitive couplings, or couplings due to mutual inductance of the coplanar type arising from imperfections in axial symmetry or between leads to the coupling coils, tend to control the output when the attenuation is large.

The output impedance of an attenuator of the mutual-inductance type is the impedance of the pickup coil, and so is inductive. For proper operation, this coil should have such low distributed capacity that its natural resonant frequency is many times greater than the highest frequency to be used.

Mutual-capacity Attenuators.[1]—In mutual-capacity attenuators, the output is controlled by varying the mutual capacity existing between two condenser plates. Typical arrangements are shown in Fig. 87, in which the plates are located in a circular

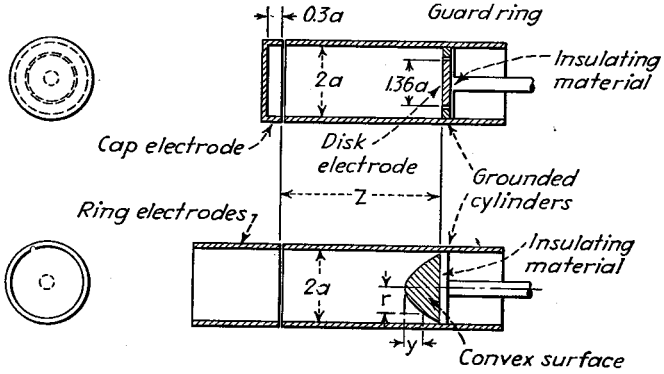


FIG. 87.—Mutual-capacity attenuator. The best shape of the convex surface is that represented by the equation $\epsilon^{-2.4y/a} = J_0(2.4r/a)$.

copper tube. When the separation between plates is large in such an arrangement, the rate of attenuation is 20.9 db per radius, *i.e.*, $(20.9Z)/a$, where Z and a are the separation and radius, as in Fig. 87[2]. With the electrode proportions shown in Fig. 87, this linear relationship between spacing and attenuation is maintained for spacings almost to the point where the electrodes touch.

Capacity attenuators do not necessarily have to be mounted in copper tubes. Another arrangement is shown in Fig. 88*a*, in which the varying mutual capacity is

[1] See Harnett and Case, *loc. cit.*

[2] In the case of a square tube, the rate of attenuation is 19.3 db per radius of the inscribed circle. It will be noted here that the radius to be used is the actual internal radius of the tube, with no allowance made for the skin depth as required in the case of the mutual-inductance attenuators.

the direct capacity between the fixed electrode *a* and the rotating semicircular plate *b*.[1] By use of the construction shown in Fig. 88, it is possible with the adjustable grounded plates C_1, C_2, etc., to make the impedance that the attenuator offers on its input side substantially independent of setting. In this way, the attenuator can be excited directly from a self-oscillating circuit without having the frequency vary excessively with attenuation changes.

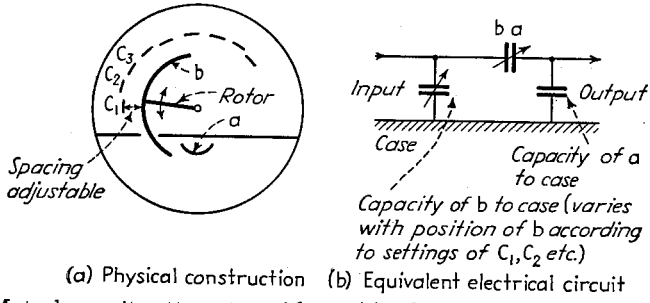


FIG. 88.—Mutual-capacity attenuator, with provision for maintaining the input impedance of the attenuator substantially constant.

The equivalent circuit of a capacity attenuator is shown in Fig. 89. Here E_1 represents the voltage applied to the attenuator input, C_1 is the capacity between input electrode and ground, C_2 the capacity between output electrode and ground, and C_m the varying mutual capacity that gives the attenuation. Under practical conditions, it is desirable to make C_2 very much larger than C_m, which means either that C_2 is augmented by a fixed condenser or that the construction is such that the required capacity is directly built into the output electrode. Under conditions where $C_2 >> C_m$, the output impedance of the attenuator is equal to the reactance of capacity C_2 plus any resistance and inductance that may be introduced by the output lead. If a concentric line is used to transmit the attenuator output to the point of use, the capacity C_2 should be large enough so that its reactance at the lowest frequency of use is not more than one-third of the characteristic impedance of the line.[2]

In mutual capacity attenuators of the type shown in Fig. 86, it is customary to engrave an arbitrary scale on the slider, with the spacing between graduations determined by calculation from the known law of attenuation. The voltage E_1 applied to the input electrode is then adjusted by means of a vacuum-tube voltmeter so that at some convenient point on the scale, preferably at 0 db, the output potential has a convenient value, such as 1 volt. In the case of capacity attenuators of the type shown in Fig. 88, the reference point is determined in the same way by applying a known voltage to the input electrode, but the law of scale variation must be determined experimentally.

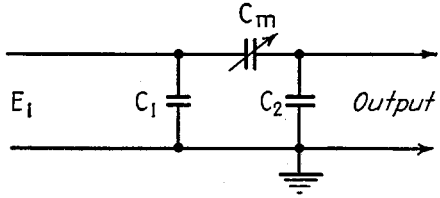


FIG. 89.—Equivalent circuit of capacity attenuators.

[1] See A Signal Generator for the Ultra-high Frequencies, *Gen. Rad. Exp.*, November, 1939.

[2] It will be noted that the voltage on the output electrode for the case $C_2 >> C_m$ is

$$\text{Output voltage} = E_1 \frac{C_m}{C_2} \qquad (44)$$

The voltage obtained with any given spacing will accordingly depend upon the value of C_2, while the variation in output with spacing will vary linearly with C_m.

Capacity attenuators are particularly attractive for use at frequencies so high that ordinary resistance attenuators are not accurate.

Miscellaneous Types of Attenuators.—Voltages down to 50 to 100 μv can be obtained with accuracy by passing a known current measured with a vacuum thermocouple through a small known noninductive resistance, as shown in Fig. 90*a*. Another method of producing moderately small known voltages consists in passing the known current through the calculated inductance formed by a short section of concentric line short-circuited at the receiving end (Fig. 90*b*). A modification of this arrangement consists in making use of a coupling loop mounted in a concentric transmission line, as shown in Fig. 90*c*. In such an arrangement, the voltage induced in the coupling loop can be calculated from the dimensions involved and from the current.

Small voltages can be obtained without the necessity of shielding by applying a carrier wave of half the desired frequency to the input of a square-law detector. When this is done, the detector output will contain, among other things, a second harmonic of the applied frequency, and also a rectified d-c current. It can be readily shown

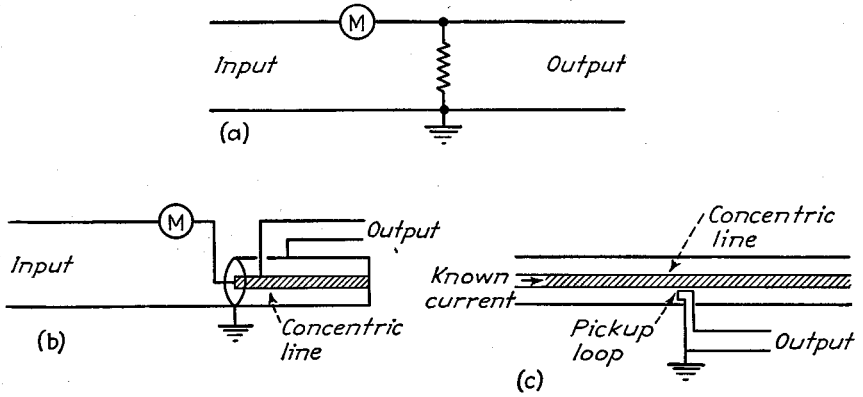


FIG. 90.—Miscellaneous type of attenuators.

by the series analysis of Par. 24, Sec. 5 that the crest amplitude of the second-harmonic current is exactly equal to the rectified d-c current, provided that the circuit impedance to these currents is small compared with the plate resistance of the tube at the operating point. Accordingly, measurement of the d-c current by a microammeter gives the amplitude of the second-harmonic current with very high precision. Known radio-frequency voltages that are quite small can accordingly be produced by passing the second-harmonic current through a low resistance, such as one ohm.

A practical circuit arrangement for utilizing the second-harmonic principle in a signal generator is shown in Fig. 91. This employs a balanced detector to reduce the amount of input frequency current flowing through the output and to provide stabilization with respect to change in electrode voltages. The normal d-c plate current is balanced out of the microammeter by adjusting the resistance R_1, which must be several hundred times the resistance of the microammeter if the meter is to indicate the rectified plate current without correction for the shunt formed by R_1. The second-harmonic type of signal generator has the very important advantages of requiring practically no shielding and of measuring the output voltage directly rather than depending upon an attenuator. At the same time, it has the disadvantage that spurious radio frequencies are present in the output, that the amount of manipulation required is greater than is the case with ordinary signal generators, and that

distortionless modulation of the output voltage is not possible, since when a wave modulated to the degree m is applied to the square-law detector, the modulation of the second-harmonic carrier output is $4/(2 + m^2)$ times the degree of modulation m of the input, and there is a second-harmonic component of the modulation frequency that is $m/4$ of the desired modulation. Another inconvenience is that the presence of modulation increases the amplitude of both the d-c output current and the second-harmonic carrier, even though it does not effect the equality of the two.

Accuracy Checks.—The accuracy of the attenuator in a signal generator can be checked in several ways. One method is to compare with a known voltage developed from a carefully made second-harmonic type of signal generator.

An alternative procedure is to compare the output voltages developed with large attenuation with output voltages obtained with small attenuation. This requires the aid of a sensitive radio receiver in which the final detector has been calibrated to act as a vacuum-tube voltmeter to measure accurately carrier amplitudes at the

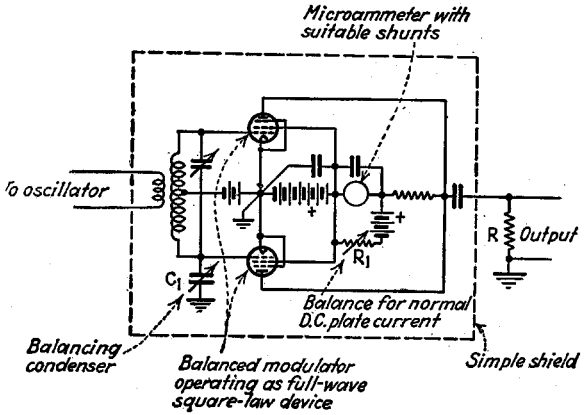


FIG. 91.—Practical form of second-harmonic type of signal generator.

detector grid over a voltage range of at least 10 to 1. The receiver must also give an output voltage that is proportional to input voltage for all signal amplitudes and sensitivity adjustments used in the measurements. The procedure is then as follows: the signal generator is coupled to the receiver through an artificial antenna and its output voltage set at some large value, say 10 millivolts, where there is little likelihood of the attenuator being in error. The receiver gain is then adjusted until the carrier at the detector is as large as the receiver can handle, say 10 volts. The signal-generator output is now reduced to 1 mv and the voltage at the detector read again. The ratio of the two detector readings gives the ratio of signal-generator outputs under the two conditions, and will be exactly 10 if the receiver is linear and the signal generator is not in error. The receiver gain is next increased so that the maximum allowable detector voltage appears in the detector with one mv input, after which the signal generator output is set to 0.1 mv and the procedure repeated. In this way, it is possible to check the attenuator step by step and to detect the existence and magnitude of errors that may be present.

38. Alignment of Radio Receivers.—The various circuits of a radio receiver can be aligned with the aid of a test oscillator capable of generating a modulated wave at the intermediate frequency and at signal frequency, together with some form of indicator to give receiver output: Such test oscillators are normally shielded and provided with some form of output control. Some of the very best are in effect

simple signal generators, and may even have the output control calibrated to give the approximate microvolt output.

The alignment of the intermediate-frequency amplifier system of a receiver is carried out by working step by step backward through the intermediate-frequency circuits from the second detector to the first detector and by always applying the test oscillator to the grid immediately preceding the circuit under adjustment and adjusting the trimmers on this circuit for maximum output. In carrying out this process, it is, of course, necessary to reduce the output of the test oscillator each time this output is applied to the grid of a tube at lower power level. The next step is to align the radio-frequency and oscillator circuits. This is accomplished by setting both receiver dial and test oscillator at a frequency somewhere near the high-frequency end of the band, connecting the test oscillator to the antenna input, and adjusting so that a moderate receiver output is obtained. The shunt-trimming condensers on the radio-frequency stages are then adjusted until the output is maximum, after which the shunt trimmer on the oscillator condenser is varied to give maximum receiver output.[1] The receiver dial and test oscillator are then set for a frequency at the low-frequency end of the band, and the series padding condenser of the oscillator is adjusted for maximum response. If the receiver covers only the regular broadcast band, it is also permissible to improve the low-frequency alignment of the radio-frequency stages by bending the end plates of the condensers. Finally, receiver and test oscillator are returned to the original high frequency, and the adjustment there is checked.

An alternative procedure that is sometimes employed is "channel analysis." The basic idea here involves tracing the signal through the receiver. The radio-frequency and intermediate-frequency systems are checked by an arrangement of tuned amplifiers feeding into a linear detector that operates some form of indicator such as a visual indicator tube. A probe with a 1 $\mu\mu$f series condenser is used to pick up the signal at any desired point, and the tuned amplifiers are adjusted to the frequency being picked up. In this way, it is possible to determine the presence of a signal at any place in the receiver and also to determine the effects produced on the magnitude of the signal by the various circuits and tubes in the receiver.

Visual Test Devices.—The alignment of receiver circuits is sometimes carried out with the aid of an arrangement that traces out the response curve visually on a cathode-ray tube. This can be accomplished by using a test oscillator that has the frequency varied over a suitable range at a low audio rate. The horizontal sweep of the cathode-ray oscillograph is then synchronized with the rate of variation of the test frequency, so that the deflection along the horizontal is a function of frequency. The vertical deflection is then obtained from the output of the receiver detector or from the radio-frequency voltage applied to the detector (either with or without further amplification).

The most common method of carrying out the operations required by a visual test instrument is to obtain a frequency variation of the test oscillator by the use of a reactance tube, to the grid of which a 60-cycle voltage is applied.[2] This oscillator is preferably fixed in frequency so that the extent of the frequency sweep is determined only by the amplitude of the control voltage applied to the reactance tube. The desired output frequency is then obtained by heterodyning with a tunable oscillator and obtaining a difference frequency with the aid of a mixer tube. A typical sche-

[1] All oscillator adjustments are preferably made while the receiver tuning dial is rocked slightly in order to take care of any slight interaction between the oscillator and radio-frequency stages that may be present.

[2] A somewhat more elaborate test oscillator of this type, suitable for the circuits of television receivers, is described by H. F. Mayer, A Visual Alignment Generator, *Electronics*, Vol. 13, p. 39, April, 1940.

matic arrangement is illustrated in Fig. 92, which operates in much the same way as a beat-frequency oscillator, but with the difference frequency in this case being the desired radio frequency.

An alternative arrangement, widely used in earlier visual test devices, makes use of a motor-driven rotating condenser, with contacts on the shaft which provide synchronization with the cathode-ray sweep.[1]

Visual test devices are particularly useful where wide band circuits are employed or where good band-pass characteristics are desired. It is impossible properly to line up such circuits by adjusting to maximum response, as is usual with sharply resonant circuits.

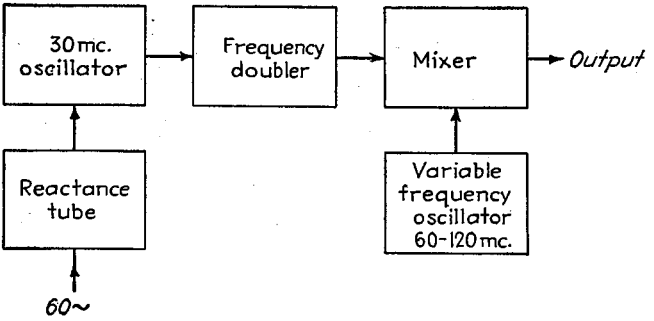


FIG. 92.—Schematic circuit arrangement of test oscillator producing a frequency-modulated wave that can be adjusted to have a carrier frequency up to 60 mc.

MODULATION

39. Measurement of Amplitude Modulation.—Modulation measurements are used in adjusting radio transmitters to proper operating conditions, to determine the amount of distortion of the modulation envelope, and in monitoring to detect over modulation.

Modulation Meter Using Rectified Envelope.[2]—This type of modulation meter is widely used for the accurate measurement of positive and negative peak modulation and for monitoring of transmitters. A typical commercial form is shown in Fig. 93. The radio-frequency wave to be investigated is rectified by a diode detector T_1, operated as a linear detector. This gives pulsating d-c output voltage that is almost an exact reproduction of the modulation envelope. The average value of this voltage gives the carrier amplitude and is read a d-c microammeter M_1. Any change in this meter reading with variation in the degree of modulation indicates carrier shift. The peak and trough values of modulation are obtained by separating out the modulation-frequency component of the rectifier output and rectifying this audio wave by a second diode detector T_2, arranged to indicate either positive or negative peaks, according to the connections. The modulation positive and negative peaks are then given by a comparison of the rectified peak voltage developed by the second diode with the carrier amplitude as read by the first diode. By adjusting the radio-frequency input voltage so that the carrier is at a predetermined level, it is possible to make the instrument direct reading in percentage modulation.

[1] An example of such an arrangement is described by G. Ulbricht, Visual Test Device, *Proc. I.R.E.*, Vol. 22, p. 89, January, 1934.

[2] See also L. F. Gaudernack, Some Notes on the Practical Measurement of the Degree of Modulation, *Proc. I.R.E.*, Vol. 22, p. 819, July, 1934; F. C. Williams, D. Phil, and A. E. Chester, A New Modulation Meter, *Wireless Eng.*, Vol. 15, p. 257, May, 1938; Verne V. Gunsolley, A Differential Modulation Meter, *Electronics*, Vol. 13, p. 18, January, 1940; H. D. M. Ellis, Measurement of Modulation Depth, *Wireless Eng.*, Vol. 18, p. 99, March, 1941.

The instrument of Fig. 93 can be arranged to flash a lamp whenever the degree of modulation exceeds a predetermined value. The required auxiliary equipment is shown in Fig. 93, and consists of an amplifier T_3 so biased that when the degree of modulation reaches the maximum allowable level, plate current begins to flow. This

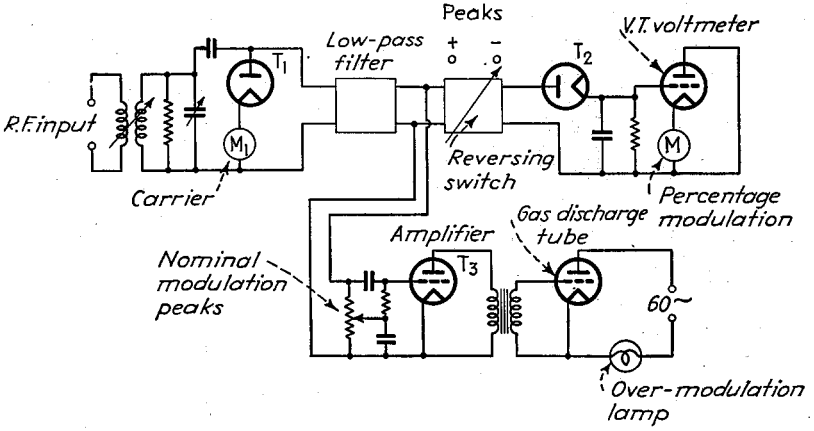


FIG. 93.—Schematic circuit diagram of a modulation meter capable of reading positive and negative peak modulations, together with overmodulation indicator.

current fires a gas-discharge tube that then causes current to pass through the lamp.

Cathode-ray Methods of Determining Modulation.—Cathode-ray oscillographs are widely used as modulation indicators. They are more generally available than are modulation meters, and are less expensive, but have the disadvantage of giving results that are less accurate.

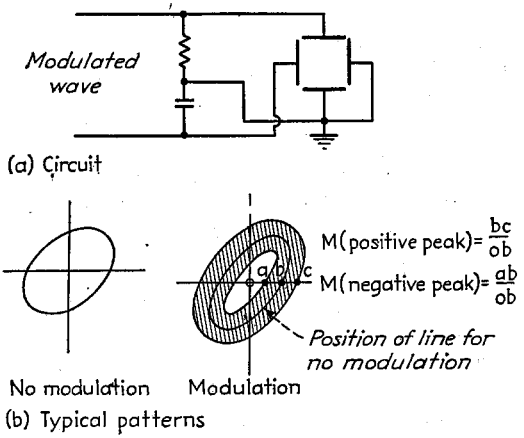


FIG. 94.—Method of measuring the degree of modulation, using a cathode-ray tube.

Cathode-ray oscillographs can be used in a number of ways to determine the degree of modulation. The peak modulation can be readily observed by simply applying the wave to be investigated to the vertical deflectors of the tube. The length of line that results with no modulation corresponds to twice carrier amplitude, while the

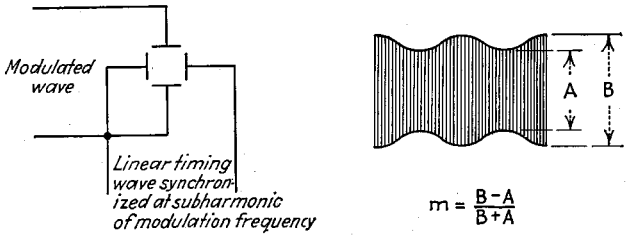
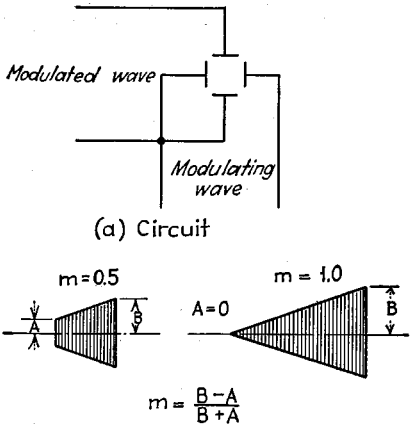


FIG. 95.—Modulated-wave envelope obtained by use of linear sweep circuit synchronized at half the modulating frequency.



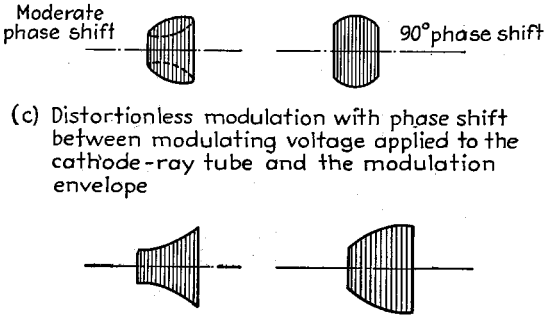


FIG. 96.—Circuit arrangement for observing modulation on a cathode-ray tube, together with typical resulting patterns.

amplitude with modulation corresponds to twice the peak amplitude of the modulation envelope.

This simple arrangement can be modified to give the modulation at the trough, as well as at the peak of the envelope, by the arrangement shown in Fig. 94*a*. Here the wave to be investigated is applied to a resistance-capacity phase splitter and the resulting components used for vertical and horizontal deflection. The result is a circular or elliptical pattern, as shown in Fig. 94*b*, which, in the absence of modulation, consists of a simple line that widens out to a band with modulation. The degree of modulation for the positive and negative peaks is calculated from distances measured along any convenient radial from the center, as shown. It will be noted that for complete modulation on the negative peaks, the eye in the center of the pattern closes completely.

When access can be had to the modulating voltage, there are additional ways in which a cathode-ray tube can be used. One possibility is to apply the radio-frequency wave to the vertical deflectors and synchronize the sweep voltage with the audio modulating wave. The result is the pattern shown in Fig. 95, which reproduces the

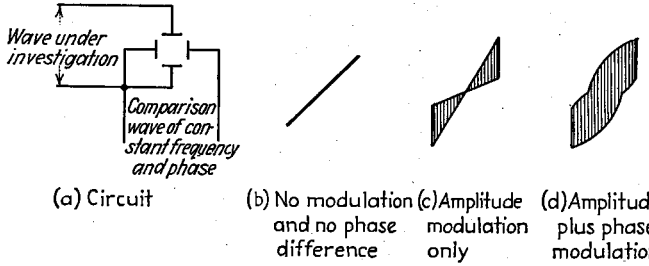


FIG. 97.—Detection of phase modulation accompanying amplitude modulation.

modulation envelope without showing the individual radio-frequency cycles. Another arrangement is to apply the modulated wave to the vertical deflectors and to use the modulating voltage to produce the horizontal deflection. When there is no phase shift between the modulation envelope and the horizontal deflecting voltage and, also, when there is no amplitude distortion in the modulation, then a trapezoid pattern with straight sides is obtained, as in Fig. 96*b*, from which the degree of modulation can be readily calculated. The presence of amplitude distortion in the modulation envelope causes the sloping sides of the trapezoid to be curved, as shown in Fig. 96*d*, while phase shift between modulation envelope and the sample of the modulating voltage used for horizontal deflection causes the sides of the pattern to be curved, and, also, under most conditions, results in a shaded ellipse appearing along the upper and lower edges, as shown in Fig. 96*c*.

40. Measurements on Frequency and Phase-modulated Waves. *Determination of Modulation Index.*—The modulation index of a frequency-modulated wave can be determined by taking advantage of the fact that with sinusoidal modulation, the carrier amplitude goes to zero whenever the modulation index has the value given in Table 3.[1]

These values of m_f correspond to $J_0(m_f) = 0$.

Detection of Phase Modulation in Transmitters.—Phase modulation often appears as an undesired by-product in transmitters making use of amplitude modulation. The presence of such phase modulation can be detected by applying the modulated radio-frequency output wave to one pair of deflectors of a cathode-ray tube and the

[1] M. G. Crosby, A Method of Measuring Frequency Deviation, *R.C.A. Rev.*, Vol. 4, p. 473, April, 1940.

TABLE 3.—VALUES OF MODULATION INDEX AT WHICH CARRIER AMPLITUDE IS ZERO

Order of Zero in Carrier Amplitude	Value of Modulation Index m_f
1	2.40
2	5.52
3	8.65
4	11.79
5	14.93
6	18.07
n $(n > 6)$	$18.07 + \pi(n - 6)$

unmodulated wave, preferably derived from the crystal oscillator, to the other deflectors. The relative phase of the two voltages is then adjusted so that with no modulation, a straight line is produced. Amplitude modulation of one of the voltages in the absence of phase shift then produces a pattern as shown in Fig. 97*c*. However, if phase modulation is present, then the pattern is modified, with one or more of the sides being elliptical rather than straight, as shown in Fig. 97*d*.

FIELD STRENGTH OF RADIO WAVES[1]

41. Measurement of Field Strength.—The strength of radio waves at frequencies below about 30 mc is normally determined by measuring the voltage induced in a loop antenna. The relation between this induced voltage and the field strength is then determined by calculation.[2]

Determination of Induced Voltage by Receiver with Calibrated Intermediate-frequency Attenuator.[3]—The voltage induced in an antenna can be conveniently measured with the aid of a receiver having an attenuator at the input of the intermediate-frequency amplifier, with the receiver so designed and operated that the output of the converter stage is proportional to the radio-frequency signal voltage acting on the converter grid up to signal amplitudes of at least one volt. Additional equipment required to measure field strength includes an auxiliary oscillator that need be only moderately well shielded, together with a vacuum-tube voltmeter capable of reading about one volt.

A typical measuring procedure with such an arrangement is as follows: The signal is tuned in and the intermediate-frequency attenuator adjusted to an attenuation α_1 that gives a convenient deflection on a direct-current microammeter located in the output of the second detector. The auxiliary oscillator is now turned on, set at the frequency of the incoming signal, and then adjusted until a voltage E (commonly one volt) appears at the grid of the converter tube, as determined by the vacuum-tube voltmeter. In this operation, the output of the auxilliary oscillator is inserted in series with the antenna. The attenuator is now readjusted to an attenuation α_2 such that the output of the second detector is the same as that originally produced by the signal. The signal voltage that was at the grid of the converter is then $(\alpha_2 - \alpha_1)$ below E, where α_1 and α_2 are in db. This voltage differs from the voltage actually induced in the antenna by the resonant rise of voltage in the loop and because of the circuits (including r-f amplifier stages, if any) between the loop and the converter grid. This ratio can be determined by removing the comparison oscillator output from in series with the antenna and applying it directly to the grid of the mixer without changing the output from that used to get the α_2 setting above. The attenu-

[1] Further information is to be found in "Standards on Radio Wave Propagation—Measuring Methods," Institute of Radio Engineers, New York, 1942.

[2] If an antenna other than a loop is used, the effective height must be determined experimentally. This is normally done by comparison with results obtained by using a loop.

[3] H. T. Friis and E. Bruce, A Radio Field-strength Measuring System for Frequencies up to Forty Megacycles, *Proc. I.R.E.*, Vol. 14, p. 507, August, 1926.

ator is then readjusted to the value α_3 that is required to give the standard output. The rise of voltage between the antenna and converter grid is then $(\alpha_2 - \alpha_3)$, and the voltage actually induced in the antenna by the signal is $(2\alpha_2 - \alpha_1 - \alpha_3)$ db below the voltage E.

Certain precautions are necessary in the design of a receiver used in such a measuring system. The beating oscillator voltage must be so arranged that this voltage is not affected by the tuning of the input circuit to the converter. The intermediate-frequency attenuator should be located immediately following the converter tube to prevent overloading of i-f stages. Attention must also be paid to the design of the input circuits and the converter operating conditions so that no overloading will occur in the section of the receiver on the input side of the attenuator.

The intermediate-frequency attenuator method of measuring induced voltage has the advantage that the accuracy is independent of signal frequency, and depends primarily upon the intermediate-frequency attenuator and the linearity of the input circuits of the receiver. Inasmuch as the input attenuator operates at a constant and not particularly high frequency, it can be easily designed to give satisfactory

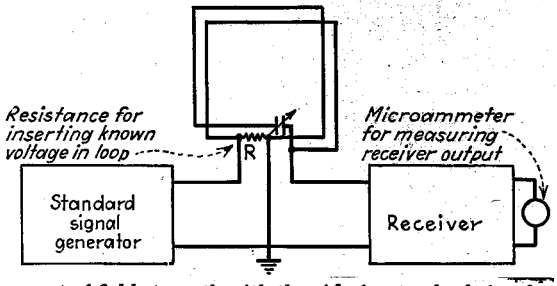


FIG. 98.—Measurement of field strength with the aid of a standard-signal generator and an ordinary receiver, by the substitution method.

performance. Also, no highly shielded oscillator is required, and the incoming signal is in effect compared with a known voltage of sufficient amplitude to be readily measured on a vacuum-tube voltmeter.

Determination of Induced Voltage by Substitution.[1]—In this method, illustrated schematically in Fig. 98, a sensitive radio receiver is connected to the antenna and tuned to the signal, with the loop oriented for maximum reception. The receiver gain is adjusted to give a convenient indication on a microammeter in the circuit of the second detector. The loop is then rotated so that no signal is received, and there is introduced in series with the loop a voltage derived from a signal generator. This voltage is adjusted to the same frequency as the signal, and the amplitude varied until the same receiver output is obtained as is produced by the signal. The signal generator output voltage then equals the equivalent induced voltage acting in series with the loop.

This method of measuring field strength is very convenient when a signal generator is available. It is commonly used at broadcast and lower frequencies, and with care can be employed at much higher frequencies.[2]

[1] C. R. Englund, Note on the Measurement of Radio Signals, *Proc. I.R.E.*, Vol. 11, p. 25, February, 1923; H. H. Beverage and H. O. Peterson, Radio Transmission Measurements on Long Wavelengths, *Proc. I.R.E.*, Vol. 11, p. 661, December, 1923; R. Bown, C. R. Englund, and H. T. Friis, Radio Transmission Measurements, *Proc. I.R.E.*, Vol. 11, p. 115, April, 1923; A. G. Jensen, Portable Receiving Sets for Measuring Field Strengths at Broadcasting Frequencies, *Proc. I.R.E.*, Vol. 14, p. 333, June, 1926.

[2] For example, field strength at very high frequencies has been measured by this method using a half-wave doublet connected to a concentric transmission line for an antenna and replacing this antenna by a concentric line leading to a signal generator for calibration purposes. See R. W. eorge, Field Strength Measuring Equipment at 500 Megacycles, *R.C.A. Rev.*, Vol. 5, p. 69, July, 1940.

Standard Field Generator.[1]—At very high frequencies, difficulty is encountered in accurately determining the loop step-up because of unbalance effects and distributed capacity in the loop. Also, in many cases, it is desirable at these frequencies to use vertical antennas, the effective height of which cannot be determined by calculation. The calibration of the receiver used in field-strength measurements can be obtained under these conditions by the use of a standard field generator.

A standard field generator consists of a compact portable oscillator with an antenna of such type that when a known current is supplied, the fields produced can be calculated from the dimensions. The combination of a standard field generator with a receiver having an intermediate-frequency attenuator permits weak signals to be measured. This is done by comparing the receiver attenuation that must be used to produce a reference output in receiving the signal of unknown strength, with the attenuation required to produce the same output when the standard field generator is producing a known field strength at the receiver antenna.

Several antenna arrangements have been used in standard field generators. One consists of a single-turn balanced loop with a thermocouple placed at the loop center and indicating the loop current (Fig. 99*a*). Another arrangement employs a short vertical antenna with a large capacity top, as shown in Fig. 99*b*, so that the effective

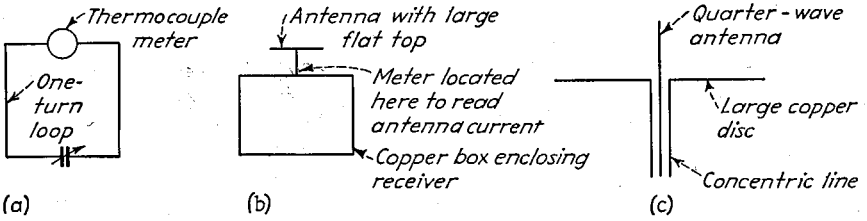


FIG. 99.—Typical antenna arrangements used in standard-field generators.

height approximates very closely the spacing between the capacity top and the copper case inclosing the oscillator. Another possibility suitable for microwaves is to employ a quarter-wave antenna, as shown in Fig. 99*c*, having a known base current.

The customary procedure followed in making use of a standard field generator at very high frequencies is to place the generator and the receiving antenna at a reasonable height above ground, preferably at least one wave length, and to vary the spacing between the two. The receiver response will vary in an oscillatory manner as indicated in Fig. 100, with peaks and troughs corresponding to conditions for which the wave reflected from the ground adds to and subtracts from, respectively, the direct ray. The mean field strength midway between the peaks and troughs, shown by the dotted line in Fig. 100, represents the intensity of the direct ray produced by the field generator, and can be calculated by conventional antenna theory from the geometry of the field-generator antenna and its current. Either vertically or horizontally polarized waves may be produced by the field generator, as the occasion requires. There is sometimes an advantage in the use of horizontal polarization, because this permits the angle of incidence with the ground to be nearly glancing before a surface wave is

[1] For further information, see J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, Ultra-short-wave Propagation, *Proc. I.R.E.*, Vol. 21, p. 427, March, 1933; J. S. McPetrie and B. G. Pressey, A Method of Using Horizontally Polarized Waves for the Calibration of Short Wave Field Strength Measuring Sets by Radiation, *Jour. I.E.E.*, Vol. 83, p. 210, 1938 (Wireless Section, *I.E.E.*, Vol. 13, p. 267, September, 1938); J. S. McPetrie and J. A. Saxton, Theory and Experimental Confirmation of Calibration of Field Strength Measuring Sets by Radiation, *Jour. I.E.E.*, Vol. 88, Part III, p. 11, March, 1941; F. M. Colebrook and A. C. Gordon-Smith, The Design and Construction of a Short-wave Field-strength Measuring Set, *Jour. I.E.E.*, Vol. 84, p. 388, 1939 (Wireless Section, *I.E.E.*, Vol. 14, p. 146, June, 1939); A Method of Calibrating a Field Strength Measuring Set, *Jour. I.E.E.*, Vol. 88, Part III, p. 15, March, 1941.

produced and also because the reflection coefficient approaches unity with horizontal polarization when the angle of incidence is small, which means that the effect of ground can be taken into account by calculation with the use of Eq. (24), Sec. 10, rather than by varying the spacing.

Accuracy of Field-strength Measurements with Particular Reference to Broadcast Frequencies.—The accuracy of field-strength measurements depends upon the frequency and the design of the equipment, and tends to be greater the lower the frequency. The relative accuracy obtained is ordinarily quite high under all conditions, particularly with the intermediate-frequency attenuator method. The absolute accuracy is less, and errors of the order of 1 to 2 db are common in equipment designed for broadcast work.

The absolute accuracy of field-strength measurements at broadcast frequencies is particularly important, because the field strength is commonly used to check the efficiency of the antenna and ground system. Investigation has shown that when such obvious sources of error as inaccurate attenuators, nonlinear converters, poor oscillator shielding, and receiver unbalance in the loop antenna, etc., have been eliminated, the largest remaining error is the distributed capacity of the loop antenna.[1] This capacity

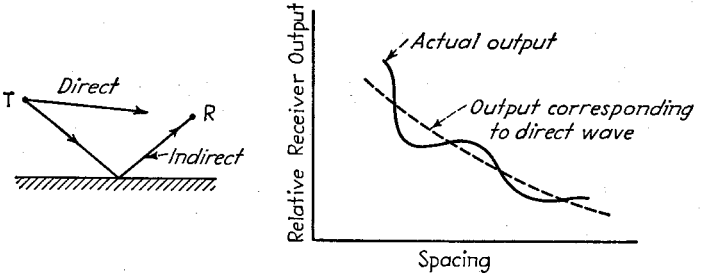


FIG. 100.—Relationship between field strength of direct wave and actual field strength produced by an elevated field generator at an elevated receiving point in the immediate vicinity, as a function of spacing.

results in a nonuniform current distribution, in which the center turns of the loop carry more current than the outer turns.[2] This effect causes a distributed induced voltage, such as is generated by the fields to be measured, to produce a different voltage across the loop output terminals than would be produced if the total induced voltage were concentrated at one point. The resulting correction factor by which one must multiply a lumped voltage introduced into the loop to get its true distributed effect is given approximately by

$$\text{Correction factor} = 1 + 0.27\left(\frac{f}{f_0}\right)^2 \tag{45}$$

where f/f_0 is the ratio of actual frequency to the natural resonant frequency of the loop. The error in field strength resulting from distributed capacity of the loop can easily reach 10 per cent. This error can be eliminated by applying the calculated correction, by using an untuned loop, or by determining the effective loop step-up ratio (*i.e.*, the loop Q) by the capacity variation method and then obtaining the

[1] An extensive study of sources of error in field-strength equipment designed for broadcast applications is given by H. Diamond, K. A. Norton, and E. G. Lapham, On the Accuracy of Radio Field Intensity Measurement at Broadcast Frequencies, *Research Paper* RP 1156, *Nat. Bur. Standards* 21, p. 795, December, 1938.

[2] Further discussion is given by Paul B. Taylor, Theory of Loop Antenna with Leakage between Turns, *Proc. I.R.E.*, Vol. 25, p. 1475, December, 1937; F. M. Colebrook, The Application of Transmission Line Theory to Closed Aerials, *Jour. I.E.E.*, Vol. 83, p. 403, 1938; also, Wireless Section, *I.E.E.*, Vol. 13, p. 273, September, 1938.

equivalent induced voltage from the potential appearing at the grid of the first tube and the effective Q of the loop circuit.

Field-strength surveys, particularly those conducted in connection with broadcast stations, are commonly made by mounting the measuring equipment in a car with the loop projecting above the top. In such an arrangement, the car distorts the radio field in its vicinity by an amount depending on the orientation of the car, loop position, etc. Errors of the order of ± 25 per cent in field strength may be expected, and, in general, tend to be independent of frequency.[1]

Continuous Recording of Field Strength.—Continuous records of field strength may be obtained by using a receiver provided with some form of automatic volume control with which is associated a continuous recorder. Arrangements that have been employed include using the current flowing through the bias resistor of an ordinary automatic-volume-control system to operate the recorder;[2] using the recorder to indicate the output of the radio receiver and obtaining the automatic volume control by connecting the recorder shaft to the volume control of the receiver in such a way that as the output increases the receiver gain is decreased, and vice versa;[3] and, finally, using arrangements in which provision is made that if the output over a period of 5 to 10 sec differs from a standard value, the receiver gain is increased or decreased as required by a definite amount, with the recorder operating from the volume-control shaft.[4]

Continuous field-strength recorders are preferably arranged so that the indication is proportional to the field strength measured in decibels. The receiver employed must be stable in its characteristics, and should have the voltage of its power supply closely regulated. The calibration can be readily checked at regular intervals by introducing known voltages in the antenna from a signal generator.

42. Measurement of Static and Noise.—Static and noise are difficult to measure satisfactorily, because such disturbances often vary widely in characteristics, and it is impossible to set up an objective criterion of the disturbing effect that will be generally applicable to all conditions. In the case of hiss type of noise or static that is a more or less continuous rumble, the disturbing effect can be measured in terms of the field strength of a just audible warble tone or the field strength of just audible telegraph signals. The average energy of the noise and static in such cases is also indicative of the intensity, and can be determined by using in the field-strength measuring equipment some form of output system that averages the response over short time intervals.[5]

Observations can be made upon individual static pulses by the use of a cathode-ray tube. Methods have been devised whereby both the direction from which a particular static impulse arrives, as well as its intensity, is shown on a cathode-ray screen.[6] The wave form of individual pulses has also been studied with high-speed cathode-ray tubes.[7]

[1] J. H. De Witt, Jr., and A. C. Omberg, The Relation of the Carrying Car to the Accuracy of Portable Field-intensity-measuring Equipment, *Proc. I.R.E.*, Vol. 27, p. 1, January, 1939.

[2] G. D. Robinson, Wide Range Scales for Fading Records by Electrical Means, *Proc. I.R.E.*, Vol. 19, p. 247, February, 1931.

[3] P. A. deMars, G. W. Kenrick, and G. W. Pickard, Use of Automatic Recording Equipment in Radio Transmission Research, *Proc. I.R.E.*, Vol. 19, p. 1618, September, 1931.

[4] W. W. Mutch, A Note on an Automatic Field Strength and Static Recorder, *Proc. I.R.E.*, Vol. 20, p. 1914, December, 1932.

[5] An example of such equipment is described by H. T. Friis, A Static Recorder, *Bell System Tech. Jour.*, Vol. 5, p. 282, April, 1926.

[6] R. A. Watson-Watt and J. F. Herd, An Instantaneous Direct Reading Radio Goniometer, *Jour. I.E.E.*, Vol. 64, p. 611, May, 1926; A. E. Harper, Some Measurements on the Directional Distribution of Static, *Proc. I.R.E.*, Vol. 17, p. 1214, July, 1929; see also Par. 5, Sec. 12 for further details.

[7] Harold Norinder, Cathode Ray Oscillographic Investigation of Atmospherics, *Proc. I.R.E.*, Vol. 24, p. 287, February, 1936.

AUTHOR INDEX

I

J

K

SUBJECT INDEX

D

E

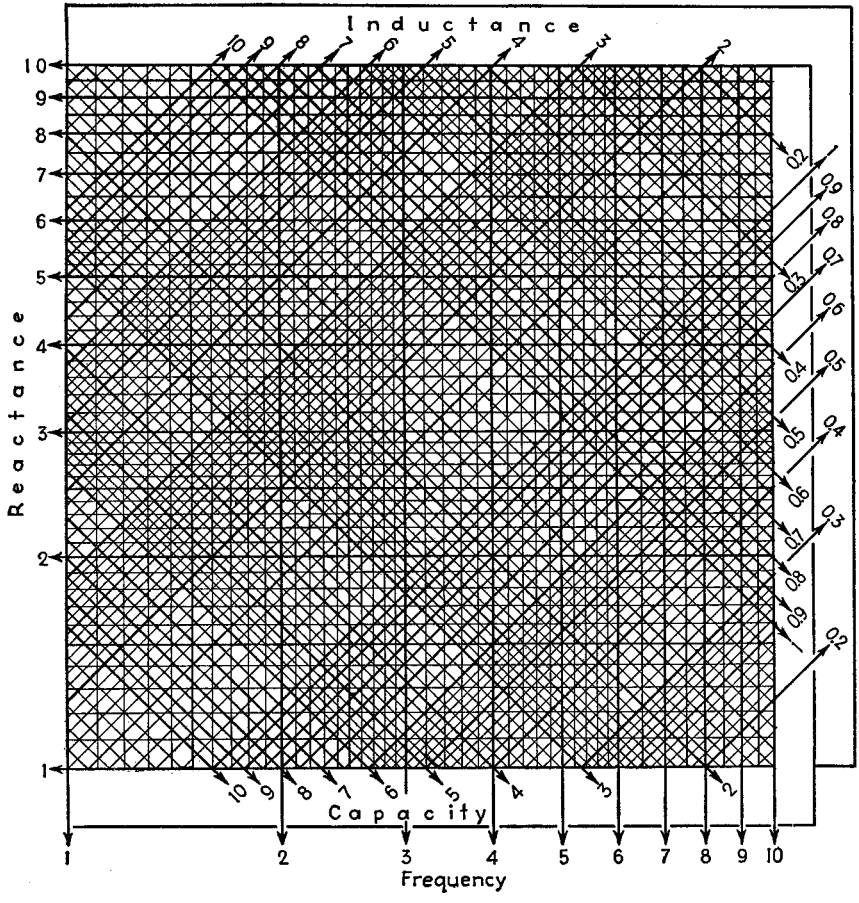
T

U

V

FIG. 1.—REACTANCE CHART*

Always obtain approximate value from Fig. 2 before using Fig. 1



Use of Fig. 1

Figure 1 is used to obtain additional precision of reading but does not place the decimal point which must be located from a preliminary entry on Fig. 2. Since the chart necessarily requires two logarithmic decades for inductance and capacitance for every single decade of frequency and reactance, unless the correct decade for L and C is chosen, erroneous results are obtained.

Typical Results

1. Find reactance of inductance of 0.00012 henry at 960 kc.

Answer: 720 ohms.

2. What capacity will have 265 ohms reactance at 7000 kc?

Answer: 86 $\mu\mu$f.

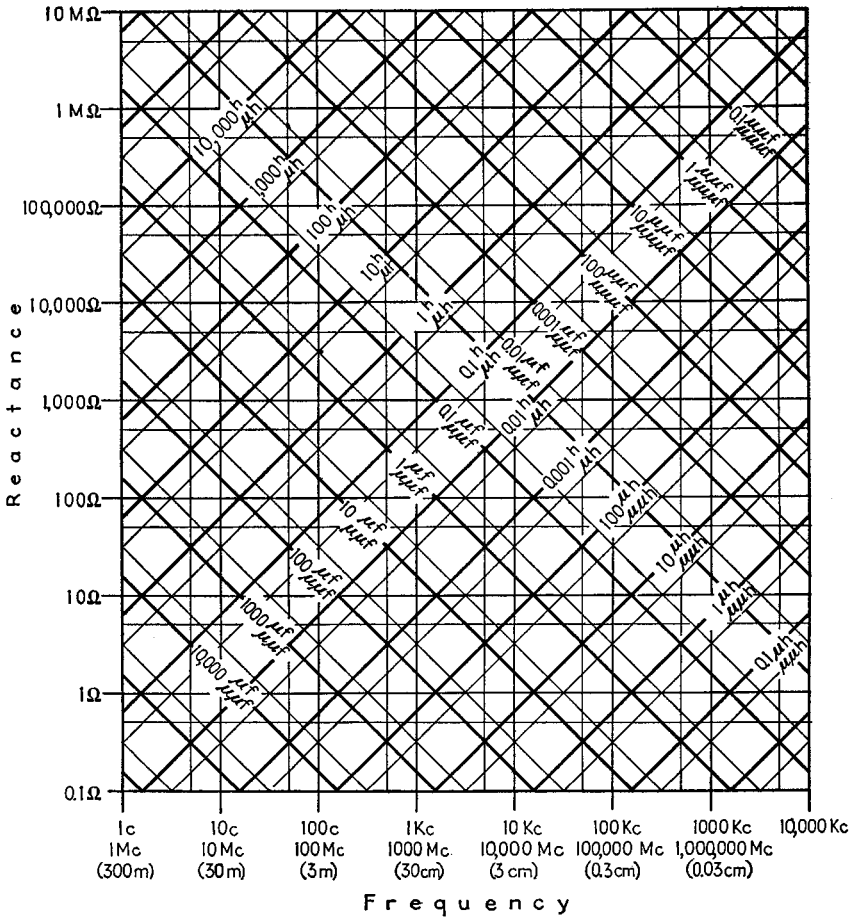
3. What is the resonant frequency with $L = 21$ μh and $C = 45\mu\mu$f?

Answer: 5.18 mc.

* Figures 1 and 2 are from General Radio Company Reactance Chart.

FIG. 2.—REACTANCE CHART*

Always use corresponding scales



The accompanying chart may be used to find

(1) The reactance of a given inductance at a given frequency.

(2) The reactance of a given capacity at a given frequency.

(3) The resonant frequency of a given inductance and capacitance.

In order to facilitate the determination of magnitude of the quantities involved to two or three significant figures the chart is divided into two parts. Figure 2 is the complete chart to be used for rough calculations. Figure 1, which is a single decade of Fig. 2 enlarged approximately 7 times, is to be used where the significant two or three figures are to be determined.

To Find Reactance

Enter the charts vertically from the bottom (frequency) and along the lines slanting upward to the left (inductance) or to the right (capacity). Corresponding scales (upper or lower) must be used throughout. Project horizontally to the left from the intersection and read reactance.

To Find Resonant Frequency

Enter the slanting lines for the given inductance and capacity. Project downward from their intersection and read resonant frequency from the bottom scale. Corresponding scales (upper or lower) must be used throughout.

* Figures 1 and 2 are from General Radio Company Reactance Chart.