PRINCIPLES OF

RADIO

SIXTH EDITION

KEITH HENNEY
GLEN A. RICHARDSON

AVERAGE CHARACTERISTICS—RECEIVING TUBES

AMPLIFIERS, DETECTORS, OSCILLATORS, ETC.

Tube Type	Purpose	Cathode Type	Filament		Plate		Screen Volts	Grid Volts	Plate Resistance, Ohms	Transconductance, Micromhos	Amplification Factor	Power Output, Watts	Load Resistance, Ohms
			Volts	Amp	Volts	Ma							
1T4 *	Super-control pentode amplifier	Filament	1.4	0.05	90	3.5	67.5	0	0.5 megohm	900	.....	.....	.....
3S4 *	Power amplifier pentode	Filament	2.8 1.4	0.05 0.1	90	6.1	67.5	−7	0.1 megohm	1425	.....	0.235	8000
6C4 *	H-F power triode	Heater	6.3	0.15	300	25	.....	−27	......	....	.....	5.5 ‡	.....
6J5	General-purpose triode	Heater	6.3	0.3	250	9	.....	−8	7,700	2600	20	.....	.....
6SF5	High-mu triode	Heater	6.3	0.3	250	0.9	.....	−2	66,000	1500	100	.....	....
2A3	Power amplifier triode	Filament	2.5	2.5	250	60	.....	−45	800	5250	4.2	3.5	2500
6K6-GT	Power amplifier pentode	Heater	6.3	0.4	250	32	250	−18	68,000	2300	.....	3.4	7600
6L6	Beam power amplifier	Heater	6.3	0.9	250	72	250	−14	22,500	6000	.....	6.5	2500
35C5	Beam power amplifier	Heater	35	0.15	110	40	110	−7.5	13,000	5800	.....	1.5	2500
6SJ7	Sharp cut-off pentode amplifier	Heater	6.3	0.3	250	3.0	100	−3	......	1650	.....	.....	....
6SK7	Super-control pentode amplifier	Heater	6.3	0.3	250	9.2	100	−3	0.8 megohm	2000	.....	.....	....
6SN7	Twin-triode amplifier	Heater	6.3	0.6	250 †	9	.....	−8	7,700	2600	20	.....	....
6SQ7	Duplex-diode, high-mu triode	Heater	6.3	0.3	100	0.4	.....	−1	0.11 megohm	900	100	.....	....
25L6	Beam power amplifier	Heater	25.0	0.3	110	49	110	−7.5	10,000	8200	.....	2.2	2000

* Miniature type. † Values for each unit. ‡ As oscillator.

RECTIFYING TUBES

Tube Type	Purpose	Filament		Typical Operating Conditions	
		Volts	Amp	Peak Inverse Volts, Max	D-C Output, Ma
5U4-G	Full-wave rectifier	5.0	3.0	1550	225
5V4-G	Full-wave rectifier	5.0	2.0	1400	175
5Y3-GT	Full-wave rectifier	5.0	2.0	1400	125
6X4 *	Full-wave rectifier	6.3	0.6	1250	70
25Z6	Rectifier-doubler	25	0.3	700	75 per plate
117Z3 *	Half-wave rectifier	117	0.04	330	90
6AL5 *	Rectifier-detector	6.3	0.3	420	9 per plate
6H6	Rectifier-detector	6.3	0.3	420	8 per plate

* Miniature type.

CONVERTER TUBES

Tube Type	Filament		Plate		Screen		Control Grid Volts	Plate Resistance, Megohms	Conversion Conductance, Micromhos
	Volts	Amp	Volts	Ma	Volts	Ma			
1R5 *	1.4	0.05	45	0.7	45	1.9	0	0.6	235
1C7-G	2.0	0.12	135	1.3	67.5	2.5	−3	0.6	300
6BA7 *	6.3	0.3	100	3.6	100	10.2	−1	0.5	900
6BE6 *	6.3	0.3	100	2.8	100	7.3	−1.5	0.5	455
6D8-G	6.3	0.15	135	1.5	67.5	1.7	−3	0.6	325
6K8	6.3	0.3	250 † 100 ‡	2.5 3.8	100	6.0	−3	0.6	350
6SA7	6.3	0.3	100	3.3	100	8.5	0	0.5	425
12BE6 *	12.6	0.15	100	2.8	100	7.3	−1.5	0.5	455

* Miniature type. † Mixer plate. ‡ Oscillator plate.

RELATION BETWEEN WAVELENGTH (METERS), FREQUENCY (KILOCYCLES), AND THE PRODUCT OF INDUCTANCE (MICROHENRIES) AND CAPACITY (MICROFARADS). *Note.* $LC = 25{,}303/f^2$ in Kc, μh, and $\mu\mu$f

Meters	f, Kc	$L \times C$	Meters	f, Kc	$L \times C$	Meters	f, Kc	$L \times C$
1	300,000	0.0000003	450	667	0.0570	740	405	0.1541
2	150,000	0.0000011	460	652	0.0596	745	403	0.1562
3	100,000	0.0000018	470	639	0.0622	750	400	0.1583
4	75,000	0.0000045	480	625	0.0649	755	397	0.1604
5	60,000	0.0000057	490	612	0.0676	760	395	0.1626
6	50,000	0.0000101	500	600	0.0704	765	392	0.1647
7	42,900	0.0000138	505	594	0.0718	770	390	0.1669
8	37,500	0.0000180	510	588	0.0732	775	387	0.1690
9	33,333	0.0000228	515	583	0.0747	780	385	0.1712
10	30,000	0.0000282	520	577	0.0761	785	382	0.1734
20	15,000	0.0001129	525	572	0.0776	790	380	0.1756
30	10,000	0.0002530	530	566	0.0791	795	377	0.1779
40	7,500	0.0004500	535	561	0.0806	800	375	0.1801
50	6,000	0.0007040	540	556	0.0821	805	373	0.1824
60	5,000	0.0010140	545	551	0.0836	810	370	0.1847
70	4,290	0.0013780	550	546	0.0852	815	368	0.1870
80	3,750	0.0018010	555	541	0.0867	820	366	0.1893
90	3,333	0.0022800	560	536	0.0883	825	364	0.1916
100	3,000	0.00282	565	531	0.0899	830	361	0.1939
110	2,727	0.00341	570	527	0.0915	835	359	0.1962
120	2,500	0.00405	575	522	0.0931	840	357	0.1986
130	2,308	0.00476	580	517	0.0947	845	355	0.201
140	2,143	0.00552	585	513	0.0963	850	353	0.203
150	2,000	0.00633	590	509	0.0980	855	351	0.206
160	1,875	0.00721	595	504	0.0996	860	349	0.208
170	1,764	0.00813	600	500	0.1013	865	347	0.211
180	1,667	0.00912	605	496	0.1030	870	345	0.213
190	1,579	0.01015	610	492	0.1047	875	343	0.216
200	1,500	0.01126	615	488	0.1065	880	341	0.218
210	1,429	0.01241	620	484	0.1082	885	339	0.220
220	1,364	0.01362	625	480	0.1100	890	337	0.223
230	1,304	0.01489	630	476	0.1117	895	335	0.225
240	1,250	0.01621	635	472	0.1135	900	333	0.228
250	1,200	0.01759	640	469	0.1153	905	331	0.231
260	1,154	0.01903	645	465	0.1171	910	330	0.233
270	1,111	0.0205	650	462	0.1189	915	328	0.236
280	1,071	0.0221	655	458	0.1208	920	326	0.238
290	1,034	0.0237	660	455	0.1226	925	324	0.241
300	1,000	0.0253	665	451	0.1245	930	323	0.243
310	968	0.0270	670	448	0.1264	935	321	0.246
320	938	0.0288	675	444	0.1283	940	319	0.249
330	909	0.0306	680	441	0.1302	945	317	0.251
340	883	0.0325	685	438	0.1321	950	316	0.254
350	857	0.0345	690	435	0.1340	955	314	0.257
360	834	0.0365	695	432	0.1360	960	313	0.259
370	811	0.0385	700	429	0.1379	965	311	0.262
380	790	0.0406	705	426	0.1399	970	309	0.265
390	769	0.0428	710	423	0.1419	975	308	0.268
400	750	0.0450	715	420	0.1439	980	306	0.270
410	732	0.0473	720	417	0.1459	985	305	0.273
420	715	0.0496	725	414	0.1479	990	303	0.276
430	698	0.0520	730	411	0.1500	995	302	0.279
440	682	0.0545	735	408	0.1521	1000	300	0.282

TRIGONOMETRIC FUNCTIONS

Angle α	Sin α	Cos α	Tan α	Angle α	Sin α	Cos α	Tan α
0	.000	1.000	.000				
1	.017	1.000	.017	46	.719	.695	1.04
2	.035	.999	.035	47	.731	.682	1.07
3	.052	.999	.052	48	.743	.669	1.11
4	.070	.998	.070	49	.755	.656	1.15
5	.087	.996	.088	50	.766	.643	1.19
6	.105	.995	.105	51	.777	.629	1.23
7	.122	.993	.123	52	.788	.616	1.28
8	.139	.990	.141	53	.799	.602	1.33
9	.156	.988	.158	54	.809	.588	1.38
10	.174	.985	.176	55	.819	.574	1.43
11	.191	.982	.194	56	.829	.559	1.48
12	.208	.978	.213	57	.839	.545	1.54
13	.225	.974	.231	58	.848	.530	1.60
14	.242	.970	.249	59	.857	.515	1.66
15	.259	.966	.268	60	.866	.500	1.73
16	.276	.961	.287	61	.875	.485	1.80
17	.292	.956	.306	62	.883	.469	1.88
18	.309	.951	.325	63	.891	.454	1.96
19	.326	.946	.344	64	.899	.438	2.05
20	.342	.940	.364	65	.906	.423	2.14
21	.358	.934	.384	66	.914	.407	2.25
22	.375	.927	.404	67	.920	.391	2.36
23	.391	.920	.424	68	.927	.375	2.48
24	.407	.914	.445	69	.934	.358	2.61
25	.423	.906	.466	70	.940	.342	2.75
26	.438	.899	.488	71	.946	.326	2.90
27	.454	.891	.510	72	.951	.309	3.08
28	.469	.883	.532	73	.956	.292	3.27
29	.485	.875	.554	74	.961	.276	3.49
30	.500	.866	.577	75	.966	.259	3.73
31	.515	.857	.601	76	.970	.242	4.01
32	.530	.848	.625	77	.974	.225	4.33
33	.545	.839	.649	78	.978	.208	4.70
34	.559	.829	.675	79	.982	.191	5.14
35	.574	.819	.700	80	.985	.174	5.67
36	.588	.809	.727	81	.988	.156	6.31
37	.602	.799	.754	82	.990	.139	7.12
38	.616	.788	.781	83	.993	.122	8.14
39	.629	.777	.810	84	.995	.105	9.51
40	.643	.766	.839	85	.996	.087	11.43
41	.656	.755	.869	86	.998	.070	14.30
42	.669	.743	.900	87	.999	.052	19.08
43	.682	.731	.933	88	.999	.035	28.64
44	.695	.719	.966	89	1.000	.017	57.29
45	.707	.707	1.000	90	1.000	.000	Infinity

Reprinted by permission from Timbie, *Basic Electricity for Communications.*

PRINCIPLES OF
RADIO

SIXTH EDITION

KEITH HENNEY

CONSULTING EDITOR, *Electronics*
FELLOW, INSTITUTE OF RADIO ENGINEERS

GLEN A. RICHARDSON

ASSISTANT PROFESSOR OF
ELECTRICAL ENGINEERING
IOWA STATE COLLEGE

JOHN WILEY & SONS, INC., NEW YORK
CHAPMAN & HALL, LIMITED, LONDON

Library of Congress Catalog Card Number 51-13005

Printed in the United States of America

PRINCIPLES OF

RADIO

Preface

Seven years is a relatively long time in a field that advances as rapidly as radio, for in that time much that was undreamed of can be conceived, developed, and brought into wide production and use. Seven years have now elapsed since the Fifth Edition of *Principles of Radio* was prepared; and, as always, some that is new now was barely thought of in 1945 and, as always, emphasis has shifted so that some things in wide use in the 1940's are now considered a bit out of date.

This edition represents a complete overhaul, Professor Richardson preparing the first draft of the new text and the senior author developing the final manuscript. The same viewpoint has been retained that made the earlier editions useful to those who must learn radio without help of a teacher—the text must be as clear as it is possible to make technical matter. It must also be practical, so that the reader not only learns principles but gets a sense of values to be found in practice as well.

The illustrations and problems are all new or revised, and only a direct comparison with the Fifth Edition will show the great many additions, deletions, and other changes in text. Much new material has been added, and all the older material has been rewritten, rearranged, or deleted.

Throughout this job of making a new book, the authors had the benefit of a thorough reading of the manuscript by Dr. J. E. Schmidt who, although not a radio man at all, has a very great interest in radio technique and at the same time a high degree of ability with the English language. His criticism has helped keep the text easy to understand.

KEITH HENNEY

Contents

1 · Fundamentals

It is essential for anyone studying radio engineering to know something about electronics, for radio apparatus of today is built almost entirely around the electron tube. Consequently, a question of fundamental importance is: what is the electron?

1.1. Electrons. Since the earliest days man has tried to find out what his universe is made of; to find out the smallest possible unit of which matter is made. He first discovered that there are very few different substances, less than 100, now called **elements.** He discovered that all known substances can be made out of these comparatively few elemental atoms, which differ from one another in chemical, physical, and electrical characteristics.

But it was a long time before man discovered that the atoms were made up entirely of a few still more basic units of matter. It is now known that the **nucleus** contains **neutrons,** which are uncharged electrically, and **protons,** which carry positive electric charges. Nearly the entire weight of the atom is made up of protons and neutrons. Around the nucleus move a number of **electrons** which are electrically negative in charge.

The entire art and science of electronics is built up around the basic fact that the electrons of atoms are electrically charged and that they can be moved about from place to place in accordance with man's wishes. Every time an electron moves it constitutes an electric current, and, therefore, it is the movements of electrons that make up all there is to electronics.

Atoms of the 90-odd elements (hydrogen, oxygen, iron, sulfur, gold, etc.) differ from one another only in the number of electrons and the make-up of the nucleus.

1.2. Atoms. Atoms are inconceivably small in comparison with every-day objects. Oil falling on a wet street tends to spread out until the oil film may be only half a ten-millionth of an inch thick. Yet the atoms of which the oil is made must be smaller than this, for there is no way to spread the film so that it is less than one atom thick.

Radio men are, in general, concerned with two kinds of substances: (1) conductors, such as metals; and (2) insulators, such as glass or mica. In the metals, the atoms are rather closely packed together. The electrons of these atoms revolve in elliptical orbits about the nucleus. The electrons in the outermost orbits are not so tightly held to the nucleus and can be detached from it. In fact, the atoms of a metal are continually exchanging electrons, so that it is possible for an individual electron to travel from atom to atom, from one end of a wire to the other, if proper conditions exist. In insulators, on the other hand, the electrons are held very closely to their parent atoms and are separated from these atoms with great difficulty.

1.3. Electrical charges. An atom, or a molecule—which is a group of atoms—carrying equal amounts of positive and negative electricity is said to be in equilibrium and tends to stay in this condition. It is possible, however, to make the molecule or atom lose or gain one or more electrons, and then it is said to be **charged** or **ionized.** If a charged particle, such as an atom or a molecule, has too few electrons, it has too little negative electricity to balance its positive charge and will pick up enough electrons to get back into equilibrium at the first opportunity. A body lacking electrons is said to be positively charged. On the other hand, if it has too many electrons it is negatively charged and will tend to lose electrons when it gets a chance. A charged body in a liquid or gas is called an **ion.**

Two electrical charges which are alike, that is, two positive or two negative charges, repel each other; two unlike charges, a positive and a negative, attract each other. The force tending to make the charges move toward or away from each other depends upon the amount of charge and varies inversely with the square of the distance between the charges.* It also depends on the nature of the substance separating the charges. Within the

* The mathematical expression for this force is known as **Coulomb's law:**

$$F = \frac{q_1 q_2}{kd^2}$$

where F is the force in dynes, q_1 and q_2 are the respective charges in statcoulombs, k is the dielectric constant, and d is the distance in centimeters between the particles. For air $k = 1$. The dielectric constant of other atom itself, where the distances are exceedingly small, the attractive force holding the atom and its component charges together may be very great.

If one could collect and place a kilogram of electrons at each of the poles of the earth (8000 miles apart), the two groups of electrons would repel each other with a force of about 200 million million tons.

1.4. Electrons in motion. The electric current. Electrons, together with their positive counterparts, the protons, and the neutrons which exist in the nucleus, are the basic building blocks of the atoms. Everything is made up of these elementary blocks. The electron is essentially an electrical charge which we arbitrarily call negative. It will be attracted to any positively charged body, and when it moves it carries its negative charge with it. If sufficient electrons move from one place to another, a definite "flow" of electricity takes place.

Note that the word "flow" does not indicate that electricity is something like water. For many years this was the belief, and so the word "flow" has come to be used for the transport of charges from one place to another. We say that an electric current flows. Actually, electric energy can be made to move from one place to another with no real transfer of electrons (charges) between these places. Radio communication is an example of such a transfer of energy. An electric current, whether passing through a wire or through the vacuum of an electron tube, is merely a movement of electric charges.

In a piece of copper wire at room temperature the electrons of the individual atoms are moving about rather slowly and haphazardly. As long as the wire is not attached to a source of electrical charge, the electrons have no preference as to the direction in which they move, and in the wire there is no tendency for the electrons to move from one end to the other. If, however, the two ends of the wire are attached to the terminals of a battery, then the electrons will move along the wire, each representing an elemental quantity of electricity. This quantity is so small that engineers are generally concerned with the motion of electrons only when large numbers of them move from one place to another. It has been estimated that all the inhabitants of the earth, count-

insulators is greater than 1. Coulomb's law may also be stated in other systems of units.

ing day and night at the highest rate of speed possible, would need two years to count the number of electrons which pass through an ordinary electric light bulb in a second.

1.5. Units. Just as a carpenter orders a board so many feet long, if he is in this country, or so many meters long, if he is in France, the electrical man must have units for the quantities he uses. The amount of electrical charge each electron represents is now well established; it is exceedingly small. The practical unit of charge is known as the **coulomb.*** There are 6.28 million million million electrons in 1 coulomb of electricity. If this number of electrons moves through an electric lamp each second, 1 **ampere** of current will flow; and if this lamp is operated from ordinary house-wiring circuits, 115 **watts** of power will be used.

Here, then, are two units: the coulomb representing a *quantity* of electricity, and the ampere representing the *rate of flow* of this quantity. The coulomb corresponds to the gallon in fluid measure, and the ampere to gallons per minute. A current of 1 ampere means a rate of flow of 1 coulomb of electricity per second.

The current flowing through a radio receiver tube is of the order of a few thousandths of an ampere. A small-town power plant may have thousands of amperes flowing from its generators. Approximate currents flowing through some common devices are shown in Table 1.

TABLE 1

Apparatus	Approximate Current, Amperes
50-watt lamp	0.5
250-watt lamp	2.5
1-hp motor	10.0
Electric iron	5.0
Filament of a battery-type radio tube	0.05
Plate circuit of electron tube	0.005

A meter to measure the flow of electricity is called an **ammeter.**

1.6. Engineers' shorthand. Engineers have a simple shorthand method of working with large numbers. For example, the

* Note that the unit of charge is given here as a coulomb, whereas it was given as a stat-coulomb in the statement of Coulomb's law on p. 2. This is an outgrowth of the various unit systems. The coulomb is the more widely used unit in general practice. A coulomb is 3×10^9 larger than the stat-coulomb.

number 6.28 million million million is expressed as 6.28×10^{18}. This many electrons flowing past a given point per second constitute the electric current known as 1 ampere. As we shall have occasion to use this shorthand system many times in this book, the reader is encouraged to master it as soon as possible. The table below will be helpful.

$$1 = 10^0 = \text{one}$$
$$10 = 10^1 = \text{ten}$$
$$100 = 10^2 = \text{hundred}$$
$$1000 = 10^3 = \text{thousand, etc.}$$

$$1 = 10^0 = \text{one}$$
$$0.1 = 10^{-1} = \tfrac{1}{10} = \text{one-tenth}$$
$$0.01 = 10^{-2} = \tfrac{1}{100} = \text{one-hundredth}$$
$$0.001 = 10^{-3} = \tfrac{1}{1000} = \text{one-thousandth, etc.}$$

The small number above the figure 10 is called the **exponent.** Numbers less than 1 always have negative exponents. Thus three-thousandths may be expressed in these several ways:

$$0.003 = 3 \times 10^{-3} = \frac{3}{1000} = \frac{3}{10^3}$$

When numbers are multiplied, their exponents are added; when numbers are divided, the exponents are subtracted. Thus 100 multiplied by four-tenths may be done in shorthand as follows:

$$\begin{aligned} 100 \times 0.4 &= 10^2 \times 4 \times 10^{-1} \\ &= 4 \times 10^1 \\ &= 4 \times 10 \\ &= 40 \end{aligned}$$

Similarly, let us divide 3000 by 150:

$$\begin{aligned} 3000 \div 150 &= (3 \times 10^3) \div (1.5 \times 10^2) \\ &= \frac{3}{1.5} \times 10^3 \times 10^{-2} \\ &= 2 \times 10 \\ &= 20 \end{aligned}$$

The rules are few and simple:

1. To multiply, add exponents.
2. To divide, subtract exponents.
3. When any number crosses the division line, change the sign of the exponent.

Example 1. Multiply 20,000 by 1200 and divide the result by 6000.

$$20{,}000 = 2 \times 10^4$$

$$1200 = 12 \times 10^2$$

$$6000 = 6 \times 10^3$$

$$\frac{20{,}000 \times 1200}{6000} = \frac{2 \times 10^4 \times 12 \times 10^2}{6 \times 10^3}$$

$$= \frac{2 \times 12 \times 10^4 \times 10^2 \times 10^{-3}}{6}$$

$$= \tfrac{24}{6} \times 10^3$$

$$= 4000$$

Problem 1. How many electrons flow past a given point per second when the current in amperes is 10? 1000? 0.5? 0.0004?

Problem 2. How many coulombs per second flow past a given point for the currents of Prob. 1?

Problem 3. How many amperes of current flow when 4.396×10^{13} electrons per second flow past a point?

Problem 4. If light travels 300 million meters per second, and a "light year" is the distance light travels in a year, how many meters does a light year represent?

Problem 5. An electron weighs 9.11×10^{-31} kilogram and carries a charge of 1.60×10^{-19} coulomb. How much charge would a kilogram of electrons carry?

In connection with such shorthand methods Table 2 of prefixes commonly used will be important.

TABLE 2

Prefix	Symbol	Meaning	Shorthand
micro	μ	one-millionth	10^{-6}
milli	m	one-thousandth	10^{-3}
centi	c	one-hundredth	10^{-2}
deci	d	one-tenth	10^{-1}
deka	dk	ten	10
hekto	h	one hundred	10^2
kilo	k or K	one thousand	10^3
mega	M	one million	10^6
megamega	MM	one million million	10^{12}

Thus a thousandth of an ampere is known as a milliampere, a million ohms is called a megohm, etc.; or, expressed in numbers, 1 ma = 10^{-3} or 0.001 amp; 1 megohm = 1,000,000 ohms.

Example 2. How many milliamperes are there in 2 amp? Since 1 ma is equal to one-thousandth ampere, 1 amp is equal to 1000 ma. Thus, 1 amp = 1000 ma = 10^3 ma. Therefore 2 amp is equal to 2×10^3 ma or 2000 ma.

How many amperes are in 2 ma? Here one must remember that there are fewer amperes in 1 ma than there are milliamperes in 1 amp because the ampere is the larger unit. Thus, 1 ma = 0.001 amp = 10^{-3} amp. Therefore 2 ma = 2×10^{-3} amp.

Problem 6. How many watts are in 5 microwatts? How many microwatts in 500 watts?

Problem 7. How many kilohms are in 20 megohms? How many milliohms in 50 kilohms?

1.7. Curve plotting. Many radio problems can be solved without any mathematics at all if one understands the technique of plotting curves. A curve is a visual means of portraying what happens to one quantity when another is varied. For example, the curve of Fig. 1 shows the distance traveled by a train moving

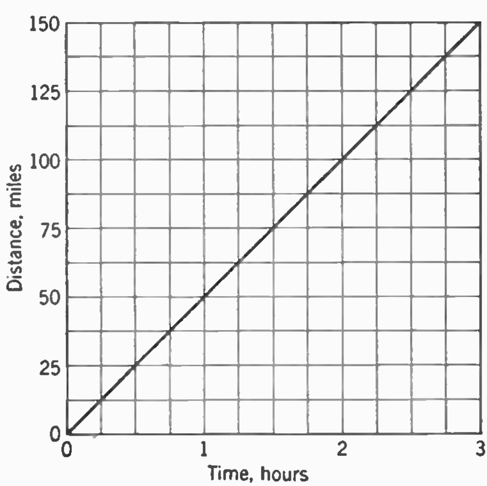


FIG. 1. Plot or curve showing distance traveled by train as a function of the time elapsed since the train started.

at a fixed rate of speed. One makes such a curve in the following manner: At zero time, the train is zero distance from its starting point. This is at the lower left-hand corner of the plot. Now let us represent time along the horizontal part of the plot, distance along the vertical. If the train goes 50 miles per hour, at the end of the first hour it will be 50 miles from the starting point (called the **origin** in curve plotting). At the end of the second hour it will be 2×50 or 100 miles away; at the end of the third hour it will be 150 miles away; and so on. All we need to do to make a plot of this kind is to put a mark at the vertical value corresponding to each horizontal value we may choose and then draw a smooth curve which best fits the points.

This curve is a visual picture of the position of the train for each portion of the time we may be interested in. The curve correlates two variables—time and distance.

If the train travels at a uniform speed, the curve correlating distance and time will be a straight line. The slope of the curve (Sect. 1.8) expresses the rate of change of distance with respect to time. This quantity is stated in miles per hour and is known as **velocity** or **speed.**

The two factors in this simple graph are known as the **variables,** one being dependent upon the other. Here the independent variable is time, and the dependent variable is distance since the distance traveled depends upon the elapsed time.

A curve such as we have been discussing has two coordinates, horizontal and vertical, which represent the independent and dependent variables. These reference lines are called the **axes.** (See Fig. 2.) The vertical axis is often called the **ordinate** or Y-axis, and the horizontal the **abscissa** or X-axis. Horizontal distances to the left of the ordinate are negative; those to the right are positive. Similarly, vertical distances below the abscissa are negative, and those above are positive. Thus we can plot both positive and negative quantities on such a chart.

1.8. Slope. The change in vertical units with a given change in horizontal units is called the **slope** of the curve. This is an important factor since it shows the *rate* at which one quantity varies with respect to the other. The actual appearance of the curve will change, depending upon the units employed, but the numerical value of the slope will not change. For example, if the vertical units are doubled in value the curve will appear flattened, and if they are halved it will appear steeper, but for both curves the actual slope (as defined by the ratio of the vertical change for a given horizontal change) will be the same. The slope is

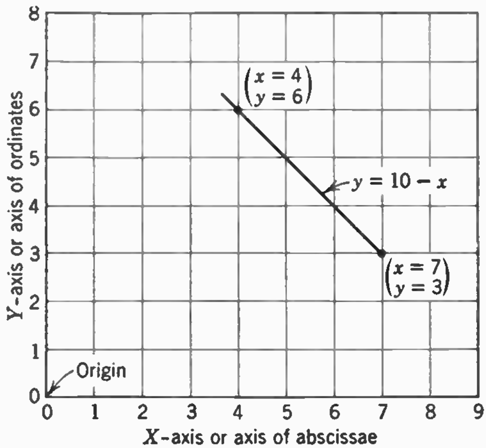


Fig. 2. Typical curve showing how one quantity, y, decreases as another, x, increases.

calculated by dividing a given *change* in vertical units by the *change* in horizontal units which caused this change, as shown in Fig. 3.

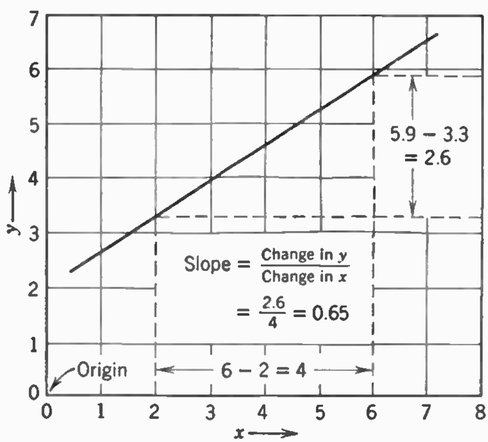


Fig. 3. Here y increases when x increases, and the rate at which y increases is indicated by the slope of the curve.

Problem 8. Calculate the slope of the curve of Fig. 2. Where will this curve cross the X-axis? the Y-axis?

Problem 9. A radio power supply has a terminal voltage of 170 volts when 10 ma of current is drawn from it. Other values of voltage and current are as follows: 110 volts at 30 ma; 80 volts at 40 ma. Plot a curve showing these relations. Determine the change of voltage per milliampere change in current. What is the voltage when no current flows? Note that for this curve the slope is negative since an increase in one variable causes a decrease of the other.

1.9. Symbols. In technical literature a number of symbols are used to represent parts of circuits. Some of the more common symbols are shown on the following pages. The symbols shown are those adopted as standard by the American Standards Association. The weight of the lines of all symbols is the same. Many other symbols have been used to represent various electrical circuit elements. Some of the most important are shown. These are marked with an asterisk (*). It is highly recommended that standard symbols be used by the student.

An electrical circuit is built up by connecting together several of these symbols in various combinations of basic series and parallel connections. In these figures each of the rectangular boxes represents one symbol. A series connection is one in which one electrical element follows another as in Fig. 4. A parallel or shunt connection is one in which the two elements are arranged side by side as in Fig. 5.

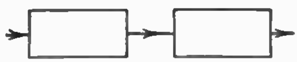


Fig. 4. Simple series circuit.

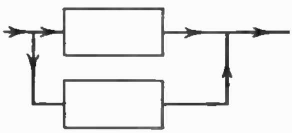
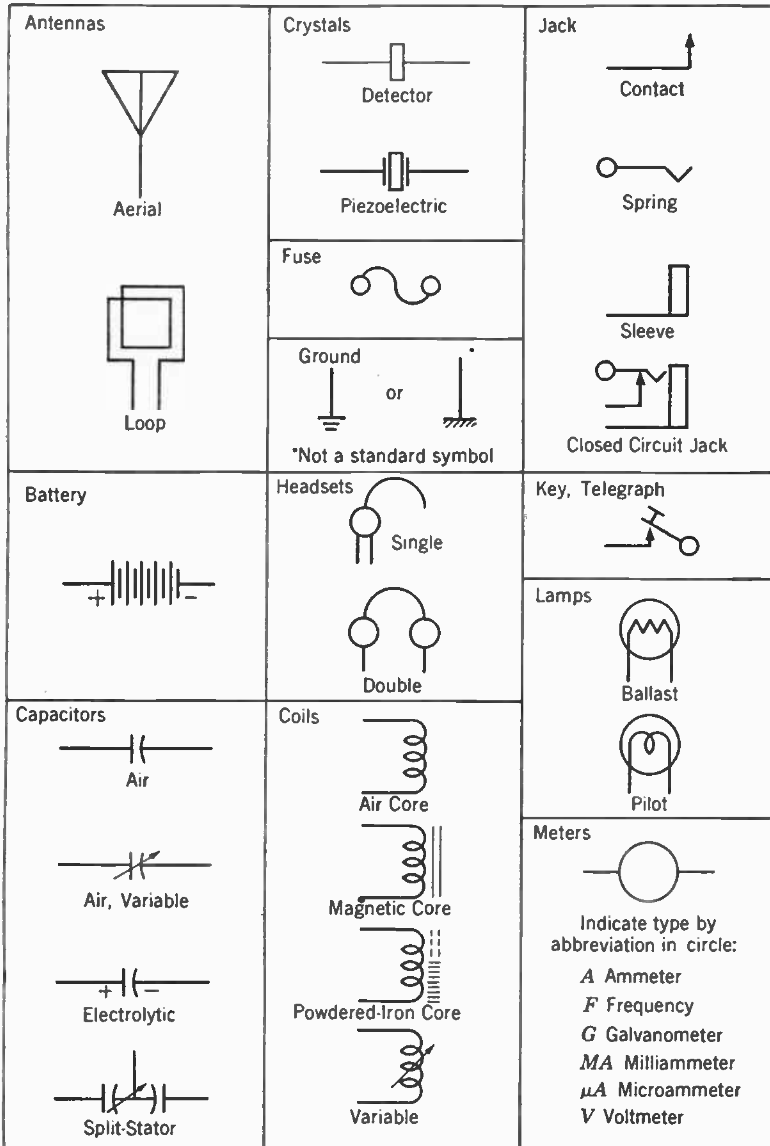
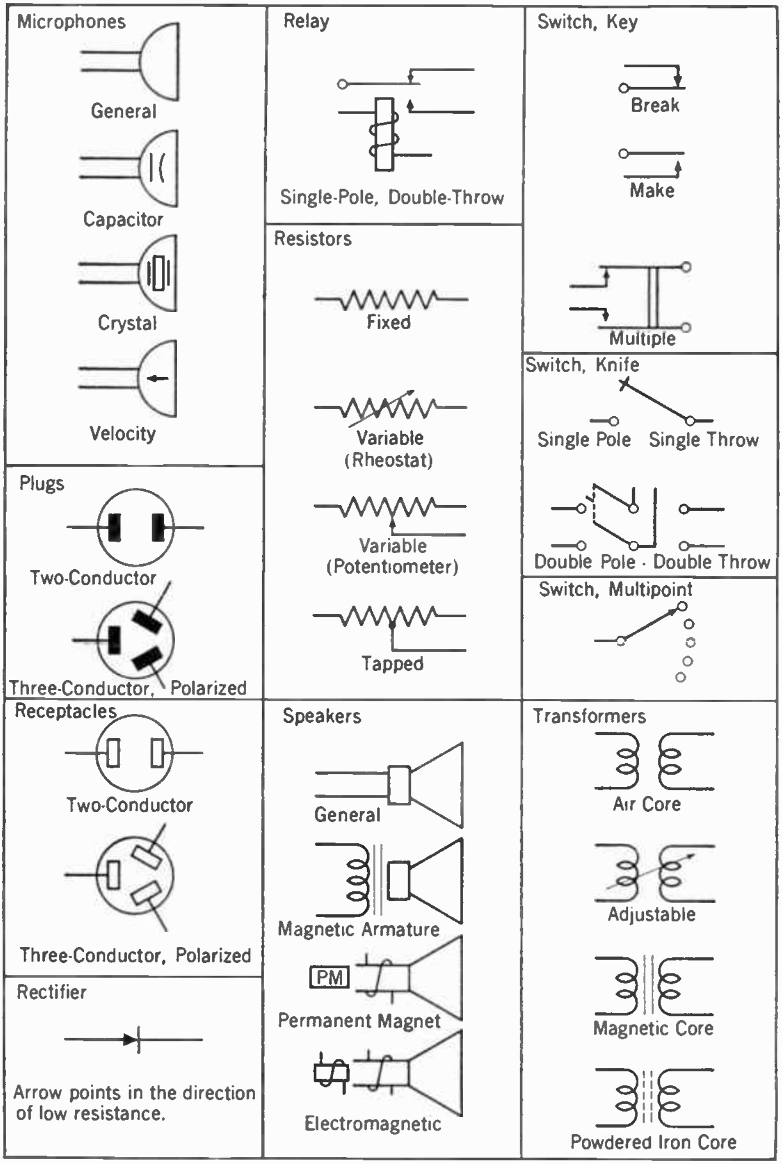
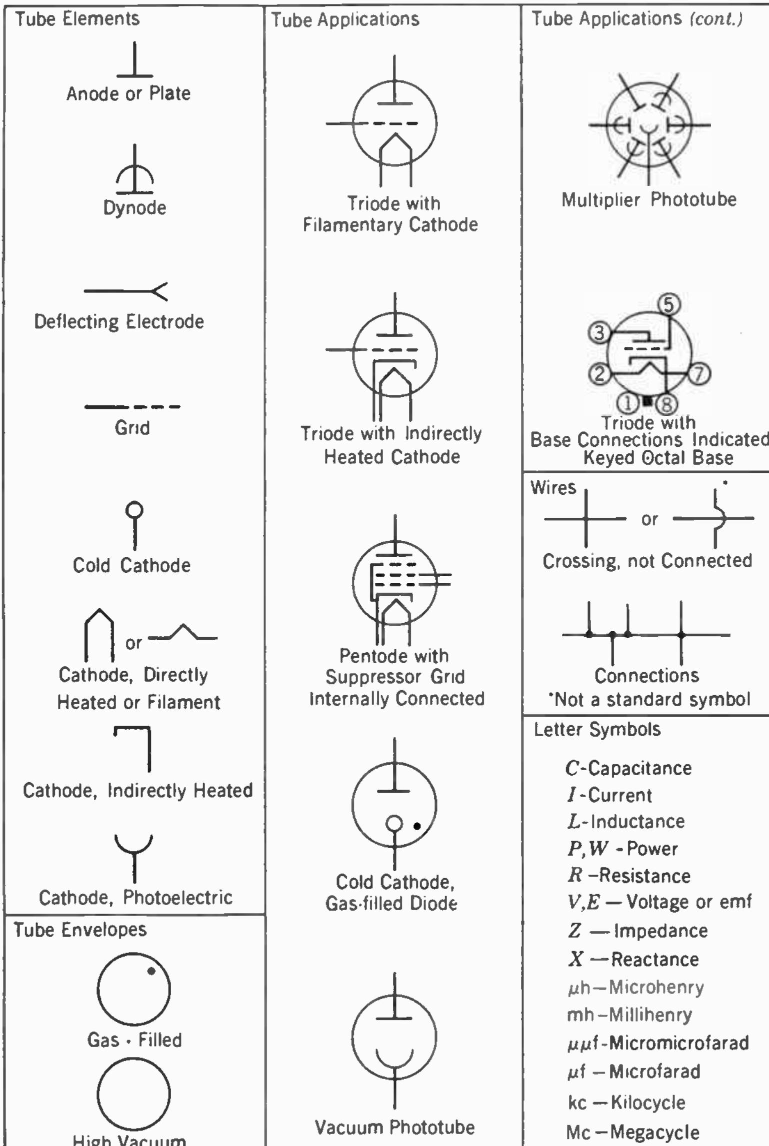


Fig. 5. A parallel or shunt circuit.

Problem 10. Draw a diagram showing an air core coil in parallel with a variable air condenser; a fuse in series with a rheostat and a milliammeter.


Antennas
Aerial
Loop
Crystals
Detector
Piezoelectric
Fuse
Ground
or
*Not a standard symbol
Jack
Contact
Spring
Sleeve
Closed Circuit Jack
Battery
+
−
Headsets
Single
Double
Key, Telegraph
Lamps
Ballast
Pilot
Capacitors
Air
Air, Variable
+
−
Electrolytic
Split-Stator
Coils
Air Core
Magnetic Core
Powdered-Iron Core
Variable
Meters
Indicate type by abbreviation in circle:
A Ammeter
F Frequency
G Galvanometer
MA Milliammeter
μA Microammeter
V Voltmeter


Microphones
General
Capacitor
Crystal
Velocity
Relay
Single-Pole, Double-Throw
Switch, Key
Break
Make
Multiple
Resistors
Fixed
Variable
(Rheostat)
Variable
(Potentiometer)
Tapped
Switch, Knife
Single Pole Single Throw
Double Pole · Double Throw
Switch, Multipoint
Plugs
Two-Conductor
Three-Conductor, Polarized
Receptacles
Two-Conductor
Three-Conductor, Polarized
Rectifier
Arrow points in the direction of low resistance.
Speakers
General
Magnetic Armature
PM
Permanent Magnet
Electromagnetic
Transformers
Air Core
Adjustable
Magnetic Core
Powdered Iron Core


Tube Elements
Anode or Plate
Dynode
Deflecting Electrode
Grid
Cold Cathode
or
Cathode, Directly Heated or Filament
Cathode, Indirectly Heated
Cathode, Photoelectric
Tube Envelopes
Gas - Filled
High Vacuum
Tube Applications
Triode with Filamentary Cathode
Triode with Indirectly Heated Cathode
Pentode with Suppressor Grid Internally Connected
Cold Cathode, Gas-filled Diode
Vacuum Phototube
Tube Applications (cont.)
Multiplier Phototube
5
3
2
7
1
8
Triode with Base Connections Indicated Keyed Octal Base
Wires
or
Crossing, not Connected
Connections
*Not a standard symbol
Letter Symbols
C-Capacitance
I-Current
L-Inductance
P, W - Power
R –Resistance
V,E — Voltage or emf
Z — Impedance
X — Reactance
μh – Microhenry
mh – Millihenry
μμf - Micromicrofarad
μf – Microfarad
kc – Kilocycle
Mc – Megacycle

2 · Direct-Current Circuits

2.1. Direction of current flow. In the previous chapter it was stated that an electric current is a motion of electrons. Now two questions naturally arise: What makes the electrons move, and in what direction does the current flow? Let us answer the second question first.

Long before scientists knew anything about the electron the convention was established that current flowed from the positive terminal of a battery (or other source), through the external circuit, and back into the negative terminal. Now it is known that the electrons move in the opposite direction. That is, electrons move from a more negative point toward a less negative point. Thus the plus-to-minus direction of current flow is merely a manner of speaking—the electrons, which carry most or all of the charge which make up the current, move in the opposite direction. The direction of current flow is now so firmly established that it would be difficult to change the convention. The fact that current flow and electron motion are in opposite directions should cause little trouble once the convention is understood.

It should be remembered that in an electrical circuit the point *from* which the electrons move is always negative with respect to the point *toward* which they move.

2.2. Electromotive force. What makes the electrons move from one place to another? If, for example, a battery is connected to a circuit containing wires and other apparatus, electrons are driven from the negative to the positive terminals of the battery through this external circuit. The force which drives these electrons is the **electromotive force** (emf) of the battery. The size of the force which determines the amount of electron flow depends on the construction of the battery. The unit of emf is the **volt.*** An instrument used to measure voltage is known as a **voltmeter.** Table 1 shows voltages of commonly used electrical sources.

TABLE 1

Apparatus	Voltage (approximate), Volts
Dry cell	1.5
Storage battery	6
Radio B battery	45
House-lighting circuit	117
"Third rail"	500

2.3. Sources of emf. The oldest source of man-made voltage is friction. Anyone who has rubbed a cat's fur on a cold winter day (or who has worn silk clothes) will remember the crackling noise and the tendency of the fur to stand up and follow the hand. A pocket comb rubbed on the coat will pick up bits of paper and other, similar light objects. These phenomena are manifestations of frictional electricity which causes electrical charges to be removed from, or added to, a substance, thus causing it to become electrically charged. Frictional electricity is not a very reliable source of emf since a very small amount of moisture prevents the continued production or storage of the charges.

A common source of voltage is the dry cell used to operate a flashlight or door bell; the lead-sulfuric acid storage battery such as is used in automobiles is another common voltage source.

The generator is a device for converting mechanical energy into electrical energy. It is used to produce very large amounts of electrical energy; its principles will be described later. Another source of emf is the barrier type of photoelectric cell. In this device, light shines on a surface coated with material which emits electrons under stimulation of the illumination. This type of photocell converts light energy to electrical energy.

2.4. Resistance. The next question which arises is: How much current flows in a circuit when it is connected to an electrical source of a known emf?

The electrons in their motion through a conductor are not un-

* A **volt** is the emf (voltage) required to force 1 **ampere** of current through a resistance of 1 **ohm.**

impeded. They constantly run into atoms and other electrons. Since there are no perfect conductors, all materials are said to have a certain **resistance.** This resistance is a measure of the difficulty electrons have in moving freely about among the atoms making up the material. Metals have less resistance than insulators; they are better conductors. Some metals have lower resistance than others. In general, pure metals have lower resistance than alloys.

Copper, for example, has low resistance, whereas some of the combinations of copper, nickel, and iron-manganese have resistances may times that of copper. The fact that copper has a low resistance and at the same time is plentiful explains why most conductors are made of this element. Silver has still lower resistance than copper, but is not so plentiful.

2.5. Factors that govern resistance. The comparative resistances of two wires of the same material and at the same temperature depend upon the length of the wires and the area of their cross sections. The longer the wire the less will be the current that passes through it if a fixed voltage is impressed across its ends; similarly, the smaller the diameter of a wire the greater the resistance (Fig. 1). More gallons of water per second flow from a 3-in. fire hose than from a 1-in. garden hose, although they may be attached to the same hydrant.

A wire 2 ft long has twice the resistance of a wire 1 ft long but of the same diameter. Of two wires the same length, the one having the smaller diameter will have the greater resistance. The resistance is inversely proportional * to the area of the cross section or to the square of the wire diameter. The copper wire table on p. 19 shows that a No. 10 wire has a diameter of 102 mils and a resistance of approximately 1 ohm per 1000 ft, whereas No. 16 wire, with one-half the diameter, has four times the resistance of the No. 10 wire.

The absolute value of the resistivity of a conductor may be indicated in several ways. The most useful to electrical engineers, since they use so much of their resistance material in the form of wires, is the **ohm per mil-foot.** This is the resistance in ohms

* "Inversely proportional" means that one quantity grows larger as another becomes smaller. This inverse relationship is seen in Fig. 5. As resistance R is increased, the current I decreases. The equation is $I = E/R$, where the voltage E is held constant.

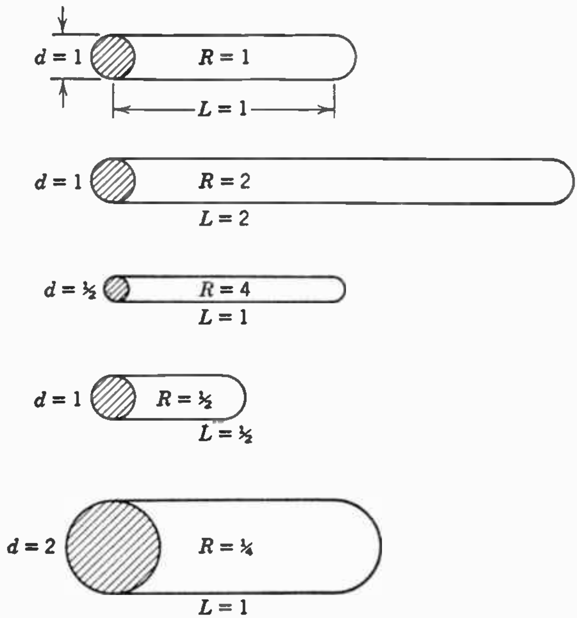


Fig. 1. Resistance depends upon the length and diameter of a conductor.

of a wire 1 mil in diameter and 1 ft long as shown in Fig. 2. A mil is a thousandth of an inch (0.001 in.). A circular mil is a unit of area. A wire having a diameter of d mils will have an area of d^2 circular mils. A mil-foot of copper (1 circular mil in cross section and 1 ft long) will have a resistance of 10.4 ohms. The resistance of any copper wire, therefore, will be $10.4L \div A$ ohms or $10.4L \div d^2$ ohms, where A is the area in circular mils, L is length in feet, and d is diameter in mils.

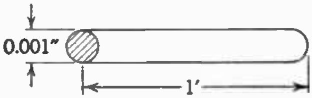


Fig. 2. A mil-foot is a wire one mil (one-thousandth inch) in diameter and one foot long.

The resistivities of several metals compared to silver are as follows:

Silver	1.00	Pure iron	6.28
Copper	1.08	German silver	20.6
Aluminum	1.77	Constantan	30.6
Nickel	4.87	Mercury	60
Platinum	6.25	Carbon, 0° C	220

Problem 1. How many times higher in resistivity is German silver than copper? than iron?

Problem 2. An iron rod is 8 in. long and 0.05 in. in diameter. What is its resistance?

Problem 3. Two wires have resistances in the ratio of 6.28 to 1. The wires are of the same diameter and length. If the lower-resistance wire is made of nickel, what is the material of the other wire?

Problem 4. Two wires have the same length. One wire is aluminum and has a diameter of 0.02 in. The other wire is constantan. What must be the diameter of the constantan wire if its resistance is the same as that of the aluminum wire?

2.6. The ohm. The unit of resistance is the **ohm.*** A 9.69-ft length of No. 30 copper wire has a resistance of about 1 ohm. Table 2 gives sizes and resistance per 1000 ft of copper wire. The resistance per foot may be obtained from such a table by dividing the resistance per 1000 ft by 1000. In this table "Turns per linear inch" is the number of turns of wire that will lie side by side per inch of winding space when the wire is covered with different insulations. "Scc" means single cotton covered; "Dcc" means double cotton covered, indicating that two layers of cotton thread are wound about the wire as insulation. Similarly "Ssc" refers to silk thread insulation, and "Enam" refers to enameled wire.

Note that decreasing the size of the wire by three numbers (diameter decreases as gage number becomes larger) doubles the resistance (approximately), and increasing the size of the wire by three numbers halves the resistance. For example, increasing the wire size from No. 20 to No. 23 increases the resistance from 10.15 to 20.36 ohms per 1000 ft; going from No. 30 to No. 27 lowers the resistance from 103.2 to 51.5 ohms per 1000 ft.

Problem 5. What is the resistance of 1 ft of No. 40 copper wire? of No. 20 aluminum wire? of No. 14 silver wire?

Problem 6. An antenna is made of No. 14 copper wire. It is 75 ft long. How much does it weigh? What is its resistance?

Problem 7. What size (B. & S. gage) of aluminum wire will have approximately the same resistance per foot as No. 30 copper wire?

Problem 8. A 60-cycle power line is 20 miles long. It consists of three No. 0000 copper conductors. What is the resistance of each conductor? If the power line is supported by poles spaced 200 ft apart, how much weight must each pole support?

Problem 9. A coil form is 2¼ in. long and has a diameter of 1 in. How many turns of No. 30 Scc copper wire can be wound on the coil form for a single-layer coil? What will be the resistance of the coil?

* The resistance unit of a **megohm** is frequently used. It is equal to 1 million ohms.

TABLE 2

Copper Wire Table
Resistance at 68° F (20° C)
Mil = 0.001 in.

Size of Wire, B. & S. Gage	Diameter of Wire, Mils	Ohms per 1000 Ft	Pounds per 1000 Ft	Turns per Linear Inch *				
				Scc	Dcc	Ssc	Dsc	Enam
0000	460	0.0490	640.5	2.14	2.10			
000	409.6	0.0618	508.0	2.39				
00	364.8	0.0779	402.8	2.68	2.62			
0	325	0.0983	319.5	3.00				
1	289.3	0.1239	253.3	3.33	3.25			
2	257.6	0.1563	200.9	3.75				
3	229.4	0.1970	159.3	4.18	4.03			
4	204.3	0.2485	126.4	4.67				
5	181.9	0.3133	100.2	5.21	5.00			
6	162	0.3951	79.5	5.88				
7	144.3	0.4982	63.0	6.54	6.25			
8	128.5	0.6282	49.98	7.35				7.6
9	114.4	0.7921	39.63	8.26	7.87			8.6
10	101.9	0.9989	31.43	9.25				9.6
11	90.7	1.260	24.92	10.3	9.80			10.7
12	80.8	1.588	19.77	11.5				12.0
13	72	2.003	15.68	12.8	12.2			13.5
14	64.1	2.525	12.43	14.3				15.0
15	57.1	3.184	9.858	15.9	14.9			16.8
16	50.8	4.016	7.818	17.9	16.7	18.9	18.3	18.9
17	45.3	5.064	6.200	20.0				21.2
18	40.3	6.385	4.917	22.2	20.4	23.6	22.7	23.6
19	35.9	8.051	3.899	24.4				26.4
20	32	10.15	3.092	27.0	24.4	29.4	28.0	29.4
21	28.5	12.80	2.452	29.9				33.1
22	25.3	16.14	1.945	33.9	30.0	36.6	34.4	37.0
23	22.6	20.36	1.542	37.6				41.3
24	20.1	25.67	1.223	41.5	35.6	45.3	41.8	46.3
25	17.9	32.37	0.97	45.7				51.7
26	15.9	40.81	0.769	50.2	41.8	55.9	50.8	58.0
27	14.2	51.47	0.610	55.0				64.9
28	12.6	64.90	0.484	60.2	48.6	68.5	61.0	72.7
29	11.3	81.83	0.384	65.4				81.6
30	10.0	103.2	0.304	71.4	55.6	83.3	72.5	90.5
31	8.9	130.1	0.241	77.5				101
32	8.0	164.1	0.191	83.4	62.9	101	84.8	113
33	7.1	206.9	0.152	90.0				127
34	6.3	260.9	0.120	97.1	70.0	121	99.0	143
35	5.6	329.0	0.0954	104				158
36	5.0	414.8	0.0757	111	77.0	143	114	175
37	4.5	523.1	0.0600	118				198
38	4.0	659.6	0.0476	125	83.3	167	128	224
39	3.5	831.8	0.0377	135				248
40	3.1	1049	0.0299	141	90.9	196	145	282

* The figures given are approximate.

2.7. Conductance. The inverse of resistance is **conductance.** Conductance expresses the property of a substance to pass an electric current, just as resistance expresses the ability of a substance to interfere with the passage of a current. The unit of conductance is the **mho.** Note that mho is "ohm" spelled backward. If the resistance is known, the conductance can be found by dividing the resistance into 1. Thus

$$\text{Conductance} = \frac{1}{\text{Resistance}}$$

2.8. The effect of temperature on resistance. The resistance of all pure metals rises with an increase of temperature because of the greater molecular agitation at higher temperatures, making it more difficult for the electrons to drift around the circuit. In certain alloys and so-called semiconductors, effects in addition to thermal agitation may cause the resistance to go either up or down as temperature increases. This is illustrated in Fig. 3.

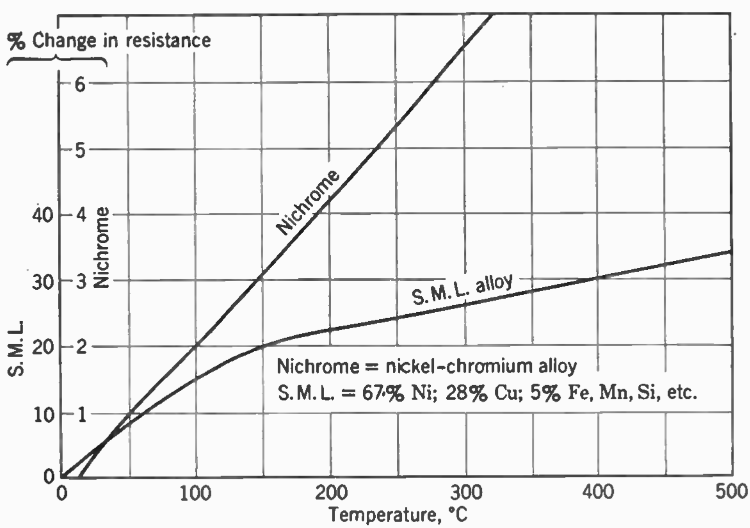


FIG. 3. How temperature affects the resistance of certain alloys.

At absolute zero, 273° below 0° centigrade, all molecular motion is supposed to stop, making the resistance of a metal practically zero. Scientists have approached to within a fraction of a degree of absolute zero.

2.9. Temperature coefficient of resistance. The amount the resistance of pure metals increases for each degree rise in temperature for each ohm at the original temperature is known as the **temperature coefficient of resistance.** This figure lies between 0.003 and 0.006 for pure metals, being 0.00393 for copper. This is the reason that the resistances of wires in wire tables are indicated as being at a given temperature; the value chosen is 20° C.

The resistance at any temperature when the resistance at a known temperature is available and when the temperature coefficient is known may be found from the formula

$$R_{t_2} = R_{t_1}[1 + \alpha(t_2 - t_1)]$$

where R_{t_1} and R_{t_2} are the conductor resistances at temperatures t_1 and t_2, and α is the temperature coefficient.

Example 1. A copper wire with a temperature coefficient of 0.00393 has a resistance of 80 ohms at 0° C. What will be the resistance at 50° C? The resistance will be increased by 80×0.00393 for each degree rise in temperature. At 50° C the resistance increase will be $80 \times 0.00393 \times 50$ or 15.72 ohms, and the resistance will then be $80 + 15.72$ ohms or 95.72 ohms. Using the formula, this works out as follows:

$$\begin{aligned} R_{t_2} &= 80[1 + 0.00393\,(50 - 0)] \\ &= 80(1 + 0.1965) \\ &= 80(1.1965) \\ &= 95.72 \text{ ohms} \end{aligned}$$

Typical temperature coefficients of resistance at 20° C for several metals and alloys used in radio apparatus are given in Table 3.

TABLE 3

Material	Coefficient
Constantan (an alloy)	0.000002
Copper, annealed	0.00393
Copper-manganese-iron	0.00012
Iron	0.0050
Nickel	0.006
Platinum	0.003
Silver	0.0038
Tantalum	0.0031

Problem 10. The field coil of a d-c motor has a "cold" resistance of 70 ohms. The "cold" resistance is measured at 20° C. If the temperature of the winding increases to 70° C when the motor is running under full load, what is the coil resistance at the operating temperature?

Problem 11. A copper shunt to be placed across a current meter has a resistance of 0.400 ohm at room temperature (20° C). If the meter is taken outdoors when the temperature is freezing (0° C), what is the resistance of the shunt? What would be the outdoor resistance of the shunt if it were made of constantan?

Problem 12. A coil of copper wire is immersed in the cooling oil of a transformer. The wire has a measured resistance of 2 ohms at room temperature (20° C). After the transformer has been loaded for several hours the resistance of the wire is found to be 2.02 ohms. What is the temperature of the transformer oil?

2.10. Ohm's law. One of the laws which govern all simple and many complex electrical phenomena is known as **Ohm's law.** This law states: Current in amperes equals emf in volts divided by resistance in ohms,* or, in electrical abbreviations,

$$I \text{ (current)} = \frac{E \text{ (voltage)}}{R \text{ (resistance)}}$$

2.11. Ways of stating Ohm's law. There are three ways of stating this fundamental law.

$$[1] \quad I = E \div R \qquad [2] \quad E = I \times R \qquad [3] \quad R = E \div I$$

These three ways of stating the same law, determined from the first statement of Ohm's law by simple mathematical transformation, make problem solving less difficult.

From these three expressions of Ohm's law, any one of the quantities can be obtained if the other two are known. Thus, from equation 1, the current in a circuit can be determined if the voltage and resistance of the circuit are known. From 2 the voltage required to force a desired current through a given resistance can be determined. Finally, from 3 the resistance of a circuit can be found if the current flowing in it under the force of a known voltage is measured. Many circuits and apparatus follow much more complex laws than Ohm's law.

Curves showing how current varies with voltage, with resistance, and with conductance are seen in Figs. 4, 5, and 6.

* In a-c circuits "resistance" is replaced by "impedance."

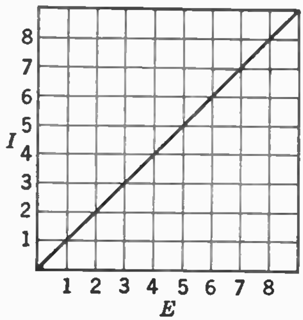


FIG. 4. In a circuit obeying Ohm's law, the current is directly proportional to the voltage.

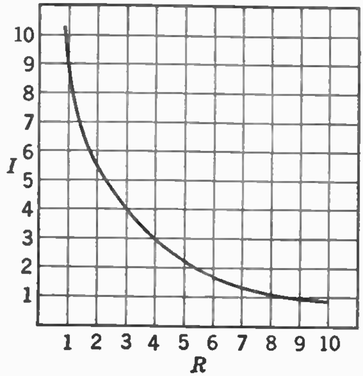


FIG. 5. Current is inversely proportional to the resistance, decreasing as the resistance increases.

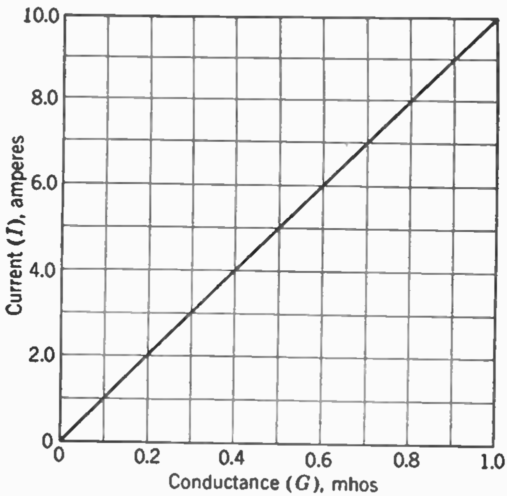


FIG. 6. The greater the conductance (lower resistance) the greater will be the current, as this curve shows.

Example 2. A radio-tube filament is heated by current flowing through it. The current comes from a battery to which the filament is connected. The battery has a voltage of 5 volts, and the resistance of the filament is 20 ohms. How much current will flow through the filament?

Using equation 1, and dividing the voltage, 5, by the resistance, 20 ohms, the current is $5 \div 20$ or 0.25 amp.

Now suppose that the resistance is 20 ohms and that 0.25 amp is required to heat the filament sufficiently to cause it to produce the proper number of electrons. What should be the voltage of a battery connected to the filament?

This time equation 2 is used, and the current, 0.25 amp, is multiplied by the resistance, 20 ohms, to obtain 5 volts.

Finally, suppose it is known from experience that the proper voltage is 5 volts and that with this voltage 0.25 amp will flow through the filament. What is the resistance of the filament? The reader should work this problem for himself, using equation 3.

NOTE. The fundamental units are amperes, volts, and ohms. Volts, milliamperes, and ohms cannot be used without getting into trouble. First, the milliamperes must be converted into amperes and then used in the formulas expressing Ohm's law.

2.12. Voltage drop. The second way of stating Ohm's law indicates that, whenever a current flows through a resistance, a difference of potential (a voltage) exists at the two ends of that resistance. For every ampere of current that flows through an ohm of resistance, a volt is lost. In other words, 1 volt is required to force 1 ampere through 1 ohm of resistance.

The term "voltage drop" is generally used to denote a loss of voltage in some part of a complete network. The sum of all the voltage drops across the series elements of a network must equal the applied voltage. If, in a network composed of three resistances in series, the voltage drops are 5, 10, and 15 volts, and a battery is connected across these three resistors, the voltage of the battery must be $5 + 10 + 15$ or 30 volts. This concept is explained more fully in the following example.

Consider Fig. 7, which shows three resistors * connected in series and placed across a source of voltage. The emf causes

* A **resistor** is a device for adding **resistance** to the circuit. Note that resistance is a property of a material or circuit, whereas a resistor is a device possessing resistance. The terms are, however, often used interchangeably.

current to flow through the three resistors. The current flowing through each resistor produces a voltage drop across each resistor, and the sum of the three voltage drops must equal the total voltage impressed across the entire series circuit. Thus, if 180 volts is impressed across the entire assembly, the voltages that will be measured by a voltmeter at other points in the series circuit are as shown.

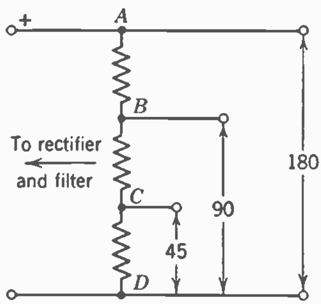


FIG. 7. A series circuit of three resistors making up a "voltage divider" across a power supply device.

Example 3. In Fig. 7 suppose that 10 ma of current is forced through the three resistors in series. What is the resistance of each of the three resistors?

Here, the third form of Ohm's law is used: $R = E \div I$. Since 180 volts appears across the circuit and since 10 ma (0.01 amp) flows through it, the total resistance will be equal to

$$R = E \div I = 180 \div 10 \times 10^{-3} = \frac{180}{10 \times 10^{-3}} = \frac{180}{10^{-2}}$$

$$= 180 \times 10^{2} = 18{,}000 \text{ ohms}$$

Now, half the total voltage drop appears across the top resistor. Therefore, the resistance between A and B is one-half the total resistance or 9000 ohms. and the resistance between B and D is also 9000 ohms. Across the 9000 ohms between B and D there appears a voltage drop of 90 volts, and in the middle of this resistance at C is a voltage drop of 45 volts or half the total voltage of 90 volts. Therefore, the point C is half way between B and D, and the resistance between B and C must be 4500 ohms. The resistance from C to D is also 4500 ohms.

A voltage which is so small that it cannot be measured with available instruments is often needed in laboratory experiments. A higher voltage, however, can be measured easily; and, if it is impressed across a series of resistors like those in Fig. 7 (known as a **voltage divider**), any desired part of the total voltage may be utilized by means of proper taps or connections or by a sliding contact. A **potentiometer** is a resistor with a sliding arm arranged so that resistances between zero and the maximum value may be obtained.

The voltage appearing across a resistance when a current flows through that resistance is known as a **voltage drop** or **IR drop.** It may be calculated by multiplying the resistance (R) in ohms by the current (I) in amperes.

Problem 13. An electric iron with a resistance of 15 ohms is plugged into a 120-volt lighting system. How much current will flow?

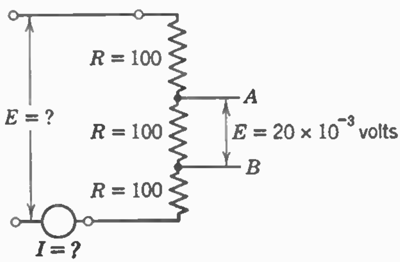


FIG. 8. A problem in Ohm's law. What current must flow through the circuit and what will be the total voltage?

Problem 14. The heater of a 6J5 tube is rated at 6.3 volts and 0.3 amp. What is its resistance?

Problem 15. Consider Fig. 8. How many milliamperes of current must be forced through the circuit to get 20 mv across the resistor A–B? How many volts in all will be needed?

Problem 16. A radio power supply furnishes 250 volts at a maximum current of 125 ma. What is the smallest resistance that should be connected across its terminals without exceeding the current rating?

2.13. Energy and power. A battery converts chemical energy into electrical energy; a generator transforms mechanical energy into electrical energy. What is meant by energy? What is power? These two terms are used rather loosely by most people, but each has a very definite meaning.

Energy is the *ability to do work.* A body may have one of two kinds of mechanical energy, either **potential** or **kinetic,** or it may have both. Potential energy is due to the position of the body; kinetic energy is due to its motion. A heavy ball on top of a flagpole has potential energy because if it falls it can do work, useful or not. It may heat the ground where it falls, or it may be used to drive a post into the ground. A cannon ball speeding through the air has energy because it can do work, useful or otherwise, if it is stopped suddenly. The target may be heated thereby, the kinetic energy possessed by the ball being converted into heat energy. The amount of damage done is a rough measure of the energy possessed by the cannon ball. This energy was originally possessed by the powder and was imparted to the ball when the powder exploded.

Whereas energy is the ability to do work, **power** is the *rate of doing work.* The horsepower, for example, is a unit of mechanical power. It is the power required to raise 33,000 lb of material 1 ft in 1 min; or 1 hp is 33,000 ft-lb per min.

All expressions for power involve the factor of time. It requires more power to accomplish a certain amount of work in a short time than in a long time. For example, a ton of material raised a foot in the air represents 2000 ft-lb of work (energy). If the job is accomplished by a crane in 1 sec of time it represents an expenditure of 2000×60 or 120,000 ft-lb per min of power. Since 1 hp is equal to 33,000 ft-lb per min, the crane has a power of $120,000 \div 33,000$ or about 3.65 hp.

Now, if a man raises the ton of material 1 ft in the air in an hour's time by pulling it up a long gradual incline with a rope and pulley, his power is $2000 \div 60$ or 33.3 ft-lb per min, or roughly one-thousandth horsepower (0.001 hp). The amount of work done in the two cases is the same—the ton of material has been raised 1 ft in the air. The *rate of doing work* has changed.

Since power is the rate of doing work, the amount of work done in a given time is the rate of doing work multiplied by the time. Thus if 1 lb is raised 1 ft per hr and the work goes on for 2 hr, 2 ft-lb of work has been done. The same amount of work would be done if 1 lb were raised 1 ft per min and if the work went on for 2 min. In this case, however, the work would have been accomplished at a faster rate, requiring more power.

Three quantities are involved: energy, which is the ability to do work; power, which is the rate of doing work; and the work done. Energy and work are rated in the same units, horsepower-hours, for example, or kilowatt-hours in electrical circuits.

In an electrical circuit, power is the product of the voltage E and the current I because, in an electrical circuit with a certain resistance, an increase in voltage (force) is required to cause an increase in current (amperes of current are a measure of "rate of flow of electrons"). If the resistance is 1 ohm, and 1 volt is impressed across it, 1 amp of current will flow. This corresponds to a flow of 6.28 million million million electrons per second. If the voltage is raised to 2 volts, the current increases to 2 amp, and the rate of flow of electrons is likewise doubled. Not only has the rate of flow of electrons (their speed) been doubled; in addition, twice as many per second are passed through the resistance.

To obtain this result, both voltage and current have been doubled; the power required is four times larger.

The unit of electrical power is the **watt.** Thus

$$\text{Power in watts} = \text{Current in amperes} \times \text{Emf in volts}$$

or

$$P = E \times I$$

A **kilowatt** is 1000 watts. The kilowatt-hour is a measure of energy or work; a smaller unit is the watt-second or **joule,** and a still smaller one used in scientific work is the **erg.** It takes 10 million ergs to make 1 watt-second.

A homeowner usually pays for electrical service in terms of how much energy or how many kilowatt-hours he uses. Except in special cases, the power he uses at any given time does not enter in the calculation of his monthly bill. A 100-watt lamp, for example, may be burned for 10 hr for the same cost as a 1000-watt electric iron heated for 1 hr. In each case 1 kilowatt-hour of electrical energy is consumed.

2.14. Power lost in resistance. According to the law called conservation of energy, *energy* can be neither created nor destroyed. It comes from somewhere and goes somewhere. Similarly, all *power,* which is the rate at which energy is used, must be accounted for. The energy required to force current through a resistor must do some work. It cannot disappear. This work results in heating the resistor, and the electrical energy is converted to heat energy. The heat appears because of the greater molecular activity which results from the flow of electrons through the material. Whenever current flows through a resistor, heat is generated, and the greater the current the greater the heat. As a matter of fact, the heat is proportional to the square of the current. If the resistor is heated faster than the heat can be dissipated by heating the surrounding air, the resistor may be damaged or destroyed. Energy has been supplied to the unit at too great a rate.

A resistor used in a radio circuit is often rated at so many ohms and as capable of dissipating so many watts. Thus a 1000-ohm, 20-watt resistor means that the resistance of the unit is 1000 ohms and that 20 watts of electrical power can be put into it without danger of burn-out.

2.15. Expressions for power. Just as there are three ways of stating Ohm's law, so there are three ways of stating the relation between watts, volts, amperes, and ohms. Thus

$$[1]\quad P = I \times E \qquad [2]\quad P = I^2 \times R \qquad [3]\quad P = E^2 \div R$$

A useful expression is $I = \sqrt{P/R}$. It may be employed in calculating the current safely passed by a resistor of a given wattage and resistance rating.

Example 4. A plate voltage power supply system supplies 180 volts to a power tube which takes 20 ma. How much power is taken? What is the resistance of the power tube?

The power supplied is $E \times I = 180 \times 0.02 = 3.6$ watts. The resistance into which this power is fed is equal to $P \div I^2 = 3.6 \div 0.0004 = 9000$ ohms, or $E^2 \div P = 32{,}400 \div 3.6 = 9000$ ohms.

The maximum current that can pass through one's body without serious results is approximately 0.01 amp. The resistance of the body varies with one's health, the surface of contact, perspiration, and similar factors. If the finger tips of the two hands are dry, the resistance from one hand to the other is about 50,000 ohms.

Problem 17. What is the "hot" resistance of a 200-watt, 115-volt lamp? of a 50-watt, 115-volt lamp? NOTE. The "hot" resistance is the resistance after the lamp has heated to normal operating temperature. The "cold" resistance is much lower.

Problem 18. The voltage drop across a resistor in an amplifier circuit must be 10 volts when 50 ma of current flows through it. What must be the size of the resistance? its power rating?

Problem 19. A 10,000-ohm resistor is connected across the terminals of a 45-volt B battery. How much power is dissipated in the resistor? What current flows?

Problem 20. How much voltage may be placed across a 10,000-ohm, 1-watt carbon resistor without danger of burning it up? Compute the voltage if the resistor is rated at 10,000 ohms, 5 watts.

Problem 21. A radio receiver consumes 50 watts on a 110-volt circuit. If the electric power rate is 6 cents per kilowatt-hour, how much will it cost to run the receiver for 10 hr?

Problem 22. An electric stove draws 20 amp on a 220-volt circuit. How much will it cost to operate the stove for 2 hr if the electric rate is 5 cents per kilowatt-hour?

Problem 23. An automobile receiver utilizes six tubes which require 0.3 amp at 6.3 volts. In addition, the plate circuits of the tubes require 60 ma at 150 volts. How much total power must a 6.3-volt storage battery supply?

Problem 24. The power supply system for a speech amplifier requires a 7500-ohm voltage-divider resistor which can pass a maximum current of 50 ma. What is the maximum voltage across this resistor? What must be its power rating?

Problem 25. The heating element of an electric heater is to be made of resistance wire which has a resistance of 0.5 ohm per ft. How many feet of wire will be required in a heater which is to consume 1000 watts on a 120-volt circuit?

Problem 26. What is the approximate current a ¼-hp motor will draw from a 110-volt line when running at full load? (1 hp = 746 watts.)

2.16. Efficiency. Efficiency is a term that is loosely employed by nearly everybody. Anything which works is said to be efficient, and one's efficiency is often confused with his energy—his ability to do work whether the work is actually carried out or not. The term "efficiency," however, has a very exact meaning when it is used in connection with mechanical or electrical systems of any kind. It is a ratio of the *useful work* done by a device to the *total energy* put into the device.

Consider a steam engine connected to a dynamo, a combination of machines for transforming mechanical energy into electrical energy. If the steam engine consumes 1 hp (746 watts) and delivers 500 watts of electrical energy, it is said to be more efficient than if it delivered only 250 watts. As a second example, consider two men, one of whom gets a lot of work done in a small amount of time and with an expenditure of little effort. The other gets the same amount of work done but with great effort, perhaps flurrying about from one thing to another instead of tackling his problem in a straightforward manner. The first man is more efficient. He *wastes* less time and energy.

Efficiency, then, is the ratio between useful work or energy got out of a machine and total energy or work put into the machine. Efficiency may also be expressed as the ratio of useful power output to total power input. Efficiency is expressed in percentage. A machine that is 100 per cent efficient has no losses; there is no friction in its bearings, or, if it is an electrical device, there is no resistance in its wires. Actually, no machine is 100 per cent efficient. Some losses are always present.

$$\text{Efficiency} = \frac{\text{Useful output}}{\text{Input}} \times 100\% = \frac{\text{Useful output}}{\text{Output plus losses}} \times 100\%$$

Problem 27. A public-address amplifier supplies 50 watts of power at an efficiency of 10 per cent. What is the power taken from a 120-volt line? the current?

Problem 28. A transformer furnishes 10 kw of power for a group of motors. If there is a 5 per cent power loss in the wiring, how much power is actually supplied to the motors? If the motors are 90 per cent efficient in the use of the power supplied to them, how much useful power can they deliver?

Problem 29. A generator can dissipate a maximum power in its windings of 200 watts without becoming overheated. If it is 95 per cent efficient, what is the maximum power output?

Problem 30. A vacuum-tube amplifier operates at 25 per cent efficiency. If the maximum power loss which the amplifier tubes can dissipate is 10 watts, what is the maximum power output which can be obtained? What is the input power for maximum power output?

Problem 31. A 5-mile telephone line is used to carry audio-frequency signals from a studio to a radio transmitter. If the line is 75 per cent efficient, and the transmitter must receive 1 mw of power, how much power must the studio amplifiers feed into the line?

Problem 32. A portable radio transmitter operates from a 6-volt, 200-amp-hr storage battery. The filaments require 30 watts of power and are operated continuously. The other circuits of the transmitter require 24 watts, and the transmitter is keyed in such a manner that power must be supplied to these circuits 60 per cent of the time. How long will the battery, when fully charged, operate the transmitter?

Problem 33. If the power radiated by a radar transmitter is approximately proportional to the fourth power of the received signal, by how much must the transmitter power be increased to double the received power? How much must the radar antenna current be increased?

2.17. Protective devices. If the wattage rating of a resistor is exceeded, that is, if energy is fed into it too rapidly, heat is generated faster than it can be lost and damage will result. In a similar manner, the amount of heat that any electrical device can dissipate without damage has a definite limit.

The simplest means of protecting a device against excessive current is a **fuse.** A fuse is merely a short length of wire which will melt and open the circuit before damage is done to the device it is protecting. Some fuses are constructed so that they "blow" very rapidly if the current exceeds their rated value. Others are constructed so that they will not blow on sudden and temporary overloads caused, for example, by a motor starting from rest. During the fraction of a second that the speed of the motor is increasing, the current required is much greater than that required when the motor is running at normal speed. Therefore a large amount of current will flow for a short time, and the fuse should not blow unless the motor is loaded too heavily or unless, for some other reason, it fails to come up to normal speed promptly. A fuse in series with the motor will melt if the current passed through it is too great for too long a time. This opens the circuit. Someone, then, must replace the fuse before the circuit will function again. A blown fuse is usually an indication of trouble. The trouble should be corrected before the apparatus is operated. In no event should the fuse be "jumpered" since then a heavy overload may cause the wiring to the motor to melt, perhaps causing a fire or other damage.

Replacing a fuse is a nuisance. In a factory a more common device used to protect machinery against overload is a circuit breaker or an overload relay. This is a mechanical switch which opens the circuit automatically when the current becomes too great. It may be arranged to close again automatically after a second or two, or it may be made so that a maintenance man must close the circuit by hand after he has cleared up the cause of the overload.

Sometimes electrical circuits must be provided with protective devices which prevent one voltage from being applied until a given time after a voltage has been applied to another part of the circuit. An example is plate voltage and filament voltage to gaseous rectifier tubes. The filament of the tube must be heated for a specified time before plate voltage is applied. This kind of protection is provided by a **time-delay relay.** After the power switch is closed the filament voltage is applied immediately. After a specified period of time an auxiliary switch is automatically closed which applies the plate voltage. Short time delays are secured with thermal elements which close the auxiliary switch as they heat up. Longer time delays are secured with electric clockwork mechanisms.

2.18. Series circuits. When two or more pieces of equipment are connected as in Fig. 9 they are said to be in **series.** The same current flows through each unit. The voltage drop across each unit is controlled by its resistance, and if one of these units has twice the resistance of the other, the voltage drop across it will be twice as great. The sum of the voltage drops across the three resistances must be equal to the voltage of the battery, for there is no other source of voltage.

In a series circuit the total resistance is the sum of the individual resistances. The current through each unit is the same as in all other units. The current is obtained from Ohm's law, equation 1.

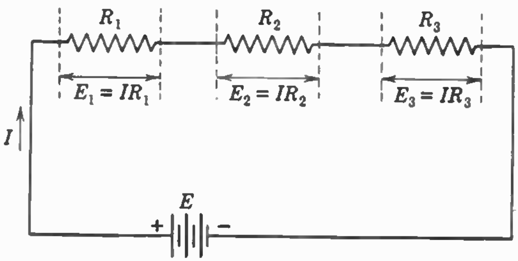


FIG. 9. Series circuit showing the voltage drops along the individual elements.

If any of the units becomes "open" the current ceases to flow. If, however, any unit becomes "shorted" the current will increase because the total resistance of the circuit has decreased.

Example 5. In Fig. 10 is a typical series circuit composed of a vacuum-tube filament R_2, a 6-volt battery, a current meter, and a rheostat or variable resistor whose purpose is to limit the flow of current through the filament of the tube. The arrow through R_1 indicates that it can be adjusted in value.

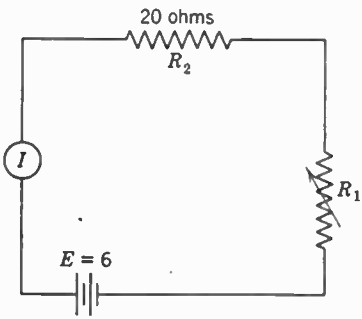


FIG. 10. Here R_2 represents a vacuum-tube filament and R_1 a rheostat for current control.

The question is, What current will flow through the circuit as the resistance of R_1 is varied? Suppose that R_1 is 4 ohms and the resistance of the meter is negligible. The resistance in the circuit is then equal to 20 plus 4 or 24 ohms, and by Ohm's law the current will be the voltage divided by the total resistance, or, using equation 1,

$$I = \frac{E}{R_1 + R_2} = \frac{6}{4 + 20} = \frac{6}{24} = 0.25 \text{ amp}$$

There are two resistances in this circuit. Current flows through them and produces voltage drops across each. What are the values of these two voltage drops? By equation 2

$$\text{Voltage drop} = IR_1 = 0.25 \text{ amp} \times 4 \text{ ohms} = 1 \text{ volt}$$

$$\text{Voltage drop} = IR_2 = 0.25 \text{ amp} \times 20 \text{ ohms} = 5 \text{ volts}$$

In other words, of the 6 volts available at the terminals of the battery, 5 volts have been used up across the 20-ohm resistance of the tube filament, and 1 volt has been used to force 0.25 amp through the 4-ohm resistance of the rheostat.

Problem 34. In an ac-dc radio receiver the filaments of the following tubes are connected in series: 6SA7, 6SK7, 6SQ7, 25L6, and 25Z6. If the filaments are to be operated from a 110-volt line, how much additional resistance must be connected in series with the filaments?

Problem 35. What are the "hot" filament resistances of the following tubes: 1T4, 2A3, 6C4, 6J5, 6L6, 25L6, 35C5? How much power does each require?

Problem 36. How much series resistance would be required to operate each of the tubes of Prob. 35 directly from a 50-volt source?

Problem 37. A fixed 1000-ohm resistor is connected in series with a variable resistor, and the combination is placed across a 45-volt battery. What must be the resistance of the variable resistor to cause a voltage drop of 30 volts across it?

Problem 38. A vacuum tube and a 50,000-ohm resistor are connected in series with a d-c supply. If 2 ma through the tube produces a voltage drop of 100 volts, and this is the condition for proper operation of the tube, what must be the voltage of the d-c supply?

2.19. Parallel circuits. A parallel circuit is represented in Fig. 11. It consists of several branches. The voltage across each branch is the same as that across every other branch and is equal to the voltage of the battery. The total current supplied by the battery is the sum of the currents taken by the branches. The resistance of the group may be found by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

where R is the resultant or total resistance, and R_1, R_2, etc., are the individual resistances.

The resultant resistance of several units in parallel is less than the lowest resistance of any of the components. If two equal resistance are in parallel, the resultant is one-half the resistance of one. This fact follows logically from the following reasoning. If two identical resistors are placed across a given voltage, they pass twice the current passed by either one of them. If, therefore, a new resistor is selected which passes as much current as the two resistors taken together, it will have half the resistance of the two identical resistors. Working this out by Ohm's law will verify the reasoning. Thus if two 10-ohm resistors are connected in parallel, the resultant is 5 ohms. What would it be if they were connected in series?

If any number of *equal* resistances are in parallel, the resultant resistance is the resistance of one unit divided by the number of units.

If two resistors whose resistances are unequal are in parallel, the resultant resistance may be calculated by dividing the *product* of their resistances by the *sum* of their resistances:

$$R = \frac{R_1 \times R_2}{R_1 + R_2}$$

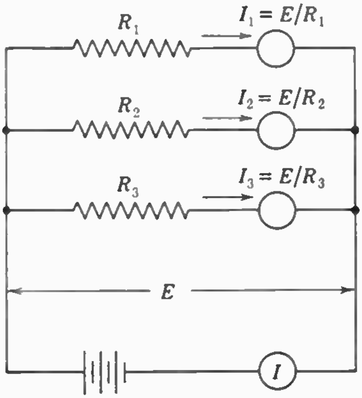


FIG. 11. A parallel circuit with meters to indicate the currents through the individual branches.

This simplified formula comes directly from the one above by application of simple algebra. The reader should check the result.

Example 6. What is the parallel resistance of two units which have resistances of 4 and 5 ohms?

This can be solved by either of the formulas given.

$$\frac{1}{R} = \frac{1}{4} + \frac{1}{5}$$

$$= 0.25 + 0.20$$

$$= 0.45$$

$$R = 1 \div 0.45 = 2.22 \text{ ohms}$$

or

$$R = \frac{R_1 \times R_2}{R_1 + R_2}$$

$$= \frac{4 \times 5}{4 + 5}$$

$$= \tfrac{20}{9} = 2.22 \text{ ohms}$$

Example 7. Suppose that, as in Fig. 12, two resistances in parallel are placed in series with a resistance of 1 ohm and across a battery of 6 volts. What current would flow from the battery and through each resistor?

The total resistance is $2.22 + 1 = 3.22$ ohms. The current flowing, then, is $6 \div 3.22 = 1.86$ amp. This current through the combined resistance of the 4- and 5-ohm resistors produces a voltage drop of $I \times R$ or 1.86×2.22 or 4.14 volts. This voltage across 4 ohms produces a current of $4.14 \div 4$ or 1.035 amp, and across 5 ohms produces a current of 0.828 amp. These two currents added together are 1.863 amp, which checks the calculation above for the total current.

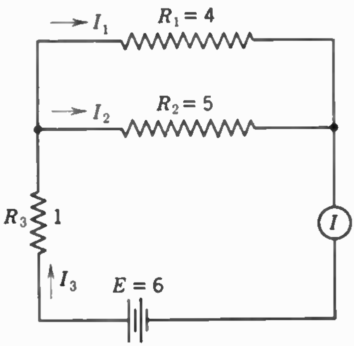


Fig. 12. A problem in series-parallel circuits.

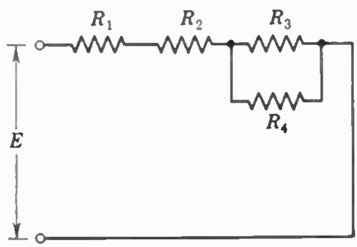


Fig. 13. Another complex circuit problem.

Problem 39. The following tubes are connected in parallel to a 6.3-volt transformer secondary: 6D8-G, 6SK7, 6SQ7, 6K6-GT, and 6X4. How much current is taken from the transformer? What is the combined resistance of the tubes?

Problem 40. A milliammeter requires 10 ma of current for a full-scale indication. Its resistance is 15 ohms. A resistor is connected in parallel with the terminals of the meter. What must be the value of this resistor to cause the meter to read full-scale when 100 ma is forced through the combination?

Problem 41. A circuit has three parallel branches of 3, 4, and 8 ohms. A current of 3 amp flows through the 8-ohm branch. What current flows through each of the other branches?

Problem 42. In Fig. 13 R_1 is the filament of a 6J5 tube, R_2 is the filament of a 25Z6 tube, and R_3 is the filament of a 35C5 tube. Resistor R_4 is adjusted to cause the proper current to flow through R_3. What must be the value of E? of R_4?

2.20. Use of Conductance. The sum of all currents passing through a number of parallel resistors will be the sum of the currents through the individual resistors. The value of the current passed by each resistor is directly proportional to the conductance ($1/R$). Therefore, by adding the conductances of all the resistances, the total conductance ($1/R_0$ or G_0) is obtained. The numerical value of current will be equal to $E \times 1/R_0$ or $E \times G_0$. Thus

$$\frac{1}{R_0} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

But

$$G = \frac{1}{R}$$

$$G_0 = G_1 + G_2 + G_3$$

and

$$R_0 = \frac{1}{G_0}$$

where R_0 and G_0 are the resultant resistance and conductance.

Conductances are useful in problems involving parallel circuits, but, since the conductance of a piece of apparatus is seldom given, conductances, as such, are not often used. The fact that the actual values of conductances are usually very small (if $R = 100$ ohms, $G = 0.01$ mho; if $R = 10{,}000$ ohms, $G = 0.0001$ mho) adds to the difficulty of using conductances.

2.21. Ohm's law in complex circuits. It must be remembered that Ohm's law applies to a circuit as a whole, or to any of its parts. If the voltage and the resistance of any part of a circuit are known, the current through that part can be calculated without regard to any other part of the circuit. In the following section, the solution of complex circuits is described.

2.22. Kirchhoff's laws. When circuits are made up of series and parallel elements, solving for the individual currents or voltages becomes somewhat complicated. All such circuits can be reduced to simpler circuits which can be solved more easily, and this reduction is a result of the direct application of Ohm's law. In some complex circuits, however, it is often easier to solve for the currents and voltages directly than to reduce the complex circuit to a simpler circuit. Two rules known as **Kirchhoff's laws** are used when the currents or voltages are solved for directly.

Kirchhoff's laws are: (1) the current flowing into any junction in a circuit must be equal to the current flowing away from that junction; (2) the algebraic sum of the sources of emf and the voltage drops around any closed circuit must be equal to zero.

The first law states that current flowing toward a point at which it may divide must equal the sum of the portions into which it divides. Otherwise, some current would be left over with no place to go. The second law states that the voltage supplied by the source of emf (a battery, for example) must be equal to the

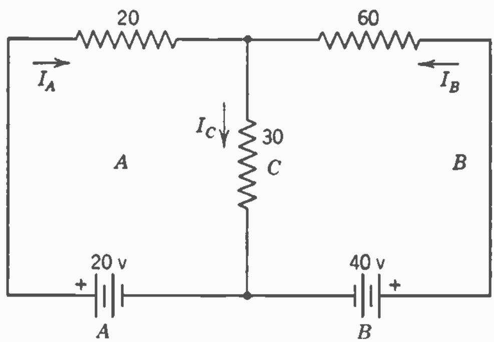


FIG. 14. An example of a circuit easily solved by use of Kirchhoff's laws.

sum of the voltage drops ($I \times R$) appearing across the several circuit components in a closed loop. All the battery voltage must be accounted for, and the voltage drops must add up to the emf produced by the battery, for there is no other source of emf.

In solving complicated circuits, only one portion need be considered at a time. For example, in Fig. 14 one might think that battery B, having higher voltage than battery A, would force some current down through the A circuit and therefore one ought to subtract the two voltages. This is unnecessary, as will be shown in the following example.

Example 8. Current through C is made up of two parts, one due to battery A and one to battery B. No other current flows through C since there is no other source of current. Then Kirchhoff's first law states that $I_C = I_A + I_B$. The junction is the point at which the two currents I_A and I_B join.

Applying the second law, the emf of battery A is accounted for by adding the two voltage drops in circuit (loop) A. These drops are $20I_A$ and $30I_C$, made up of I_A flowing through 20 ohms and I_C through 30 ohms. Similarly the emf of battery B produces two voltage drops in circuit B. Thus

$$I_C = I_A + I_B \quad \text{(Kirchhoff's first law)} \qquad [1]$$

$$20 = 20I_A + 30I_C \quad \text{(Kirchhoff's second law)} \qquad [2]$$

$$40 = 60I_B + 30I_C \quad \text{(Kirchhoff's second law)} \qquad [3]$$

Now there are three unknown currents, I_A, I_B, and I_C, and three equations expressing the relations among them. Rewrite equation 1 as

$$I_A = I_C - I_B$$

Substitute this value for I_A in equation 2:

$$20 = 20I_C - 20I_B + 30I_C$$

or

$$20 = 50I_C - 20I_B \qquad [4]$$

and from equation 3:

$$40 = 30I_C + 60I_B$$

Now multiply by 3 both sides of equation 4:

$$60 = 150I_C - 60I_B \qquad [5]$$

and add equations 3 and 5. Thus

$$40 = 30I_C + 60I_B \qquad [3]$$

$$60 = 150I_C - 60I_B \qquad [5]$$

$$100 = 180I_C + 0$$

or

$$180I_C = 100$$

$$I_C = 100 \div 180 = 0.555 \qquad [6]$$

From equation 3:

$$40 = 60I_B + 30I_C$$

Substitute for I_C its value 0.555:

$$40 = 60I_B + 30(0.555)$$

$$40 = 60I_B + 16.7$$

$$40 - 16.7 = 60I_B$$

or

$$I_B = 23.3 \div 60 = 0.389 \qquad [7]$$

From equation 1:

$$I_A = I_C - I_B$$

$$= 0.555 - 0.389 = 0.166 \qquad [8]$$

Therefore

$$I_A = 0.166 \text{ amp}$$

$$I_B = 0.389 \text{ amp}$$

$$I_C = 0.555 \text{ amp}$$

The current through the middle branch (I_C) is made up of contributions from circuits A and B. To determine the share of each contribution, another rule called the **superposition theorem** may be used. The superposition theorem states that the current through the C branch due to battery A may be obtained by shorting out battery B (on paper, of course. In an actual circuit battery B would be removed and the terminals to which it was connected would be shorted together), and calculating the current that would flow. With battery B shorted out, battery A forces current through a 20-ohm resistance and through two resistances of 30 and 60 ohms in parallel. Since the resultant of two resistances in parallel is equal to the

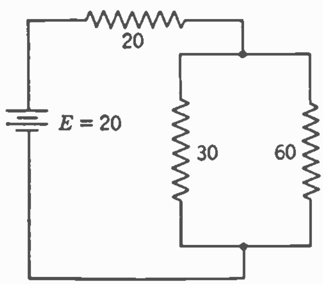


FIG. 15. Use of superposition theorem in solving left-hand part of the circuit of Fig. 14.

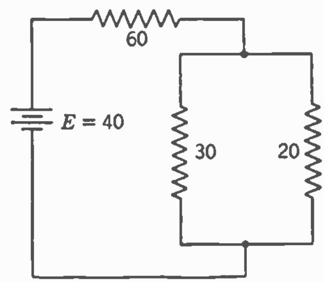


FIG. 16. Representation of the right-hand portion of the circuit in Fig. 14.

product of the two divided by their sum, the current through C due to A is solved as follows:

$$I_A' = \frac{20}{20 + \dfrac{30 \times 60}{30 + 60}} = \frac{20}{20 + \dfrac{1800}{90}} = \frac{20}{20 + 20} = \frac{20}{40} = 0.5 \text{ amp}$$

This is the current flowing out of battery A. But not all of it flows in the 30-ohm resistor. Part flows in the 60-ohm resistor in parallel with the 30-ohm resistor C. The current through the individual portions of the parallel circuit will be inversely proportional to their resistances and will numerically be equal to the total current I_A' multiplied by the resistance of the resistor in which the desired current does *not* flow, divided by the sum of the two resistances.

Thus the current through the 30-ohm resistance will be equal to the total current flowing (I_A') multiplied by $60 \div 90$:

$$I_C' = I_A' \times \tfrac{60}{90} = \tfrac{2}{3} \times I_A' = \tfrac{2}{3} \times 0.5 = 0.333 \text{ amp}$$

Similarly, the current through C due to battery B will be found to be 0.222 amp, and the sum of these currents, 0.555 amp, equals the current found by Kirchhoff's laws. (The reader should solve for the current in C due to battery B.)

Note. In solving problems of this type the engineer must assign a definite direction of current flow in each circuit element and then stick to it. The choice is strictly arbitrary. Usually conventional direction of current flow is assigned where the direction is obvious. That is, current is assumed to flow away from the positive terminal of a battery. This choice, however, is not necessary for a solution of a problem, and time should not be wasted trying to determine ahead of time which way the current actually flows. If a wrong choice is made, the resultant current will carry a minus sign. The minus sign merely means that current in the

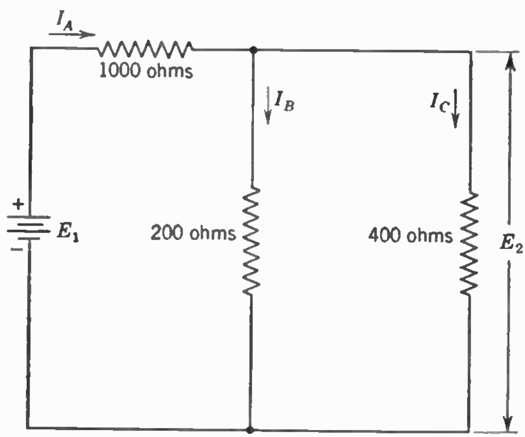


Fig. 17. A problem in series and parallel circuits.

element under consideration actually flows in a direction opposite to that originally assumed. Any minus signs appearing in the course of the solution must be carried along through the remainder of the solution.

Problem 43. In Example 8, reverse the connections of battery B and solve for the currents.

Problem 44. If, in Fig. 17, $I_A = 30$ ma, what is the current I_B in the 200-ohm resistor? What is the voltage E_1?

Problem 45. If, in Fig. 17, E_2 is 20 volts, what must be the value of E_1? If resistors are available in 0.5-, 1-, 5-, and 25-watt sizes, determine which wattage rating should be specified for each resistor.

2.23. Sources of direct current. Electrical energy does not exist in nature in a form useful to man. It must generally be transformed from some other form of energy. For example, the mechanical energy of a steam or gasoline engine may be transformed into electrical energy by means of a generator.

The commonest sources of direct current useful to radio workers are the **battery,** the **generator,** and the **rectifier.** The battery is a device which converts chemical energy into electrical energy; the generator converts mechanical energy into electrical energy; and the rectifier converts a-c energy into d-c energy. Direct-current generators are described in Chapter 4, and rectifiers and associated apparatus are described in Chapter 12.

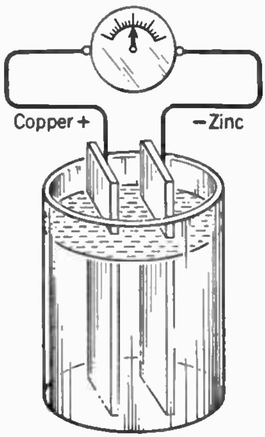


Fig. 18. Plates of copper and zinc immersed in a solution of copper sulfate will produce an electrical voltage.

A **battery** is made up of one or more units called **cells.** The essentials of a cell are three: two elements called **electrodes,** usually of different conducting materials; and a chemical solution known as the **electrolyte,** which acts upon one of the electrodes more than it does on the other. In this action, one of the electrodes is usually eaten up, and when this conductor, generally a metal, is gone, the battery is exhausted; it must be thrown away or the metal replaced. If the metal can be replaced by sending a current through the cell from some outside source, that is, by reversing the process through which the cell was exhausted, the cell is known as a **secondary** or **storage** cell. A cell that must be thrown away when one of the electrodes is eaten up is called a **primary** cell. The dry cell is a well-known example of a primary cell. A simple primary cell is shown in Fig. 18.

2.24. Dry cells. The common dry-cell battery is a very familiar source of emf for vacuum-tube circuits. It is widely used in portable electronic equipment and in some field telephone sets, or in any service where the current required and the periods of current drain are not very great.

The dry cell is commonly made up of a zinc container within which is a sal ammoniac (ammonium chloride) electrolyte mixed with some porous material like manganese dioxide and powdered carbon. In the center of this paste is a carbon rod which forms the positive terminal. The zinc case is the negative terminal. The emf of a fresh battery is 1.5 volts, and the cell can be used until the emf falls to 1.13 volts or even lower. In use, the zinc case is consumed by the electrolyte; when the case is badly eaten away, the cell must be discarded, as the moisture will dry out of the electrolyte by exposure to the air. Also, the container may swell and some of the electrolyte may leak out and damage the surrounding equipment. It should be noted that a dry cell is not truly "dry."

The life of the dry cell depends, naturally, upon the rate at which power is taken from it. On the shelf, no power being required from the cell, the life is approximately 12 months. The life during use depends upon the current required. If the cell is operated at low current drain, the life during use may be fully as long as the shelf life. If a large current is required the life may be only a few weeks. Typical figures on the life of cells may be misleading since improvements in manufacturing are being made constantly.

Dry cells are made in several sizes, including small units which are built up into sources of rather high voltage by placing a number of individual cells in series. A 45-volt B battery has 30 cells in it, each delivering 1.5 volts.

Cells will not deliver full voltage at low temperatures, but freezing does not harm them. The best way to determine the condition of a dry cell is to measure its terminal voltage with a high-resistance voltmeter while it is supplying power to the circuit in which it is to operate.

2.25. Storage batteries. There are two common types of storage batteries: (1) a cell using sulfuric acid as the electrolyte, with a positive plate of lead peroxide and a negative plate of spongy lead; and (2) the nickel-iron alkaline Edison cell. The lead-acid battery is by far the more common and is the type used in almost every automobile. The great virtue of the storage battery is that it may be recharged after it has run down. On the other hand, it is bulky, heavy, and expensive.

Alkaline cells stand mechanical shock better than lead cells, but they cost more. They have longer life than lead cells if both are handled properly, and they weigh less and occupy less space than lead cells of the same capacity.

If the lead-acid battery is carefully treated, not overdischarged before it is recharged, it will last from 5 to 10 years. If it is abused or subjected to high current drains, its life may be much shorter.

The lead-acid cell delivers, fully charged, a voltage of approximately 2.1 volts; the alkaline cell, about 1.45 volts. A short circuit is harmful to the lead cell but not to the alkaline cell. The lead cell, however, will deliver high currents in emergencies, such as starting an automobile on a cold morning.

Alkaline cells may be stored in a discharged and short-circuited condition. Lead cells should be recharged immediately if they have been accidentally short-circuited, and they should be stored fully charged. A charged lead cell will freeze at $-61°$ F; a discharged cell, at $+18°$ F.

The condition of the charge of a lead cell is best tested by means of a hydrometer, a device for measuring the specific gravity (weight per unit volume) of the electrolyte. Since the electrolyte is heavier in a fully charged battery than in a discharged one, the specific gravity is a measure of the condition of the battery.

Since the individual cells of a lead-acid battery have a voltage of about 2 volts, batteries may be built to supply 6, 12, and 24 volts by connecting the necessary number of cells in series.

2.26. Battery characteristics. The term **ampere-hour** is used to express the amount of electricity taken from, or put into, a battery. As the name implies, an ampere-hour represents a current of 1 amp flowing for a period of 1 hr. A 100-amp-hr storage battery will (theoretically, at least) deliver 1 amp for 100 hr or 100 amp for 1 hr. The ampere-hour capacity of batteries decreases as the current drain increases.

One might think that an unlimited current could be drawn from a battery if its terminals were short-circuited. This is not true. A low-resistance ammeter placed across a dry cell gives a definite current reading—it is not unlimited. There must be some resistance in the circuit greater than that of the ammeter and connecting wires. For example, a new dry cell will deliver about 30 amp through wires of very low resistance. An old cell may deliver only an ampere or two.

The additional resistance which limits the current from a cell to a certain maximum is the internal resistance of the cell. This resistance depends upon the construction of the cell, the electrode and electrolyte material, the separation of the electrodes, and the age of the cell. The older the cell, the greater the resistance. The current delivered by a cell is

$$I = \frac{E}{r + R}$$

where r = internal resistance of cell.
R = external resistance of circuit.

Cells which have a large internal resistance deliver but small currents; low-resistance cells deliver large currents. If the internal resistance of a cell is known to be much smaller than the external resistance of the circuit, the current can be calculated as $I = E/R$. The error will be small.

A test of a cell with an ammeter is, in reality, a means of determining the condition of the cell by measuring the internal resistance. When the cell gets old or has become exhausted by too heavy current drain, its internal resistance becomes high and an ammeter reads only small current when placed across it.

The storage battery is a very low-resistance device. The terminal voltage of each cell is about 2.1 volts when fully charged, and the internal resistance is about 0.005 ohm. Placing an ammeter across such a cell is dangerous. The meter will probably be ruined.

Example 9. A dry cell on short circuit (zero external resistance) delivers 30 amp. Its terminal voltage is 1.5 volts on open circuit. What is its internal resistance?

By Ohm's law

$$I = E \div r$$

$$30 = 1.5 \div r$$

$$r = 1.5 \div 30 = 0.05 \text{ ohm}$$

Example 10. The emf of a battery is 6 volts. When 100 ohms is placed across it the voltage falls to 5 volts. What is the internal resistance of the battery?

In Fig. 19, R = 100 ohms, r = the internal resistance of the battery, and the voltage drop across the 100 ohms is 5 volts, which leaves a 1-volt drop in the internal resistance of the battery. The current through the external 100-ohm resistor is, by Ohm's law,

$$I = 5 \div 100 = 0.05 \text{ amp}$$

This current must also flow through the internal resistance of the battery, and there it causes a voltage drop of 1 volt.

$$E = I \times r$$

$$1 = 0.05 \times r$$

$$r = 1 \div 0.05 = 20 \text{ ohms}$$

Note that the internal resistance of the battery is represented as being in series with the voltage and the external resistance. The reason is that all the current must flow through the internal resistance of all such voltage generators, and hence the resistance of the device is represented in series with the remainder of the circuit. The actual terminals of the battery are

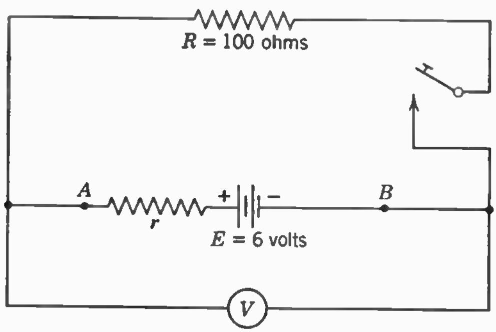


FIG. 19. In this diagram, r represents the internal resistance of the battery.

shown as points A and B in Fig. 19. Care must be taken to place the voltmeter in the proper place in the diagram drawn to represent the circuit. It must be placed on the *actual* terminals of the battery.

The voltage of the cell on open circuit is its **emf.** Under load the voltage falls and is then labeled as the **pd** (potential difference).* The pd, then, depends on the load current. The emf of high-resistance cells can be measured only by high-resistance meters, those that take but little current from the cell. When the voltage of a cell or battery is mentioned, its emf is assumed unless otherwise stated.

2.27. Cells in series. Cells, both dry and wet, may be connected in several ways. When the positive terminal of one cell is

* For fresh batteries the emf and pd are nearly the same and are often assumed to be exactly equal. If a battery is old, then the effect of internal resistance must be considered. This can be done most easily by considering the actual battery to be made up of a perfect battery with no internal resistance in series with a resistor whose value is equal to the internal resistance of the actual battery. The pd is then the voltage measured across the terminals of this equivalent circuit.

connected to the negative terminal of the next cell, as in Fig. 20, the cells are said to be connected in series. Under these conditions, the voltage at the two ends of the series of cells is the sum of the individual cell voltages. At the same time, the total internal resistance of the combination is the sum of the internal resistances of the individual cells, and the current must flow through this total resistance.

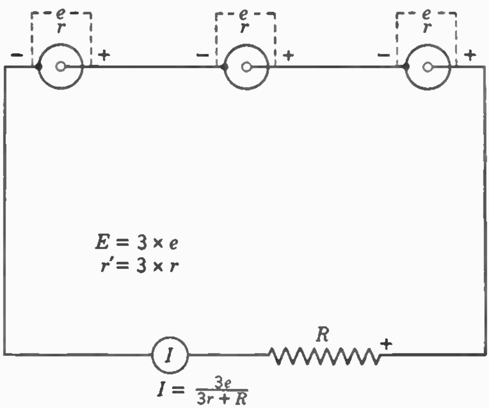


Fig. 20. Three cells in series have a terminal voltage and an internal resistance equal to the sum of the individual voltages and resistances.

If the terminals of a battery made up of several cells in series are connected with a resistance R, the current that will flow may be obtained by Ohm's law, assuming similar cells, as

$$I = \frac{Ne}{Nr + R}$$

where N = number of cells.
e = emf of each cell.
r = internal resistance of each cell.

2.28. Cells in parallel. Often more current at a given voltage is required than may be obtained from a single cell. Then the cells may be connected in parallel by connecting together all the positive terminals of a group of cells. Similarly, all the negative terminals are connected together. (See Fig. 21.) With this connection the terminal voltage of the combination is the same as the terminal voltage of each cell, but the internal resistance of each cell has been divided by the number of cells, N, and has become $r \div N$, assuming all the cells are similar.

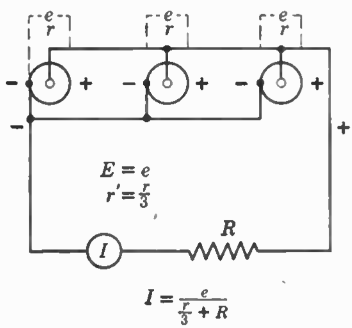


FIG. 21. Placing cells in parallel reduces the internal resistance and enables a greater current to be obtained.

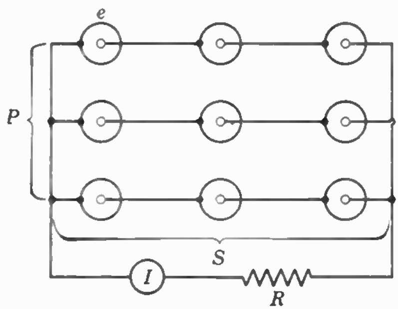


FIG. 22. A series-parallel arrangement of cells enables more current and more voltage to be obtained than from a single cell.

If the ends of the battery are connected with a wire whose resistance is R, the current that will flow is

$$I = \frac{e}{(r \div N) + R}$$

Cells may also be connected in a series-parallel arrangement to obtain both higher voltage and higher current. In Fig. 22 are P sets of S similar cells in series, and the sets themselves are connected in parallel. If the battery shown in Fig. 22 is connected to a wire whose resistance is R, the current that will flow is

$$I = \frac{Se}{(rS/P) + R} = \frac{Ne}{rS + PR}$$

where $N = P \times S$.

Problem 46. A battery has an emf of 45 volts and an internal resistance of 20 ohms. What voltage will a voltmeter read when placed across the terminals of the battery if the resistance of the voltmeter is 50,000 ohms? if the voltmeter resistance is 100 ohms?

Problem 47. A military expedition is to have 100 receivers, each requiring 90 volts at 15 ma. Individual B battery cells deliver 1.5 volts, and at the 15-ma drain they have a life of 780 hr. If each 90-volt battery costs

\$2.00, estimate the cost for energy per kilowatt-hour. If the individual cells have a diameter of 1.25 in. and are 4 in. high, estimate the minimum cargo space required. Each cell weighs 0.366 lb. What total weight must be carried?

Problem 48. A radio transmitter requires a filament voltage of 10 volts. Assume that lead-acid batteries have a fully charged voltage of 2.05 volts and a discharge voltage of 1.75 volts, and that the corresponding voltages for alkaline cells are 1.45 and 1.0 volts.

How many lead-acid cells are required? How many alkaline cells are required?

If the filament current is 3.25 amp, what is the maximum value of a variable resistance to be placed in the battery-filament circuit to adjust the current to the proper value? Give answers for both lead-acid and alkaline batteries.

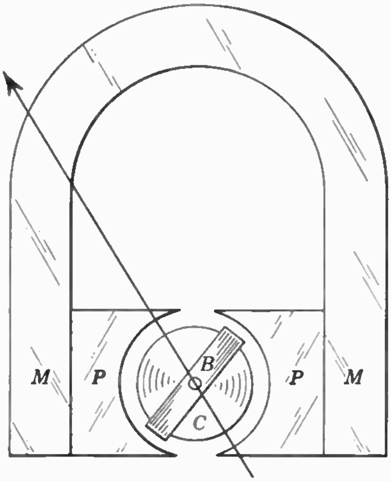
3 · Electrical Meters and Measurements

BASIC METERS

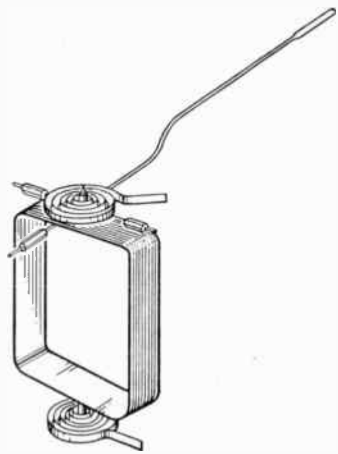
In many ways electricity is an extremely peculiar "substance." It cannot be detected in the usual way in which most things are detected, for it cannot be seen, heard, or smelled. And yet it is necessary that there be ways of measuring electrical quantities, for without methods of measuring it would be difficult to put electricity to useful work.

Since electric currents or voltages cannot be detected directly, advantage is taken of the effects which they produce. Four kinds of effects are important: thermal, chemical, magnetic, and mechanical. For example, a fine, high-resistance wire gets hot when current flows through it. Two dissimilar metals (copper and zinc, for example) placed in a solution of one of them (copper sulfate) give off gas bubbles when a wire connects them externally. The gas bubbles are the product of the chemical action which generates a voltage and causes current to flow. When a wire carrying an electric current is brought near a compass needle, the needle will change from its habitual north-south position. Finally, two adjacent metal plates, charged to a high voltage difference, are attracted to each other. Any one of these fundamental effects of electricity may be used to detect its presence and, with proper design of instruments taking advantage of these effects, can be used to measure the flow of current or the magnitude of a voltage. A hot-wire ammeter, for example, is merely a wire which expands when heated by a current flowing through it. A needle is attached to the wire and is pulled across the scale by a spring as the wire gets hot. An electrostatic voltmeter consists of a fixed and a movable plate. The movable plate is connected to a pointer which moves across a scale to indicate the voltage impressed across the plates.

3.1. D'Arsonval meter. The d'Arsonval movement is employed on practically all the modern meters for the measurement of direct current and voltage. This instrument, illustrated in Fig. 1, consists of a coil of fine wire wound on a bobbin, *B*, surrounding a soft-iron core, *C*, and mounted so that the assembly revolves in a magnetic field set up by a permanent magnet, *M*. A pointer



(a) General Details



(b) Details of Moving Coil

Fig. 1. Essential parts of a d'Arsonval moving-coil movement. (Courtesy, Weston Electrical Instrument Corp.)

is attached to the coil of wire. When current flows through the wire, the coil revolves and the needle is carried across a scale. Springs which connect the coil to the meter terminals retard the movement of the coil and cause the amount of the deflection to be proportional to the strength of the current.*

D'Arsonval meters can be made sensitive enough to measure a microampere and to detect (without measuring accurately) even smaller currents. Since such extremely sensitive meters are not very rugged and are easily damaged, the most sensitive meters for general use give a full-scale deflection for about 20 microamperes of current through the movable coil.

* Since the coil and associated moving parts have considerable mechanical inertia, the needle tends to hold a steady position even if the current is pulsating. The indication, thus, is the average or d-c value of the current even in the presence of a-c components.

3.2. Thermocouple instruments. In another type of instrument advantage is taken of the fact that two dissimilar metals heated at a junction between them will produce a small but measurable emf. The basic elements of a thermocouple meter are shown in Fig. 2. In this instrument the dissimilar metals are copper and constantan. The thermal junction of these two wires is heated by a current flowing through a heater wire placed close to the junction. The emf produced by the heated thermocouple causes a current to flow which is measured by a d-c meter of the d'Arsonval type. Two milliamperes is about the smallest current that can be indicated by this type of instrument.

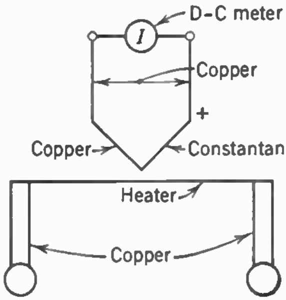


FIG. 2. Elements of thermocouple meter.

Thermocouple instruments work equally well on alternating or direct currents since either kind of current causes the heater to warm up. Since the amount of heat produced is proportional to the square of the heater current, this type of instrument inherently has a non-linear scale on which the low-current readings are bunched together. A thermocouple meter is generally calibrated using direct current, and with proper design will indicate r-f currents up to 60 or 70 megacycles.

3.3. Rectifier-meter instruments. A copper oxide rectifier may be used as shown in Fig. 3 to transform an alternating current to a direct current, which is then measured by a d'Arsonval meter. The copper oxide unit consists of several small flat plates close together, and the capacitance between them is rather high. At the higher frequencies some of the current is by-passed around the rectifier unit by this capacitance, and the meter is thus limited to rather low frequencies. Such meters are usually accurate over the audio-frequency range, but their

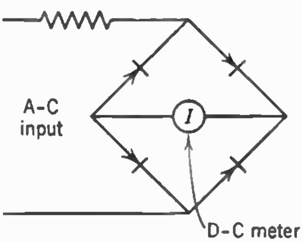


FIG. 3. Copper oxide rectifier meter.

accuracy begins to fall off rather rapidly above about 10,000 cycles.

If, instead of the copper oxide rectifier, a rectifier unit having low capacitance is used, currents of quite high frequencies may be measured with a rectifier meter. A crystal rectifier meets this requirement. It consists of some substance such as silicon or germanium with a small "cat-whisker" contact as shown in Fig. 4. This type of rectifier was widely used as a detector in the early days of radio, then practically disappeared. In recent years its use has again become very widespread in the detection and measurement of very high-frequency currents in radar and other microwave equipment. In contrast to the earlier types, modern crystal rectifiers are fairly rugged and are mechanically stable.

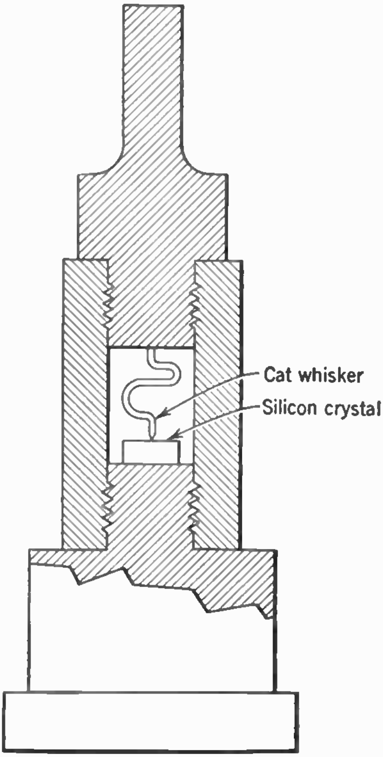


Fig. 4. Details of crystal detector rectifier.

3.4. Iron-vane meters. In another type of instrument, used mostly on low-frequency alternating current, the magnet and coils are reversed in the roles they play. The magnet is made of soft iron and is the movable part,. whereas the coil remains stationary as seen in Fig. 5. Currents as low as 15 ma may be measured in this way. The meter is limited in its usefulness to frequencies below about 200 cycles because the inductance of the coil cuts down the current flow and reduces the "turning power" of the coil as the frequency is increased.

3.5. Dynamometer-type meter. In a dynamometer type of meter two coils are used. One coil revolves inside the other coil as seen in Fig. 6. Current is conducted to the movable coil much as in the d'Arsonval meter. All or part of the same current flows through both coils. The instrument will work on alternating current, since the reversals of current occur in both coils at the

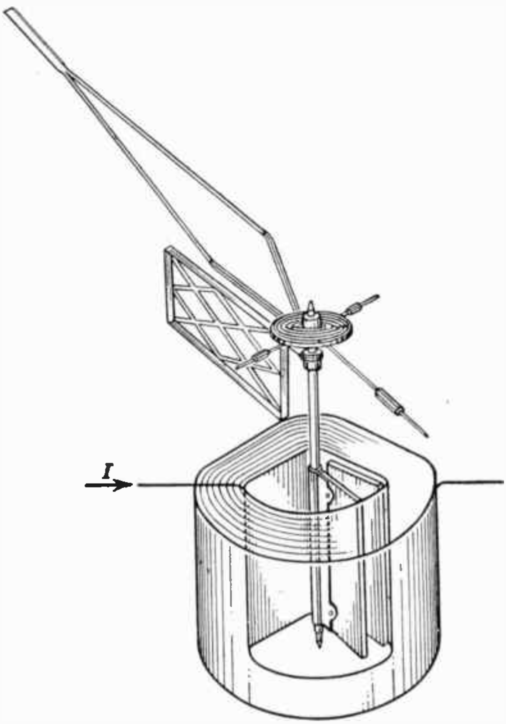


FIG. 5. Iron-vane type of meter for alternating currents.

same instant and a torque (turning power) is produced in the same direction over an entire a-c cycle. The deflection is approximately proportional to the square of the current. Like the iron-vane meter, the dynamometer instrument is normally limited in usefulness to frequencies below about 200 cycles. With special compensation, the frequency range may be extended to about 2000 cycles.

USE OF METERS

3.6. Ammeters. Meters used to measure current are called **ammeters.** They are connected in series with the source of the current and the load. Passing too much current through them easily damages them, bending the pointer, injuring the movement

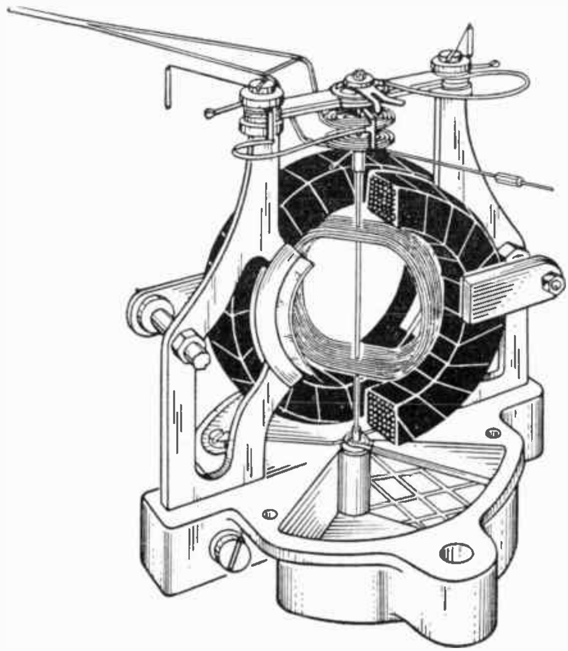


Fig. 6. Dynamometer type of instrument for alternating current. (Courtesy, Weston Electrical Instrument Corp.)

mechanically, or actually burning out the moving coil. Meters may be protected by fuses or by shunting them with low-resistance conductors which allow most of the current to by-pass the meter itself.

In addition to their use for the protection of ammeters, shunts may also serve to convert a low-current meter into a higher-current meter. When this is done the resistance of the shunt must bear a special relation to the resistance of the meter coil. If the coil resistance is known, the value of the shunting resistance can be calculated by the use of parallel-circuit formulas. The value is

$$R_s = R_m\left(\frac{X}{1 - X}\right)$$

where R_s is the shunt resistance.

R_m is the meter resistance.

X is the proportion of the total current that is to flow through the meter, that is, one-tenth, one-half, etc.

The manner in which ammeters may be adapted to read currents higher than originally intended may be shown by the following example. Suppose a meter is available which has a resistance of 28 ohms and reads full-scale when 1 ma of current flows through it. Now, if it is shunted with a resistance of 0.57 ohm, the total current taken by the meter and its shunt will be 50 ma, but only 1 ma will go through the meter and 49 ma through the shunt, because the shunt has so much less resistance than the meter. The original readings of the meter may now be multiplied by the proper factor, 50 in this case, for an indication of the total current taken by the meter and its shunt.

If large currents are to be passed by the shunt it is important that the material from which the shunt is made has a low temperature coefficient of resistivity so that the resistance does not change appreciably with temperature. Manganin is often used because of its low temperature coefficient.

Problem 1. A 0- to 1-ma milliammeter has a resistance of 50 ohms. What shunt resistance is required to convert the meter to a 0- to 20-ma meter? What will be the voltage drop across the meter when maximum current flows through it?

3.7. Voltmeters. Ammeters have low resistance. They are connected in series with the apparatus, taking current from the source as shown in Fig. 7. **Voltmeters,** on the other hand, must indicate the voltage across some part of the circuit. They must

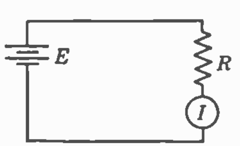


FIG. 7. Proper connection for measuring current.

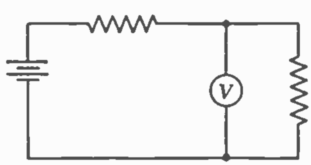


FIG. 8. Voltmeter goes across circuit being measured.

not permit much current to flow through them because this current would be taken away from the circuit. The voltmeter is connected across the element whose voltage is to be measured, as in Fig. 8.

A voltmeter is, in reality, an ammeter with a high resistance. This high resistance may be built into the moving coil of the meter (if it is of the d'Arsonval type), but more commonly is attained by connecting a series resistor to one of the terminals. This series resistor is called a **multiplier.** The following example will show how a voltmeter is obtained from an ammeter.

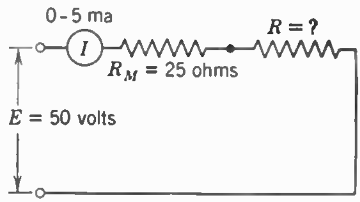


FIG. 9. Use of series resistance to measure voltage with a current meter.

Example 1. A 0- to 5-ma ammeter is to be used to measure a full-scale voltage of 50 volts. The resistance of the ammeter itself is 25 ohms. How much resistance must be added to that of the meter? See Fig. 9.

The total resistance, including that of both the meter and the multiplier, which will permit only 5 ma to flow when the voltage is 50 volts, is

$$R_{\text{total}} = \frac{E}{I} = \frac{50}{5 \times 10^{-3}} = \frac{10}{10^{-3}} = 10^4 = 10{,}000 \text{ ohms}$$

The multiplier resistance, then, is

$$R = R_{\text{total}} - R_{\text{meter}} = 10{,}000 - 25 = 9975 \text{ ohms}$$

3.8. Sensitivity of meters. A sensitive current-measuring meter is one which will measure very small currents and which has a low resistance, that is, one which will give a great deflection of the pointer with very little current. There is always some voltage drop across an ammeter, and consequently the voltage supplied to the remainder of the circuit (load) is somewhat lower than the source voltage. For this reason it is desirable to have ammeters of low resistance.

A sensitive voltmeter is one which will give a large needle deflection for a small current and which at the same time has a high resistance. The sensitivity of voltmeters is often rated in **ohms per volt,** and this value is higher for more sensitive meters. The ohms per volt sensitivity is obtained by dividing the resistance of the meter and its multiplier by the full-scale voltage reading.

Thus, for Example 1, the sensitivity is 10,000 ÷ 50 or 200 ohms per volt. An alternative method is to divide 1 volt by the full-scale current that the meter takes. Thus, $1 \div 5 \times 10^{-3} = 200$ ohms per volt, which is the same as before.

Problem 2. A 0- to 100-μa meter has a resistance of 400 ohms. How large must be the multiplier resistance if the meter is to be used as a 0- to 10-volt voltmeter? What will be the ohms per volt sensitivity of the voltmeter?

Problem 3. What is the lowest full-scale voltage reading which can be obtained from the microammeter of Prob. 2?

3.9. Voltmeter loading effects. Many of the resistors used in vacuum-tube circuits have resistances of several thousands or hundreds of thousands of ohms. If a voltmeter is temporarily connected across one of the resistors to measure the voltage, the indicated voltage may be quite different from that existing before the meter was connected. This effect of the voltage changing when the meter is connected in the circuit is known as **loading.** As a general rule, loading will not be appreciable if the resistance of the voltmeter is at least ten times the resistance across which the voltage is to be measured. Often more nearly accurate readings can be obtained on a multirange instrument by switching it to a higher range scale than is actually needed in order to take advantage of the higher resistance on the higher scale. A sure indication of loading is found when the indicated voltage is considerably lower as the meter is switched from a high range scale to a lower range scale.

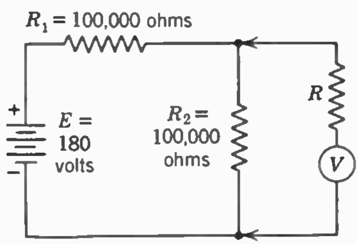


FIG. 10. An example of voltmeter "loading" so that different voltages are obtained.

Example 2. Suppose a voltmeter is available which has a sensitivity of 1000 ohms per volt and which has 100-, 500-, and 1000-volt scales. The voltage across R_2 in Fig. 10 is read on all three scales. Compare the readings to the voltage (calculated) across R_2 before the meter is connected.

The voltmeter, on the 100-volt scale, has a resistance of 1000 × 100 or 100,000 ohms. This resistance is in parallel with R_2, and their combined resistance is

$$\frac{100{,}000 \times 100{,}000}{100{,}000 + 100{,}000} = \frac{100{,}000 \times 100{,}000}{200{,}000} = 50{,}000 \text{ ohms}$$

The voltage read on the voltmeter is

$$\frac{50{,}000}{100{,}000 + 50{,}000} \times 180 = \frac{50{,}000 \times 180}{150{,}000} = 60 \text{ volts}$$

Similarly, the voltmeter on the 500-volt scale has a resistance of 1000 × 500 or 500,000 ohms. The combined resistance of R_2 and the voltmeter is

$$\frac{100{,}000 \times 500{,}000}{100{,}000 + 500{,}000} = \frac{100{,}000 \times 500{,}000}{600{,}000} = 83{,}333 \text{ ohms}$$

and the voltage read on the voltmeter is

$$\frac{83{,}333}{100{,}000 + 83{,}333} \times 180 = \frac{83{,}333 \times 180}{183{,}333} = 81.7 \text{ volts}$$

On the 1000-volt scale the voltmeter resistance is 1,000,000 ohms and the combined resistance of R_2 and the voltmeter is 91,000 ohms. The voltage read on the 1000-volt scale is 85.7 volts.

Since the resistances of R_1 and R_2 are equal, the voltage across each is one-half of 180 or 90 volts.

	Voltage across R_2, Volts
No voltmeter	90
1000-volt scale	85.7
500-volt scale	81.7
100-volt scale	60

3.10. Wattmeters. Since power in watts is the product of the emf in volts and the current in amperes, a **wattmeter** must be arranged to read the product of the voltage across a device and the current through it. A dynamometer-type instrument may be employed as a wattmeter by using one coil as a potential (voltage) coil and connecting it in shunt with the load (across the line). The other coil serves as a current coil and is connected in series with the load. When used as an a-c wattmeter, a dynamometer instrument indicates $EI \cos \theta$, which is the formula for a-c power (Sect. 7.21). The manner in which a wattmeter is connected to measure power to a load is shown in Fig. 11.

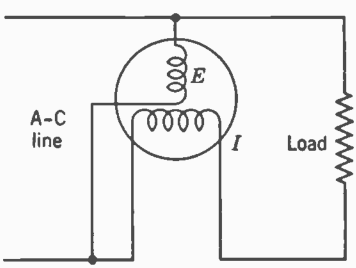


Fig. 11. Connections of a wattmeter into a circuit.

RESISTANCE MEASUREMENTS

The most important and most frequent single measurement in radio practice is the measurement of resistance. Many resistors are used in radio receivers, transmitters, and other electronic equipment, and the actual resistance of these units is often important. Furthermore, the ohmic resistance of many units, used for purposes other than as resistors, is an important indication of the condition of the unit, that is, whether the unit is normal or whether it needs to be repaired or replaced. Methods of measuring resistance, then, are important and useful.

3.11. Ammeter-voltmeter method of measuring resistance. The simplest method of measuring resistance follows from Ohm's law. The resistance of a device is the ratio of the voltage across the device to the current through it ($R = E \div I$). If the voltmeter utilized has a much higher resistance than that of the device being measured, its inclusion in the circuit need not be considered. Otherwise, the loading effect of the voltmeter must be taken into account. Since the resistance of d-c ammeters is usually low enough to be considered negligible, trouble may often be avoided by connecting the voltmeter across both the unknown resistance and the ammeter in series as shown in Fig. 12. With this connection, the current drawn by the voltmeter does not flow through the ammeter and a better value for the resistance is obtained.

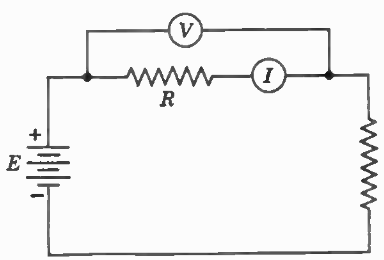


FIG. 12. Ammeter-voltmeter method of measuring resistance.

Example 3. Consider the circuit of Fig. 12. A voltmeter V across the device whose resistance is unknown reads 75 volts, and the current meter I indicates a current of 0.05 amp. What is the unknown resistance?

$$R = E \div I = 75 \div 0.05 = 1500 \text{ ohms}$$

3.12. The ohmmeter. A simple instrument very commonly used for resistance measurements is a direct-reading **ohmmeter** consisting of an ammeter and a battery. The circuit of a series-type ohmmeter is shown in Fig. 13, where R is the unknown resistance whose value is to be measured. With R short-circuited, resistor S inside the instrument case limits the current taken by the meter from the battery to about full-scale deflection. The deflection is made exactly full-scale by means of the variable shunt resistor B across the indicating meter. When R is placed in the circuit, the deflection of the instrument decreases to correspond to the new value of the current flowing. This current is, of course, less than it was with the unknown resistance short-circuited, and the meter can be calibrated directly in terms of ohms rather than amperes. Note that full-scale deflection corresponds to zero resistance, whereas no deflection indicates an open circuit.

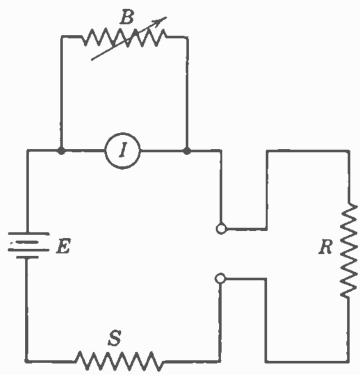


Fig. 13. Circuit of a series-type ohmmeter.

The useful range of this type of meter is usually considered as 10 times the half-scale resistance indication. The calibration marks for deflections below about third-scale are too crowded to be of much use. Ranges may be 1000, 10,000 and 100,000 ohms. Use of a sensitive instrument or more voltage will permit the measurement of several megohms by this method. Low ranges, such as 100 or 10 ohms, are usually obtained by shunting the unknown resistance across the meter instead of placing it in series with the meter. This type of ohmmeter (shunt ohmmeter) deflects up-scale for large resistances, since large resistances will shunt very little current around the meter.

Ohmmeters are used to test the continuity of a circuit, as, for example, a tube filament, a transformer winding, or a line cord. They may also be used to detect short circuits, as in a capacitor or choke, or between a tube cathode and its heater.

The accuracy of resistance measurements made with an ohmmeter is not generally very good. However, the ease with which the instrument can be used far outweighs the low accuracy for all except precision measurements.

3.13. Wheatstone bridge. Resistances are often measured by what is known as the **comparison method,** that is, by comparing them with resistance units whose values are known. For example, the current through an unknown resistance, R_1, as in Fig. 14, can be measured with the ammeter I. Then a variable calibrated resistance R_2 can be switched in the circuit and adjusted to give the same current. The two resistances are then equal in value.

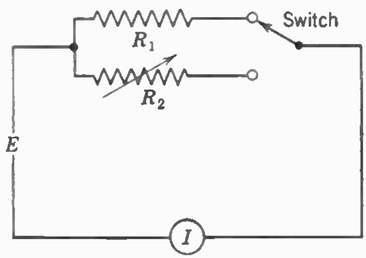


Fig. 14. Comparing one resistance with another.

Another method employs a **Wheatstone bridge.** Such a method is standard practice when high precision is required. In diagrammatic form a Wheatstone bridge is represented in Fig. 15. R_1 and R_2 are resistors whose resistances are accurately known, R_3 is the unknown re-

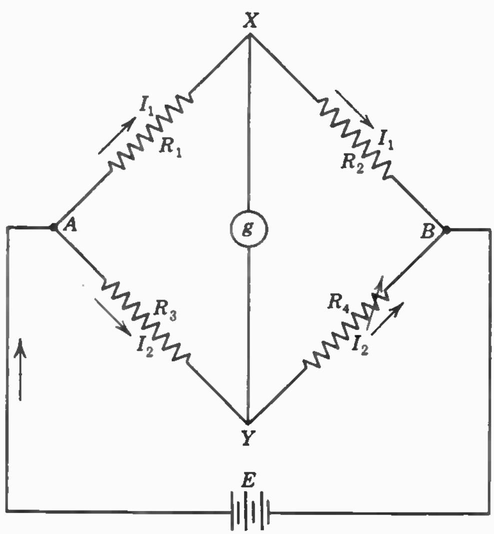


Fig. 15. Circuit of the Wheatstone bridge for measuring resistance.

sistance, and R_4 is a calibrated variable resistance to which the unknown is compared. A Wheatstone bridge is used as follows: A current is led into the bridge arrangement of resistances at the points A and B, and a sensitive current-indicating meter g is placed between points X and Y. The values of R_1, R_2, and R_4 are adjusted until the meter g shows that no current flows through it; that is, there is no difference of voltage between the points X and Y which would force current through the meter. In other words, X and Y are at the same voltage.

The total current divides at A and flows into the arms of the bridge, forming the currents I_1 through R_1 and R_2 and I_2 through R_3 and R_4. If there is no potential difference between X and Y, the voltage drop along R_1 is equal to the voltage drop along R_3. Thus

$$I_1R_1 = I_2R_3 \tag{1}$$

Similarly

$$I_1R_2 = I_2R_4 \tag{2}$$

Dividing equation 1 by equation 2,

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \tag{3}$$

Suppose that R_1 and R_2 are equal in value. Then equation 3 becomes

$$1 = \frac{R_1}{R_2}$$

or

$$1 = \frac{R_3}{R_4}$$

or

$$R_3 = R_4 \quad (\text{for } R_1 = R_2)$$

and to find the value of the unknown resistance R_3, R_4 is adjusted until no current flows through the meter. Then R_3 and R_4 are equal. Suppose, however, that the unknown resistance is much larger than any value that can be obtained by adjusting R_4. For example, let it be ten times as large. Then it is necessary only to replace R_1 by a resistor ten times as large as R_2, and equation 3 becomes

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} = 10$$

$$R_3 = 10R_4$$

Now R_4 is adjusted until no current flows through the meter. The resistance of this standard R_4 is multiplied by 10 to obtain the value of the unknown resistance R_3.

Resistances R_1 and R_2 are called the **ratio arms;** R_4, the standard resistance, is usually a **resistance box,** that is, a box containing several resistance units. The box is equipped with switches so that any of the accurately known resistances may be utilized.

Problem 4. An ohmmeter is constructed as shown in Fig. 13 except that the shunt B is omitted. Voltage E is 6 volts. The resistor S is adjusted so that a full-scale current of 100 μa flows when resistor R is shorted. With R in the circuit the meter reads 25 μa. What is the value of R?

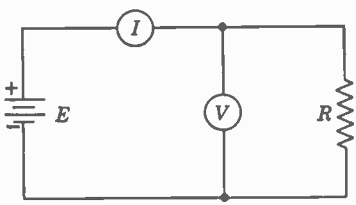


FIG. 16. Example of the use of a high-resistance voltmeter.

Problem 5. An unknown resistance is measured on a Wheatstone bridge. The bridge balances when the resistors have the following values: $R_1 = 10$ ohms, $R_2 = 50$ ohms, and $R_4 = 437$ ohms. What is the value of the unknown resistance?

Problem 6. Let R in Fig. 16 represent the resistance of a tube in a circuit. If the indication on a high-resistance voltmeter is 100 volts when the ammeter indicates 2 ma, what is the resistance of the tube? What would be the resistance of the tube if a voltmeter with a sensitivity of 500 ohms per volt reads 100 volts on its 150-volt scale and the same total current flows through the ammeter?

Problem 7. A 100-μa microammeter is to be used to make a 5000-ohms-per-volt voltmeter. What must be the value of the shunt if the resistance of the meter itself is 300 ohms? (Remember that any 5000-ohms-per-volt meter must pass 200 μa at full scale.)

Problem 8. Suppose that the wattmeter in Fig. 11 is used to measure the power taken by a resistance load. On a 110-volt line the wattmeter reads 50 watts. What is the resistance of the load?

Problem 9. In Fig. 17 suppose that E and R represent a rectifier circuit. The internal resistance R is known to be 10,000 ohms. A 1000-ohms-per-volt voltmeter reads 220 volts on the 250-volt scale when connected as shown. What is the value of E?

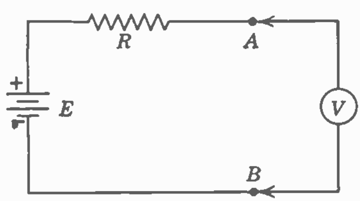


FIG. 17. The terminal voltage, V, depends upon the internal resistance.

Problem 10. In Fig. 17 it is desired to make measurements from which both E and R may be determined. The voltage is first measured across AB with a voltmeter having a total resistance of 100,000 ohms and found to be 500 volts. The voltmeter is then shunted with a 100,000-ohm resistor and the voltage drops to 475 volts. What are the values of E and R?

Problem 11. A d-c microammeter requires a full-scale current of 50 μa and a resistance of 2000 ohms. It is to be used as the foundation meter for a multirange meter having the following scales: 0 to 1, 0 to 10, 0 to 100 ma, and 0 to 10, 0 to 50, 0 to 100 volts. Specify the shunting resistor for each current range and the multiplier resistor for each voltage range.

4 · Magnetism and Electromagnetism

Batteries have been described in a previous chapter as a common source of electrical power. A second important source of electrical power is the generator. Since the generator depends upon electromagnets for its operation, the phenomenon of magnetism must be carefully investigated.

4.1. Magnetism. Everyone is familiar with the common horseshoe magnet. It is made of a piece of hardened steel which has been magnetized and, as is well known, will attract bits of iron and steel. One end is identified as the north pole, and the other as the south pole. Two magnets will repel each other if their north (or south) poles are brought together and will attract each other if the south pole of one is brought near the north pole of the other. The amount of attraction or repulsion follows the same sort of law as do electric charges: the attraction or repulsion is proportional to the strength of the individual magnets (poles) and inversely proportional to the square of the distance between them. Thus

$$F = \frac{m_1 m_2}{\mu d^2} \; *$$

* In this expression F is the force in dynes, m_1 and m_2 are the strengths of the poles of the two magnets in gilberts, μ is the permeability, and d is the distance in centimeters between the poles. The permability of most materials, except iron, cobalt, and nickel, and some of their alloys, is very nearly equal to 1. Iron, cobalt, and nickel and various alloys are so-called magnetic materials and may have values of permeability ranging up to several thousand. As with Coulomb's law, the above formula may be expressed in other systems of units.

Permeability is a measure of the ability of a material to produce magnetic effects; it is the *ratio* of the magnetic effect produced by a material, iron, for example, to the effect produced by air in a given magnetic circuit. The nature of permeability is discussed in more detail in later sections.

Soft iron loses its magnetism easily and is not used for permanent magnets. Instead, hardened steel or certain alloys of nickel and iron, such as Alnico, which is composed of aluminum, nickel, cobalt, and iron, are employed when magnetism must be retained over a long period.

4.2. Terrestrial magnetism. Everyone knows that the earth is a great magnet, but no one knows precisely why. It is thought that the earth's core is iron and that it might have become magnetized in some manner. But the core is supposed to be very hot, perhaps molten, and iron loses its magnetism when heated to high temperatures. On the other hand, the magnetic poles shift about, a fact that is consistent with a molten core but not with a solid one.

During periods of intense sunspot activities serious disturbances appear on wire and radio communications circuits. These disturbances are known as magnetic storms, and it is quite likely that the earth's magnetism is partly due to a magnetized core and partly to the effect of streams of electrons and protons produced in the sunspots.

Currents flowing near the earth's surface during these storms produce high voltages in telephone cables and lines and mask the signals; radio propagation suffers from the same causes.

4.3. Electromagnetism. One of the most important discoveries of all those concerning the phenomena of electricity was that a coil of wire carrying an electric current acts like a magnet. It was this discovery that tied electric current and magnetism together as different aspects of the same science. The action of a permanent magnet can be easily demonstrated by distributing some fine iron filings on a sheet of cardboard or a piece of glass placed over a magnet. As the cardboard or glass is tapped lightly it will be found that the filings orient themselves along lines, called **magnetic lines of force,** that are assumed to leave the north pole of the magnet and enter the south pole. A typical pattern which is obtained is seen in Fig. 1. Similarly, a coil of wire carrying an electric current will be found to have magnetic lines of force like those of a permanent magnet.

The existence of a magnetic field made up of lines of force can be demonstrated by bringing an ordinary mariner's compass near a wire carrying a current. The compass needle will swing from its north-south direction. If the coil of wire is wound on a hollow form, such as a cardboard cylinder as shown in Fig. 2, and if a steady current flows through the wire, it will be found that the magnetic effect of the coil is increased very appreciably when an

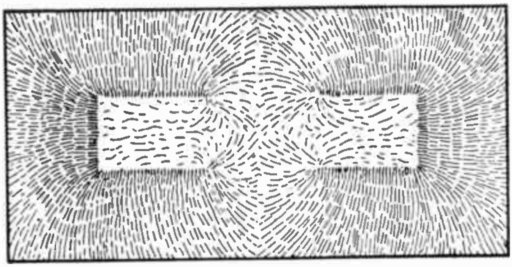


FIG. 1. How iron filings show the lines of force surrounding a bar magnet.

iron core is placed within the coil form. Magnetic lines of force are carried with much less opposition through iron than through air. The ratio of the magnetic effect of the coil with and without

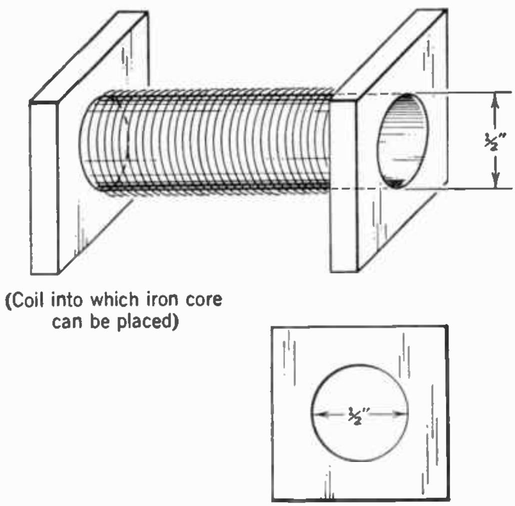


FIG. 2. A cardboard tube wound with wire becomes a magnet when current flows through the wire.

the iron is a measure of the **permeability** of the iron, that is, of its ability to carry magnetic lines of force.

4.4. Magnetic quantities. What is it that makes a magnet attract pieces of iron? What exists in the space surrounding a magnet? Anyone asking these questions is squarely up against a phenomenon of nature. One might as well ask, "Why does a stone fall toward rather than away from the earth?" As yet no one knows precisely what magnetism is. The best one can do is to accept magnetism as a fact and then learn all that is necessary to make proper use of it.

The space surrounding a magnet is a **magnetic field** made up of lines of force. These lines take their place as indicated by the iron filings as described above. Their position can be explored by means of a compass needle, as well as by other ways. Lines of force, then, cannot be detected by the normal senses. They are detected only by effects that they produce. It may well be that lines of force have no real physical existence. However, they do help explain many magnetic phenomena and are important for that reason.

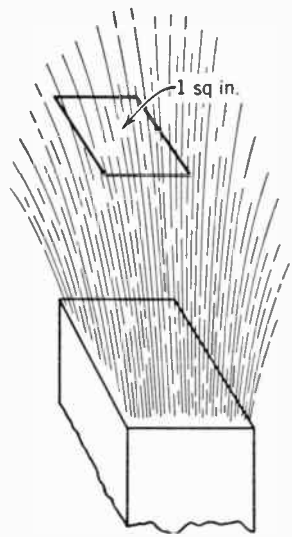


Fig. 3. Flux density is the number of lines through a given area; here 1 sq in.

In any magnetic circuit, certain quantities must be known in order to design properly apparatus using the circuit. For example, the strength of the magnetic field may be expressed by the total number of lines of force, or by the number of lines per unit area, that is, the number of lines that go through a square inch or square centimeter. The total number of lines is known as the **flux,** and the lines per unit area as the **flux density.** Flux density is the total flux divided by the area in which the flux exists. Figure 3 illustrates this concept of flux density.

Another quantity that must be known is the magnetic force required to set up a given number of lines in a material having a given magnetic quality, just as, in an electric circuit, the voltage required to produce a given current through a material of a given resistance must be known.

4.5. Magnetomotive force. Corresponding to emf in an electric circuit is **magnetomotive force** (mmf) in a magnetic circuit, and corresponding to the resistance is the magnetic quantity

reluctance. The magnetic quantity flux corresponds to current. The unit of mmf is the **gilbert;** it is the magnetic "pressure" required to produce 1 line of force in a circuit having a reluctance of 1 unit. (There is no generally accepted name for the unit of reluctance.) In an electromagnet the mmf is proportional to the product of the current in the coil and the number of turns, that is, to the **ampere-turns.**

In an electric circuit, the relations between the "cause" and the "effect" are

$$\text{Current } I = \frac{\text{Electromotive force } E}{\text{Resistance } R}$$

In a magnetic circuit, the corresponding relations are

$$\text{Flux} = \frac{\text{Magnetomotive force } \mathfrak{F}}{\text{Reluctance } \mathcal{R}}$$

Consider a centimeter cube of air, that is, a cube of air 1 cm on an edge. If a mmf of 1 gilbert is applied to opposite faces, 1 line of force will go through the cube. The reluctance of this cube of air, therefore, is 1 unit. If the cube is replaced by a column of air 1 cm wide, 1 cm deep, but 2 cm long, the reluctance will be 2 units. If, however, the cross-sectional area is increased, the reluctance will be decreased. Similarly, the resistance of a piece of copper wire 1 ft long and having a cross-sectional area of 1 circular mil is 10.4 ohms. A wire 2 ft long will have twice this resistance, and if the cross-sectional area is increased (by using a larger wire) the resistance will be decreased.

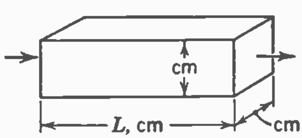


FIG. 4. Reluctance depends directly upon the length and inversely upon the area of the material.

In a magnetic circuit all materials do not have the same reluctance. However, the reluctances of different magnetic materials may be compared if the reluctances of pieces of the materials of the same size and shape are compared. The shape usually used for comparison is a cube of material 1 cm on an edge. The reluctance per centimeter cube is called the **reluctivity.** The reluctance of any larger or smaller piece of the material may be figured from its reluctivity. Thus, from Fig. 4,

$$\text{Reluctance} = \frac{\text{Reluctivity} \times L}{A}$$

where reluctivity = reluctance per centimeter cube.
= 1 for air.
L = length of magnetic circuit in centimeters.
A = cross-sectional area of magnetic circuit in square centimeters.

Example 1. What is the reluctance of an air gap having a cross section of 1.5 cm on a side and a length of 1.6 cm as shown in Fig. 5?

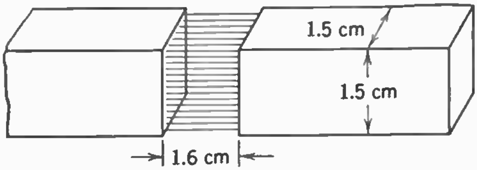


FIG. 5. Typical problem in magnetic circuits: what is the reluctance of the air gap?

Solution:

$$\text{Reluctance} = \frac{\text{Reluctivity of air} \times \text{Length of gap}}{\text{Cross-sectional area of gap}}$$

$$\frac{1 \times 1.6}{1.5 \times 1.5} = \frac{1.6}{2.25} = 0.71 \text{ unit of reluctance}$$

Now, how much magnetomotive force is required to produce a given flux in a given air gap? In a magnetic circuit the relation between mmf, flux, and reluctance is

$$\text{Magnetic flux} = \frac{\text{Magnetomotive force}}{\text{Reluctance}}$$

or, symbolically,

$$\Phi = \mathfrak{F} \div \mathfrak{R}$$

Example 2. In the air gap of Example 1, how many lines of force will be produced by a mmf of 320 gilberts?

Solution:

$$\Phi = 320 \div 0.71$$
$$= 450 \text{ lines of force (approximately)}$$

Magnetic calculations are rarely as accurate as similar electrical calculations, because current can be very accurately and precisely directed by conductors. Lines of magnetic force cannot be exactly and precisely directed and controlled; some of them escape through paths which cannot be calculated.

Note that reluctivity and reluctance are related but are not the same thing, just as resistivity and resistance are not the same in an electric circuit. Reluctivity always refers to a given geometrical configuration, viz., the centimeter cube. The same material drawn out into a fine wire would have a much greater reluctance than it had in the form of the centimeter cube. The analogy here to an electric circuit is clear. The resistivity of a centimeter cube (or a circular mil-foot) of copper is a fixed and known amount. But the resistance of this amount of copper depends upon how much it is drawn out.

4.6. Magnetic field in a solenoid. If a coil of wire is wound on a cardboard cylinder, and if the coil is of such dimensions that the length is greater than the diameter by a factor of 10 or more, and if a current is passed through the wire, it will be found that the number of lines of force (flux) passing through the center of the coil depends upon the number of turns of wire and the magnitude of the current through the wire. It also depends upon the reluctance of the air inside the coil, of course. However, the units of the magnetic quantities are chosen so that the reluctivity of air is equal to 1. Thus

$$\text{Flux} = \frac{\text{Magnetomotive force}}{\text{Reluctance}}$$

$$= \frac{0.4\pi NI}{\mathcal{R}} = \frac{1.26NI}{\mathcal{R}}$$

where N is the number of turns of wire in the coil.
I is the current in amperes.
0.4π is a proportionality factor.

The product of N and I is known as the ampere-turns in the coil, and thus

$$\text{Flux } \Phi = \frac{1.26 \times \text{Ampere-turns}}{\text{Reluctance}}$$

Now the reluctance is equal to the reluctivity times the length of the material in centimeters divided by the cross section of the material in square centimeters. Here, the material on which the coil is wound is air and cardboard, and the reluctivity of each is equal to 1. The flux at the center of a coil of solenoidal dimensions, that is, a coil of the shape described above, is

$$\Phi = \frac{1.26NIA}{L}$$

where A = cross-sectional area in square centimeters.
L = length in centimeters.

One-half the total number of lines of force pass out of the ends of the coil, and the other half constitute the **leakage flux,** that is, the flux which escapes between turns of the coil as seen in Fig. 6. The flux passing out of the ends and the sides of the

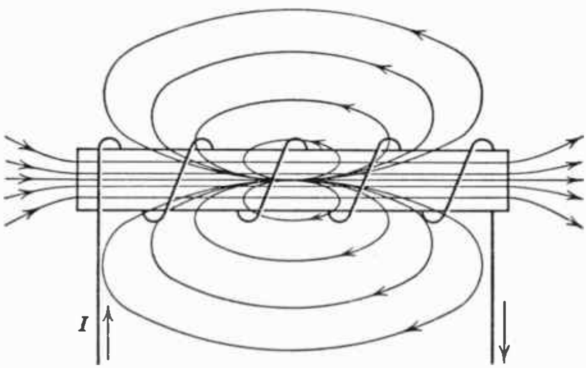


Fig. 6. Lines of force which do not link all the turns of wire make up the "leakage flux."

coil returns through the air space surrounding the coil.

The mmf produced by a current of I amperes flowing through a coil which is long compared to its diameter and having N turns is equal to $1.26NI$ gilberts.

For the special condition that the coil has circular cross section,

$$A = \pi r^2 = \tfrac{1}{4}\pi d^2 = 0.7854d^2$$

where r = radius of core in centimeters.
d = diameter of core in centimeters.

Then

$$\Phi = \frac{0.989NId^2}{L}$$

Problem 1. How many gilberts are required to produce a flux of 1000 lines across an air gap 0.2 cm long and having a cross section of 4 by 4 cm? if the air gap is replaced with a material having a reluctivity of 0.001?

Problem 2. A coil has a rectangular cross section 4 cm by 5 cm. It is 30 cm long and is wound with 1000 turns of wire. What must be the current in the wire to produce 500 lines of flux at the center of the core?

Problem 3. A coil form is 1.5 in. in diameter and 4 in. long. The form is wound with a single layer of No. 24 enameled wire. How many turns of wire will there be on the form? What flux will be produced in the center of the coil if 0.5 amp flows through the wire?

4.7. Relative permeability. The reluctivity of a material is seldom used in design work; in its place the *relative* permeability is employed. Relative permeability is the number of lines of force that will be produced in a centimeter cube of material when the mmf between opposite faces is 1 gilbert. The statement that the relative permeability of a material is 2000 simply means that 2000 times as many lines of force will be produced in it by a given mmf as will be produced in air. The relative permeability of air is chosen as 1 in fixing the size of the fundamental units; that of certain alloys of iron may be as high as many thousand. The term "relative permeability" arises because of the choice of air as the reference material; it is often shortened to "permeability."

In terms of permeability rather than reluctivity the formula for reluctance becomes

$$\text{Reluctance} = \frac{L}{\text{Permeability} \times A} = \frac{L}{\mu \times A}$$

where the Greek letter mu (μ) is the symbol for permeability.

Example 3. A bar of material having a relative permeability of 400 has a length of 6 cm and a cross-sectional area of 4 sq cm. Along the longer dimension is a mmf of 0.24 gilbert. How many lines of force (at the center) will be produced in this bar?

Solution: The reluctance of the bar is

$$\mathcal{R} = L \div \mu A$$

therefore

$$\Phi = \frac{\mathfrak{F}}{\mathfrak{R}} = \frac{\mathfrak{F}}{L \div \mu A} = \frac{\mathfrak{F}\mu A}{L}$$

$$= \frac{0.24 \times 400 \times 4}{6} = 64 \text{ lines}$$

In solving problems in magnetic circuits, one must remember that all the fundamental units are referred to a unit cube of air. Thus 1 gilbert of mmf is required to produce 1 line of force through a length of 1 cm and across 1 sq cm of air, as the relative permeability of air is chosen as 1. If a material has a relative permeability of 400, 1 gilbert will produce 400 lines per square centimeter of cross-sectional area in a piece 1 cm long.

If the circuit is made up of two materials having different reluctances, a solution for each portion of the circuit must be made separately. Thus, if an iron core has an air gap in it, the number of gilberts for the iron portion must be determined, and then the gilberts necessary for the air portion; finally, the two are added together to determine the total magnetizing force, $\mathfrak{F}$. If two or more magnetic paths are in series, the total reluctance will be the sum of the individual reluctances.

If two or more magnetic paths are in parallel, the total reluctance will be the reciprocal of the sum of the reciprocals of the individual reluctances.

Reluctance in series: $\mathfrak{R} = \mathfrak{R}_1 + \mathfrak{R}_2 + \mathfrak{R}_3 + \cdots$, etc.

Reluctance in parallel: $\dfrac{1}{\mathfrak{R}} = \dfrac{1}{\mathfrak{R}_1} + \dfrac{1}{\mathfrak{R}_2} + \dfrac{1}{\mathfrak{R}_3} + \cdots$, etc., or $\mathfrak{R} = \dfrac{\mathfrak{R}_1\mathfrak{R}_2}{\mathfrak{R}_1 + \mathfrak{R}_2}$ when there are only two reluctances in parallel.

Here again the reader should note the similarity to the electric circuit where resistances in series and parallel are combined in exactly the same way.

Problem 4. An iron bar is 2 cm on a side and 20 cm long. It has a relative permeability of 500. What is its reluctance?

Problem 5. A cylinder of iron with a relative permeability of 400 is bent into the form of a ring with a small air gap as shown in Fig. 7. If 20 ma of current flows through the coil of wire wound on the core, how many turns will be necessary to produce 500 lines of flux in the core?

HINT. Determine the reluctance of the two parts of the circuit separately. Add to determine the total reluctance, and then determine the mmf required. Remember that 1 gilbert = 1/1.26 ampere-turn, and that the ampere-turn is the larger unit.

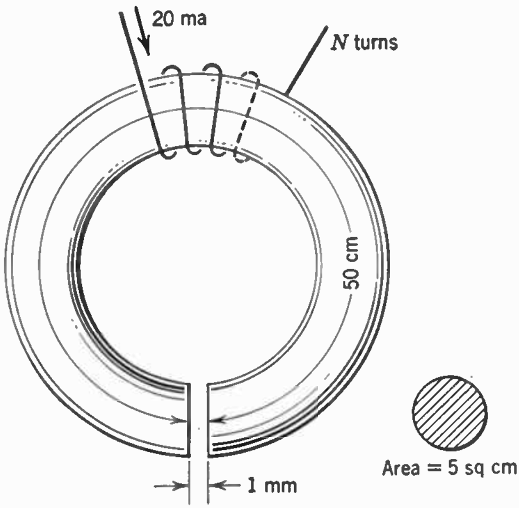


FIG. 7. An iron bar bent in the form of a ring presents an interesting problem in magnetic circuits.

Problem 6. In Prob. 5 how many turns would have been required if there had been no air gap?

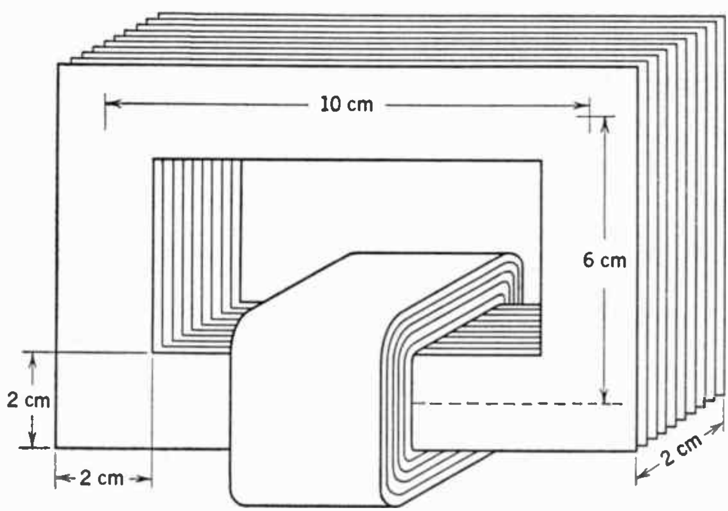


FIG. 8. Typical problem in transformer design.

Problem 7. In Fig. 8 is a drawing of a transformer core. The relative permeability is 400, and 500 turns of wire through which 1 amp flows are wound on the core. Find the magnetizing force in gilberts and the flux in the core.

4.8. Flux density. Since magnetic circuits are made in many shapes and of many magnetic materials, a means of comparing their magnetic characteristics is required. For example, the force necessary to produce a given flux depends upon the length of the magnetic circuit, its cross-sectional area, and its permeability. How, then, can two magnetic circuits made of different materials and having different dimensions be compared? The procedure is to reduce everything to unit dimensions. This method introduces the term **flux density** (B), which is the number of lines per unit area. The unit for flux density is the **gauss,** defined as the number of flux lines per square centimeter.

The flux density in any material, therefore, is the total flux, Φ, divided by the area, A:

$$B = \Phi \div A$$

where B = flux density in gausses.
Φ = the total number of lines of flux.
A = cross-sectional area in square centimeters.

Another unit is the magnetomotive force required to produce a given number of lines through a 1-cm length of a given material. This is called the **magnetizing force.** The unit is the **oersted,** defined as 1 gilbert per centimeter. It is equal to the total mmf in the circuit divided by the length of the magnetic circuit. Thus

$$\text{Magnetizing force } H = \frac{\mathfrak{F}}{L}$$

where H is in oersteds,* $\mathfrak{F}$ is in gilberts, and L is in centimeters. Magnetizing force is often called **magnetic field intensity.**

4.9. *B-H* curves. If a given piece of magnetic material is subjected to varying magnetizing forces, the effects produced, expressed in lines per square centimeter (gausses), may be plotted as a curve. Such a curve is called a ***B-H* curve,** a **saturation curve,** or simply a **magnetization curve.** A typical *B-H* curve is shown in Fig. 9.

* Up to 1930, the oersted was used as the unit of reluctance. Older books, therefore, will use the oersted in this manner. An oersted is 1 gilbert per centimeter. Often it is useful to express the magnetizing force in ampere-turns per inch. The relationship is 1 ampere-turn per inch = 0.495 oersted. The oersted is the larger unit.

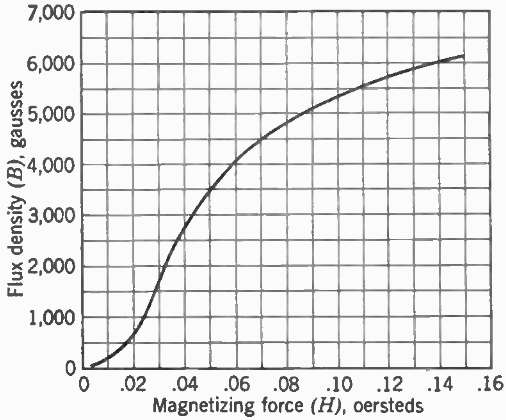


FIG. 9. *B-H* curve of Hypernik.

Since the permeability, μ, of a material is the number of lines of force per unit area set up by 1 unit of magnetizing force, then

$$\mu = B \div H$$

Since B is the flux per square centimeter, the total flux through any area is the product of B and the area A. Thus

$$\Phi = BA$$

From a typical *B-H* curve, one can find the permeability of the material at any magnetizing force H by dividing a value of B by the corresponding value of H. It can be seen at once that the permeability is not a constant but varies with the amount of magnetization. At values of B of the order of 100 or less, the permeability is fairly constant for most magnetic materials, but at higher values of flux density, permeability varies widely. Manufacturers of magnetic material supply *B-H* curves so that engineers can tell the value of H required to produce any specified value of B.

To calculate the mmf required to produce a given number of lines of force through a given length of material, merely multiply the value of H in oersteds (which is the mmf required to produce a given number of lines through a 1-cm length of the material) by the length in centimeters. Thus

$$\mathfrak{F} = HL$$

where $\mathfrak{F}$ is in gilberts, H is the magnetizing force in oersteds, and L is the length in centimeters.

These several magnetic quantities are summarized in the table.

Quantity	Electrical Analog	Symbol	Unit	Definition
Magnetomotive force	Electromotive force	$\mathfrak{F}$	Gilbert	1 line through a reluctance of 1 unit
Flux	Amperes	Φ	Lines of force	Total number of lines
Reluctivity	Resistivity	..	....	Reluctance per centimeter cube
Reluctance	Resistance	$\mathfrak{R}$	....	Reluctivity $\times L \div A$
Permeability	Dielectric constant	μ	....	$B \div H$
Magnetizing force	Volts per centimeter	H	Oersted	1 gilbert per centimeter $= \mathfrak{F} \div L$
Flux density	Amperes per square centimeter	B	Gauss	1 line per square centimeter $= \Phi \div A$

Magnetomotive force ($\mathfrak{F}$) and magnetizing force (H) are causes producing the effects of flux (Φ) and flux density (B) through the opposing quantity reluctance ($\mathfrak{R}$).

Problem 8. What is the flux density in the core in Prob. 7?

Problem 9. A bar of high-permeability ($\mu = 1000$) alloy is 50 cm long, and has a cross-sectional area of 5 sq cm. A flux of 5000 lines is produced in the bar. What magnetizing force, in oersteds, is required? in ampere-turns per inch? What is the flux density?

Problem 10. A bar of Hypernik (Fig. 9) is 20 cm long and has a cross-sectional area of 4 sq cm. What mmf is required to produce a flux density of 3500 gausses?

Problem 11. Using values from Fig. 9, plot a curve of relative permeability vs. flux density for Hypernik. Is your curve substantially the same as that shown in Fig. 10?

4.10. Saturation. In the typical B-H curve of Fig. 9 it is seen that, for low values of magnetizing force H, the flux density rises rapidly; at higher values of H the flux density rises less rapidly, and a point is reached where further increase in H produces no noticeable increase in flux density. At this point the material is said to be **saturated.** Above this point still greater flux can be obtained only by applying very much larger values of H. As a matter of fact, the permeability of materials in this region of saturation may decrease to that of air, which is 1. Curves showing the variation of the permeability of two typical magnetic materials are shown in Fig. 10.

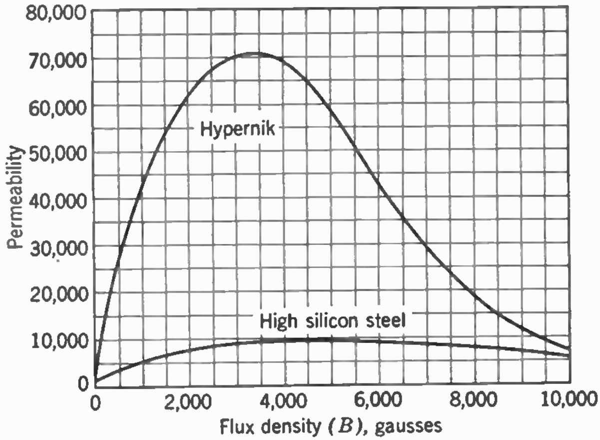


Fig. 10. Curves of two types of magnetic materials, showing how the permeability can be increased by use of alloys and heat treatment.

4.11. Hysteresis. When iron or other magnetic material is operated in an a-c circuit, the magnetizing force is continually changing: increasing to a maximum, then reversing its direction and increasing to a maximum in the opposite direction. If the flux density is measured at points along a complete cycle the result will be a curve like that in Fig. 11. Note that the material initially has no magnetization, and that, as H is increased, B increases steadily, then more slowly up to a final point. Now, if the magnetizing current is decreased in value, the values of flux density will not retrace the upward curve but will trace out a new curve, and when the point is reached at which H is equal to zero some flux remains in the material. This is called **residual flux.** To reduce the value of B completely to zero a considerable magnetizing force must be applied in a direction opposite to that which produced the original flux density. The flux density B seems to lag behind the magnetizing force H.

This lagging effect is called **hysteresis;** it is a result of the fact that not all the energy is usefully employed in producing flux. The area enclosed by the hysteresis curve is actually a measure of the work done in overcoming the molecular resistance to magnetization. This power loss depends upon the material, the flux density, and the frequency of the alternating current.

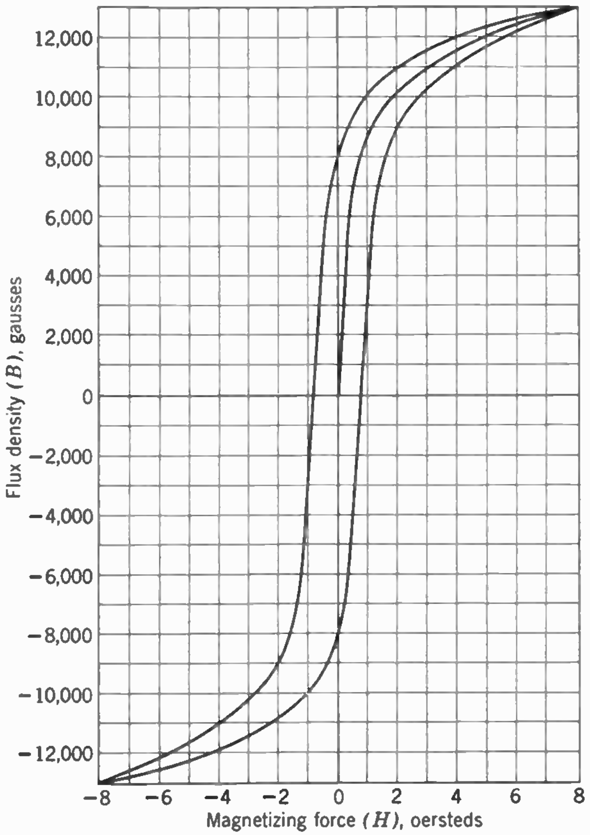


FIG. 11. Typical hysteresis curve. Note that as H in oersteds is increased, B in gausses increases at different rates, and that as H is decreased, B follows a different curve.

Since the loss increases with frequency, care must be taken to choose magnetic materials which are suited for a particular job. Much effort has been expended to develop materials having low hysteresis losses and other desirable properties.

4.12. Magnetic alloys. Iron, nickel, cobalt, aluminum, and a few other elements in various combinations produce magnetic alloys of most interesting and useful properties. Permalloy, for example, has a very high permeability, a magnetizing force of 0.06 oersted producing a flux density of 6000 gausses. Alnico is an alloy widely used for permanent magnets for loud speakers, microphones, and relays. Hypernik is a magnetic material used for power transformers where minimum weight and size are important.

4.13. Magnetic shielding. Since lines of magnetic force "prefer" to flow through iron or other magnetic materials rather than air, it is possible to protect a piece of apparatus from a magnetic field by enclosing it in a box made of iron or other magnetic material. Since the total number of lines of force through the space occupied by the object (a measuring instrument, for example) is relatively constant, placing a magnetic box around it merely shifts the paths of the lines so that most of them thread through the iron rather than through the space in which the object lies. Since the air is still a medium in which lines of flux may exist, it is not possible to exclude all the lines of force from the object, but the number of them that go through it can be greatly reduced by the iron box. Such a box is called a **magnetic screen** or **shield.**

4.14. Faraday's discovery. Another very important discovery that helped bring about the development of present-day electrical machinery was made by the celebrated English experimenter Faraday. The experiment which led to Faraday's discovery may be performed by anyone who has a coil of wire, a bar magnet, and a sensitive current indicator such as a galvanometer. The coil should have such dimensions that the bar magnet may be thrust into its center as shown in Fig. 2. If the terminals of the coil are connected to the galvanometer, it will be found that a current will be indicated when the magnet is thrust into the coil; that no current flows when the magnet is stationary even though inside the coil; that a current opposite in direction to the first is generated when the magnet is withdrawn from the coil; and that the magnitude of the current depends upon the *rate* at which the magnet is pushed into or withdrawn from the coil.

Now what is happening that makes it possible to produce an electric current by a motion of a bar magnet with respect to a coil of wire? The essential phenomenon is the change of position of the magnet and the coil. From the discussion of a previous section, it is known that magnetic lines of force surround the bar magnet. When the magnet is thrust into the coil, these lines of force move with respect to the individual turns of wire in the coil. The lines of force are said to "cut" the coil of wire, and it is this cutting of lines of force that "induces" an electric current in the conductor.

Such was Faraday's discovery. He investigated the matter further, however, and made other discoveries of vast importance.

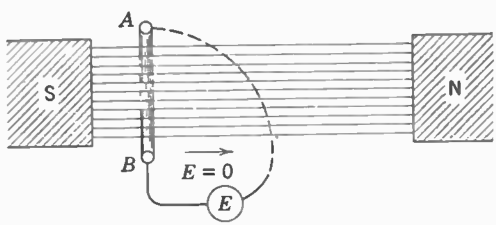
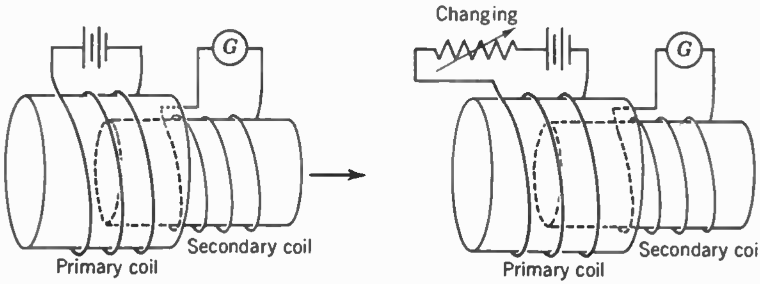


Fig. 12. If A–B is a conductor and if it is moved so that it "cuts" the lines of force between the magnet poles, then a voltage will appear across the ends of the conductor. If A–B moves in the direction of the lines of force, then no voltage will be produced.

For example, if two coils of wire are in close proximity but are not electrically connected (perhaps one coil inside the other as shown in Fig. 13), then if an electric current from a battery flows through one coil (the "primary"), a current will be produced in the second coil (the "secondary") when the relative position of the two coils is changed. The coil carrying the current acts exactly like a bar magnet, inducing a current in the second coil.

Now, instead of actually moving the position of one coil with respect to the other, let the relative positions be maintained but let the current through the primary be altered. As long as the primary current is changing, a current will appear in the secondary coil; but as long as the primary current is constant, no current will be induced in the secondary coil.

The two coils present an exact analogy to the single coil and the bar magnet. The coil carrying the current acts like a magnet; lines of force emanate from it just as they do from the bar magnet. When these lines of force cut the turns of wire of the secondary coil, a current will be produced. If, however, the positions of the coils are not changed, and if an unvarying current flows through the primary, there will be no secondary current. A change in the lines of force with respect to the secondary coil is necessary. This change can be achieved either by moving one coil with respect to the other or by changing the amount of current flowing in the primary. Changing the current changes the number of lines of force and, in turn, the number linking the secondary coil.



(*a*) Secondary Coil Moved (*b*) Primary Current Changed

Fig. 13. Current is induced in the secondary coil if the position of the secondary coil is changed or if the primary current is changed.

There are few scientific discoveries more important than this discovery of Faraday. A large share of the electrical applications of modern times is dependent upon the phenomena of **electromagnetic induction** outlined above.

4.15. The electric generator. An electric generator is a device for converting mechanical energy to electrical energy. A crude device for making this conversion was discussed in the previous section. A more workable arrangement is shown in Fig. 14. It consists of a turn of wire (a conductor) which is rotated mechanically between two magnets. Since induced currents are produced only when the lines of magnetic force are cut by the conductor, it is seen that, as long as the conductor moves parallel with the lines of force, it will not cut any of the lines and no current will be produced. On the other hand, when the conductor moves at right angles to the lines of force, the greatest number of lines will be cut per unit time and the greatest current will be produced.

Let the coil be rotated. At position (*a*), Fig. 14, the conductor moves nearly parallel to the lines of force. No current will flow in the wire. As the wire rotates, it begins to move more and more at right angles to the lines, and more and more current will be produced in the coil, as at (*b*). Soon, however, the wire *AB* begins to approach the position (*a*) when it is again moving parallel with the lines of force, and the current begins to decrease in value.

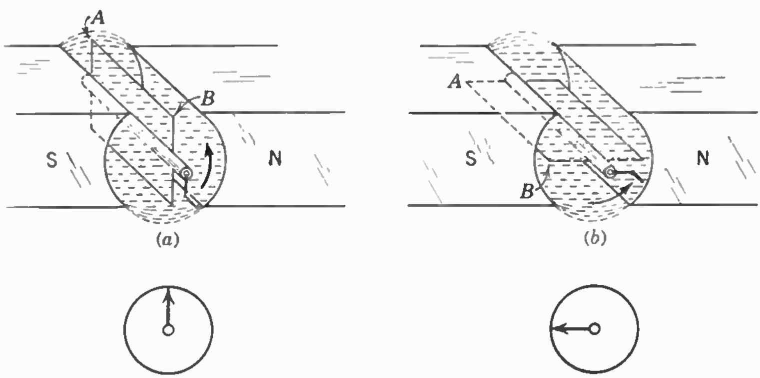


FIG. 14. Elements of a simple electric generator, in this case producing alternating current.

At another portion of the complete rotation, the coil will again be cutting lines of force at right angles, but in a direction opposite to the direction in which it was cutting them at (*b*). Now the current in the coil will be found to flow in the direction opposite to its direction in (*b*).

Twice in the cycle of events (one revolution) there will be zero induced current; twice in the cycle there will be current maxima; these maxima will be in different directions; and between the zero-current positions the current will have values between zero and maximum. Thus the current rises and falls as the coil is rotated in the electric field.

The elements of an electric generator producing alternating current are described above. An actual generator is much more complex. The magnet is replaced by a heavy iron core wound with insulated wire through which a direct current flows, producing a strong magnetic field. The moving coil is wound on a slotted iron form with many turns of wire in the slots.

4.16. Alternating-current definitions. Since a circle, like a compass, can be divided into 360°, the position of the rotating member of a generator may be described in terms of the angle through which it has moved rather than by "position (*a*)" or "position (*b*)," or the position may be described in terms of the time it has been moving since the start of its rotation. If it makes a complete rotation it has gone through 360°; if it has turned half way through a complete revolution it has turned 180°; and so on.

If the current at each position is plotted against the angle of rotation, expressed in degrees, a curve like that in Fig. 15 will result. Note that the current rises and falls about its zero-current value.

One complete revolution is known as **a cycle.** The number of cycles per second is known as the **frequency** of the induced voltage or current. The time required for a complete revolution is called the **period.**

Frequency = Cycles per second.

$$\text{Period} = \text{Time of one cycle} = \frac{1}{\text{Frequency}}.$$

Although frequency is correctly rated in cycles per second, abbreviated as **cps,** engineers seldom use the "per second" part of the expression, the assumption being that it is the number of complete changes of current direction per second that is being considered. In this book the simple term "cycles" is used rather than cycles per second.

4.17. Work done by alternating current. Some may wonder whether or not a current that is continually reversing its direction—never getting anywhere, so to speak—is useful. It is.

Consider a board that is to be sawed. It matters little whether a circular saw or a common handsaw is used. If a circular saw is used, the cutting teeth move continuously in one direction. This is analogous to direct current. If a handsaw is used, the saw is moved back and forth or up and down. Nevertheless, the board is sawed. This is analogous to alternating current.

The alternating current which lights homes in the United States is usually 60-cycle current. In some few communities 25- or

50-cycle current is supplied. In radio installations for use on shipboard and on aircraft, generators often produce 400- or 500-cycle current. At the high-power radio stations of the Radio Corporation of America at Rocky Point, Long Island, are huge alternators which generate radio-frequency currents with frequencies of 20,000 cps. Smaller generators which produce frequencies as high as 100,000 cps have been built for radio transmitters, but these have been supplanted by vacuum-tube generators. By means of vacuum tubes, alternating currents having frequencies of many millions of cycles may be generated.

4.18. Direct-current generator. When current is taken out of a generator by means of **collector rings,** illustrated in Fig.

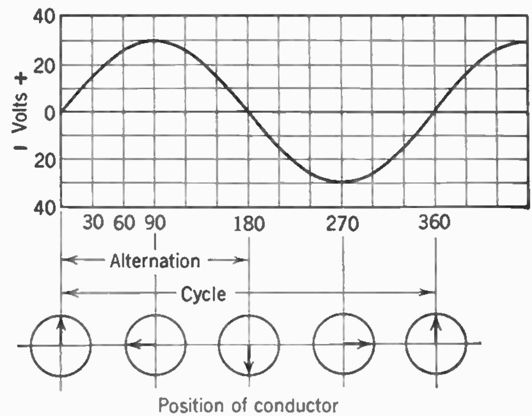


FIG. 15. How the current rises and falls in the output of an a-c generator. It actually follows a "sine" wave curve.

16, the resulting current is alternating, as shown in Fig. 15. If current flowing continuously in the same direction is desired, a device called a **commutator** is used instead of the collector rings. A commutator is a switch, or valve, which keeps the output current flowing in the same direction by reversing at the proper time the position of the external wires with respect to the rotating wire. In this manner the current flows through the external circuit in a single direction, although the current in the conductor which cuts the lines of force must reverse each time the conductor passes through 180° and reverses its direction with respect to the direction of the magnetic field.

The commutator serves the same purpose as a valve in a pump which may be employed to keep water flowing upward whether the pump handle is worked down or up. A machine that sends out current that flows in a given direction is called a **direct-**

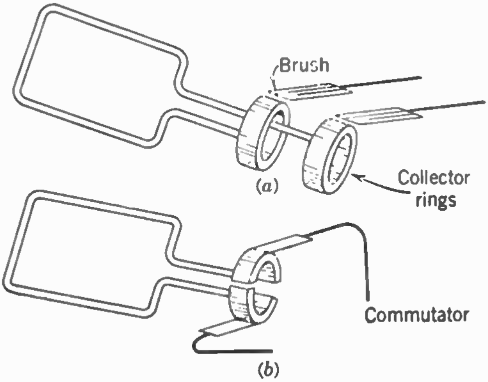
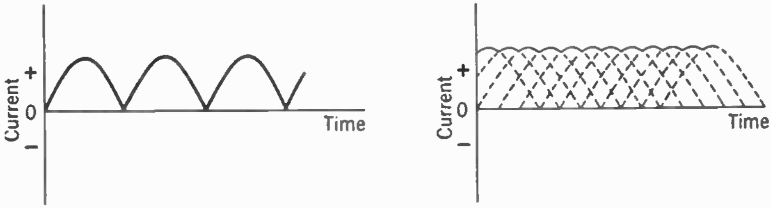


Fig. 16. An alternator has collector rings for taking the current out of the machine; a d-c generator has a single ring with segments which reverse the current as the rotor moves so that, in the external circuit, current always flows in one direction.

current (d-c) generator, and naturally, the current is known as direct current.

The current produced in the simple d-c generator described above, while flowing in the same direction all the time, is pulsat-



(a) Pulsating Direct Current from a Single Coil and Commutator

(b) Direct Current Produced by a Typical D-C Generator Employing Many Coils

Fig. 17. Pulsating nature of the output of a d-c generator.

ing in nature, as shown in Fig. 17(*a*). In normal practice many coils are wound on the rotating core of the generator, and these are connected to a complex commutator such that the output current looks more like that pictured in Fig. 17(*b*).

5 · Inductance

5.1. Coupled circuits. Consider the two coils P and S in Fig. 1. These coils are said to be "coupled" when lines of magnetic force from one coil go through the other. When P is attached to a battery and the switch is closed, there is a momentary movement of the needle of the meter which is connected across S. The needle then returns to zero. When the switch is opened, the needle moves again but in the opposite direction. As long as the current in the primary P is steady in value and direction, there is no movement of the needle. A deflection of the needle indicates a momentary flow of current in the secondary coil; this current flows only when the primary current is changing, that is, starting or stopping, not when the primary current is fixed in value and direction.

5.2. Lenz's law. Two fundamental facts about this phenomenon of coupled circuits should be noted. The first is that, when lines of magnetic force couple two coils together, and some change in these lines takes place, a voltage is induced in the circuit. The change may result from a variation of current in the primary coil, or from relative movement of the two coils. The second fundamental fact is expressed in *Lenz's law:* the induced current is in such a direction that it opposes the change that produced it.

When the battery is attached to P by closing the switch, lines of force begin to thread through the turns of wire in S. This movement of the lines of force through S induces a voltage across the coil, and a current flows in the coil and in the apparatus connected to it. The current in the second coil is in such a direction that its field, that is, its lines of force threading through the primary, induces a counter-voltage in the primary opposite in direction to the battery voltage.

When the battery is disconnected by opening the switch, the lines of force from the primary current collapse back on the primary coil and, in crossing the secondary turns in a direction opposite to that taken when the current in the primary is increasing, induce a voltage in the secondary in such a direction that it tends to keep the primary current flowing.

If it were not for the phenomenon expressed in Lenz's law, electrical equipment would burn up. If the current induced in

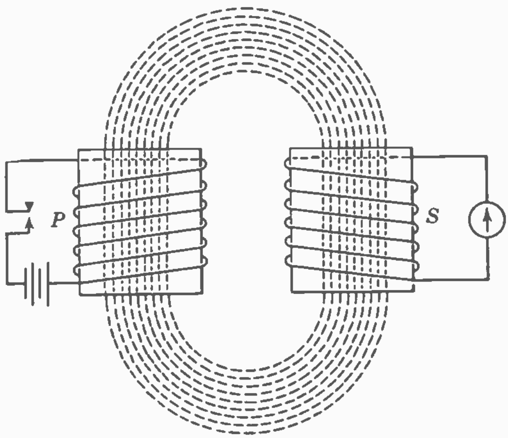


FIG. 1. If the coils P and S are coupled so that lines of force from P go through S, a current indicator will show a flow of current in S when the key is momentarily closed or opened.

the secondary by a change of lines of force from the primary produced a voltage in the primary in the same direction as the battery voltage, currents in both the primary and secondary would increase indefinitely, and these excessive currents would destroy the apparatus.

If a large coil, a battery, and a quick-acting current meter are placed in series with a switch, the meter will be seen to reach its final value slowly and not at the instant the switch is closed. If the battery were suddenly removed from the circuit after a steady value of current had been reached, the current through the coil would not fall to zero instantly; it would take as long to decrease to zero as it took to rise to its final "steady-state" value. The actual time of build-up may be quite short, a fraction of a second, but in large iron-core coils, such as the field of a large generator, several seconds may be required for the final value of current to be attained.

The fact that current takes longer to reach its final value in a circuit in which there is a coil of wire indicates that something about the coil tends to prevent any change in the current. This phenomenon is of fundamental importance.

The property of an electrical circuit which tends to prevent any change in current flowing through it is called *inductance.* It has a mechanical analogy in *inertia.* A flywheel requires considerable force to get it up to speed, and after it is started it will continue to run for some time after the driving force is removed. It does not stop suddenly. Considerable force is required to stop it, and the more rapidly one wants to stop it, the more force he must apply.

Inertia is evident in a mechanical system only when some change in motion is attempted. It is not the same as friction, which is always present. The inductance of an electrical circuit manifests itself when the current through it is changing. It is not to be confused with resistance, which is always present. Current flowing in a circuit containing only resistance stops as soon as the driving force (voltage) is removed. If inductance is added in series to the circuit, the resistance remaining the same, a measurable time will be required for the current to drop to zero when the voltage is removed.

5.3. Self-inductance. Inductance may be added to a circuit by winding a length of wire into a compact coil. For example, 1000 ft of No. 20 copper wire strung up on poles would have a resistance of about 10 ohms, and the current into it would reach a final value very soon after a battery was connected to its two ends. If, however, the wire is wound up on a spool, the time required for the current to reach its final value will be considerably longer. The resistance has not changed; the longer time is due to the inductance added to the circuit.

A single coil can have inductance and can have a voltage induced across its terminals just as though it were the secondary coil shown at S in Fig. 1.

When the current starts to flow through the coil, lines of force begin to thread their way through the coil, thereby cutting adjacent turns of wire and, according to Lenz's law, inducing in each turn a voltage in such a direction that it tends to oppose the building up of the current from the battery. When the battery connection is broken, these lines of force collapse, again cutting the turns of wire and thereby inducing voltages which tend to keep the battery current flowing.

5.4. Magnitude of inductance and induced voltage. The greater the number of turns of wire in a small space, or the higher the permeability of the core on which the wire is wound, the

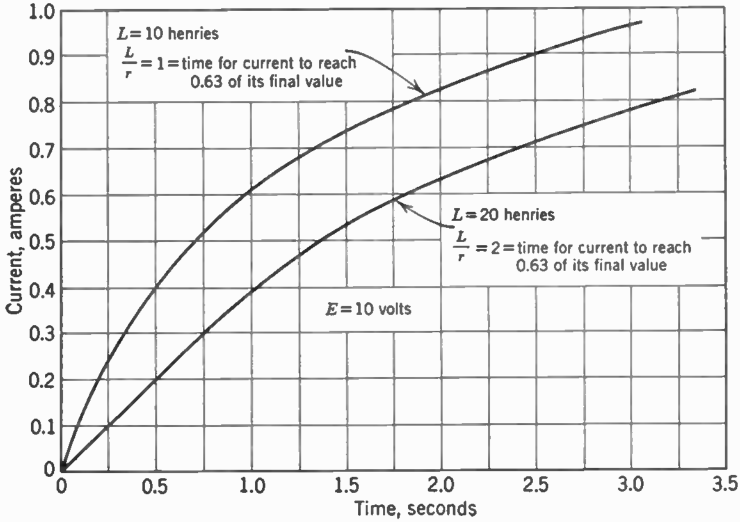


FIG. 2. Manner in which current through a circuit containing inductance and resistance builds up at a rate dependent upon L/R.

greater will be the inductance of the coil and the longer the time required for the current to reach its final value. The inductance of a coil is proportional (approximately) to the square of the number of turns on the coil; the induced voltage is proportional to the inductance and the rate of change of current through the coil. The time required for the current through a coil to reach its final value increases as the inductance is made larger and decreases as the resistance of the coil becomes larger. The general manner in which the current through an inductance and resistance in series increases is shown in Fig. 2. Since the current tends to keep flowing when an inductive circuit is broken, the induced voltage across the coil must be in the same direction as the battery voltage. Thus, if a coil has 100 volts from a battery across it, and the current is suddenly broken, the voltage at the instant of break across the coil will be 100 plus additional induced voltage. Breaking the current in a highly inductive circuit may set up a very large voltage across the coil. The spark which bridges the switch contact when the circuit is broken is an indication of a high voltage. This is a practical demonstration of Lenz's law.

It must be borne in mind that there is no induced voltage whatever so long as the current through the coil does not change. The instant it changes, there will be a "back emf" across the coil which opposes the change in current. The larger the coil, and the quicker the current is increased or decreased, the greater will be the induced or back emf.

The *rate of change*, not the absolute value of the current, is the important factor. When current is measured in amperes and time in seconds, the rate of change is measured in amperes per second. Thus, if the current is 1 amp and it drops to zero in 1 sec, the rate of change is 1 amp per sec. If it drops to zero in $\frac{1}{100}$ sec, the rate of change is 100 amp per sec.

A coil with a large inductance may, when the circuit to it is broken, produce a large enough voltage and be able to supply enough current to be dangerous to a person who comes in contact with the terminals. Therefore, care should be exercised when handling large inductors.

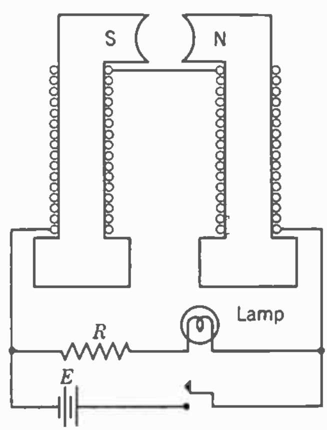


Fig. 3. An experiment illustrating Lenz's law and the effect of breaking a highly inductive circuit.

Example 1. In Fig. 3 is an inductor made up of the field winding of a generator. Across the inductor are placed a flash lamp, resistor, and battery as shown. The current is adjusted by varying the resistor R so that it is just insufficient to light the lamp. Now when the switch is opened the lamp will suddenly light (and may burn out) because of the momentary increase in voltage due to the collapsing field of the coil.

In radio circuits the coils are usually wound on non-magnetic or specially treated powdered iron cores. In audio * and power circuits iron is utilized to build up large inductances in small spaces and with a minimum of copper wire.

If an a-c voltage is placed across a coil, the current through the coil will be much less than if a d-c voltage of the same value is placed across it. This is due to the counter-voltage or back emf induced across the coil by the effects just described.

5.5. The unit of inductance. When a current through a coil changes at a rate of 1 amp per sec and causes an induced voltage of 1 volt across the coil, the inductance of the coil is said to be 1 **henry.** This unit is named in honor of Joseph Henry, an American experimenter who discovered the phenomenon of electromagnetism at about the same time as Michael Faraday.

The qualitative facts of electromagnetism may be expressed as

$$E_{av} = L\frac{\Delta I}{\Delta t}$$

where E_{av} is the average induced voltage.

Δ is the Greek letter delta, indicating "a change in."

I = current in amperes.

t = time in seconds.

The portion of the expression $\Delta I/\Delta t$ may be translated as "the rate of change of current with respect to time." The induced voltage is, therefore, proportional to the rate at which the current changes and is equal to amperes change per second multiplied by a factor L known as the inductance.

Coils added to circuits for the purpose of increasing the inductance are properly called **inductors,** although many engineers use the terms "inductance" (a property of a circuit) and "inductor" (a piece of apparatus) interchangeably.

The inductance of a coil is really a measure of the flux lines linking with the turns of the coil. The unit of inductance, the **henry** (h), is that inductance which causes 10^8 (one hundred

* Audio circuits involve alternating currents which, when converted into sound waves as by a loud speaker, can be heard by the human ear, that is, roughly from 50 to 15,000 cycles. Radio frequencies are much higher and are inaudible.

million) interlinkages per ampere change in current. Thus

$$L = \frac{N\Delta\Phi \times 10^{-8}}{\Delta I} \text{ henries} \qquad [1]$$

where N = the number of turns on the coil.
Φ = magnetic flux linking these turns due to I.
I = amperes of current through the coil.
10^{-8} = a constant term to bring L into practical units.

In Chapter 4 it was determined that the flux in a long coil, or solenoid, is given by

$$\Phi = \frac{0.4\pi NI}{\mathcal{R}}$$

where $\mathcal{R}$ is the reluctance.

$$\mathcal{R} = l \div \mu A$$

where l = length of the winding in centimeters.
A = area of cross section in square centimeters.
μ = permeability of core on which coil is wound.

Therefore,

$$\Phi = \frac{0.4\pi NIA\mu}{l} = \frac{1.26NIA\mu}{l}$$

If this value of Φ is inserted in formula 1, an expression for the inductance in henries of a coil whose length is large compared to its diameter will be found. Thus

$$L = \frac{1.26N^2A\mu}{10^8 l}$$

where μ = permeability (μ = 1 for air).
A = cross section of coil in square centimeters.
l = length of winding in centimeters.

5.6. Energy stored in a magnetic field. Since it is a property of inductance to oppose any change of current, it follows that energy is required to increase the current through an inductance. As would be expected, the amount of energy required depends upon the final value of current, providing, of course, that the initial current is zero. The expression for the amount of energy in the magnetic field around an inductance is given by

$$W = \tfrac{1}{2}LI^2 \text{ joules}$$

where L is in henries and I is in amperes.

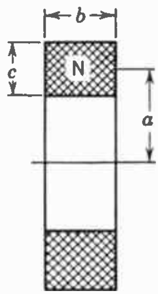
If an inductor is carrying a current and the connections to it are suddenly opened, the magnetic field must give up its energy since the current must soon drop to zero. As the circuit is opened a voltage is developed across the terminals of the inductor large enough to cause a spark to bridge the opened terminals. The energy of the magnetic field is then dissipated in this spark and in the resistance of the circuit.

The high voltage which appears when the circuit to a current-carrying inductor is opened is taken advantage of in many electrical applications. For example, fluorescent lamps will not start on normal line voltage, although they operate satisfactorily on this voltage once they are started. The starting mechanism normally consists of an inductor (ballast) and a starter switch. The starter switch, when closed, permits current to flow through the filaments at the two ends of the lamp and through the ballast. In a few seconds, after the filaments are warm, the starter switch is opened and the "inductive kick" of voltage is enough to cause the lamp to light. The ballast is usually arranged so that it remains in the circuit in series with the lamp and therefore the current taken by the lamp is limited to the proper value. The filaments are not used except during the starting period. The starter switch may operate on an automatic basis, or it may be a manual switch.

5.7. Typical inductors. The coils used in radio apparatus vary from small coils having inductances of the order of microhenries (μh) to very large coils having 100 or more henries of inductance. Broadcast-frequency tuning coils have inductances of about 300 μh and may be from ½ to 3 in. in diameter wound with No. 30 to No. 20 wire of 50 to 100 turns or so. There are a number of complicated formulas by which one can calculate the inductance of coils of various sizes and shapes. The ones in Fig. 4 are accurate enough for practical purposes.

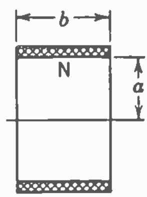
These formulas show that the inductance increases as the square of the number of turns. Thus, if a coil of 3 units inductance has its number of turns doubled, the inductance will have increased four times or to 12 units. This is true provided that there is good coupling between turns; that is, if the coil is on an iron core this rule is quite accurate, but if the coil is wound on a core of air the rule is only approximately true. It becomes more nearly a fact the closer together the turns of wire are placed.

Problem 1. A coil like that shown in Fig. 4(*a*) is composed of 1000 turns wound on a coil form such that *b* and *c* are both ¾ in. and *a* is 2½ in. What is the inductance of the coil?



$$L = \frac{0.315a^2N^2}{6a + 9b + 10c}\ \mu\text{h}$$

(*a*)



$$L = \frac{0.394a^2N^2}{9a + 10b}\ \mu\text{h}$$

(*b*)

Fig. 4. Typical coil forms and formulas by which the inductance may be calculated. Dimensions are in centimeters.

Problem 2. Calculate the inductance of a coil such as shown in Fig. 4(*b*) if it is made of 100 turns wound on a form ½ in. in diameter and 1 in. long. This coil is similar to many used in modern broadcast receivers.

Problem 3. A coil form is 2 in. in diameter and 3½ in. long. A single layer of No. 24 copper wire is wound on the form. If the turns are wound as closely as practicable, what will be the inductance of the coil? What will be the approximate d-c resistance?

Problem 4. A power supply filter inductor has an inductance of 8 henries and initially is carrying a current of 0.2 amp. If the current is decreased to zero in 0.1 sec, what average voltage is induced across the terminals of the inductor?

Problem 5. An inductor has a core of annealed sheet iron which has a relative permeability of 3000. The core has a cross-sectional area of 6 sq cm, and the mean length of the core is 30 cm. It is wound with 1000 turns of wire. What is the inductance?

5.8. Coupling. The closer together the two coils, *P* and *S*, Fig. 1, the greater the number of lines of force produced by the primary current that link with the turns of the secondary, and the better the **coupling** is said to be. Also, the higher the permeability of the medium in which the lines go, the better the coupling.

The voltage across the secondary of such a two-coil circuit as that shown in Fig. 1 depends on the inductances of both coils, their proximity, the permeability of the medium between the coils, and the rate at which the primary current changes. All the factors except the rate of change of the primary current are grouped together and are called the **mutual inductance** of the circuit.

The secondary voltage, then, is equal to

$$M \times \text{Rate of change of primary current with time}$$

where M is the mutual inductance and is rated in henries.

5.9. Magnitude of mutual inductance. Formulas in Fig. 4 show that the inductance of a coil depends upon the square of the number of turns. Doubling the turns increases the inductance four times. Consider two coils built alike and having the same inductance. If they are connected in series as in Fig. 5(*a*), and

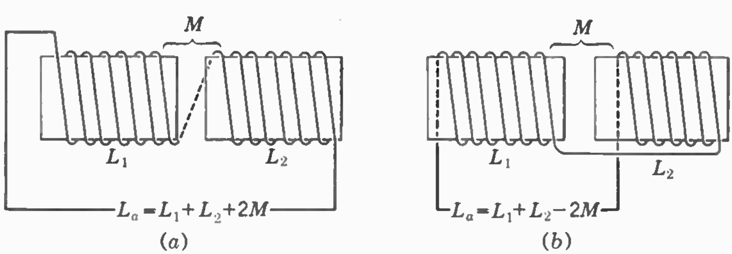


FIG. 5. Coils may be connected so that their fields aid or "buck." In one (*a*) the total inductance is increased; in the other (*b*) the total inductance is less than the sum of the component inductances.

if the coils are perfectly coupled, the total inductance would be equal to that of a single coil of double the number of turns. In other words, the total inductance of two coils connected "series-aiding" and with perfect coupling will be four times the inductance of a single coil. If the connections to one coil are reversed, the total inductance will be zero because the lines of force from one coil will encounter the lines of force from the other coil, which are in the opposite direction. The coils are now connected "series-opposing."

Consider the series-aiding case shown in Fig. 5(a). The total inductance is made up of the inductance of coil 1, that of coil 2, the mutual inductance due to the lines of force from coil 1 which go through coil 2, and the mutual inductance associated with the lines from coil 2 which go through coil 1; these two mutual inductances are equal even though the coils are not identical. Thus,

$$L_a = L_1 + L_2 + 2M \qquad [2]$$

If the two coils are identical, $L_1 = L_2$, and

$$L_a = 2L_1 + 2M$$

and if the coupling is perfect, $2L_1 = 2M$, and

$$L_a = 4L_1$$

Under these conditions, if the coils are connected so that their fields oppose each other, the net inductance will be zero. Thus,

$$L_a = 2L_1 - 2M$$

and since $2L_1 = 2M$,

$$L_a = 0$$

Now suppose that some of the lines from one coil do not link with the turns of the other coil (this is always the case in reality). Now L_a will be less than it would be if the coupling were perfect, and L_o will be greater than zero. Under any conditions, however,

$$L_a = L_1 + L_2 + 2M \quad \text{(connected series-aiding)}$$

$$L_o = L_1 + L_2 - 2M \quad \text{(connected series-opposing)}$$

These two formulas give us an easy way to measure the mutual inductance between two coils. Connect them series-aiding and measure the total inductance; then connect them series-opposing and measure the total inductance. Now subtract L_o from L_a. Thus,

$$\begin{aligned} L_a - L_o &= L_1 - L_1 + L_2 - L_2 + 2M - (-2M) \\ &= \quad 0 \quad + \quad 0 \quad + \quad 4M \end{aligned}$$

The mutual inductance, therefore, will be ¼ of the difference between L_a and L_o.

Now it will be useful to have some measure of the amount of coupling between two coils.

If two coils are connected series-aiding so that complete flux linkage results, the total measured inductance will be found to be equal, not only to equation 2 but also to

$$L_a = L_1 + L_2 + 2\sqrt{L_1 L_2}$$

and so for complete coupling

$$M = \sqrt{L_1 L_2}$$

The expression $M/\sqrt{L_1 L_2}$ is called the **coefficient of coupling** (k) since it is a measure of the completeness of the coupling between the individual coils. Since

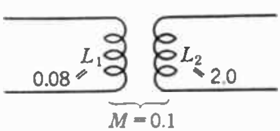
$$\frac{M}{\sqrt{L_1 L_2}} = k$$

$$M = k\sqrt{L_1 L_2}$$

and for complete coupling $k = 1$.

Values of k vary from 5 per cent (0.05) or less in the case of air-core coils to as high as 98 per cent (0.98) where iron is used as the core.

The mutual inductance depends upon the two coils L_1 and L_2, and the coupling between them, that is, $M = k\sqrt{L_1 L_2}$; the coefficient of coupling between the two circuits depends upon the closeness of the two coils and on the medium separating them. An example is shown in Fig. 6. If the medium is iron, which makes it easy for the flux from one coil to link the other, the coefficient of coupling will be high. The maximum possible value of k is 1.0. This value, called ***unity coupling***, is approached in iron-core transformers. Of course, if the coils are situated so that there is no coupling, either because they are too far apart or because they are placed in iron boxes, the total inductance in the circuit will merely be the sum of the individual inductances.



$$k = \frac{0.1}{\sqrt{0.08 \times 2.0}} = 0.25$$

FIG. 6. Example showing dependence of coefficient of coupling, k, on mutual inductance.

5.10. Measurement of inductance. Inductance may be measured, like resistance, by means of a Wheatstone bridge. The Wheatstone bridge method consists essentially of comparing the unknown inductance to a known inductance. Resistances are used as the ratio arms, A and B in Fig. 7. When there is no

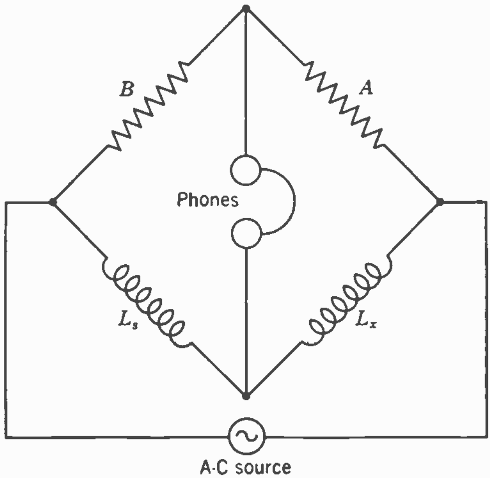


Fig. 7. Wheatstone bridge arranged to compare inductances. Here L_s is a known "standard" inductance to which the unknown inductance, L_x, is to be compared. One of the ratio arm resistances must be adjustable unless L_s is adjustable. The bridge is balanced when no sound is heard in the telephones, and at this adjustment the unknown inductance can be determined by the formula in the text.

sound in the headphones the inductances are equal if the ratio arms are equal; if the ratio arms are not equal the unknown inductance is given by the equation

$$L_x = L_s \times \frac{A}{B}$$

Mutual inductance is measured on a bridge by the following method: The inductance of the individual coils may be measured first by the method outlined above. Then the two coils are connected series-aiding, and the total inductance is measured. This permits the calculation of M from the formula $L_a = L_1 + L_2 + 2M$. The same general result could be obtained by connecting the coils series-opposing. It is not necessary to measure the individual inductances first, provided the inductance when the coils are connected series-aiding (L_a) and then series-opposing (L_o) can both be measured. Then

$$4M = L_a - L_o$$

$$M = \frac{L_a - L_o}{4}$$

Problem 6. Two coils each have a self-inductance of 200 μh. Their coefficient of coupling is 0.2. What is the mutual inductance? What is the total inductance if the coils are connected series-aiding? if they are connected series-opposing?

Problem 7. Two identical coils are wound on a common coil form. The coils are connected in series. The total inductance with one series connection is measured and found to be 300 μh. The connections on one coil are reversed and the inductance is now found to be 220 μh.

(*a*) What is the mutual inductance?

(*b*) What is the coefficient of coupling?

Problem 8. A radio-frequency pentode-tube amplifier uses a transformer to couple from one tube to the following stage. The primary has an inductance of 250 μh, and the secondary an inductance of 200 μh. The mutual inductance is 90 μh. What is the coefficient of coupling?

Problem 9. A laboratory inductor has two coils, one of which rotates inside the other. Each coil has a self-inductance of 200 mh. The coils are connected in series. What must be the mutual inductance and the coefficient of coupling when the total inductance is (*a*) 50 mh? (*b*) 500 mh?

Indicate for each case whether the coils are series-aiding or series-opposing.

5.11. The transformer. A transformer is merely a coupled circuit. It may have a primary and a single secondary coil, or it may have two or more secondary coils. Transformers for use on power-line frequencies (60 cycles) are commonly employed to transform one voltage into another. Transformers for audio-frequency use also perform this voltage-transforming function. However, transformers for radio-frequency circuits are generally used in conjunction with capacitors to obtain desirable frequency-response curves, and their voltage-transforming properties are of secondary importance. This and the following sections of this chapter will discuss transformers used on power-line and audio-frequency circuits. A discussion of radio-frequency transformers will be found in Chapter 15.

A simple power or audio-frequency transformer consists of two windings on an iron core, as in Fig. 8. The purpose of the iron core is to insure that the magnetic flux set up by current flowing in the primary links with the secondary. However, even with iron cores some flux produced in the primary does not link with the secondary turns. The lines of force (flux) which do not link with the secondary are called **leakage lines,** and the inductance associated with them is called **leakage inductance.**

The **primary** coil of a transformer is the coil which is connected to the source supplying power to the transformer. The **secondary** coil is connected to the device requiring power. For example, the primary of a door-bell transformer is connected to the house-lighting circuit. The secondary is connected to the

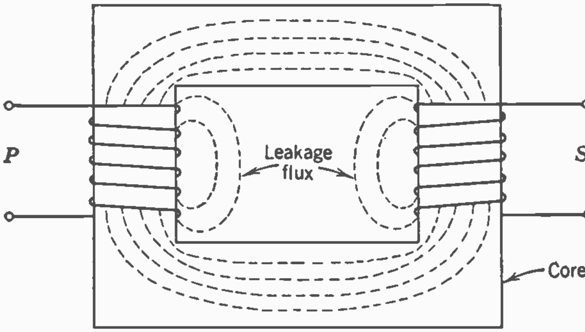


FIG. 8. Lines of force which do not link both windings are of no value; they are called, collectively, the leakage flux.

door bell which requires power. The relative size of the voltages on the two coils does not determine which is the primary and which is the secondary.

Since power and audio-frequency transformers are designed so that the amount of leakage flux is small (coupling is nearly unity), it is possible to establish a simple relation between the primary and secondary voltages. If there are twice as many secondary turns as primary turns, the secondary voltage will be twice the primary voltage. The turns ratio will be 1:2 or ½. The relation between primary and secondary turns and the respective voltages is:

$$\frac{E_p}{E_s} = \frac{n_p}{n_s} = N \quad \text{(turns ratio)} \qquad [3]$$

This simple relation is subject to slight correction for practical transformers because the coupling is actually slightly less than 1.0 and because of voltage drops in the winding resistances.

However, these corrections are generally small and may usually be neglected for power and audio transformers.

By using the proper ratio of turns, voltages either greater or less than the primary voltages may be secured at the secondary terminals.

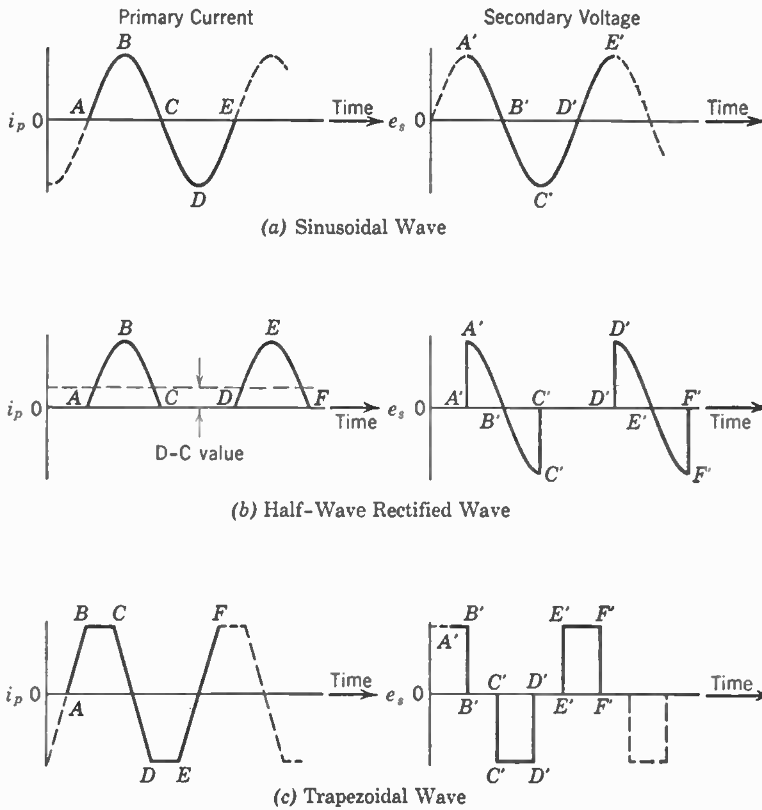
5.12. Waveforms in transformers. Since many transformer applications utilize voltages with complex waveforms, it is desirable to determine how these waveforms are changed in transformation. The relation

$$e_s = M \times \text{Rate of change of primary current with time}$$

can be used to determine the voltage induced in the secondary for a given waveform of current in the primary.

Several typical waveforms are shown in Fig. 9. Figure 9(*a*) illustrates waveforms for a sine-wave primary current such as would usually be the case in transformers operating from power lines. Suppose that the circuit has been connected long enough for conditions to have been established. In the interval from *A* to *B* the current is increasing with time. Its *rate of change* is positive. Thus it induces a positive voltage in the secondary. However, the current increases less rapidly (its *rate of change* is less) as point *B* is approached and the secondary voltage, while still positive, becomes smaller, and is zero when the primary has the value at point *B*. From *B* to *C* the current is decreasing (*rate of change* is negative) and the secondary voltage is negative. At point *C* the rate of decrease is greatest and so the negative voltage is also the greatest. Between *D* and *E* the current is again increasing (*rate of change* is positive) and the secondary voltage is positive. It is interesting to note that, with a sinusoidal waveform of primary current, the induced voltage has the same shape as the primary current, but the wave is displaced along the time axis so that secondary voltage maxima occur at the same time as primary current minima, and vice versa.

The reasons for the difference in time between the maximum primary current and the maximum secondary voltage are as follows: The maximum secondary voltage will be produced when the primary current is changing at the greatest rate, that is, the greatest amperes change per second. At the time of maximum primary current, the current *changes* hardly at all, as its curve is rather flat on top. At this point the secondary voltage will be a minimum. When, however, the primary current goes through zero, its direction changes from positive to negative and its rate of change (slope) is maximum. At this point the greatest secondary voltage will be produced.



(a) Sinusoidal Wave

(b) Half-Wave Rectified Wave

(c) Trapezoidal Wave

Fig. 9. Waveforms of primary current i_p and secondary voltage e_s in transformers.

Consider a pendulum swinging freely. When it reaches a vertical position, its bob hanging straight down, it is moving at its maximum velocity—that is, its rate of change of distance with time is greatest. Its position in space is about to change from, say, right to left. Now, when the bob reaches its maximum distance from the vertical, it is momentarily at a standstill and its velocity or speed (change of distance with time) is least.

If the bob were magnetized and if a coil of wire connected to a current meter were so placed that the lines of force from the bob flowed through the coil, it would be found that the greatest current would be produced when the pendulum was moving at its maximum speed and that the minimum current (zero current, actually) would correspond with the greatest distance the bob reached on its flight; that is, at the point where its velocity was zero.

The waveform of the secondary voltage can always be predicted by remembering that the least voltage will be produced when the rate of change of the primary current is least, and the greatest voltage will be produced when the primary current is changing at the greatest rate with respect to time. Also, the secondary voltage will be positive when the rate of change of primary current is positive; it will be negative when the rate of change of primary current is negative.

Figure 9(b) shows a primary current such as might be obtained from a half-wave rectifier. The waveform is the same as that of Fig. 9(a) except that the lower half is missing. The induced voltage for the positive half wave of current is the same as for the first illustration. However, from C to D the rate of change of current is zero and there is no voltage induced in the secondary. The second positive loop of current is a repetition of the first, and so the induced voltage from D' to F' has the same waveform as that from A' to C'. It is important to note that, whereas the primary current has a d-c value as marked on the figure, the induced voltage in the secondary is centered about the axis and has no d-c value. This is an illustration that transformers do not transform d-c voltages. If the primary current contains both d-c and a-c components, only the a-c component is effective in producing a secondary voltage.

Figure 9(c) shows a primary current with a trapezoidal waveform. Suppose that the analysis is started at point A. Here the current is increasing from A to B and produces a positive voltage in the secondary. The value of current is constant from B to C (*rate of change* is zero) and so the voltage induced is zero during this interval. From C to D the current is decreasing (*rate of change* is negative) and so the secondary voltage is negative. From D to E there is again no change in the value of the current and the secondary voltage is zero. Note that it makes no difference if the interval during which the rate of current change is zero is along the axis as from C to D in Fig. 9(b), or if the rate of change of current is zero at some constant current value such as from B to C in Fig. 9(c). In either case the induced voltage is zero during the interval.

The last two illustrations show the secondary voltage changing instantaneously from one value to another. These are idealized cases. In a circuit containing an actual transformer a finite time is required for this change, even though the time may be very short if the transformer is properly designed.

The primary current and primary voltage will have approximately the same waveform if the load on the secondary is a resistor. However, if the secondary load includes a capacitor or inductor, the waveform of primary voltage and current may be quite different. A more exhaustive study of the properties of capacitors and inductors is required before their effect on these waveforms can be treated.

Problem 10. Suppose that a battery is connected in series with a switch and large resistance to the primary terminals of a transformer. Sketch the waveform of the voltage induced in the secondary winding if the battery switch is closed, then opened again after a few seconds. NOTE. The purpose of the series resistance is to insure that the primary current and voltage have approximately the same waveform. Even with this resistance, the current will take a short time to rise to its final value.

Problem 11. A filament transformer has a 115-volt primary and a 5-volt secondary. The secondary has 10 turns. How many turns must there be on the primary?

Problem 12. A power transformer for a radio receiver has a 110-volt primary winding. It has three secondary windings: a high-voltage winding rated at 500 volts, 100 ma; a rectifier-tube winding rated at 5 volts, 3 amp; and a 6.3-volt, 2-amp winding.

(a) What is the turns ratio (n_p/n_s) of each secondary winding referred to the primary?

(b) What is the total primary current? HINT. Calculate the primary current required by each secondary winding, then add the three resultant currents.

Problem 13. A transformer is often designed so that it may be connected to each of several supply voltages and still furnish the proper secondary voltage to the load. This is accomplished by tapping the primary as shown in Fig. 10. Suppose that a filament transformer has a 10-turn,

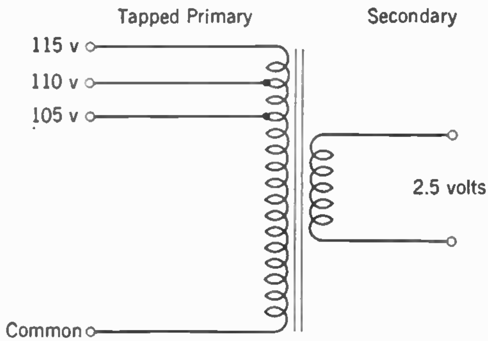


FIG. 10. Transformer with tapped primary.

2.5-volt secondary. What must be the number of primary turns for use on a 105-volt supply line? on a 110-volt line? on a 115-volt line?

5.13. Power in transformer circuits. Since a transformer does not add electric power to the circuit but merely changes or transforms one voltage to another, the power entering the primary side must be available at the secondary terminals. Of course, any loss in the transformer itself must be subtracted from the power which may be supplied to the load. In a well-designed power transformer there is little internal loss; practically all the primary power appears on the secondary side; and the product of the primary volts and amperes must equal (nearly) the product of the secondary volts and amperes. Thus, the primary power is

$$P_p = E_p I_p$$

and the secondary power is

$$P_s = E_s I_s$$

and, assuming no loss of power in the transformer,

$$E_p I_p = E_s I_s$$

whence

$$\frac{I_s}{I_p} = \frac{E_p}{E_s} = N \quad \text{(turns ratio)}$$

which shows that the secondary **voltage** is greater in a transformer with a low turns ratio and the secondary **current** is less.

If N is less than 1, that is, if there are more secondary turns than primary turns, the secondary voltage will be greater than the primary voltage but the secondary current will be less than that taken by the primary from the supply line. The transformer acts like an electrical lever and provides a means for obtaining a high voltage at a low current from a low-voltage, high-current source.

Since heat losses in a transmission line are proportional to the current squared (I^2R), it is common practice to transmit large blocks of electric power at high voltages and relatively low currents. There will be 100 times the power loss in a 110-volt line than in a 1100-volt line of the same resistance if both transmit the same power. Transformers make it possible to generate a-c power at comparatively low voltage, to transmit it long distances at high voltages, and then to utilize it at low voltages.

5.14. The autotransformer. It is not necessary for the transformation of voltage that the primary and secondary windings be distinct. Figure 11 represents what is known as an **autotransformer** in which the secondary is part of the primary. The voltage across the secondary turns, however, bears the same relation to that across the primary part as though there were two separate windings. The ratio of voltages is the ratio of the respective number of turns possessed by the secondary and primary. Since one side of the a-c supply line is usually grounded, one primary terminal of the autotransformer will be grounded and care must be used to prevent a direct short circuit such as would occur if the "hot" (ungrounded) secondary terminal should be connected to ground.

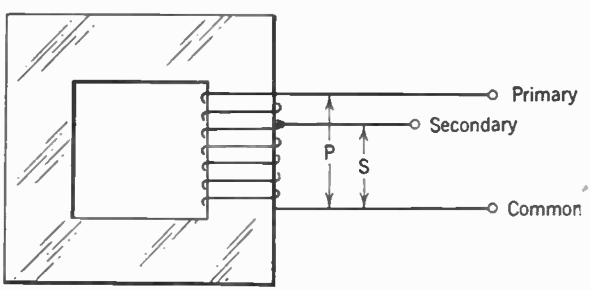


FIG. 11. Single tapped winding may be used as an autotransformer.

5.15. Transformer applications. A transformer is often used when both alternating and direct currents flow through a circuit and it is desired to keep the direct current out of the circuit to which the secondary is connected. No direct current can go across the transformer when the two windings are distinct, although the a-c voltage variations are transferred to the secondary by the effects already described. If no change in the a-c voltage is desired, the turns ratio is made unity; that is, the number of primary and secondary turns is made the same. An **output transformer** which couples a power tube in a power amplifier to the loud speaker serves this isolating purpose. The power tube has direct current as well as alternating current flowing in its plate circuit. Since the direct current should not flow in the loud-speaker voice coil winding, a transformer is used to isolate the speaker from the direct current of the tube. A good transformer for this service will transmit all frequencies in the audible range with an efficiency of about 80 per cent.

When no current is taken from the secondary of a transformer, the primary acts merely as a large inductance across the line and the current which it takes is rather small. The energy associated with this current is used in two ways, one of which is for supplying the energy which heats the transformer and its core. The other part maintains the magnetic field of the primary. This second part consumes no power from the line because at each reversal of the current the energy of this field is given back to the circuit.

When a secondary load is connected to the transformer, it begins to draw current from the secondary and more power is taken from the source supplying the primary. This additional power is that required by the load, together with that required to supply losses in the primary and secondary coil resistances.

Transformers of one kind or another are used in practically all radio, television, and radar equipment.

Problem 14. An a-c welding transformer has a 110-volt primary and a 5-volt secondary. The primary is connected to a standard 110-volt supply line, and the secondary is connected to the welding contacts. Suppose that a certain welding job requires 1000 amp.

(*a*) What is the primary current?

(*b*) How much power must the line furnish if the transformer is 90 per cent efficient?

Problem 15. An electric fan, designed to operate on 55 volts, is to be used on a 110-volt line with an autotransformer (Fig. 11) so that the proper voltage is supplied to the fan. The total number of turns on the autotransformer is 500. Show how the motor is connected, and mark the number of turns between the tapped point and each end of the autotransformer.

Problem 16. Work Prob. 15 if the motor requires 208 volts for proper operation. The line voltage is again 110 volts.

5.16. Transformer losses. Transformers are not perfect. Not all the power taken from the line by the primary results in power in the secondary load circuit. There are two sources of power loss: (1) losses in the windings (which have resistance and therefore heat up when current passes through them), and (2) core losses. The winding losses are called the **copper losses.** They are equal to the I^2R loss in the primary plus a similar loss in the secondary. These losses, in watts, may be found from the resistances of the windings, which are determined by measuring the current flowing through the individual windings when a known d-c voltage is across the winding. This will give a value for R from Ohm's law. Then the currents taken by the primary and secondary when the load is applied to the secondary may be measured with an a-c ammeter. From these values the I^2R losses may be found.

The second source of power loss is in the core. The magnetic flux expanding and contracting about the core induces a voltage in the secondary. But the same lines of force which cause voltages to be induced in the secondary also produce voltages in the iron core upon which the windings are placed. These voltages produce currents in the core, the currents produce heat (I^2R), and another heat loss must be added to the copper losses. In addition to these loss currents, known as **eddy currents,** there is another power loss in the core which is due to the basic principles of magnetism. It is supposed that, when iron is magnetized, the position of the molecules is changed so that most of the north poles point in one direction and the south poles in the opposite direction. When the direction of flux changes every half cycle, the orientation of the molecules must change so that the north poles now point in the opposite direction. The continuous shifting of the positions of the molecules in the core causes the core to heat up, and the heat represents still another loss of power. This wastage of power is known as **hysteresis loss.**

Since all the losses must be supplied by the source supplying the power, the losses represent power that cannot be utilized by the load. Special pains are taken to keep these unwanted losses as small as possible, for example, by using copper wire as large as practicable to keep the winding resistances low and by making the core of laminations (thin, insulated sheets of iron) rather than of solid iron.

Eddy-current and hysteresis losses may be determined by measuring the power input and power output of the transformer. The difference is the total loss. Then the copper losses may be calculated as indicated above. The copper losses subtracted from the total losses give the sum of the eddy-current and hysteresis losses, that is, the core losses.

Iron cores of one type or another are universally used in power transformers. In transformers operating at higher than commercial power frequencies (60 cycles) the use of iron becomes more difficult and the difficulty increases as the frequency increases. Hysteresis effects become very great because of the rapid reversal of the magnetic flux direction. Radio-frequency coils contain either non-magnetic core material or core material of very finely powdered iron which is mixed with a binder and then pressed into the desired form. The binder insulates one small particle of iron from another and thus reduces the eddy-current losses. The iron is specially selected and heat-treated to make the hysteresis losses low. These powdered iron cores have fairly low losses and may be used in r-f coils for broadcast and even higher frequencies. They increase the inductance of a given coil appreciably and enable engineers to get high inductances in a small space.

5.17. Transformers in vibrator power supplies. It is not necessary that the primary of a transformer be connected to a source of alternating current to produce an alternating voltage in the secondary. All that is necessary is for the primary current to be interrupted or reversed in direction. A battery, for example, may be placed in series with the primary of a transformer and a vibrator or other current-interrupting device as shown in Fig. 12. At each break of the primary battery current, a voltage will be induced in the secondary. The same step-up or step-down in voltage is secured as where alternating current is utilized. The secondary voltage may be rectified and filtered to supply high-voltage d-c power from low-voltage batteries.

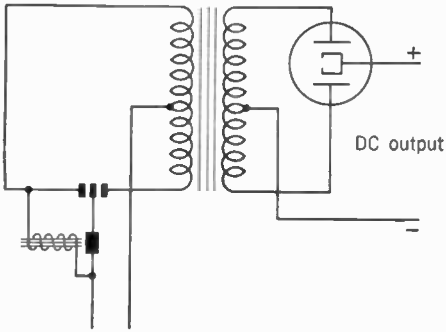


Fig. 12. The use of a vibrator, a transformer, and a rectifier to secure high-voltage direct current from a low-voltage source.

5.18. Variable inductors. Several means are available for varying the inductance of a coil. One is the simple expedient of having connections brought out from several points on the winding. Another method is to remove the insulation from the wire along a part of the winding and make contact with a slider. Also, merely inserting the iron core different distances into the coil will vary the inductance. Still another method consists of using two coils in series and changing the orientation of one coil with respect to the other; thus the coupling between the windings may be changed so that the total inductance, made up of the primary, secondary, and mutual inductance, may be varied from nearly zero to a value appreciably greater than the sum of the two individual inductances.

5.19. Air gaps in iron-core inductors. If the magnetizing force applied to an iron-core coil is increased indefinitely, the magnetic flux will not increase indefinitely but will reach a **saturation** value. Further increase in magnetizing force (as by increasing the current through the coil) will not produce any more flux and therefore will not produce any further counter-voltage. The large currents which flow cause the coil to overheat. The difficulty may be overcome by inserting a small air gap in the iron core. Even a gap of only a few thousandths of an inch has a very useful effect for the simple reason that air cannot be saturated by flux. Increasing the current in the coil increases the flux in the air gap, even though the flux in the iron does not increase. Changes in current, therefore, continue to produce changes in flux.

The air gap is useful where direct current as well as alternating current must flow through the coil. The direct current may produce sufficient flux to almost saturate the coil. Then, at the peaks of the alternating currents in the coil, the flux does not increase proportionately to the increase in current, and the core saturates. An air gap may be used to overcome this trouble. An important application of coils which carry both direct and alternating current is the "choke coils" used to filter out hum and noise in rectifier power supplies.

The manner in which the inductance of an iron-core coil varies with direct current is shown in Fig. 13, and the effect of various lengths of air gaps is shown in Fig. 14.

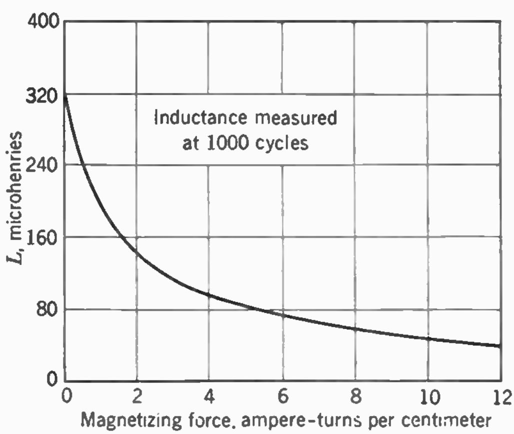


Fig. 13. Effect of saturating the iron core of an inductance as the direct current through it increases.

Problem 17. A large power transformer has an input power of 10,000 watts and delivers 9100 watts from the secondary to the load. The total copper loss in both primary and secondary windings is 650 watts. What is the core loss?

Problem 18. A 2200- to 110-volt transformer furnishes power to a 110-volt, 1.1-kw load. The resistance of the primary winding is 100 ohms, and that of the secondary winding is 0.5 ohm. What is the total copper loss in the transformer? Note. The actual voltage applied to the primary

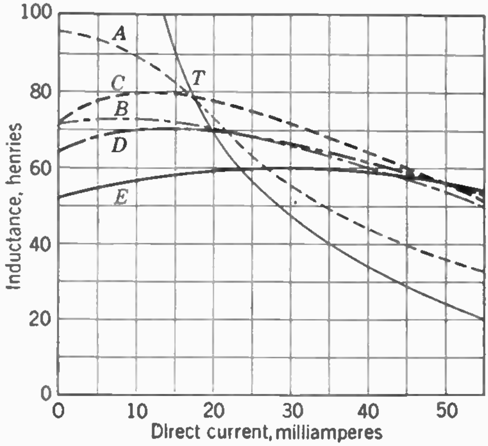


Fig. 14. Variation of inductance with air gap.

T = no air gap.
A = average air gap.
B = air gap at one end, 0.01 in.
C = air gap at both ends, 0.005 in. each.
D = air gap at both ends, 0.0075 in. each.
E = air gap at both ends, 0.01 in. each.

must be somewhat greater than 2200 volts to compensate for the voltage drop across the resistance of the windings. This effect may be neglected in the problem.

Problem 19. A 120-volt a-c generator is to be used to supply power to a town several miles away. The town requires 500 amp at 120 volts. The transmission line between the generator and the town has a resistance of 0.01 ohm.

(*a*) How much power is lost in the line? What is the actual voltage supplied to the town? Assume that the current is 500 amp even though the voltage is lower than 120 volts.

(*b*) Repeat part (*a*) if the voltage at the generator is stepped up to 24,000 volts with a transformer, and then applied to the same transmission line. It is stepped down to 120 volts at the other end of the line. Neglect any losses in the transformers.

6 · Capacitance

6.1. Capacitance. In the preceding chapter inductance was likened to inertia. In any circuit in which the current tends to change, an inductor tends to decrease or slow down the current variations. If current increases, owing to an increase in voltage, inductance tends to prevent or delay this increase; if the current decreases, owing to a decrease in voltage, inductance tends to prevent or delay this decrease. Inductance is one property of a circuit; capacitance * is another.

When current flows through a coil of wire, a magnetic field is produced. Energy resides in this field. Similarly, energy can be stored in a condenser. A condenser tends to prevent any change in the voltage across its terminals. If the voltage tends to decrease or increase for any reason, the capacitance of the circuit tends to prevent or retard this decrease or increase. The proper combination of inductance and capacitance will force the current and voltage of a circuit to remain almost perfectly constant in spite of varying conditions which, in a d-c circuit, would change the voltage and current.

6.2. Types of condensers. There are many types of condensers but all have the same essential elements, two conducting electrodes separated by an insulating medium. The electrodes serve as storage plates for the accumulation of electrical charges (electrons), and the insulating material prevents the plates from being short-circuited except through an external connection.

Figure 1 shows a simple condenser. The two plates may, for example, be made of brass, and the insulating material may be a sheet of glass. The insulating material is usually called the **dielectric**.

* **Capacitance** is the preferred word to describe this property of a circuit element; a device having capacitance is a *capacitor*. Engineers frequently use the terms "condenser" for a capacitor and "capacity" for capacitance.

Condensers for use in radio circuits may be divided into two groups: variable air condensers and fixed condensers. The general construction and use of each type will be described briefly. A more complete description of their uses will be found in later chapters.

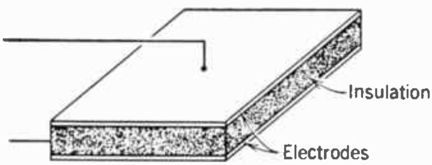


FIG. 1. Elements of a simple condenser.

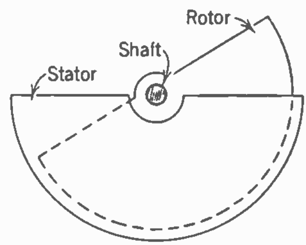
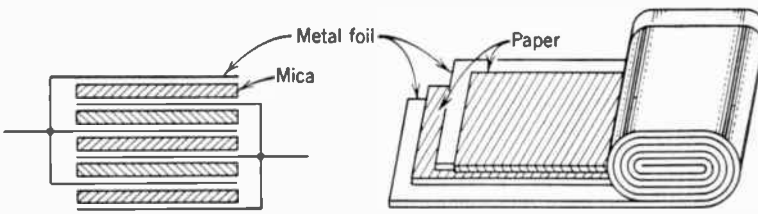


FIG. 2. Variable air condenser.

Variable air condensers are made up of a number of approximately semicircular metal plates. Half these plates are assembled in a fixed mounting, and the other half are attached to a shaft. The plates are spaced so that those of the fixed group (**stator**) and those of the rotating group (**rotor**) are interleaved. Air serves as the insulating dielectric. By rotating the movable plates the two groups may be adjusted so that they are completely or only partially meshed. The most common application of these condensers is to the tuning circuits of radio receivers and similar apparatus.

The term "fixed condenser" means that none of the plates are movable; all are fixed in position. A wide variety of fixed condensers is obtainable. The most common are mica, paper, oil-filled, and electrolytic. A mica condenser is normally made by stacking alternate rectangles of mica and metal foil. The pieces of metal foil are divided into two separate electrical groups by connecting alternate pieces together as shown in Fig. 3(a). These two groups are then connected to two terminals. Generally the assembly is cased in some insulating material and impregnated with wax to prevent moisture from entering the assembly. Mica condensers have very good electrical characteristics and are used in many applications. Their chief disadvantages are their relatively high cost and low capacitance.

Paper condensers are usually made by forming two long strips of metal foil and two strips of insulating paper into a spiral roll. Figure 3(*b*) illustrates the general construction. The two metal strips are attached to terminals, and the assembly, wax- or oil-impregnated, is mounted in a cardboard tube or encased in plastic insulation. Paper condensers are low in cost and may be made to have high capacitance. Paper condensers are the most widely used of all fixed types and are found in practically all kinds of electronic equipment. Oil-filled condensers are usually paper condensers cased in a hermetically sealed can containing a good quality of insulating oil.



(*a*) Mica Condenser (*b*) Paper Condenser

FIG. 3. Simple fixed condensers.

Electrolytic condensers consist of a spirally wound strip of metal foil, such as aluminum, immersed in a liquid or paste electrolyte. The metal foil is often embossed to provide more surface area. The electrolyte itself is an electrical conductor, but under the action of electrical current it forms a thin insulating film on the metal foil. This film is very thin, and the resulting capacitance of the combination may be quite high. The polarity of the voltage determines whether the thin insulating film is formed on the metal foil or on the encasing metal can. The condensers are designed for use with the film on the metal foil, and the condenser must therefore be used with the proper potential on its terminals. The metal foil is the positive terminal. These condensers cannot be used on a-c circuits * since the insulating film is then not formed on either electrode. They may, however, be used in circuits containing both d-c and a-c voltages (or pulsating d-c voltages), provided the d-c voltage is larger than the a-c voltage. The operating potential is limited to about

* A special type of electrolytic condenser is used in the starting circuit of some single-phase a-c motors. This type is not used in radio circuits.

600 volts. The big advantage of electrolytic condensers is that a high capacitance may be built into a small volume. This is particularly true of low-voltage units with about 25-volt ratings. Electrolytics are used chiefly in rectifier-filter networks and for cathode by-pass condensers.

6.3. The charge on a condenser. Several questions arise when one considers the electrical behavior of a condenser. Can it pass current? What is the charge on a condenser and how is the charge stored? How much capacitance * does a condenser possess?

When a condenser is connected to the terminals of a battery, there is a momentary rush of electrons to one of the electrodes and a corresponding rush of electrons away from the other electrode. There is no passage of electrons through the insulating medium which separates the electrodes. However, since the passage of electrons through the wires connecting the battery to the condenser constitutes a current, a voltage appears across the condenser terminals. If a d-c voltmeter (which should have very high resistance so that the voltmeter current can be neglected) is connected across the condenser, it will be found that the voltage starts from zero and rises to the battery voltage. If the connecting wires have very low resistance, the voltage comes to its final value very quickly. If, on the other hand, the connecting wires have high resistance, it will be found that the voltage rises to its final value very slowly. If the battery is removed after the voltage reaches its final value, that is, after the condenser is charged, it will be found that the voltage remains at this final value, or that it decreases very slowly as the charge is drained off through the voltmeter.

The term "charged," then, means that a condenser has too many electrons on one of its plates and too few on the other. The condenser is in a state of unequilibrium. If a wire is connected from one terminal to the other, a spark will pass, indicating that the maldistribution of electrons is being corrected. Now the condenser is discharged.

* **Capacitance** is a technical term describing the ability of a condenser to store energy in an electric field between its plates. The basic unit of capacitance is the **farad,** and this represents the capacitance of a device which will develop 1 volt across its terminals when 1 ampere flows into these plates for 1 second.

If a condenser is placed in an a-c circuit, there is a flow of electrons from the plate that is connected to the positive side of the line at the moment; then, when the polarity of the voltage changes, the electrons again move and charge the condenser in the opposite direction. This to-and-fro motion of the electrons may be just as effective in doing work as if the electrons moved through the insulator separating the plates. For example, these electrons moving to and fro through the wires connected to the terminals of the condenser cause the wires to become heated.

So long as the electrons are unevenly distributed on the two plates the condenser is charged. In this condition it possesses energy, which is like the energy possessed by a metal ball on top of a flagpole. The kind of energy possessed by the ball is potential energy; it is due to the position of the ball. The energy possessed by the condenser when charged is also potential energy. Nothing happens until the condenser discharges; then it may set fire to a piece of paper, may puncture a hole in a sheet of glass, or may give a person a severe shock. Thus the condenser, just like the ball on top of the pole, has the ability to do work—which is the definition of energy. The energy resides in the distorted placement of the electrons, a condition that will equalize itself at the first opportunity.

6.4. How much charge on a condenser? The charge (each electron carries a definite quantity of electricity or "charge") that flows into a condenser when it is connected to a battery is a perfectly definite quantity depending upon only two factors, the electrical size of the condenser (capacitance) and the voltage of the battery. Naturally, the larger the plates of the condenser the greater the number of electrons or charge that can be stored in it; and the greater the voltage tending to produce a state of unequilibrium, the greater the charge that can be forced into the condenser.

The quantity of charge in a condenser is rated in **coulombs.** It should be noted that it is proper to speak of both the charge *on* a condenser and the charge *in* a condenser. Heretofore we have concentrated upon the movement of charges to the plates of a condenser, and in that sense the charge is "on" the plates of the condenser. The voltage which appears across the condenser does, however, exert a distorting force on the dielectric material between the plates and so disturbs its equilibrium. In this sense, the charge is "in" the condenser.

The capacitance of a condenser depends upon three things: (1) the size of the plates, (2) the dielectric constant, k, of the insulating medium, and (3) the spacing between the plates. The larger the area of the plates, or the greater k, the greater the capacitance. The capacitance is greater when the plates are closer together.

The equation which expresses the relation among charge, voltage, and capacitance, is

$$Q \text{ (coulombs)} = C \text{ (capacitance)} \times E \text{ (voltage)} \qquad [1]$$

The unit of capacitance is the **farad** (f) named after Michael Faraday; 1 farad is the capacitance of a condenser whose voltage is raised 1 volt when 1 coulomb of charge is added to it, or it is the capacitance of a condenser to which 1 coulomb of charge can be added by the application of an external emf of 1 volt. This is a very large unit, and in radio circuits the millionth part of a farad (microfarad, μf) or even the micromicrofarad ($\mu\mu$f) is the customary unit.*

Equation 1 may be written

$$C \text{ (farads)} = \frac{Q \text{ (coulombs)}}{E \text{ (volts)}} \qquad [2]$$

Here the capacitance is considered as the ratio of the charge on a condenser to the voltage across it. A third way of writing the relation is

$$E \text{ (volts)} = \frac{Q \text{ (coulombs)}}{C \text{ (farads)}} \qquad [3]$$

A discharged condenser contains no stored electrical energy; there are just as many electrons on one set of plates as there are on the other set of plates; there is no strain across the insulating material between the plates; there is no voltage across it.

6.5. Time required to charge a condenser. How long does it take to charge a condenser? Suppose that an uncharged con-

* It is useful to note that 1 ampere (a unit for rate of flow of electricity) is equal to the movement of 1 coulomb of electricity per second. Conversely, a coulomb represents the amount of charge transported when 1 ampere of current flows for 1 second.

denser is connected across a battery. When the battery is first connected, there will be a rush of electrons through the circuit. Electrons will leave the plate connected to the positive terminal of the battery, since the negatively charged electrons are attracted toward a positive body; and, at the same time, electrons from the negative terminal of the battery will move to the condenser plate connected to this terminal. Each electron which arrives at one of the condenser plates represents a certain amount of electricity. As the electrons continue to pile up on one condenser plate, and to leave the other, a voltage is built up across the condenser. This voltage has a polarity such that it opposes the flow of current from the battery, and when enough electrons have moved through the circuit to cause the voltage built up on the condenser to equal the battery voltage, the current ceases to flow.

The first rush of current is high since there is then no counter-voltage across the condenser, and the *rate* at which the voltage across the condenser builds up is high. Near the end of the charging time the current is low and the rate of increase of voltage is low.

Time is required to change the charge on a condenser and to build up the voltage across it; the condenser tends to prevent any change of voltage across itself. This fact is very important. The phenomena connected with these momentary effects, called "transients," are employed in many radio, radar, and industrial electronic circuits.

For example, if a condenser and a resistor are connected in series across the battery, the condenser at the start acts as if it were shorted—there is no counter-voltage yet built up on the condenser to oppose the flow of current. The current at the moment is equal to E/I, according to Ohm's law, and all the voltage is impressed across the resistor. The current flowing into the condenser charges it; its voltage rises and "bucks" the battery voltage so that the current decreases. The voltage drop across the resistor, therefore, decreases. When the condenser is finally charged and the current ceases, there is no voltage across the resistor and all the battery voltage is impressed across the condenser.

The time required for the voltage of a condenser to reach 63 per cent of its final value, on charge or discharge, is equal to the product of the capacitance and the resistance, in farads and ohms, respectively. This RC product is called the **time constant** of the circuit.

A more extensive treatment of the electrical properties of condensers during the charging and discharging period will be found in Chapter 22.

Example 1. A condenser of 15 μf is attached to a 220-volt d-c circuit. What quantity of electricity flows into the condenser? If it is charged through a resistance of 10,000 ohms, how long will it take to put 63 per cent of its total charge into it?

$$\begin{aligned} Q &= C \times E \\ &= 15 \times 10^{-6} \times 220 \\ &= 3300 \times 10^{-6} \\ &= 0.0033 \text{ coulomb} \\ t &= RC \\ &= 10{,}000 \times 15 \times 10^{-6} \\ &= 15 \times 10^{-2} \\ &= 0.15 \text{ sec} \end{aligned}$$

Problem 1. A condenser holds 0.00044 coulomb of charge when attached to 110 volts. What is its capacitance?

Problem 2. What voltage is necessary to put 0.09 coulomb into a 2-μf condenser?

Problem 3. A 350-$\mu\mu$f condenser is attached to a 300-volt source. How much charge flows into it?

Problem 4. A condenser and a 1,000,000-ohm resistor are connected across a 110-volt d-c source. A voltmeter is placed across the condenser. The time required for the voltage to reach 63 per cent of its final value after voltage is applied to the circuit is measured with a stopwatch and found to be 1⅗ sec. What is the size of the capacitance?

6.6. Energy stored in a condenser. The amount of energy that can be stored in a condenser is given by

$$\text{Energy} = \tfrac{1}{2}CE^2$$

where energy is in joules, C in farads, and E in volts. This energy represents the amount of work done in charging the condenser. It does not include any energy loss in the resistance of the wires connecting the condenser to the source of voltage. The above amount of energy is also the amount released when the condenser is discharged.

The unit of energy or work is the **joule.** It is the amount of work required to force 1 coulomb of electricity through a 1-ohm resistor. If a 1-μf condenser is charged to a potential of 500 volts, the energy is

$$\tfrac{1}{2} \times 1 \times 10^{-6} \times (500)^2 = 0.125 \text{ joule}$$

A joule is also known as a **watt-second.** It is a measure of work done, or energy, not power. Power is work per unit of time.

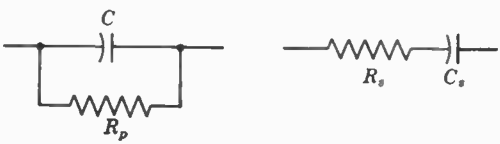
6.7. Electrostatic and electromagnetic fields. The energy existing in an inductive circuit is said to exist in the electromagnetic field surrounding the inductor. The field is made up of magnetic lines of force, and it can be explored with a compass or by sprinkling iron filings on a piece of paper. The energy in an inductor is equal to $\frac{1}{2}LI^2$ joules, where L is in henries, and I is in amperes.

The energy in a condenser is said to exist in the electrostatic field, where the electrical strain exists, that is, in the insulating materials in the vicinity of the charged plates. The electrostatic field cannot be explored by a magnetic substance, but it can be explored with a tiny charged body which will be attracted to one set of condenser plates and repelled from the other. The energy in a condenser is $\frac{1}{2}CE^2$.

Some circuits, an antenna system, for example, include both capacitance and inductance. In such circuits energy can be stored in both the electrostatic and the electromagnetic fields.

6.8. Imperfect condensers. The discussion of condensers up to this point has tacitly assumed that the condenser was perfect, that there was no resistance in the condenser itself. This is not, of course, true for actual condensers. The metal plates themselves have some resistance. However, this is usually so low that it may be neglected. Of more importance is the finite amount of resistance of the insulation separating the electrodes. This resistance, even though very large, will pass some current and therefore introduces a heat loss in the condenser. An imperfect condenser can be represented as shown in Fig. 4(a), in which C is the capacitance of the condenser and R_p represents the resistance of the insulating dielectric. For high-quality condensers the value of R_p is extremely high, many millions of ohms. For electrolytic condensers, on the other hand, the value of R_p is one-tenth to one-hundredth of a megohm.

An alternative representation of an imperfect condenser is shown in Fig. 4(*b*). In this representation C_s is approximately the same as the true capacitance C in Fig. 4(*a*), except that it may be considerably different for very poor condensers and for electrolytic condensers. R_s is the effective series value of R_p and may be calculated by methods developed in Chapter 7. Unlike R_p, R_s is very small for good condensers, and is larger for poorer condensers. This second representation is very convenient for



(*a*) Parallel (*b*) Series

FIG. 4. Ways of representing an imperfect condenser.

many circuit calculations and is equivalent electrically to the first representation.

6.9. Condensers in a-c circuits. A perfect condenser, as suggested in the preceding section, is one which is an absolute nonconductor to direct currents—that is, it has an infinite d-c resistance—and which has no a-c resistance. All the energy that is put into a perfect condenser is used in setting up an electrostatic field. A perfect condenser, once charged, would keep its charge indefinitely, or until its terminals were connected together.

A perfect condenser placed across an a-c line would take no average power from the line, since for one half cycle it would take energy from the line but during the next half cycle it would return this energy to the line. If the condenser has resistance (and all do), the current taken from the line to charge the condenser also flows through the resistance and in so doing produces a power loss there equal to I^2R_s. For this reason, a condenser having resistance or a circuit made up of a resistor and a condenser will consume power from an a-c source.

Experiment 1. Charge an electrolytic filter condenser of about 2- to 10-μf capacitance by connecting it to a 110-volt d-c circuit, using care to keep the polarities correct. Then discharge by connecting a heavy wire across its terminals; then charge again and allow to stand for a half hour and then discharge. Charge again, and permit to stand for an hour and then discharge it. The relative size of the spark gives an idea of how poor a condenser it may be from the standpoint of leakage. As another experiment, charge the condenser as before; then place a high-resistance voltmeter across its terminals and note the rate at which the voltage drops.

A good condenser may retain its charge for many hours after being charged. The leakage of the charge takes place not only across the insulating material through which its terminals are brought out but also through the wax or other material filling the condenser, and through the dielectric itself. Leakage is caused by an actual movement of electrons between the terminals and through the insulation.

A considerable amount of energy can be stored in a condenser if it is charged to a high voltage. A person touching the terminals of such a charged condenser may receive a severe shock. *Caution should always be exercised in handling condensers which may be charged.*

Example 2. Since a condenser blocks the flow of d-c current but permits a-c current to flow through it, a condenser may be employed to separate direct and alternating currents existing in the same circuit. For example, Fig. 5 shows a resistance-capacitance network for coupling the output of one tube to the input of another tube in an amplifier. The purpose of C is to prevent E_{bb}, the plate supply voltage of the first tube, from being impressed upon the grid of the second tube. Any leakage of direct current through C will

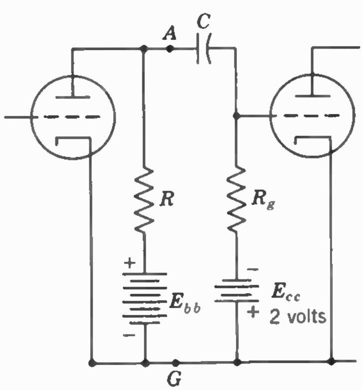


Fig. 5. Leakage of current through condenser C causes a voltage drop across the input of the following tube.

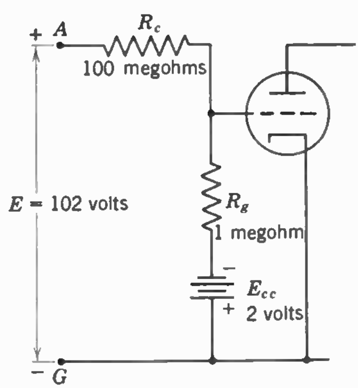


Fig. 6. The circuit equivalent of Fig. 5 for elements to right of AG.

impress upon the grid of the second tube a positive voltage which is detrimental to the operation of the circuit.

Figure 6 is a d-c equivalent circuit for the circuit elements to the right of AG. The condenser has been replaced by its d-c resistance R_c, assumed to be 100 megohms. The voltage from G to A is assumed to be 102 volts as shown. The problem is to find what current flows through the circuit and what voltage drop this causes across the grid leak R_g. The net voltage across the series circuit composed of R_c and R_g is 100 volts as shown in Fig. 7. One hundred volts across 100 megohms (the size of R_g will be negligible in comparison to the size of the leakage resistance) produces 1 μa of current. This 1 μa flowing through 1 megohm produces a voltage of 1 volt. The polarity of this voltage across R_g is opposite to that of E_{cc}, and the net voltage applied between the cathode and grid of the second tube is reduced from 2 to 1 volts. In addition a certain amount of noise is developed by the direct current flowing through the condenser and grid leak. This noise is impressed on the grid of the second amplifier tube and may attain large value when amplified by succeeding stages. Condensers used in resistance-coupled amplifiers must have a high d-c resistance.

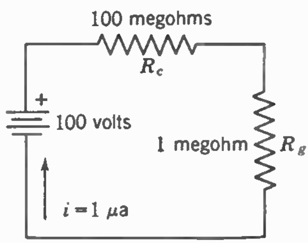


Fig. 7. The battery represents the voltage across R_c and R_g in Fig. 6.

6.10. The nature of the dielectric. If two square metal plates 10 cm on a side are suspended in air about 1 mm from each other, the capacitance will be about 88.5 $\mu\mu$f. If a 1-mm sheet of mica is placed between the plates, the capacitance will be increased about six times. If other insulating substances are used, the capacitance will have different values. Each substance, in fact, will result in a certain value of capacitance, depending upon what is called the **dielectric constant** of the substance. The table gives the value of the dielectric constant, k, of several materials.

Material	Dielectric Constant, k	Material	Dielectric Constant, k
Air	1.0	Quartz	5
Bakelite	4–8	Rubber, hard	2.3–2.5
Glass	6–10	Shellac	3.1
Mica, clear	5.5–7	Steatite	6.1
Paper	2–4	Transformer oil	2.3–2.5
Paraffin	2.1	Varnished cambric	4
Polystyrene	2.6	Water (pure)	81
Porcelain	5.7–7.5	Wood (dry)	2–8

6.11. Condenser formulas. Formulas have been worked out by which it is possible to compute the capacitance of condensers of various forms. For example, the capacitance of a flat plate condenser may be computed from the formula

$$C = \frac{8.85 \times A \times k \times (N - 1)}{10^8 \times d} \text{ microfarads}$$

or

$$C = 0.0885 \frac{kA(N - 1)}{d} \text{ micromicrofarads}$$

where A = area of one side of one metallic plate in square centimeters.
k = dielectric constant of insulator separating the plates.
N = number of plates.
d = thickness of the dielectric in centimeters.

Formulas for other types of condensers may be found in the Circular of the National Bureau of Standards C74, "Radio Instruments and Measurements," p. 235.

The value of the dielectric constant k has nothing to do with the ability of an insulating material to withstand high voltages without rupture. Some materials with a high dielectric constant may have low tolerance to high voltages; other materials which can be used in high-voltage condensers may have a low dielectric constant. Mica, with a medium value of k, has a highly prized ability to withstand high voltages. Furthermore, the condition of the dielectric material affects its breakdown voltage. Temperature, moisture, and mechanical pressure to which the insulating material is subjected are among the factors that affect voltage breakdown.

Example 3. How many plates 16 by 20 cm in area and separated by paraffined paper ($k = 2.1$) 0.005 cm thick are required for a condenser of 24-μf capacitance?

$$C = \frac{8.85Ak}{10^8 d}$$

$$= \frac{885 \times 16 \times 20 \times 2.1}{10^{10} \times 0.005}$$

$$= 0.0119 \ \mu\text{f per plate}$$

$$\text{Number of plates} = \frac{24}{0.0119} = 2200$$

Problem 5. Express in micromicrofarads: $\frac{1}{5000}$ farad, 2 μf, 0.00075 μf. Express in microfarads: 2 farads, 300 $\mu\mu$f, 0.01 $\mu\mu$f. Express in farads: 250 μf, 0.05 μf, $\frac{1}{10}$ μf.

Problem 6. A certain condenser has a capacitance of 0.01 μf and a leakage of 1000 megohms. It is charged to a potential of 500 volts; then the voltage source is removed. How long will it take for 63 per cent of the charge to leak off?

Problem 7. A condenser in a transmitter station has a capacitance of 0.002 μf and is charged from a 20,000-volt source. What energy goes into the condenser?

Problem 8. What will be the capacitance of a condenser made of two metal plates 1 cm by 2 cm separated by a 0.1-mm sheet of mica with a dielectric constant of 6?

Problem 9. A variable air capacitor has a capacitance of 350 $\mu\mu$f. What will be the capacitance if the capacitor is immersed in transformer oil with a dielectric constant of 2.4?

Problem 10. A variable air condenser is made up of 10 plates, each having an area of 20 sq cm. The plates are separated 3 mm. What is the capacitance when the plates are completely meshed?

Problem 11. Two large metal plates separated by air have a capacitance of 40 $\mu\mu$f. When the space between the plates is filled with a block of dry wood, the capacitance is measured and found to be 140 $\mu\mu$f. What is the dielectric constant of the wood?

Problem 12. Metal-foil plates 2 by 3 in. are separated by mica 0.1 mm thick having a dielectric constant of 6. What is the capacitance of a condenser made up of a stack of 101 metal-foil plates and 100 mica spacers?

6.12. Condensers in parallel and series. When condensers are connected in parallel, as shown in Fig. 8, the resultant capaci-

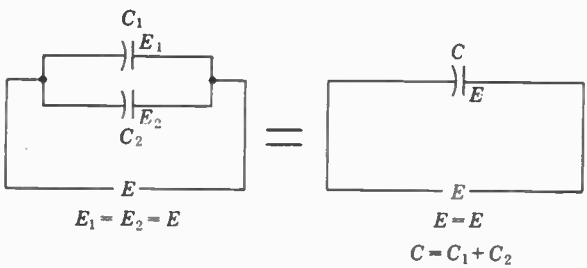


Fig. 8. Condensers in parallel.

tance is the sum of the individual capacitances for the simple reason that the voltage which is applied stores as much charge in each condenser as it would store if it were applied to each condenser separately.

When condensers are connected in series, however, a different condition exists. Figure 9 shows two condensers connected in series across a source of voltage, E. Since the current which

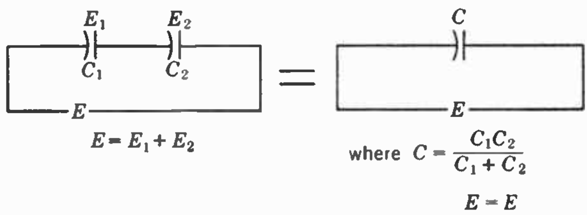


FIG. 9. Voltages across condensers in series.

charges the condensers must be the same in the two units, the same amount of charge (current is the rate at which charges flow) exists in each condenser. The voltage across the combination, E, is equal to the charge divided by the resultant capacitance of the series circuit. The voltage across each condenser is equal to the charge divided by the capacitance. Thus

$$E = Q \div C$$

$$E_1 = Q \div C_1$$

$$E_2 = Q \div C_2$$

$$E = E_1 + E_2$$

or

$$E = Q \div C = \frac{Q}{C_1} + \frac{Q}{C_2}$$

or

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

or

$$C = \frac{C_1C_2}{C_1 + C_2} \quad \text{(two condensers in series)}$$

If several condensers are connected in series the resultant capacitance is given by

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \quad \text{(several condensers in series)}$$

Thus it can be seen that condensers in series act (so far as their resultant capacitance is concerned) like resistances in parallel; condensers in parallel act like resistances in series. When condensers are connected in parallel the voltage is the same across each; when they are connected in series the voltage across each varies inversely as the capacitance. If two equal condensers are in series, half the total voltage is impressed across each condenser. If one condenser has half the capacitance of the other, that one will have two-thirds of the total voltage and the other will have one-third of the total voltage across it.

The resultant capacitance of several condensers in parallel is always greater than any single capacitance; in series the resultant is less than the capacitance of the smallest of the group. In series the voltage across each condenser varies inversely as its capacitance and is always less than the total voltage across the combination. When a condenser is to be placed in an a-c circuit whose voltage is more than the condenser can tolerate, it is necessary only to use two or more condensers in series so that the voltage across individual members of the group is below the danger limit, and with individual capacitances such that the total capacitance in the circuit is the value desired.

When condensers are used across d-c circuits, as in filters, the d-c resistance of the condenser becomes of importance. If two condensers in series across a certain d-c voltage have equal capacitances but different d-c resistances, the voltage drop across the two condensers will be different. The voltage across one of them may be sufficient to cause its insulation to be punctured. This difficulty may be avoided by placing a separate resistor in parallel with each condenser as shown in Fig. 10. If the voltage is to be divided equally between the condensers, R_1 and R_2 are made equal. If the condensers are paper or mica, each resistance may be several megohms. If the condensers are of the electrolytic type, the resistances should be smaller since the d-c resistance of electrolytic condensers is rather low.

Example 4. In a radio circuit a 0.0005-μf variable condenser is available but the circuit calls for a 0.00035-μf condenser. What fixed condenser may be used to reduce the maximum capacitance of the circuit to the proper value? How should it be connected?

Solution: Since the total capacitance is to be reduced, the fixed condenser must be connected in series with the variable condenser. The total capacitance is given as 0.00035 μf. This is equal to

$$0.00035 = \frac{C_1 \times C_2}{C_1 + C_2}$$

$$= \frac{0.0005 \times C_2}{0.0005 + C_2}$$

whence

$$C_2 = 0.001166 \ \mu\text{f or } 1166 \ \mu\mu\text{f}$$

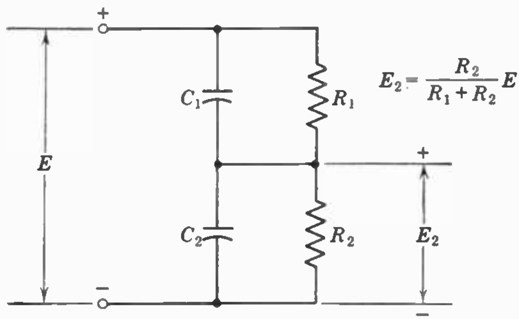


FIG. 10. The resistors cause the d-c voltage, E, to divide across the condensers in any desired manner.

6.13. Negative-temperature-coefficient condensers. The dielectric constant of materials used in most condensers is not seriously affected by temperature, and the capacitance may be assumed to be constant even though the temperature varies, except when considerable precision is required. There is, however, a special group of condensers in which are used dielectric materials, such as titanium oxides, which are very sensitive to temperature. These condensers usually have a negative temperature coefficient; that is, their capacitance decreases as the temperature increases. By proper design these condensers can be made to have rapid or slow variations in capacitance with a change in temperature.

Negative-temperature-coefficient condensers are often used as frequency-compensating elements. As various elements of the circuit become heated during operation the values of capacitance, inductance, and resistance change slightly. This may cause the frequency to drift from the correct value. The proper application of negative-temperature-coefficient capacitors can hold this frequency drift to small values.

6.14. Condenser tests. If a voltmeter, a battery, and a condenser are connected in series, a momentary deflection of the voltmeter needle will be noted as the connection is made if the condenser is good. If the condenser is leaky, a constant deflection will be noted. (This test does not work for an electrolytic condenser since a d-c current will flow through an electrolytic condenser even if it is good.) If the condenser is fairly large, 1 or 2 μf, and no deflection is noted, the condenser is probably open. If the voltmeter continues to register the full battery voltage, the condenser is short-circuited.

If a condenser that is not short-circuited is placed momentarily across a battery, and then a pair of phones is placed across the condenser, a click will indicate that the condenser is good. The click results from the discharge of the condenser through the phones.

6.15. The earth as a condenser. The earth is considered as being at zero potential. All objects not connected to earth have a higher voltage than the earth; they have a capacitance with respect to the earth. The purpose of "grounding" metallic parts of radio receivers or transmitters is to make sure that they are at the same potential so that there will be no currents flowing through the capacitance between these objects and the metal chassis of the receiver or transmitter. The chassis, in turn, is usually grounded either by actually connecting it to earth or by connecting it electrically to one of the power lines that furnish power to operate the equipment.

Problem 13. Some paper condensers are made up of a thin metallic foil pressed to a dielectric of paper which has the form of a long sheet about 2 in. wide. Suppose that the dielectric constant of such paper is 3 and that it is 0.0025 in. thick. A condenser is to have a capacitance of 0.1 μf; the metal foil is 1¾ in. wide. Two sheets of foil and two sheets of paper are used. How long will each sheet of foil be?

Problem 14. What is the resultant capacitance of a 0.1- and a 0.05-μf condenser connected in series?

Problem 15. What is the maximum voltage which can be placed across the series combination of Prob. 14 without causing the voltage across the 0.1-μf condenser to exceed 200 volts? Neglect leakage resistance.

Problem 16. Three condensers have capacitances of 0.1, 0.2, and 0.5 μf, respectively. What will be the resultant capacitance if they are connected in parallel? if they are connected in series?

Problem 17. A "band spread" condenser is one in which a large rotation of the rotor corresponds to a relatively small change in capacitance. One way to make such a band spread condenser is to connect a fixed condenser in series with a variable condenser. Suppose that a variable condenser with a maximum capacitance of 350 $\mu\mu$f and a minimum of 35 $\mu\mu$f (a change of 10 to 1 in capacitance) is connected in series with a 500-$\mu\mu$f fixed condenser. What is the maximum and the minimum capacitance of the combination? What is the ratio of maximum to minimum capacitance?

Problem 18. Two paper condensers, one with a capacitance of 0.5 μf and the other with a capacitance of 0.3 μf, are to be placed in series across 1000 volts. The voltage rating of the 0.5-μf condenser is 300 volts and of the 0.3-μf condenser is 700 volts. If a 500,000-ohm resistor is placed in parallel with the 0.5-μf condenser, what size resistor must be placed in parallel with the 0.3-μf condenser to insure that its voltage rating will not be exceeded?

Problem 19. What would be the voltage across each condenser in Prob. 18 if the parallel resistors were omitted? Neglect leakage resistance.

7 · Properties of Alternating-Current Circuits

The two kinds of electric current in common use are **direct currents,** which have a more or less constant value and which flow in the same direction all the time; and **alternating currents,** which are constantly varying in both magnitude and direction.

7.1. Definitions used in a-c circuits. When the voltage (or current) has started from zero and risen to its maximum value in one direction, decreased to zero and risen to its maximum value in the opposite direction, and finally come back to its starting value, zero, it is said to have completed a **cycle.** Ordinary house-lighting current which has a frequency of 60 cycles per second goes through this cyclic change in magnitude and direction 60 times a second. The **frequency** is the number of times a second a cycle is completed. In a-c circuits the element of time must be considered. In d-c circuits time does not enter; the direction of the current is always the same.

Alternating currents used in present-day electrical equipment cover an extremely wide frequency range. Sixty cycles is the common power frequency; tones generated by audio-frequency oscillators for testing purposes may go from almost zero frequency up to around 20,000 cycles per second. (The human ear can hear tones up to about 15,000 cycles per second. The upper frequency depends upon the person and decreases with his age.) Radio-frequency waves as low as 15,000 cycles are used. They are generated in the long-wave high-power radio stations carrying on transoceanic communication. Radio frequencies up to 10 billion cycles (10,000 megacycles) and above are now used.

Table 1 gives the conventional terminology which has been applied to certain ranges of frequencies.

In a-c circuits the voltages and currents are continuously varying. It will be helpful in the discussion which follows to have a

TABLE 1

Frequency Range: Kilocycles	Frequency Range: Megacycles	Name	Abbreviation
Below 30		Very low frequency	VLF
30–300		Low frequency	LF
300–3,000	0.3–3	Medium frequency	MF
3,000–30,000	3–30	High frequency	HF
30,000–300,000	30–300	Very high frequency	VHF
300,000–3,000,000	300–3,000	Ultra high frequency	UHF
3,000,000–30,000,000	3,000–30,000	Super high frequency	SHF
	30,000–300,000	Extremely high frequency	EHF

means of determining the value of these voltages and currents at any instant.

Consider the circle in Fig. 1, in which the radius or arm rotates counterclockwise at a constant speed. Let the position of this radius at any instant represent the position of the rotating part of an a-c generator, and the length of the radius represent the maximum value of the voltage produced by the generator. The height of the end point of the arm above the horizontal (or zero) line represents the instantaneous value of the voltage at any instant.

Suppose that the generator starts from rest at zero time. The voltage is zero; now the arm begins to rise as the rotor of the generator moves, and when the rotor has gone through one-quarter of a complete rotation or cycle (90°), the voltage has risen to its maximum value and the end point of the arm has reached its maximum height above the zero axis. At the end of another quarter cycle the arm is again on the zero axis but pointing in the opposite direction and moving down instead of up; in another quarter cycle the arm points down perpendicularly, and its distance from the axis is again a maximum.

If the heights of the end point of the arm are plotted against time expressed in seconds, or in degrees, a curve like that shown at the right in Fig. 1 will result. This curve represents the variation in voltage produced by an a-c generator throughout the cycle of 360°. The angle through which the arm has moved at a given time, reckoned from its zero position, is called the **phase angle** (or **phase**). Thus, when the arm is vertical, it has gone through 90° of its rotation and the phase is said to be 90°. This is merely a method of expressing time in degrees of rotation instead of seconds, and degrees can be converted into seconds if the speed of revolution is known. The value of the voltage (or current) at any instant is known as the **instantaneous** value and of course varies from zero to a maximum value in one direction, through zero again, and then to a maximum in the opposite direction.

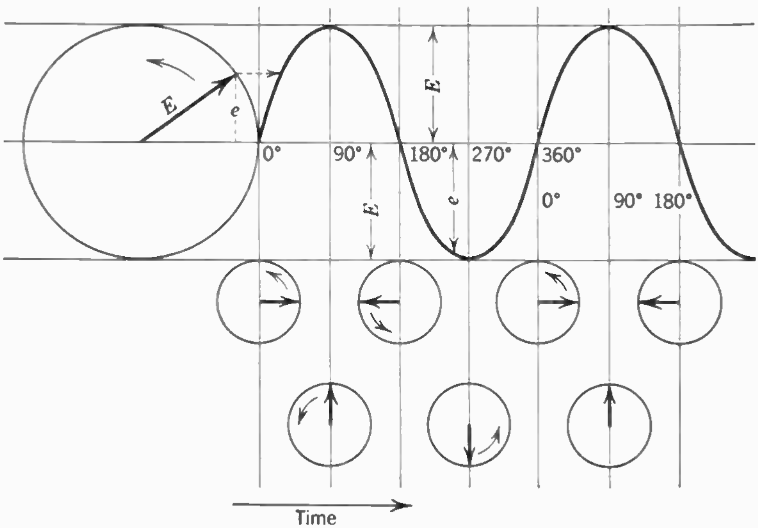


Fig. 1. As the vector E rotates, the vertical height, e, of the end of the vector varies according to the curve known as a sine wave. When the vector has the position as shown in the first small circle, the instantaneous value e is zero. At other times the instantaneous voltage has the other values as shown.

7.2. Sine waves. The undulating curve plotted in Fig. 1 is known as a **sine wave.** The ratio of the vertical height of the end of the radius to the length of the radius is known as the **sine** (abbreviated "sin") of the angle between them. Thus, $\sin \theta = e/E$, where θ is the Greek letter theta. This equation is read "sine theta equals e/E" and means simply that the sine of the angle θ is equal to e divided by E. Sine waves have very important properties, as will be seen later; they are representative of many natural phenomena such as the rise and fall of the tides and the motion of a pendulum. From the expression for the sine of an angle may be obtained the value for e, the instantaneous value of the voltage, if the maximum voltage E is known.

If one chooses to start counting time, or measuring angles, at the point marked 90° in Fig. 1, the wave is then known as a **cosine** wave. Thus, a cosine and a sine wave are the same except for the starting point, and a cosine wave has a 90° phase in relation to a sine wave.

Another expression for angles in circular measure is the **radian.** The radian is defined as follows: If the length of the circumference of a circle is measured, it will be found to be 6.28 times as long as the radius of the circle, regardless of the size of the circle. That is, the circumference of any circle divided by its radius is 6.28, and the circumference divided by the diameter is 3.14. Now the ratio between the circumference and the diameter of any circle is known by the Greek letter π (pi), and the numerical value is 3.1416 (approximately $3\frac{1}{7}$) for all circles. If an angle is marked out in a circle such that the length of its arc is equal to the radius (see arc PQ, Fig. 2), this angle is called 1 radian. There are, therefore, 2π radians in a circle, and any portion of the circle may be measured in radians as well as degrees. *A radian is equal to 57.3°.*

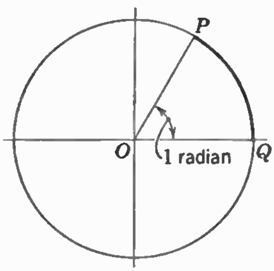


FIG. 2. The arc PQ, if stretched out straight, is as long as the radius OP. The angle at O is said to be 1 radian.

Table 2 shows useful relations among radians and degrees.

TABLE 2

RADIANS	DEGREES
$\pi/12$	15
$\pi/6$	30
$\pi/4$	45
$\pi/2$	90
π	180
2π	360

7.3. Triangle functions. Consider the right triangle in Fig. 3. Let the angle between CB and AB be called theta (θ); then the side AC is called the **opposite side,** CB the **adjacent side,** and AB the **hypotenuse.** The relations among the lengths of these sides and the angle θ are as follows:

$$\frac{AC}{CB} = \tan\theta \quad \text{or} \quad AC = CB\tan\theta \tag{1}$$

$$\frac{AC}{AB} = \sin\theta \quad \text{or} \quad AC = AB\sin\theta \tag{2}$$

$$\frac{AB}{CB} = \sec\theta \quad \text{or} \quad AB = CB\sec\theta \tag{3}$$

The tangent (tan), sine (sin), and secant (sec) are called functions of the angle θ; if any two of the three functions of

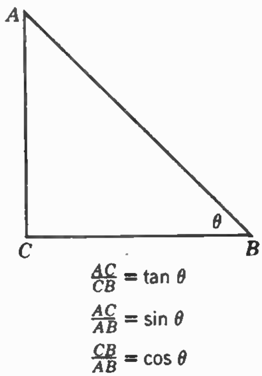


FIG. 3. Trigonometric functions.

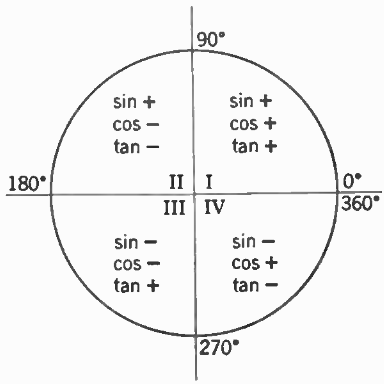


FIG. 4. Signs of the functions of angles in each quadrant of a circle.

the right triangle and the angle involved are known, the other function may be found by means of the table at the end of this book. The cofunctions are, respectively, the cosine (cos), cotangent (cot), and cosecant (csc). They are defined as follows:

$$\cos\theta = \frac{1}{\sec\theta} \qquad \cot\theta = \frac{1}{\tan\theta} \qquad \csc\theta = \frac{1}{\sin\theta}$$

The triangle (trigonometric) functions most often used in this book are the sine, cosine, and tangent.

The function of an angle greater than 90° can be found from the equation:

Function $(N \times 90° + A)$ = Function of A if N is even; or cofunction of A if N is odd [4]

The numerical sign (+ or −) of the functions varies from quadrant to quadrant as shown in Fig. 4.

Example 1. What is the sine of an angle of 195°? 120°?

Solution: This angle of 195° is equal to twice 90° plus 15°. Thus $195° = 2 \times 90° + 15°$, and $N = 2$. Since N is even, the sine of 195° is equal to the sine of 15°, except that a minus sign is required since the angle is in quadrant III. From the table at the end of the book $\sin 15° = 0.259$; therefore $\sin 195° = -0.259$.

The sine of 120° is equal to $\sin (1 \times 90° + 30°) = \cos 30° = 0.866$, since here N is an odd number.

Problem 1. (*a*) Express the following numbers of radians in degrees: $\pi/3$, $\pi/9$, $2\pi/3$, $\pi/5$, $\pi/7$. (*b*) Express the following numbers of degrees in radians: 40°, 135°, 75°, 29°.

Problem 2. (*a*) Using the table of trigonometric functions at the end of the book, evaluate the following: sin 40°, cos 10°, tan 75°. (*b*) Evaluate, using equation 4: cos 135°, sin 300°, tan 210°. Be sure to affix the proper sign to your answers.

7.4. Means of expressing instantaneous values. The instantaneous value of an a-c voltage or current may be expressed as follows:

$$e = E \sin \theta \quad \text{or} \quad i = I \sin \theta$$

where e or i = the instantaneous value.
E or I = the maximum value.
θ = the phase angle in degrees or radians.

Small letters denote instantaneous values; capital letters denote maximum values.

Figure 5 shows the instantaneous voltage, e, for four typical cases, 45, 90, 135, and 333°. The actual value of e may be computed from the above formulas if the value of E is given.

Since the sine of an angle of 0° is zero, the instantaneous value of the voltage at 0 phase is zero; since the sine of an angle of 90° is 1, the instantaneous value of the voltage at this point in the cycle is equal to the maximum value; and so on.

The three methods of representing an a-c voltage or current are:

1. By a graphical illustration such as Fig. 1, called a sine wave.
2. By an equation, such as

$$e = E \sin \theta \quad \text{or} \quad i = I \sin \theta$$

3. By pictures shown in Fig. 5, known as **vector diagrams.**

Such a line as E in Fig. 6 which moves about a circle is called a **vector;** the vertical distance of its end point from the hori-

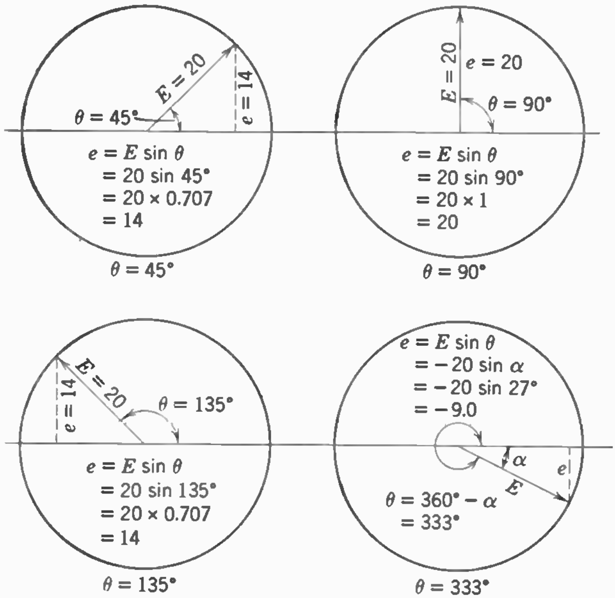


FIG. 5. At various times in the cycle the instantaneous value e of the vector E is as shown in these vector diagrams.

zontal axis is called its **vertical component.** The angle θ between the horizontal and the vector is called the **phase angle;** the value of the vertical component may be found by multiplying the maximum value of E by the sine of the phase angle.

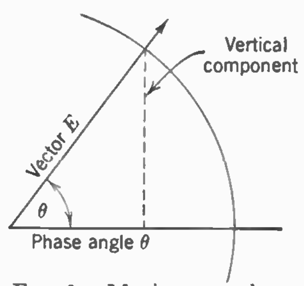


FIG. 6. Maximum value of the vector is E; the value at any instant is $e = E \sin \theta$.

Example 2. Illustrate the three methods of representing an a-c voltage whose maximum value is 20 volts.

Method 1. On graph or cross-section paper, let 12 divisions to the right of a vertical line represent 1 alternation (1 positive loop) and 20 divisions vertically from a horizontal line represent the maximum value of the voltage. The voltage starts at zero, increases to a maximum of 90° or 6 divisions, then decreases to zero at 12 divisions, etc. What is its value at other times? These values can be determined by using a table of sines. At the 30° phase angle the instantaneous value, or the height of the tip of the rotating arm above the time axis, is $e = E \sin 30° = 20 \times 0.5 = 10$ volts. Other values can be found in a similar manner and the entire sine wave plotted. The result will look much like Fig. 1.

Method 2. Lay off on cross-section paper a length, such as 20 divisions, equal to the maximum value of the voltage. Using this length as a radius, draw an arc of a circle as in Fig. 6. The instantaneous value of the voltage at any time in the cycle is found by measuring the vertical distance of this point on the circle to the horizontal axis. Thus at the 30° phase angle the vertical distance is 10 because $\sin 30° = 0.5$, and the instantaneous value is equal to the maximum value multiplied by the sine of 30° or $20 \times 0.5 =$ 10 volts.

Method 3. A voltage of maximum value of 20 volts may be represented mathematically by

$$\begin{aligned} e &= 20 \sin \theta \\ &= 20 \sin 30° \\ &= 20 \times 0.5 = 10 \text{ volts} \end{aligned}$$

Problem 3. The maximum value of an alternating voltage is 120 volts. What is its value at the following phase angles: 30°? 45°? 60°? 90°? 150°? 180°? 270°? 360°?

Problem 4. The instantaneous value of an alternating voltage is 200 volts at 30°. What is its maximum value? What is its value at 240°?

Problem 5. The instantaneous value of an alternating voltage is 150 volts at 50°. Plot its sine wave to some convenient scale.

Note. In all preceding discussion in this chapter voltages and currents can be spoken of with the same laws in mind. The form of a sine wave of current looks exactly like a sine wave of voltage except that the scale chosen to represent its maximum value may be different. The vector diagram looks the same because it is necessary only to label the rotating arm I instead of E and the mathematical formula reads $i = I \sin \theta$ instead of $e = E \sin \theta$. The answers to the above problems would be numerically the same if current rather than voltage had been specified.

7.5. Effective value of alternating voltage or current. Since the voltage (or current) in an a-c system rises to a maximum in one direction, then to a maximum in the other direction, the average value is zero; the time the voltage is in one direction is equal to the time the voltage is in the other direction. The needle of the usual d-c voltmeter will not move from its zero position or will merely vibrate about its zero position when placed in an a-c circuit (Sect. 3.1) even though the voltage may be several volts. Therefore, some measure other than average values must be used to compare a-c and d-c strengths. The measure actually used is that of their respective heating abilities. An alternating current is said to be equal in value to a direct current of so many amperes when it produces the same heating effect when flowing through a given resistance. This value is known as the **effective** value of the alternating current; it is equal to the maximum value multiplied by 0.707 or divided by $\sqrt{2}$. Thus

$$I_{\text{eff}} = 0.707 I_{\text{max}} = \frac{I_{\text{max}}}{\sqrt{2}} \quad \text{or} \quad I_{\text{eff}} = \frac{I_{\text{max}}}{1.414}$$

where I_{eff} = effective value of an alternating current.
I_{max} = maximum value.

The effective value is also known as the **root-mean-square** or **rms** value for the following reasons: The heating effect of a direct current is proportional to the square of the current. If the instantaneous values of an a-c current over a cycle are squared, then these squares are averaged, and finally the square root of the average is extracted, the final value will be equal to the direct current that will produce the identical heating effect. The value of current calculated in this manner is 0.707 times the maximum value. Since this value is the square *root* of the average (or *mean*) of the *squares*, it is called "root-mean-square" or "rms" current. Root-mean-square voltage is defined in the same manner.

$$E_{\text{rms}} = E_{\text{eff}} = 0.707 E_{\text{max}} = \frac{E_{\text{max}}}{\sqrt{2}}$$

and

$$E_{\text{max}} = \frac{E_{\text{eff}}}{0.707} = E_{\text{eff}} \times \sqrt{2}$$

Values given for a-c voltage and current are generally considered to be effective values unless otherwise indicated.

Example 3. What is the effective value of an alternating voltage whose maximum value is 100 volts?

$$E_{\text{eff}} = 0.707 \times 100 = 70.7 \text{ volts}$$

Problem 6. What is the maximum value of a 120-volt rms house-lighting circuit?

Problem 7. What is the effective value of an alternating current whose maximum value is 20 amp?

Problem 8. If the instantaneous value of an alternating voltage is 100 volts at 30°, what is its effective value?

Problem 9. The rms value of an alternating current is 500 ma. What is its instantaneous value at 60° phase angle?

7.6. Adding alternating currents or voltages. The two radius vectors I_1 and I_2 in Fig. 7 rotate at the same speed and

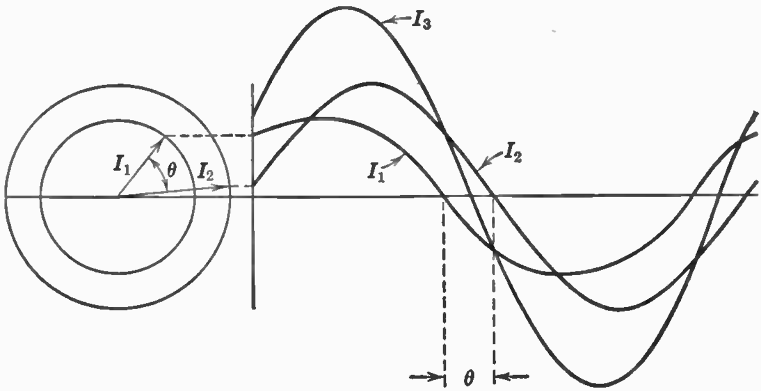


FIG. 7. Curve I_3 is the sum of I_1 and I_2 at each instant.

represent two alternating currents which have the same frequency (since the speed is the same) but which reach their maximum values at different times. The sum of these two currents may be obtained by adding at each instant the vertical heights of the two vectors and plotting these values as a function of time in seconds, or in degrees, or in radians. The result will be the new sine wave, I_3, as shown. The phase difference between the two sine waves representing I_1 and I_2 is constant, but, since the two maxima do not occur at the same instant, the maximum value of the new sine wave is not the sum of the maxima of each of the original waves. The phase angle must be taken into account. Only when the phase difference is 0 or 180° is the resultant maximum the algebraic sum of the maxima of the components (Fig. 8).

The maximum of the new wave formed by the addition of two currents I_1 and I_2 can also be obtained from a vector diagram.

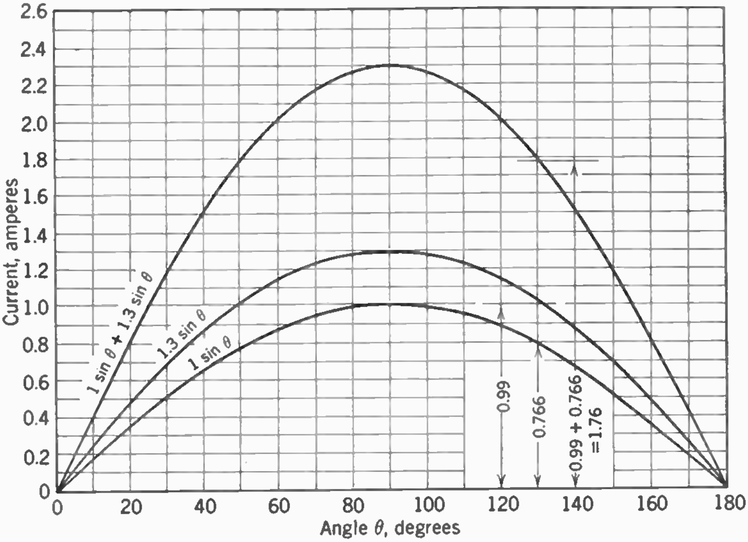


FIG. 8. How two sine waves of the same frequency but different amplitudes may be added by means of graph paper.

Let the two vectors, I_1 and I_2, form two sides of a parallelogram as shown in Fig. 9. The diagonal shown in the completed figure represents the maximum value of the resultant wave.

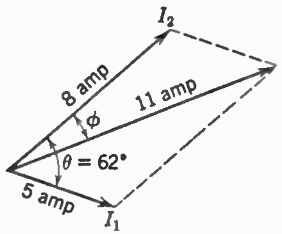


FIG. 9. Vector diagram of two sine waves combined to produce a new sine wave of 11 amp maximum value.

If currents I_1 and I_1 have different frequencies, the maximum value of the resulting wave *cannot* be obtained by completing the parallelogram as described above, since the two vectors rotate at different speeds. However, the two sine waves may be added in a manner similar to that described above. The resulting wave will not be a sine wave. As a matter of fact, all that can be said of the resulting wave at this point in the discussion is that it is non-sinusoidal and is made up of two component waves of different frequencies. Figure 10 shows the result of adding one wave of current to another of double frequency.

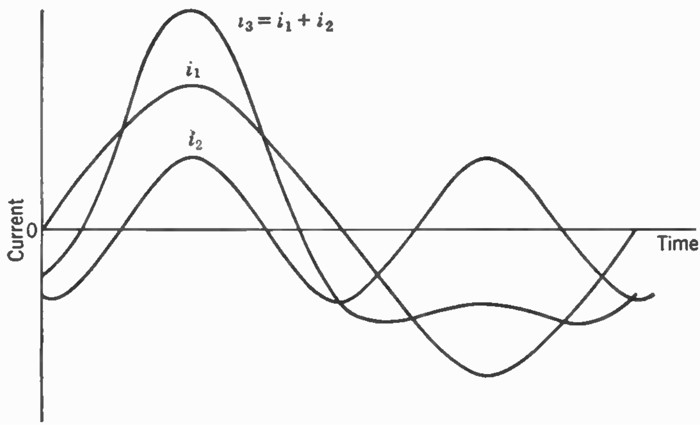


FIG. 10. The sum of two currents of different frequencies is a non-sinusoidal wave of current.

Problem 10. (*a*) On the same sheet of graph paper draw the two sine waves representing

$$i_1 = 10 \sin \theta \text{ amp}$$

$$i_2 = 5 \sin (\theta + 30°) \text{ amp}$$

(*b*) Add i_1 and i_2 graphically to obtain i_3.

Problem 11. Given:

$$i_1 = 3 \sin \theta \text{ amp}$$

$$i_2 = 4 \sin (\theta + 90°) \text{ amp}$$

(*a*) Draw a vector diagram of i_1 and i_2.

(*b*) From the vector diagram in part (*a*) determine the maximum value of i_3, the sum of i_1 and i_2.

Problem 12. Given:

$$i_1 = 10 \sin \theta \text{ amp}$$

$$i_2 = 5 \sin 3\theta \text{ amp}$$

Draw the sine waves for i_1 and i_2, then add to obtain i_3, the sum of i_1 and i_2. Note that i_3 is not a sine wave since i_2 has a frequency three times that of i_1.

7.7. Phase relations between voltage and current. Whenever an a-c voltage forces a current through a resistance, the waveforms of the voltage and current look alike; so do their mathematical formulas and their vector diagrams. This is explained by the fact that the current and voltage start at zero at the same instant, rise to a maximum value at the same instant, and carry on throughout their respective cycles in perfect step, or **in phase;** the **phase angle** between the voltage and current is zero. Note that this is a somewhat different use of the term "phase angle" from the one met with in Sect. 7.1. There, phase angle was used to denote how far the vector representing voltage (or current) had advanced from its zero position. Here, phase angle denotes the angular *difference* between voltage and current; the zero position is not of interest.

When an inductance or capacitance or any combination of these quantities with each other or with resistance is in the circuit, other phenomena take place which differ entirely from those in a d-c circuit. For example, when an a-c voltage forces a current through an inductor, the current does not attain its maximum value at the same instant as the voltage, but at a later time; when the inductor is replaced by a capacitor, the opposite is true: the maximum value of the current is reached earlier than is the maximum value of voltage.

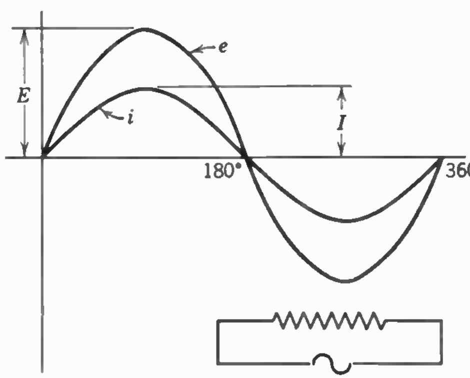


FIG. 11. Current and voltage in phase, true of a circuit containing only resistance.

CASE I. *Current and voltage in phase.* Figure 11 shows the voltage and current in phase, as would be the case in a resistive circuit. In addition to representing the voltage and current as sine waves, they may also be shown on a vector diagram as in Fig. 12. The voltage and current vectors may or may not have the same length (their length depends upon the choice of scales), but they do rotate at the same speed; they have the same frequency.

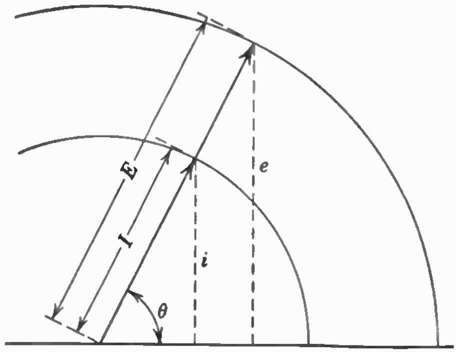


FIG. 12. Current and voltage in phase. Here the maximum value of the voltage is greater than the maximum value of the current, or it is drawn to a different scale.

In a resistive a-c circuit Ohm's law ($I = E \div R$) states the relations among current, voltage, and resistance, just as it does in a d-c circuit.

Example 4. A 55-ohm resistor is placed across a 110-volt, 60-cycle line. What current will flow through it at the 30° phase angle?

Solution: The maximum value of the voltage is

$$E = E_{\text{eff}} \times \sqrt{2}$$

$$= 110 \times 1.41 = 155 \text{ volts}$$

Since a resistance does not cause a shift of current in respect to voltage, the current is given by Ohm's law:

$$I = E \div R$$

$$= 155 \div 55 = 2.82 \text{ amp}$$

This is the maximum current. The current at any phase angle is

$$i = I \sin \theta$$

or, for a 30° phase angle,

$$i = 2.82 \sin 30° = 2.82 \times 0.5 = 1.41 \text{ amp}$$

CASE II. *Current lagging behind voltage.* Consider the case for which the current attains its maximum value at a later time than does the voltage. This is illustrated in Fig. 13. Here the current maximum is 67.5° behind the maximum value of voltage.

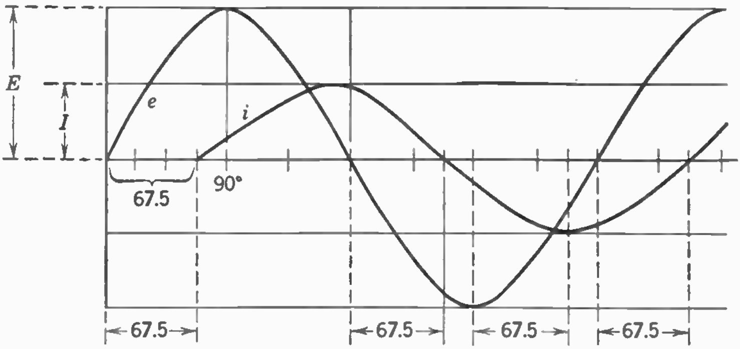


FIG. 13. Current and voltage in an inductive circuit where the maximum current lags behind the maximum voltage.

Therefore the current is said to **lag** behind the voltage by 67.5°. The current and voltage could be represented by two vectors as in Fig. 14. Here, the current vector is 67.5° behind the voltage vector; the current lags.

Example 5. A current lags a voltage by 60°. The maximum value of the current is 40 amp. What is the instantaneous value of current when the voltage has a phase angle of 75°?

Solution: Lay off on graph paper a vector diagram similar to Fig. 14. The voltage vector is drawn at a 75° angle, and the current vector is drawn 60° behind the voltage vector, or at a 15° angle. The vertical component of the current is then its instantaneous value at this instant.

The problem may also be solved by the mathematical formula

$$i = I \sin (75° - 60°)$$

$$= 40 \sin 15° = 40 \times 0.26 = 10.4 \text{ amp}$$

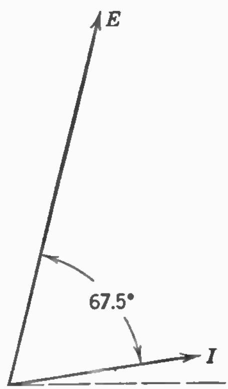


FIG. 14. Vector diagram showing lagging current. Same conditions as Fig. 13.

The cause of the lagging current is inductance which makes the maximum of current take place later than the maximum of voltage. If a circuit is purely inductive (resistance negligible), the difference between voltage and current maxima is 90°. If there is resistance, the difference is less than 90°. This matter is discussed in more detail in succeeding sections.

CASE III. *Current leading the voltage.* In this case the maximum value of the current is reached before the corresponding maximum of voltage is reached. The voltage lags behind the current, or, as it is usually stated, the current **leads** the voltage.

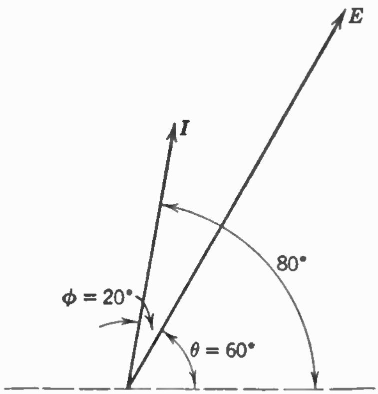


FIG. 15. Current leads the voltage at an angle ϕ of 20°. Voltage is at 60° phase.

A vector diagram for this condition is shown in Fig. 15. The instantaneous values of voltage and current are

$$e = E \sin 60°$$

$$i = I \sin 80°$$

or

$$i = I \sin (60° + 20°)$$

The cause of leading current is capacitance, which makes the maximum of current occur before the maximum of voltage. If the resistance associated with the capacitance is negligible, the current leads the voltage 90°. If the resistance is not negligible, the angle of lead is less than 90°.

The formulas applicable when the current and voltage are not in phase are

$$e = E \sin \theta$$

$$i = I \sin (\theta \pm \phi)$$

where ϕ is the phase *difference* between E and I. The plus (+) sign is used if the current leads the voltage, the minus (−) sign if the current lags the voltage.

Example 6. The effective current in an a-c circuit is 70 amp. The current leads the voltage by 30°. What is the instantaneous current when the voltage has a phase angle of 10°?

Solution: The maximum value of the current is

$$I_{max} = I_{eff} \times 1.41$$

$$= 70 \times 1.41 = 98.7 \text{ amp}$$

By the equation

$$\begin{aligned} i &= I \sin (\theta + \phi) \\ &= 98.7 \sin (10^\circ + 30^\circ) \\ &= 98.7 \sin 40^\circ \\ &= 98.7 \times 0.643 = 63.5 \text{ amp} \end{aligned}$$

The phase relation of a voltage (or current) in one part of a circuit in respect to that of a voltage (or current) in another part of a circuit is often of interest. For example, if the phase angle of the vector representing one voltage in a circuit is 0° and that of another vector representing a second voltage is 30°, it may be said that the "phase shift" is 30°. More specifically, the phase shift is 30° and the second voltage *leads* the first voltage.

7.8. Inductive reactance. Why does the maximum value of the voltage occur ahead of the maximum value of the current in Case II? The maximum induced voltage across an inductance will occur when the current through the inductance is changing most rapidly, since the induced voltage is equal numerically to the value of the inductance multiplied by the rate of change of the current. The current is changing most rapidly as it goes through zero. This current may be likened to the motion of a pendulum. When the pendulum is exactly vertical its velocity is greatest; when it is at the peak of its swing to the right or left its velocity is zero for a fraction of a second, since at this moment it comes to its greatest swing and then reverses direction. At the bottom of the swing, gravity has had its greatest effect and the bob is, at this moment, moving with its greatest speed. Similarly, in an a-c circuit, the current is changing most rapidly at the moment it is going through zero.

When alternating current flows through an inductor, there will be an alternate increase and decrease in the strength of the magnetic field about it. When the field is increasing, a counter-emf will be produced (according to Lenz's law), which opposes the increase in current. During this period energy is stored in the magnetic field. On the next quarter cycle the field collapses. This process is repeated during the negative half cycle. It is for this reason that the maxima (or any other corresponding values) of voltage and current in an inductive circuit do not occur at the same instant, the maximum of current occurring after the maximum of voltage.

If a coil of negligible resistance is placed in series with a resistance in an a-c circuit as in Fig. 16, less current will flow than if the coil were not connected. The effect of the coil which re-

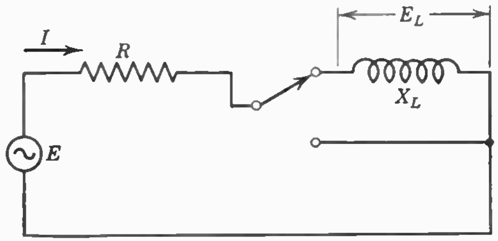


FIG. 16. The inductive reactance of a coil impedes the flow of a-c current.

duces the current is called the **inductive reactance** and is proportional to the inductance of the coil and to the frequency. The expression for inductive reactance is

$$X_L = 2\pi \times f \times L = 6.28 \times f \times L = \omega L \text{ ohms}$$

where f = frequency in cycles per second.
L = inductance in henries.
ω = Greek letter omega * = $2\pi f = 6.28f$.

If the coil in Fig. 16 has negligible resistance, the current through it is given by

$$\text{Current} = \frac{\text{Voltage across coil}}{\text{Reactance}} \quad \text{or} \quad I = \frac{E_L}{X_L}$$

Example 7. An inductive reactance of 20 ohms is connected to a 110-volt (rms), 60-cycle source. What is the maximum and the effective current? What is the instantaneous current when the voltage phase angle is 150°? What is the size of the inductance?

Solution:

$$I_{\text{eff}} = \frac{E_{\text{eff}}}{X_L} = \frac{110}{20} = 5.5 \text{ amp}$$

$$I_{\text{max}} = I_{\text{eff}} \times 1.41 = 7.8 \text{ amp}$$

* The use of ω here as equal to $2\pi f$ should not be confused with its use as an abbreviation for ohms. It is unfortunate that the same symbol has been associated with both ohms and $2\pi f$. The present-day tendency is to use the capital omega, Ω, as an abbreviation for ohms.

From the vector diagram of Fig. 17

$$\begin{aligned} i &= I \sin (\theta - 90°) \\ &= 7.8 \sin (150° - 90°) = 7.8 \sin 60° \\ &= 7.8 \times 0.866 = 6.75 \text{ amp} \end{aligned}$$

The size of the inductance is

$$\begin{aligned} L &= \frac{X_L}{\omega} = \frac{X_L}{2\pi f} \\ &= \frac{20}{2\pi \times 60} = \frac{20}{377} = 0.053 \text{ henry} \end{aligned}$$

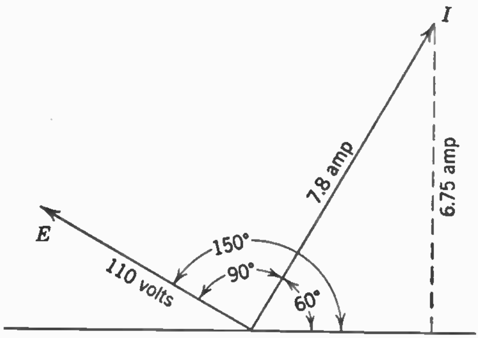


Fig. 17. Current lagging voltage by 90°.

Problem 13. Calculate the reactance of 1 henry, 1 mh, and 1 μh at the following frequencies: 100 cycles, 1000 cycles, 1 Mc.

Problem 14. A coil in a 50-volt, 1-Mc circuit passes 20 ma. What is the reactance of the coil? What is its inductance?

Problem 15. A 0.5-henry coil passes 0.2 amp of 60-cycle current. What is the voltage across the coil? What is the reactance of the coil?

Problem 16. The voltage across a 1-henry coil is 50 volts. At what frequency will the coil pass 0.1 amp?

Problem 17. A 0.1- and a 0.2-henry coil are connected in series across a 110-volt, 60-cycle line. How much current do the coils draw from the line? What would the current be if the coils were connected in parallel?

7.9. Capacitive reactance. When current flows into a condenser, it places a charge on the condenser plates and the voltage increases. When the current starts to flow into an uncharged condenser, there is no opposing voltage and the rate of current flow is maximum. As current continues to flow into the condenser, the voltage across the condenser rises and opposes the flow of current. As a result of this action the current flow is decreased, and, when the voltage across the condenser is equal to the impressed voltage, the current is zero.

The maximum of current, then, corresponds in time to the minimum of voltage, and current is said to **lead** the voltage. The minimum of current occurs when the maximum of voltage is reached. In a purely capacitive circuit (no resistance and no inductance) these maxima are 90° apart. If there is resistance in addition to capacitance the phase difference is less than 90°.

If a condenser having negligible resistance is placed in an a-c circuit as in Fig. 18, less current will flow than if the condenser

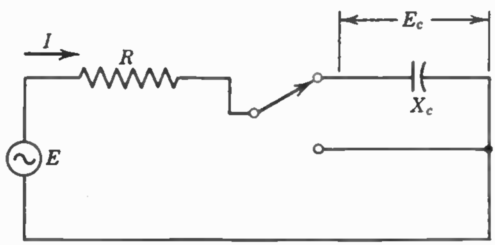


FIG. 18. The capacitive reactance of a condenser impedes the flow of a-c circuit.

is switched out of the circuit. The opposition a condenser offers to the flow of current is called the **capacitive reactance** and is inversely proportional to the capacitance and to the frequency. Thus

$$\text{Current} = \frac{\text{Voltage across condenser}}{\text{Capacitive reactance}} \quad \text{or} \quad I = \frac{E_c}{X_c}$$

Capacitive reactance is given by

$$X_c = \frac{1}{2\pi \times f \times C} = \frac{1}{6.26 \times f \times C} = \frac{1}{\omega C} \text{ ohms}$$

where f = frequency in cycles per second.
C = capacitance in farads.
$\omega = 2\pi f = 6.28 \times f$.

Example 8. A condenser has a capacitive reactance of 5 ohms at 60 cycles. It is placed in a 60-cycle circuit in which the instantaneous value of the voltage at a 20° phase angle is 48 volts. What is the maximum current through the condenser? What is the instantaneous current through it at the 20° voltage phase angle? What is the capacitance of the condenser?

Solution:

$$e = E \sin 20°$$

$$48 = E \times 0.342$$

$$E = 140 \text{ volts}$$

$$I = \frac{E}{X_c} = \frac{140}{5} = 28 \text{ amp}$$

$$i = I \sin (20° + 90°)$$

$$= 28 \sin 110° = 28 \cos 20°$$

$$= 28 \times 0.940 = 26.32 \text{ amp}$$

$$C = \frac{1}{\omega X_c} = \frac{1}{2\pi \times 60 \times 5} = \frac{1}{377 \times 5}$$

$$= 0.000530 \text{ farad or } 530\ \mu\text{f}$$

Problem 18. Calculate the reactance of a 2-μf, a 0.002-μf, and a 200-$\mu\mu$f condenser at 60 cycles, 60,000 cycles, 600 kc, and 6 Mc.

Problem 19. A condenser in a 110-volt (rms), 60-cycle circuit passes a current of 1 amp. What is the reactance of the condenser? What is its capacitance?

Problem 20. At what frequency will a 0.1-μf condenser pass 100 ma of current if the voltage across the condenser is 100 volts?

Problem 21. A 0.01-μf condenser passes 1 amp of 1-Mc current. What is the voltage across the condenser? What is its reactance?

Problem 22. How much current will a 0.5-μf condenser pass when a potential of 200 volts is across it if the frequency is 1000 cycles?

Problem 23. How much current will a 0.2-μf and a 0.5-μf condenser take from a 110-volt, 60-cycle line if connected in parallel? if connected in series?

Problem 24. (*a*) What is the reactance of a 0.2-mh coil at 1000 cycles? at 1 Mc?

(*b*) What is the reactance of a 0.1-μf condenser at 1000 cycles? at 1 Mc? Compare results in parts (*a*) and (*b*).

7.10. Comparison of inductive and capacitive reactances. Coils and condensers have opposite effects upon an alternating current. A coil causes the current to lag behind the voltage; a condenser causes current to lead the voltage. The reactance of an inductor increases as frequency increases; the reactance of a capacitor decreases as frequency increases.

A coil which passes considerable current at 60 cycles may pass practically no current at 1000 kc. On the other hand, a condenser which passes considerable current at 1000 kc may pass very little current at 60 cycles.

In circuits in which it is desired to pass a low-frequency current and to exclude a high-frequency current, a shunt (parallel) condenser and series coil are used as shown in Fig. 19(*a*). The series coil prevents the flow of the high-frequency current because of its high reactance at high frequencies, and the capacitor shunted across the circuit provides a low-impedance path for

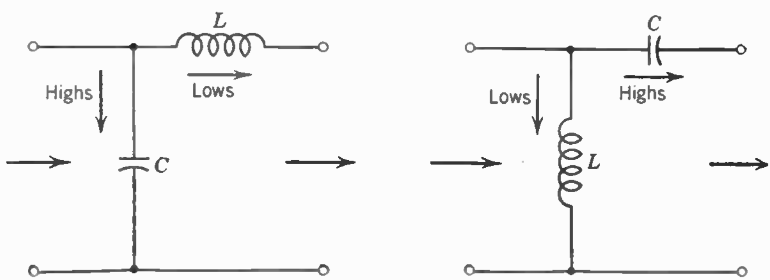


FIG. 19. Combinations of a coil and a condenser to pass low- or high-frequency currents.

the high frequencies, but the high reactance of the condenser at low frequencies prevents these frequencies from passing through this shunt circuit. The coil has low reactance to the low-frequency currents and passes these currents easily.

The coil and condenser can be interchanged as in Fig. 19(*b*) to provide an arrangement which easily passes high frequencies and at the same time presents a high reactance to low frequencies.

By extension of the ideas discussed above, various combinations of coils and condensers can be used to pass only low frequencies, high frequencies, or an intermediate band of frequencies, or to exclude a band of frequencies. Such devices are called **filters.**

Condensers that are shunted across a circuit to provide a low-reactance path for the higher-frequency currents are called **by-pass capacitors.** Coils that are used in series with the circuit to prevent the flow of the higher-frequency currents are known as **choke coils.**

7.11. Measurement of capacitance. The capacitance of a condenser may be measured by comparing the unknown capacitance with a known capacitance by means of a Wheatstone bridge as shown in Fig. 20. A and B are precision resistors known as ratio arms and are usually adjustable, and C_s is an adjustable precision condenser. The bridge is adjusted so that no sound is heard in the phones, then

$$C_x = \frac{B}{A} C_s$$

The bridge method is widely used for the precision measurement of capacitance.

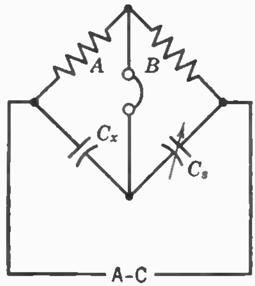


FIG. 20. Wheatstone bridge for measuring capacitance. When bridge is balanced, no sound is heard in headphones, and $C_x = (B/A)C_s$.

A simple method of measuring capacitances ranging from about 0.01 to 10 or more microfarads is to connect them in series with an a-c ammeter across a known a-c voltage. The condenser should be tested for an open or short circuit as described in Sect. 6.14 before applying the voltage. If one knows the voltage, current, and frequency, the capacitance can be calculated. The equation is

$$C = \frac{I \text{ (milliamperes)} \times 1000}{6.28 \times f \times E} \mu\text{f}$$

Small capacitances may also be measured with a Q-meter.

Problem 25. A Wheatstone bridge is used to measure a capacitance. At balance, the bridge elements have the following values: $A = 1000$ ohms, $B = 5000$ ohms, and $C_s = 0.054$ μf. What is the size of the unknown capacitor?

Problem 26. A condenser is placed in series with an a-c ammeter and connected across a 110-volt, 60-cycle line. The ammeter indicates 0.08 amp. What is the capacitance in microfarads?

7.12. Combinations of resistance with capacitance or inductance. Coils and condensers always have some resistance associated with them. This is particularly true of coils, which must be wound of wire which has at least a small resistance.

The resistance of good condensers may generally be neglected. Since both reactance and resistance impede the flow of current, the procedure for calculating the current when both resistance and reactance are in the circuit must be discussed.

Because inductances and capacitances act differently from resistances in an a-c circuit, the reactance in ohms *cannot* be added directly to the resistance in ohms to determine the resultant effect of a combination in impeding the flow of current. *Reactances and resistances must be added vectorially, not algebraically.*

7.13. Impedance. A combination of resistance and reactance is called an **impedance.** Thus, impedance is the combination of all effects, both resistance and reactance, in a circuit which impede the flow of current. The unit of impedance is the ohm.

As discussed in previous sections, the effect of capacitance and inductance is to cause the current to lead or lag the voltage by 90° whereas a resistance causes no phase shift of current in respect to voltage. Thus the effect of either capacitive or inductive reactance acts at right angles to the effect of resistance. Two factors which act at right angles to each other may be combined and the resultant determined from a formula found in plane geometry which states that "the square of the hypotenuse of a right triangle is equal to the sum of the squares of the other two sides." Then, for impedances,

$$Z^2 = R^2 + X^2$$

and, if $R = 3$ ohms and $X = 4$ ohms,

$$Z^2 = 9 + 16 = 25$$

whence

$$Z = \sqrt{25} = 5 \text{ ohms}$$

This result is called the **vectorial sum.** It is this value of Z that determines how much current will flow. The algebraic sum, obtained by simple addition—as when two resistances are combined in series—would be, in this example, 7 ohms, and the current computed from this value would be in error. It does not matter in the computation whether X is positive or negative, since the square of either a plus or minus quantity has a positive sign. Thus

$$(X)^2 = (-X)^2 = X^2$$

The resultant of combining a resistance and a reactance can be found graphically. Lay off on a horizontal line a number of units corresponding to the number of ohms of resistance. Then, if the reactance is inductive, erect a perpendicular and lay off on it a number of units equal to the number of ohms of inductive reactance. These two lines are the two sides of a rectangle as shown in Fig. 21. The length of the diagonal line represents the resultant impedance in ohms. This procedure is, of course, based on the fact that the effects of resistance and reactance upon current act at right angles to each other.

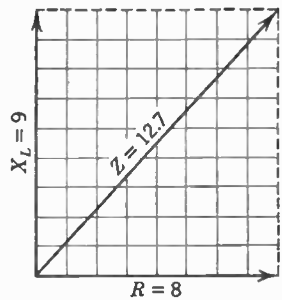


FIG. 21. How to compute impedance by vectors, using cross-section paper.

Because capacitive reactance has an effect opposite to that of inductive reactance, the line representing it should be pointed downward rather than upward. Graph paper is of great aid in solving a-c problems in this manner.

Example 9. An alternating current of 8 amp flows through a coil whose inductance is 0.043 henry and whose resistance is 5 ohms. What voltage is required if the frequency is 60 cycles?

Solution: The current in such a circuit is

$$I = E \div Z$$

$$Z = \sqrt{R^2 + X^2}$$

and

$$X = 2\pi f \times L = 6.28 \times 60 \times 0.043 = 16.25 \text{ ohms}$$

$$Z = \sqrt{5^2 + 16.25^2} = \sqrt{289} = 17 \text{ ohms}$$

whence

$$E = IZ = 17 \times 8 = 136 \text{ volts}$$

Example 10. What is the impedance at 60 cycles in a circuit in which there is a 1.66-μf condenser in series with an 800-ohm resistor?

Solution:

$$X_c = \frac{1}{2\pi fC} = \frac{10^6}{6.28 \times 60 \times 1.66} = 1590 \text{ ohms}$$

$$\begin{aligned} Z &= \sqrt{R^2 + X^2} \\ &= \sqrt{800^2 + 1590^2} \\ &= \sqrt{(64 \times 10^4) + (253 \times 10^4)} \quad \text{(approx.)} \\ &= 10^2\sqrt{317} = 1780 \text{ ohms} \end{aligned}$$

7.14. General expressions for impedance. If a series a-c circuit includes resistance, inductive reactance, and capacitive reactance, the impedance is computed as follows: Since inductive and capacitive reactance effects are opposite, a negative sign is affixed to the capacitive reactance, a positive sign to the inductive reactance. That is, a capacitive reactance is a negative reactance of so many ohms; an inductive reactance is a positive reactance of so many ohms. Before capacitive and inductive reactances may be combined vectorially with resistance, their *algebraic* sum is taken to obtain the net reactance. Thus the general expression for impedance of elements in series is

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

and after the capacitive reactance has been combined with the inductive reactance (it is actually subtracted because the signs of the two reactances are different) the form of the impedance becomes as before

$$Z = \sqrt{R^2 + X^2}$$

Here X represents the *net* reactance—the algebraic sum of the capacitive and inductive reactances. As discussed previously the sign before X^2 is positive. This is true regardless of whether X_L or X_C is the larger.

Example 11. What is the impedance of a circuit consisting of a capacitive reactance of 6 ohms, an inductive reactance of 10 ohms, and a resistance of 4 ohms, all in series?

The vector diagram for this example is shown in Fig. 22. Here X_L points upward at an angle of 90° from the resistance, and X_C points downward at an angle of 90°. The resultant of adding the reactances is a positive 4 ohms.

If, however, the values of X_L and X_C are interchanged, then the resultant is a negative 4 ohms and points downward.

In Case 1:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$
$$= \sqrt{4^2 + (10 - 6)^2}$$
$$= \sqrt{4^2 + 4^2} = 5.7 \text{ ohms}$$

In Case 2:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$
$$= \sqrt{4^2 + (6 - 10)^2}$$
$$= \sqrt{4^2 + (-4)^2}$$
$$= \sqrt{4^2 + 4^2} = 5.7 \text{ ohms}$$

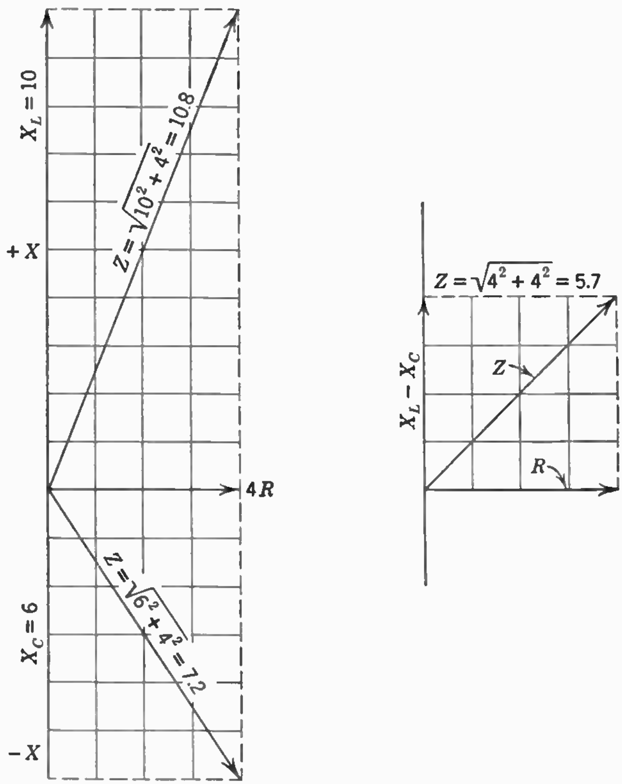


FIG. 22. At left is resultant of combining reactances and a resistance to form two Z vectors. At right, the two reactances have been added to form a net reactance of 4 ohms, which is then combined with a resistance of 4 ohms to form an impedance of 5.7 ohms.

Problem 27. The electrical elements of a broadcast receiving antenna may be represented approximately by a coil and a condenser in series (L_A and C in Fig. 23). The voltage picked up by the antenna is E. The coil L_s is a part of the input circuit to the receiver. At a frequency of 800 kc, what will be the current through the series circuit? (There is no mutual inductance between L_A and L_s.)

Problem 28. A 1000-ohm resistor, 100-μh coil, and 100-$\mu\mu$f condenser are connected in series. What will be the impedance at 1 Mc?

Problem 29. A series circuit is made up of a 25-ohm resistor and a 0.07-henry inductor. What is the impedance at 60 cycles? at 1000 cycles?

Problem 30. A 1000-ohm resistor is connected in series with a 2.5-μf condenser. What is the impedance at 60 cycles? at 200 cycles?

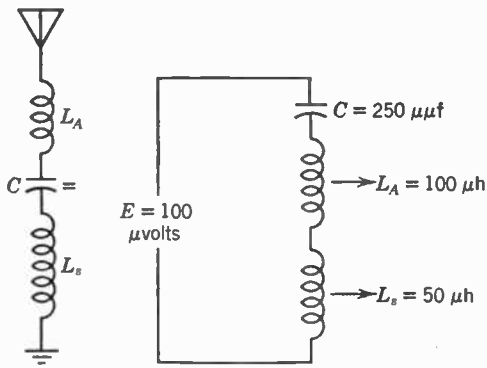


FIG. 23. An antenna and its equivalent circuit.

7.15. Series a-c circuits. In Fig. 24 is an inductance in series with a resistance. The current flowing in the circuit may

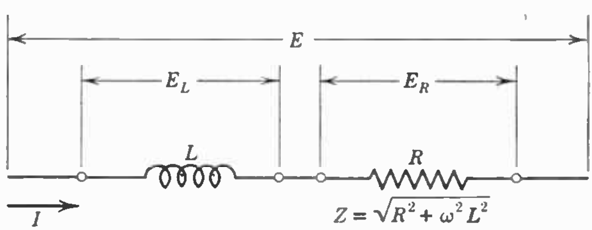


FIG. 24. A series circuit: $\omega = 6.28 \times f$.

be found by dividing the voltage across the circuit by the impedance, that is, $I = E \div Z$, in which

$$Z = \sqrt{R^2 + X^2}$$

which is quite different in numerical value from $R + X$. For example, if $R = 3$ and $X = 4$, the vector sum $Z = 5$, whereas the arithmetical sum $= 7$.

In an a-c circuit the voltage across an impedance, a reactance, or a resistance is equal to that impedance, reactance, or resistance in ohms multiplied by the current in amperes.

Voltage across a resistance	$E_R = I \times R$
Voltage across an inductance	$E_L = I \times X_L$
Voltage across a capacitance	$E_C = I \times X_C$
Voltage across L and C	$E_{L+C} = I(X_L - X_C)$
Voltage across an impedance	$E_Z = I \times Z$

These voltages are shown in Fig. 25.

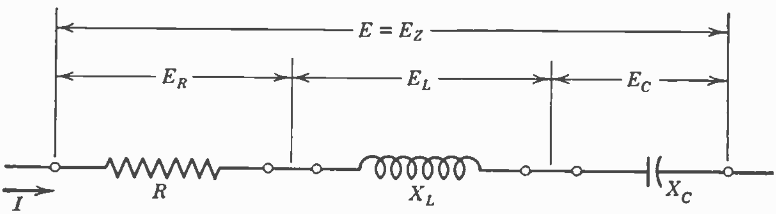


Fig. 25. Voltages across elements of an impedance.

The voltage across two resistances in series is the *algebraic* sum of the individual voltages. The same is true of voltage across two reactances, remembering that a capacitive reactance has a negative sign and that the voltage across it is negative in respect to that across an inductance. The voltage across two impedances in series, however, must be determined by adding the individual voltages *vectorially*, because the impedance is the vector sum of resistance and reactance.

The circuit of Fig. 24 will serve as a typical example. Here the current is

$$I = \frac{E}{\sqrt{R^2 + X^2}}$$

and

$$E = I\sqrt{R^2 + X^2}$$

$$E^2 = I^2(R^2 + X^2) = I^2R^2 + I^2X^2$$

$$= E_R{}^2 + E_X{}^2$$

whence

$$E = \sqrt{E_R{}^2 + E_X{}^2}$$

Therefore the resultant voltage across a resistance and a reactance is the *vector* sum of the individual voltages.

Example 12. If, in Fig. 24, $E = 15$ volts, $R = 3$ ohms, and $X = 4$ ohms, what are the current and the voltage across R and X?

Solution:

$$I = \frac{E}{\sqrt{R^2 + X^2}} = \frac{15}{\sqrt{9 + 16}} = \frac{15}{5} = 3 \text{ amp}$$

$$E_R = IR = 3 \times 3 = 9 \text{ volts}$$

$$E_X = IX = 3 \times 4 = 12 \text{ volts}$$

$$E = \sqrt{E_R{}^2 + E_X{}^2} = \sqrt{81 + 144} = \sqrt{225} = 15 \text{ volts}$$

Experiment 1. Connect a 1000-ohm, 10-watt resistor in series with a 2-μf condenser and place the combination across a 110-volt, 60-cycle source. With an a-c voltmeter measure the voltage across the resistance, the voltage across the condenser, and the total voltage. Draw a vector diagram to see whether the voltages across the resistor and condenser add vectorially to give the total applied voltage.

7.16. Vector diagrams of a series circuit. In a resistive circuit, the voltage and current are in phase; their maximum values occur at the same instant. If the circuit is purely reactive (no resistance), there is a 90° phase difference between the current and voltage. If there are both resistance and reactance in the circuit, the angle between the current and voltage is less than 90°, the exact value depending upon the relative values of R and X.

Vector diagrams furnish a convenient means of studying the relations existing in such complex series circuits. If the vector diagram is laid off on cross-section paper, the work is greatly simplified.

Since the current is the same in all parts of a series circuit, the current can be taken as the "reference vector" and the positions of the voltage vectors referred to the current vector. As an example, consider a circuit made up of 5 ohms resistance and 4 ohms inductive reactance carrying 1 amp of alternating current. In Fig. 26 the reference vector I is drawn along a horizontal line. The voltage across a resistor is the product $I \times R$, so the same line which represents I can be used to represent $E_R = IR$. The horizontal line is scaled so that one square represents 1 volt. Note that the voltage considered thus far is the voltage across the resistance; it is the only voltage in the circuit which is in phase with the current.

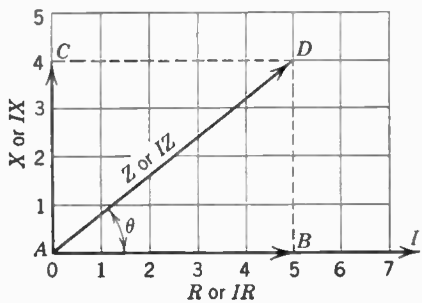


FIG. 26. Vector diagram of a series circuit containing 5 ohms of resistance and 4 ohms of inductive reactance.

The voltage across the inductance is 90° ahead of the current through it. The vector representing the reactive voltage, $E_X = IX$, is erected perpendicular to the current vector. Since $IX = 1 \times 4 = 4$ volts, the length of this voltage vector is 4 units.

The total voltage across the complete circuit is the vector sum of the voltages E_R and E_X. This vector sum is obtained by drawing the diagonal of the completed parallelogram (a rectangle in this case). The length of the diagonal gives the magnitude of the total voltage. Its direction determines its phase angle in respect to the current. With cross-section paper and a compass it is a simple matter to determine the value of the voltage, in this example approximately 6.4 volts.

The 6.4 volts required to force 1 amp through this series circuit must be the voltage across the entire circuit, and the line representing this voltage must also be representative of the impedance of the circuit. Thus, the vector diagram has IR as the horizontal vector, IX as the vertical vector, and IZ as the vector representing the total applied voltage. Each of these vectors is multiplied by the same constant factor, I. If I is divided out of each, there remain R, X, and Z. Not only does the vector diagram represent the voltages across the parts of the series circuit, it also can be used to represent the parts themselves. Furthermore the angle between the vector representing Z and R is the angle by which the voltage across the circuit and the current through it differ in phase. Clearly in this example the voltage leads the current (current lags voltage), and the angle between the directions of the IR and IZ vectors is the angle of lead. In the figure $BD \div AB$ is the tangent of the angle θ, or

$$\frac{BD}{AB} = \tan\theta$$

and, since $BD = AC = IX$ (or voltage drop across X) and $AB = IR$ (or voltage drop across R),

$$\frac{E_X}{E_R} = \frac{IX}{IR} = \frac{X}{R} = \tan\theta$$

The reactance and resistance in ohms are known, and so the tangent of the angle may be looked up in a table. When the tangent of an angle is known but not the angle, the expression is written $\theta = \tan^{-1}(X/R)$ and is read "θ (theta), the angle whose tangent is X over R."

Similarly, $X \div Z = \sin\theta$, and $R \div Z = \cos\theta$.

In Fig. 26 $\theta = \tan^{-1} 4/5 = \tan^{-1} 0.8$. From the table at the end of the book it is found that the angle whose tangent is 0.8 is between 38° and 39°. A better value is given by the following method. The total difference between tan 39° and tan 38° is $0.810 - 0.781 = 0.029$. Tan θ is $0.800 - 0.781 = 0.019$ unit greater than tan 38°. Taking the ratio of the differences,

$$\frac{0.019}{0.029} = 0.655$$

Then

$$\theta = 38° + 0.655° = 38.655°$$

This method of determining values between those given in a table is called **interpolation.**

The effect of a resistance in series with a reactance is to decrease the phase difference between the current and voltage, that is, to bring them more nearly into phase. In a pure reactance circuit, the angle is 90°; when resistance is added this angle decreases. In a pure resistance circuit the angle is zero, and the current and voltage are in phase; they reach their maximum values at the same instant. If the reactance is capacitive, the procedure is the same as above except that the X vector is directed downward.

Note that in Fig. 26 the three vectors can be used to form a voltage or impedance triangle, DAB. This is a right-angle triangle for which simple relations among the sides and the angle θ exist. If one side and the angle θ, or any two sides are known, all the other sides and angles can be computed. Using the impedance triangle, since the current I is common to all sides, the following relations can be written:

$$\cos\theta = R \div Z; \quad \sin\theta = X \div Z; \quad \tan\theta = X \div R$$

or

$$R = Z\cos\theta; \quad X = Z\sin\theta; \quad Z = R \div \cos\theta = X \div \sin\theta$$

Case I. *X and R given, to find θ and Z.* Look up in a table of trigonometric functions the angle whose tangent is $X \div R$; find the sine (or cosine) of this angle and divide it into X (or R) to obtain Z. Z may also be computed from $Z = \sqrt{R^2 + X^2}$.

Case II. *X and Z given, to find θ and R.* Here $\sin\theta = X \div Z$ or $\theta = \sin^{-1}(X \div Z)$ (read as "theta is the angle whose sine is equal to $X \div Z$"). The value of θ and of $\cos\theta$ can be looked up in a trigonometric table. Then $R = Z\cos\theta$. R may also be computed from $R = \sqrt{Z^2 - X^2}$.

Case III. *R and Z given, to find θ and X.* This is similar to Case II. The equations are

$$\theta = \cos^{-1}(R \div Z)$$

$$X = Z\sin\theta$$

$$= \sqrt{Z^2 - R^2}$$

Example 13. A series circuit has 46 ohms of resistance and 26.6 ohms of inductive reactance. By what angle does the current lag the voltage? What is the impedance?

Solution:

$$\theta = \tan^{-1}(26.6 \div 46) = \tan^{-1} 0.577 = 30°$$

$$\cos 30° = 0.866; \quad \sin 30° = 0.5$$

$$Z = R \div \cos\theta = 46 \div 0.866 = 53.2 \text{ ohms}$$

$$= X \div \sin\theta = 26.6 \div 0.5 = 53.2 \text{ ohms}$$

also

$$Z = \sqrt{R^2 + X^2} = \sqrt{46^2 + 26.6^2}$$

$$= \sqrt{2824} = 53.2 \text{ ohms}$$

Problem 31. What is the phase difference between voltage and current in each of the following series circuits? Indicate lagging current with a minus sign. (*a*) Pure resistance; (*b*) pure capacitive reactance; (*c*) pure inductive reactance; (*d*) 10 ohms resistance. 15 ohms capacitive reactance; (*e*) 20 ohms resistance, 10 ohms inductive reactance; (*f*) 50 ohms capacitive reactance, 30 ohms inductive reactance; (*g*) 10 ohms resistance, 20 ohms capacitive reactance, 15 ohms inductive reactance; (*h*) 25 ohms resistance, 50 ohms each of inductive and capacitive reactance.

Problem 32. Two resistors, one of 10 ohms and the other of 15 ohms resistance, are connected in series across a 60-cycle, 120-volt circuit. What current flows? What is the current if the frequency is increased to 500 cycles?

Problem 33. If the 15-ohm resistor in Prob. 32 is replaced with a coil which has a 60-cycle reactance of 15 ohms, what is the current? What will be the current at 200 cycles? What is the inductance of the coil?

Problem 34. A coil and resistor are in series in an a-c circuit. The voltage across the coil is measured and found to be 50 volts, that across the resistor 30 volts. What is the total voltage?

Problem 35. Two inductances are in series with two capacitors and two resistors. The inductors have reactances of 4 and 5 ohms, the condensers have reactance of 8 and 10 ohms, and the resistances are 3 and 4 ohms. The voltage across the combination is 120 volts. What is the current, what voltage appears across each separate element, and what is the phase difference between the current and the total voltage?

Problem 36. A current of 1.5 amp flows through a series circuit composed of 10 ohms inductive reactance, 15 ohms capacitive reactance, and 5 ohms resistance.

(*a*) Draw a vector diagram showing the current and all the voltages.

(*b*) What is the voltage across each element and across the entire circuit?

(*c*) What is the total impedance?

Problem 37. In a series circuit there are 30 ohms inductive reactance and 18.75 ohms resistance. It is desired to increase the phase angle to 70°. How can this be done? How can the phase angle be decreased to 30°? The resistance is to remain constant.

7.17. Characteristics of a series circuit. 1. The voltage across a series combination of resistance and reactance is the *vector* sum of the voltages across the separate units.

2. The combined resistance of several resistances in series is the *algebraic* sum of the individual resistances.

3. The combined reactance of several reactances in series, whether inductive or capacitive, is the *algebraic* sum of the individual reactances.

4. The impedance, or combined effect of a series resistance and reactance, is the *vector* sum of the total resistance and the net reactance.

5. The impedance of several impedances in series must be obtained by breaking down the individual impedances into their corresponding resistances and reactances, adding the resistances, and adding the reactances with due regard for the negative sign associated with capacitive reactances. The total impedance is then equal to

$$Z = \sqrt{(R_1 + R_2 + \cdots)^2 + (X_1 + X_2 + \cdots)^2}$$

Example 14. Table 3 shows the resultant obtained by combining elements of 3 and 4 ohms, respectively, which may be resistance or reactance, or combinations.

TABLE 3

Combination	Sum		Resultant
1. $R = 3$; $X = 4$	$\sqrt{9 + 16}$	$= Z$	5
2. $R = 3$; $R = 4$	$3 + 4$	$= R$	7
3. $X_L = 4$; $X_C = 3$	$4 - 3$	$= X$	1
4. $X_L = 3$; $X_C = 4$	$3 - 4$	$= X$	−1
5. $X_L = 3$; $X_L = 4$	$3 + 4$	$= X$	7
6. $X_C = 3$; $X_C = 4$	$-3 - 4$	$= X$	−7
7. $R = 3$; $X_L = 4$; $X_C = 3$	$\sqrt{9 + (4 - 3)^2}$	$= Z$	3.16
8. $R = 3$; $X_L = 3$; $X_C = 4$	$\sqrt{9 + (3 - 4)^2}$	$= Z$	3.16
9. $R = 3$; $X_L = 3$; $X_C = 3$	$\sqrt{9 + (3 - 3)^2}$	$= R$	3

Note in 7 and 8 above that the resultant has the same magnitude although the conditions are different. This occurs because a negative number when squared yields a positive result. The angle of the impedance for the net inductive reactance of 7 is positive, whereas it is negative in 8, which has a net capacitive reactance.

In a series circuit the reactance which is the greater determines the reactance of the resultant or equivalent series circuit.

For example, if there are 4 ohms capacitive reactance and 11 ohms inductive reactance, the circuit as a whole will be inductive.

7.18. Series resonance. When, in a series circuit, the capacitive and inductive reactances are equal, their effects cancel and the impedance of the circuit is controlled by the resistance alone. This is illustrated in part 9 of the preceding example. Since reactances vary with frequency, and capacitive reactance decreases while inductive reactance increases with frequency, this cancellation of reactance may occur as a result of changing frequency, or it may occur as a result of adjusting one of the reactances. This phenomenon is known as **series resonance** and is so important that the following chapter is devoted to it and to a similar phenomenon which occurs in parallel circuits.

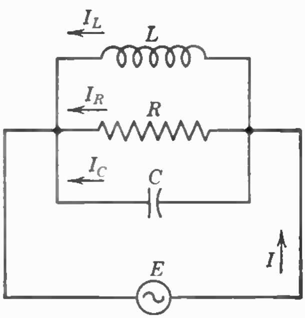


FIG. 27. In this parallel circuit I may be very small compared to the branch circuit currents I_L or I_C.

7.19. Parallel circuits. In a circuit like that in Fig. 27, in which several reactances or combinations of resistance and reactance may be connected in parallel, the following rules hold:

The *voltage* across each branch is the same and is equal to the applied voltage.

The *current* taken by any branch is equal to the applied voltage divided by the impedance of that branch.

The *total current* taken from the voltage source is the vector sum of the currents through each branch.

The *impedance* offered to the flow of current by the combination is the applied voltage divided by the total current.

To illustrate these rules assume in Fig. 27 that $E = 120$ volts, $X_C = 8$ ohms, $X_L = 5$ ohms, and $R = 3$ ohms. The branch currents, the total current, and the resultant impedance are found as follows:

$$I_C = \frac{E}{X_C} = \frac{120}{8} = 15 \text{ amp (through condenser)}$$

$$I_R = \frac{E}{R} = \frac{120}{3} = 40 \text{ amp (through resistor)}$$

$$I_L = \frac{E}{X_L} = \frac{120}{5} = 24 \text{ amp (through inductor)}$$

$$I = \sqrt{I_R^2 + (I_L - I_C)^2} = \sqrt{1681} = 41 \text{ amp}$$

$$Z = \frac{120}{41} = 2.93 \text{ ohms}$$

Thus the impedance is the ratio of the voltage across the circuit to the total current through it. Often it is necessary to compute the impedance of parallel circuits when the voltage is not given. The procedure then is to assume a convenient value of voltage, find the currents that would flow, and then divide the voltage by the total current to obtain the impedance, or

$$Z = \frac{E}{I}$$

The choice of a value of voltage is arbitrary. The impedance will be the same for whatever voltage is used.

Example 15. What is the impedance of 630 ohms capacitive reactance shunted by 100 ohms of resistance?

Solution: Assume a voltage of 100 volts.

$$I_C = 100 \div 630 = 0.159 \text{ amp}$$

$$I_R = 100 \div 100 = 1.0 \text{ amp}$$

The total current is

$$I = \sqrt{I_C^2 + I_R^2} = \sqrt{0.159^2 + 1^2}$$

$$= \sqrt{1.025} = 1.013 \text{ amp}$$

$$Z = \frac{E}{I} = \frac{100}{1.013} = 98.5 \text{ ohms}$$

7.20. Vectors in parallel circuits. Since the voltage across the various branches in a parallel circuit is the same, while the currents may differ, the impressed *voltage* is taken as the reference vector in solving problems by vector diagrams. This contrasts with the case in series circuits, in which the current is common to all elements and hence is taken as the reference vector.

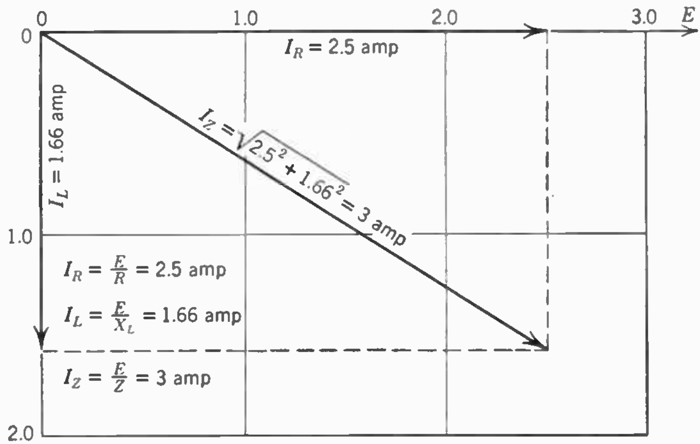


FIG. 28. Vector diagram of the parallel circuit of Example 16.

In a parallel circuit, as in a series circuit, if there is resistance only, the current and voltage are in phase. The total current taken from the source by the individual resistive branches is merely the arithmetical sum of the currents taken by the individual branches. If, however, there is reactance as well as resistance, the resultant current is found in much the same way that the resultant voltage is found in the series circuit. The method is illustrated by the following example:

Example 16. A resistance of 40 ohms is shunted by an inductance with 60.4 ohms reactance. A voltage of 100 volts is impressed across the combination. Draw a vector diagram and determine the resultant impedance of the circuit.

Solution: Lay off a horizontal line to represent the reference vector E as in Fig. 28. Since the voltage and current are in phase in a resistor, the horizontal line is also the direction of the vector for the current I_R through the resistance. The length corresponds to $E \div R = 100 \div 40 = 2.5$ amp. The current through an inductance lags the voltage by 90°. Therefore, the vector for the current I_L points straight down. Its length corresponds to $E \div X_L = 100 \div 60.4 = 1.66$ amp.

Completing the parallelogram as shown and drawing the diagonal line, the length of the diagonal is found to be 3 units, which corresponds to 3 amp. The resultant impedance is $Z = E \div I = 100 \div 3 = 33.3$ ohms.

7.21. Power in a-c circuits. In a d-c circuit the power is the product of the voltage across the circuit and the current through it. If, in a d-c circuit, a potential of 100 volts forces 1 amp of current through a device, the power used is 100 watts.

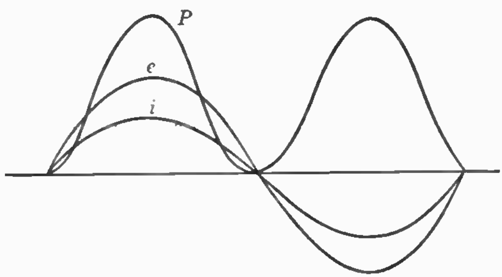


Fig. 29. In a resistive circuit, all power delivered by the generator is consumed by the external circuit. Voltage and current are in phase. The curve of instantaneous power has twice the frequency of the voltage and current curves.

In a-c circuits the voltage and current are not always in phase. On the contrary, there may be a large phase difference between them. The questions then arise: Is the actual power in an a-c circuit the product of voltage and current? Does the phase difference between voltage and current affect the amount of power taken by a device?

In all electrical circuits the power at any *instant* is the product of the instantaneous voltage and current. In the resistive case shown in Fig. 29 for which voltage and current are in phase the instantaneous power is shown by the curve P. All the P-curve lies above the axis and the instantaneous power is positive at all times.

When the voltage and the current are not in phase as in Figs. 30 and 31, a part of the power curve lies below the axis and this part represents negative power; that is, power is being returned by the circuit to the source rather than being used by the circuit. Negative power is indicated by the shaded areas. In Fig. 30

the positive and negative areas of the power curve are equal and as much power is returned as is used; there is no net power supplied by the source. In Fig. 31 there is only a small amount of negative power: the source supplies power most of the time.

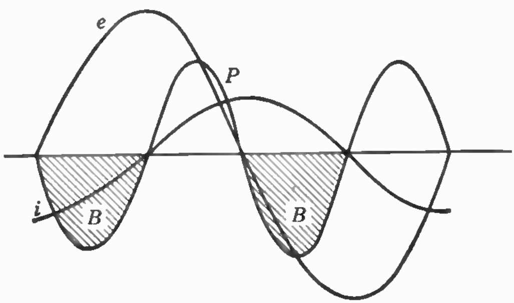


FIG. 30. In a purely reactive circuit, the average power is zero, as much being delivered back to the generator (shaded areas) as is delivered to the circuit. The curve of instantaneous power is a double-frequency curve.

The power actually consumed in a circuit is considered positive power, whereas power returned to the generator by the circuit is called negative power. All the power fed to a resistor by a generator is consumed by the resistor, and so it is all positive

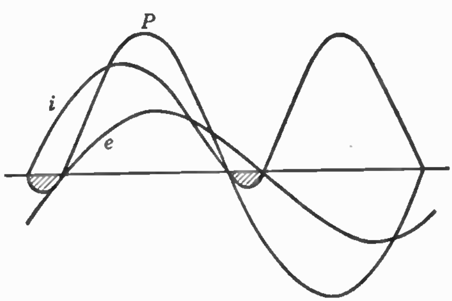


FIG. 31. In this case the circuit is slightly capacitive; some power is stored in the condenser, but most of it is used up in the circuit resistance.

power. On the other hand any power that is fed to a pure capacitance or a pure inductance on one half of a cycle is returned to the generator the next half cycle, and the average power consumed is zero. If the load is composed of both resistance and reactance, some power is consumed by the resistance, and the average power is positive.

The ability of a reactance to return instantaneous power to the generator arises from the fact that a reactance can store energy. Consider first an inductance with negligible resistance. As the voltage across the inductance rises, the inductance tends to keep the current from increasing, and power must be supplied by the generator to cause the current to grow. As more power is supplied the current becomes greater and more energy is stored in the magnetic field of the inductor ($W = \frac{1}{2}LI^2$). As the voltage wave continues through its maximum and then decreases toward zero, the current continues to rise and more and more energy is stored in the field. As the voltage returns to zero, the maximum current is flowing and there is maximum energy storage in the field. As the voltage reverses and starts to increase in the opposite direction, the current continues to flow in the original direction even though it does decrease in magnitude. It is during this interval when the current flows in one direction while the voltage is in the opposite direction that power (and energy) is returned to the generator. In a purely inductive circuit the same amount of energy is returned *to* the generator in one half cycle as was taken *from* the generator the previous half cycle, and the net energy (and power) required by the circuit is zero. The intervals during which power is being returned to the generator in an inductive circuit are the shaded areas marked B in Fig. 30.

A capacitor stores energy, which depends upon the voltage across it ($W = \frac{1}{2}CE^2$). Thus the maximum storage of energy in a condenser occurs at the maximum of voltage. At this time the current through the condenser is zero and immediately thereafter starts to increase in the opposite direction. Again, as with the inductor, the interval during which the voltage is in one direction and the current is in the other direction is the interval during which energy is being returned to the generator. The net power consumed by a pure capacitance is zero.

When a circuit is composed of both resistance and reactance, the phase shift between voltage and current is less than 90° as discussed in previous sections. In this case the time during which positive power is taken from the generator is correspondingly greater than the time during which the circuit returns power to the generator. Thus there is a net flow of power from the generator. This is illustrated in Fig. 31 for a resistor-condenser circuit.

The instantaneous power taken by a circuit is continually changing even though a net quantity of positive power is required. The average power that is required by the load circuit is of extreme importance, since it is this average power that determines the size of the generator and the amount of useful work that can be done in the load circuit. The **average** power, also called **actual** or **true** power, is

$$P = I^2R \text{ watts}$$

The quantity which results from the multiplication of the voltage and current is called the **apparent** power. Its relation to the average or true power depends on the ratio of resistance to reactance in the load circuit. The apparent power, then, is

$$P_A = EI \text{ volt-amperes}$$

Note that the units of apparent power are **volt-amperes,** not watts.

The ratio of the average to the apparent power is known as the **power factor** (pf):

$$\text{pf} = \frac{P}{P_A}$$

$$= \frac{I^2R}{EI} = \frac{IR}{E}$$

Since $E = IZ$

$$\text{pf} = \frac{IR}{IZ} = \frac{R}{Z} = \frac{R}{\sqrt{R^2 + X^2}}$$

The power factor of a circuit, then, is the ratio of its resistance to its impedance. In radio circuits one is usually more interested in the reactance and the designer keeps the resistance as low as possible; in power circuits the opposite is often true. A condenser or inductor is normally placed in a radio circuit because of the useful properties of such a device. Resistance tends to obscure these useful properties and so must be as small as possible.

A pure inductance or a pure capacitance will not take any power from the line; thus, a pure capacitance or inductance when plugged into a lamp socket will not cause the electric watt-hour meter to turn. However, there is resistance in the line leading back to the power company's generator. The current which flows through the reactance must flow through this resistance and in so doing consumes power. Thus, the power company must supply some power even though the customer is not himself consuming power. Of course, if there is appreciable resistance in the customer's line from the meter to the lamp socket, some power will be consumed in this resistance and the meter will indicate the amount of energy represented by this power.

Another useful expression for average power can be derived as follows:

$$P = P_A \times \text{pf} = P_A \times \frac{R}{Z}$$

but

$$\frac{R}{Z} = \cos\theta \quad \text{and} \quad P_A = EI$$

then

$$P = EI\cos\theta$$

where E and I are the *rms* values of voltage and current and θ is the angle between voltage and current.

Since $\cos\theta$ may have any value between 0 and 1, the power factor may be expressed as a percentage. For example, 90 per cent power factor has several meanings: pf = 0.9; $\cos\theta = 0.9$; $R/Z = 0.9$. All these expressions are, of course, equivalent.

Unity power factor characterizes a circuit containing only resistance, or a circuit for which the equivalent impedance is a pure resistance. In a purely reactive circuit the power factor is zero.

Example 17. A 220-volt a-c motor takes 50 amp from the line. A wattmeter shows that the motor uses 9350 watts. The apparent power, however, is $E \times I$ = 11,000 volt-amperes. The power factor is average power divided by apparent power or 9350 ÷ 11,000 or 0.85 or 85 per cent.

Power P taken by motor from line = $EI\cos\theta$ = 9350 watts

Current taken from line = $P \div E\cos\theta$ = 50 amp

In-phase current = $P \div E$ = 42.5 amp

Therefore, 50 − 42.5 = 7.5 "extra" amp flow through the line and from the generator supplying the system. The current would be only 42.5 amp if the motor operated at unity power factor. The extra current represents a waste because of the additional voltage and power loss in the resistance of the supply line.

Example 18. What is the power loss in the line and what voltage must be supplied by the generator in Example 17 if the line resistance is 0.1 ohm? If the motor is changed so that it operates at unity power factor and takes the same power from the line, what will be the generator voltage and the power loss in the line?

Solution:

CASE I. $pf = 0.85$.

$$\text{Voltage drop in line} = IR = 50 \times 0.1 = 5 \text{ volts}$$

$$\text{Generator voltage} = 220 + 5 = 225 \text{ volts}$$

$$\text{Power loss in line} = I^2R = 50^2 \times 0.1 = 250 \text{ watts}$$

CASE II. $pf = 1.00$. The current taken from the line in this case will be 42.5 amp as calculated in the previous example. Then

$$\text{Voltage drop in line} = IR = 42.5 \times 0.1 = 4.25 \text{ volts}$$

$$\text{Generator voltage} = 220 + 4.25 = 224.25 \text{ volts}$$

$$\text{Power loss in line} = I^2R = 42.5^2 \times 0.1 = 180.6 \text{ watts}$$

Note that the change in power factor has decreased the power loss in the line by about 70 watts.

Example 19. A single-phase induction motor operates at 440 volts and delivers 12 hp. The motor is 89.5 per cent efficient and has a power factor of 84 per cent. What current and power are taken from the line?

$$\text{Power taken by motor} = \frac{\text{Power delivered}}{\text{Motor efficiency}} = \frac{12 \times 746}{0.895} = 10{,}000 \text{ watts}$$

This is the real (average) power taken from the line.

$$\text{Current taken from the line} = \frac{\text{Real power}}{E \cos \theta} = \frac{10{,}000}{440 \times 0.84}$$

$$= 27.0 \text{ amp}$$

If the power factor of a circuit is known, the ratio of the resistance to the impedance is also known, but the resistance itself cannot be computed without additional information. A motor may have a power factor which is more or less independent of the current drawn or the power delivered. That is, its ratio of resistance to impedance is nearly constant. The current taken from the line will depend upon the power the motor is called on to deliver. The **effective** or **equivalent resistance** of the motor is that resistance which would require the same amount of power from the line, that is, $R_{eq} = P \div I^2$. Now, since the ratio of R to Z is known (pf $= R/Z$), both Z and X can be computed. Most of the equivalent resistance, R_{eq}, of the motor results from power supplied to the load. As a matter of fact, if the resistance of the motor windings is a very large fraction of R_{eq}, the motor will be very inefficient.

The concept of equivalent resistance discussed above is not restricted to motors. The same ideas apply in any complex circuit. If the power absorbed by the circuit and the current taken by the circuit are known or may be measured, the equivalent resistance for the circuit may be computed by $R_{eq} = P \div I^2$.

Problem 38. A 10-μf condenser is shunted across a 2000-ohm resistor. What will be the ratio of current in the condenser to that in the resistor at 100 cycles? This is a typical "by-pass" arrangement. Hint. Assume a convenient value of voltage.

Problem 39. A typical choke coil has an inductance of 15 henries and a resistance of 200 ohms. What is the power factor of the coil at 120 cycles?

Problem 40. A 0.001-μf mica condenser has a power factor of 0.0002 (0.02 per cent) when operated at 1 Mc. What is its equivalent series resistance at this frequency? Note that power factor $= R/X$ (approx.) for very small values of power factor.

Problem 41. The reactances of a coil and of a condenser are each 500 ohms at 1000 cycles. Compute the reactance of each at the following frequencies: 100 cycles, 250 cycles, 500 cycles, 2000 cycles, 10,000 cycles. Hint. Use simple proportion.

Problem 42. A resistor and condenser are in series. The resistance and reactance in ohms are the same at a certain frequency. What is the ratio of resistance to impedance, Z/R, at this frequency? What will be the ratio at a frequency twice as great?

Problem 43. The name-plate data on an a-c motor follow:

$\frac{1}{4}$ hp	60 cycles
120 volts	pf $= 0.9$

What is the full-load current of the motor?

Problem 44. A coil with a reactance of 30 ohms, a condenser with a reactance of 60 ohms, and a resistor with a resistance of 50 ohms are connected in parallel, and the combination is placed across a 220-volt source. Draw a vector diagram, determine the current through each branch, and determine the total current taken by the network.

Problem 45. Given:

$$e = 100 \sin \theta \text{ volts}$$

$$i = 2 \sin (\theta + 30°) \text{ amp}$$

Draw the sine waves of voltage and current. Then sketch the instantaneous power curve. (See Fig. 31.) Shade areas of negative power. Is this the current in an inductive or capacitive circuit?

Problem 46. (*a*) What is the maximum energy stored in a 1-μf condenser when 110 volts (rms) are connected across its terminals? How many amperes (rms) flow if the frequency is 1000 cycles?

(*b*) What is the maximum energy stored in a 10-henry coil when connected across a 110-volt (rms), 60-cycle source?

7.22. Speed of radio transmission. Radio waves traveling through space, like light, travel at the rate of 186,000 miles per second (300,000,000 meters per second). Although this is a very high velocity, it is a perfectly definite value and is not infinite. When radio waves travel through materials like glass whose dielectric constant is greater than unity, the speed of transmission is reduced. The velocity in any dielectric medium is given by

$$v = \frac{c}{\sqrt{k}}$$

where v is the velocity in the medium, c is the velocity in space, and k is the dielectric constant. If c is in miles per second, then v will be in miles per second, and so on.

The speed with which electrical waves move along wires is always somewhat less than the speed in space, and may be as low as 50,000 miles per second or even less on some electrical cables. The speed on open-wire telephone lines at high frequencies is very close to the free-space value.

Important applications have been made of the fact that radio signals travel at known speeds. For example, a radio signal sent out from a transmitter may be reflected back to the region of the transmitter if it strikes an expanse of metal such as an airplane. The time required for the signal to go out to the plane and to be returned to the transmitter is a measure of the distance of the airplane from the transmitter. This is the principle upon which radar is based.

A radio signal travels 186,000 miles per second, or 0.186 mile per microsecond. This is equal to $0.186 \times 1760 = 327$ yards per microsecond. Since it takes the wave as long to go out as to return, if the time interval between sending out the signal and receiving it back is 10 μsec, the reflecting object is 1635 yards distant.

7.23. Wavelength. When higher and higher frequencies are used, as is the trend today, the numbers designating these frequencies become cumbersome and it is more convenient to express the frequency characteristic of a voltage or current wave in another fashion, the **wavelength.** The wavelength method is also a convenience in the design of antennas, wave guides, and other similar apparatus which must have definite physical sizes determined by the frequency of operation.

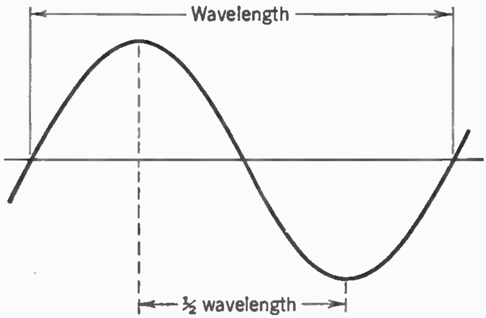


FIG. 32. A wavelength is the distance between any two corresponding portions of two waves. Thus, the wavelength may be the distance between two positive (or two negative) peaks, or between the points shown above.

Since the frequency is the number of complete cycles per second and since radio waves travel at a fixed speed, it follows that a complete cycle occupies a certain amount of space. Figure 32 pictures one of the cycles. The distance apart, in space, of two corresponding parts of two waves, say the two positive or negative crests, or the points where the two waves cross the zero axis in a given direction, constitutes the wavelength. The wavelength of radio waves is given by

$$\text{Wavelength} = \frac{\text{Velocity of radio waves}}{\text{Frequency}}$$

$$= \frac{300 \times 10^6}{f\ \text{(cycles)}} = \frac{300 \times 10^3}{f\ \text{(kilocycles)}}\ \text{meters}$$

The customary symbol for wavelength is the Greek letter lambda (λ).

The wavelength of waves traveling slower than 300×10^6

meters per second, such as those on a telephone line, is calculated by dividing the actual velocity of the waves by the frequency. Unless otherwise specified or implied, values of wavelength are given as the free space values.

Table 4 shows the correspondence between frequency and wavelength (in free space) for several values of each.

TABLE 4

Frequency, Megacycles	Wavelength, Meters
3	100
30	10
300	1
3000	0.1

Problem 47. What free-space wavelength corresponds to 60 cycles? 1500 kc? 20 Mc? 6000 Mc?

Problem 48. What frequency corresponds to a free-space wavelength of 60 meters? 75 cm? 1 cm?

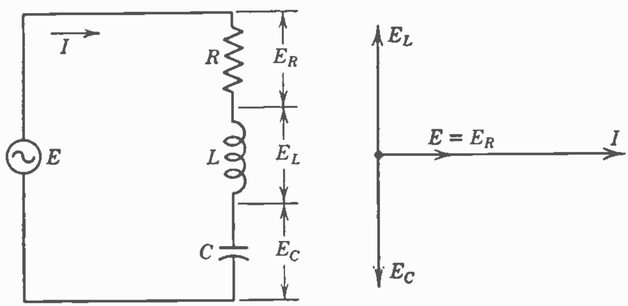
Problem 49. A 30,000-cycle wave on a transmission line travels at nine-tenths the speed of light. What will be the actual wavelength of the wave on the line?

Problem 50. An antenna is to be cut to a length of one-half wavelength and is to operate at a frequency of 30 Mc. Waves travel along the antenna at 0.95 times the speed of light. How long must the antenna be?

8 · Resonance

Some of the most important circuits in radio receivers and transmitters are those in which series or parallel resonance occurs. In transmitting and receiving systems resonance is used to build up large voltages and currents at certain desired frequencies and to discriminate against undesired signal frequencies by keeping their voltages and currents low. When one tunes a radio receiver, he actually is adjusting the receiver circuits so that a condition of resonance exists. Resonance is one of the most interesting of a-c circuit phenomena.

8.1. Series resonance. A simple circuit in which series resonance can occur is shown in Fig. 1(*a*). Although resonance



(*a*) Series-Resonant Circuit (*b*) Voltages at Resonance

Fig. 1. A circuit in which series resistance may occur.

can be attained by varying either the frequency or one of the reactances, varying the frequency will be considered first.

At zero frequency, that is, at direct current, there will be zero current in such a circuit because a condenser will not pass direct current. At very low frequencies, the reactance of the condenser is very high, so that little current flows. At very high frequencies the reactance of the coil becomes very great and again little current will flow. At intermediate frequencies the reactances of the coil and of the condenser have moderate values and more current flows. When the frequency is such that the reactances are exactly equal, the current is maximum; the circuit is **series resonant.**

The reason for this situation is simple—the effects of capacitive reactance and inductive reactance are exactly opposite, capacitive reactance decreasing and inductive reactance increasing as the frequency is increased. Furthermore, capacitance produces a *negative* reactance, inductance a *positive* reactance. The net reactance in a series circuit, therefore, is the difference between the two reactances. If these reactances are exactly equal, their difference is zero, and therefore there is no net reactance in the circuit. At resonance this is just what has happened: the positive inductive reactance has been balanced by the negative capacitive reactance, and under these conditions all that limits the flow of current is the resistance of the circuit.

When a series circuit is resonant, the current through and the voltage across it are in phase, the current is a maximum, the impedance is a minimum, and the voltages across the coil and condenser are equal and opposite in sign and may be much greater than the impressed voltage. As a matter of fact the voltage across the condenser may be great enough to puncture it. A typical vector diagram of the voltages in a series circuit at resonance is shown in Fig. 1(*b*).

The voltage across the condenser at resonance is numerically equal to the voltage across the entire circuit multiplied by the factor X_L/R or X_C/R ($X_L = X_C$ at resonance). The factor X_L/R or X_C/R, where X_L and X_C are the reactances at the resonant frequency, is called the Q of the circuit, or, more precisely, the Q_0 of the circuit, where the subscript denotes that the Q is to be calculated at resonance. This factor, Q_0, is an important and useful quantity and will appear many times in subsequent discussions. It is discussed in greater detail in Sect. 8.7.

Suppose that, instead of the frequency, the inductive reactance in Fig. 1(*a*) is varied. This may be done by varying $L(X_L = 2\pi fL)$. When the inductive reactance is zero the impedance is equal to $\sqrt{R^2 + X_C{}^2}$ and is labeled Z_1 in Fig. 2(*a*). As inductance is added to the circuit the impedance is given by

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

If a small amount of inductive reactance is added, the resultant impedance is shown as Z_2 in Fig. 2(*b*), and it is noted that the value of Z_2 is less than the value of Z_1. If the added inductive reactance is equal to the capacitive reactance as in Fig. 2(*c*), the impedance becomes equal to the resistance R alone, and $Z_3 = R$. If still more inductive reactance is added, the impedance, which

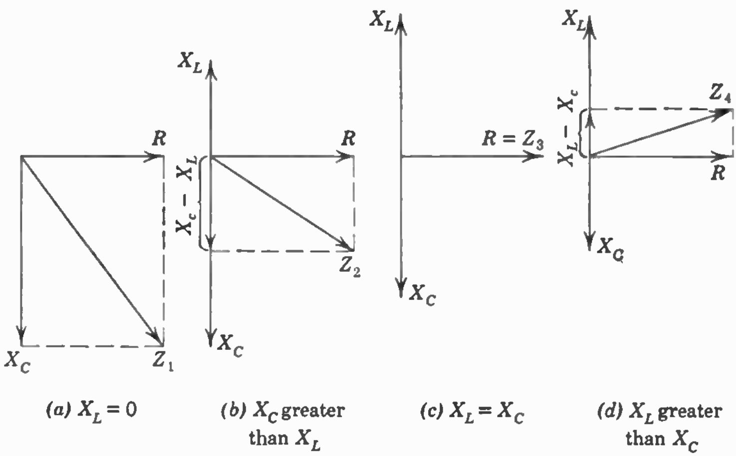


FIG. 2. Effect of varying X_L in a series *RLC* circuit.

has been decreasing, begins to increase as in Fig. 2(*d*). Thus, the impedance has its minimum value when the inductive and capacitive reactances are equal, that is, when the circuit is series resonant. The manner in which the impedance varies for the several cases is shown in Fig. 3.

If the voltage applied to a series circuit capable of resonance is held constant, then the current must follow an inverse relation in respect to the impedance; that is, as the impedance decreases the current must increase and vice versa. A graph of I against frequency is shown in Fig. 4. This is called a **resonance curve.** A similar curve would result if the capacitance or the inductance instead of the frequency had been varied. The curve is symmetrical about the resonant point when inductance is the variable element; it is somewhat non-symmetrical when frequency or capacitance is varied. The major differences in the curves for different choices of the variable element appear away from the resonant "hump." Near the region of resonance the curves are nearly the same whichever factor is varied.

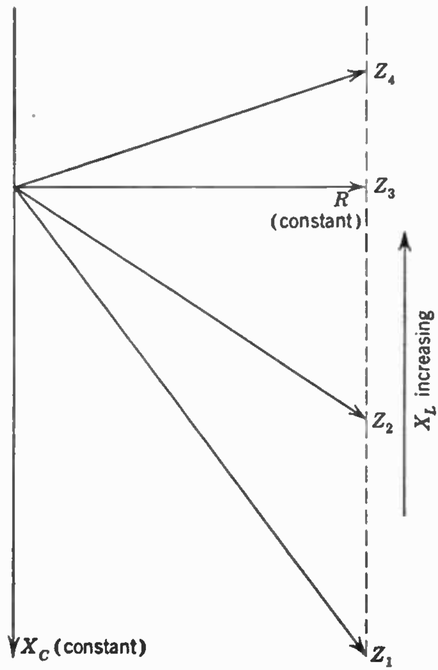


FIG. 3. Manner in which impedance varies in a series RLC circuit in which X_L varies.

The amount of current that will flow in a circuit at resonance is determined by the resistance alone if the applied voltage is constant. The lower the resistance, the higher the current at resonance and the sharper will be the resonant rise in current. Therefore, in circuits in which series resonance is a desirable characteristic the resistance should be low.

A good idea of what goes on in a series-resonant circuit may be obtained from an experiment which is a simplified version of what goes on when a radio receiver is tuned to a station.

Experiment 1. Connect in series, as shown in Fig. 5, a coil of about 200 μh, a variable capacitance of about 1000 $\mu\mu$f, a resistance of about 10 ohms, and a radio-frequency ammeter. Couple the inductance loosely to a r-f generator so that at resonance the maximum current will produce a reading at the top of the r-f ammeter scale. [The circuit is equivalent

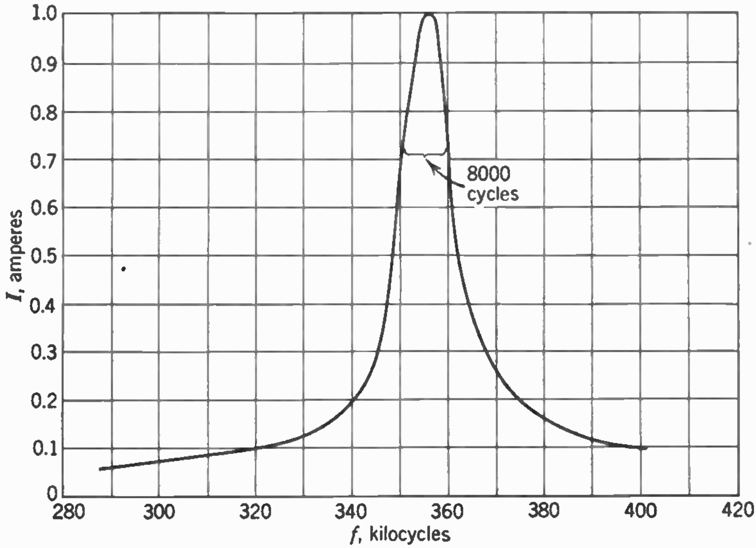


FIG. 4. The resonance curve of a circuit like that of Fig. 1.

to that in Fig. 1(a) if the voltage induced in L is represented as a generator in series with L.] Then vary the capacitance, recording the current that flows at each value of capacitance. (If the condenser is fitted with a knob

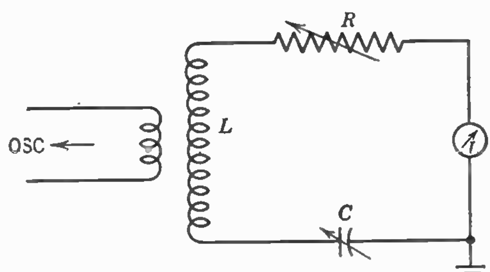


FIG. 5. When L is coupled loosely to an oscillator and C is varied, the current indicated at I will go through a maximum like that in Fig. 4.

which has a dial on it, the dial readings may be recorded rather than the capacitance.) Plot these data as a curve as in Fig. 6. Now set the condenser to the resonant value so that maximum current flows. Then vary the frequency of the generator above, through, and below resonance, plotting the data secured. Change the value of the resistance and repeat both experiments.

Calculate the voltage across the condenser and across the coil and the phase angle between current and voltage, and plot the data for both experiments outlined above.

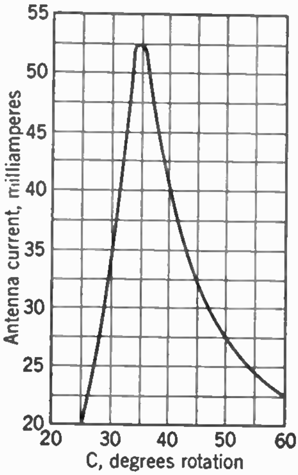


Fig. 6. How the antenna current of a radio station varies as the antenna condenser is varied.

8.2. Characteristics of a series-resonant circuit. The impedance of a series circuit capable of resonance varies over a wide range as the frequency is changed. At frequencies lower than resonance the impedance is mainly capacitive reactance; at resonance the impedance is a pure resistance; above resonance the impedance is mainly inductive reactance. The further the frequency is from the resonant value, either above or below, the greater will be the reactance (also the impedance) and the smaller the current that will flow.

At frequencies below resonance, where the capacitive reactance predominates, the current leads the voltage in phase; at resonance the current is in phase with the voltage; above resonance the current lags behind the voltage. Thus, the transition from leading to lagging current occurs as the frequency increases and passes through the resonant value. The voltage referred to in the above discussion is, of course, the total voltage across the circuit.

At all frequencies, the voltage across the inductance is 90° ahead of the current, as indeed it must always be, and the voltage across the condenser is 90° behind the current. Between these two reactive voltages, then, is a 180° phase difference; that is, the two voltages are exactly out of phase. Their resultant voltage may be found by subtracting the lesser from the greater. The resultant of the voltages across the reactances and the resistance, which is the total applied voltage, is the vector sum of the voltages across the elements. Thus

$$E = \sqrt{E_R^2 + (E_L - E_C)^2}$$

At any frequency other than the resonant value, one of the two reactive voltages is greater than the other. When added vectorially to the IR drop in the circuit, the resultant is the voltage which is impressed across the entire circuit. At resonance, however, the two voltages are equal in magnitude and opposite in phase so that the result of combining the reactive voltages is zero; the applied voltage is equal to the drop across the resistor.

The vector sum of the reactive and resistive voltages is equal to the impressed voltage.

At resonance the reactances are equal to each other and equal to $\sqrt{L \div C}$, that is, $X_L = X_C = \sqrt{L \div C}$. For example, the reactance of a capacitance of 1000 $\mu\mu$f and an inductance of 200 μh at resonance is

$$X_L = X_C = \sqrt{\frac{L}{C}}$$

$$= \sqrt{\frac{200 \times 10^{-6}}{1000 \times 10^{-12}}} = \sqrt{0.2 \times 10^6}$$

$$= 0.447 \times 10^3 = 447 \text{ ohms}$$

At resonance the inductive and capacitive reactances in the general equation for impedance $Z = \sqrt{R^2 + (\omega L - 1/\omega C)^2}$ cancel out, that is, $\omega L - 1/\omega C = 0$, so that the resultant impedance is the resistance alone,

$$Z = R \quad \text{(at resonance)}$$

Example 1. What are the voltages and phase relations in a series circuit like that of Fig. 5 at a frequency of 370 kc if I, X_C, X_L, and R are as follows:

$$I = 0.274 \text{ amp}$$

$$X_C = 430 \text{ ohms}$$

$$X_L = 466 \text{ ohms}$$

$$R = 10 \text{ ohms}$$

then

$$E_R = I \times R = 0.274 \times 10 = 2.74 \text{ volts}$$

$$E_C = I \times X_C = 0.274 \times 430 = 118 \text{ volts}$$

$$E_L = I \times X_L = 0.274 \times 466 = 128 \text{ volts}$$

$$E_{R+L} = \sqrt{2.74^2 + 128^2} = 128 \text{ volts} \quad \text{(approx.)}$$

$$\phi_{R+L} = \tan^{-1} \frac{128}{2.74} = \tan^{-1} 46.6 = 88.77°$$

$$E_{R+C} = \sqrt{2.74^2 + 118^2} = 118 \text{ volts} \quad \text{(approx.)}$$

$$\phi_{R+C} = -\tan^{-1} \frac{118}{2.74} = -\tan^{-1} 43 = -86.68°$$

$$E_{L+C} = I(X_L - X_C) = 0.274 \times 36 = 9.8 \text{ volts} = E_X$$

$$E = I\sqrt{R^2 + X^2} = \sqrt{E_R{}^2 + E_X{}^2} = \sqrt{2.74^2 + 9.8^2} = 10.2 \text{ volts}$$

$$\phi_{R+L+C} = \tan^{-1} \frac{X}{R} = \tan^{-1} \frac{X_L - X_C}{R} = \tan^{-1} \frac{466 - 430}{10}$$

$$= \tan^{-1} 3.6 = 74.48°$$

8.3. Reactance diagrams. One of the best means of showing how the reactance of a circuit varies with frequency is to draw a plot known as a **reactance diagram.**

The reactance of an inductor increases directly with frequency ($X_L = \omega L$), whereas the reactance of a capacitor varies inversely with frequency ($X_C = 1/\omega C$). It must be remembered that capacitive reactance carries a negative sign. Reactance diagrams of an inductor alone and of a capacitor alone are shown in Figs. 7 and 8.

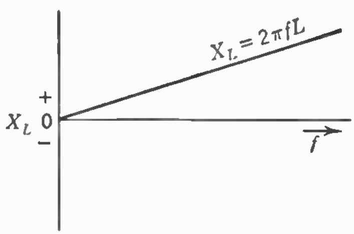


FIG. 7. Reactance curve for an inductance.

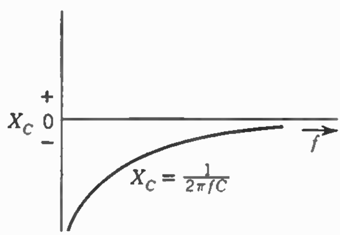


FIG. 8. Reactance curve for a capacitance.

Consider a series circuit made up of a 300-μh coil and an 84.3-$\mu\mu$f condenser. The total reactance of a series circuit is the difference between the inductive and capacitive reactances ($X = X_L - X_C$). The inductive reactance X_L and the capacitive reactance X_C as functions of frequency are plotted in Fig. 9. The difference, which is the total reactance, is found by subtracting the capacitive reactance at a given frequency from the inductive

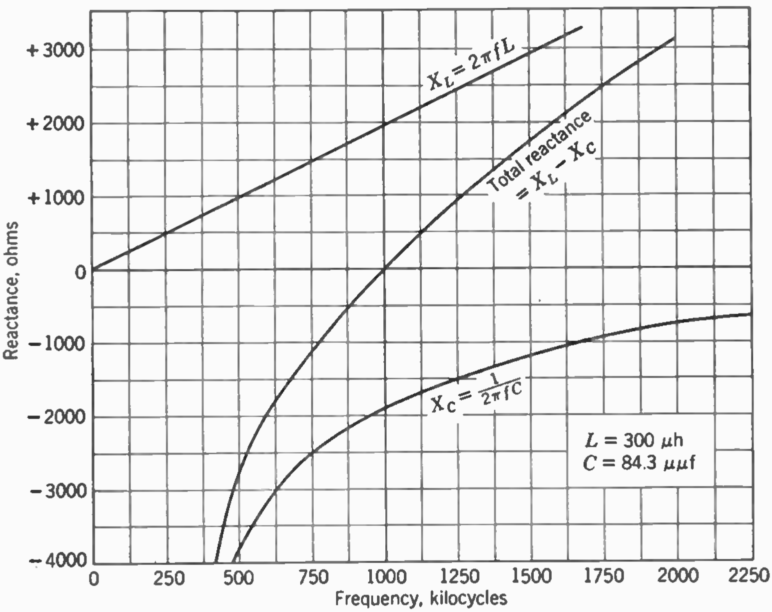


FIG. 9. A plot of the reactance of a coil and a condenser in series. X_L rises in direct proportion to f; X_C rises inversely as f. The sum, $X_L - X_C$, changes from a high negative value where X_C is greater than X_L, through zero where $X_L = X_C$, to a high positive value where X_L is greater than X_C.

reactance at the same frequency. Note that the total reactance curve passes through zero at 1000 kc, the resonant frequency of the circuit. The final curve shows very clearly how the reactance varies with frequency, and that the net reactance is capacitive below resonance and is inductive above resonance.

Reactance diagrams do not give completely accurate information about the total impedance of a circuit. However, if the resistance of a circuit is low, the reactance and impedance are almost equal, so that not much accuracy is lost if the reactance is taken as the total impedance. This approximation does not, of course, hold at frequencies very close to resonance, where the resistance of the circuit, even though small, is the controlling factor.

8.4. Effect of resistance. It has already been mentioned that resistance plays an important role in determining the cur-

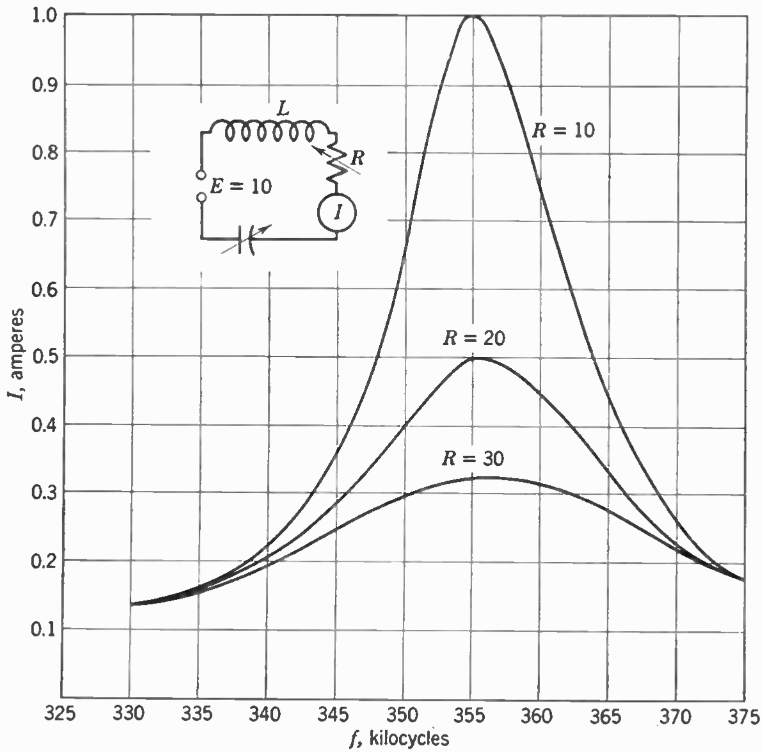


FIG. 10. Effect of resistance on a resonance curve. Note that the current far from resonance is not changed so much as the resonant current.

rent which flows in a series-resonant circuit near the resonant frequency. The curves of Fig. 10 show the effect of adding resistance to the circuit of Fig. 5. The voltages across the condenser and the inductor depend upon the resistance of the circuit, particularly in the vicinity of resonance. The current, controlled entirely by the resistance at resonance, in turn produces greater voltages across the reactances as the resistance is made smaller, the total applied voltage remaining constant. If E is the voltage impressed across the entire circuit, the voltage across the condenser is $E \div (\omega CR)$ and that across the inductor is $E \times (\omega L/R)$. The voltage across either L or C is equal to EQ_0, where $Q_0 = \omega_0 L/R = 1/\omega_0 CR$, and ω_0 is the value at resonance.

8.5. Power into a series-resonant circuit. No power is dissipated as heat, or in other ways, in a pure inductance or capacitance, but energy stored during one half cycle in a magnetic or electrostatic field is returned to the generator in the next half cycle. Power is expended only in the resistance of the circuit. The power is equal, as usual, to

$$P = I^2 \times R \text{ watts}$$

where R is the total resistance of the circuit. In Fig. 10, when R is 10 ohms, the current at resonance is 1 amp, and the power is 10 watts.

In other words, all the real power taken from the generator is used in heating up the resistance. No power is required to maintain the magnetic and electrostatic fields of the coil and condenser.

. In actual circuits all the resistance is not isolated as in the demonstration problems. All coils have some resistance; so do all condensers, although the resistances of modern condensers, other than the electrolytic type, are quite small. These resistances take power from the generator and reduce the maximum height of the resonance curve.

8.6. Resonant frequency. The condition for series resonance—that the reactances of the circuit add up to zero—is fulfilled when

$$X_L - X_C = 0 \quad \text{or} \quad X_L = X_C$$

$$\omega L = \frac{1}{\omega C} \quad \text{or} \quad \omega^2 = \frac{1}{LC}$$

and, since $\omega = 2\pi f$,

$$(2\pi f)^2 = 4\pi^2 f^2 = \frac{1}{LC}$$

$$f^2 = \frac{1}{4\pi^2 LC}$$

and so we arrive at the familiar expression for the **resonant frequency** of a circuit as

$$f = \frac{1}{2\pi\sqrt{LC}}$$

in which f = frequency in cycles.
L = inductance in henries.
C = capacitance in farads.
π = the Greek letter pi and is equal to 3.1416 ⋯ or approximately $3\frac{1}{7}$.

Example 2. An inductance of 0.25 henry and a capacitance of 0.001 μf are connected in series. At what frequency will series resonance occur? *Solution:*

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{0.25 \times 0.001 \times 10^{-6}}}$$

$$= \frac{1}{6.28\sqrt{25 \times 10^{-2} \times 1 \times 10^{-3} \times 10^{-6}}}$$

$$= \frac{1}{6.28\sqrt{25 \times 10^{-11}}} = \frac{1}{6.28\sqrt{2.5 \times 10^{-12}}}$$

$$= \frac{1}{6.28 \times 1.581 \times 10^{-6}}$$

$$= 10.1 \times 10^{3} \text{ cycles}$$

$$= 10.1 \text{ kc}$$

The expression for the resonant frequency shows that the frequency depends upon the product of L and C, and not upon either of them alone. If L is doubled, C can be halved and the resonant frequency will not be changed.

Problem 1. A 500-mh coil has a resistance of 150 ohms. What series condenser will cause the combination to resonate at 500 cycles? If 75 volts is placed across the combination, what current flows at resonance? What power is dissipated in the circuit?

Problem 2. What happens to the voltage across a condenser of a series-resonant circuit if the capacitance is reduced by one-half, resonance being maintained by other means which also keep the original current constant?

Problem 3. A variable condenser has a range of 9 to 1 from maximum to minimum capacitance, for example, from 270 $\mu\mu$f to 30 $\mu\mu$f. What is the ratio of the highest to lowest frequency it will tune a given coil to resonance?

Problem 4. A coil and condenser are in series across a 110-volt, 60-cycle circuit. Series resonance occurs when the condenser has 4 μf of capacitance. The voltage across the condenser is measured and found to be 300 volts.

What are the resistance and inductance of the coil? What is the current at resonance? What is the Q_0 of the circuit?

Problem 5. A series circuit contains a 10-mh coil and a 0.001-μf condenser. Plot reactance curves of the inductive reactance, the capacitive reactance, and the total reactance.

Problem 6. A 0.1-henry coil, a 10-μf condenser, and a resistance of 25 ohms are in series. The applied voltage is 100 volts. What is the series-resonant frequency? What voltage is across the condenser at the resonant frequency? at 300 cycles?

Problem 7. A loud speaker may be coupled to a power tube through a condenser as in Fig. 11. If the speaker has an inductance of 0.5 henry and

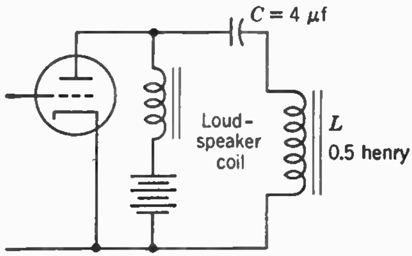


Fig. 11. An example of a resonant circuit.

the condenser a capacitance of 4 μf, to what frequency will the combination become resonant?

Problem 8. What are the voltages and phase relations in a series circuit like that of Fig. 5 at a frequency of 360 kc if $I = 0.2$ amp, $X_C = 500$ ohms, $X_L = 400$ ohms, and $R = 8$ ohms? See Example 1.

Problem 9. Work Prob. 8 if $R = 80$ ohms.

Problem 10. A certain antenna may be represented by 100 μh in series with 400 $\mu\mu$f and a resistance of 40 ohms. What will be the current if a distant station transmitting on 600 kc produces 1000 μv across the ends of the antenna? How large must the coil be to produce resonance at this frequency? What will be the current when resonance is established?

Problem 11. Plot a resonance curve similar to that of Fig. 10 if $L = 500$ μh, $C = 0.002$ μf, $R = 10$ ohms, and $E = 25$ volts. What is the power input to the circuit at resonance? at 140 kc?

8.7. *Q* factor.

The Q of a circuit has been mentioned already in this chapter. It is so widely used in design work that a more extensive discussion is merited. The term may be applied to a single element, such as a coil, with its own associated resistance, or it may be applied to a circuit made up of several impedance elements.

In a simple series circuit consisting of R and L, or R and C, Q

is merely the reactance at a given frequency divided by the resistance. The resistance may be the element's own resistance or may include an added series resistance. Since reactance varies with frequency, Q also varies with frequency.

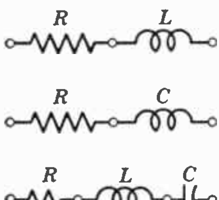
$$Q = \frac{\omega L}{R} \qquad \text{(inductive circuit)}$$

$$Q = \frac{1}{\omega CR} \qquad \text{(capacitive circuit)}$$

If a series circuit contains both capacitive and inductive reactance along with resistance, the Q of the entire circuit is generally defined only at the resonant values of the reactances. Then

$$Q_0 = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR} \quad \left(\omega_0 L = \frac{1}{\omega_0 C} \text{ at resonance}\right)$$

The subscripts on Q_0 and ω_0 denote that the values are taken at the resonant condition. One use of Q_0 was described in Sect. 8.4. Figure 12 shows three series circuits and the formulas which apply to each.



$$Q = \frac{2\pi fL}{R} = \frac{\omega L}{R} = \frac{X_L}{R}$$

$$Q = \frac{1}{2\pi fCR} = \frac{1}{\omega CR} = \frac{X_C}{R}$$

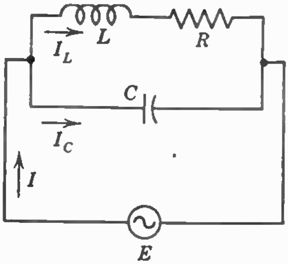
$$Q_0 = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

$$\omega_0 = 2\pi f_r$$

f_r = resonant frequency

FIG. 12. The Q factor of typical series circuits.

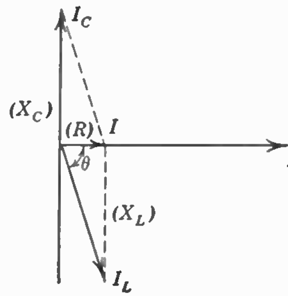
The most common parallel circuit met with in radio practice is the one shown in Fig. 13. Here, as in the case of the series-resonant circuit, Q is usually defined only for the resonant condition and $Q_0 = \omega_{ar}L/R$, where $\omega_{ar} = 2\pi f_{ar}$ and f_{ar} is the parallel-resonant frequency. Since X_C may be slightly different from X_L at resonance in a parallel circuit, Q_0 is only approximately equal to $1/\omega_{ar}CR$. Note that in the diagram no resistance has been shown in the capacitive branch, because the kind of condensers normally used in this type of circuit has very little resistance and it can generally be neglected.



$$Q_0 = \frac{2\pi f_{ar} L}{R} = \frac{\omega_{ar} L}{R}$$

$$I_L = \frac{E}{X_L} = \frac{E}{\omega L} = \frac{E}{2\pi f L}$$

$$I_C = \frac{E}{X_C} = E\omega C = E \times 2\pi f C$$



(*a*) A Typical Parallel-Resonant Circuit (*b*) Vector Diagram of Currents at Resonance

Fig. 13. A circuit in which parallel resonance (anti-resonance) may occur and a vector diagram of currents and voltage. The vector diagram also serves as an impedance diagram.

8.8. Parallel resonance. Many of the circuits used in radio involve resonance in a branched or parallel circuit. Figure 13 shows a typical parallel circuit composed of an inductance shunted by a condenser, the combination forming an **anti-resonant** circuit. The term "anti-resonant" is used to distinguish the phenomenon from that occurring in a series circuit. Often, however, both are referred to simply as "resonant circuits."

The effects of varying the frequency of the voltage across the circuit are widely different from the effects in a series circuit. In a series circuit, the current becomes very large at resonance and the resultant series impedance of the circuit becomes small. In the parallel circuit the equivalent impedance at resonance is high and the total current is small. In the series circuit the same current flows through the coil and condenser and the voltages across these elements may differ. In the parallel circuit the same voltage is across both branches, but the currents in them differ.

Consider first the case in which *neither the coil nor the con-*

denser contains appreciable resistance. The voltage across the branches is equal to the applied voltage. The current taken by each branch is the ratio between the voltage and the reactance of that branch. Thus

$$I_L = \frac{E}{X_L} = \frac{E}{\omega L}$$

$$I_C = \frac{E}{X_C} = E\omega C$$

$$I = I_L - I_C = E\left(\frac{1}{\omega L} - \omega C\right) \quad \text{(circuit resistance neglected)}$$

As the frequency is increased, more and more current is taken by the capacitive branch, less and less by the inductive branch. In a parallel circuit it is the currents (not the voltages as in the series circuit) which are out of phase; at any frequency their algebraic sum combined *vectorially* with the shunt resistance current (if any) gives the total current taken from the generator. In the simple example discussed here, the resistance is neglected and the algebraic sum of the two currents gives the total current. Since these two currents are 180° out of phase, adding them actually means subtracting I_C from I_L.

At resonance the currents taken by the two branches are equal and the current taken from the generator is zero, since the branch currents algebraically add to zero. If there is resistance in the circuit, the generator current will be small but not zero. The larger the resistance in one of the branches, the larger the total current will be at resonance.

The impedance of the circuit as a whole, that is, the impedance into which the generator must feed current, is the ratio between the voltage and current:

$$Z = E \div I$$

Therefore, if no current flows, the circuit has infinite impedance. Actually there is always some resistance in the circuit, and the resistance may be in series with one or both branches or in shunt with the network. The effect of the resistance is to cause some current to flow at resonance and the impedance does not go to infinity, although it may become very large.

Figure 13 shows a parallel circuit in which the coil resistance is not neglected, and a vector diagram of the currents and applied voltage. The voltage E is taken as the reference and drawn as a horizontal line. The current through the condenser I_C leads the voltage by 90° and so is drawn vertical and upward. The current through the inductive branch lags the voltage, but not by 90° because of the resistance in this branch. The vector sum of I_L and I_C is the total current I. Note carefully that, whereas the total current into a parallel-resonant circuit is very low, the currents through the individual branches may be quite high. If R is small, $I_L = I_C = IQ_0$, approximately, where Q_0 is the Q of the inductive branch at the resonant frequency. Thus, if Q_0 of the coil is 150, as it is for many r-f coils, each of the branch currents will be approximately 150 times the total current at the resonant frequency. These branch currents are often called the **circulating currents.**

The reasons for the above characteristics of a parallel circuit at resonance follow. The impedance offered to an external source of voltage at resonance is quite high as has already been discussed. Thus, the net current taken by a parallel circuit at resonance is small. On the other hand, the impedance of the *individual* branches may be much smaller than the total impedance which results when the branches are considered together. The currents through these individual branches are almost 180° out of phase, the capacitive current leading and the inductive current lagging the voltage across the parallel circuit by about 90°. These two currents must be quite large if their *vector* sum is to equal the total current to the network.

It can be shown that (Sect. 8.9)

$$I = \frac{I_L}{Q_0} \quad \text{(approx.)}$$

or

$$I_L = Q_0 I \quad \text{(approx.)}$$

where I is the total current to the parallel circuit, I_L is the current in the inductive branch, and Q_0 is the value at the resonant frequency. The above equation holds when the resistances of the inductive and capacitive branches are low. However, no matter how large these resistances, the branch currents are somewhat higher than the total current, even though the simple expression above no longer gives the correct value of current. Under the condition that the resistances are low, the currents taken by the inductive and capacitive branches are nearly equal in magnitude.

For example, if the coil in a parallel-resonant circuit (Fig. 13) has a Q_0 at resonance of 100, and the net current flowing to the circuit is 1 ma, the currents in each of the branches will be very nearly equal to 100 ma ($IQ_0 = 1 \times 100 = 100$ ma). Suppose that the voltage E across the circuit is 10 volts. Then the total impedance of the circuit is $10 \div (1 \times 10^{-3}) = 10{,}000$ ohms. The impedance of each branch, on the other hand, is $10 \div (100 \times 10^{-3}) = 100$ ohms, or 100 times less than the total impedance. This illustrates how condensers and coils in parallel may be used to produce an effective impedance many times larger than their own individual impedances.

Experiment 2. Connect as in Fig. 13 the coil and condenser used in Experiment 1. If sufficient r-f ammeters are available, read the current in the two branches as well as the current from the generator as the frequency is changed. Then fix the generator frequency and adjust the capacitance until minimum current flows. Plot the currents against frequency and against the capacitance. The generator in this experiment may be a small oscillating tube with 4 or 5 watts output. This amount of power is sufficient to produce in the branches currents of at least 100 ma.

In series circuits reactance diagrams were found to be useful in determining the manner in which the reactance varied with frequency. These diagrams could be used, even for several reactances in series, because the net reactance of a group of series reactances is their algebraic sum, and this sum may be computed using graphical methods. In parallel circuits the total reactance is not the algebraic sum of the reactances of the branches. Instead, the reactance is computed, *after* the resultant current has been determined, by dividing the applied voltage by this current. For this reason, it is more instructive to draw plots of branch currents against frequency in parallel circuits.

The manner in which the branch currents, and the total current, vary with frequency in a circuit such as that of Fig. 13 is shown in Fig. 14. The current through the inductive branch is high at low frequencies and low at high frequencies; that in the capacitive branch is low at low frequencies and high at high frequencies. Since the currents are 180° out of phase (resistance neglected), the inductive current is drawn as a positive current and the capacitive current as a negative current. The result of adding the two currents algebraically is the total current. At low frequencies the total current is determined largely by the amount of current in the inductive branch; at high frequencies it is almost equal to the current in the capacitive branch. In the middle range of frequencies the total current is much less in magnitude

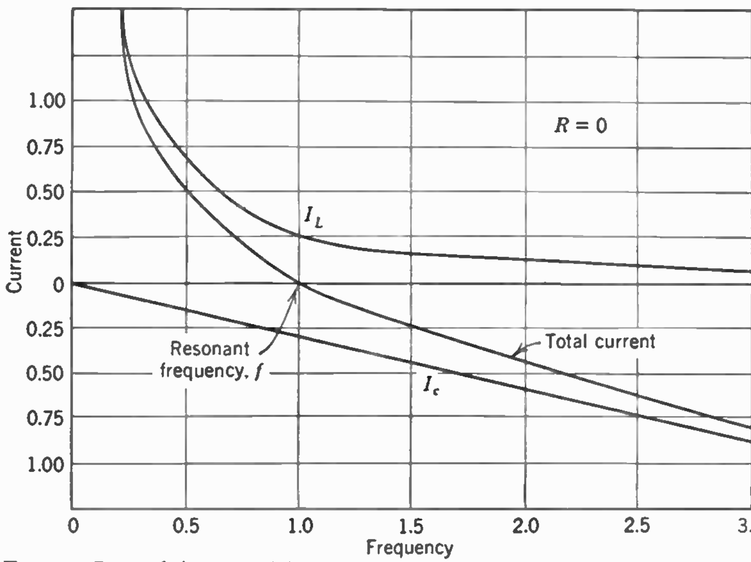


Fig. 14. In studying parallel circuits, it is simpler to plot the current taken by the reactive branches than to plot the individual reactances as is done in series circuits, for example in Fig. 9.

than either branch current, and at the resonant frequency the total current is zero. As with reactance diagrams for series-resonant circuits, this current plot is not highly accurate in the region near the resonant frequency, 1.0 in the diagram, but is very accurate at frequencies above and below the resonant frequency unless the coil resistance is unusually high.

8.9. Resonance conditions in parallel circuits. The phenomena occurring in a parallel circuit near the resonant frequency are rather complex. As a result there are several criteria which could be used to denote the condition of anti-resonance. Three of these are:

1. Inductive and capacitive reactances are equal.

2. Minimum current flows from the line. This is the condition for **maximum impedance.**

3. Parallel circuit acts like a pure resistance. This is the **unity power factor** condition.

The frequencies, or reactances, at which these three conditions exist but are slightly different provided that the resistances in the coil and condenser are small, as is the case in practical radio systems. It is common practice to denote condition 3 as the condition for parallel resonance where there is a need to differentiate among the possible conditions. Under any of these three conditions, in a practical circuit, the resultant impedance of a parallel circuit at resonance is approximately a pure resistance.

The development of the exact equations for the impedance of a parallel circuit containing resistance is beyond the scope of this book.

Approximate relations, accurate enough for most radio problems, are as follows [the vector diagram is shown in Fig. 13(*b*)]. The total current taken by the circuit is

$$I = \frac{I_L R}{X_L} = \frac{I_L}{Q} \quad \text{(approx.)}$$

but

$$I_L = \frac{E}{Z_L} \quad \text{if } X_L \text{ is much greater than } R$$

then

$$I = \frac{E}{QX_L}$$

but the impedance is $E \div I$ or

$$Z = QX_L = \frac{X_C X_L}{R}$$

Because R is small, X_C and X_L are nearly equal, so

$$R_{\text{ar}} = Z = \frac{L}{CR} = \frac{\omega_{\text{ar}}^2 L^2}{R} \quad \text{(approx.)}$$

where R_{ar} is the impedance of the circuit at resonance; it is a pure resistance.

The resonant frequency is the same as for a series circuit provided R is low:

$$f_{ar} = \frac{1}{2\pi\sqrt{LC}}$$

Example 3. The following are given for the circuit of Fig. 13:

$$L = 200\ \mu\text{h}$$

$$R = 10 \text{ ohms}$$

$$f_{ar} = 356{,}000 \text{ cycles}$$

$$E = 10 \text{ volts}$$

Compute the size of the condenser, the line current, the circulating current, and the total impedance. Use the approximate equations developed above.

Solution:

$$f_{ar}^2 = \frac{1}{(6.28)^2 LC}$$

$$C = \frac{1}{(6.28)^2 L \times f_{ar}^2}$$

$$= \frac{1}{(6.28)^2 \times 200 \times 10^{-6} \times 356{,}000^2} = 0.001\ \mu\text{f}$$

$$X_L = 2\pi \times 356{,}000 \times 200 \times 10^{-6} = 447 \text{ ohms}$$

$$Q = \frac{X_L}{R} = \frac{447}{10} = 44.7$$

$$I_L = \frac{E}{X_L} = \frac{10}{447} = 0.0224 \text{ amp or } 22.4 \text{ ma}$$

This is the circulating current. The line current is $I_L \div Q$ or

$$I = \frac{22.4}{44.7} = 0.5 \text{ ma}$$

The total impedance of the circuit is

$$R_{ar} = QX_L = 44.7 \times 447 = 19{,}880 \text{ ohms}$$

The impedance can also be calculated by

$$R_{ar} = \frac{L}{CR} = \frac{200 \times 10^{-6}}{0.001 \times 10^{-6} \times 10} = 20{,}000 \text{ ohms}$$

The slight difference in the computed values of R_{ar} arises from the approximate equations used in the computations.

A parallel-resonant circuit, when used in radio transmitters and receivers, is often called a **tank circuit.**

8.10. Power relations in a parallel-resonant circuit. The impedance of an anti-resonant circuit is a pure resistance at the resonant frequency. Thus, the generator voltage and current are in phase and the power taken from the generator is equal to the square of the current from the generator times the impedance of the circuit, or

$$P = I^2 R_{ar} = I^2 \frac{L}{CR}$$

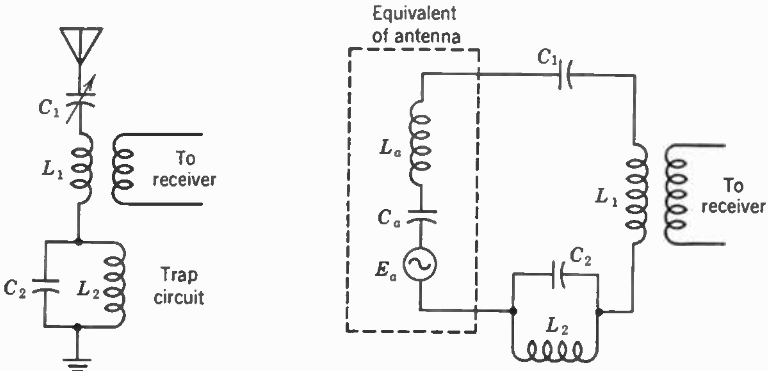
where P = power from generator.
I = current taken from generator.
L = inductance of the anti-resonant circuit.
C = capacitance of the anti-resonant circuit.
R = resistance of the anti-resonant circuit.
R_{ar} = effective resistance of the anti-resonant circuit at resonance.

If the inductance of a tank circuit is coupled to a load, such as an antenna, through a secondary coil, then the load will draw some power from the tank. This power must be supplied by the generator since the tank circuit itself supplies no power—the power is merely transferred through the tank from the generator to the load. When the antenna takes power, the coil in the tank circuit acts as though its resistance had been increased. For example, suppose that 5 watts is taken by the tank from the generator and that 1 amp flows to the tank circuit. Now couple the coil to an antenna which may take 5 watts. A total of 10 watts now flows from the generator, and so far as the generator is concerned it behaves as though the resistance of the tank had been doubled since twice the power is now taken from it. This assumes, of course, that the circulating current remains constant, which may be made approximately true in such a circuit by proper design.

The value of R, the tank circuit resistance, must therefore represent not only the series resistance of the coil, the leads, and the condenser, but also the resistance "reflected" into the tank by the load which it feeds.

8.11. Tuned circuit applications. Series- and parallel-resonant circuits are very widely used in radio apparatus. Either circuit may be made highly frequency selective; both may produce either high currents or high voltages at the resonant frequency. The important distinction between the two is that a series circuit has its minimum impedance at resonance; a parallel circuit has a high impedance at resonance.

A circuit used as the input circuit for a radio receiver as shown in Fig. 15(*a*) will serve as an example of both series and parallel



(*a*) Actual Circuit (*b*) Equivalent

Fig. 15. The anti-resonant circuit in series with the antenna rejects undesired signals by making the series impedance to these signals very high.

resonance. A simple equivalent circuit for the antenna itself is shown in Fig. 15(*b*). In operation, it is desired to build up the maximum possible voltage at a particular frequency across L_1, and at the same time discriminate against signals of other frequencies which may be picked up on the antenna. The condenser C_1 is used to tune the *entire* circuit to *series* resonance at the desired frequency so that this maximum voltage does appear across L_1.

Suppose that the desired signal is arriving from a distant station and is rather weak and that a strong local station with a frequency near that of the distant station is also being picked up on the antenna. In such a case the parallel-resonant "trap" circuit can be tuned to the frequency of the strong but undesired signal so that, at this frequency, the current through the whole circuit is small, and the voltage across L_1 is likewise small. Thus, the combination of a series- and a parallel-resonant circuit serves to eliminate the undesired signal and to build up the strength of the weak signal for good reception. The two coils L_1 and L_2 are physically separated; there is no mutual coupling between them.

The anti-resonant circuit in the above illustration is called a **rejector** circuit because it rejects signals of the frequency to which it is tuned. It is also known as a **wave trap.** The series-resonant circuit is called an **acceptor** because it accepts signals at the resonant frequency.

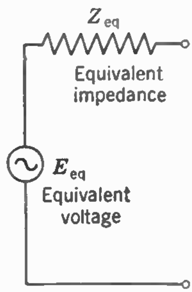


FIG. 16. The equivalent circuit of a voltage source is a constant - voltage generator and an equivalent impedance in series.

In a large group of radio applications the resonant circuit is connected across the terminals of the source, which may be the output of a vacuum tube, the secondary of an antenna coil, or a variety of other devices. For convenience, this source may be considered a generator (because it produces voltage across the load circuit) in series with an impedance (because it does possess some impedance). Such an equivalent circuit is shown in Fig. 16. Heretofore, the examples relating to resonant circuits have assumed a constant voltage across the terminals of the circuit. The impedance of the **equivalent generator** must be taken into account in applying resonant circuits to specific jobs.

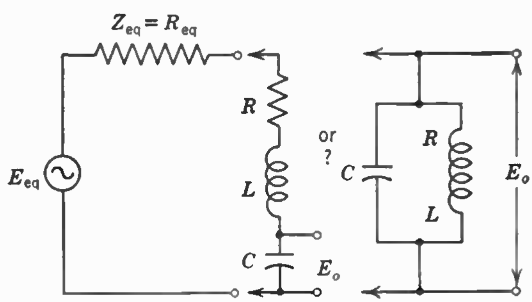


FIG. 17. A series-resonant circuit is used with a generator with a low equivalent impedance; an anti-resonant circuit is used with a generator with a high equivalent impedance.

Consider Fig. 17 and suppose that it is desired to use a circuit which will give a voltage output, E_0, which is high at the resonant frequency and falls off rapidly at frequencies above and below the resonant value. Which circuit should be used, the series- or parallel-resonant combination? Actually the question cannot be answered until the equivalent impedance of the generator is specified.

In a series-resonant circuit the sharpness of the resonance curve is determined largely by the series resistance. If the resistance is low the curve rises rapidly, but if the resistance is high there is little rise of current at the resonant frequency. The equivalent impedance of the generator, a resistance in this example, must be added to the resistance of the series circuit. If the generator and circuit resistance are both low, a series-resonant circuit will produce a sharp rise in current at resonance, and therefore a large voltage, E_0, across C.

The impedance of a parallel circuit is highest at resonance and falls off rapidly as the frequency is changed to values either above or below this value. Suppose the equivalent generator has zero internal impedance. Then the voltage across the parallel circuit must remain constant at any frequency. However, suppose that the generator impedance is resistive and equal to the impedance of the parallel circuit at resonance, that is, $Z_{eq} = R_{ar}$. Now, as the frequency moves away from the resonant value, the impedance of the resonant circuit drops, the current from the generator rises, and more of the total voltage appears across Z_{eq} and less across the impedance of the parallel circuit, and the circuit can have a sharp resonance curve. The sharper resonance curve is accompanied by a lower voltage at resonance. This is not usually serious. Thus, a generator with a high internal impedance requires an anti-resonance circuit in order to obtain a frequency-selective device. The effective resistance of most tubes is quite high, and so parallel-resonant circuits, not series circuits, are most commonly used as frequency-selective loads for tubes.

The general shapes of the curves of series and parallel circuits with varying generator resistances are shown in Fig. 18.

Problem 12. A parallel tuned circuit is required which has 100,000 ohms total resistance at the anti-resonant frequency of 500 kc. The condenser is to have a capacitance of 200 $\mu\mu$f. What must be the inductance and resistance of the parallel coil?

Problem 13. A pentode tube gives the greatest voltage amplification when worked into a high impedance load. A 750-$\mu\mu$f condenser is at hand. What size inductance is required to provide a parallel-resonant circuit as a load at 400 kc? If the resistance of the coil is 20 ohms, what is the effective resistance of the parallel circuit at resonance? This is the load resistance for the tube.

Problem 14. A wave trap is to be placed in series with an antenna (Fig. 15) to reduce interference from a station transmitting on a frequency of 1 Mc. A 200-μh, 10-ohm coil is available. What size condenser should be shunted across the coil? If the resistance of the condenser is negligible, what impedance will the wave trap offer to the offending signal?

Problem 15. What is the Q_0 of the coil in Prob. 14?

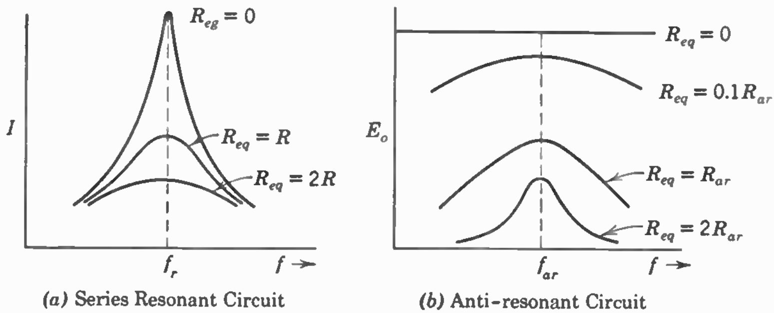


FIG. 18. The resonance curve of a series circuit is sharp for low values of the equivalent impedance of the voltage source; the curve for an anti-resonant circuit is sharper if the equivalent impedance of the source is high.

8.12. Sharpness of resonance. The effect of resistance is to reduce the maximum current flowing in a series-resonant circuit, to make less pronounced the minimum of current flowing into a parallel-resonant circuit from an external source, and to decrease the impedance (L/CR) of the parallel circuit.

Since the maximum current is desired in a series circuit, and the maximum impedance in a parallel one, the inclusion of any more resistance than necessary is deleterious.

As an example, consider the antenna illustrated in Fig. 15. The wave trap will be neglected for the present. Suppose that the total inductance $(L + L_a)$ is 200 μh, and the total capacitance (C and C_a in series) is 1000 $\mu\mu$f at a resonant frequency of 356 kc. Assume a voltage E of 10 mv. What is the effect on this system if it has a resistance of 2.5 ohms and of 10 ohms?

The current at resonance (356 kc) in the 2.5-ohm case is 4 ma, whereas at 370 kc the current is down to 0.284 ma, a ratio of 14 to 1. In the 10-ohm case, the resonant current would be 1 ma—one-fourth of its value with the lower resistance—and the current at 370 kc, that is, 14 kc off resonance, would be 0.274 ma. The ratio of current at resonance to current at 370 kc in this case is 3.65 to 1.

If, then, the antenna has impressed on it from equally distant and equally powerful radio stations two voltages—one of 356 kc, the desired frequency, and one of 370 kc, the unwanted frequency —14.1 times as much current flows at the desired frequency as at the unwanted. In the 10-ohm case, however, not only is the desired current cut to one-fourth of its other value but the ratio of wanted to unwanted current has been decreased to 3.65. The low-resistance antenna is said to be more selective, and its selectivity is decreased when resistance is added to it. The sharpness of resonance, or the selectivity, is proportional to the Q of the circuit. Low-resistance, high-Q circuits are highly selective.

8.13. Width of the series-resonance curve. A series-resonance curve such as in Fig. 19 cannot truly be said to have width

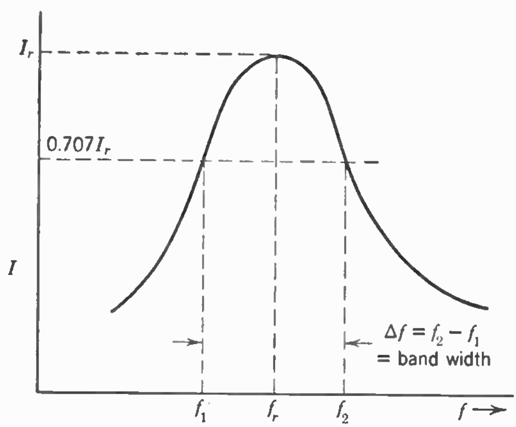


Fig. 19. The band width of a series-resonant curve is the number of cycles between points on the curve at which the current is equal to 0.707 times the maximum value, I_r.

because the sides are sloping. Nevertheless it is very desirable to have a means of comparing one resonance curve with another. Therefore, the "width" of a resonance curve is defined as *the total change in frequency about the resonant frequency which is required to reduce the current to 0.707 times its resonant value.*

The power into the resistance R, at these frequencies, is $0.707^2 I_r^2 R = \frac{1}{2} I_r^2 R$, and these frequencies on the resonance curve are also called the **half-power points.**

The total reactance in the circuit at any frequency is $X = (X_L - X_C)$. The current that flows is

$$I = \frac{E}{\sqrt{R^2 + X^2}}$$

At resonance

$$I = I_r = \frac{E}{R}$$

If the frequency is increased above the resonant value f_r to a new frequency f_2 such that $X = (X_L - X_C) = R$, then

$$I = \frac{E}{\sqrt{R^2 + X^2}} = \frac{E}{\sqrt{R^2 + R^2}} = \frac{E}{\sqrt{2R^2}}$$

$$= \frac{E}{\sqrt{2} \times R} = \frac{1}{\sqrt{2}} \times \frac{E}{R} = 0.707 I_r$$

and

$$Z = \frac{E}{I} = \sqrt{2}R$$

The same result is secured if the frequency is reduced the proper amount to f_1. The difference in the two frequencies, which is the band width of the resonant curve, is given by

$$\Delta f = f_2 - f_1 = \frac{R}{2\pi L}$$

If the terms of the equation are all divided by f_r there results

$$\frac{\Delta f}{f_r} = \frac{f_2 - f_1}{f_r} = \frac{R}{2\pi f_r L} = \frac{1}{Q_0}$$

The last equation says that the *percentage*, or *ratio* of frequency discrimination, is inversely proportional to the Q_0 of the circuit; that is, the higher the Q_0 the greater the selectivity, that is, the narrower the band width, on a percentage basis.

Example 4. What will be the half-power band width of a series-resonance curve if $L = 200$ μh, $R = 10$ ohms, and $f = 356{,}000$ cycles?

$$\Delta f = f_2 - f_1 = \frac{R}{2\pi L} = 7960 \text{ cycles}$$

Problem 16. What is the band width of a series circuit resonant at 356 kc if the coil has 200 μh inductance and 30 ohms resistance? What is the size of the series condenser? What are Q_0 and the percentage frequency discrimination of the circuit?

Problem 17. Repeat Prob. 16 if the resistance is changed to 6 ohms. Compare results in Probs. 16 and 17.

Problem 18. A circuit is to pass 0.707 of its maximum current at a point 3 kc off resonance, which occurs at 1.5 Mc. The condenser is to have a capacitance of 400 $\mu\mu$f. Calculate the amount of resistance which must be in the circuit.

Problem 19. If the expression $Q = \omega L/R$ of a coil remains constant over a fairly wide band of frequencies, that is, the effective or r-f resistance increases directly with frequency, does the selectivity of a tuned circuit utilizing such a coil differ at different frequencies? Do the band widths at 500 kc and 1500 kc differ? If so, how?

Problem 20. Suppose that increasing the size of an inductance by a factor of 4 increases the resistance by a factor of 2. The circuit is to tune to the same wavelength. What is the ratio of the selectivities?

8.14. Ratio of L/C. Since the selectivity of a resonant circuit is proportional to Q_0, and Q_0 is equal to the inductive (or capacitive) reactance at the resonant frequency divided by the circuit resistance, it is apparent that increasing L and decreasing C to maintain the same resonant frequency will increase the selectivity. This is shown below.

$$Q_0 = \frac{X_L}{R} = \frac{X_C}{R} = \frac{\omega L}{R} = \frac{1}{\omega CR} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

If, however, the coil resistance increases as fast as the inductance increases, or faster, then the benefit of using high L and low C will not be secured.

8.15. Skin and proximity effect. The d-c resistance of a coil may be only a fraction of an ohm. Yet the effective resistance * of that coil, when placed in a high-frequency circuit, will appear to have a value many times the d-c resistance. This increase of resistance is mainly due to skin effect, although some of the increase may be accounted for by the proximity effect.

* This refers to the apparent *resistance* of the coil. The *impedance* of the coil is made up of this resistance and the inductive reactance ($X_L = 2\pi fL$).

The current in a wire, whether wound into a coil or not, changes rapidly in both strength and direction when placed in a r-f circuit. Small counter-voltages are induced within the wire itself according to Lenz's law. The magnetic lines originate at the center of the wire and expand outward, being more concentrated at the center than near the surface of the conductor. The greater induced voltages near the center oppose the flow of current more than is the case near the surface. The result is that the current is crowded toward the surface (skin) of the wire [Fig. 20(*b*)]

(*a*) D-C Current is Uniformly Distributed over Cross Section of Conductor

(*b*) High-Frequency Currents Crowd to Surface Because of Skin Effect

(*c*) Current Distribution is Distorted When Two Conductors Are in Proximity

FIG. 20. Skin effect causes currents to crowd toward surface of conductor; proximity effect causes distorted current distribution.

and the apparent resistance of the wire is increased. This effect is known as **skin effect,** and explains why the r-f resistance of a wire or coil is higher than the d-c resistance. The r-f resistance increases approximately as the square root of frequency.

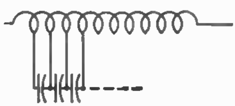
Over the normal tuning range in which a coil is used, the r-f resistance does not vary greatly (perhaps 2 to 1), but the effect is important, nevertheless. It should be noted that skin effect does not depend on the placement of the wire; it may be strung out in a straight line, or it may be wound in a coil.

The skin effect is so pronounced at very high frequencies that all the current effectively flows almost on the surface of the wire. High-frequency coils are often made of hollow copper tubing, with little gain in resistance over that of a solid conductor, so that cooling water may be forced through the center. Radar apparatus, operating at frequencies of 3000 Mc and above, is often silver-plated to provide a low-resistance path for the current. The base metal has practically no effect on the resistance.

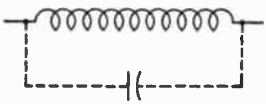
A second effect which depends upon the closeness (proximity) of current-carrying wires also acts to increase the resistance.

Currents flowing in two or more adjacent wires set up magnetic fields which distort the distribution of current in the wires [Fig. 20(*c*)] and cause crowding that results in an increase of the effective resistance. This effect is known as the **proximity effect.**

8.16. Distributed capacitance of coils. Whenever two objects which conduct current are insulated from each other and are at different voltages, electrostatic charges can be stored in them; they constitute a condenser. The capacitance of this condenser depends upon the closeness of the objects, the dielectric



(*a*) Distributed Capacitance of a Coil



(*b*) Equivalent Circuit

FIG. 21. "Distributed" capacitance exists between each pair of turns of a coil. This capacitance limits the usefulness of coils at high frequencies.

constant of the insulation between them, and their shape. In a coil of wire each turn is insulated from the adjacent turns and a difference of potential exists between two turns. Therefore a coil is not a pure inductance but may be thought of as a coil shunted by a capacitance made up of the sum of many small turn-to-turn capacitances. These small capacitances and the equivalent circuit of the coil are shown in Fig. 21. The equivalent circuit is composed of a coil and condenser in parallel, and at some frequency the circuit will be anti-resonant. Above the anti-resonant frequency the circuit actually acts like a condenser rather than a coil. This inherent capacitance of a coil is known as the **distributed capacitance** since it is more or less evenly distributed along the whole length of the wire from which the coil is wound.

In most circuits a coil must be used at frequencies considerably lower than the frequency at which the coil goes into self-resonance. Therefore, to extend the frequency at which coils may be used, much effort has been expended toward developing special methods of winding to reduce the distributed capacitance. If the coil has only a few turns it is usually wound as a solenoid, perhaps with some spacing between turns to reduce the capacitance. To obtain large inductances in a small space and with a minimum of capacitance, it has become customary to make multilayer coils by the use of peculiar types of windings called bank windings and universal windings. A bank winding is shown in Fig. 22. Note that turn No. 3 is on top of the first two turns, and so on. This is to keep the potential difference between adjacent turns as low as possible. It would be much higher, for example, if turn No. 13 was adjacent to turn No. 1.

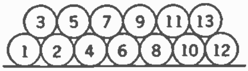


FIG. 22. Bank-winding type of coil.

8.17. Comparison of series and parallel tuned circuits. In a series tuned circuit the line current is high, the circuit impedance is low, the voltages across coil and condenser are high, the current through L and C is high, and the actual impedance at resonance is equal to R, the resistance of the circuit.

In a parallel tuned circuit the line current is low, the total circuit impedance is high, the currents through the coil and condenser are high, the voltage across the coil and condenser is equal to the impressed voltage, and the actual impedance at resonance is equal to L/CR.

Below is a tabular comparison of the currents, voltages, and impedances in series- and parallel-resonant circuits.

Circuit	Current Line	Current Branch	Total Impedance	E_X	Z
Series	High	....	Low	High	R
Parallel	Low	High	High	Line	L/CR

Problem 21. A 250-μh coil has a distributed capacitance of 20 $\mu\mu$f. What is the upper limit of frequency at which the coil still behaves like an inductor?

Problem 22. The r-f resistance of a certain coil varies directly as the square root of frequency; that is, if the frequency is increased four times, the resistance is increased two times ($\sqrt{4} = 2$). Suppose that a coil has an inductance of 159.2 μh, and at 1 Mc has a r-f resistance of 5 ohms. Plot a graph of Q against frequency up to 4 Mc. Repeat for a coil of the same inductance which has a resistance of 5 ohms at all frequencies.

9 · Properties of Coils and Condensers

Coils and condensers form the nucleus of almost every radio circuit. To gain an understanding of their role in the reception of radio messages, let us look at a simple receiving system.

9.1. Tuning a receiver. A simple receiving circuit consists of an antenna-to-ground system connected to a coil and detector, such as a crystal of galena or silicon or other sensitive mineral which has the property of separating the audio tones from a radio wave. A pair of headphones in series with the detector makes audible the audio-frequency signals which the detector extracts from the radio wave. A small condenser across the phones will by-pass the radio-frequency currents around the phones but will not by-pass the audio-frequency currents, which must go through the phones. (The condenser has low reactance to r-f currents but high reactance to a-f currents.)

The strength of the signal in the phones may be made greater by tuning the antenna-to-ground system to the frequency of the desired wave. (See Sect. 8.11.) This is done by varying C in Fig. 1. When the circuit is resonant, a large current flows through the inductance. The voltage across it ($X_L \times I$) will be large, and the response from the crystal will be greater than before.

The voltage across the inductance can be amplified first and then impressed across the detector circuit. This amplification may take place in several stages so that very weak signals may be received. If desired, the signals may be amplified again after detection by means of a-f amplifiers.

Tuning the antenna system also increases the selectivity of the system. Signals whose frequency is lower than the resonant frequency of the antenna meet with a high impedance in the tuning condenser; signals of high frequency meet with a high impedance in the coil; signals of the desired frequency find a minimum of impedance, and so the filtering action of the tuned circuit is advantageous.

If a second coil is loosely coupled to the antenna coil and then tuned to the desired frequency as in Fig. 2(*a*), an even greater degree of selectivity can be achieved. The primary circuit in series with the antenna works the same as before; it is tuned to series resonance at the desired frequency. The voltage induced in the secondary coil L_2 is equal to the impedance of the mutual inductance multiplied by the primary current ($\omega M \times I_1$). This induced voltage can be represented as an equivalent generator as in Fig. 2(*b*). Thus, the secondary circuit is a series-resonant circuit in which the output voltage is taken from across C_2. In such a series circuit the voltage across the condenser may be considerably greater than the driving voltage, which is the equivalent generator voltage E_2 in this case, if the Q_0 of the circuit is reasonably high. In addition to the increased voltage supplied to the detector, the second resonant circuit provides an

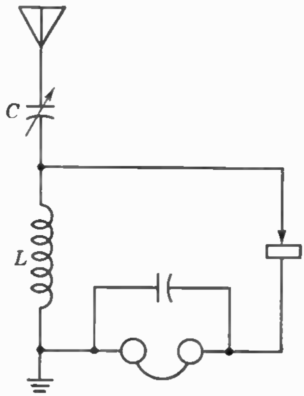
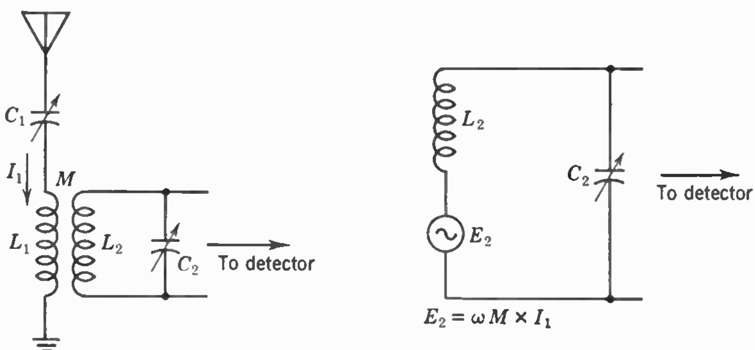


Fig. 1. A simple radio receiver.



(*a*) Actual Circuit

(*b*) Equivalent Secondary Circuit

Fig. 2. A resonant circuit coupled to the antenna coil supplies more voltage to the detector and gives greater selectivity.

increased amount of selectivity; the composite circuit provides rather effective discrimination against unwanted signals.

If, in addition, a r-f amplifier preceding the detector is tuned to the desired signal, the selectivity may be made even greater. Because of the present congestion of radio stations, a high degree of selectivity is a prerequisite to good reception.

9.2. The frequency meter. An instrument for determining the frequency of a signal is called a **frequency** (or **wave**) **meter.** Although several types of circuits are used for the determination of frequency, all except a very few utilize the resonant properties of coils and condensers in the frequency-measuring circuit. A discussion of the operation of a simple frequency meter will serve to illustrate many of the important properties of coils and condensers.

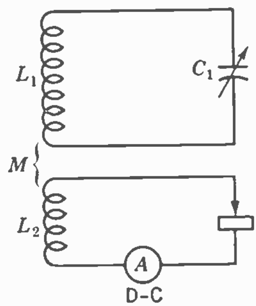


Fig. 3. A simple frequency meter.

The circuit of a simple and effective frequency meter is shown in Fig. 3. The resonant circuit, L_1C_1, is the heart of the apparatus. Only when this circuit is tuned to resonance with a signal whose frequency is to be determined will an appreciable current flow. Also, the higher the Q of this circuit, the sharper the response of the instrument. Current in this circuit is coupled to coil L_2, which is in series with a crystal detector and a d-c milliammeter or microammeter. The detector rectifies the a-c signal and thus furnishes direct current to the indicating meter. The detector could be omitted and the d-c meter replaced with an a-c meter. However, since d-c meters are much more sensitive than a-c meters, an a-c meter is seldom used. The coupling between the two coils, L_1 and L_2, is kept very low so that the detector circuit does not introduce (reflect) much resistance into the L_1C_1 circuit and broaden its response curve. The coil L_1 is generally fixed in value, and C_1 is varied to obtain resonance. Several coils are usually provided so that a wide range of frequencies may be measured. If the coils are made so that the larger coils have exactly four times the inductance of the next smaller, the wavelengths to which the larger coil will tune will be twice those of the next smaller coil or the frequencies will be one-half of those of the next smaller coil.

A wavemeter is used by holding it close to the source of the signal whose frequency is to be measured, such as, for example, that of a r-f oscillator. The meter condenser is then tuned to obtain maximum current indication on the d-c meter. The condenser dial, which is usually calibrated either in frequency or in wavelength, then indicates the frequency. It is important for an accurate determination of the frequency that the wavemeter be held no closer to the signal source than absolutely necessary to obtain an indication. Otherwise, the power taken by the meter may cause a shift of the frequency of the signal source.

A modification of the circuit in Fig. 3 omits the detector circuit and inserts a small pilot lamp in series with L_1C_1. As the condenser is tuned through resonance the lamp lights, and when it reaches its maximum brilliance the circuit is in resonance. The lamp has the effect of reducing the Q of the circuit and of broadening the response curve.

Still another modification of the circuit omits the detector circuit but does not alter L_1C_1. The meter then operates as a **reaction- (absorption-)** type frequency meter. It can be used only in cases in which the signal source incorporates a sensitive meter, such as a meter in the plate circuit of the source, which indicates when power is being taken. In use, a reaction frequency meter is held close to the output of the signal source and tuned until the meter in the signal source indicates that additional power is being required. What happens, of course, is that the frequency meter absorbs power only when tuned to resonance with the signal source. It is particularly important when using this type of meter to keep the coupling as loose as possible in order not to disturb the signal frequency any more than necessary. The frequency meter in Fig. 3 could be used in the same manner without alteration.

9.3. Grid-dip frequency meter. A sensitive meter in an oscillator circuit rises or dips sharply when any power is taken by an external circuit. Such an oscillator can be used as a wavemeter. Provisions are often made for the modulation of the oscillator with an a-f signal so that the oscillator output includes both a r-f signal and an a-f signal. Figure 4 shows the circuit of a relatively simple meter of this type.

The **grid-dip** meter (so called because the meter is in the grid circuit of the oscillator) can be used to determine the frequency to which an external coil and condenser resonate, for example, a coil-condenser combination intended for use in a radio circuit.

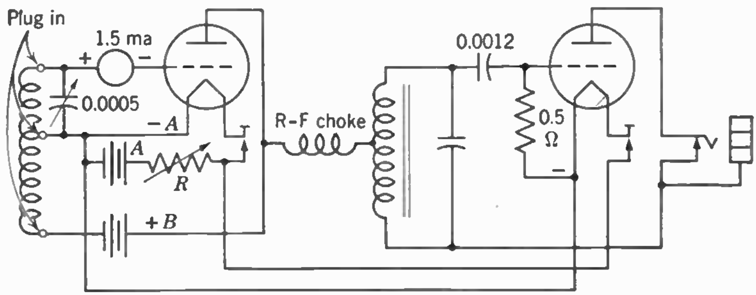


FIG. 4. Circuit diagram of a modulated oscillator useful as a simple grid-dip wavemeter. The r-f oscillator has a d-c meter in its grid circuit and employs a series of plug-in coils. The a-f modulating oscillator may use an audio output transformer or other center-tapped inductance of the correct value. When the r-f oscillator is coupled to an external tuned circuit, a sharp dip in the grid current will occur when the two circuits are tuned to the same frequency.

TABLE I

Coil	f, kc	λ, meters	Kilocycles per Dial Degree
A	2500–6660	45–120	31.6
B	1430–3750	80–210	23.3
C	750–1820	165–400	10.7
D	485–1130	265–620	6.5

Coil	Turns	Size Wire	Diameter, in.	Length of Winding	L, mh
A	15	21	$2\frac{11}{64}$	$1\frac{5}{8}$	0.014
B	30	21	$2\frac{11}{64}$	$1\frac{1}{2}$	0.055
C	60	21	$2\frac{11}{64}$	$1\frac{5}{8}$	0.217
E	90	27	$2\frac{11}{64}$	$1\frac{3}{8}$	0.495

The frequency meter is held close to the circuit whose resonant frequency is to be measured, and the frequency of the oscillator adjusted until a sharp dip in grid current is observed. The condenser calibration then indicates the resonant frequency of the circuit being measured.

A second use for such a meter is to determine the frequency of another oscillator or r-f generator. A standard radio receiver is used to pick up the signal from the r-f generator whose frequency is to be determined. Then the wavemeter is tuned so that it is also picked up by the receiver. When the oscillator and wavemeter have almost the same frequency, a beat-note will be heard in the receiver. As the wavemeter frequency is changed, the pitch of the beat-note will go up and down, and it will be found that if the wavemeter frequency is increased slowly the beat-note will start at a high pitch, which will become lower as the wavemeter frequency is increased. At a certain frequency the beat-note disappears, but if the frequency is still further increased, the pitch of the beat-note reappears and becomes higher as the wavemeter frequency is increased. The beat-note disappears (zero beat) when the oscillator and wavemeter frequencies are the same. Therefore, the frequency of the oscillator may be determined accurately. The beat-note is often called a "heterodyne" note and the meter a "heterodyne" frequency meter.

A grid-dip meter may also be used to calibrate a radio receiver. This usage requires that the signal output of the wavemeter be modulated with an a-f note. Such a modulating circuit is shown in Fig. 4. After the wavemeter is set to a desired frequency, the receiver is tuned to give a maximum sound output. The receiver is then in tune with the wavemeter, and its dial may be marked with the correct frequency.

9.4. Calibrating a grid-dip meter. A frequency meter, to be most useful, must be properly calibrated. This may be done in several ways. If the meter is of the grid-dip (heterodyne) type, the process is simple; all one needs is a source of known frequency and a receiver. If the frequency meter is of the type described in Sect. 9.2, a calibrated r-f oscillator is required.

The following experiment demonstrates how to calibrate a grid-dip meter using a radio receiver:

Experiment 1. Tune the receiver to a broadcast station whose frequency is known. Then turn on the frequency meter and tune until a whistle is heard in the receiver. Adjust for the lowest-pitched note obtainable. Now the frequency meter is at the same frequency as the known station. Next tune the receiver to another station whose frequency is known and repeat the procedure. After the process has been repeated for several known frequencies, a curve can be plotted showing the calibration of the frequency meter.

9.5. Standard frequency broadcasts. In this country standard frequency signals are broadcast by the National Bureau of Standards station, WWV, Beltsville, Maryland. These signals are broadcast on the following frequencies: 2.5, 5, 10, 15, 20, 25, 30, and 35 Mc. The signals are modulated alternately with 440- and 600-cycle a-f tones. Also broadcast are standard time intervals of 1 second, and 1, 4, and 5 minutes. Time announcements are made at 5-minute intervals by voice and International Morse Code. The frequencies are kept within extremely close limits. This station should be heard on one of its frequencies at almost any time anywhere in the United States. A sister station, WWVH, in Hawaii broadcasts a similar service.

For many purposes, broadcast and short-wave stations are good standards of frequency.

9.6. Properties of coils. Various properties of a coil such as its distributed capacitance, natural wavelength, and r-f resistance may be investigated by performing the three following experiments:

Experiment 2. *Distributed capacitance and natural wavelength of a coil.* Wind on a form, about 3 in. in diameter, a coil of about 60 turns of rather large wire, preferably with silk or enamel insulation so that the distributed capacitance of the coil will be rather large. Connect it across a variable condenser whose maximum capacitance is about 500 $\mu\mu$f. Start with the maximum capacitance of the condenser and measure the resonant frequency of the coil-condenser combination by use of a grid-dip wavemeter. Then decrease the capacitance and repeat several times, say at 400, 300, 200, etc., $\mu\mu$f. Plot the results against C as shown in Fig. 5, that is, λ^2 (wavelength squared) against capacitance.

The equation of the resulting straight line is

$$\lambda^2 = 3.55LC_d + 3.55LC$$

where L is the true inductance of the coil in microhenries, C_d is the distributed capacitance of the coil in micromicrofarads, and C is the condenser setting in micromicrofarads. The equation states that the wavelength (in meters) squared is proportional to the capacitance in the circuit. The slope of the line divided by 3.55 is the inductance of the coil, that is,

$$L = \frac{\text{Slope}}{3.55}$$

and

$$\text{Slope} = \frac{\lambda_1{}^2 - \lambda_2{}^2}{C_1 - C_2}$$

where λ_1^2 and C_1 correspond to a particular point on the curve; likewise, λ_2^2 and C_2. The method is illustrated in Fig. 5.

The straight line, extended from the last condenser reading, crosses the λ^2-axis at some distance above zero. This point of crossing gives the natural wavelength squared of the coil itself and therefore the resonant wavelength to which the coil with no additional capacitance will tune.

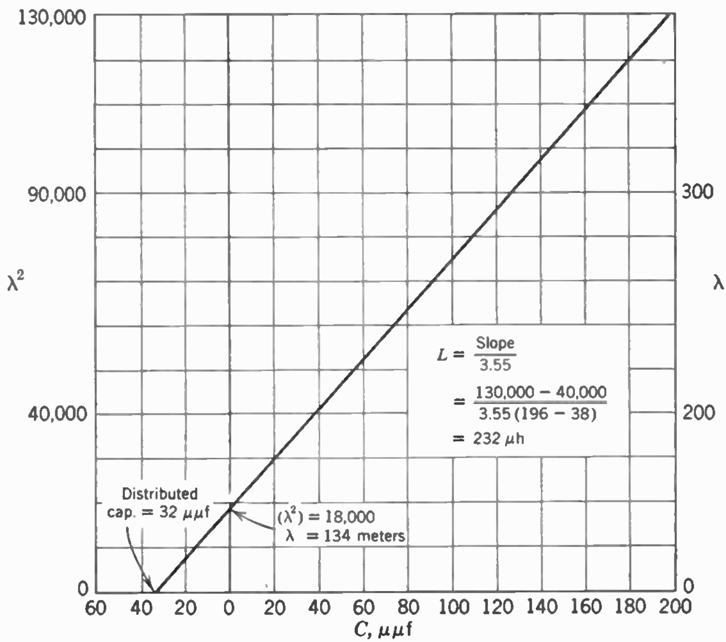


Fig. 5. A method of determining the distributed capacity of a coil.

The point where the line crosses the capacitance axis gives the distributed capacitance C_d of the coil. This value multiplied by the inductance as obtained above gives the LC product, which when fitted into the proper formula also gives the natural wavelength of the coil.

The data can be checked by disconnecting the condenser and then determining the natural frequency of the coil alone by the same method used to determine the resonant frequency with the condenser in place. This gives another method of determining the natural wavelength which can be compared to the values obtained by the methods above.

The effect of resistance upon the sharpness of resonance and the selectivity of the circuit has been mentioned. Also discussed was the increase of resistance with frequency. The resonance curve furnishes one method of measuring the resistance of a given circuit, provided that the inductance of the coil is known, or can be measured or calculated.

Experiment 3. *R-f resistance of a coil. Method 1.* Connect a coil, a condenser, and a low-current r-f milliammeter in series. Couple the coil loosely to a r-f generator with about 5 watts output as shown in Fig. 6.

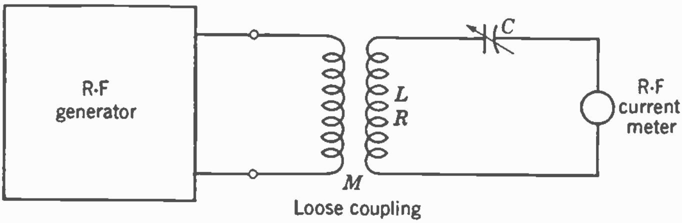


FIG. 6. Circuit for measuring r-f resistance of a coil.

Adjust the frequency of the generator to resonate with the series circuit as indicated by a maximum reading of the milliammeter. Change the generator frequency above and below the resonant value until two frequencies (f_1 and f_2) are found for which the current is 0.707 its value at resonance (f_r). Calculate the band width at this point and the resistance of the circuit from the equation

$$R = 2\pi L(f_2 - f_1)$$

Subtract from this calculated value of R the resistance of the milliammeter. The remaining resistance is the resistance of the coil, leads, and condenser. Most of the resistance is in the coil.

This method depends upon the generator current output being constant at the three frequencies f_1, f_2, and f_r. If the output varies widely, the results will not be accurate. If a second r-f milliammeter is available it can be placed in series with the output coil of the r-f generator, and the generator adjusted at each frequency to supply the same current to this meter.

Experiment 4. *R-f resistance of a coil. Method 2.* A second method necessitates the use of a series of accurately known resistances of negligible inductance and capacitance, and a variable condenser.

Small lengths of high-resistance wire (manganin) are to be preferred for frequencies higher than 1000 kc. Their d-c and r-f resistances are practically the same.

Connect the apparatus in series and couple rather loosely to a r-f generator. With the resistance box short-circuited ($R = 0$), tune the circuit to resonance. Then add enough resistance to the circuit to reduce the current to one-half its resonant value, retuning the r-f generator to resonance, if necessary. Then, since the current has been halved, the resistance has been doubled, for at resonance only resistance is effective in controlling the size of the current. In other words, the added resistance is equal to the resistance of the circuit. Again subtract the resistance of the milliammeter. The remaining resistance is the coil resistance plus the small resistance of the leads and condenser.

Repeat the experiment at several frequencies, and calculate the Q and plot against frequency and wavelength.

If only one or two resistance units are available, say 5 and 10 ohms, and not a continuously variable standard of resistance like a decade box, the resistance of the circuit above may be determined by noting the current at resonance, and the current when some resistance has been added, retuning to resonance after adding the resistance if necessary. Then the currents, according to Ohm's law, are

$$I_1 = \frac{E}{R_1}$$

$$I_2 = \frac{E}{R_1 + R_2}$$

where I_1 = current at resonance with no added resistance.
I_2 = current at resonance with R_2 added.
R_1 = resistance of the circuit.
R_2 = added resistance.

Then

$$R_1 = \frac{R_2 I_2}{I_1 - I_2}$$

The lower the resistance of the r-f milliammeter used in the above experiments, the greater will be the accuracy of the measurements.

9.7. Condenser capacitance. The capacitance of a condenser may be measured as described in the following experiment:

Experiment 5. *Capacitance of a condenser.* Connect a calibrated variable condenser across a r-f coil. With the condenser set near its maximum capacitance measure the resonant frequency of the combination. Attach the unknown condenser across the calibrated condenser and tune the calibrated variable condenser so that the resonant frequency is the same as before. The difference in the readings of the calibrated condenser is the capacitance of the unknown condenser. For example, suppose that resonance is obtained by the variable condenser alone when set to 400 $\mu\mu$f, and is obtained at 340 $\mu\mu$f when the unknown condenser is connected. The difference, 400 − 340 = 60 μμf, is the capacitance of the unknown. The capacitances of the coil and its leads do not enter into the calculations since they are in the circuit at all times. The leads of the unknown condenser should, however, be kept short if accurate results are to be obtained.

This method requires that the unknown capacitance be smaller than the calibrated condenser. A method for measuring the capacitance when the unknown is greater is discussed as an illustration of the use of the Q-meter (Chapter 21).

9.8. Antenna wavelength, capacitance, and inductance.

The natural wavelength, capacitance, and inductance of an antenna may be measured as described in the following three experiments:

Experiment 6. *Natural wavelength of an antenna.* Connect in series with an antenna an inductance which can be adjusted in even steps, say a coil of 20 turns with taps at each turn. Measure the frequency to which the antenna tunes with the entire coil in the circuit by loosely coupling the coil to a grid-dip meter. Then reduce the inductance by one turn, and repeat. Repeat until accurate readings are no longer possible. Plot wavelength, λ (meters) $= (3 \times 10^5)/f$ (kilocycles), against added turns of wire. Extend the curve until it crosses the wavelength axis. This point of crossing gives the natural wavelength of the antenna.

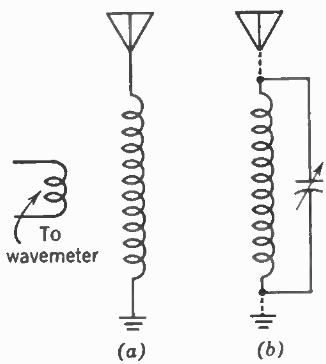


Fig. 7. To measure the capacitance of an antenna.

Experiment 7. *Antenna capacitance.* Connect a coil in series with the antenna and find the resonant frequency of the combination with a wavemeter. Then replace the antenna by a variable condenser as in Fig. 7 and tune the condenser to obtain resonance at the same frequency as before. The capacitance of the condenser is equal to the capacitance of the antenna.

Experiment 8. *Antenna inductance.* Connect a known inductance L_1 in series with the antenna and measure the resonant frequency f_1. Repeat, using a different inductance L_2 to obtain another resonant frequency f_2. The two frequencies are related as follows:

$$f_1 = C \times \frac{1}{\sqrt{(L_1 + L_a)C_a}}$$

$$f_2 = C \times \frac{1}{\sqrt{(L_2 + L_a)C_a}}$$

where L_a = antenna inductance in microhenries.
C_a = antenna capacitance in micromicrofarads.
C = a constant, the same in both equations.

Squaring both equations and solving for L_a,

$$L_a = \frac{f_1{}^2L_1 - f_2L^2{}_2}{f_2{}^2 - f_1{}^2}$$

The frequency in the above equation may be expressed in either kilocycles or megacycles, whichever is convenient.

9.9. Coil-condenser applications. Combinations of coils and condensers perform useful services throughout the electric communication art. Not only is the selectivity of a radio receiver produced by the properties of a resonant circuit, but such circuits may also be used to eliminate undesirable bands and to change the frequency response of a system in a desired manner.

For example, a filter may be used with a phonograph amplifier to reduce the noise caused by needle scratch. This filter, in one form, is a series-resonant circuit placed across the line between the pickup and the amplifier it feeds into, as shown in Fig. 8.

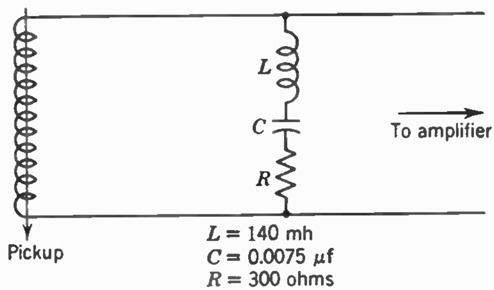


Fig. 8. Use of a series-resonant circuit as a low-impedance trap to reduce noise from needle scratch.

The resonant circuit tunes to about 4900 cycles, and, since it is a series-resonant circuit, its impedance to frequencies near 4900 cycles will be very low. It will produce a low-impedance path for the needle-scratch noise, the frequency of which is in this general region, and thus the needle scratch is reduced. Of course, any music frequencies in this region will also be reduced.

9.10. Frequency filters. Since coils and condensers have varying effects upon a circuit, one having a reactance which increases with frequency and the other a reactance which decreases with frequency, combinations of L and C may perform many useful functions. Filters are combinations of L and C * that may be used to transmit only a required portion of a wide frequency spectrum, or, conversely, to attenuate certain frequencies.

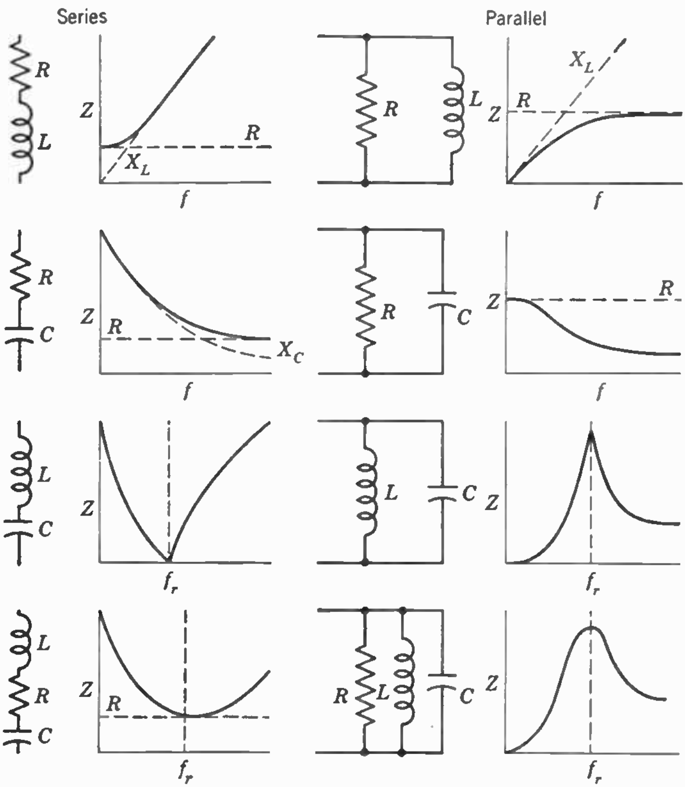


FIG. 9. Impedance diagrams of series and parallel combinations of L, C, and R.

A filter that passes low frequencies but attenuates (discriminates against) frequencies above the cut-off frequency is known as a **low-pass** filter; a combination of L and C that passes high frequencies but attenuates frequencies lower than the cut-off

* Combinations of L and R, or C and R, may also be used to discriminate against certain bands of frequencies as shown in Fig. 9. Such combinations, by themselves, do not have the sharp "cut-off" characteristics obtainable in L and C networks.

frequency is called a **high-pass** filter. If a band is attenuated while all other frequencies both higher and lower than this band are transmitted, the filter is a **band-elimination** filter; if a band of frequencies is transmitted and all other frequencies are attenuated, the unit is called a **band-pass** filter.

If sharper cut-off is desired than that obtainable with one filter section, several sections of these elementary filters may be connected in cascade (one following the other), each attenuating still further the undesired frequencies transmitted by the one before. For applications in which the filter characteristics must meet exacting specifications, very special filter networks have been devised. These networks are usually quite complicated. The basic ideas used, however, are the same as those discussed in this section.

The values of L, C, and the terminal resistance R, between which the filter works, for the simple filters shown in Fig. 10, are given in Table 2.

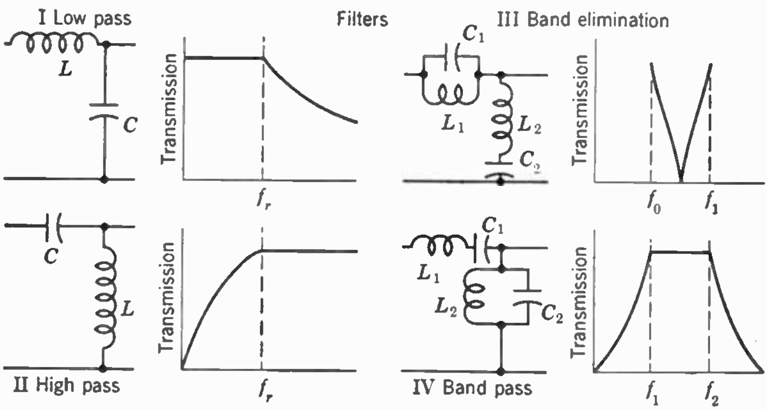


FIG. 10. Examples of the use of coils and condensers as filters.

TABLE 2

VALUES OF L AND C FOR FILTERS

I. $L = R/\pi f_r$; $C = 1/\pi f_r R$

II. $C = 1/4\pi f_r R$; $L = R/4\pi f_r$

III. $\begin{cases} L_1 = (f_1 - f_0)R/\pi f_0 f_1; & C_1 = 1/4\pi(f_1 - f_0)R \\ L_2 = R/4\pi(f_0 - f_1); & C_2 = (f_1 - f_0)/\pi R f_0 f_1 \end{cases}$

IV. $\begin{cases} L_1 = R/\pi(f_2 - f_1); & C_1 = (f_2 - f_1)/4\pi f_2 f_1 R \\ L_2 = (f_2 - f_1)R/4\pi f_2 f_1; & C_2 = 1/\pi(f_2 - f_1)R \end{cases}$

9.11. Frequency regulation. A simple example of the use of the diverse effects of L and C is shown in Fig. 11. As long as the frequency is equal to the series-resonant frequency of L and C, the two tubes are supplied with voltages of the same magnitude. Equal currents are produced in the output transformer. The currents in the transformer primary are opposite in direction (since the voltages supplied to the two tubes are 180° out of phase) and cancel each other so that the secondary current is

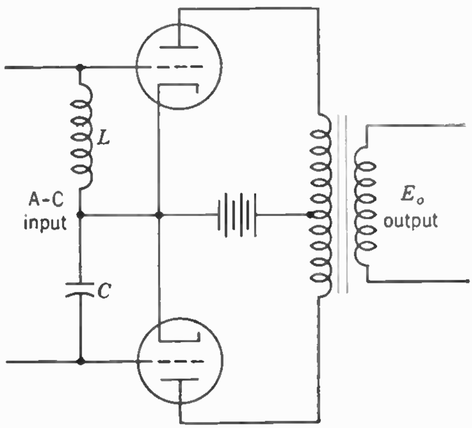


Fig. 11. Use of L and C as a frequency-regulating network.

zero. If, however, the input frequency rises, more voltage will be applied to the L tube than to the C tube, and more current will flow through it. Conversely, if the frequency drops, a greater voltage will be impressed on the C tube and more current will flow in it. These variations in current, since they are in opposite directions, can be utilized in a circuit to maintain the input frequency at a fixed value. Thus the circuit may be used in a frequency-control system.

Problem 1. A certain coil and condenser (LC_1) tune to a frequency f_1. The condenser is replaced by another, C_2. The circuit now tunes to a frequency f_2. Prove that $f_2 \div f_1 = \sqrt{C_1} \div \sqrt{C_2}$.

Problem 2. In an experiment to determine the r-f resistance of a 100-μh coil, the frequency of resonance with a certain condenser is found to be 1500 kc. The current falls to 0.707 times the resonant value when the frequency is 1480 and 1520 kc. What is the r-f resistance of the coil? What is the Q at the resonant frequency? Neglect the meter resistance and assume that the lead and condenser resistances are negligible.

Problem 3. Work Prob. 2 if the meter resistance is 10 ohms. Compare the results of the two problems.

Problem 4. A coil is connected in parallel with a condenser. The combination is coupled loosely to a r-f generator. At resonance the current is measured with a r-f ammeter with 20 ohms resistance and found to be 100 ma. A 15-ohm resistor is inserted in the circuit and the current drops to 80 ma. What is the resistance of the coil-condenser circuit alone?

Problem 5. A coil-condenser combination tunes to 1000 kc when the condenser is 400 $\mu\mu$f. When an unknown condenser is placed in parallel the circuit tunes to 800 kc. What is the unknown capacitance?

Problem 6. A coil-condenser combination tunes to 800 kc when the condenser is 400 $\mu\mu$f. When an unknown condenser is placed in series the circuit tunes to 1000 kc. What is the unknown capacitance?

Problem 7. The following data are recorded for a test as outlined in Experiment 2:

λ, Meters	C, $\mu\mu f$
100	100
125	180
150	290
175	400

Plot λ^2 against C and calculate the natural frequency, distributed capacitance, and inductance of the coil.

Problem 8. Data are taken for an antenna as outlined in Experiment 6.

f, Mc	Turns
0.75	20
0.858	16
1.0	12
1.33	8
1.5	4

Determine the natural wavelength and natural frequency of the antenna.

Problem 9. Suppose the "scratch" filter in Fig. 8 is to be tuned to a scratch frequency of 3500 cycles by shunting an additional condenser across C. How large must the added condenser be?

10 · Vacuum Tubes

The radio tube is a marvelous device. It makes possible the performing of operations, amazing in conception, with a precision and a certainty that are astounding. It is an exceedingly sensitive and accurate instrument. . . . Its future possibilities, even in the light of present-day accomplishments, are but dimly foreseen; for each development opens new fields of design and application. . . . The importance of the radio tube lies in its ability to control almost instantly the flight of millions of electrons . . . with a minimum of control energy.*

10.1. General principles of operation. The principles of operation of a vacuum tube are not complex. In the simplest tubes only two elements are required, a source of electrons, called the **cathode,** and an electrode to collect the electrons, called the **anode** or **plate.** A third element is often inserted between the cathode and plate to exercise control of the number of electrons that can pass to the plate. This element is called a **grid** and is an open grid or mesh of wires. Additional grids, which perform various auxiliary functions, are often inserted in the tube between the first or control grid and the plate.

A tube will pass current only when the plate is positive with respect to the cathode. A grid between cathode and plate can exercise a large measure of control of the amount of current that flows. As the voltage between cathode and grid is made more and more negative, fewer electrons flow to the plate, and finally the flow of electrons is entirely cut off. This control is exercised almost instantaneously, a change in grid voltage reflecting itself immediately in a change in plate current, and with little expenditure of power in the control circuit. Thus, the engineer is provided with a device for controlling the flow of an electric current which may be as small as a fraction of a microampere or as great as many amperes.

* From the RCA Receiving Tube Manual RC-14.

A tube will produce direct current from alternating current, or vice versa. It will act as a voltage or current amplifier or will perform as a frequency changer. It will merge audio- and radio-frequency currents for transmission as a radio signal or will separate them in a radio receiver. It will convert voltage changes into changes of light intensity, and vice versa. It will convert mechanical motion into an electrical signal. It will act as a voltmeter, ammeter, or frequency meter. It will control industrial machinery with a precision possible in no other way. The electron tube is indeed a modern miracle.

10.2. The Edison effect. In his search for more efficient filaments for incandescent lamps, Thomas A. Edison discovered that the glass bulbs of the lamps were blackened in the form of a shadow of the filament. An investigation showed that particles were leaving one side of the filament and crossing the evacuated space to the other leg of the filament. The particles that missed the filament hit the glass and blackened it. Subsequent research disclosed the fact that the particles carried an electric current which could be collected by a positive metallic plate placed within the bulb. The particles were electrons, although Mr. Edison did not know it at the time. The phenomenon by which the bulbs were darkened is now known as the **Edison effect.**

This discovery lay dormant for several years and was only a scientific curiosity described by physics lecturers. Then Fleming, in England, made use of the ability of the filament to release electrons to a near-by positive electrode to devise a more sensitive radio detector than had existed up to that time. The electrons could flow only away from the filament, current could flow in only one direction, and rectification occurred. Several years passed during which the Fleming "valve" was widely used for detection, but during this time little or no improvement was effected. Then Lee de Forest, in the United States, made a most important invention. He introduced a third element, the control grid, between the filament (cathode) and the electron collector (anode or plate). This invention provided a device in which the plate current was under the control of an auxiliary electrode and opened wide avenues of usefulness for the vacuum tube.

Much work remained to be done before the radio tube (de Forest called it the "audion") reached its present stage of development, but his basic idea made possible modern-day radio broadcasting, the control of many industrial processes, radar, electronic computers, and a host of other fascinating applications.

10.3. Electron emission. The heart of a tube is the source of electrons. There are several ways in which free electrons are obtainable.

1. Thermionic emission. When a metallic filament such as tungsten is heated to a sufficiently high temperature (2700° C for pure tungsten), some of the outer electrons escape from the tungsten atoms, break through the surface of the filament, and are free to be moved about by electric fields. This process resembles the escape of molecules of water from the surface of water heated to the boiling point. At the surface of water, or of tungsten, exists a barrier through which the moving molecules, or electrons, cannot pass under ordinary circumstances; but when they have sufficient energy imparted to them, as by heating, they break through the "surface tension" and escape.

2. Secondary emission. When an electron is accelerated to a sufficiently high velocity, it may have enough kinetic energy imparted to it to knock one or more electrons out of any material with which it comes in contact, either a metal conductor or an insulator. A positively charged electrode situated near the source of these "secondary" electrons will collect them. In actual tubes the secondary electrons may be attracted back to the electrode from which they came, as from the plate, or they may be collected by another electrode which is positively charged. In many tubes these secondary electrons give rise to undesirable effects, and design steps are taken to reduce their number and to control their movements. In a few tubes, such as electron multipliers, the desired operation is based on the principle of secondary emission.

3. Photoelectric emission. When light of proper wavelength is allowed to fall upon certain metals, electrons are released from the surface of the metal as a result of the energy imparted by the light. Here, then, is another electron source. Such sources are used in phototubes and in certain types of television camera tubes.

4. Field emission. Electrons may be literally pulled out of the surface of metals if a sufficiently high potential is established between two electrodes. The number of electrons emitted in this fashion depends upon the electric field intensity, that is, on the volts per meter. This high electric field intensity may actually occur with relatively low potentials between two tube elements if the elements present sharply curved surfaces to each other. The energy which causes electron emission in this case is derived from the high electric field. Field emission is not often used as a means of obtaining electrons in tubes; rather it is a phenomenon which is guarded against by properly shaping the electrodes so that high fields are avoided.

Practically all tubes used in radio equipment employ thermionic emission. Relatively few use any of the other types of emission, and these few are generally special-purpose tubes.

10.4. Types of thermionic cathodes. Thermionic cathodes are of two general types: **directly** and **indirectly heated.** In the directly heated type the filament itself serves as the source of the electrons. It is heated by a battery or from an a-c source. In the second type the filament merely acts as a heater for raising the actual electron emitter to the proper operating temperature; it is insulated from and placed within a metallic sleeve or cylinder which is coated with an emitting substance. The cylinder then becomes the cathode.

In the usual circuit, the plate of a tube is connected through a load resistor and a battery to the cathode. In this manner, the electrons which pass from cathode to plate *inside* the tube have a path *outside* the tube to return from plate to cathode. This connection to the cathode is called the **plate** (or **anode**) **return.**

If a tube employing a directly heated cathode is operated from an a-c source, care must be taken to return the electrons from the anode to the center of the filament rather than to one end. Otherwise, the cyclic change in voltage of one end of the filament with respect to the other, as the current changes its direction, produces hum in the output signal of the tube. In a very few types of tubes the center of the filament is available at a terminal in the tube base. Usually, however, the electrical center of the filament is attained by supplying the filament from a center-tapped secondary winding of a transformer. The anode return is made to this center tap, and hum is minimized. In circuits in which a very low hum level is required, these tubes may be heated by batteries or by direct current obtained from rectified and filtered alternating current.

An indirectly heated cathode usually has no electrical connection to the heater, and the filament or heater can be operated from an a-c source without having alternating voltages appear on the cathode. The cathode, therefore, acts as though all portions of it were at the same potential, and it is called a **unipotential cathode.** No hum results from the use of a tube employing this type of cathode if the tube is properly designed and operated. Some hum may be introduced by improper operation as a result of the small capacitance existing between the heater and the cathode.

10.5. Emitter materials. Thermionic electron emitters are usually constructed of pure tungsten, thoriated tungsten, or certain metallic oxides such as those of barium and strontium. Pure tungsten filaments operate at high temperature, consume considerable filament power, but are mechanically sturdy. If the tungsten filament is impregnated with thorium during manufacture, the operating temperature required for copious electron emission is materially reduced. Less heating power is required than for a pure tungsten filament, but the filament is not as sturdy and the operating temperatures are more critical. If a metal sleeve, such as nickel, is coated with certain alkaline earths (barium or strontium oxide), the temperature required for emission is further reduced. These oxide cathodes are widely used in receiving-type tubes because of their low filament power requirements. They are easily damaged, however, and are not adapted for use in high-voltage tubes. Indirectly heated tubes almost invariably employ an oxide-coated cathode.

In general, receiving tubes and small transmitting tubes, up to around 30 watts plate dissipation, have oxide-coated cathodes; transmitting tubes up to about 1000 watts plate dissipation have thoriated-tungsten filaments; and larger tubes have pure tungsten filament-type emitters. Practically all mercury-vapor and other gaseous tubes have oxide-coated cathodes.

Tungsten filaments are operated at a white heat, thoriated tungsten filaments at a yellow heat, and oxide-coated cathodes at a dull red heat.

Directly heated cathodes (filament-type) are more efficient than heater-cathode (unipotential) types employing the same emitter material. An oxide-coated cathode is more efficient than a thoriated tungsten cathode, which in turn is more efficient than a pure tungsten cathode.

10.6. Diodes. The simplest radio tube consists of a cathode (either directly or indirectly heated) which acts as a source of electrons and an anode which acts as a collector. The characteristics of such a tube, known as a **diode** or two-element tube, can be studied by connecting it in the circuit shown in Fig. 1. When

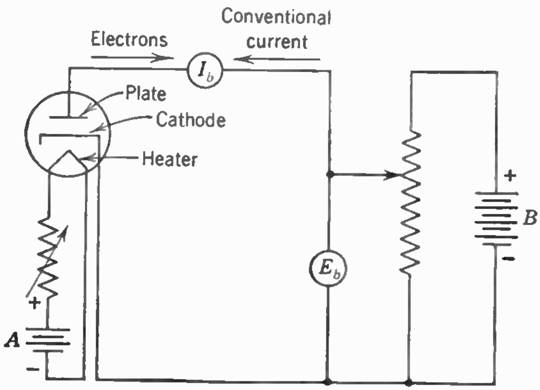


FIG. 1. Connections for studying a two-element or diode tube. Here the two elements are the cathode and the plate, the heater serving only to heat the cathode to emitting temperature.

the cathode is heated by means of a battery or other source of electrical power, and the plate circuit is completed, current will flow. The amount of current * flowing depends upon two factors: (1) the temperature of the cathode and (2) the voltage between plate and cathode.

If the filament voltage is held constant and the plate-cathode voltage is increased gradually, it will be found that, at first, the tube current rises at about the same rate as the plate voltage rises. Ultimately, however, an increase in plate voltage does

* The direction that electrons flow inside the tube is always from the cathode towards the plate; none flow in the opposite direction. Thus, electrons leave the tube at the plate terminal and enter it at the cathode terminal. Conventional current, the direction of which is opposite to electron flow, flows into the plate terminal. The term "current" will be used as a general term in the description of tube circuits. When current direction is of importance, "electron flow" will be used and clearly stated as such.

not result in a corresponding increase in plate current; the current, in fact, levels off to an almost constant value. When this happens, the plate is attracting all the electrons that are being emitted by the cathode. An increase in plate voltage does not materially increase the number of electrons, or the current; it merely speeds up their flight from cathode to plate. This phenomenon is called **temperature saturation.** If the filament

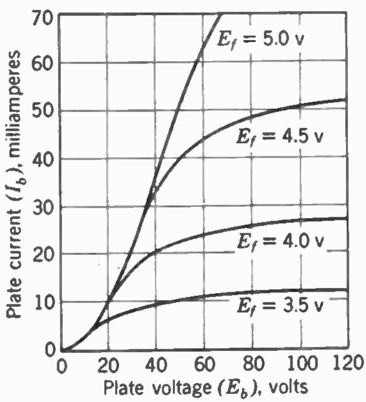


FIG. 2. Curves of typical diode. Leveling of upper portions of curves is caused by temperature saturation.

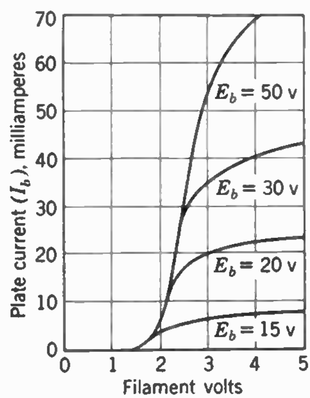


FIG. 3. Curves of typical diode. Leveling of upper portions of curves is caused by voltage saturation.

voltage is reduced, the value of saturation current is less because the cathode furnishes fewer electrons at the lower temperature. The general shape of the curves is shown in Fig. 2.

If, instead of varying the plate voltage, it is held constant at some positive value and the filament voltage is increased, the plate current will be zero or very small until a certain value of filament voltage is reached. Beyond this value of filament voltage the plate current increases very rapidly, but levels off as the filament voltage is increased still more. The leveling off in this case occurs when the plate attracts all the electrons it can at the particular plate-cathode voltage. This limiting value of current is called the **voltage saturation** current and results from space-charge effects discussed below. This limiting value becomes higher if the plate voltage is made higher. The general appearance of the curves is shown in Fig. 3.

10.7. Space charge. Electrons not collected by the plate tend to congregate in the space between cathode and plate and form a negatively charged cloud called a **space charge.** This space-charge cloud is densest near the cathode. Electrons attempting to leave the cathode now find this negative barrier between them and the plate and are, of course, repelled by it. Many of the electrons return to the cathode, some pass on to the plate, and others remain in the space-charge cloud. The negative cloud has the same effect as decreasing the positive plate voltage, and the only way to increase the flow of electrons to the plate is to increase the plate voltage.

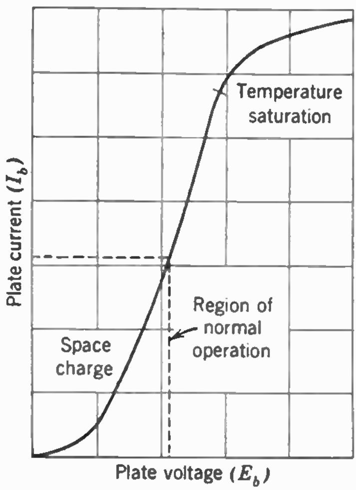


FIG. 4. A typical diode characteristic, showing normal region of operation and the regions in which space charge and temperature saturation are effective.

The effect of space charge on tube characteristics is best seen in Fig. 4. At the lower end of the curve, where the plate voltage is low, space charge has its greatest effect and causes considerable curvature. Further up on the curve, space charge has less effect and the curve straightens out. Then, as temperature saturation effects appear, the curve flattens out again.

So long as there is an adequate supply of electrons the space charge controls the plate current in such a way that the plate current varies very closely as the $3/2$ power of plate voltage. That is, $I_b = KE_b^{3/2}$, where K is a constant which takes into account the size and spacing of the electrodes. This equation is often called **Child's** or **Langmuir's law.**

A manufacturer ordinarily rates a tube so that the maximum allowable plate current is one-third or less of the saturation current when the filament (or heater) is operated at rated voltage. Then, as the tube is used, the cathode may lose some of its emission capabilities and still supply an adequate number of electrons. In this way the useful life of the tube is made longer. The normal operating region is shown by the dashed lines in Fig. 4. It is only under special circumstances that operation extends into the temperature-saturated region.

10.8. Tube nomenclature. To identify easily various voltages and currents in a tube circuit, a group of letter symbols is used. For example, the filament battery is called the **A** battery; the plate supply battery is called the **B** battery; and if there is a grid battery it is known as the **C** battery. The various voltages, currents, power, etc., associated with different tube elements are identified according to the following general rules: *

Instantaneous values of current, voltage, and power which vary with time are represented by the lower-case letter of the proper symbol. EXAMPLES: i, e, p, i_g, e_b.

Maximum, average (d-c), and root-mean-square values are represented by the upper-case letter of the proper symbol. EXAMPLES: I, E, P, I_b, E_b.

If necessary to distinguish between maximum, average, or root-mean-square values, maximum values may be represented by the subscript m, average values by the subscript av, and root-mean-square or effective values by the upper-case letter without subscript. EXAMPLES: E_m, I_{p_m}, E_{av}, $I_{p\mathrm{av}}$, E, I, E_g.

The electrode abbreviations to be used as subscripts are:

- j general (for any electrode)
- f filament
- h heater
- k cathode
- g grid (c also used)
- p plate or anode (b also used)
- s metal shell, or other self-shielding envelope
- d deflecting, reflecting, or repelling electrode (electrostatic type)

Further letter symbols will be introduced in connection with the discussion of particular tube connections.

10.9. Diode operation. If a plate battery is connected in series with a resistance and a diode, as in Fig. 5, current flows through the circuit and represents power taken from the battery. The power consumed in the resistance R_L is equal to the product of the current through and the voltage across the resistance.

* "Standards on Abbreviations, Graphical Symbols, Letter Symbols, and Mathematical Signs," 1948, The Institute of Radio Engineers.

The power taken from the battery is equal to $E_{bb}I_b$, and this is greater than the power consumed in the resistance. Where does the rest of the power go?

A rectifier tube in a radio receiver or transmitter gets hot in operation. It can be observed that the more direct current demanded from a tube, the hotter it gets. This means, then, that not all the heating of the tube results from the power supplied to the filament, which is constant; some of the heat is supplied by current flowing in the plate-cathode circuit. This additional heating effect can be accounted for by assigning a value of resistance to the electron path between cathode and plate. This effective resistance results from the fact that energy must be supplied the electrons to accelerate them and move them to the anode. They release the kinetic energy due to their motion as heat as they strike the anode. This effective resistance is called the **static** (or d-c) **plate resistance** (r_b) of the tube. It is the ratio of the voltage to the current at a particular point on the tube characteristics, or $E_b \div I_b$. The power lost in the tube, then, is equal to the voltage across the tube itself multiplied by the current through the tube; or to $I_b^2 \times r_b$. The various power relations in Fig. 5 are

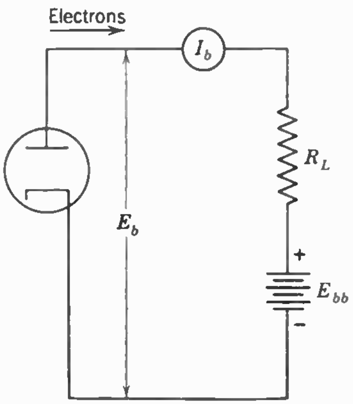


FIG. 5. Power from the battery is used both in the resistor R_L and in the tube.

$$\text{Battery power} = E_{bb}I_b$$

$$\text{Power in resistance} = I_b^2 \times R_L$$

$$\text{Power in tube} = \text{Battery power} - \text{Power in resistance}$$

$$= E_{bb}I_b - I_b^2R_L = I_b^2r_b$$

The static plate resistance can be found for any plate current by dividing the power consumed in the tube by the current squared, as well as from $E_b \div I_b$.

In Fig. 6 is shown the plate characteristic of a typical receiver rectifier, the 35Z5-GT/G. The value of the plate resistance of the tube at any value of plate current may be found by dividing the voltage by the current. This may be done for several values of plate current, and the resulting resistance plotted against plate current to determine graphically the manner in which plate re-

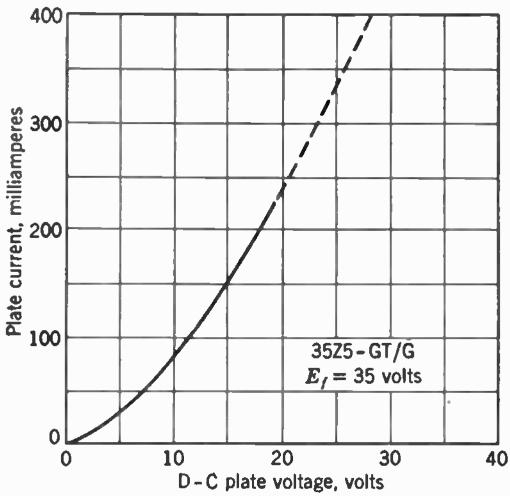


FIG. 6. Plate characteristics of a typical diode. The filament voltage is sufficient to supply more electrons than will be collected by the plate.

sistance varies with current. It should be noted that plate current does not follow Ohm's law; that is, doubling the voltage does not double the current. As a matter of fact, the current varies approximately as the $\frac{3}{2}$ power of the voltage. Thus, doubling the voltage increases the current $2^{3/2}$ times; multiplying the voltage by 3 increases the current $3^{3/2}$ times.*

It will be noted that the plate resistance varies with the current through the tube. If the tube is acting as a rectifier, that is, if alternating voltage is impressed between cathode and plate,

* These fractional powers need not be confusing. For example, 2^3 equals 2 multiplied by itself 3 times; $2^{1/2}$ equals the square root of 2. Therefore $2^{3/2}$ equals the square root of 2 cubed. Thus, $2^{3/2} = (2 \times 2 \times 2)^{1/2} = \sqrt{2 \times 2 \times 2} = \sqrt{8} = 2.8$. The reader should take values from Fig. 6 and see how closely a 35Z5 tube follows the $\frac{3}{2}$ power law.

no current flows on the portion of the cycle which makes the plate negative in respect to the cathode. Under this condition the plate resistance is infinite. Throughout the half cycle in which the plate is positive, the current through the tube is varying and the tube resistance is varying also.

The static plate resistance of high-vacuum rectifiers is fairly high—ranging from around a hundred to a few thousand ohms. It is made low by placing the anode and plate close together. This method is subject to the limitation that if the elements are too close together there is danger both of short circuits and of the tube arcing over internally. It is advantageous to have the resistance as low as possible, since power loss in the tube subtracts from the amount of power available for an external load.

10.10. Gaseous tubes. If, after a tube has been evacuated, a gas such as mercury vapor is introduced, the tube characteristics are changed markedly. The static plate resistance is very much lower, and the voltage drop across the tube becomes decidedly lower and is essentially independent of the plate current. The tube glows with a characteristic color when current flows. The color is bluish for mercury vapor and is different for other gases.

Mercury vapor consists of atoms of mercury. Electrons leaving the cathode collide with the mercury atoms and liberate electrons from them. The mercury atom, having lost an electron, is positively charged. It is large and heavy and moves slowly toward the cathode. Large numbers of these positively charged mercury atoms (called **ions**) are found in the region of the cathode (Fig. 7) where their positive charge effectively neutralizes the negative space charge of the electrons and removes this inhibiting influence upon the number of electrons that can move to the plate. The number of electrons that can move to the plate is limited by the number supplied by the cathode, and by the impedance in the external circuit. There is some voltage drop across the tube, of the order of 15 volts, which is the potential required to ionize the mercury vapor.

A gas tube such as described above employs a heated cathode, and the current from the tube is obtained by thermionic emission from this cathode. If more current is required from the tube than the cathode can supply, the doubly charged positive ions which result may bombard the cathode hard enough to destroy it. For this reason, *the load resistance in a gas-tube circuit must always be high enough to protect the tube from excessive current.*

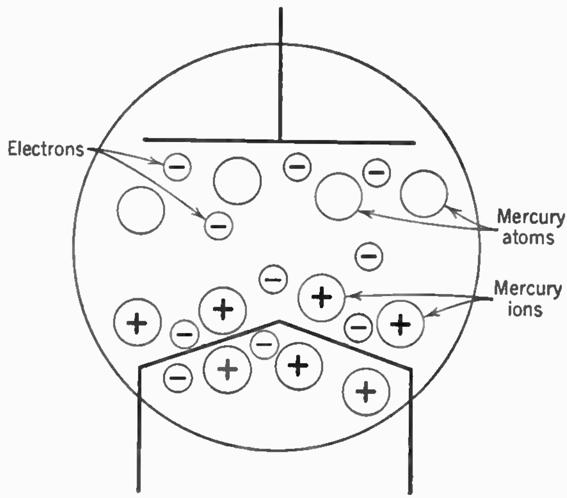


Fig. 7. A gaseous rectifier tube. The positive mercury ions gather near the negative cathode and neutralize the space charge effect of the emitted electrons.

10.11. Peak inverse voltage. How much voltage can be placed across a tube? The amount is limited by the construction of the tube. Excessive voltage will cause arcing and consequent damage. Spacing between cathode and plate, the kind of insulation, the placement of the terminals in the tube base, and the effect of heat on the electrodes and glass all govern the maximum voltage permissible. On the half cycle that the tube conducts current, if the tube is operating as a rectifier, the voltage drop across the tube is fairly low. It is actually the impressed voltage minus the voltage drop across the load. But during the half cycle in which the tube is not conducting, the total impressed voltage is across the tube, since there is no voltage drop across the load because no current is flowing.

The voltage across the tube during the half cycle when the tube is not conducting is known as the inverse voltage, and the maximum or peak value that may safely be placed across a tube is the **maximum peak inverse voltage.** It is an important tube rating.

10.12. Triodes. If a third element, like an open-mesh grid or similar structure, is placed between cathode and plate, the versatility of the tube increases enormously. Such a tube is called a **triode** or three-element tube. Now the tube will not only rectify, that is, change alternating to direct current, but it will also perform the many other purposes for which the electron tube is so well known.

The commercial names of various types of high-vacuum and gas-filled tubes are shown in Table 1.

TABLE 1

Number of Tube Elements *	Type	Commercial Name	
		High-Vacuum	Gas-Filled
Two	Diode	Kenotron	Phanotron
Three	Triode	Pliotron	Thyratron
Four	Screen-grid; Beam power	Pliotron	Shield-grid thyratron
Five	Pentode	Pliotron	....

* The heater is not counted unless it is also the cathode.

10.13. The purpose of the grid. The primary function of a grid is controlling the effect of the space charge on the amount of current that flows to the plate.

Suppose that the grid is made positive with respect to the cathode. Since it is physically nearer the cathode than is the plate, a small positive potential will have the same effect as a large positive potential on the plate. This positive potential on the grid cancels part of the effect of the space charge and the plate current is increased.

Suppose, however, that the grid is made negative with respect to the cathode.* Then, the effect of the space charge is increased, and the plate current becomes smaller. A comparatively small negative voltage on the grid may cause the plate current to drop to zero.

Electrons move from the cathode to the plate at very high

* It is general practice to refer all voltages to the cathode. Thus, when it is said "the grid is negative" it means "the grid is negative with respect to the cathode."

speeds. Therefore, there is practically no time lag between the time that the grid voltage is changed and the time that a corresponding change is noted in the plate current. Only at very high frequencies does this small time lag become significant.

10.14. Characteristic curves of triodes. In most circuits the filament voltage of the tubes is held at such a constant value that more electrons are available than are needed. Then, the factors which control the amount of current flow from the tube (the number of electrons attracted to the plate) are the grid and

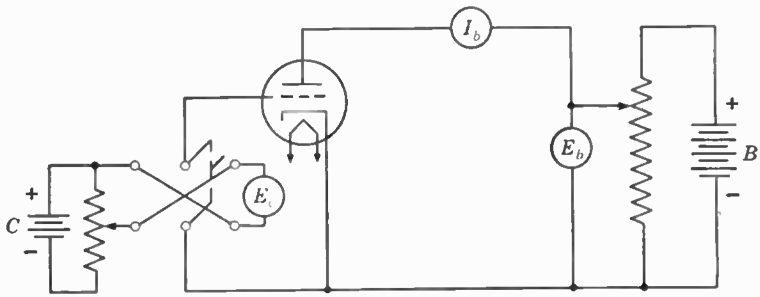


Fig. 8. Circuit for studying triode characteristics. The DPDT switch reverses the polarity of the grid voltage. The filament is held at its rated value, since that is the normal operation.

plate voltages. Raising either of these voltages in a positive direction will increase the current flow.

The characteristic curves of a triode may be obtained by placing a tube in a circuit as in Fig. 8. If either voltage is held constant and the other varied, the characteristic curve showing how the plate current varies as one voltage is changed is obtained. Meters in the circuit measure the various voltages and the plate current. The milliammeter (I_b) should always be placed closest to the plate of the tube, with the voltmeter (E_b) on the side nearest the plate battery, as shown in the figure. This is important, since otherwise the milliammeter will also indicate the current taken by the voltmeter, and this current may easily be as great as that taken by the tube.

Experiment 1. *Effect of grid bias upon plate current.* Set up the apparatus as shown in Fig. 8, using in succession several common triodes. Fix the filament or heater voltage at the value specified in a tube manual. Set the grid voltage to a negative value of around −8 or −10 volts and the plate voltage to 25 volts. Record the plate current as the grid voltage (known as the **bias**) is varied from a value negative enough to make the plate current zero to a value 1 or 2 volts positive. The DPDT switch in the grid circuit permits changing the polarity of the grid without changing the grid meter connection. Increase the plate voltage to 50 volts and repeat. Repeat for two or three other plate voltages, using care that the plate current does not exceed values recommended in the tube manual. Plot these data like those in Fig. 9.

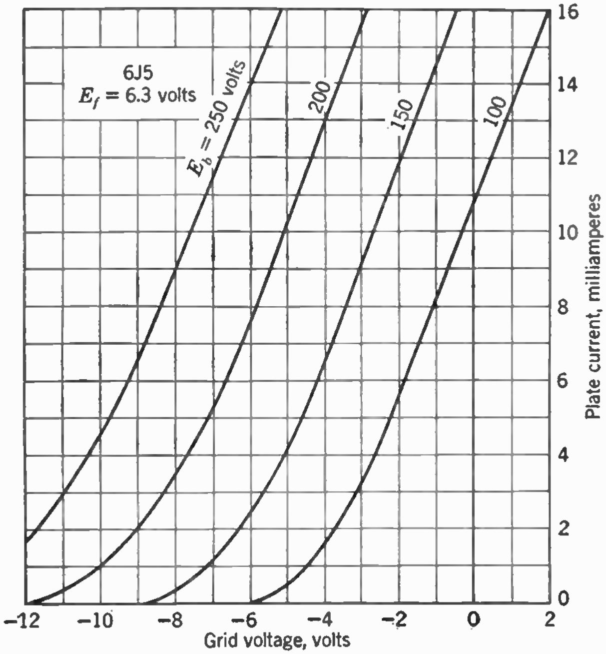


FIG. 9. Typical E_c-I_b curves. Note that the several curves corresponding to different plate voltages are parallel and have long, straight portions.

10.15. Grid voltage-plate current curves. These curves, also called the **transfer characteristics** or simply the E_c–I_b *curves*, yield several interesting and important facts regarding a triode. Referring to Fig. 9, it is seen that at high negative grid voltages for a particular value of E_b, say 100 volts, there is little or no plate current. As this negative voltage is decreased, some electrons get to the plate. The plate current begins to flow and increases at a rather slow rate, then more rapidly, then in a steep and almost straight line; finally, if the curve can be extended far enough without damaging the tube, the curve flattens out. (This flat portion is not shown in the figure.) Increasing the plate voltage to a higher value, say 150 volts, and again varying the grid voltage produce a new curve which is essentially parallel to the first, but moved to the left. Increasing the plate voltage a like amount again produces a new curve displaced an equal distance to the left of the second line. Such a graphic collection of data, known as a **family of curves,** gives a complete story of the effect of grid voltage upon plate current.

A milliammeter in the grid circuit will show that no grid current flows except when the grid is positive and even then it is a small fraction, one-tenth or less, of the plate current.

10.16. Plate voltage-plate current curves. The E_b–I_b curves are essentially parallel over their straight portions as

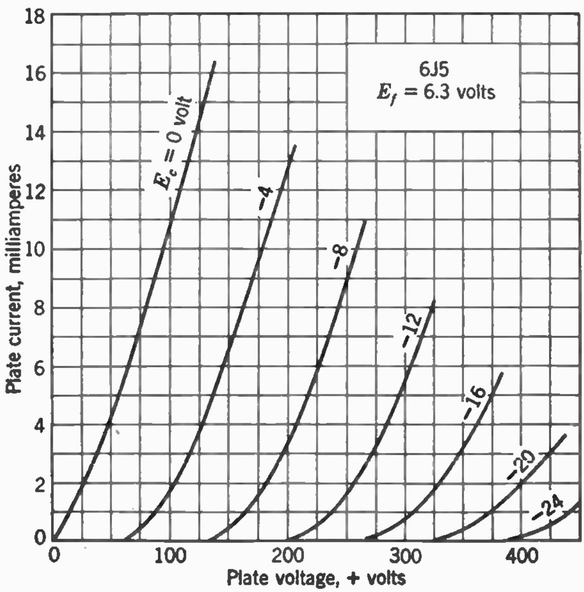


Fig. 10. Typical plate family of curves for a triode.

shown for a typical case in Fig. 10. If the grid voltages (bias voltages) chosen are in equal steps, the plate-current curves will be approximately equal distances from each other.

Note that the curves in Fig. 10 give the same information that is given in Fig. 9. For example, the plate current for $E_b = 200$ volts, $E_c = -8$ volts is 3.5 ma in Fig. 10. It has the same value in Fig. 9 since the two sets of curves are for the same tube. For some purposes it will be found that one set of curves is more convenient than the other.

Experiment 2. *Plate characteristics.* Set up the apparatus as in Experiment 1. Set the grid voltage at some value, say -6 volts for an ordinary receiving tube, and record the value of the plate current as the plate voltage is changed from 0 to perhaps 100 volts in 10-volt steps. Then change the grid voltage to -10 volts and repeat. Repeat for other values of grid voltage. Plot these data as a set of curves as in Fig. 10.

These data could be obtained by picking values from the curves plotted in Experiment 1.

10.17. Constant-current curves. The data in Figs. 9 and 10 can be replotted as in Fig. 11. To obtain this curve, combinations of plate and grid voltages have been selected that give curves of constant plate current. The curve for $I_b = 0$ ma is of particular importance. It is called the **cut-off** curve and gives the combinations of plate and grid voltage which cause the plate current to be zero.

10.18. Amplification factor. It has already been noted that the grid is more effective than the plate in controlling plate current. The reason is that the grid is closer than the plate to the cathode and to the region in which space charge is densest. How much more effective this control is can be determined from Fig. 11. Consider the curve of $I_b = 4$ ma at the point where $E_c = -8$ volts and $E_b = 205$ volts. If the current is changed to 8 ma by varying the grid voltage, the new value of E_c is -6 volts, or a change of 2 volts. If, however, the new current is obtained by increasing the plate voltage, the new value of E_b is 245 volts, or a change of 40 volts. The ratio of the two voltage *differences* is $40 \div 2 = 20$, and this ratio indicates the relative effectiveness of the grid and plate in controlling plate current. This ratio is called the **amplification factor,** μ (mu). Thus

$$\mu = -\frac{\text{Change of plate voltage}}{\text{Change of grid voltage}} = -\frac{\Delta E_b}{\Delta E_c} \quad \text{(at a given } I_b\text{)}$$

The minus sign is used since the grid and plate voltages must be varied in opposite directions to hold the plate current constant. This makes the sign of μ *positive.*

The amplification factor should not be confused with the actual voltage amplification realizable from a tube. Normally the actual voltage amplification is considerably less than the amplification factor.

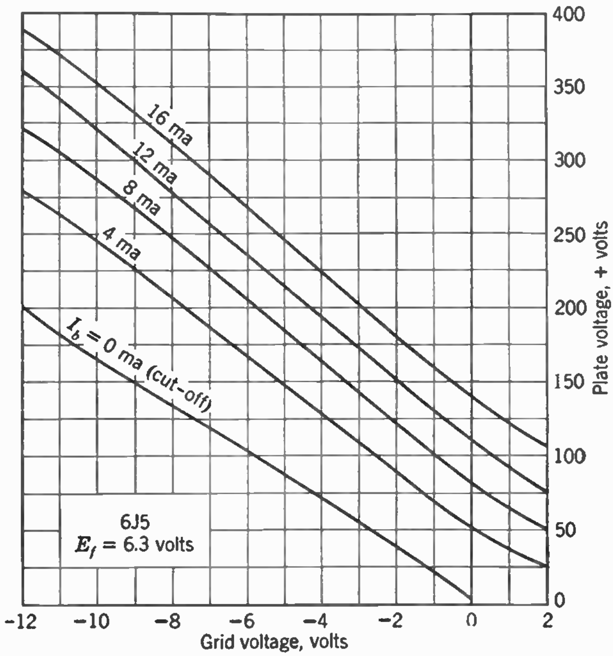


Fig. 11. Typical constant-current characteristics of a triode.

10.19. Dynamic plate resistance. Like a rectifier tube, the triode imposes a resistance to the flow of electrons or current. The static (d-c) plate resistance is the ratio between a given value of plate voltage and the corresponding value of plate current and can be most easily obtained from the E_b–I_b curves for a triode. However, since a triode is most often used with an alternating voltage impressed on the grid, there is also an alternating component of plate voltage and current. For this reason, the resistance of the tube to an alternating current is of importance. This resistance is called the **dynamic plate resistance,** r_p.

Ordinarily the tube is operated with certain fixed grid (bias) (E_c) and plate (E_b) voltages. Then, in addition, an alternating voltage is applied to the grid (sometimes to the plate) and alternating current flows in the plate circuit.

If the plate voltage of a triode is varied slightly, say a volt or so, and the corresponding change in plate current is noted, then the ratio of these *changes* is called the dynamic plate resistance. Thus

$$r_p = \frac{\text{Change in plate voltage}}{\text{Change in plate current}} = \frac{\Delta E_b}{\Delta I_b} \quad (E_c \text{ constant})$$

Example 1. Determine the static and the dynamic plate resistance of a 6J5 (Fig. 10) tube at the point $E_b = 150$ volts, $E_c = -8$ volts.

Solution: The plate current at the given point is 6.5 ma. and the static (d-c) plate resistance is simply

$$r_b = \frac{E_b}{I_b} = \frac{150}{6.5 \times 10^{-3}} = 23{,}100 \text{ ohms}$$

The dynamic plate resistance can be determined by taking 25-volt increments of plate voltage above and below 150 volts. Thus

$$\text{At } E_b = 175 \text{ volts,} \quad E_c = -8 \text{ volts,} \quad I_b = 9.6 \text{ ma}$$

$$\text{At } E_b = 125 \text{ volts,} \quad E_c = -8 \text{ volts,} \quad I_b = 3.8 \text{ ma}$$

Then

$$r_p = \frac{\Delta E_b}{\Delta I_b} = \frac{175 - 125}{(9.6 - 3.8) \times 10^{-3}} = \frac{50}{5.8 \times 10^{-3}} = 8620 \text{ ohms}$$

Note that the dynamic plate resistance is considerably lower than the static plate resistance. This is usually the case.

The dynamic plate resistance is, in reality, the slope of an E_b–I_b curve at a particular point. Since this slope changes from point to point, the dynamic plate resistance is not a constant and its value for a particular tube must be specified at particular plate and grid voltages.

10.20. Grid-plate transconductance. Another important tube constant remains, the **grid-plate transconductance** or **mutual conductance.*** This factor tells how much plate-current change is caused by a given grid-voltage change when the plate

* "Transconductance" is a general term which expresses the change in current to one tube element caused by a voltage change on another element and may be applied to a tube with several grids. "Grid-plate transconductance" is a special case, and one of much interest. "Mutual conductance" has the same meaning but is not the preferred term.

voltage remains constant. Thus

$$g_m = \frac{\text{Change in plate current}}{\text{Change in grid voltage}} = \frac{\Delta I_b}{\Delta E_c} \quad (E_b \text{ constant})$$

Thus, if a change of 1 volt on the grid of a tube produces a change of plate current of 1 ma, and the plate voltage is held at a fixed value, the grid-plate transconductance

$$g_m = \frac{1 \times 10^{-3}}{1} = 1 \times 10^{-3} \text{ mho or 1000 micromhos}$$

The **mho** is the unit of conductance. Note that it is "ohm" spelled backward. The micromho (one-millionth mho) is most often used in connection with tubes.

The grid-plate transconductance, amplification factor, and dynamic plate resistance of a tube are related, at any particular operating point, by the equation

$$\mu = g_m \times r_p$$

Experiment 3. *Amplification factor.* Set up the apparatus as in Experiment 1, and set the plate and grid voltages to, say, $E_b = 200$ volts, $E_c = -6$ volts. Note the plate current. Change the grid voltage by a small amount such as 0.5 volt and then change the plate voltage so that the plate current is the same as before. The *change* in plate voltage divided by the *change* in grid voltage required to hold the plate current constant gives the amplification factor. Repeat for other values of plate and grid voltage, some for which the plate current is almost zero, others for which the plate current is near the maximum rating for the tube. Compare results.

This same circuit can also be used for the experimental determination of r_p and g_m. The changes in voltages (currents) should be as small as can be accurately read on the meters.

10.21. Normal values of tube parameters. The amplification factor (μ), dynamic plate resistance (r_p), and grid-plate transconductance (g_m) are very important tube factors or quantities. They are often called tube constants, but they are not actually constant since they vary with the operating voltages. A better term is "parameter." The typical manner in which μ, r_p, and g_m vary is shown in Fig. 12. Note that, of the three, μ is the most nearly constant, r_p becoming very high and g_m very low at low values of plate current.

The amplification factor for triodes ranges as low as 3 or 4 for power triodes to as high as 100 for certain high-mu triodes. A large class of triodes has amplification factors of the order of 20.

The dynamic plate resistance ranges from a low of 800 ohms for certain power triodes to as high as 100,000 ohms for high-mu triodes. In general, the higher the amplification factor, the higher the dynamic plate resistance.

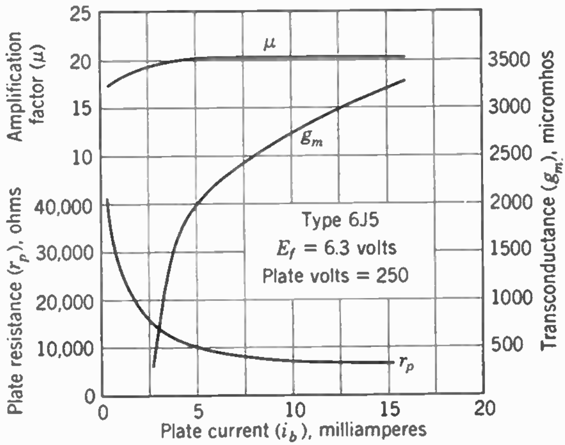


FIG. 12. Dependence of r_p, g_m, and μ on the plate current. Note that μ is almost constant in value and that g_m increases as r_p decreases.

For triodes g_m usually ranges from about 800 to 3000 micromhos.

10.22. Slopes of characteristic curves as tube parameters. All three tube parameters (μ, g_m, and r_p) can be obtained from any one of the three characteristic curves. However, it will aid in an understanding of how the parameters vary with operating voltages if it is recognized that the magnitudes of the parameters are determined by the slopes of the various characteristic curves. Some of these have already been illustrated. Table 2 gives the particular curves from which each parameter can most easily be determined.

Voltage changes as small as consistent with accurate reading of the graphs should be used in evaluating the tube parameters. Too large changes decrease the accuracy of the results.

TABLE 2

Tube Parameter	Equation	Slope of Which Curve?	Held Constant
Amplification factor, μ	$\mu = -\dfrac{\Delta E_b}{\Delta E_c}$	E_b–E_c	I_b
Transconductance, g_m	$g_m = \dfrac{\Delta I_b}{\Delta E_c}$	I_b–E_c	E_b
Dynamic plate resistance, r_p	$r_p = \dfrac{\Delta E_b}{\Delta I_b}$	I_b–E_b	E_c

10.23. Crystal diodes. In recent years there have been a large number of applications of crystal diodes. These are vastly improved descendants of the "cat whisker-crystal" detector used in the early radio receivers. The sensitive element is a small block of some semiconducting mineral such as germanium or silicon. A contact is made to one surface with a fine, pointed wire. The other surface is imbedded in a soft-metal base which forms a good electrical contact. Once the elements are adjusted for the proper sensitivity the whole unit is usually filled with an insulating compound which also holds the elements in place. The mechanical features of a typical crystal diode are shown in Fig. 4, Chapter 3.

The current-voltage characteristics of a typical crystal diode are shown in Fig. 13. Note that the "reverse" voltage and current scales multiply the values ten times over the "forward" scales. This is to show the reverse current-voltage relations in some detail. In the forward (conducting) direction, the characteristics are much like those of a high-vacuum diode except that there is usually somewhat more curvature.

Crystal diodes may be used in practically all applications for which high-vacuum diodes are employed. The fact that they do pass a small current in the reverse direction, whereas the reverse current of a high-vacuum diode is zero, must be taken into account in the design of apparatus in which they are used. Crystal diodes have the advantage of small size and weight and low capacitance. They have the disadvantage of low current-carrying ability, less uniformity than diode tubes, and relatively low mechanical ruggedness. Recent improvements have resulted in units which approach in ruggedness and uniformity the general run of high-vacuum tubes.

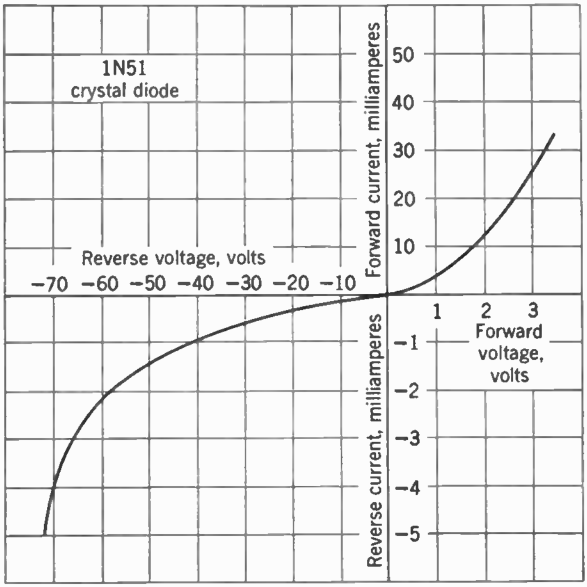
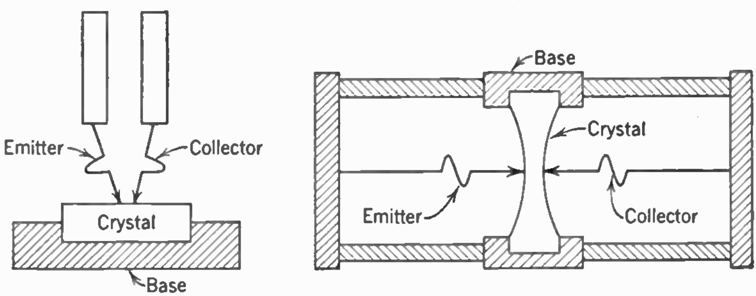


FIG. 13. Typical current-voltage characteristics of a crystal diode. Note that "reverse" scales are ten times larger than "forward" scales.

10.24. Transistors. The **transistor** is a new amplifying device barely out of the experimental stage. It utilizes a crystal as in a crystal diode, except that there are two fine contacts, very



(a) Simple Transistor (b) Coaxial Type

FIG. 14. Mechanical construction of typical transistors.

close together, on one surface. The other surface is imbedded in a conducting metal to form a base contact. A coaxial type places the two contacts on opposite sides of a thin wafer of the semiconducting material. Both types are illustrated in Fig. 14.

A typical circuit connection for a transistor is shown in Fig. 15. Note that the input circuit is biased positive, and that the output circuit is operated at a negative d-c voltage. This is opposite to the normal connection of a triode tube. As a result, the input im-

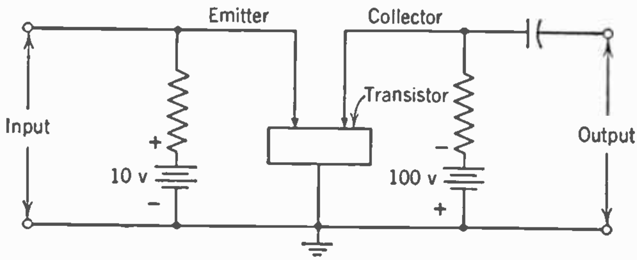


FIG. 15. Typical circuit connection of a transistor amplifier.

pedance tends to be fairly low. At the present time the voltage gain realized from a transistor amplifier is rather low and the noise level rather high. Further development will probably overcome these disadvantages, at least in part. The small size and lack of filament heating power are distinct advantages.

Problem 1. Calculate the static plate resistance of a 35Z5-GT/G rectifier tube (Fig. 6) at various values of plate current. Plot the resistance against plate current.

Problem 2. What plate current flows in a 6J5 tube at the following operating points?

(a) $E_b = 250$ volts, $E_c = -12$ volts.
(b) $E_b = 200$ volts, $E_c = -4$ volts.
(c) $E_b = 200$ volts, $E_c = -10$ volts.

Problem 3. Compute the amplification factor of a 6J5 triode at:

(a) $E_b = 300$ volts, $E_c = -9$ volts.
(b) $E_b = 200$ volts, $I_b = 4$ ma.

Problem 4. Compute the grid-plate transconductance of a 6J5 triode at:

(a) $E_b = 200$ volts, $E_c = -8$ volts.
(b) $E_b = 150$ volts, $I_b = 12$ ma.

Problem 5. Using graphical methods, determine the static and the dynamic plate resistances of a 6J5 triode at the following points:

(a) $E_b = 100$ volts, $E_c = 0$ volt.
(b) $E_b = 225$ volts, $I_b = 5.9$ ma.
(c) $E_c = -4$ volts, $I_b = 6$ ma.

11 · The Tube as an Amplifier

The most important property of a triode (and of the more complex tetrode and pentode) is its ability to amplify voltages or currents or power. Small a-c voltages impressed on the grid may produce higher a-c voltages in the plate circuit. The same may be true of current or power. How is this possible?

11.1. Basic triode circuit. The basic * amplifier circuit is shown in Fig. 1. The cathode-grid and cathode-plate direct volt-

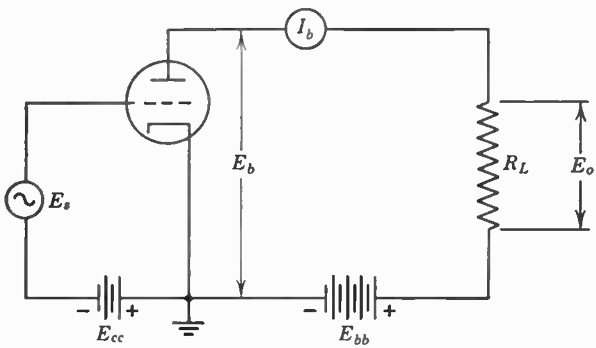


FIG. 1. Basic amplifier circuit with resistance load R_L.

ages are fixed by the batteries E_{cc} and E_{bb}, respectively. It should be noted that, while the d-c cathode-plate voltage or, more simply, the "plate voltage" (since the cathode is always taken as reference) is *fixed* by E_{bb}, the *actual* voltage is E_b because of the voltage drop across R_L. The resistor in the plate circuit, R_L, is called the **load resistor.** The input voltage E_s is an

* There are several modifications of this circuit, some of which will be discussed in later sections. However, an analysis based on the circuit of Fig. 1 may be extended to practically all cases.

alternating voltage. The a-c output voltage is E_o. If E_s is zero, no alternating voltage will appear across the load resistor because there will be no alternating current through the load.

When, however, alternating voltage is applied to the grid, the changing potential of the grid allows more and then less plate current to flow; this constitutes an alternating component of current and an alternating voltage appears across R_L. If operating conditions are properly chosen, the value of the alternating voltage across the load will be greater than the alternating voltage applied to the grid. The tube is then amplifying.

Consider, for the moment, that the alternating grid voltage E_s is zero but that the bias voltage E_{cc} is varied slowly above and below its normal value. For each value of grid voltage there will be a corresponding plate current, and by plotting these values a curve like that in Fig. 2 is obtained. The same result would be obtained if the value of E_{cc} remained fixed and an alternating voltage E_s was supplied. Note that the tube has both direct and alternating currents in the output circuit. The grid current is practically zero since the grid is maintained at a negative potential with respect to the cathode and will not collect any substantial number of electrons. At some instants the grid voltage is more negative than the bias voltage; at other instants it is less negative. The alternating grid voltage and the alternating portion of the plate current rise and fall in step, more current flowing when the grid is less negative and less current flowing when the grid is more negative. The alternating grid voltage and alternating plate current are in phase when the load is a resistance.

The plate current can be analyzed in two ways: (1) it is made up of a pulsating current, increasing and decreasing with respect to an average value in unison with the grid voltage changes; (2) it is made up of two component currents, (*a*) **a d-c component** whose value is fixed by the load resistor and the d-c grid and plate voltages, and (*b*) an **a-c component** whose peak and rms values are determined by the circuit conditions.

In Fig. 2 the d-c plate current is 6.6 ma, whereas the a-c component is 1.4×0.707 or 1.0 ma (rms). If the frequency of the grid signal is 60 cycles, the a-c component of plate current will also have a frequency of 60 cycles.

The point P on the curve is called the **operating point.** It is determined by the fixed values of grid-bias and plate voltage and by the size of the load resistance.

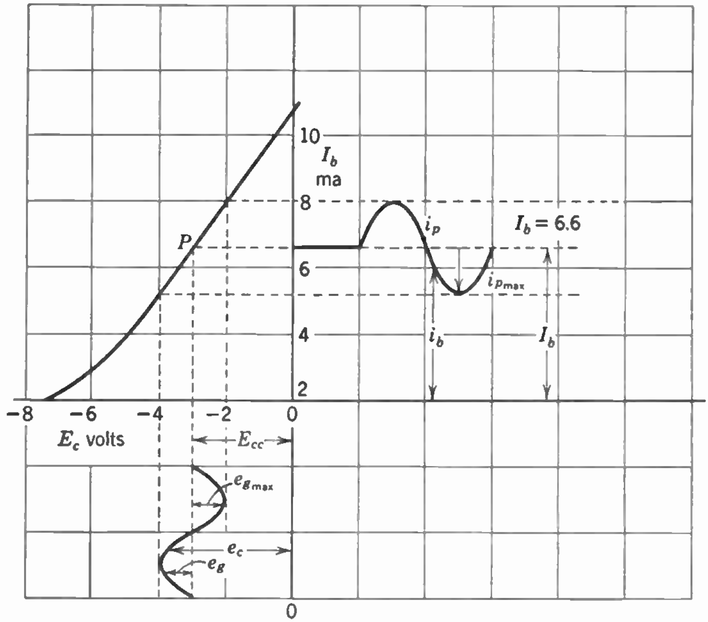


FIG. 2. Alternating voltages superimposed on the steady (bias) grid voltage produce a pulsating plate current which rises and falls from its steady (d-c) value.

11.2. Choice of operating point. Two conditions determine the placement of the operating point on the tube characteristic curve*: (1) the amount of power the tube can safely dissipate, and (2) the amount of distortion of the input waveform that can be tolerated in the plate circuit. The power dissipated in the tube is the product of the plate voltage and plate current ($P_p = E_b \times I_b$). The distortion factor will be considered later.

11.3. References for tube voltages. All tube voltages are measured with reference to the cathode; that is, the grid bias is the direct voltage which exists between cathode and grid, and the plate voltage is the direct voltage between cathode and plate.

* The *dynamic* curve. See Sect. 11.6.

The plate voltage is less than the plate supply voltage (E_{bb}) by the amount of voltage drop in the load resistor. With filament-type tubes operated from direct current the most negative end of the filament is considered the reference point.*

The low-potential ends of the grid and plate circuits are connected to the cathode of an indirectly heated tube, to the negative terminal of a filament-type tube when operated from direct current, and to the center tap of the filament transformer winding or to the center of a relatively low resistance across the filament when operated from alternating current. These low-potential leads are called the "common return" of the grid and plate circuits. Electrons flowing in the plate circuit return to the cathode through this lead.

11.4. Nomenclature for triodes. Several letter symbols will appear regularly in connection with triode circuits. These will be used consistently to refer to particular kinds or components of voltages and currents. The most important of these symbols follow:

Grid and signal voltages

- E_{cc} grid-bias battery direct voltage.
- E_c total cathode-grid direct voltage.
- e_c total instantaneous cathode-grid voltage $= E_c + e_g$.
- e_g instantaneous value of a-c component of cathode-grid voltage.
- E_g rms value of e_g (assumed sinusoidal).
- e_s instantaneous value of a-c grid circuit (signal) voltage.
- E_s rms value of e_s (assumed sinusoidal).

NOTE. E_{cc} and E_c are often, but not always, the same. Likewise, E_s and E_g.

Plate and output voltages

- E_{bb} plate circuit d-c supply voltage.
- E_b total cathode-plate direct voltage.
- e_b instantaneous total cathode-plate voltage $= E_b + e_p$.
- e_p instantaneous value of a-c component of cathode-plate voltage.
- E_p rms value of e_p.
- e_o instantaneous value of a-c component of load voltage.
- E_o rms value of e_o.

* The grid bias is the voltage from grid to cathode; this is not the same as the voltage from grid to ground in all cases.

Plate currents

- i_b instantaneous total plate current $= I_b + i_p$.
- I_b average or d-c value of plate current.
- i_p instantaneous value of a-c component of plate current.
- I_p rms value of i_p.

Figure 3 shows where the various voltages and currents appear in an amplifier. Instantaneous, direct, and rms values are all shown on the figure.

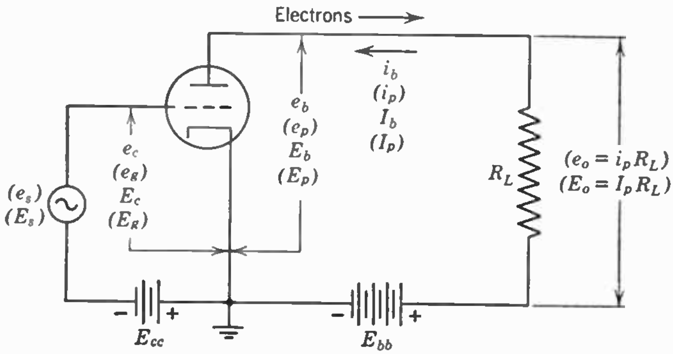


FIG. 3. Unbracketed lower-case letters i_p, e_c, etc., indicate instantaneous total values; bracketed lower-case letters (i_p), (e_g), etc., indicate instantaneous value of a-c component; unbracketed capital letters, E_b, E_c, etc., indicate d-c or average values; and bracketed capital letters (E_p), (I_p), etc., indicate rms a-c values. Brackets are not used in general practice.

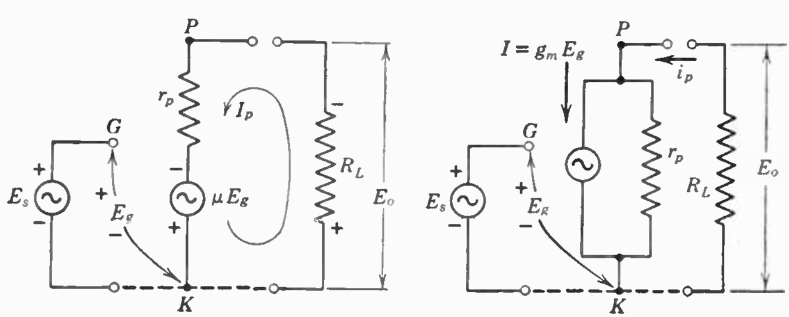
11.5. Amplifier with resistance load. The amplifier circuit of Fig. 1 will reproduce faithfully the waveform of the input signal in the output voltage and give voltage amplification only if certain conditions are met. The most important condition is that the bias voltage E_{cc} and the signal voltage E_s must be adjusted so that the alternating plate current has values along the straight-line portion of the tube characteristic. Also important is the fact that distortion caused by a slight curvature of the characteristic may be almost eliminated if the load resistance R_L is made much greater than the dynamic plate resistance r_p of the tube.

Consider again the circuit in Fig. 1, this time rather critically. The d-c plate voltage E_b, which is actually across the tube, is not the voltage E_{bb} of the B battery. It is less than this value by the voltage drop $(I_b \times R_L)$ in the resistor R_L. The direct voltage actually at the plate, then, is $E_b = E_{bb} - I_b R_L$. If, for example, the B battery voltage is 180 volts, R_L is 100,000 ohms, and I_b is 0.5 ma, then $I_b \times R_L$ equals $0.0005 \times 100{,}000$ or 50 volts, and E_b equals 180 — 50 or 130 volts.

It can be seen that any variation in I_b (the *variation* would be called i_p) causes a variation in the voltage drop across R_L and therefore a variation in the plate voltage E_b (the *variation* in the plate voltage would be called e_p, and the total plate voltage e_b is equal to $E_b + e_p$). Thus, any change in the grid voltage produces a change in the plate voltage.

It is not convenient to use the static characteristic curves (E_b–I_b, E_b–E_c, and E_c–I_b) discussed in Chapter 10 for the direct determination of a-c plate current and output voltage because of the effect of the load resistor and the voltage drop across it. To get around this difficulty a load line is used. The load-line method is most suitable when a large signal is applied to the grid as in power amplifiers, and so the discussion of this method will be delayed until later in the chapter. For amplifiers in which small grid signals are employed, as is the case in most voltage amplifiers, an equivalent circuit for the tube is most useful.

11.6. Equivalent circuits for tubes. Since a change in the alternating plate voltage in a triode amplifier is caused by a (usually) smaller change in grid voltage, it is possible to draw a



(*a*) Series Equivalent (*b*) Parallel Equivalent

Fig. 4. A-c equivalent circuits of tubes. D-c elements are not shown. Dashed lines indicate connections for amplifier shown in Fig. 1.

series-equivalent circuit as in Fig. 4(*a*). The voltage in the plate circuit is replaced by the fictitious, but equivalent, generator μE_g. The plate load is between the plate (P) and cathode (K); the input voltage E_s is between the grid (G) and cathode

(K). This is an *a-c equivalent circuit,* so the d-c voltages and currents are not shown. The dashed lines indicate the connection of the equivalent tube circuit for the amplifier of Fig. 1. The use of the equivalent circuit of the tube itself is not restricted to this particular amplifier. In any case, the batteries and other d-c elements are not shown, and the equivalent tube is connected to the same points in the circuit as was the actual tube.

There are two restrictions on the use of the equivalent tube: (1) the grid signal and bias must be such that operation is over a linear portion of the E_c–I_b curve, and (2) the grid must not go positive during any portion of its cycle, that is, no grid current can flow. The circuit as shown assumes a sinusoidal input voltage. This is not necessary but is the usual assumption in the analysis of circuits.

The equivalent circuit for tetrodes and pentodes is the same as for a triode except that the appropriate values of μ, g_m, and r_p must be used. The same general restrictions apply.

The polarity marks on the circuit of Fig. 4(a) are a-c polarity marks and indicate only that at a particular instant the voltage of E_s, for example, is assumed to be positive in the direction indicated, and that the other polarities at that instant are assumed to be as shown. It is important to note that *the instantaneous polarity* of E_s *and* μE_g *must be opposite;* that is, if the upper terminal of E_s is positive, then the upper terminal of μE_g must be negative. Then the proper phase relations will be maintained. This is necessary because there is 180° phase shift through the tube in the circuit.

A parallel-equivalent circuit is shown in Fig. 4(b). In this circuit the equivalent generator is a *constant-current* generator in parallel with the dynamic plate resistance. The series and parallel circuits are equivalent in all ways, and the one is used which makes the analysis of a particular circuit easiest.

11.7. Voltage amplification. The most direct application of the circuit of Fig. 4(a) is in the calculation of a-c plate current and of the voltage amplification. The a-c plate current is the voltage of the equivalent generator divided by the total resistance of the circuit. That is

$$I_p = \frac{\mu E_g}{r_p + R_L} \qquad [1]$$

The voltage across the resistor is $I_p \times R_L$, and its instantaneous a-c polarity is such that the bottom end of the resistor is positive. The **voltage amplification** or **gain** is the output voltage divided by the signal voltage, or

A_v (voltage amplification or gain)

$$= \frac{E_o}{E_s} = -\frac{I_p \times R_L}{E_s}$$

$$= -\frac{\mu E_g R_L}{E_s(r_p + R_L)} = -\frac{\mu R_L}{r_p + R_L} \quad (E_g = E_s, \text{ in this case}) \quad [2]$$

The minus sign is introduced in the expression for gain since the output voltage is 180° out of phase with the input voltage. It represents the *phase shift* through the tube. The presence of this phase shift is of considerable importance as will be seen in some of the later sections.

The same results as above may be obtained by use of the parallel-equivalent circuit of Fig. 4(*b*). The total impedance of the resistors r_p and R_L in parallel is

$$R_t = \frac{r_p \times R_L}{r_p + R_L}$$

The constant current $g_m E_g$ flowing through this total resistance produces a voltage equal to $g_m E_g R_t$, and this is the output voltage. The gain, then, is

$$A_v = -\frac{g_m E_g R_t}{E_s}$$

$$= -\frac{g_m E_g}{E_s} \times \frac{r_p \times R_L}{r_p + R_L}$$

but $\mu = g_m r_p$ and $E_s = E_g$ in this circuit, so

$$A_v = -\frac{g_m r_p R_L}{r_p + R_L} = -\frac{\mu R_L}{r_p + R_L} \quad [3]$$

which is the same result as was secured with the series-equivalent circuit.

The equation for gain shows that the maximum voltage amplification occurs when R_L is much greater than r_p, and that the *max-*

imum value that the gain can attain is equal to the amplification factor of the tube.

The normal gain realized in a triode circuit is perhaps 75 to 80 per cent of the amplification factor. If the size of R_L is increased while holding the size of the plate supply voltage E_{bb} constant, the amplification factor of the tube tends to decrease and the plate resistance increases so that the gain does not go up as rapidly as expected. If, on the other hand, the plate supply voltage is increased so that the d-c plate-cathode voltage of the tube is held constant so as to keep μ and r_p constant, a point is reached where the supply voltage has to be so high as to be impractical.

11.8. Power output. The power output or power in the load resistance can be calculated as follows:

$$P = I_p^2 R_L$$

$$I_p = \frac{\mu E_g}{r_p + R_L}$$

then

$$P = \frac{(\mu E_g)^2 R_L}{(r_p + R_L)^2}$$

The power output is maximum when the two resistances are equal, that is, when

$$r_p = R_L$$

Then the power

$$P = \frac{\mu^2 E_g^2}{4 r_p^2} \quad (r_p = R_L)$$

This power, which is fed to the load resistance, must come from the plate supply battery because the tube itself generates no power—it merely acts as a valve to take small voltages at the input and convert them to larger voltages and to a-c power in the load resistance. The power in the input circuit must come from the circuit to which it is attached.* The tube therefore releases

* Even though the grid terminal is shown open-circuited in the equivalent circuit it does not actually have infinite resistance and therefore requires a small amount of power. The power required is very small as long as no grid current flows.

power from the batteries in a form which is a replica of the power utilized in the input circuit to which the tube is attached as long as the bias and signal voltages are properly chosen. The tube itself consumes power from the plate supply battery. This power heats the plate and is wasted. It is equal to $E_b \times I_b$. Note that the power output is proportional to the *square* of the input voltage. Doubling the input voltage E_g increases the power output four times but, since the d-c plate current is constant, does not increase the heat loss in the tube. Actually, the amount of power lost in the tube decreases as the signal is increased. This is discussed in Sect. 11.16.

11.9. Power amplification. Because of the high grid resistance (grid resistance would be infinite except for a very small current in the insulation between grid and cathode pins) the power amplification of the circuit in Fig. 1 is very high as long as no grid current flows. Often there is a high-valued resistor in parallel with the grid. Then the power in this resistor represents substantially all the grid input power, and the power amplification is this a-c power divided into the a-c output power.

11.10. Phase of E_g, E_p, and I_p. When the grid voltage in Fig. 1 increases in a positive direction, the plate current increases. When this happens a greater voltage drop occurs across the load resistance and *less* voltage appears from plate to cathode of the tube. Thus the plate voltage decreases when the grid voltage increases in a positive direction. These two voltages, E_g and E_p, are 180° out of phase for a resistance load.

If a sine wave which starts from zero and rises in a positive direction is applied to the grid of an amplifier with a resistive load, a sine wave of voltage which starts at zero and drops in a negative direction appears across the plate load. A negative grid pulse, on the contrary, produces a positive pulse in the plate circuit. These phase relations are shown in Fig. 5.

Example 1. A tube is connected in a circuit as shown in Fig. 1 except that a 500,000-ohm resistor is connected in parallel with the cathode-grid circuit. The amplification factor of the tube is 8, and its dynamic plate resistance r_p is 1000 ohms. The plate load resistance R_L is 2000 ohms. If the grid-signal voltage is 20 volts (rms), what is the a-c plate current, the a-c output voltage, voltage amplification, and power amplification?

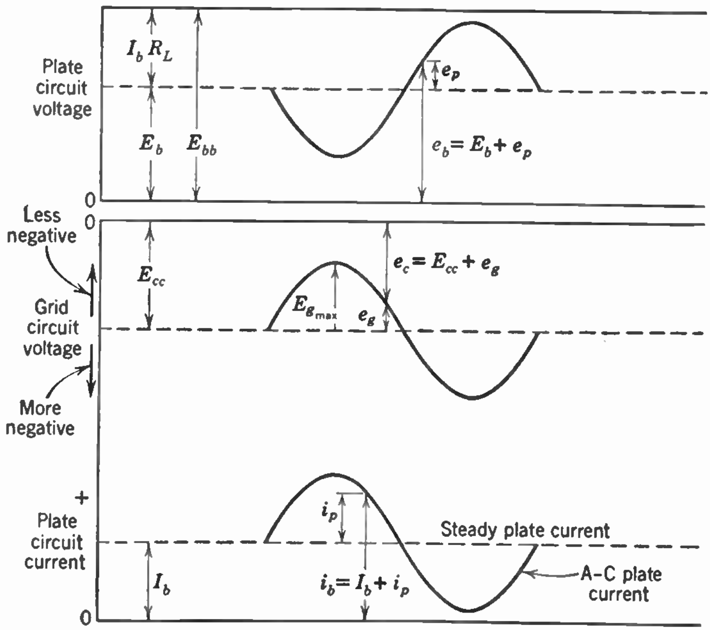


FIG. 5. Phase of grid, plate, and load voltages. As e_g becomes more negative, e_p becomes less negative. The load voltage, e_L, also differs from the grid voltage by 180°.

Solution:

$$I_p = \frac{\mu E_g}{r_p + R_L} = \frac{8 \times 20}{1000 + 2000} = \frac{160}{3000} = 53.3 \text{ ma}$$

$$E_o = I_p R_L = 53.3 \times 10^{-3} \times 2000 = 106.6 \text{ volts}$$

$$\text{Voltage amplification} = \frac{E_o}{E_s} = -\frac{106.6}{20} = -5.33$$

$$\text{Grid power} = \frac{E_g{}^2}{500{,}000} = \frac{20^2}{500{,}000} = \frac{400}{500{,}000} = 0.0008 \text{ watt}$$

$$\text{Output power} = I_p{}^2 \times R_L = (53.3 \times 10^{-3})^2 \times 2000$$

$$= 5.7 \text{ watts}$$

$$\text{Power amplification} = \frac{\text{A-c output power}}{\text{A-c grid power}}$$

$$= \frac{5.7}{0.0008} = 7120 \text{ times}$$

Problem 1. A sinusoidal grid voltage of 5 volts (peak) is applied to a tube which has a load resistance of 10,000 ohms. The μ of the tube is 8; the plate resistance is 10,000 ohms. What are the peak and rms values of the a-c component of plate current? What a-c power is developed in the load resistance? What is the rms output voltage?

Problem 2. What is the voltage amplification in Prob. 1? What will be the voltage amplification and power output if the load resistance is changed to 50,000 ohms?

Problem 3. A 6J5 tube is to be operated with $E_b = 150$ volts (Fig. 9, Chapter 10). Estimate the value of grid-bias voltage which will place the operating point in the middle of the linear range of the E_c–I_b characteristic.

Problem 4. A type 6SF5 high-mu triode has a plate resistance of 66,000 ohms and an amplification factor of 100 under certain operating conditions. What will be the voltage amplification for a load resistance of 200,000 ohms?

Problem 5. A 2A3 power triode has a plate resistance of 800 ohms and an amplification factor of 4.2 at the operating point $E_b = 250$ volts, $E_c = -45$ volts. Plot a curve of a-c power output against grid voltage as the rms grid voltage is varied from 0 to 30 volts in 5-volt steps. The load resistance is 2200 ohms.

Problem 6. What is the largest rms grid-signal voltage that can be applied to the tube in Prob. 5 without driving the grid positive?

Problem 7. The power output of a tube when worked into a load resistance equal to the plate resistance of the tube is $(\mu^2/4r_p) \times E_g^2$. Dividing this expression by E_g^2 gives $\mu^2/4r_p$, which is a "figure of merit" that gives power output per (volt input)^2. Make a table of such values for representative tubes used at the present time, getting the data from a tube manual.

Problem 8. The 6C4 is a high-frequency power triode of the miniature type. When $E_b = 250$ volts and $E_c = -8.5$ volts, the d-c plate current is 10.5 ma, the amplification factor is 17, and the plate resistance is 7700 ohms. A 25,000-ohm load resistor is to be used with the tube in an amplifier circuit. What must be the value of the d-c supply voltage E_{bb} if the tube is to operate under the given conditions? What is the grid-plate transconductance?

Problem 9. A certain triode has an amplification factor of 20 and r_p of 6000 ohms. The plate load resistor is replaced by a 20-henry choke with a Q of 10 at 200 cycles. What will be the output voltage E_o across this choke for a 200-cycle grid voltage E_s of 2 volts? Note. $E_o = I_p\sqrt{R_L^2 + X_L^2}$ and $I_p = \mu E_g/\sqrt{(r_p + R_L)^2 + X_L^2}$.

11.11. Distortion due to curved characteristic. The waveform of the plate current and output voltage of an amplifier may be different from that of the grid-signal voltage because of operation on a non-linear portion of the characteristic curve. This kind of distortion is called **non-linear** or **amplitude** distortion and is caused by an incorrect grid bias, by a too large grid signal, or both.

Figure 6(a) shows a case in which the bias is too negative so that the plate current is cut off (reduced to zero) in a portion of the negative cycle. Even in the portion of negative cycle in which current flows, its waveform is rounded as a result of the curvature near the bottom of the E_c–I_b characteristic. If the signal voltage is reduced, the current can flow the entire cycle but there will still be some waveform distortion as a result of the

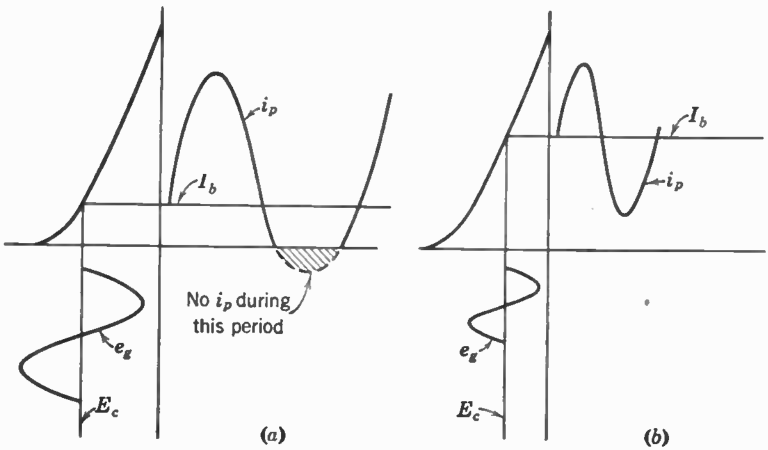


Fig. 6. Wrong bias (a) produces output unlike the input voltage waveform. In (b) the bias is correct.

curved characteristic. The correct value of bias voltage is shown in Fig. 6(b). Note that it is approximately half way between zero and the value for plate-current cut-off. Even in this case, distortion would occur if the value of the signal voltage was increased to such a value that the instantaneous grid voltage went more negative than the cut-off value.

When a part of the wave is cut off as in Fig. 6(a), the instantaneous current rises more in the first half cycle than it drops in the second half cycle and the d-c or average value of the current is increased. This represents an increase in the direct current taken from the plate supply battery, and a d-c meter in the plate circuit will show this increase. If the a-c signal voltage in Fig. 6(a) is zero, the d-c meter will indicate the value I_b. As the signal voltage is increased it will be noted that the value of direct current increases. This is a sure sign of distortion.

Distortion of the waveform may occur if the bias is such that the grid voltage goes positive and grid current flows. The amount of distortion from this cause depends to a large extent upon the equivalent impedance of the source connected to the grid. If the impedance is low, this source can furnish the current required by the grid with little drop in voltage across the source impedance. However, if the source impedance is high, its voltage drops during the interval the grid is taking current and the full voltage is not supplied to the grid. The appearance of the wave is much the same as that in Fig. 6(*a*) except that the flattened portion is on the top.

The correct design and operation of an amplifier, then, require that a number of conditions be met. Not only must the grid bias be the proper value, but the signal voltage must not be so large that the grid is instantaneously driven into a curved portion of the characteristic or beyond cut-off. If the grid is driven positive the source of grid power must be able to furnish that power without a decrease in voltage. And for maximum freedom from distortion, the load resistor must be correctly related to the tube resistance r_p.

11.12. Voltage vs. power amplifiers. So far the terms "voltage amplifier" and "power amplifier" have been used without definition. Actually there can be no hard and fast definition that will distinguish between the two classes of amplifiers. Both amplify the power that is used to drive the grid. Both usually also amplify the grid voltage. What, then, is the difference between a voltage and a power amplifier?

A voltage amplifier is operated so that the maximum amplification of voltage is achieved and with very little distortion. The plate load resistance is usually considerably higher than the plate resistance of the tube. The grid-signal voltage is normally quite small. A power amplifier, on the other hand, is operated to furnish a large amount of output power. To do this the plate load resistance is comparable in value to the tube resistance, and the grid-signal voltage is high compared to that in a voltage amplifier. The grid of a power amplifier is often driven so that there is an appreciable amount of distortion. The amount of allowable distortion is determined by how much can be tolerated for a particular application.

A difference also exists between tubes designed primarily for voltage and for power amplification. A voltage-amplifier triode usually has an amplification factor of 15 or greater and a relatively high plate resistance. The physical size of the anode is rather small since the tube does not have to dissipate much heat. A power-amplifier tube generally has an amplification factor below 10 and the plate resistance is rather low. The physical size of the anode is fairly large since the tube must dissipate considerable power as heat.

11.13. Load lines. How much d-c plate current flows in an amplifier when a certain size load resistance is used? This question would be easy to answer if the resistance (static, in this case) of the tube were constant. But the plate resistance depends upon the amount of plate voltage actually across the tube, even though the grid voltage is fixed. The amount of voltage across the tube cannot be determined until the plate current is known. Since the plate resistance and the voltage across the tube are unknown, and they depend upon each other, the plate current cannot be determined by methods already described.

One possibility for finding the plate current and for determining the operation of the tube when an a-c voltage is applied to the tube would be to draw a dynamic set of characteristics for the particular load resistance. This would mean that each value of load resistance would require a separate set of characteristic curves. Such a scheme, although it would work, would require so many separate characteristics that it would be very unwieldy. A method based on the static characteristics would be much more useful. Such a method is the one using **load lines.**

All that is required for the use of the load-line method is a family of E_b–I_b curves. The value of the load resistance and the value of the plate supply voltage or the operating point must be known. The load line gives the division of voltage across the load and tube for all values of plate current produced by various grid voltages. There are several ways of plotting this load line, but the simplest is as follows.

The total plate supply voltage is divided between that across the load resistance and that across the internal resistance of the tube. If the tube resistance is zero, then all the supply voltage E_{bb} appears across R_L. On the other hand, if the R_L is zero all the voltage appears across the tube. In the actual case part of the voltage appears across each. The load line is a method of subtracting, graphically, the voltage across the load resistor R_L from the supply voltage E_{bb} to give the voltage across the tube. When the voltage across the tube is known, the plate current can be read from the characteristic curves.

To construct a load line, start with a known value of voltage

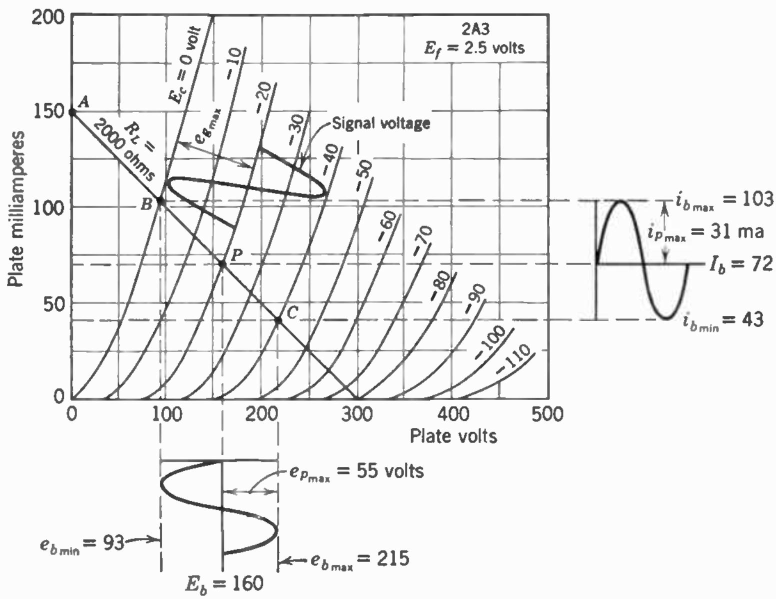


Fig. 7. The straight line AC is the load line. The value of plate current for various values of grid voltage is found along this line.

such as $E_{bb} = 300$ volts in Fig. 7. Assume a load resistance R_L of 2000 ohms. Divide the supply voltage by the load resistance ($E_{bb} \div R_L = 300 \div 2000 = 150 \times 10^{-3}$ amp or 150 ma) to determine the point A where the load line crosses the I_b-axis.* Then draw a straight line from E_{bb} to point A. Since this load line is specifically for a 2000-ohm load, the point of operation in this particular case must be somewhere along this line. If the bias voltage is fixed at −20 volts, the operation must also be on this bias line and the intersection of the bias and load lines

* This gives the current that would flow if the tube were shorted and R_L were the only resistance in the circuit.

is the operating point P. The plate current I_b is seen to be 72 ma.

Suppose that a 20-volt (peak) signal is applied to the grid. When the a-c grid voltage is at the positive peak of its cycle, the total grid voltage is zero (point B) and the plate current can be read from the graph. Likewise, when the grid voltage is at its negative peak, the total voltage is -40 volts and the corresponding plate current can also be read from the graph at point C. Values of plate current for any time on the grid-voltage cycle can be determined in a similar manner. The values of both plate current and plate voltage can be traced out for the entire cycle as shown in Fig. 7.

Suppose, however, that instead of the plate supply voltage E_{bb} being known, the d-c plate voltage E_b and bias voltage E_{cc} are specified. The construction of the load lines proceeds in much the same manner as before. A convenient value of E_{bb}, preferably one that is easily divisible by R_L, is assumed and a temporary load line drawn as before. The point where E_b and I_b intersect is marked and a line parallel to the temporary load line is drawn through this point. The proceduce for determining the plate current for various values of alternating grid voltage is then the same as before.

The *slope* of a load line does not depend upon the choice of E_{bb}, E_b, or E_{cc}. It is determined solely by the load resistance and the characteristic curves of the tube. The *placement* of the load line, however, is determined by the choice of operating conditions. This means that the slope of the load line may be determined and a temporary load line drawn. The actual load line may then be drawn so that it goes through the chosen operating point.

The load lines discussed above are **d-c load lines.** When the plate circuit contains a load with an a-c impedance which is not equal to the d-c impedance, another load line, called the **a-c load line,** must be drawn. Such load lines will be discussed in connection with transformer-coupled amplifiers.

11.14. Power and gain calculations from load lines. The method for calculating output power by the use of load lines can be illustrated by reference to Fig. 7. The total plate current i_b is shown and has a maximum value of 103 ma and a minimum value of 43 ma. The difference, $103 - 43 = 60$ ma, is the peak-to-peak value, and the maximum or "single peak" value is 30 ma.* The plate voltage e_b has a maximum value of 215 volts and a minimum value of 93 volts, a peak-to-peak value of 122 volts, and a peak value of 61 volts. The power output can be calculated from

$$P = \tfrac{1}{8}[(e_{b_{max}} - e_{b_{min}}) \times (i_{b_{max}} - i_{b_{min}})]$$

In this case

$$P = \tfrac{1}{8}[(215 - 93)(0.103 - 0.043)]$$
$$= \tfrac{1}{8} \times 122 \times 0.060 = 0.916 \text{ watt}$$

The voltage amplification can also be calculated from the load line. The formula is

$$A_v = -\frac{(e_{b_{max}} - e_{b_{min}})}{2e_{g_{max}}}$$

In Fig. 7 the amplification is

$$A_v = -\frac{215 - 93}{2 \times 20} = -\frac{122}{40} = -3.05$$

11.15. Distortion calculations. The distortion caused by non-linear tube characteristics is called **non-linear** or **amplitude** distortion. It has also been called harmonic distortion. The effect of this kind of distortion is to produce new frequency components in the output voltage that were not in the grid-signal voltage. These new frequencies are harmonically related. That is, they are always some multiple of the frequency of the grid signal. If the signal frequency is 1000 cycles, the second harmonic † is 2000 cycles, the third is 3000 cycles, etc. In a triode only the second harmonic is important. All higher harmonics are usually so small as to be negligible.

The production of new frequencies as described above is a real thing. These harmonic frequency components of voltage

* Half of the peak-to-peak value is used rather than the value shown as $i_{p_{max}}$ in the diagram. This method eliminates most of the error in calculation due to distortion of the waveform.

† The term "first harmonic" is not used. It is the basic frequency of the wave, the grid signal in this example, and is called the **fundamental** frequency.

(or current) can be separated with appropriate apparatus and their values measured. Distortion is purposefully introduced in some circuits to produce an output rich in harmonics, but harmonics are undesirable in the usual amplifier.

The second harmonic in a triode may be calculated from

$$\text{Second harmonic} = \frac{\frac{1}{2}(i_{b\text{max}} + i_{b\text{min}}) - I_b}{(i_{b\text{max}} - i_{b\text{min}})} \times 100\%$$

From Fig. 7, the second harmonic distortion is found to be

$$\%\text{ second} = \frac{\frac{1}{2}(103 + 43) - 72}{103 - 43} \times 100$$

$$= \frac{(\frac{1}{2} \times 146) - 72}{60} \times 100 = \frac{73 - 72}{60} \times 100$$

$$= \tfrac{100}{60} = 1.67\%$$

This means that, for every signal put on the grid having a peak value of 20 volts, there will be in the output a second harmonic of this input signal frequency equal to 1.67 per cent of the value of the fundamental. Since it is generally considered that 5 per cent distortion is acceptable to the average listener, the distortion calculated above is well within good design limits. It would, of course, be larger if the grid were driven harder.

In calculations during the design of an amplifier it is often found that the first calculation gives too high a value of distortion. Possible methods of decreasing the distortion are to reduce the value of the grid-signal voltage, to increase the plate voltage, to increase the plate load resistance, to change the bias voltage, or to use combinations of these possibilities.

In view of the foregoing discussion it is apparent that any amplifier designed to give an appreciable power output will have some distortion. The problem then arises: How can maximum power be obtained and distortion still be kept within prescribed limits? The condition for maximum power output ($R_L = r_p$) is not the solution, since this value of load resistance almost always gives more distortion than can be tolerated. Both power output and second harmonic distortion usually decrease as the load resistance is made larger, other operating conditions being held the same. Fortunately, the distortion generally decreases more rapidly than the power output, and so an amplifier with reasonably low distortion can be designed by increasing the load resistance to perhaps two or three times the tube resistance.

Most power amplifiers utilize a transformer rather than a resistor in the plate circuit. Curves of power output and second harmonic distortion plotted against effective load resistance are shown in a later section on transformer-coupled amplifiers.

11.16. Power diagrams. A power diagram of a tube with a resistance load will serve as a good review of some of the material already discussed and will introduce some new ideas.

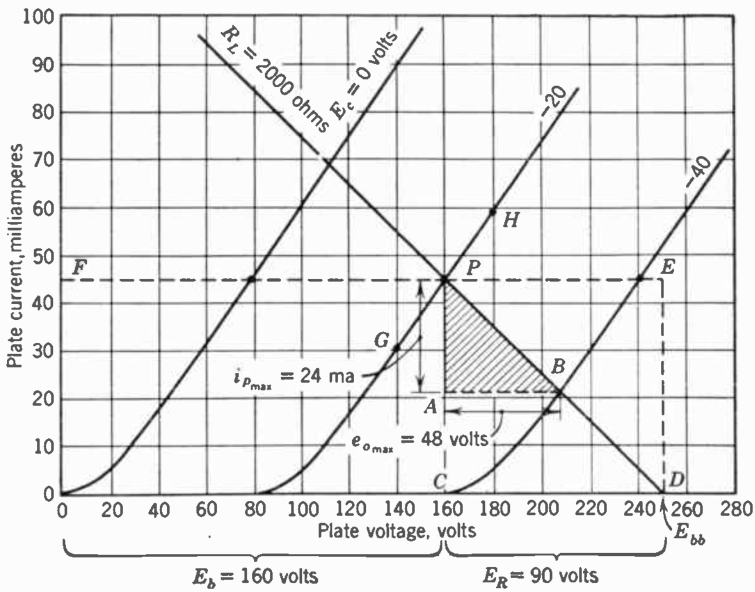


FIG. 8. Power diagram of an amplifier with a resistance load of 2000 ohms.

Consider the E_b–I_b curves in Fig. 8, which is a typical family of triode curves although it does not represent any particular tube now available. The diagram was drawn under the assumption that at the operating point P the plate voltage E_b was 160 volts when the plate current I_b was 45 ma, and that the load resistance R_L was 2000 ohms. The load line is drawn through the operating point and is found to cross the E_b-axis at 250 volts. This is the value of the plate supply voltage E_{bb} which must be furnished. The grid-bias voltage E_{cc}, from the curve, is −20 volts. A sinusoidal grid signal of 20 volts (peak) is assumed. Then the instantaneous grid voltage varies between 0 and -40 volts, and the corresponding plate currents are 69 and 21 ma. With no a-c grid voltage (zero input) the voltage across the tube is 160 volts, that across the load is 90 volts, and the plate current is 45 ma.

The d-c power lost in the load is the product of the voltage across the load resistance and the current through it. The d-c power lost in the tube, which must be dissipated by the plate of the tube, is the product of the voltage across the tube and the current flowing. Thus,

$$\text{D-c power in load} = E_R \times I_b = \text{Area } CPED = CD \times CP$$

$$\text{D-c power in tube} = E_b \times I_b = \text{Area } OFPC = OC \times OF$$

and the total power supplied by the battery must be the sum of these powers, that is,

$$\begin{aligned}\text{Total power from B battery} &= E_{bb} \times I_b \\ &= \text{Area } OFED = OD \times OF\end{aligned}$$

When a 20-volt (peak) grid signal is applied to the tube, the peak value of the alternating current through the load is given by the line AP, and the peak value of the a-c voltage across the load is given by the line AB. Then

$$e_{o\max} = \text{Peak a-c voltage across load} = AB$$

$$i_{p\max} = \text{Peak a-c current through load} = AP$$

Since the a-c power in a resistance circuit is the product of the rms current and voltage, the rms values of the a-c load voltage and current must be determined. They are

$$E_o = \frac{e_{o\max}}{\sqrt{2}} \quad \text{and} \quad I_p = \frac{i_{p\max}}{\sqrt{2}}$$

From these values the a-c power to the load may be obtained as

$$\begin{aligned}\text{Power output} = E_o \times I_p &= \frac{e_{o\max}}{\sqrt{2}} \times \frac{i_{p\max}}{\sqrt{2}} = \frac{e_{o\max} \times i_{p\max}}{2} \\ &= \frac{AB \times AP}{2} = \text{Area of triangle } APB\end{aligned}$$

This power is dissipated in the load in addition to the d-c power lost there, and the total power is the sum of the two components. The average current taken from the battery has not changed—and yet the load has an additional amount of power used up in it. Where does this power come from? Clearly it must come from the power dissipated by the plate of the tube. *When an a-c voltage is placed on the grid, then a-c power is developed in the load, less power is dissipated as heat at the tube plate, and the tube will actually run cooler when it is delivering a-c power to the load than when standing idle, that is, with no a-c input grid voltage.* This statement applies to a Class A amplifier (the type under discussion) with no distortion. It is not true for most other types of amplifiers.

The values of power can be determined from the graph in Fig. 8. They are

D-c power lost in load (no a-c grid voltage)
$= CD \times CP = 90 \times 45 \times 10^{-3} = 4.05$ watts
D-c power lost in tube (no a-c grid voltage)
$= OC \times OF = 160 \times 45 \times 10^{-3} = 7.2$ watts
D-c power from B battery
$= OD \times OF = 250 \times 45 \times 10^{-3} = 11.25$ watts
Maximum a-c voltage across load (maximum a-c grid voltage = 20 volts)
$= AB = 208 - 160 = 48$ volts
Maximum a-c current through load (maximum a-c grid voltage = 20 volts)
$= AP = 45 - 21 = 24$ ma
A-c power in load (maximum a-c grid voltage = 20 volts)
$= \frac{1}{2}AB \times AP = \frac{1}{2}\, 48 \times 24 \times 10^{-3} = 0.576$ watt
Net power lost in tube (maximum a-c grid voltage = 20 volts)
$= 7.2 - 0.576 = 6.62$ watts

The formula for the power output of a tube is

$$\text{Power output} = \frac{\mu^2 E_g{}^2 R_L}{(r_p + R_L)^2}$$

The above calculations can be checked by finding the value of μ and r_p from the curves in Fig. 8. The plate resistance is the reciprocal of the slope of the E_b–I_b line about the operating point, or the slope of the line GH. This is

$$r_p = \text{Slope of line } GH = \frac{(180 - 140)}{(59 - 31) \times 10^{-3}} = \frac{40}{28 \times 10^{-3}}$$

$$= 1430 \text{ ohms}$$

The amplification factor $\mu = -\Delta E_b/\Delta E_c$ at a constant value of I_b. Taking these increments along the line FE from the intersections with the 0 and -40 grid voltage lines, and reading the corresponding values of plate voltage, there results

$$\mu = -\frac{\Delta E_b}{\Delta E_c}(I_b \text{ constant}) = -\frac{(242 - 78)}{(-40 - 0)} = \frac{164}{40} = 4.1$$

Then the power output can be calculated as

$$\text{Power output} = \frac{(4.1)^2 \times (20 \times 0.707)^2 \times 2000}{(1430 + 2000)^2}$$

$$= 0.572 \text{ watt}$$

This value agrees very closely with the value obtained above. More difference than this can normally be expected when values of voltage and current are read from graphs.

From a set of the E_b–I_b curves the data may be obtained from which to design an amplifier and to calculate its power output, the losses in the various tubes, the percentage distortion, etc. The only other information about the tubes that is needed is their filament voltage and their maximum plate voltage, current, and plate dissipation ratings. This information is usually given in tube manuals or in data sheets supplied by tube manufacturers.

Problem 10. On a set of 2A3 characteristics show a 3000-ohm load line and the operating point P for the following operating conditions:

(*a*) $E_{bb} = 250$ volts, $E_c = -20$ volts.
(*b*) $E_b = 200$ volts, $I_b = 50$ ma.
(*c*) $E_b = 250$ volts, $E_c = -40$ volts.

Problem 11. A 2A3 power triode (Fig. 7) is operated under the following conditions: $E_{bb} = 300$ volts, $E_{cc} = E_c = -40$ volts, $R_L = 1500$ ohms, $e_{g_{max}} = 30$ volts. Construct the load line and calculate the power output, second harmonic distortion, and rms value of the output voltage.

Problem 12. A 6J5 triode (Fig. 10, Chapter 10) is operated at a plate supply voltage (E_{bb}) of 300 volts and a grid-bias voltage of -8 volts. The plate load resistor is 30,000 ohms. Compute the power output and second harmonic distortion if the a-c grid signal has a maximum value of 8 volts. Use graphical methods.

Problem 13. What is the rms value of the a-c component of plate current in Prob. 12? the rms value of the output voltage? the voltage amplification?

11.17. Screen-grid tubes. The grid and plate of a triode are two metallic conductors, more or less parallel to each other, and insulated from each other. In addition to their normal functions in the tube, already discussed, they also possess a capacitance. The total *grid-plate capacitance* is made up of the capacitances between the grid and plate themselves, between the connecting wires, and between the tube prongs. For most triodes this total capacitance ranges from 3 to 15 $\mu\mu$f. At audio frequencies, the reactance of this capacitance is so high that its effect can usually be neglected, but at higher frequencies the reactance becomes low enough that it may seriously affect the operation of the tube in a circuit. Alternating current will flow from a point of high a-c potential (plate) to a point of low potential (grid). The voltage produced by this current will add to or subtract from the grid voltage, depending upon the phase of the voltage; thus there will be "feedback" from output to input.

If the amplification of the tube and its circuit is 10, and if one-tenth of the output voltage is fed back to the grid, as much voltage is impressed on the grid from the output as was impressed by the grid driving circuit. This additional voltage is amplified in the tube, along with the externally applied voltage, and one-tenth of the new value of output voltage is again returned to the grid. Soon the amplifier will break into uncontrolled oscillation and the tube becomes worthless as an amplifier. The grid loses control so far as amplifying and reproducing the original signal are concerned.*

If, now, another grid is placed between the control grid and the plate and maintained at zero a-c potential by connecting a large condenser between it and ground as in Fig. 9, the capacitance between grid and plate is reduced materially. The feedback from plate circuit to grid circuit is reduced, and much greater amplification without instability is possible. The added grid, called the

* Special "neutralizing" circuits can be used to prevent triodes from breaking into oscillation even at high frequencies. These circuits are too complex and expensive to use in most radio receivers. They are used in transmitters and are discussed in Chapter 17.

screen grid, is maintained at a positive d-c potential and so tends to neutralize the space charge and accelerate the electrons.

Some of the electrons in a screen-grid tube strike the screen and form a screen current, but the greater proportion get through the mesh of this grid and go on to the plate. Since the screen current subtracts from the total possible plate current, the efficiency is impaired, but by proper design the effect of the control grid on the plate current is as great as if the screen were not present.

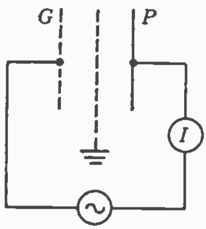


Fig. 9. A grounded shield between plate and grid reduces a-c current flow from plate to grid.

Not only does the screen protect the control grid from serious reaction from the output circuit but it produces a change in the tube characteristics as compared to those of a triode. Since the plate is shielded from the grid, it follows that it is shielded to some extent from the cathode. Plate voltage, therefore, is not so effective in controlling plate current as it is in a triode. On the other hand, the effect of the control grid is not impaired (except by those few electrons which are lost on the way to the plate). The ratio of the effectiveness of the grid voltage to plate voltage in controlling plate current is increased, and the amplification factor is increased several-fold. The amplification factor of screen-grid tubes ranges up to 400 or 500. The voltage gain usually obtainable from screen-grid tubes is about one-fourth of the amplification factor.

Since the plate resistance of the tube is the ratio of a change in plate voltage to the corresponding change in plate current, it follows that the plate resistance may become very large since large changes in plate voltage produce only small changes in plate current.

The screen-grid tube (tetrode), therefore, has high r_p (up to about 1 megohm), high μ, and average g_m.

In a normal amplifier application (Fig. 10) the screen is connected to a positive potential, usually somewhat lower than the d-c plate potential, and the a-c components of current in the screen circuit are by-passed to the cathode by a condenser connected close to the screen terminal. This by-pass condenser must be large enough so that its reactance is low at all frequencies of interest.

The plate characteristics of a typical screen-grid tube are shown in Fig. 11. For plate voltages higher than the screen voltage (90 volts in the diagram) a tetrode has rather flat curves.

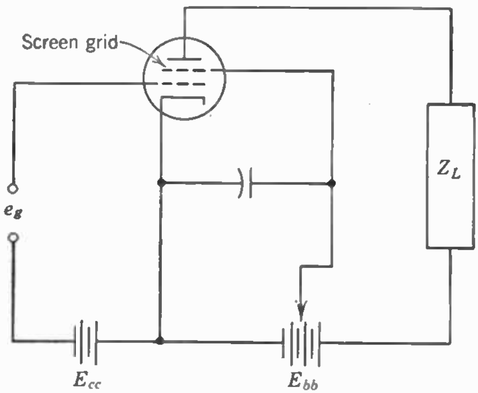


Fig. 10. Screen grid has a positive d-c potential with respect to cathode and is by-passed to cathode to provide low-impedance path for alternating currents.

The portion of the curves for plate voltages lower than the screen voltage is very irregular. Over a part of this portion of the curves

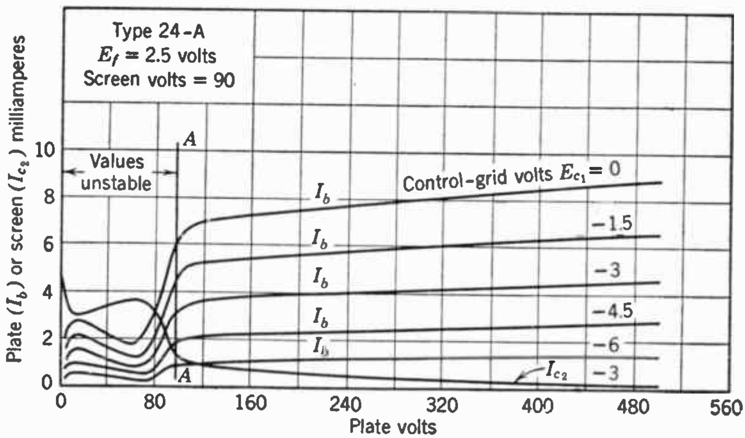


Fig. 11. Characteristic curves of typical screen-grid tube.

the current actually decreases as the plate voltage increases. How does this happen?

The screen in Fig. 11 is at 90 volts potential. An electron leaving the cathode is attracted by the screen. Since the mesh of the screen is rather coarse the electron most likely will miss the screen and go on to the plate. If the plate voltage is around 10 volts or so, the velocity of the electron as it strikes the plate may be enough to knock out another electron. This secondary electron has little energy or velocity and is attracted to the screen which, at the moment, is the most positive tube element. Under certain conditions the oncoming electron may knock off two or three secondary electrons. It is this phenomenon of **secondary emission** that results in the irregular curves and the dip of plate current in Fig. 11.*

If the plate voltage is raised higher than the screen voltage, the secondary electrons will be attracted back to the plate. This accounts for the leveling off of the curves for plate voltages above 90 volts.

The use of tetrodes is almost entirely limited to voltage amplification in which the grid signal is small. Large signals cause the plate-current excursions to include a part of the irregular portion of the curves and give unsatisfactory and unstable operation. Even for voltage amplification, the tetrode has been largely replaced by the pentode.

11.18. Pentode tubes. The secondary emission effects in a tetrode seriously limit its usefulness, yet its low grid-plate capacitance and its high amplification factor are distinct advantages in many applications. The question, then, is, How can the advantages be maintained, and the disadvantages eliminated? An answer to the question is found by inserting an additional widely spaced grid between the screen grid and the plate. This third grid is connected directly to the cathode and so is at a negative potential with respect to the plate. Now an electron may strike the plate and cause secondary electrons to be ejected, but these electrons are attracted back to the plate rather than to the screen grid. The third grid, because of its negative potential with respect to the plate, suppresses the effects of secondary emission and is called a **suppressor grid**. Compare the characteristics of the tetrode in Fig. 11 and the pentode in Fig. 12. Note that the

* Secondary electrons may be knocked off the plate of a triode, also, but they are attracted back to the plate since there is no other positive electrode, except in special circuits in which the control grid is driven highly positive.

curves for the pentode do not have the very irregular shapes found in the tetrode.

Pentodes are used both as voltage and as power amplifiers. A power pentode such as the 6F6 produces considerable power output with low grid excitation voltage; voltage amplifiers like the 6J7 produce high-voltage amplification compared to that of triodes.

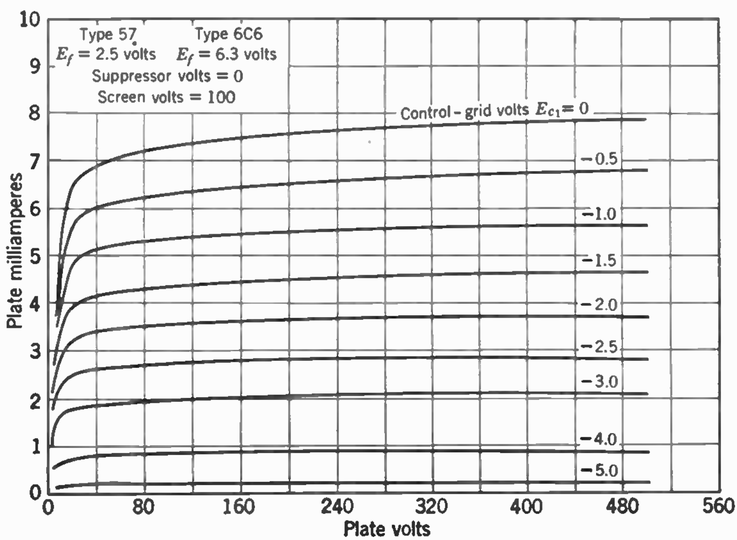


FIG. 12. Characteristics of a typical pentode tube used for voltage amplification.

11.19. Variable-mu tubes. Early receivers utilizing screen-grid tubes suffered from many troubles. Among others was the production of cross modulation (cross talk) by strong undesired signals. These tubes had such a short range of grid voltage over which they could work that a strong signal would force the grid voltage to the point where the alternating plate current would be cut off during a part of the cycle. This resulted in severe distortion, evidenced by blurbs or gasps of undesired modulation crashing through the desired program, or by hum entering from the power supply.

Late in 1930 a new screen-grid tube made its appearance and became quite important the following year. This was the

variable-mu or **super-control** tube. It had a very long, fairly linear characteristic at the bottom of the E_c–I_b curve. Strong negative voltages on the grid did not force the plate current to zero, and cross modulation was largely prevented. A comparison of super-control and regular characteristics is shown in Fig. 13, and the characteristics of a typical super-control tube are shown in Fig. 14.

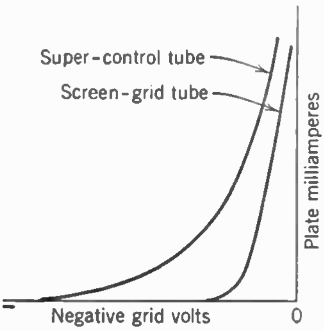


FIG. 13. Difference between variable-mu (super-control) tube and sharp cut-off tube. The former has a "remote" cut-off.

Because of the very long even characteristic, the super-control tube is nicely adapted for automatic volume control, for i-f and r-f amplifiers, and for recording signal strength or loud-speaker output, or other measurements, where large ranges of current, voltage, etc., are to be measured.

In this tube the grid is wound, not regularly or evenly along the length of the winding space, but with a wide spacing at one portion and close spacing at another. The E_c–I_b characteristic may be controlled in manufacture by varying the pitch of the winding or the manner in which the wide and close spacing is put on the grid-wire supports.

The super-control tube is also known as a **remote cut-off tube.**

11.20. Beam power tubes. Although the pentode is an advancement over the screen-grid tetrode, there is need for improvement. At the lower plate voltages there is still enough curvature to the characteristic to produce distortion unless the tube is seriously limited in its voltage or power output. It would be an advantage to straighten out the E_b—I_b curves. This is done in the beam power tubes in a most ingenious way.

Instead of placing an actual gridlike electrode within the tube structure, electrons are forced to travel in prescribed paths and with such density that a "virtual" low-voltage electron-repelling screen (the suppressor) is in the desired region. The mechanical details are shown in Fig. 15. The screen, therefore, which repels secondary electrons back to the plate does not exist in the form of wires, but rather in the form of a collection of other electrons.

The second grid, corresponding to the screen grid in a tetrode, is spiral-wound like the control grid, and each turn is shaded from the cathode by a turn of the control grid. For this reason not many electrons hit the screen and little screen current flows.

The plate characteristics of a 6L6 beam power tube are shown

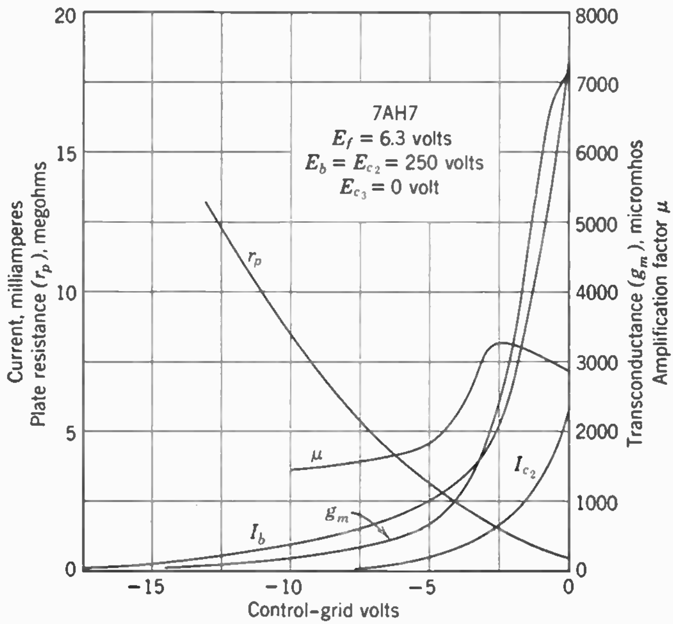


FIG. 14. Characteristics of a variable-mu or super-control tube.

in Fig. 16. The curves are flatter, and the "knee" occurs at a lower voltage than is the case for a comparable power pentode. The tube is very efficient since so little loss of current occurs in the screen circuit, and it has a very high sensitivity, which means that a large amount of power is delivered for a small input grid voltage.

11.21. Grid nomenclature. When tubes have several grids, it is customary to speak of the grids according to their position from the cathode. Thus the grid nearest the cathode is the No. 1 grid, the next grid is called the No. 2 grid, and so on. Grid voltages are labeled accordingly, E_{c_1} being the bias on the grid nearest the cathode.

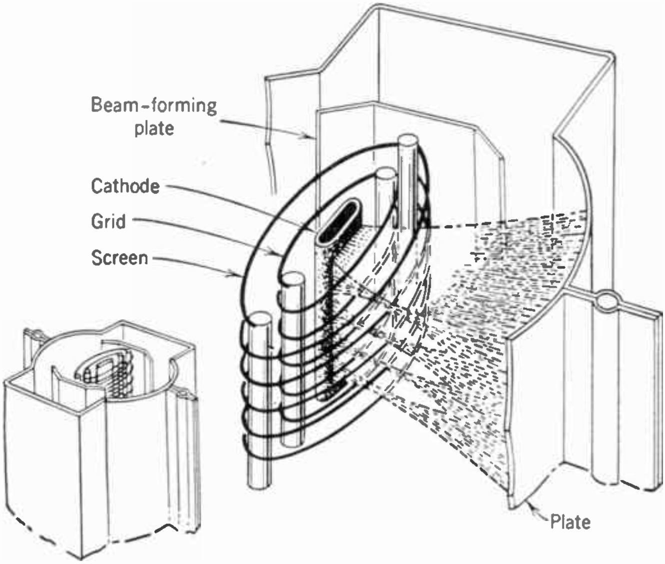


FIG. 15. Construction of beam power tube, showing beam effect on electron flow. (RCA)

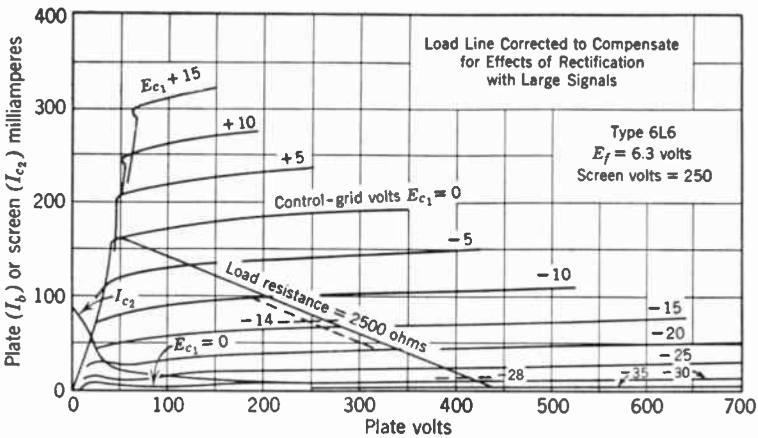


FIG. 16. Plate voltage-plate current curves of beam power tube.

11.22. Power output for pentode and beam power tubes. The general procedure for obtaining power output and distortion of pentodes and beam power tubes is much the same as for triodes. The calculations are made from the E_b–I_b family. In Fig. 17,

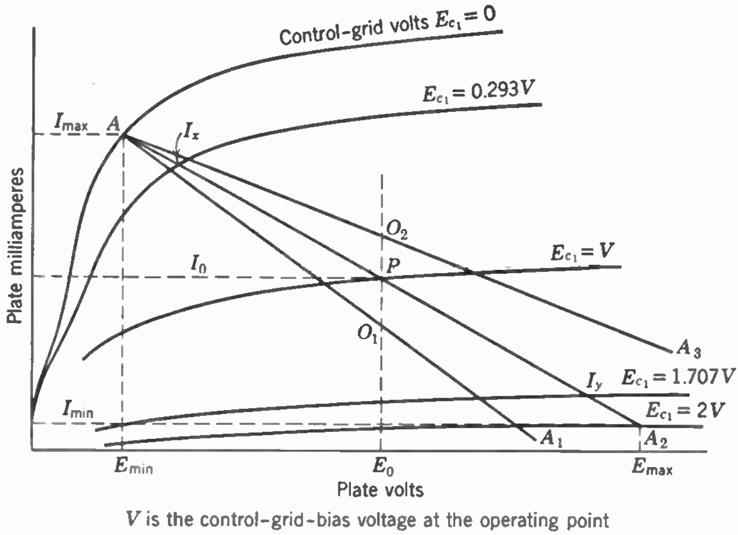


FIG. 17. Method of determining correct load for a pentode power tube.

from a point A just above the knee of the zero-bias curve draw arbitrarily selected load lines to the zero-plate-current axis. These lines should be on both sides of the operating point P, whose position is located by the desired operating plate voltage E_o (also E_b), and one-half the maximum signal plate current. Along any load line AA_1 measure the distance AO_1. Along this load line lay off line O_1A_1 equal in length to AO_1.

For best operation the change in bias from A to O_1 and that from O_1 to A_1 should be nearly equal. If they are not, select some other load line arbitrarily, until one is found which makes the equal-bias change a fact. The load resistance for the tube may then be found by

$$R_L = \text{Load resistance} = \frac{E_{\max} - E_{\min}}{I_{\max} - I_{\min}}$$

This value of load resistance may be substituted in the following formula and the power output obtained:

$$P = \text{Power output} = \frac{[I_{\max} - I_{\min} + 1.41(I_x - I_y)]^2 R_L}{32}$$

In these formulas, I is in amperes, R is in ohms, E is in volts, and P is in watts.

11.23. Distortion calculations for pentodes and beam power tubes. The following formulas are useful for calculating distortion from tubes of the pentode and beam power output types:

Second harmonic distortion

$$= \frac{I_{\max} - I_{\min} - 2I_0}{I_{\max} - I_{\min} + 1.41(I_x - I_y)} \times 100\%$$

Third harmonic distortion

$$= \frac{I_{\max} - I_{\min} - 1.41(I_x - I_y)}{I_{\max} - I_{\min} + 1.41(I_x - I_y)} \times 100\%$$

11.24. Gain from high-mu tubes. One of the expressions for gain given in Sect. 11.7 was

$$A_v = -\frac{g_m E_g R_t}{E_s}$$

$$= -g_m R_t \quad (E_g = E_s)$$

where R_t is the value of the load resistance and dynamic plate resistance in parallel, or

$$R_t = \frac{r_p \times R_L}{r_p + R_L}$$

In a high-mu tube such as a pentode or tetrode the plate resistance r_p is usually much larger than any practicable value of load resistance R_L. For this condition the value of R_t is equal (nearly) to R_L. A good expression for gain of a pentode or tetrode then is

$$A_v = -g_m R_L \quad (r_p \text{ much greater than } R_L)$$

This final result shows that gain may be increased either by making the effective load resistance larger or by using a tube with a higher g_m. Special tubes have been developed and are widely used in television, where the value of the load resistance must be small, often as low as a few thousands of ohms, to secure adequate band width. These special tubes feature high transconductances. For example, the 6AC7/1852 has a transconductance of 9000 micromhos.

11.25. Multi-unit tubes. Often more than one tube is mounted in a single glass or metal envelope. These are called **multi-unit tubes.** In some such tubes the individual units are entirely distinct with separate cathodes and with shields to prevent electrons from one unit straying to the other; other tubes share the cathode; and in still other tubes, extra grids are added to exercise additional control functions, as in converter tubes.

In the 6SA7, for example, there are seven electrodes. This tube is employed as a converter tube in a superheterodyne circuit. It mixes signals coming from the antenna with signals generated in an oscillator within the tube. These two signals are at different frequencies, and in the output of the tube occur signals of still another frequency. For descriptions of these tubes and circuits see sections on superheterodyne receivers.

11.26. Metal tubes. The tube elements may be mounted within a glass or metallic envelope. Modern receivers use both types. The metal-envelope tubes are smaller and better shielded than those in glass envelopes.

Several kinds of bases and sockets are utilized, but, since the tube operations are not controlled by the kind of base, this matter will not be discussed here.

11.27. Miniature tubes. In recent years many tubes have been made in the miniature type. These tubes are smaller than conventional tubes. There is no separate base; the pins are sealed through the bottom of the tube envelope. Miniature tubes permit the design of smaller equipment. Their characteristics are much the same as those of larger tubes of the same types, and the upper frequency limit of operation is higher. Their power rating is relatively low.

11.28. Cathode-heater connections. In heater-type tubes, it is common practice to connect the cathode to the center tap of a transformer winding supplying heater voltage or to the center of a 50-ohm resistor across this winding as shown in Fig. 18. It is sometimes desirable to have the heater somewhat positive with respect to the cathode to prevent any electrons from flowing from heater to cathode and thereby producing a-c hum. This voltage is of the order of 10 volts. In general, however, it is desirable to have no high potential between cathode and heater. Since the electrical insulation between heater and cathode will withstand only relatively small voltages, the manufacturer often sets an upper limit on the voltage which should exist between these elements.

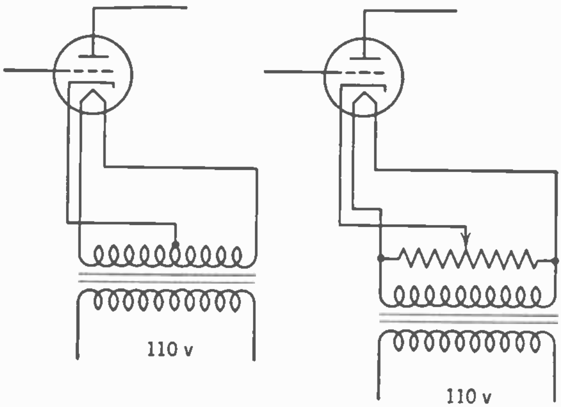


FIG. 18. Cathode of heater-type tubes is connected to filament circuit in two ways as shown. This minimizes hum.

The importance of the proper cathode-heater connection is illustrated in Fig. 19. This is a typical cathode-keying circuit often used in amateur radio transmitters. The plate current is interrupted by the key to form dots and dashes. When the key is open, no plate current flows and the output of the tube is reduced to zero. At the same time the cathode assumes the potential of the plate supply battery; that is, it is E_{bb} volts above ground. When the key is closed the cathode is at ground potential. If the center tap of the heater winding of the filament transformer (or one side of the winding if a center tap is not provided) is connected above the key at point A, the d-c potentials of the heater and cathode go up and down together as the key is opened and closed, and there is no potential difference between them. However, if the heater connection is made directly to ground, the potential difference between heater and cathode can rise to

E_{bb} volts and may easily destroy the insulation between the heater and cathode.

11.29. Tube operation. Tube filaments or heaters may be operated from batteries or from d-c or a-c lines. The heater-type tube may be operated on alternating or on direct current with little danger of hum appearing in the output. Filament-type tubes cannot be operated on alternating current except in the

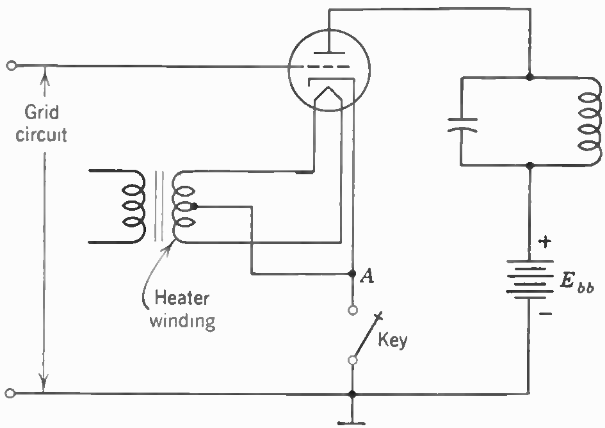


FIG. 19. The heater and cathode are connected at A so that the least possible potential difference exists between them as the key is opened and closed.

final output stage, since otherwise hum produced in the filament will be amplified by succeeding stages and cause noise in the output.

If several tubes are to be operated in a circuit, the electron-emitting filaments or the heaters may be wired in series or in parallel. If the source of filament power is of low voltage, such as a battery or a step-down transformer, the filaments are wired in parallel. Each tube gets the same voltage, and the current required from the source is the sum of the currents taken by the individual tubes. If the source of filament power is of high voltage, a house-lighting circuit for example, the tubes are frequently wired in series. Each tube, then, gets the same current, and the total voltage required to heat the filaments properly is the sum of the voltages appearing across each tube.

Filament voltages vary from 1.4 to 117 volts. The 1.4-volt tubes are designed to operate directly from a dry cell; the 2-volt tubes operate directly from a single storage cell; the 6.3-volt tubes may be operated directly across a 6-volt storage battery. If these tubes are to operate from higher voltage sources, means must be provided to limit the flow of current through the filaments. Thus, if a 1.4-volt tube is to be operated from a 2-volt battery, a series resistance must be provided. If several tubes are to be operated with their filaments in series across a line whose voltage is greater than the sum of the tube voltages required, a voltage-dropping resistor must be employed.

The required resistance may be calculated from the formula:

$$\text{Required resistance} = \frac{\text{Supply volts} - \text{Voltage required by tubes}}{\text{Current required by the tubes}}$$

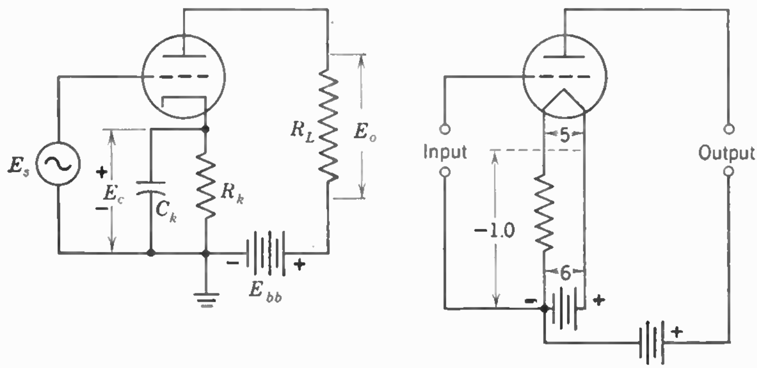
If one 6SA7, one 6SK7, one 6B8, one 25A6, and one 25Z6 are to be operated in series across a 117-volt line, the resistance equals $(117 - 68.9) \div 0.3$, or 160 ohms.

Tubes must be selected that require the same filament current if they are to be connected in series. Otherwise, the tubes requiring the lowest current must be shunted by a resistor of the proper size to prevent the tube filament from burning out.

11.30. Grid-bias arrangements. A tube operated from batteries may get its grid bias from a C battery. Batteries, however, are inconvenient, and their use is avoided except where they are the only practical source of power.

One of the simplest ways of placing a steady negative voltage on the grid is by means of a resistor through which the cathode current of the tube flows. This is called **cathode** or **self**-bias. Such a circuit is shown in Fig. 20(a). The d-c component of plate current flows down through R_k (the electron flow is upward), and a d-c voltage is produced across R_k. The bottom end of R_k is negative in respect to the top end. The grid bias is the d-c voltage from cathode to grid, and, in this circuit, this d-c voltage is negative. The value of R_k is such that the d-c cathode current produces a voltage drop across this resistor which is of the proper value for biasing the tube.

In a triode the cathode and plate currents are the same; in a tetrode, pentode, or beam power tube the cathode current is the sum of the d-c plate and screen currents. The condenser C_k by-passes alternating components of the plate current around the resistor and prevents these currents from producing a-c voltages



(a) Cathode Bias

(b) Bias Is Developed across Filament Voltage-dropping Resistor.

Fig. 20. Methods of biasing the grid negative.

which would be fed back to the grid circuit. The reactance of the condenser ($X_c = 1/2\pi f C_k$) at the lowest frequency to be amplified should be one-tenth or less of the value of R_k.

Example 2. Suppose that a tube has a d-c plate current of 4 ma when the supply voltage is 250 volts and the plate load resistor is 25,000 ohms. What value of bias resistor is required if the tube bias is to be −2 volts?

Solution: The value of R_k is simply the bias voltage divided by the d-c plate current, or

$$R_k = \frac{2}{4 \times 10^{-3}} = 500 \text{ ohms}$$

The d-c plate current will actually be slightly smaller after the bias resistor is inserted, and the bias voltage will be somewhat smaller than required. However, if the value of the cathode resistor is small compared with the value of the plate load resistor this difference is not serious.

Sometimes the voltage across the voltage-dropping resistor required in the filament circuit of a directly heated tube can be utilized for bias voltage. This arrangement is illustrated in Fig. 20(b).

In the case of series filaments, operated from a d-c source, advantage may be taken of the voltage drop across tube filaments to bias one or more grids. Thus, in Fig. 21 the filaments are wired in series. The same current flows through each filament, and the total voltage required is the sum of the individual tube filament voltages, plus the voltage drop across the 10-ohm resistors. If each tube requires, for example, 250 ma at 5 volts, then the total voltage across the tube filaments is 5 × 5 or 25

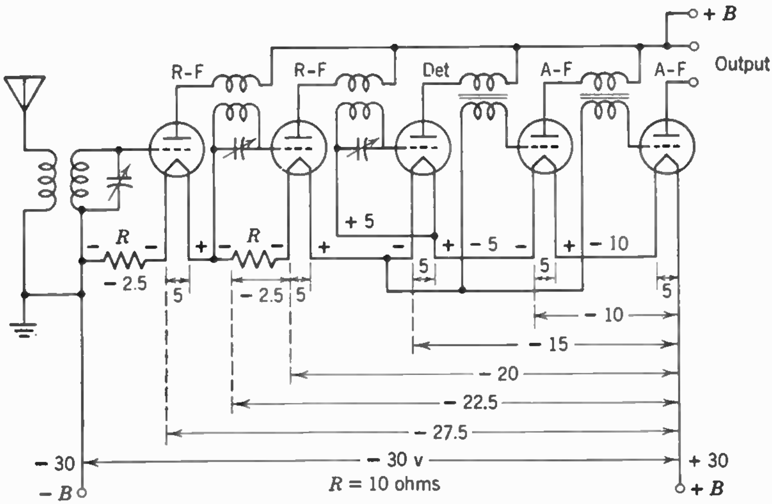


FIG. 21. Typical series-filament circuit in which voltage drops across filaments or series resistors are used as C bias.

volts. To this must be added the voltage drop across the resistors in the grid circuits of the first two tubes. These voltages are each 2.5 volts and the total is 5 volts. Then the total required voltage is 30 volts.

The drops across the filaments themselves may also be utilized for biasing some of the tubes. For example, the final power output tube requires a bias voltage of -10 volts. This is obtained by attaching its grid to the negative side of the filament of the third tube. The voltage drop across the filaments of the third and fourth tubes is 10 volts, and so the grid of the final tube is -10 volts negative with respect to the negative side of its own filament. A positive bias of 5 volts for tube 3 is obtained by returning its grid to the positive side of its own filament. Tube 4 receives a bias of -5 volts from the negative side of the filament of tube 3.

Problem 14. A 6J7 pentode has a plate resistance of 1 megohm and a grid-plate transconductance of 1185 micromhos when $E_b = E_{c_2} = 100$ volts and $E_{c_1} = -3$ volts. The load resistor is 250,000 ohms. Compute the gain, (*a*) neglecting the effect of the plate resistance, and (*b*) taking into account the effect that the plate resistance has on the gain. Repeat for a load resistance of 100,000 ohms.

Problem 15. A 6V6 beam power tube has a plate current of 29 ma and a screen current of 3 ma when the plate and screen voltages are both 180 volts. If the tube is to be operated with -8.5 volts bias from a cathode biasing network, what must be the size of the cathode resistor R_k? What must be the minimum value of the cathode condenser C_k if the lowest frequency to be amplified is 100 cycles?

Problem 16. The filaments of the following tubes are connected in series: 25Z6, 25L6, 6SQ7, two 6SK7's, and 6BA7. What must be the value of the voltage-dropping resistance if this string of tubes is to be operated from a 117-volt line? What must be the minimum power rating of the resistor?

Problem 17. A 6.3-volt tube requiring 0.15 amp filament current is connected in series with a group of tubes which require 0.3 amp. What should be the size of the resistor shunted across the 0.15-amp tube so that its filament current will be the proper value?

Problem 18. Using data from tube manuals and manufacturers' data sheets, list typical values of μ, g_m, and r_p for voltage- and power-amplifier triodes, tetrodes, beam power tubes, and voltage- and power-amplifier pentodes.

12 · Rectifiers and Power Supply Apparatus

Among the many useful functions of vacuum tubes is the conversion of alternating currents into direct currents for charging batteries, for furnishing power for radio receivers and transmitters, and for various electrochemical applications. Battery chargers require high currents at low voltages; radio apparatus requires high voltages at low currents. High-vacuum rectifier tubes for x-ray apparatus operate safely at voltages as high as several hundred thousand volts. Gaseous rectifiers will pass several thousand amperes for electroplating, welding, and other applications. Three-phase gaseous rectifiers perform very useful industrial functions.

The essential characteristic of a rectifier is that it permits current to flow in only one direction. The output current is pulsating and so must be filtered to remove the alternating components before it can be used for certain applications. The study of power supply apparatus involves the study of rectifier tubes, rectifier circuits, and filter circuits.

RECTIFIER TUBES

12.1. High-vacuum rectifiers. When high-voltage, low-current d-c power is desired, high-vacuum tubes are ordinarily used. If greater current is necessary gaseous rectifiers are required. If still greater currents are needed, mercury-pool rectifiers are employed. Figure 1 shows characteristic curves of receiver-type rectifier tubes, giving the voltage drop across the tube for various currents flowing through it. Note that the voltage drop varies with voltage for high-vacuum tubes and is a constant value of about 15 volts (or less) for gaseous tubes.

12.2. Mercury-vapor tubes. When greater currents are needed than can be safely handled by high-vacuum tubes, or

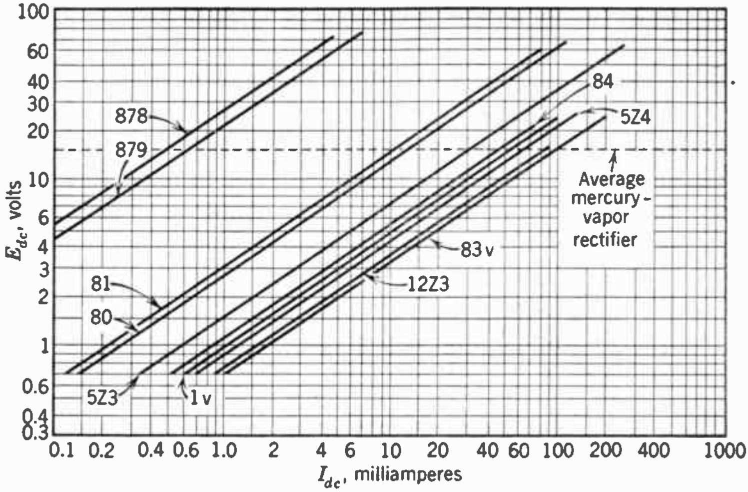


FIG. 1. Voltage drop versus current through several types of rectifier tubes. Note the dotted curve for the mercury-vapor rectifier.

when greater efficiency is required, mercury-vapor or other types of gaseous tubes must be employed. The voltage drop across gaseous rectifier tubes is essentially constant at about 15 volts for current values over the entire operating range. Gaseous tubes are especially useful where currents are in excess of 100 ma.

Since the voltage drop in a gaseous rectifier tube is constant, some means must be provided to limit the current flowing through it in the event of a short circuit in the load. Otherwise the cathode of the tube will be damaged (see Sect. 10.10). Gas tubes also create radio interference at times. This trouble is due to the fact that they start conducting abruptly on the half cycle which makes the plate positive, and this steep wavefront or abrupt starting resembles turning on and off an electric switch. Some interference ("hash") also results from erratic motion of the gas molecules themselves. A choke coil of about 1 mh connected in series with and close to the plate and placed within a metallic shield which also contains the tube will help reduce the difficulty. Condensers connected from the outer terminals of the transformer secondary windings to the center tap, plus the use of chokes as described, will usually eliminate most of this trouble.

To prevent excessive current flow in the event of a short circuit, series resistors are sometimes placed in the plate leads, or the transformer supplying the tube may be designed to have relatively high magnetic flux leakage at high currents which will also limit the short-circuit current. Fuses in the transformer primary winding are essential so that power will be removed as quickly as possible when a short circuit occurs.

Mercury-vapor tubes have heavy cathodes in order to supply a very large number of electrons. Considerable time is required for these cathodes to heat up to the required emitting temperature. If plate voltage is applied and current taken from the tube before the cathode has reached the required temperature, an abnormally high voltage drop appears across the tube, which may accelerate positive ions to a velocity sufficient to damage the cathode coating by impact. Time-delay relays are often employed to keep the load circuits open until the cathode has had time to reach its operating temperature.

Gaseous rectifiers have good voltage regulation,* that is, the output voltage remains nearly constant over wide ranges of current, and are employed in circuits where considerable fluctuation of output current is experienced. In radio receiver circuits the current taken from the power supply system is fairly constant; but in a transmitter, for example, where the output is keyed, the current requirements vary widely from instant to instant. If the internal drop within the rectifier tube varies with current output (as it does in high-vacuum tubes), the voltage supplied to the transmitter varies from instant to instant. This will not be the case in gas-tube rectifiers except for the small voltage drop across the resistance of the transformer windings and of the filter chokes.

12.3. Controlled rectifiers. A most important addition to the vacuum-tube family has been made in the **thyratron,** which is a gaseous rectifier with a control grid. The characteristics of a thyratron present many interesting points. For example, if the grid is made negative by the proper amount no current will flow to the plate even though it is positive with respect to the cathode.

* **Voltage regulation** is defined as follows:

$$\text{Voltage regulation} = \frac{\text{No-load voltage} - \text{Full-load voltage}}{\text{Full-load voltage}}$$

The term is often used in a general way to denote how rapidly output voltage drops as load current increases. "Good" voltage regulation means that the numerical value is small, that is, the full-load value is nearly as large as the no-load value.

At any given positive plate voltage, however, there is a value of grid voltage which will permit current to flow—and when current flows it is limited only by the external circuit, not by the tube itself. In this respect the thyratron is the same as the gas diode.

There is another interesting characteristic of the thyratron. Once the plate current begins to flow, the grid loses control over it and negative voltages on the grid make no difference—the plate current continues to flow. The only way to stop the flow of electrons from cathode to plate is to remove the positive voltage from the plate. The voltage may be removed by opening a switch if it is a d-c voltage; if it is an alternating voltage, the current stops flowing when the alternating voltage goes to zero and does not start again until the plate becomes positive.

Suppose that an a-c voltage is placed on the plate and a fixed bias on the grid of a thyratron. Whenever the plate has reached the point in the positive half cycle at which its positive voltage is high enough to overcome the inhibiting effect of the grid voltage, the tube will "fire"; current will flow. Then, at the end of the positive half cycle, the plate voltage is zero and the plate-current flow ceases. During the negative half cycle, of course, no plate current flows. By adjusting the bias on the grid, the point in the positive half cycle at which current starts may be controlled—but current will flow throughout the rest of that half cycle.

Now, if alternating voltage is placed on the grid as well as the plate and if the phase between grid and plate is adjusted, any amount of average current may be taken from the tube, from zero to the maximum rated current. The thyratron is a gaseous rectifier, with low internal voltage drop, which can supply many amperes of current under the control of the user by adjusting the phase between voltages on grid and anode. A high-vacuum triode acts like a rheostat and a rectifier in series. A thyratron acts like a latching relay plus a rectifier plus an opposing battery.

Mercury-vapor tubes are rather sensitive to operating temperature and must be used in the temperature range specified by the manufacturer. The dependence on temperature is seen in Fig. 2, which shows the characteristics of a typical mercury-vapor thyratron.

Thyratrons which employ a second or shield grid are also made. The chief purpose of the shield grid is to reduce the number of electrons collected by the control grid and so raise the grid-cathode impedance. In most applications the shield grid is operated

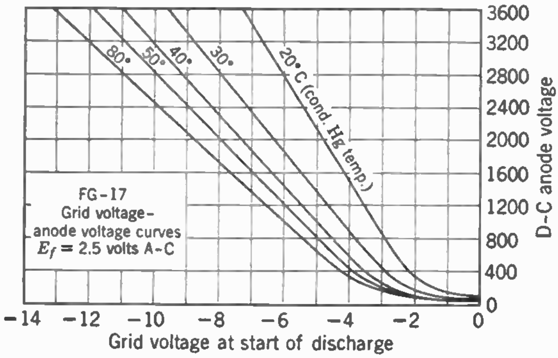


FIG. 2. Characteristics of mercury-vapor thyratron. Note that the temperature affects the starting voltage.

at zero potential. This grid can, however, be operated at other potentials as shown in Fig. 3. The effects of temperature are much the same as for a simple thyratron. The general features

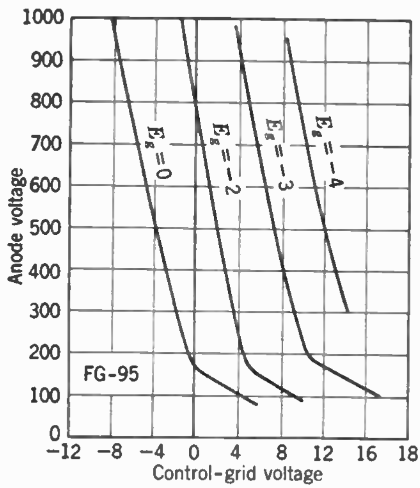


FIG. 3. Characteristics of shield-grid thyratron. Note that various starting voltages are possible, depending upon voltage of shield grid.

of operation of a shield-grid thyratron are the same as those of a three-element thyratron.

12.4. Relay tubes. The OA4G is a gaseous triode designed for use primarily as a relay tube in carrier-actuated equipment. The tube is filled with an inert gas or vapor. The electrodes are a cathode, starter-anode (sometimes called the grid), and an anode. The cathode is cold; that is, there is no heated filament to supply electrons. Instead, the electrons are supplied from a small amount of active material, usually a cesium compound, on the unheated cathode. A relatively small voltage on the starter-anode starts the discharge between this electrode and the cath-

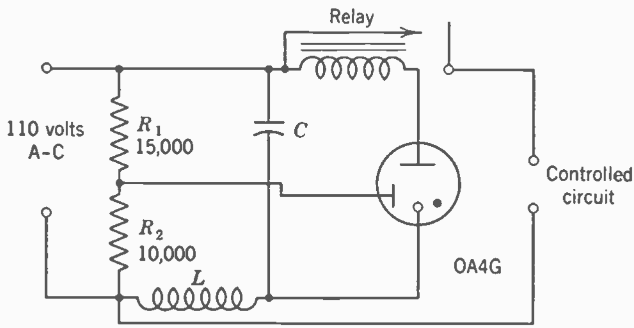


Fig. 4. Use of gas tube in a remote-control circuit.

ode. This discharge causes complete ionization of the inert gas and initiates the main discharge between cathode and anode. It is the current that flows in the anode circuit which performs the work of the tube, such as closing a relay. In use, the cathode material becomes spattered over the other two electrodes, and it is possible for the initial discharge to occur between any two of the electrodes. The circuit is usually arranged so that the discharge proceeds as above.

In the remote-control function in connection with carrier-actuated equipment, which is the primary use of a relay tube, the starter-anode is maintained at a voltage just below that required for breakdown. In Fig. 4 this voltage is supplied from the power lines and is taken from across R_2 of the voltage divider R_1–R_2. A tuned circuit made up of L and C has a considerable voltage developed across it when a carrier of the proper frequency is impressed on the line at some distant point. The voltage across L increases the potential between cathode and starter-anode, and the increase of potential starts the discharge which flows between cathode and anode and closes the relay. The tube ceases to pass current when the carrier is removed.

12.5. Deionization time. A thyratron ionizes (fires) very rapidly once the plate voltage is high enough. Once a tube has fired the negative grid is surrounded by positive ions. These ions are one of the two products of ionization, that process by which electrons are knocked out of mercury (or gaseous) molecules to leave the molecule positively charged. A part of the positive ions neutralizes the negative voltage on the grid, preventing it from exercising any further control. Other positive ions neutralize the space charge of the tube. When the plate voltage is cut off, or goes to zero as with an a-c voltage, the ions recombine with electrons and become neutral molecules again. A few microseconds are required for this recombination process, and the required time is known as the **deionization time.** If this time is not permitted to elapse before the anode again becomes positive, the effect of the grid will still be neutralized and it will not exercise its normal control upon the starting of plate-current flow. Thus, the anode voltage must be zero or negative long enough for the tube to deionize and permit the grid to regain control. The deionization time of mercury-vapor tubes is of the order of 100 to 1000 microseconds, and such tubes are therefore not suited for use at frequencies above a few thousand cycles.

Some gaseous tubes containing inert gas have a deionization time low enough to permit operation up to around 30,000 cycles. Typical of this type of tube is the 884, often used in the sweep-oscillator circuits of cathode-ray oscillographs.

RECTIFIER CIRCUITS

12.6. Half-wave rectifier. The simplest rectifier circuit is shown in Fig. 5. The a-c power source is connected in series with the tube and the load. Current flows through the circuit on the half cycle which makes the plate positive, but not on the other half cycle. If I_m is the maximum or peak value of the current passed during the conduction half cycle, the average value (which would be read by a d-c ammeter) is $0.318\ I_m$ for a resistance load. In addition to the d-c component, the output current contains many a-c components. The relative size of these components is shown in Table 1.

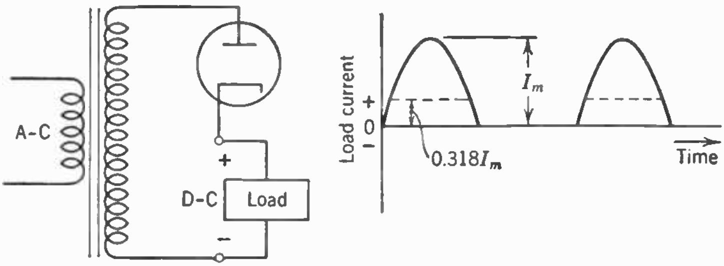


FIG. 5. A half-wave rectifier circuit. The load current is for a resistance load.

TABLE 1

CURRENT COMPONENTS IN RECTIFIERS
RESISTANCE LOAD

COMPONENT	HALF-WAVE	FULL-WAVE
D-c	$(1/\pi)I_m$	$(2/\pi)I_m$
f	$\frac{1}{2} \times I_m = (\pi/2)I_{dc}$	None
$2f$	$(\frac{2}{3\pi})I_m = \frac{2}{3}I_{dc}$	$(\frac{4}{3\pi})I_m = \frac{2}{3}I_{dc}$
$3f$	None	None
$4f$	$(\frac{2}{15\pi})I_m = \frac{2}{15}I_{dc}$	$(\frac{4}{15\pi})I_m = \frac{2}{15}I_{dc}$

NOTE. Frequency f is that of the power source. Frequency components higher than those shown in the table are correspondingly smaller.

The half-wave rectifier has three disadvantages: (1) The output contains an a-c component of the same frequency as the power source. If this frequency is low, say 60 cycles, it is difficult to filter or smooth it out so that the load has only direct current flowing through it. (2) The a-c power supply is furnishing power only half the time so that the peak load is high compared to the average load. (3) If the circuit employs a transformer as shown in Fig. 5, and this is often the case, the d-c pulses are taken through the secondary winding and cause saturation and a consequent lowering of the efficiency of the transformer. These disadvantages limit the use of the half-wave rectifier to low-current loads, such as cathode-ray-tube circuits, and to applications where there is no transformer. In the latter case the unit using the tube can be operated from either an a-c or d-c source.

12.7. Full-wave rectifier. If two tubes are connected as in Fig. 6 a full-wave rectifier results. Now each tube conducts on each alternate half cycle so that current flows through the load

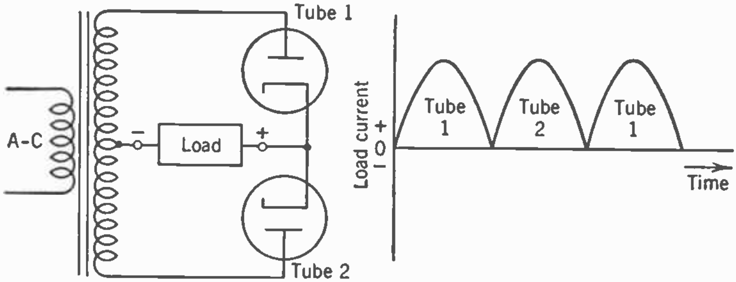


FIG. 6. Full-wave rectifier which passes current on both halves of the a-c cycle. The load current is for a resistance load.

all the time. Under these conditions the lowest-frequency a-c component has twice the frequency of the power source and is fairly easy to filter. The ratio between the peak power and the average power taken from the transformer is lower than for a half-wave circuit. Since current pulses flow in opposite directions in the two halves of the secondary winding, saturation effects are avoided.

Many rectifier tubes for use in full-wave rectifier circuits contain two plates and a common cathode in the same glass or metal envelope. A typical example is the 5Z3.

12.8. Voltage-doubler circuits. It is possible to have two tubes in a circuit which will produce a d-c output voltage with a value approximately twice the peak voltage of the a-c line from which the circuit derives its energy. The full-wave voltage-doubling circuit of Fig. 7 requires two rectifier tubes with their outputs connected in series. This type of rectifier circuit is most valuable in applications where it is impossible or undesirable to use a power transformer. It is possible to obtain a useful output of 240 volts at 40 ma from a 117-volt a-c line from this circuit.

Consider tube *B*. When its plate is positive, current flows through the tube and charges C_2 to the polarity shown. Now consider tube *A*. When its plate is positive, current flows through the tube and charges C_1 to the indicated polarity. Note that the polarities of the two condensers are such that their voltages add. The total load voltage is the sum of the voltages across the condensers. If no current is taken from the rectifier, the condensers charge to the full peak voltage of the a-c input. When current is drawn by the load, the d-c terminal voltage drops somewhat.

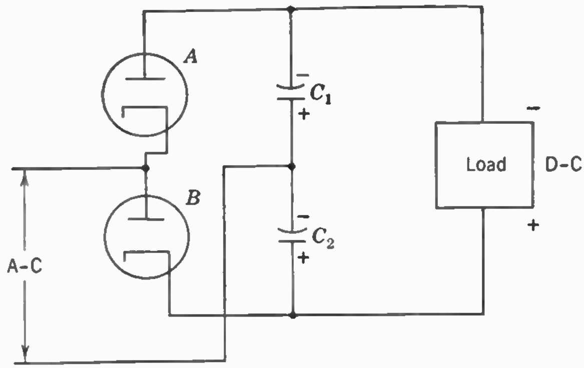


FIG. 7. Use of two diodes in a full-wave voltage-doubling circuit.

Thus, the voltage regulation of voltage-doubler circuits is poor. About 75 ma is the maximum useful current obtainable from doubler circuits.

A special tube such as the 25Z5 or 25Z6 is commonly used in voltage-doubler circuits. These tubes have two separate diodes in one envelope. Only the heater is common. In each of these two tubes the heater requires 25 volts and is ordinarily connected in series with the heaters of the other tubes of a receiver. The series combination, along with an appropriate voltage-dropping resistor, if needed, is connected across the a-c line. A typical circuit using a 25Z5 is shown in Fig. 8.

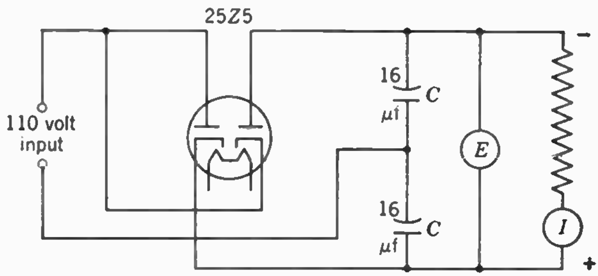


FIG. 8. A rectifier with two plates and two cathodes will double the voltage obtainable from a single diode.

The circuit of Figs. 7 and 8 has two disadvantages: it cannot be grounded, nor can the d-c output be connected to one side of the a-c line. Hum may be produced because of the high voltage between heaters and cathodes. A circuit which overcomes the difficulty is shown in Fig. 9. One side of the output is connected to the line. It is called a **half-wave doubler** because rectified current flows to the load only on alternate half cycles of the input voltage. The left-hand diode charges C_1 which discharges in series with the line and the other diode on the next half cycle. Condenser C_2 helps smooth the output current. The regulation is not as good as that of the full-wave doubler of Fig. 8.

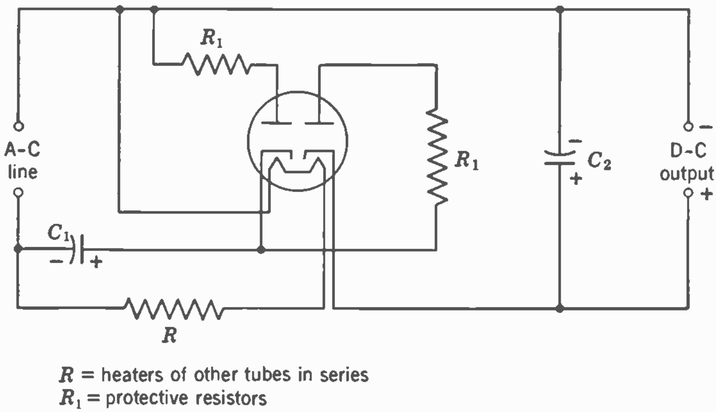


FIG. 9. Half-wave voltage doubler in which one output terminal is at line or ground potential.

Circuits can be devised which will triple or quadruple the input voltage. The regulation of these circuits is so poor that they are not often used.

Voltage-doubler tubes have other interesting applications. Since the cathodes and anodes are separate, these tubes really constitute two complete rectifiers. Thus one half of a tube may be used to supply plate voltage and the other half to supply grid-bias voltage; or one side may supply plate voltage and the other half supply energy for a dynamic loud-speaker field coil.

12.9. Rectifier-tube ratings. High-vacuum rectifiers have fairly high internal resistance, and the voltage across the tube must rise as more current is taken by the load. This makes the tube, in a measure, self-protecting from excessive currents. Gaseous and vapor tubes, on the other hand, have a low and constant voltage drop of about 15 volts across them. Larger load currents do not require a corresponding increase in voltage, and these tubes will draw too much current unless adequately protected as described in a previous section. In addition, mercury-vapor tubes have characteristics which depend to some extent upon the temperature of the tube. Flow of plate current should be delayed until the cathode has had time to warm up to the proper temperature. In addition to this time delay, the mercury vapor or other gas must be allowed time to reach the proper temperature. Thus, a tube rating, applied most often to gas tubes, is the **cathode-heating time.**

The **maximum peak inverse voltage** is the highest peak voltage that may be applied safely to the tube in the direction opposite to that in which it conducts current (plate negative). Even with the same a-c supply voltage the inverse voltage on a tube varies from circuit to circuit. Care must be taken in a particular circuit that the peak inverse voltage rating of the tube is not exceeded.

The **maximum peak current** is the highest instantaneous current that a rectifier tube can safely stand in the direction in which it is designed to pass current. This peak current may occur each cycle. If a condenser is used as the input element of a filter, the peak current is often many times the average load current, and care should be exercised that the peak current does not exceed the current rating of the tube. An inductor used as the first element of a filter tends to reduce the peak current passed by the tube to a value roughly twice that of the d-c load current.

The **maximum average plate current** is the highest value of average (d-c) current that should be allowed to pass through the tube. This current may be read on a d-c meter placed in series with the tube.

12.10. Dry disk rectifiers. Certain combinations of metals, such as copper sulfide and magnesium, copper and copper oxide, or iron and selenium, have the useful property of passing current readily in only one direction. Their voltage-current characteristics are much the same as those of crystal rectifiers discussed in Sect. 10.23. The copper oxide rectifier is typical in its simplicity and reliability. It consists of a sheet of copper, on one side of which has been formed a coat of cuprous oxide (Cu_2O). Properly made, this combination has relatively low resistance in the direction oxide-to-copper, and very high resistance in the reverse direction.

The units are generally made in the form of washers, a typical unit having an outside diameter of 1½ in. These washers are assembled in any desired series and parallel arrangement on mounting bolts. Soft-metal washers are placed between the oxide layer and the adjacent metal surface for the purpose of improving the contact with the oxide. The surface of the oxide is graphitized for the same reason.

The rectifier operates electronically, not electrolytically. Rectification commences almost instantly on the application of voltage, with no forming or transient condition interposed. Current is carried uniformly over the active area. Furthermore, the elements may be operated in parallel. Operation in series presents no difficulties, because the disks divide the voltage with approximate uniformity.

Dry disk rectifiers can be used in either half-wave or full-wave circuits. The bridge connection is common since it simplifies the transformer design and permits units to be operated directly from the a-c line without an intervening transformer if so desired. Battery charging, battery elimination, magnet operation, loud-speaker excitation, relay operation, etc., in fact almost any d-c application, can be successfully handled by these simple rectifiers. They are often used in measuring instruments. Thus a d-c movement can utilize the rectified current produced by a copper oxide disk when alternating current is applied. In meter applications the disk is kept as small as possible so that it will have low capacitance.

In recent years a series of dry plate rectifiers have been developed which operate directly from a 117-volt a-c line. These are often used in place of a rectifier tube in ac-dc radio receivers.

FILTER CIRCUITS

12.11. General considerations. Since a rectifier passes current only when the anode is positive, and since alternating voltages are applied to it, the output is not a smooth flow of direct current but a pulsating one. Although this current may be applied directly to some uses, most radio applications require that the pulsations be smoothed out so that a steady direct current results. The smoothing takes place in a **filter** which may be made up of shunt capacitances, series inductances, and series resistances, in various combinations.

12.12. Shunt capacitor filter. The simplest filter is a condenser shunted across the load resistor. The action of the condenser can best be explained by reference to Fig. 10. The diagram shows conditions after the circuit has been in operation long enough so that initial transient conditions have disappeared. As current flows through the tube on the positive half cycle, the condenser is charged. On the negative half cycle a part of this charge flows through the load. This results in an output voltage waveform as shown in the figure. The filter condenser is usually large enough so that it does not have time to discharge completely before the next positive half cycle of voltage appears, and the

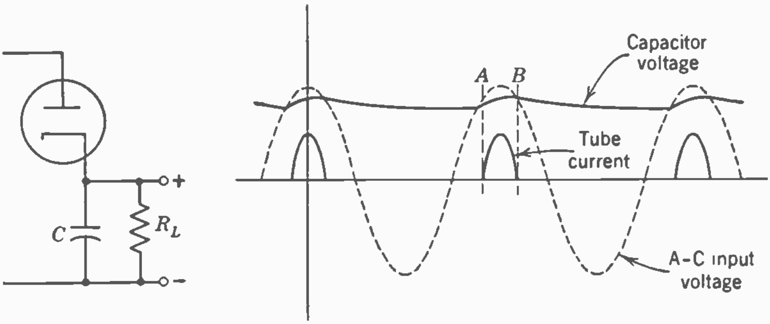


FIG. 10. Current and voltage for a shunt capacitor filter. The time constant R_LC is greater than the period of the a-c wave.

cathode of the tube is therefore held at some positive voltage. For this reason, plate current does not start to flow until the applied voltage becomes higher than the condenser voltage, that is, the plate of the tube becomes positive in respect to its cathode at a point such as A. As soon as the applied voltage drops below the condenser voltage, as at B, plate current ceases to flow. These short pulses of plate current may be quite high, particularly if the condenser is large and the load resistance small. Therefore, the maximum peak plate-current rating of the tube must be taken into account in the design of a circuit using a shunt capacitor filter.

This type of filter does not result in good smoothing of the output voltage wave except in applications in which the condenser and the load resistance are large. A high load resistance means small load current. Another way of stating these conditions is that the time constant ($t = RC$) of the condenser and load resistance must be large. For a circuit with a given time constant the filtering action will be better if the frequency of the input voltage is raised.

A mechanical analogy to the storage of charge in a capacitor in relatively short time intervals is a pump. If the handle is worked slowly the water all flows out during the downward motion of the handle. No water flows while the handle is being raised in preparation for the next downward stroke. If, however, the handle is worked at a faster rate so that the complete up-and-down cycle is rather short, water will flow continuously since it takes longer for the water to flow out of the pump than it takes to raise the pump handle. This corresponds to raising the frequency in a rectifier circuit. If the size of the spout opening is reduced while the handle is operated at a constant rate, the water will not flow out so fast and the flow will be smoother. This corresponds to making the time constant of the RC combination longer.

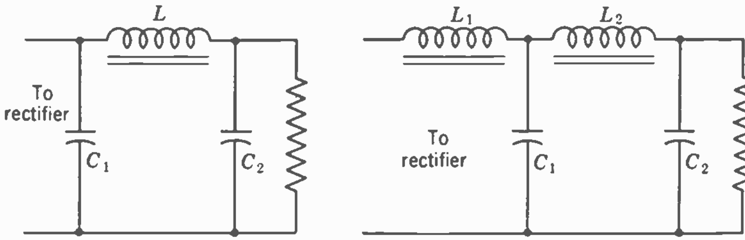
If the rectifier is a full-wave system, the condenser is charged twice as often as in a half-wave system, the discharge time is shorter and so the voltage drop between rechargings will be less. The variation in the voltage in a full-wave rectifier occurs at twice the rate of that of a half-wave system. This "ripple" voltage is easier to filter because of its higher frequency.

On the half cycle in which the tube does not conduct, the voltage across the tube is equal to the line voltage plus the voltage to which the condenser has been charged. For this reason the tube must be insulated to withstand voltages approximately twice the output d-c voltage. Furthermore, since the tube conducts only during a short period near the voltage peaks, the transformer supplies power to the rectifier only part of the time, even in a full-wave circuit.

A rectifier using a shunt capacitor filter has poor voltage regulation. That is, the d-c voltage output drops rapidly with increasing load current. This poor regulation coupled with excessive ripple for large currents makes this type of filter unsuited for applications which require fairly large currents with little ripple. The chief application of this kind of filter is to the rectifiers supplying voltage to the high-voltage anodes of cathode-ray and similar tubes which require very little current—a few milliamperes at most.

12.13. Filters using chokes and condensers. Filters for use with rectifiers supplying direct currents above a few milliamperes are divided into two general types, depending upon whether the shunt condenser or the series choke is connected next to the rectifier terminals. The first type is a **condenser-input** filter, the second a **choke-input** filter. Typical examples of each are shown in Fig. 11.

The condenser-input filter is almost always used when the load currents are relatively small since the output voltage of this system, like that of the shunt capacitor filter, approaches the peak of the applied a-c voltage. The output voltage of a choke-input filter approaches the average value, or 0.636 times



(*a*) Condenser Input (*b*) Choke Input

FIG. 11. Typical power supply filters.

the peak value of the a-c input voltage. If the second type is properly designed, its voltage regulation is better, and for this reason it is used in transmitters where large currents are required.

The first condenser of a condenser-input filter acts much the same as in the simple capacitor filter. Current flows to it in short, high pulses. Between the peaks of the voltage wave the condenser discharges through the choke into the second condenser and the load. Since the choke tends to hold the current through itself constant, the current supplied to the final condenser and the load is nearly constant, and there is little ripple in the final output voltage. The second condenser acts to bypass any remaining ripple components.

The first choke of the choke-input filter tends to keep the same current flowing to the load throughout the cycle and in so doing tends to drop the output voltage to the average of the input voltage as described above. Thus the peak tube current is considerably less than in a comparable filter of the condenser-input type. The first condenser C_1 charges and tends to hold its voltage constant. The second choke and condenser further aid

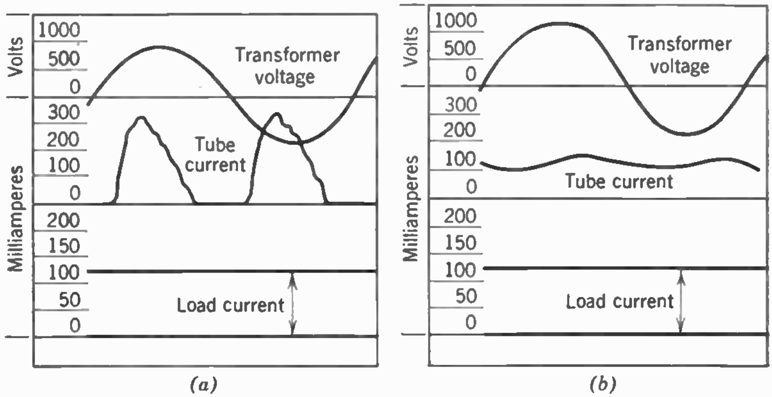


FIG. 12. High instantaneous current required by condenser-input filter (*a*) compared to the lower currents passed by the inductance-input filter (*b*).

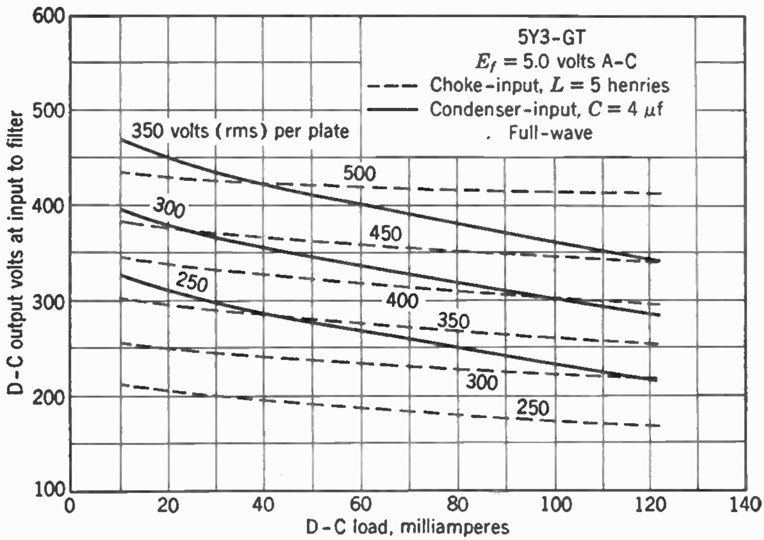


FIG. 13. Comparison of condenser-input and choke-input filter systems. Voltage drops due to resistance of filter chokes must be added to these regulation curves if complete regulation data are desired.

the filtering action so that there is very little remaining ripple in the output wave.

A comparison of the tube currents in the two types is shown in Fig. 12, and typical voltage-current curves in Fig. 13.

12.14. Filter design. The output of a rectifier may be considered as being made up of a steady direct current plus an a-c component (ripple). The shunt condensers of the filter offer low impedance to the ripple and very high impedance to the pure direct current. Thus the a-c components are shunted or bled off by the condensers. The series inductances offer high impedance to the flow of a-c currents but do not impede the flow of direct current. A proper combination of L and C will produce an output voltage as free of ripple as desired. If a single filter section does not produce low enough ripple, two or more filter sections may be connected in cascade.

In a choke-input filter with a single choke and condenser, the ratio of the ripple voltage (a-c voltage across the load) to the alternating voltage applied to the input to the filter is approximately $1/(2\pi f)^2LC$. Thus

$$\frac{\text{Ripple across load}}{\text{A-c voltage to input}} = \frac{1}{(2\pi f)^2LC}$$

From this equation the value of the LC product for any frequency and any desired attentuation of the alternating voltage may be determined. If the filter has two identical chokes and condensers as in Fig. 11, the relation becomes $1/(2\pi f)^4L^2C^2$. For a circuit in which the 120-cycle ripple is the strongest (as in full-wave circuits operated from a 60-cycle source) the relations become.

$$\frac{\text{Ripple}}{\text{Input a-c voltage}} = \frac{1}{0.57LC} \quad \text{(one section)}$$

$$= \frac{1}{(0.57LC)^2} \quad \text{(two identical sections)}$$

where L = henries per section.
C = microfarads per section.

Example 1. A choke of 10 henries and a condenser of 10 μf are employed in a single-section choke-input filter. What is the ratio between the ripple voltage and the applied input alternating voltage?

$$\frac{\text{Ripple voltage}}{\text{Applied voltage}} = \frac{1}{0.57 \times 10 \times 10} = \frac{1}{57}$$

or the ripple will have 1/57 of the voltage applied to the input to the filter or roughly 2 per cent. Now, if two identical sections are employed, the ratio will be $(1/57)^2$ or 1/3250, which is about 0.3 per cent.

The above discussion implies that the filtering action depends only on the product of L and C and that any set of values would be useful as long as the product remained the same. This is not quite true. The purpose of the choke is to prevent a change in the current flowing to the load. There is a minimum value of inductance which can accomplish this purpose. For values of inductance lower than this critical value the condenser will completely discharge between charging periods, the ripple will rise, and the voltage regulation will be very poor.

The minimum value of inductance which should be used in a 60-cycle full-wave rectifier is $R_L \div 1130$, where R_L is the effective load resistance, which must include the resistance of the load plus the resistance of the choke. A much larger inductance would be required in a half-wave circuit. However, choke-input filters are rarely used in such circuits.

Another possibility of trouble in the selection of the L and C values is that they may produce series resonance. In such a case the ripple voltage will be very high. This can be checked with the resonance formula, $\omega = 1/\sqrt{LC}$, in which $\omega = 2\pi f$. The frequency f is 60 cycles for half-wave circuits and 120 cycles for full-wave circuits operating from a 60-cycle source. LC products which will produce resonance near these frequencies should be avoided.

12.15. Rectifiers using r-f power source. The transformers and filter components for high-voltage power supplies are rather expensive. In applications for which the required current is low, as for rectifiers supplying cathode-ray and television tubes, it may be less expensive to supply the rectifier with r-f power from an oscillator. The r-f voltages permit the use of air-core transformers and chokes and smaller condensers. Thus the components are not only less expensive; they are also smaller. It is practical to use such power supplies to furnish voltages at 10,000 volts or more at currents of a few milliamperes. Such circuits are discussed in more detail in Chapter 23.

12.16. Swinging choke. If the output current required from a power supply system varies widely, as in a keyed transmitter, the voltage supplied will vary because at high currents there is a greater voltage drop in the system than at low currents. Much greater freedom from regulation troubles will be had if a gaseous rectifier is used than if a high-vacuum tube is employed, for the simple reason that the voltage drop across the gaseous tube itself is relatively independent of the current passed through it.

Another method of keeping the output voltage more nearly independent of the load current is to use a **swinging choke** as the input choke of a filter such as shown in Fig. 11(*b*). A swinging choke has high inductance with low currents through it and low inductance when the currents are high. When the inductance is low, as it is at high current drains, the filter resembles a condenser-input system. When the inductance is high, as when the current taken from it is low, the filter resembles an inductance-input filter. The net effect is to keep the output voltage nearly constant as the current varies over a wide range.

The "swinging" effect of the choke is a matter of design, so that the inductance varies widely with the current through it. The design involves the dimensions of the air gap in the iron core. With a very small air gap the core approaches saturation (low inductance) at high values of current, but at low currents the choke will have a high inductance.

Chokes used in filters should always be rated at so many henries inductance with so many milliamperes of direct current flowing through them. A typical swinging choke might be rated at 6 henries at full or rated current and 15 henries at 10 per cent of the rated current.

12.17. Multisection filters. If the ripple in the output of a filtered power supply is too high for the purpose at hand, it may be reduced by employing additional filter sections. These sections may be identical with the input section, or they may be different. Rectifier-filter design does not lend itself to mathematical rigor, and most power supply systems are designed by trial or on the basis of past experience. The values of L and C are usually greater than necessary so that more than sufficient filtering is secured. Naturally, if the output of the filter is to supply the plate circuit of an amplifier tube whose a-c output is to be further amplified (a microphone amplifier, for example), the ripple must be smaller than if the amplifier output is fed directly to a loud speaker. In the first case, the ripple in the plate voltage due to the poorly filtered power supply may be nearly as large as the signal voltage itself; both the ripple voltage and the signal voltage will be amplified in the succeeding stages and the output will contain considerable hum. For low-level stages, that is, amplifying stages in which the signal voltage may be only a fraction of a volt, the ripple should be as low as 0.005 per cent of the a-c input voltage to the rectifier. In the final amplifier stages, where the signal voltage may be 50 or 100 volts, a much larger value of ripple can be tolerated; it may be as high as 1 per cent. In general, more than sufficient filtering is employed.

Often in an amplifier in which a multisection filter is required to reduce the ripple to a low value for the low-level stages, the plate voltage for the final amplifier stage is taken from an intermediate point along the filter, say after the first choke. This has the advantage of providing a somewhat higher value of voltage for the final stage. In addition, since the final stage often draws two or three times more current than the whole remainder of the amplifier, the current ratings of the chokes following the point to which the final stage is tapped may be much smaller. This results in smaller and less expensive components in the final portions of the filter.

12.18. Bleeder. In the discussion of choke-input filters it was stated that this type of filter gave better voltage regulation than a condenser-input filter. This statement was not entirely true. A typical curve of d-c output voltage plotted against d-c load current is shown in Fig. 14. Note that at very low plate currents the output voltage rises to approximately the maximum value of the input voltage. For currents greater than a critical value I_k, the output voltage is nearly constant at a value $2/\pi$ of the maximum value of the input voltage. It was the latter portion of the curve that was referred to when good regulation was claimed for this type of filter.

The high voltage for low currents results because, under these conditions, the filter behaves almost as if the choke were absent. The condenser voltage drops only a few volts between successive peaks of the rectified voltage wave, and only short pulses of current are required to recharge it. These pulses are so short and small in magnitude that the inductance of the choke has little effect. As the current taken by the load rises, the pulses of current are longer and larger. When the current to the load rises to the critical value I_k, the current through the choke and to the condenser is no longer a series of pulses. Instead, it is a continuous current, with a d-c and an a-c component, and the choke

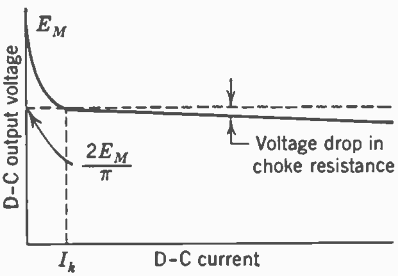


FIG. 14. Typical regulation curve for a choke-input filter.

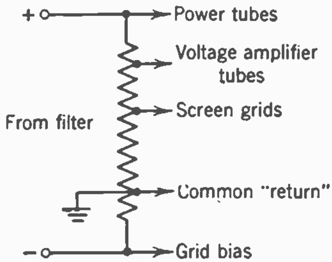


FIG. 15. Voltage divider for supplying various voltages to an amplifier.

is more effective in keeping the current constant. The output voltage drops to a lower value as shown in Fig. 14 and is nearly constant for currents larger than I_k. The gradual drop of voltage for currents greater than I_k results from voltage drop across the resistance of the choke.

If a choke-input filter is to have good regulation characteristics, a separate resistor called a **bleeder** is connected permanently across the output terminals. The size of the bleeder is selected so that it draws a current approximately equal to I_k. Then, the current supplied to another circuit is at a nearly constant voltage. The value of I_k is determined largely by the inductance of the choke. The larger the inductance, the smaller the value of I_k. The value of I_k for most designs is about $1/10$ of the total current available from the rectifier.

In addition to the advantages discussed above, a bleeder also serves to discharge the filter condensers when the power to the rectifier is switched off, and thus acts as a safety measure.

12.19. The voltage divider. A rectifier power supply is usually designed with a particular application in mind, and so the total voltage of the output is applied to the amplifier tube circuits. Sometimes, however, the power tubes require higher voltages than do the voltage amplifier tubes. Also lower voltages may be required for screen grids. Bias voltages, when required, are negative.

The screen-grid and other positive voltages lower than the total voltage may be supplied from taps on the bleeder. Bias voltage may be obtained from the negative end of the bleeder, if a tap on the bleeder is grounded as shown in Fig. 15.

12.20. Engineering the voltage divider. The lower the resistance of the bleeder (usually called a "voltage divider" when it is tapped), the better will be the regulation, but the greater

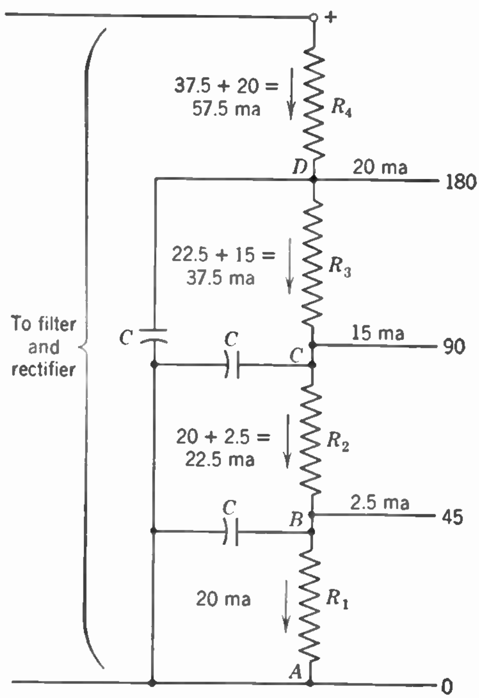


FIG. 16. Designing the voltage divider.

the load on the rectifier tube. In general, a voltage divider is engineered as follows: In Fig. 16, suppose that the current flowing through resistance R_1 is 20 ma when no current is being taken from the taps. This is "waste" current. It flows whether there are any loads connected or not. At point B, a voltage of 45 volts and a current of 2.5 ma are desired. The value of resistance is, according to Ohm's law, $E \div I = 45 \div (20 \times 10^{-3})$ or $R_1 =$ 2250 ohms. This 2.5-ma current is added to the 20 ma taken by R_1, and the sum, or 22.5 ma, flows through R_2. Since 90 volts is required at point C the resistance R_2 will be $(90 - 45) \div (22.5 \times 10^{-3})$ or 2000 ohms. If the load connected to the tap at C takes 15 ma, and the voltage at point D is to be 180 volts, the value of R_3 is $(180 - 90) \div (22.5 \times 10^{-3} + 15 \times 10^{-3})$ or 2400 ohms. The current through R_4 is 57.5 ma as shown. The value of R_4 can be determined once the total voltage across the divider is known.

The entire resistance from A to D is 6650 ohms, with taps at 2250- and 4250-ohm points. The greatest power, in this example, will be consumed in the 2400-ohm resistor R_3. Its power is $(37.5 \times 10^{-3})^2 \times 2400$ or 2.5 watts. R_2 and R_1 consume 1.01 and 0.9 watts, respectively. Thus, if the entire divider, up to R_4, is wound with wire large enough and on a frame that can dissipate the heat corresponding to 5 watts, there will be no trouble. Voltage dividers are available which consist of a single winding of resistance wire on a heat-resisting form. By means of sliders the correct voltages can be obtained easily.

Condensers across the various resistances of the divider are employed to keep the voltages from varying even though currents taken from the taps change from instant to instant.

Problem 1. A two-section choke-input filter employs two chokes, each with an inductance of 2.5 henries at a current of 200 ma, and two 8-μf capacitors. What will be the percentage ripple at a load current of 200 ma if the filter is used on the output of a full-wave rectifier supplied with 60-cycle power?

Problem 2. A filter employs a series choke which has a resistance of 120 ohms. It is connected to a rectifier with a constant direct output voltage of 250 volts. By how many volts does the output voltage of the filter drop as a result of the choke resistance as the load current increases from 10 to 30 ma?

Problem 3. A power supply has an output voltage of 250 volts. An amplifier requires the following voltages and currents: 250 volts at 40 ma, 180 volts at 10 ma, and 90 volts at 5 ma. Design a voltage divider which will furnish these voltages. Specify resistance and power rating of each section of the divider. Assume a "no-load" current of 10 ma.

Problem 4. Suppose that the voltage supplied to the voltage divider in Fig. 16 is 220 volts. What must be the resistance and power rating of R_4?

12.21. Grid-bias requirements. The steady negative voltage required by the grid of an amplifier must be "clean"; that is, it must be steady in value regardless of the alternating voltages placed upon the grid circuit. The source of bias voltage should be adequately by-passed with condensers and should have good voltage regulation for use with Class B audio amplifiers, Class B linear r-f amplifiers, and grid-bias-modulated Class C amplifiers. The regulation is unimportant with Class A amplifiers (since there is normally no grid current) or with Class C plate-modulated amplifiers.

12.22. Grid-bias sources. Three methods of supplying grid bias to an amplifier were discussed in Chapter 11. These were (1) using a C battery between the lower end of the grid-cathode input circuit and the cathode, (2) methods utilizing the voltage drop along a resistance in the filament lead of a battery-operated circuit, and (3) self-bias, in which the bias voltage is developed across a resistor in the cathode circuit through which the plate current and screen current (if any) flow. The possibility of obtaining grid bias from a tap on the rectifier voltage divider was mentioned above. There are other possibilities, all of which find application at one time or another. Five methods are shown in Fig. 17.

In some circuits the grid is forced positive at some portion of the input cycle of applied voltage. During these instants, the grid will attract electrons and, therefore, will become negative with respect to the cathode. This negative charge means that the grid has biased itself. Unless some means is provided for these electrons to leave the grid, the bias remains through the period in which the grid does not draw current, and so the next batch of electrons will merely add to the existing negative grid voltage. In this manner the bias on the grid may build up to the point where the plate current is too low or even zero. Then the tube is said to be "blocked" or cut off.

To provide an escape for these electrons, a "leak" resistance is connected from grid to cathode and a condenser is connected across this grid leak. The condenser is charged by the flow of electrons to the grid to a voltage given by the product of the grid current and the leak resistance. The condenser is large enough so that the a-c components are filtered and so that the bias will have a steady d-c value and not fluctuate up and down as the input a-c voltage varies.

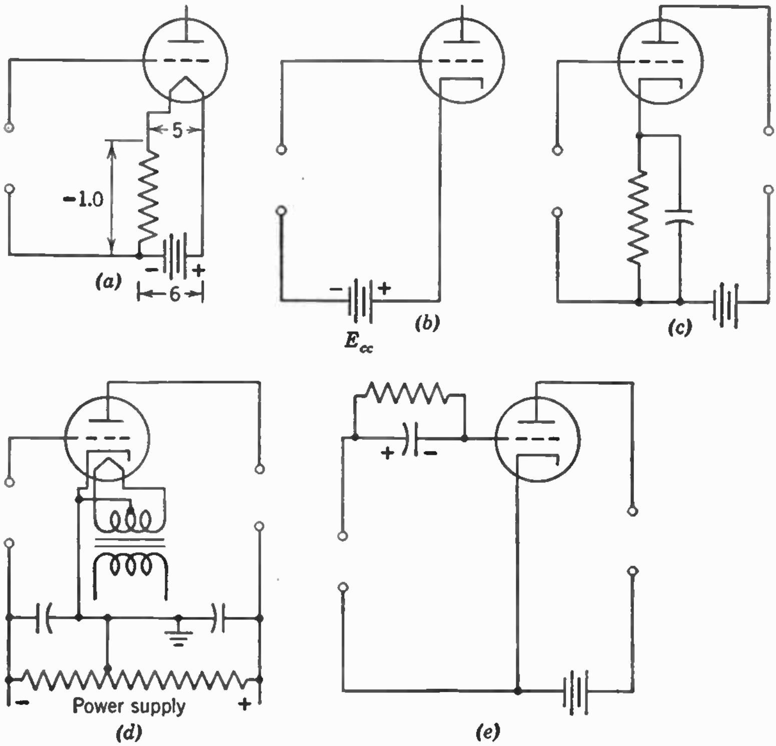


Fig. 17. Methods of supplying grid bias. (*a*) Voltage drop along a resistor in filament lead. Here the grid is 1 volt negative with respect to the negative end of the filament. (*b*) Conventional C battery. (*c*) Cathode bias produced by plate current flowing through a resistor. (*d*) Tap on power supply voltage divider. Here the grid is negative with respect to the cathode by the voltage drop to the left of the cathode connection. (*e*) Grid leak and condenser. Grid must go positive at some portion of the excitation cycle so that it collects electrons. This current produces a voltage drop along the grid-leak resistor.

This method of biasing a tube is often utilized in oscillators or in r-f amplifiers, and is commonly called **grid-leak bias.**

Amplifiers with fixed bias (C battery) and those with cathode or self-bias (cathode resistor) behave the same in most respects. However, there is one important difference. If for some reason the d-c value of the plate current should increase, the voltage drop along the cathode resistor in a self-biased circuit increases. As a result the bias increases, thus limiting the plate current. On the other hand a fixed bias, such as supplied by a battery or a separate power supply system, has no such self-regulating feature since the voltage supplied by it is generally independent of plate current.

More power at a specified maximum amount of distortion can be obtained from an audio amplifier operated with fixed bias than from one operated with cathode bias, if the same plate voltage and the same excitation are employed. This is because large input voltages produce greater average (d-c) plate currents as a result of some distortion in the waveform. These larger average or d-c plate currents increase the bias in a cathode-biased amplifier and move the operating point nearer the cut-off point. With fixed bias, this is not true. The operating point remains fixed and greater power output may be obtained without exceeding a specified amount of distortion.

12.23. Ballast tubes. A current-regulator or ballast tube consists of a metallic wire filament, which has a large temperature coefficient of resistance, immersed in a gas. If, as a result of an increase of voltage, the current through the wire tends to increase and the filament to get warmer, the resistance increases and the current actually changes very little. Such a tube is often placed in series with a load, such as a radio receiver power transformer or tube filament, so that, as the line voltage changes, the tube with its ballasting action keeps the current and voltage supplied nearly constant. For example, the B4 ballast tube passes 1.24 amp at 105 volts and 1.36 amp at 125 volts. A change of 19 per cent in voltage produces a change of only 9.7 per cent in current. This current flows through the primary of the power transformer, or other load, and has half the variation of the line voltage.

Tubes of this type are available in several voltage ranges, for example, 5 to 8, 3 to 10, 15 to 21, 40 to 60, and 105 to 125 volts. Ordinarily the load must be designed to operate at a voltage lower than the normal line voltage and consequently the voltage required by the ballast tube plus that required by the load will equal the line voltage. The tube operates slowly and cannot regulate rapid fluctuations.

12.24. Surge protector tubes. Surge protector tubes are two-element gaseous discharge tubes which are connected across a line or a load for protection against excessive voltages. When the voltage exceeds the breakdown voltage of the tube, a discharge takes place, and, as is characteristic with such discharges, the voltage across the terminals is more or less constant.

12.25. Voltage-regulator tubes. The ballast tube is a *current* regulator. It will operate successfully on both alternating and direct voltages. Another useful tube is the *voltage* regulator, which consists generally of two electrodes in a gas. A discharge occurs between the electrodes. The voltage across the discharge is approximately constant for a considerable variation in discharge current. In this characteristic the voltage regulator resembles merceury-vapor and gaseous rectifier tubes already described.

The voltage-regulator tube is placed in parallel with the load so that variations in the applied d-c voltage are smoothed out with the result that the load has a constant voltage applied to it. The tube may be employed to reduce ripple voltages or to improve regulation when the load current varies. The tube operates quickly. It is designed for a definite operating voltage, and the current through it must be between specified minimum and maximum values. A typical curve for such a tube is shown in Fig. 18.

12.26. Regulator-tube-circuit design. In Fig. 19 is shown a ballast tube connected in series with the primary of a power

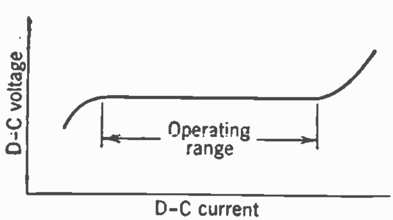


Fig. 18. Typical characteristics of a voltage-regulator tube.

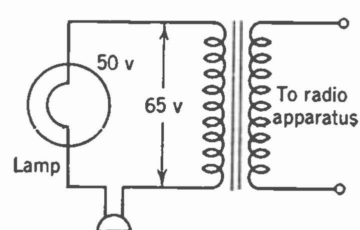


Fig. 19. Use of current regulator or ballast tube.

transformer. The circuit is designed so that 65 volts appears across the primary of the transformer, and 50 volts across the tube. When the line voltage increases, more drop occurs across the tube but the current through it stays nearly constant so that the drop across the transformer remains nearly the same. Likewise, the voltage across the transformer remains nearly constant for decreases in line voltage.

A circuit with a voltage regulator is shown in Fig. 20. The OC3/VR105 tube, which is one of several voltage-regulator tubes, maintains a voltage drop of 105 volts across its terminals for d-c currents ranging from 5 to 40 ma. More precisely, the voltage rises only 1 volt as the current changes from 5 to 30 ma; it rises 2 volts if the current goes on to 40 ma. If the load requires a

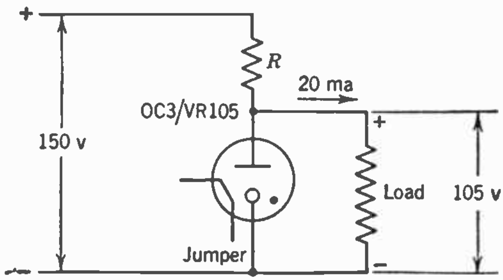


FIG. 20. Circuit using a voltage-regulator tube.

maximum current of 20 ma at 105 volts, the value of the resistance R in series with the tube may be calculated as follows. Assume, for best regulation, that the total current will be 30 ma—20 ma for the load and 10 ma for the tube. Since the voltage drop across R must be 150 − 105 or 45 volts, R is $(45 \div 30) \times 10^{-3}$ or 1500 ohms.

Since the voltage drop across the tube remains almost constant regardless of current changes, an increase in the applied terminal voltage results in an increased voltage drop across R. On the other hand, an increase in the load current above the 20-ma design value results in a decrease in the tube current, and vice versa. Thus the load voltage remains substantially constant.

Ripple on a d-c voltage represents changes in applied voltage. The effect of the action of the voltage regulator is to decrease greatly the ripple across the load. Two similar voltage-regulator tubes can be connected in series to obtain a higher regulated voltage.

The "jumper" shown in Fig. 20 is often connected in series with the power supply transformer, or in some other position in the circuit so that voltage to the load will be removed if the regulator tube is removed from its socket.

12.27. Vibrator power supplies. Radio transmitters and receivers require d-c voltages higher than can conveniently be obtained from batteries. The advent of automobile and other forms of mobile apparatus led to the development of special power supply systems for these purposes. One system involves the use of a **vibrator.**

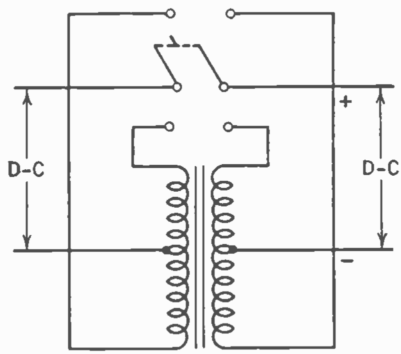


Fig. 21. Elements of vibrator power supply. As the switch is moved up and down, the direction of current flow through the windings reverses. The switch contacts on the right reverse the direction of the secondary current in synchronism with the reversals in the primary so that the polarity of the output terminals does not change.

The simplest type of vibrator power supply operates as follows: A switch periodically interrupts the flow of direct current in the primary of a transformer, the secondary of which furnishes alternating voltage of the required amplitude to a rectifier-filter system. Whenever the current in the primary is interrupted a voltage is induced in the secondary, and this voltage may be much higher than the primary voltage if a step-up transformer is used. The purpose of the rectifier-filter is to convert the secondary voltage to direct voltage and to remove the ripple as in a conventional rectifier power supply.

The principle of operation of a second type, called a **synchronous vibrator,** is illustrated in Fig. 21. The switch is arranged so that when the electrons in the secondary tend to flow upward and out of the "+" terminal, for example, instead of downward and out of the "−" terminal, the switch changes a set of contacts so that the electron current continues to flow from left to right. These two switches, the primary and the secondary, can be mechanically coupled together so that they operate together in synchronism. In this manner the rectifier tube can be eliminated. The filter, however, is still necessary.

In practice the switches are replaced by vibrating metal reeds which carry contacts and which are attracted by an electromag-

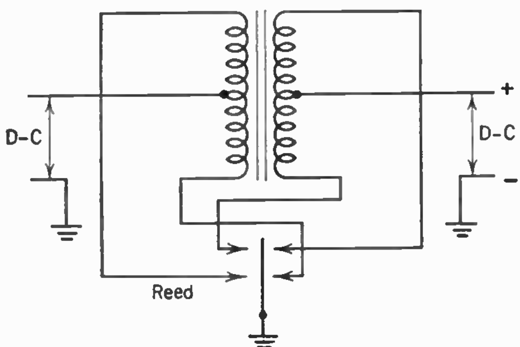


FIG. 22. Reed vibrated by electromagnet takes place of switch in Fig. 21 to produce high-voltage direct current from low-voltage direct current.

net toward one or the other of a stationary set of contacts. In this matter the desired reversals of current flow are obtained. Such a mechanical vibrating switch is called a vibrator; it can be arranged to act as a half-wave rectifier, as a full-wave rectifier, or simply as a current reverser for the primary circuit. A synchronous vibrator is shown in Fig. 22. Vibrators are available which will operate from 6 volts, as from an automobile storage battery, or from higher direct voltages. The contacts on vibrators for use on voltages much above 20 volts are generally subject to rapid wear and their life is rather short. Also, generally speaking, the higher the current through the contacts the shorter their life.

The transformer for a vibrator power supply must be designed especially for this service since the primary switch does not produce a pure sine wave of voltage. The contacts on the vibrator must be carefully designed and made and must be maintained in proper condition; thoughtful consideration must be given to the time of contact in the primary cycle of current reversal, the time of contact on the secondary side, and the proper use of condensers and resistors to produce the desired waveform. Much research has gone into the vibrator method of transforming one kind of power to another, leading to the present successful mobile radio transmitters and receivers.

A very good description of the vibrator systems will be found in the *MYE Technical Manual,* published by P. R. Mallory & Co. in 1942.

12.28. Dynamotors. Vibrator power supplies are not well suited for mobile applications in which considerable amounts of high-voltage d-c power are required. The dynamotor, however, can serve for such applications. The dynamotor consists of a d-c motor, driven from a low-voltage battery, coupled to a d-c generator which supplies a direct voltage of the required value. The motor and generator coils are placed on the same rotor, but separate commutators serve the low- and high-voltage sides. Dynamotors are made in various sizes to accommodate small receivers and rather large mobile transmitters. They are widely used in aircraft applications. Since the secondary voltage is pulsating in nature, it must be filtered in much the same manner as in a rectifier power supply.

Problem 5. An OD3/VR150 tube maintains 150 volts across its terminals for direct currents from 5 to 40 ma. A certain load requires 150 volts at 25 ma. A power supply is available which has an output voltage of 200 volts. What should be the value of the series voltage-dropping resistor? What should be its power rating? Design for a maximum total current of 40 ma.

13 · Audio Amplifiers

13.1. Classification of amplifiers. Amplifiers may be classified according to (1) frequency range, that is, audio, radio, video; (2) voltage or power amplification; (3) mode of operation (Class A, B, etc.); (4) circuit connections.

An audio amplifier may be designed to amplify all frequencies, for example, from 20 to 20,000 cycles. Sometimes it may be called upon to amplify only a limited band in this region, say 250 to 2500 cycles as for telephone communication, or a narrow band in the vicinity of 400 cycles for aircraft signaling. A radio-frequency amplifier usually handles only a narrow band of frequencies and is, therefore, said to be a frequency-selective amplifier. The band of frequencies may be centered at any frequency up to hundreds of megacycles. A video amplifier is called upon to cover a wide frequency range, perhaps several megacycles. A d-c (direct-coupled) amplifier is useful in amplifying very low-frequency signals, often so low in frequency that they are nothing more than slowly changing direct currents or voltages.

The distinction between voltage and power amplifiers was discussed in Sect. 11.12. It should be noted again that the distinction between the two types is often not very great.

13.2. Modes of operation. The mode of operation of an amplifier refers to the fraction of a complete cycle of grid voltage that the plate current flows.

A **Class A** amplifier is one in which the grid-bias and alternating grid voltages are such that the plate current flows in a tube at all times, even at the most negative portion of the negative half cycle of the signal. Under normal Class A operation the grid-bias voltage (E_{cc}) is set so that operation occurs over the linear portion of the dynamic E_c–I_b curve, or E_{cc} is set at about half the cut-off voltage [see Fig 1(a)]. The alternating grid voltage must not be large enough to drive the total instantaneous grid voltage beyond cut-off at any time during a cycle. Usually the grid is not permitted to go positive. However, this is not a requirement of Class A operation. The waveform of the plate

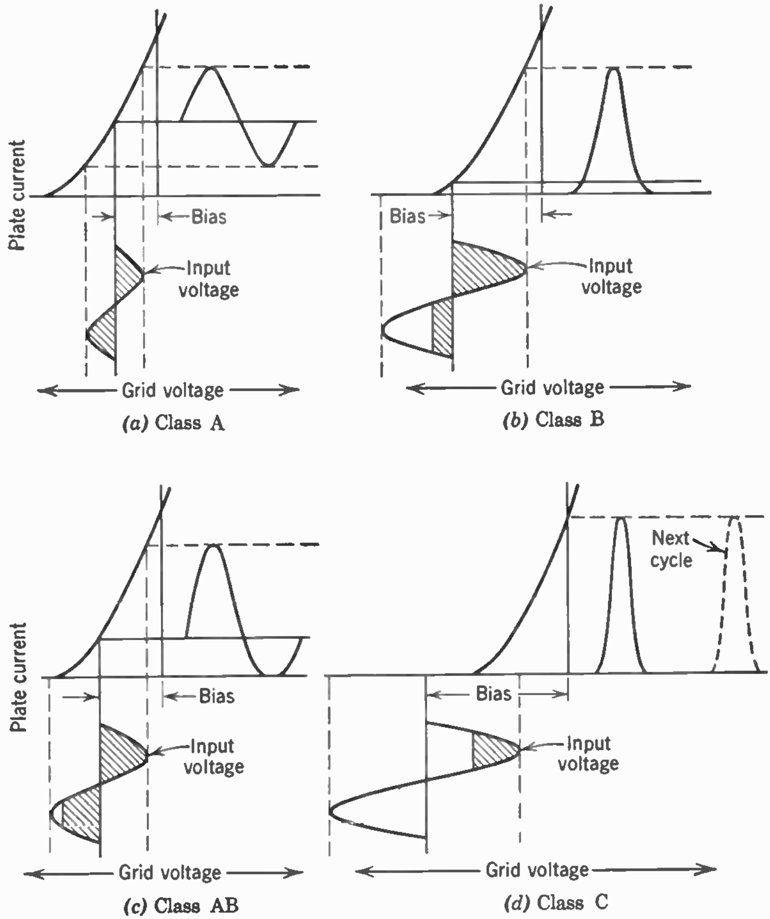


FIG. 1. Class A, B, AB, and C operation of amplifiers. Plate current flows only during shaded portion of grid voltage.

current is similar to that of the grid signal and the distortion is low. The power-converting efficiency is low, ranging in practice from 5 to 20 per cent. The voltage and power amplification are high, but the available power output is low.

A **Class B** amplifier is one in which the grid bias is approximately equal to the cut-off value so that the plate current is approximately zero for no exciting signal on the grid, and so that the plate current flows for approximately the entire positive half cycle of the alternating voltage applied to the grid. The waveform of the plate current as seen in Fig. 1(*b*) approaches closely that of a half-wave rectifier. Since nearly all the negative half cycles of plate current are absent, the distortion is high. A rather large grid-driving voltage is required if, as is often the case, the grid is driven positive. The voltage amplification is lower than with Class A, but the power output is larger. The efficiency is also higher, being of the order of 40 to 60 per cent. Class B amplifiers are not used for audio amplifiers except in push-pull arrangements, discussed later. In a push-pull circuit the distortion is fairly low.

A **Class C** amplifier is one in which the grid bias is appreciably greater than the cut-off value so that the plate current is zero for no grid signal, and so that plate current flows for appreciably less than one-half of each cycle when an alternating grid-signal voltage is applied. The efficiency may be as high as 70 to 85 per cent, and the power output is correspondingly larger than that of a Class A or B amplifier. The large amount of distortion in the output current makes such an amplifier unsuited for a-f amplification; but it is well suited for r-f amplifiers which have parallel-resonant circuits as plate loads.

A **Class AB** amplifier is one for which the grid bias is intermediate between that required for Class A and for Class B, and plate current flows for more than 180° but less than 360° of the grid-voltage cycle. Higher power output can be obtained than with Class A operation, but at the expense of somewhat more distortion. Class AB amplifiers are generally used in push-pull circuits.

The grid is usually driven so that the instantaneous voltage is positive for a part of the cycle and grid current flows in Class B and C amplifiers, whereas the grid is usually kept negative in a Class A amplifier. If it is desirable to indicate whether or not the grid is driven positive, the subscript 1 indicates that the grid never becomes positive, and the subscript 2 indicates that the grid does become positive during a part of the cycle and some grid current flows. Then, "Class A_1" means that the operation is

Class A and the grid is never driven positive. "Class B_2" indicates that the tube is operated under Class B conditions and that the grid signal drives the grid positive for a part of the cycle.

13.3. Operating conditions for audio amplifiers. For an amplifier to operate at audio frequencies it must meet the following requirements:

1. The amplification versus frequency curve must be sufficiently flat for the particular application.
2. The distortion (harmonic production) in the output must not be greater than is satisfactory or tolerable.
3. The amplification must be sufficient to deliver the required power from the given input voltage.
4. The noise and hum must be lower than some predetermined limit.
5. The amplification should not vary much with ordinary changes in filament temperature, plate voltages, etc.
6. The amplifier must work out of a certain impedance and transmit its power to another (usually different) impedance.

13.4. Multistage amplifiers. A single stage of amplification is rarely capable of satisfying the above requirements. Often it is possible to get high amplification in a single stage but at the expense of excessive distortion or a poor frequency characteristic. Therefore, means must be provided for using the output of one stage to drive another stage. Preliminary stages in a multistage **(cascade)** amplifier are designed to produce high voltage amplification with little regard for power efficiency. The final stage is usually the power amplifier and is designed to deliver the desired power to the load with as high efficiency as is practical.

13.5. Resistance-capacitance coupled amplifier. The simplest method of connecting two amplifying stages together is by means of an *RC* network. The circuit of such an **RC amplifier** is shown in Fig. 2. The signal voltage fed to the first tube is amplified and the output voltage of this tube appears across R_L. This alternating voltage is used to drive the grid of the second tube.

The purposes of the condenser C are (1) to block the d-c plate voltage of tube V_1 from the grid of tube V_2 and (2) to pass the a-c voltages to the second tube. Its capacitance should be large enough so that the flow of alternating currents is not appreciably impeded. The grid resistor R_g connects the grid to the bias battery E_{cc_2}. The size of R_g should be much larger than that of R_L, usually around ten times, so that its shunting effect is not enough to reduce the gain. That is, the equivalent effect of the two re-

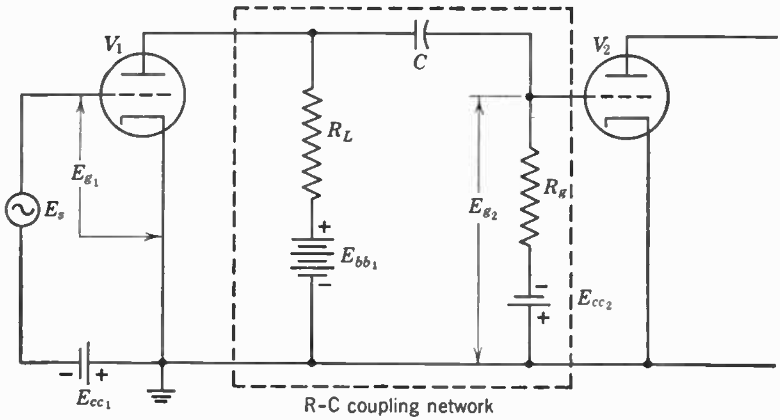


FIG. 2. Resistance-capacitance coupling of two amplifier stages.

sistances in parallel $[R_L R_g/(R_L + R_g)]$ is high so that a high-resistance load is in the plate circuit of the tube. The a-c equivalent circuit of the first tube and the coupling network is shown

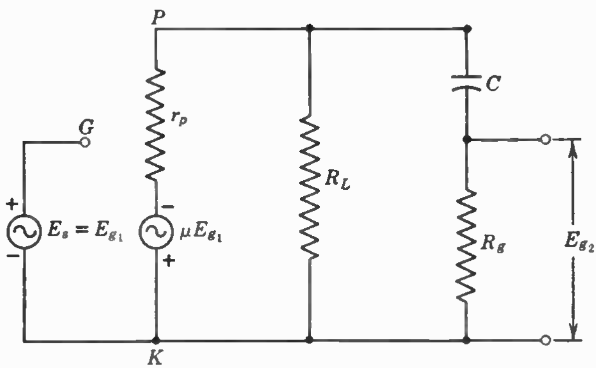


FIG. 3. A-c equivalent of circuit of Fig. 2.

in Fig. 3. It is clearly seen that the voltage across R_L is shared by C and R_g. Since the maximum possible voltage is desired across R_g, the voltage drop across C is kept low by using as large a condenser as possible. The upper limit of the size of C is fixed by two considerations: the physical size, which increases as the capacitance becomes larger, and the leakage resistance, which becomes less as C is made larger. A low leakage resistance is undesirable since it may permit direct current to leak through the condenser and establish a positive voltage on the grid of the second tube. (See Fig. 5, Chapter 6, and the accompanying discussion.) Often the condenser C discharges at an audible rate, such as one or two times per second, so that the desired signals are modulated with the voltages produced by the periodic charge and discharge of C. If this occurs at a low but audible rate, the noise set up is called "motorboating."

In a triode amplifier of the RC type, the effective load impedance in the plate circuit of the first tube—made up of R_L, X_c, R_g, and the inherent capacitances in the wiring, inside the tubes, between tube and socket terminals, etc.—should be several times the plate resistance r_p of the tube. In amplifiers using pentodes or tubes with very high plate resistance, the load impedance must simply be as high as possible or convenient.

The voltage amplification for the first stage is the ratio of the a-c voltage on the grid of the second tube to that on the grid of the first tube, or

$$A_v = \frac{E_{g_2}}{E_{g_1}}$$

$$= -\frac{\mu R_{eq}}{r_p + R_{eq}} \quad \text{(triode amplifier)}$$

$$\doteq -g_m R_{eq} \quad \text{(pentode amplifier)}$$

where R_{eq} is the total resistance of R_L and R_g in parallel. The voltage drop across C is assumed negligible. Its effect is discussed in the next section.

The gain calculated from the above expressions is called the **mid-frequency** gain. The frequency is assumed high enough so that the reactance of C is low and the a-c voltage drop across it is negligible, but not so high that the shunting effect of the wiring and tube capacitances is low enough to reduce the gain. A frequency at which X_c equals $\frac{1}{10}R_g$ can be considered the low-frequency limit of frequencies for which the mid-frequency expression applies.

It is common practice to use values of C of the order of 0.01 μf and R_g values of 0.5 to several megohms.

13.6. Low-frequency gain. At frequencies lower than those considered above, the reactance of C becomes higher and must be taken into account. Across this reactance will appear a voltage which subtracts from that which can be applied to the grid of the second tube. The amplification now depends upon the relative values of X_c and R_g (as well as R_L) since the larger X_c is compared to R_g the more voltage is lost across C and the less appears across R_g. These two elements must be taken into account together. If R_g is large, C can be small; if R_g is low, C must be large.

Actually, if X_c is equal to R_g, both expressed in ohms, the gain will be 70.7 per cent of the maximum possible when X_c can be neglected, or of the mid-frequency value. If X_c is $\frac{1}{3}R_g$, the amplification will be 95 per cent of the maximum value.

13.7. High-frequency gain. Loss in gain at high frequencies occurs as a result of the shunting effect of the tube and wiring capacitances. Of course the reactance of X_c at these frequencies is so small as to be negligible, but by the same argument the reactance of any inherent capacitances shunted across R_L or R_g may be small enough to reduce the effective load impedance into which the first tube works and thus reduce its gain.

The higher R_L and R_g, the lower the frequency at which these stray capacitances tend to reduce the gain. This means that the gain can be kept relatively constant up to rather high frequencies only if low values of R_L and R_g are used. This is accomplished at the expense of lower gain at all frequencies.

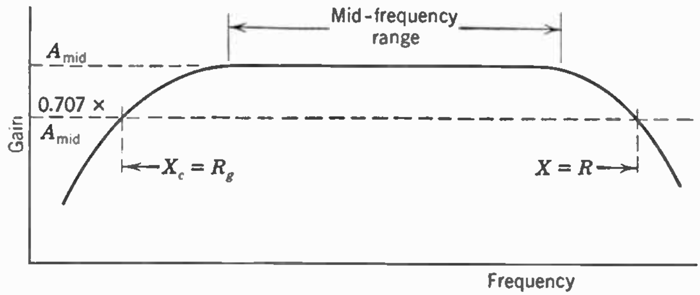
The actual gain at a high frequency (at which the reactance of the shunting capacitance becomes effective) compared to the mid-frequency gain may be found from

$$\frac{\text{Gain at high frequency}}{\text{Mid-frequency gain}} = \frac{1}{\sqrt{1 + (R/X)^2}}$$

where R is the effective resistance of the grid leak R_g, the plate load R_L, and the plate resistance r_p all in parallel, and X is the reactance of the stray capacitance shunting the circuit at the high frequency for which gain is to be calculated.

At the frequency at which the stray shunting capacitances have a reactance equal to three times the value in ohms of the three resistances in parallel, the amplification will be 95 per cent of the mid-frequency value. If the frequency is raised until the reactance is equal to the effective resistance of the circuit, the amplification will be 70.7 per cent of the mid-frequency value.

A typical curve of gain plotted against frequency for an *RC* amplifier is shown in Fig. 4. The number of cycles included in



X = reactance of stray wiring and tube capacitance
R = resistance of R_L, R_g, and r_p in parallel

FIG. 4. Typical frequency-response curve of an *RC* amplifier.

the mid-frequency range is, of course, determined by the frequencies at which the series capacitance C and the shunting stray capacitances become effective in reducing the gain.

13.8. Over-all amplification. The gain per amplifier stage has already been defined as the ratio of the voltage supplied to the grid of the driven tube (or load, if the last tube is considered) to the voltage on the grid of the driving tube. Thus, the output of one stage is the input to the next, and the gain of several stages is simply the product of the gain of the individual stages. If an amplifier has three stages which have gains of 20, 10, and 5, then the over-all amplification is $20 \times 10 \times 5$ or 1000. That is, the output voltage will be 1000 times the voltage on the grid of the first tube. The voltage gain of any number of stages can be computed in a similar manner.

13.9. Plate voltage requirements. One of the objections to the use of *RC* coupling in triode amplifiers is that rather high d-c plate supply voltages are required. The condition that makes the gain high, namely, that the load resistance R_L should be much higher than the plate resistance of the tube, also causes a large part of the d-c voltage to be lost across this load resistance. Consequently, the plate supply voltage must be quite high if the proper d-c voltage is to appear across the tube. Suppose a triode amplifier is to be operated at a plate voltage E_b of 100 volts, and that the static plate resistance is 2000 ohms and the plate current is 2 ma at this voltage and with a specified grid-bias voltage. A plate load resistance of 50,000 ohms is to be used. The plate supply voltage then must be

$$E_{bb} = E_b + I_bR_L = 100 + 2 \times 10^{-3} \times 50{,}000$$
$$= 100 + 100 = 200 \text{ volts}$$

To get 100 volts on the plate of this tube requires a plate supply of 200 volts! What if the tube is operated at a lower voltage? Then the plate resistance of the tube is increased, and the load resistance R_L must be made larger if the gain is to be kept the same. The result is that nearly the same plate supply voltage is required as in the first case.

In spite of this disadvantage, *RC* coupling is widely used in voltage amplifiers because of the excellent frequency response obtainable and because of the low cost of the condenser and resistors which make up the coupling network.

In amplifiers employing pentodes and other high-resistance tubes the proportionate loss of voltage in the plate load resistor is not so great since the load resistor is generally much smaller in ohms than the d-c resistance of the tube.

13.10. Impedance-coupled amplifier. An amplifier in which the plate load resistor is replaced with a low-resistance inductor having a high value of reactance at the frequencies at which the amplifier is to work is called an **impedance-coupled amplifier.** Actually, any interstage coupling network is an impedance network, but the term has, by common usage, been employed to describe the network shown in Fig. 5.

Suppose the choke in Fig. 5 has an inductance of 100 henries and a d-c resistance of 1000 ohms. Now the direct plate current encounters relatively little opposition in the 1000-ohm resistance. The alternating current, however, must flow through the high inductive reactance and the resistance in series. Across this impedance, $\sqrt{R^2 + \omega^2 L^2}$, will appear the amplified a-c voltage which may be used to drive another amplifier stage.

The loss of d-c voltage between the plate supply and the plate is 1 volt per ma of d-c plate current, and if the d-c plate current is 2 ma, only 102 volts of supply voltage will be required to put

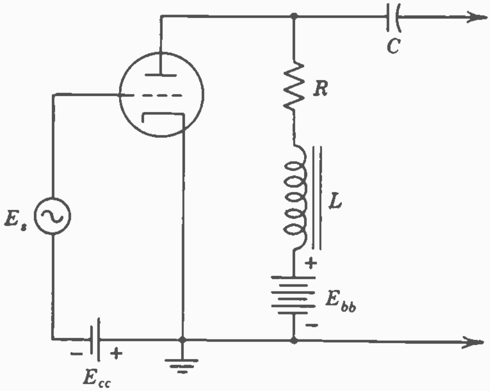


FIG. 5. Elements of impedance-coupled amplifier.

100 volts on the plate of the tube. In modern amplifiers, impedance coupling is rarely used except for power tubes having high plate current. Since the reactance is a function of frequency, the gain also varies with frequency and the frequency-response curve of an impedance-coupled amplifier is decidedly inferior to that of an amplifier employing *RC* coupling.

Problem 1. What is the mid-frequency gain of an *RC* amplifier, given the following: $\mu = 20$, $r_p = 7700$ ohms, $R_L = 25{,}000$ ohms, $R_g = 0.5$ megohm? What will be the gain if $R_g = 25{,}000$ ohms?

Problem 2. If the coupling condenser C in Prob. 1 is 0.01 μf, what is the lowest frequency for which 95 per cent of mid-frequency gain will be obtained? Work for both values of R_g.

Problem 3. The mid-frequency gain of a 6SJ7 pentode amplifier working into a plate load of 100,000 ohms and a grid leak (R_g) of ½ megohm is 104. The plate resistance is 0.9 megohm. Stray capacitance plus tube capacitances adds up to 30 $\mu\mu$f. What is the voltage gain at 50 kc? at 1 Mc?

Problem 4. A type 6SF5 high-mu triode passes 0.3 ma when the plate-cathode voltage is 200 volts and the grid bias is −2 volts. If the tube is to be operated at the above operating point and with a 200,000-ohm plate load resistance, what must be the power supply voltage?

Problem 5. A tube has a μ of 12 and an r_p of 12,000 ohms. It is used in an impedance-coupled amplifier in which the plate choke is 20 henries and the resistance is 1000 ohms. What is the voltage amplification at 50 cycles? at 200 cycles?

Problem 6. A tube draws 1.5 ma through a 50,000-ohm plate load resistor. If the plate supply voltage is 250 volts, what is the voltage across the tube?

13.11. Miller effect. Not only do the tube and wiring capacitances play havoc with amplifier gain at high frequencies, but the effect of the grid-plate capacitance of the second tube in a circuit such as in Fig. 2 is multiplied by the gain of that tube—a very interesting and unfortunate fact.

The capacitances which cause trouble at high frequencies are:

1. Grid-cathode capacitance, C_{gk}.
2. Plate-cathode capacitance, C_{pk}.
3. Grid-plate capacitance, C_{gp}.
4. Stray capacitances in wiring, etc.

The capacitance C_g across the input to a tube is nearly equal to

$$C_g = C_{gk} + C_{gp}(A_v + 1)$$

where A_v is taken as the *magnitude* of the voltage amplification. This increase in effective input capacitance is due to the **Miller effect.** The various interelectrode capacitances of a tube are shown in Fig. 6.

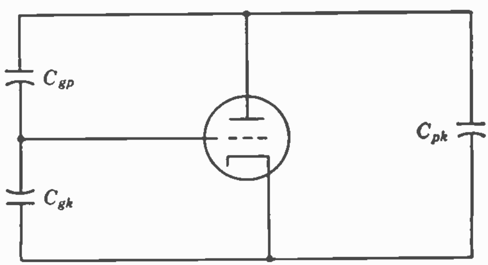


FIG. 6. Tube interelectrode capacitances.

A 6J5 triode has the following interelectrode capacitances: $C_{gk} = 3.4\ \mu\mu f$, $C_{pk} = 3.6\ \mu\mu f$, and $C_{gp} = 3.4\ \mu\mu f$. The amplification factor, μ, of a 6J5 is 20. Suppose that in a particular circuit the voltage gain realized is 15. Then

$$C_g = 3.4 + 3.4(15 + 1) = 3.4 + (3.4 \times 16)$$
$$= 57.8\ \mu\mu f$$

The value of C_g is rather high—much higher than would be expected—and it reduces the gain of an amplifier at high frequencies.

Since the increase of the input capacitance depends on the gain of the tube, the increase will be lower if the gain is made lower, as by using a smaller plate load resistor or by substituting a tube with a lower μ. Such changes unfortunately result in lower amplification, and more tubes or other means must be used to return the amplification to the desired value.

The effect of the screen grid in tetrodes and pentodes is to reduce C_{gp} to a very low value, and so the input capacitance of these types of tubes is not changed much by the Miller effect. The grid-plate capacitance of a 6SJ7, a typical voltage-amplifier pentode, is only 0.005 $\mu\mu$f.

Problem 7. A tube with a μ of 30 and an r_p of 50,000 ohms is used with a 200,000-ohm load resistance and a 0.5-megohm grid-leak resistance in the RC coupling network. Suppose that in construction a path of soldering flux is placed across the grid-leak resistor so that the half megohm is effectively shunted by 100,000 ohms. What is the resulting voltage amplification? What is the amplification without the shunting resistance of the flux? What is the percentage loss due to the flux?

Problem 8. The grid-plate capacitance of a 2A3 power triode is 16.5 $\mu\mu$f and the grid-filament capacitance is 7.5 $\mu\mu$f. What is the effective input capacitance of the tube when used in a circuit in which it has a gain of 3?

13.12. Practical high-gain amplifiers. The use of pentodes makes it relatively simple to obtain high voltage gains in single- or two-stage RC coupled amplifiers. Since the gain of a single stage may be 100 or more and since the gain of several stages is the product of the individual stage gains, two tubes and the appropriate coupling system will produce a voltage amplification of 100×100 or 10,000. As much amplification can be secured with a single pentode amplifier as is possible with two stages of triode amplification, with greater simplicity of apparatus and much saving of space.

A typical stage of pentode amplification is shown in Fig. 7. Higher gains will be obtained with higher values of R_L, but the gain will begin to fall off sooner at the high-frequency end. The value of the screen-dropping resistor R_d is chosen so that the d-c voltage applied to the screen has its proper value. The screen voltage will be the plate supply voltage E_{bb} minus the voltage drop across R_d; $(E_{bb} - I_{sg}R_d)$. The screen by-pass capacitor C_d keeps any a-c voltage from developing between screen to ground.

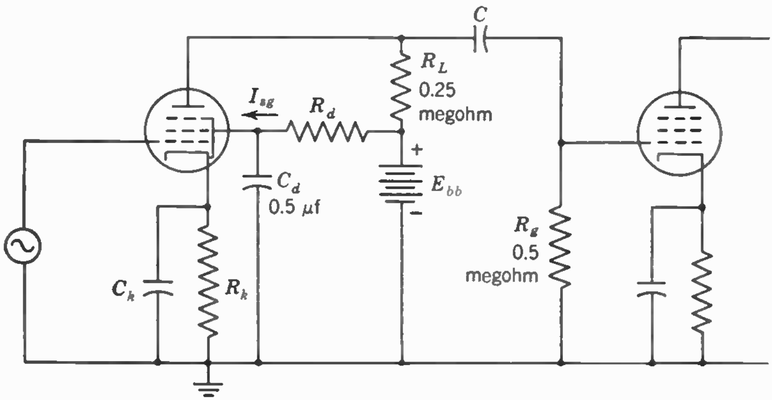


FIG. 7. Typical RC amplifier using pentode tube.

13.13. Amplifier instability. The larger the coupling capacitor, the better the low-frequency response in an RC amplifier. Some of the difficulties of attempting to extend the low-frequency response by this method have already been discussed. The most serious difficulty is the possibility that the grid of the following stage may collect enough electrons to cause the bias to change and possibly to cause "motorboating." This can be prevented if the charge can leak off to ground rapidly enough. The leakage path is through R_g. If the product of C and R_g is so great as to cause appreciable time for this charge to leak off, the second tube may "block"; that is, the plate current may be cut off as a result of the grid first being driven positive so that grid current flows, and then highly negative as a result of the resulting voltage drop across R_g. A practical value for the RC product is 0.05, and therefore there is a practical limit to the low-frequency response that is compatible with high gain and good stability. Mica capacitors have higher leakage resistance than paper capacitors, and the RC product using mica types can be somewhat higher. However, there is again the practical limit of physical size since mica capacitors are generally much larger than paper capacitors having the same capacitance value.

If two or more stages secure their plate and screen voltages from a common source of positive voltage, as from a rectifier-filter system, care must be taken that each of these individual circuits is filtered so that none of the a-c signal currents flow through the power supply system. If the a-c impedance of the power supply is extremely low so that no a-c voltages are developed across it, the additional filtering is not required. However, this a-c impedance is not low enough in most cases. The

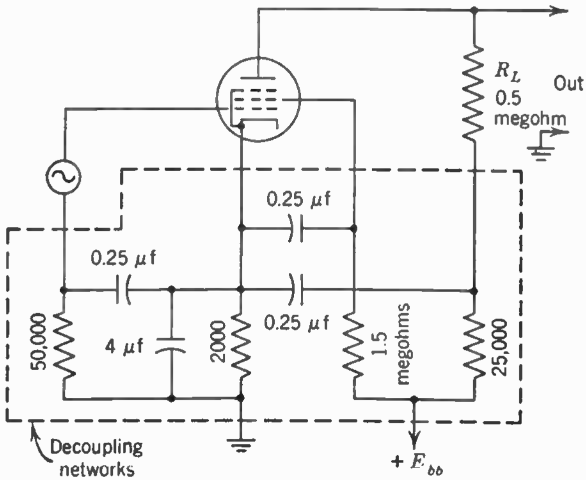


Fig. 8. Shunt capacitors and series resistors in plate, screen, and grid leads reduce amplifier instability.

filtering of the individual circuits is accomplished by connecting by-pass condensers to the cathode directly from the screens and from the plate supply end of the plate load resistors. Often, additional series resistors or chokes are inserted in the plate and screen leads to aid further in keeping the a-c voltages out of the power supply system as seen in Fig. 8. If the filtering is inadequate, the a-c voltages developed across the power supply may be reintroduced into the plate circuit of an early stage of the amplifier and cause instability in the form of a very low-frequency oscillation. The filtering networks described above are often called **decoupling** networks.

13.14. Transformer-coupled amplifier. The a-c output voltage of an amplifier employing a resistance or impedance load can never be greater than the μ of the tube multiplied by the grid-input voltage, and it approaches this value only when the impedance of the load is very much higher than the plate resistance of the tube. Suppose, however, that a transformer is used as the coupling network as shown in Fig. 9. The a-c voltage across the secondary will be increased over that across the primary by the step-up turns ratio of the windings, and so the a-c voltage developed in the plate circuit may not only be passed on to the following tube, but may also be multiplied by the transformer. Since the a-c impedance measured at the primary terminals,

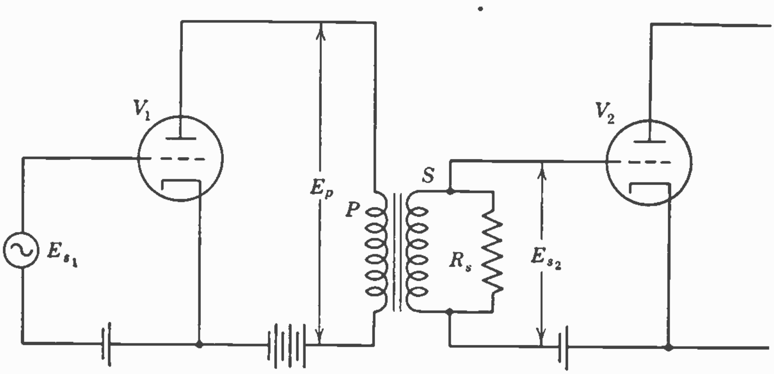


FIG. 9. Transformer-coupled amplifier.

which is the load for the tube, can be made very high, the total amplification can be made to approach μN, where μ is the amplification factor of the tube and N is the turns ratio, that is, the number of turns on the secondary divided by the number on the primary.

If the secondary circuit takes no current or power the greatest voltage will appear across the secondary when a very high turns ratio is used, but this is not true when power is taken—and some always is. The upper value of turns ratio is also limited by the frequency response of the transformer which usually becomes worse as the turns ratio is increased. A practical step-up value is 2.5 or 3 times.

The maximum power in the secondary circuit will be obtained when the turns ratio is given by the expression

$$N^2 = \left(\frac{n_s}{n_p}\right)^2 = \frac{R_s}{R_p}$$

where R_p and R_s are the resistances between which the transformer works—R_p is the dynamic plate resistance of the tube, r_p, in this case. With such a turns ratio the voltage across the secondary is given by

$$E_{s_2} = \frac{\mu_1 E_{g_1} N}{2} \tag{1}$$

If the resistance of the load across the secondary is 0.5 megohm, and the plate resistance of the tube out of which the transformer works is 20,000 ohms, the proper turns ratio is

$$N = \sqrt{\frac{500{,}000}{20{,}000}} = \sqrt{25} = 5$$

and the voltage gain from equation 1 is

$$\frac{E_{s_2}}{E_{s_1}} = \mu_1 \times 5 = 60 \quad \text{if } \mu_1 = 12$$

Often the circuit is operated with no physical resistor in the secondary. Then the value of R_s is the input impedance of the second tube and may be several megohms. Then the gain is approximately $\mu_1 \times N$.

The above discussion assumes that the transformer is perfect; that is, it has no d-c resistance, no magnetic leakage, and infinite primary and secondary reactances. The values of gain calculated above are somewhat higher than would be obtained with an actual transformer.

Transformer coupling is not widely used in modern voltage-amplifier stages since *RC* coupling gives much better frequency response, and adequate gain can be obtained with pentodes. Another factor is the cost of a transformer, which is higher than the components of an *RC* coupling network.

Not only can the transformer be used to step up the voltages as described above, but also to step the voltage down to match a fairly high tube resistance to a lower secondary resistance. It is in this service that transformer coupling finds its widest application in modern amplifiers. The low-resistance secondary load may be an amplifier which draws grid current, or it may be the voice coil of a loud speaker. The amplifier feeding into the transformer is then generally classed as a power amplifier.

Transformers are often a necessity when an amplifier is coupled to its load since the load is almost always much lower in impedance than that of the output tube or tubes. These are known as **output transformers** and are designed to connect a single power tube to a loud speaker or to a 500-ohm line, or push-pull tubes to a speaker or line. The transformers may have considerable direct current in the primary and must cover a wide frequency range with high alternating currents in both primary and secondary. It is not an easy matter to make a good output transformer, especially of the single-tube type.

13.15. Advantages and disadvantages of transformer coupling. The advantages of transformer coupling have already been mentioned. They include the ability to obtain a voltage gain higher than the amplification factor of the tube, and the possibility of matching to either a high- or a low-impedance load. Another important advantage is that all the d-c plate supply voltage, except a few volts lost across the d-c resistance of the transformer primary, is supplied to the plate of the tube.

The chief disadvantages, other than the iron losses and losses in the resistances of the primary and secondary, are the rather poor frequency response and the fact that transformers are subject to hum pickup unless carefully shielded. The poor frequency response at low frequencies results from the low reactance of the primary winding. This reactance, of course, decreases as the frequency is lowered. The distributed capacitance of the transformer windings and of the tubes resonates with the leakage inductance of the transformer at some higher frequency to cause a peak in gain, followed by a rapid decrease as the frequency is raised to still higher values. This peak occurs as low as 5000 cycles for some transformers. A typical frequency response is shown in Fig. 10. By careful design the useful frequency range can be extended to cover most of the a-f band and hum pickup can be reduced, but the resulting transformer is rather expensive.

Problem 9. A power tube has an r_p of 1000 ohms and a μ of 8. It is to be coupled to a 10-ohm speaker by means of an output transformer. What must be the turns ratio if the effective load on the tube is to be 2000 ohms? What will be the power output if the grid-signal voltage is 20 volts? Assume that there are no power losses in the transformer.

Problem 10. In Prob. 9 what turns ratio should be used to transfer maximum power to the load? What will be the power in this case?

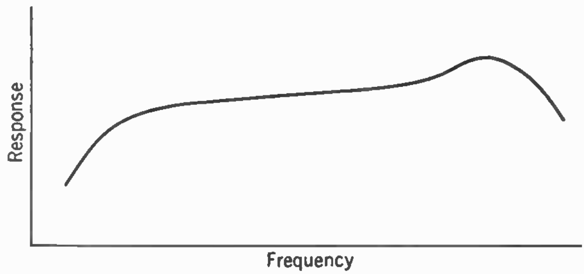


FIG. 10. Typical frequency response of a transformer-coupled amplifier.

Problem 11. A tube whose r_p is 10,000 ohms works into a load resistance of 360,000 ohms. What is the proper turns ratio of the output transformer for maximum power output? What is the voltage gain if $\mu = 20$?

Problem 12. Repeat Prob. 11 if the tube works into a load resistance of 2500 ohms.

Problem 13. A 6SJ7 pentode has a g_m of 1575 micromhos and an r_p of 0.7 megohm when the plate voltage and screen voltage are both 100 volts, the grid bias is -3 volts, and the suppressor grid is 0 volt. The plate current is 2.9 ma and the screen current is 0.9 ma. A 100,000-ohm plate load resistor is to be used.

(*a*) What is the voltage amplification?

(*b*) What plate supply voltage is required?

(*c*) If the screen is supplied its voltage through a resistor connected to the plate supply voltage source, what must be the size of the screen-dropping resistor?

(*d*) What must be the size of the cathode resistor if self-bias is used? (Both screen and plate current flow through this resistor.)

13.16. Load lines for transformer-coupled amplifiers. The d-c load line for a transformer-coupled amplifier is drawn for the d-c resistance of the primary and, because of the low value of this resistance, is nearly a vertical line as shown in Fig. 11. It is drawn vertically upward from the value of E_{bb} until it intersects the grid-bias line corresponding to the value of E_{cc}. This point of intersection is the operating point P. The a-c load line must be drawn through P. It is drawn for a value of resistance corresponding to

$$R_p = R_s\left(\frac{n_p}{n_s}\right)^2$$

This a-c load line may be immediately drawn through the operating point, but it is usually easier to draw an auxiliary load line and then draw the actual load line through the operating point and parallel to this auxiliary load line.*

Power output and distortion calculations are made in the

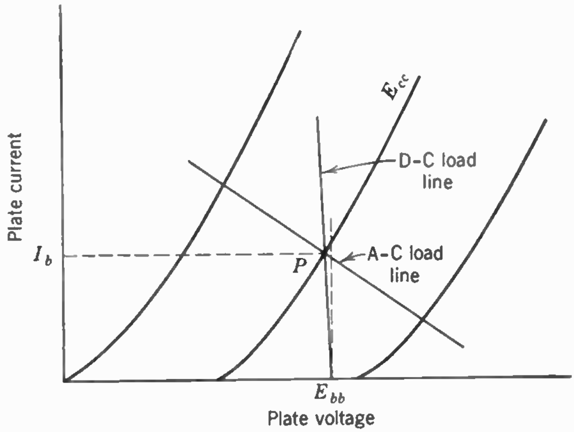


FIG. 11. D-c and a-c load lines for a transformer-coupled amplifier.

same manner as described in Chapter 11. It must be kept in mind, however, that values for the calculations must be taken from the a-c load line. A typical curve of power output and dis-

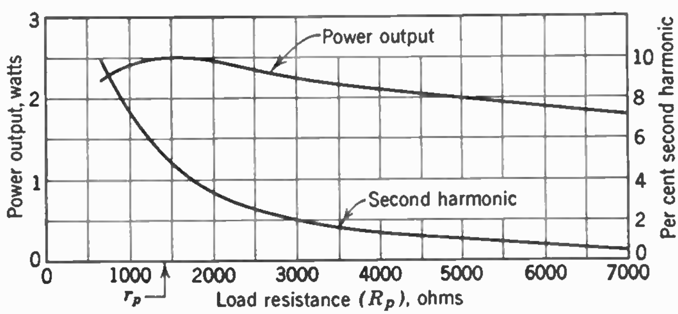


FIG. 12. Power output and second harmonic distortion plotted against load resistance for a typical triode transformer-coupled amplifier.

tortion plotted against the load resistance R_p for a triode amplifier is shown in Fig. 12. Note that the power output is maximum when R_p equals r_p but that the distortion is very high. Since the

* The general details of the construction of load lines are found in Chapter 11.

distortion decreases much more rapidly than does the power output, a load resistance R_p about twice the value of r_p gives nearly maximum power output at a fairly low distortion. If less distortion is desired, the value of R_p may be increased still further (by changing the turns ratio or increasing R_s). The lower distortion is achieved at the expense of power output. There is no such thing as "undistorted" power output from a power amplifier, although the distortion may be made very low.

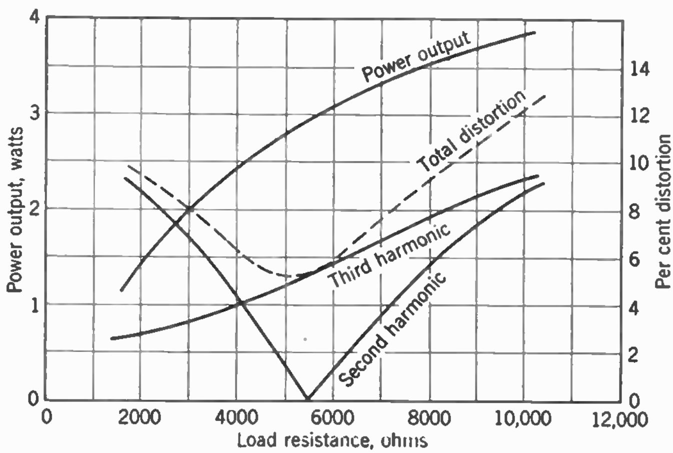


FIG. 13. Effect of load resistance on distortion and power output in a typical pentode or beam power tube transformer-coupled amplifier.

The a-c load resistance for a pentode amplifier employing transformer coupling is much more critical than that for a triode. A typical curve of power output and distortion is shown in Fig. 13. As the load resistance is increased the distortion falls to a minimum value, then increases instead of continuing to decrease as in a triode amplifier. The power output is not maximum at the value of load resistance that gives minimum distortion. As a matter of fact, the load resistance which would yield maximum power output would also give an intolerably large amount of distortion. Thus, pentodes (also beam power tubes) are much more critical in their requirements as to the impedance into which they should work. The recommended operating conditions should be referred to when designing a power amplifier using a particular pentode or beam power tube.

13.17. Push-pull amplifier. Where considerable audio output is desired with good tone fidelity, as in high-quality receivers or in public-address systems, two power-amplifier tubes are connected in a **push-pull** circuit as shown in Fig. 14. The advantages of this circuit are several, as outlined below.

Distortion due to second harmonics is reduced to a low value compared to that from a single-tube circuit; the usable power output is more than twice that obtainable with a single tube; because of the elimination of troublesome second harmonics the input signal voltage can be raised somewhat with correspondingly greater output; equal values of direct current flow in opposite directions in the two halves of the output transformer primary, which lessens saturation and decreases the amount of iron required; hum voltages originating from the power supply are canceled and therefore less filtering of the power supply is necessary.

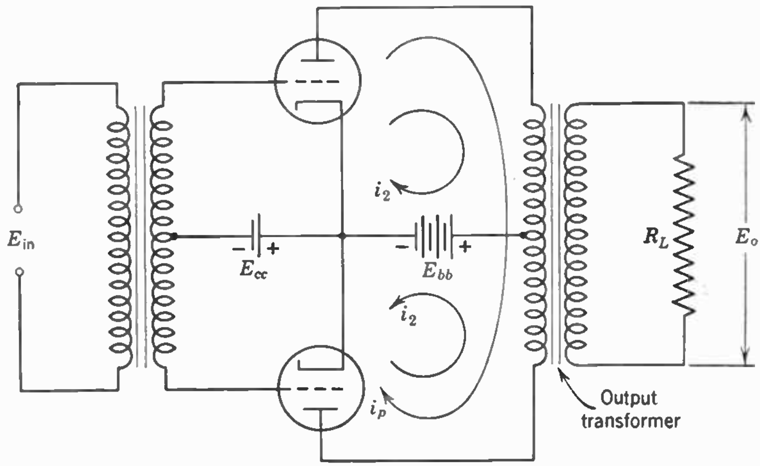


Fig. 14. Push-pull amplifier.

Push-pull operation is not limited to any particular type of power tube. Either low-μ triodes of the 2A3 type, or pentodes and beam power tubes like the 6F6 and 6L6, may be operated in this manner. Frequently a combination of push-pull and parallel operation of tubes is utilized; two tubes in parallel are on each side of the amplifier, making four tubes in all. The two parallel tubes give twice the output of a single tube and the push-pull connection multiplies the usable output by another factor of two or more. Therefore a power output of greater than four times that obtainable from a single tube of the same type is secured. The same result could be obtained by substituting a tube of four times the power rating of the original single tube, but such tubes usually require high plate voltages, are large, and increase the expense of the power supply considerably.

The reason for the name "push-pull" can be seen by referring to Fig. 14. When the grid of the upper tube is at its most positive alternating voltage value, the grid of the lower tube has its most negative value. This is because the polarity at the opposite ends of the center-tapped input transformer must be opposite. Then the upper tube has its maximum plate current at the same time that the lower tube has its minimum plate current. A half cycle later these conditions are reversed. Thus, when the plate current of the upper tube is "pushed" to its maximum value, that in the lower tube is "pulled" to its minimum value, and so on.

The current to each tube flows through its half of the center-tapped primary of the output transformer. When the current to the upper tube is increasing, the current to the lower tube is decreasing. The increasing current to the upper tube induces a voltage in the secondary of the transformer. The *decreasing* current to the lower tube also induces a voltage in the secondary of the transformer, and, since this current is flowing in a direction *opposite* to that of the current in the upper tube, the induced voltage in the secondary is in the *same* direction. In effect, the *sum* of the currents to the two tubes induces a secondary voltage E_o.

The current drawn from E_{bb} by one tube is increasing at the same time that the current drawn by the other tube is decreasing. These increasing and decreasing currents exactly cancel each other, and the net a-c current drawn from the battery is zero. The result is the same as though the a-c component of current flowed alternately from the plate of the top tube into the plate of the bottom tube, and vice versa. This is the current i_p shown in the diagram. Note that no net a-c current flows from E_{bb}. Each tube, of course, takes direct current from E_{bb}.

Second harmonic currents, i_2's (also fourth, sixth, and other even harmonics), generated within the tubes because of distortion have directions at one instant as shown on the diagram. Note that these two currents, which are equal in magnitude since the tubes are assumed identical, flow in *opposite* directions in their respective halves of the transformer and therefore induce equal and opposite voltages, really *zero* voltage, in the secondary. These even harmonic currents flow together through the E_{bb} line.

Since the second harmonic currents do not appear in the output—and these are the most prominent in a triode amplifier—the load resistance per tube can be adjusted to the value which gives maximum power output. Consequently the usable power output per tube is greater than for a single-tube circuit, for which the load resistance is usually two or three times the value for maximum power output.

The impedance measured across the total primary winding of the output transformer should be twice the dynamic plate resistance of a single tube at the operating point in order to obtain maximum power output. This results in each tube being loaded with a resistance equal to its own plate resistance, which is the condition for maximum power output. The equation is

$$R_{pp} = 2r_p = \left(\frac{n_p}{n_s}\right)^2 R_L$$

in which n_p is the number of turns on the *entire* primary winding, n_s is the number of secondary turns, and R_{pp} is the total impedance seen across the outside terminals of the primary.

The push-pull amplifier, then, is a device for eliminating the even-harmonic distortion which occurs when tubes are worked too far down on the curved part of their characteristic. Since the distortion is less, the tubes can be worked harder, having greater input voltages impressed on them, with consequently greater output.

Cancellation of even harmonic distortion is not achieved unless the two tubes are identical. Even two tubes of the same type may have somewhat different characteristics, and means are often provided in push-pull circuits to adjust the grid bias of the two tubes independently. Usually, if the grid bias is adjusted so that the d-c plate currents to the two tubes are equal, the tubes will be properly "balanced" for satisfactory operation.

Occasionally two tubes are connected in **push-push.** This connection is the same as push-pull except that the two grids are driven in phase instead of 180° out of phase. The result is that the signal voltage is canceled in the transformer and only the even-harmonic distortion voltages are transferred to the secondary or output side. Such an arrangement is obviously no good for normal audio amplifiers, but has a few special uses.

13.18. Class B amplification. Suppose that the grid bias in Fig. 14 is increased so that the plate current is near zero for no a-c grid signal; that is, the bias of each tube is adjusted for Class

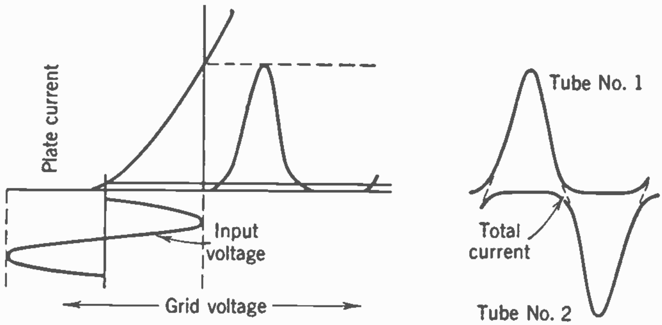


Fig. 15. Plate and total currents in a Class B push-pull amplifier.

B operation. Then there will be very little plate-current flow until an alternating voltage is applied to the grids. Even then the plate current flows only when the alternating grid voltage of a tube is on its positive half cycle. Since the grid voltage supplied to the two tubes in push-pull is 180° out of phase, first one tube conducts, then the other. The resulting currents may be added graphically as seen in Fig. 15 to obtain the total current in the primary winding of the output transformer. Note that the curvature at the lower ends of each current pulse is largely canceled in the addition process. However, there is still some distortion in the composite wave. It is considerably greater than in a well-designed Class A amplifier.

The grids of a Class B amplifier are often driven hard enough that they become positive with respect to their cathodes for a part of the cycle. This is known as Class B_2 operation. Now the grid circuit requires appreciable power which must be supplied by the preceding stage. The amplifier which supplies power to the grids is called a **driver.**

A Class B push-pull amplifier is considerably more efficient than a Class A amplifier using the same tubes. Much higher power output can be secured with a medium value of plate voltage. The distortion is fairly high. The two tubes must be well matched, and the bias voltage must be well regulated.

For Class B amplification special triode tubes are available which have a high amplification factor and operate at zero bias. With no grid excitation, and with zero bias, the plate current is very low. As the grid of one of these tubes is driven positive the plate current increases much as in any Class B amplifier. In fact, a **zero-bias Class B amplifier tube** can be considered to be one that is designed so that the cut-off bias is approximately zero rather than several volts negative as it is for many tubes. The most important advantage of this type of tube is that a well-regulated grid-bias source is not required.

A typical zero-bias Class B tube is the 6N7, which includes two triodes in the same envelope. With 300 plate volts and zero grid bias the plate current of one of these triodes is 17.5 ma. In a push-pull circuit, a plate-to-plate load resistance of 8000 ohms, and a peak-to-peak grid voltage of 58 volts (peak voltage per grid of 29 volts) the maximum plate current per tube is 97 ma. The total direct current taken from the power supply by both tubes is 35 ma with no excitation and is 70 ma for full excitation. Under these conditions the power output is 10 watts at 4 per cent distortion. The peak grid current is 20 ma per tube. Figure 16 shows the operation characteristics of a 6N7 in Class B service.

13.19. Phase inverters. A push-pull amplifier requires that the two tubes have their grids excited 180° out of phase with each other, the voltage on one increasing while the other decreases. It is easy to feed such an amplifier from a single-ended amplifier by means of an input transformer since the secondary can be divided in the center as shown in Fig. 14. If it is desired to eliminate the input transformer and use RC coupling, then phase-inversion circuits must be used.

A two-tube **phase-inverter** circuit is shown in Fig. 17. The input signal to the phase inverter is fed to the grid of tube A. This tube amplifies the voltage in the normal manner and supplies its output voltage to one of the push-pull grids at G_1. The grid-leak resistor R_{g_1} is tapped and a part of the voltage is fed to the grid of tube B. The output of this tube is fed to the other

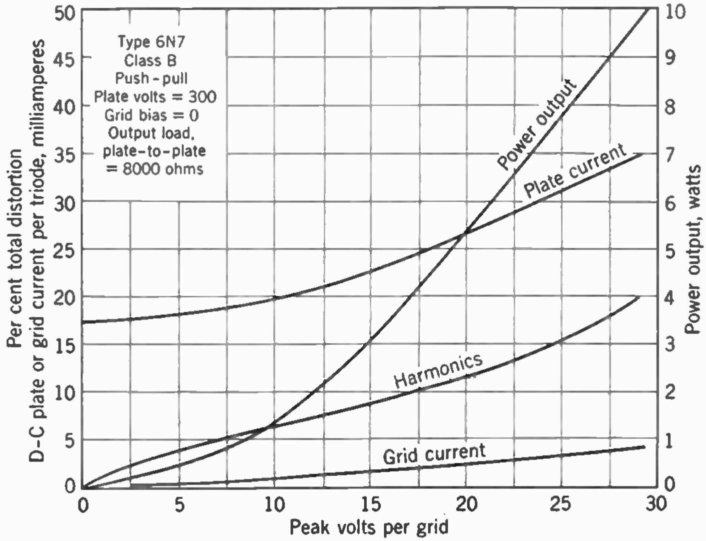


FIG. 16. Operating characteristics of a 6N7 Class B push-pull amplifier.

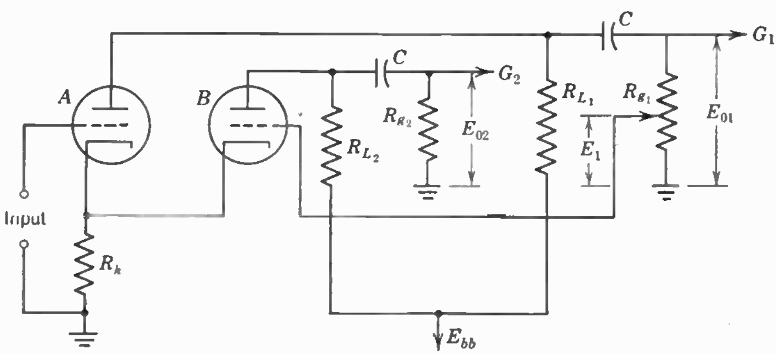


FIG. 17. A phase-inverter circuit used to drive a push-pull amplifier.

push-pull grid at G_2. Tubes A and B usually employ similar coupling networks, and so their gain is the same. This requires that the value of the voltage fed to the grid of B must be the same as that fed to A, or that the total output of A divided by the voltage fed to the grid of B, that is, $E_{o_1} \div E_1$, must equal the gain of tube A (also of tube B since the gains are equal). The phase shift through tube A is 180° as discussed in previous sections. The voltage fed to B is shifted another 180°, or a total of 360° in respect to the original input voltage. As a result, the voltages supplied to the push-pull grids are 180° out of phase, as they must be for proper operation.

13.20. Parallel tubes in amplifiers. In some cases it is desirable to connect two similar tubes in parallel to double the available power output. Often two groups of paralleled tubes are operated in push-pull as described in Sect. 13.17. The main advantage of parallel operation is that lower plate voltages are required than would be the case if a larger single tube, capable of twice the power output, was used.

The connection of two similar tubes in parallel is shown in Fig. 18. The two tubes together have half the plate resistance, twice

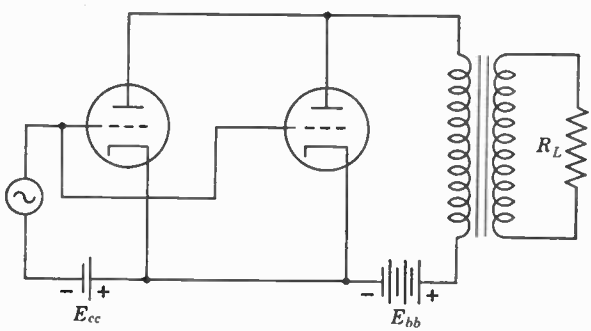


FIG. 18. Two tubes in parallel.

the grid-plate transconductance, and the same amplification factor as a single tube. The effective plate load resistance should be approximately half that used for a single tube. Under these conditions, and with the same grid-bias and excitation voltages and the same plate voltage, the power output of two tubes is twice that of a single tube. The percentage distortion is the same.

The plate (E_b–I_b) characteristics of a single tube can be used for the graphical calculation of the operation of two tubes in parallel. All that is necessary is to double the values of plate current. Then the calculations are made as before.

13.21. Feedback in amplifiers. When part of the output voltage of an amplifier is reintroduced into the input, it is said to be "fed back" and the circuit is called a "feedback circuit." Feedback has both advantages and disadvantages, depending upon how it is employed.

For example, suppose that an amplifier stage has a gain of 10 and that one-tenth of the output voltage is fed back in phase with the input. Now there is as much voltage on the grid due to the feedback as there is due to excitation from the external source. The additional voltage is amplified and reappears in the output. Again 10 per cent is fed back to the input. After a few cycles of this process the amplifier becomes overloaded and will probably break into uncontrolled oscillation. The amplifier, then, is said to "sing" or howl because it will probably produce an audible tone in the loud speaker, the frequency of which is determined by the sizes of the various electrical elements in the circuit.

If, on the other hand, a fraction of the output, say 5 per cent this time, is fed back into the input 180° out of phase with the original excitation voltage, the net voltage on the grid is reduced and the over-all gain becomes smaller. The gain does not, however, go to zero since on the next cycle of the process the actual voltage fed back is less than before (because of the decreasing gain), whereas the grid signal from the external source has not changed. After conditions have stabilized the resulting over-all gain is given by

$$A_{fb} = \frac{A_v}{1 - A_v\beta} \; *$$

where A_{fb} is the gain with feedback applied, A_v is the gain without feedback, and β is the fraction of the output voltage that is fed back 180° out of phase with the original excitation voltage and is called the **feedback factor.**

In the first case the process is called **positive feedback** or **regeneration,** and in the second case it is called **inverse** or **nega-**

* Here, as before, A_v carries a negative sign if the amplifier load is a pure resistance. Thus the sign in the denominator is positive for the usual numerical calculation.

tive feedback or **degeneration.** The effect of regeneration, even if not carried to the extreme as in the above example, is to exaggerate any non-uniformity in the frequency-gain characteristic of the amplifier. If the gain at some frequency is higher than at other frequencies, then more voltage is fed back. The result is that there is even more gain in proportion at this frequency than before.

13.22. Application of negative feedback. Negative feedback tends to decrease the non-uniformity of the gain-frequency characteristic of an amplifier. At a frequency for which the amplifier has too much gain more voltage is fed back. This results in a greater decrease in gain at this than at other frequencies. The result is that the frequency response is made more nearly uniform.

Negative feedback also tends to reduce waveform distortion arising within the tube and caused by operation over a non-linear part of the tube's characteristic. For example, if the output voltage of an amplifier is held constant by increasing the input voltage as negative feedback is applied, the distortion will be reduced in the same proportion as the over-all gain is reduced; that is, the new values of gain and distortion will both be equal to the values without feedback divided by $1 - A_r\beta$. This reduction in distortion arises as a result of the distortion components of voltage being fed back to the grid 180° out of phase so that they tend to cancel those generated in the tube. Thus, if an amplifier has a gain of 100 and 10 per cent distortion and negative feedback is employed which reduces the gain to 10, the new value of distortion will be 1 per cent.

On the surface it appears that this reduction of distortion cannot be taken advantage of since it is accompanied by a similar decrease in gain. However, since most of the distortion arises in the final power stages, which have large voltages in their grid and plate circuits, the gain lost as a result of negative feedback may be made up in low-level voltage-amplifying stages which are comparatively free of distortion. Negative feedback is a very useful tool in producing amplifier outputs which are nearly free of waveform distortion.

Negative feedback will not reduce distortion produced before the stage in which feedback is applied; it will decrease the distortion only in the feedback stage.

The use of inverse feedback has the following purposes: (1) to reduce frequency and waveform distortion; (2) to reduce variation in amplifier gain caused by variations in power supply voltage, variations in circuit components due to aging, and variations due to substituting new tubes for old ones; and (3) to reduce noise and hum produced in the amplifier. Also, some types of feedback reduce the effective plate resistance of the amplifier tube so that the load into which the tube works is increased in proportion. This tends to reduce distortion since it straightens out the dynamic tube characteristic. In addition, if the tube works into a load of varying impedance, such as a loud speaker, these variations have less effect upon the frequency characteristic and produce less distortion.

Negative feedback will not correct distortion occurring ahead of the point at which feedback is applied. It may be applied, however, over several stages so that an entire multistage amplifier may secure its benefits.

13.23. Practical negative-feedback circuits. The simplest method of producing negative feedback in a self-biased amplifier

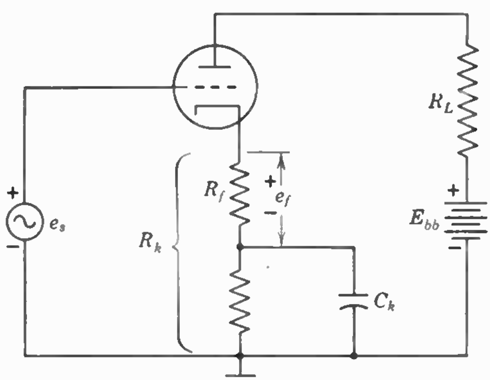


FIG. 19. Negative-feedback voltage is produced across resistor R_f. This is called current feedback.

is to remove the by-pass condenser from part or all of the cathode resistor. The a-c voltage across the unby-passed resistance R_f is 180° out of phase with the grid signal e_s as can be seen in Fig. 19. The polarity marks are shown for the part of the cycle during which e_s is going positive. The plate current is then increasing and produces a voltage drop e_f across R_f of the polarity shown. There is little a-c voltage across C_k since its action is to by-pass a-c currents. The total cathode-grid a-c voltage is $e_s - e_f$, the minus sign arising since the voltages are out of phase. This form of feedback will reduce distortion generated within the tube but will increase the effective plate resistance of the tube. The feedback factor β for this circuit is $R_f \div R_L$. This type of feedback is commonly termed **current feedback** since it is proportional to the a-c plate current.

Another feedback circuit is shown in Fig. 20. Here two re-

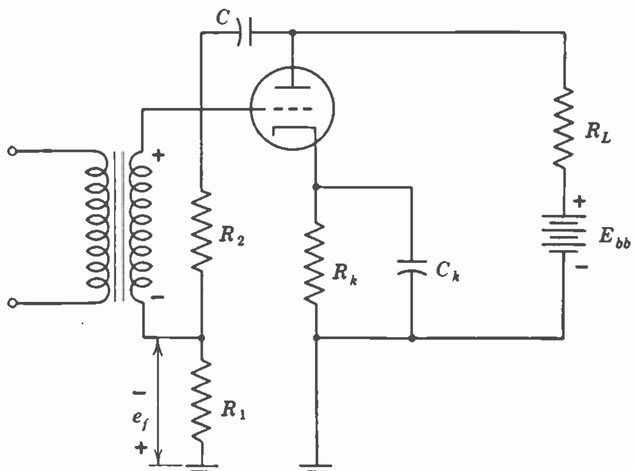


FIG. 20. A voltage-feedback circuit.

sistors R_1 and R_2 are connected in series with a blocking condenser across the output circuit. The feedback voltage is across R_1. This circuit employs **voltage feedback** since the voltage fed back is proportional to the output voltage.

There are many other feedback circuits using a variety of connections. Some work over a single tube, whereas others are used across several tubes in a multistage amplifier.

Problem 14. Suppose an amplifier has a gain of −70 at 100 and 5000 cycles and a gain of −100 at 1000 cycles. What will be the gain at these three frequencies with 10 per cent negative feedback ($\beta = 0.1$)? Compare the ratio of gains at the extreme frequencies to that at 1000 cycles for both the case with and the case without feedback. Does negative feedback flatten the gain-response curve?

13.24. Cathode follower. In all the amplifiers discussed up to this point the output voltage was taken from across the plate of the tube and the load network was in series with the plate lead. These are **plate-loaded amplifiers.** Another type of amplifier takes the output from across the cathode circuit. There is no plate load resistor. Such circuits are called **cathode followers.**

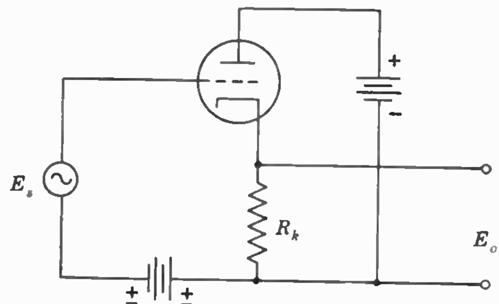


Fig. 21. Cathode-follower circuit.

The circuit of a cathode follower is shown in Fig. 21. Since the output voltage is also a part of the total grid voltage this circuit has negative feedback. In fact, since all the output voltage is fed back, the feedback factor is 100 per cent—the voltage gain can never be greater than unity and is usually 0.8 to 0.9 in practical circuits. Even though there is no voltage amplification in a cathode follower, there may be considerable power gain.

A cathode follower has the advantageous features usually associated with degenerative amplifiers. In addition, it has some special features. One of the most important is that the input impedance is much higher than that of the conventional plate-loaded amplifier. A second is that the output impedance is quite low, usually about 500 to 600 ohms. Thus, the cathode follower can be used to couple a very high impedance to a low impedance; it can serve as an impedance transformer. A further advantage is the fact that an increase in grid-signal voltage also increases the amount of feedback voltage, and the circuit can accept high signal voltages without overloading. Still other advantages are that there is no phase shift between input and output voltages and one side of the output is at ground potential.

Because of its many desirable features the cathode follower is used in broad-band video amplifiers, as the input stage of cathode-ray oscillograph amplifiers, and in many other applications. The cathode resistor is replaced with a parallel-resonant circuit when a cathode follower is used as an r-f amplifier.

13.25. Grounded-grid amplifier. If the grid of an amplifying tube is connected directly to ground, the grid serves to isolate the capacitance coupling the plate and cathode, and thus largely prevents undesirable feedback. Such a **grounded-grid** circuit is shown in Fig. 22. The input voltage is across the cathode

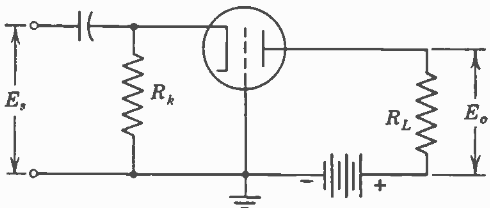


FIG. 22. Grounded-grid amplifier.

resistor R_k, and the output voltage is taken across the plate load resistor R_L. As with the cathode follower there is no phase shift through a grounded-grid amplifier. The circuit acts as though the amplification factor was equal to $(\mu + 1)$, and so slightly more gain is obtainable than with a conventional plate-loaded amplifier.

The chief use of this kind of amplifier is at high frequencies where the effect of interelectrode tube capacitances becomes serious. Then the cathode and load resistors are replaced by parallel tuned circuits.

13.26. The decibel. An amplifier may be described in several ways. It may be described by saying that it has a voltage amplification or gain of 160 times. This gives no information about the amount of power the amplifier can transmit to an antenna or to a loud speaker. Or the amplifier may be described as having an output of 10 watts. This gives no information about the voltage required on the input to deliver this amount of power to the load.

The voltage amplification and power output are useful and often necessary facts to know; but when two amplifiers, or two sounds of different intensities, are to be compared the matter must be looked into more closely. Suppose that a certain audio amplifier delivers 16 watts to a loud speaker and that this amount of power is not enough. The power output must be increased. By how much must it be increased before the normal human ear can distinguish a difference? Suppose that the power output is increased by 2 watts. This is a sizable amount of audio power, but the ear would not be able to note any difference. Suppose the power is increased another 2 watts. Now the ear just begins to notice that a change has taken place.

It is a fact that a given volume of sound must be increased approximately 25 per cent before the ear can detect the difference. When this increase has been made, the new value must be increased by another 25 per cent of its value before the increase in sound power can be detected. Through a wide range of sound intensities, equal percentage increments must be added to each succeeding intensity before the human ear detects that a change has been made.

Would 10,000 people shouting sound 100 times as loud as 100 people shouting? The answer is no—the greater sound will impress the average ear as being only about 20 times as loud.

If two sounds bear an intensity (power) ratio to each other of 1.25 to 1, the ear will detect that one is louder than the other. Now this is true for a very wide range of absolute values; that is, if two sounds of 10 and 12.5 watts are compared by the ear, the same *difference* in loudness will be noted as if the two sounds were actually 1.00 and 1.25 watts, in spite of the fact that in the first case the difference is 2.5 watts and in the second case only 0.25 watt. Clearly it is not the absolute value that counts; it is the ratio.

Now we might use this ratio of 1.25 to 1 as a loudness unit. To determine the relative loudness of two sounds, we would have to determine the number of times 1.25 would go into the ratio of the power in the two sounds. The result would give us the number of times louder one sound was than the other. This is a bit awkward and so a simpler system has been devised.

The **decibel** is a unit which approximately expresses this ratio of 1.25 to 1. When using decibels, each time a sound (or the power output of an amplifier) is increased by the ratio of approximately 1.25 to 1, we say that we have added 1 decibel.

Here we have one important advantage of the decibel as a convenient unit—we can add decibels, whereas actual powers or loudnesses must be multiplied.

If two sound intensities P_1 and P_2 (or the power outputs of two amplifiers) are to be compared according to the ability of the ear to detect intensity differences, the relative value of the two intensities may be determined by

$$N_{\mathrm{db}} = 10 \log_{10} \frac{P_1}{P_2}$$

where P_1 is greater than P_2.

The factor 10 comes into this expression because the original unit was the **bel** (named for Alexander Graham Bell), which is the logarithm of 10 to the base 10. The decibel is one-tenth of a bel and is used in preference to the bel since a change of sound intensity of 1 decibel corresponds very closely to the ratio of 1.25 to 1 discussed above. A change of 1 decibel is about the minimum change in sound intensity that the ear can detect.

The decibel is a *logarithmic* unit. Each time the power of an amplifier, for example, is increased by a factor of 10 the change is 10 decibels (abbreviated **db**). The table below gives some easily remembered values of decibels and their corresponding power ratios.

N_{db}	Approximate Power Ratio
3	2.0
4	2.5
6	4
7	5
9	8
10	10
20	100
23	200
30	1000

A convenient rule of thumb is that a 3-db increase represents a power ratio of 2 to 1. In the above table 3 db represents a power ratio of 2 and 6 db a ratio of 4, or the amount of power has been doubled again. The power ratio for 9 db (6 + 3 db) is 8, which is double the value for 6 db.

To determine the number of decibels by which two powers differ, the ratio of the two powers is first determined; then this ratio is looked up in a table of logarithms to the base 10, and the figure obtained from the table is multiplied by 10.

Consider again the sound intensities produced by 100 persons shouting and by 10,000 persons shouting. The ratio in this case is 10,000 ÷ 100 or 100. The logarithm of any number A to the base 10 is merely the number of times 10 must be multiplied by itself to be equal to the number A. In the example above, 100 represents 10 multiplied by itself, and the logarithm of 100 to the base 10, therefore, is 2. The number of decibels expressing the relative loudness of 10,000 people shouting to 100 is

$$
\begin{aligned}
N_{\mathrm{db}} &= 10 \log_{10} (10{,}000 \div 100) \\
&= 10 \log_{10} 100 \\
&= 10 \times 2 = 20
\end{aligned}
$$

or, if the number of people is doubled to 20,000,

$$
\begin{aligned}
N_{\mathrm{db}} &= 10 \log_{10} (20{,}000 \div 100) \\
&= 10 \log_{10} 200
\end{aligned}
$$

Now enters an aspect of logarithms which must be kept in mind constantly. It must be remembered that we are interested in the number of times 10 must be multiplied by itself to produce the given number. We know that 10 multiplied by itself produces 100, and when 10 is multiplied by itself 3 times, 1000 is the result. Clearly 200 is produced by multiplying 10 by itself somewhere between 2 and 3 times. Here is where the logarithm table comes in handy. We look up 2 (which is the significant part of 200) in the log table and find that it is 0.3. Then we state that the logarithm to the base 10 of 200 is 2.3. Here the 2 indicates the original number is somewhere between 100 and 1000; and 0.3 gives the exact place where 200 falls, on a logarithmic basis, between 100 and 1000.

To arrive at the correct figure, we must use the log table and our head at the same time. We must determine for ourselves what the number before the decimal point is—in this example, 2—then use the table to determine the value of the figure after the decimal point.

Therefore

$$N_{db} = 10 \log_{10} 200$$
$$= 10 \times 2.3 = 23$$

13.27. Voltage and current ratios. Strictly speaking, the decibel should be used only when expressing the ratios of powers. In special cases, and with care, the unit can be used to express the ratio of voltages or currents. Suppose that two amplifiers are feeding current into equal resistances. The currents are different. How can decibels be used in this case? All that is necessary is to find the ratios of the powers in terms of the currents and resistances and multiply the logarithm of this number by 10. Thus

$$P_1 = I_1{}^2R$$
$$P_2 = I_2{}^2R$$
$$N_{db} = 10 \log \frac{P_1}{P_2} = 10 \log \frac{I_1{}^2R}{I_2{}^2R}\,*$$
$$= 20 \log \frac{I_1}{I_2}$$

If the resistances are not equal, the equation becomes

$$N_{db} = 20 \log \frac{I_1\sqrt{R_1}}{I_2\sqrt{R_2}}$$

The equations when the voltages across the resistances are known are

$$N_{db} = 20 \log \frac{E_1}{E_2} \qquad (R_1 = R_2)$$

$$N_{db} = 20 \log \frac{E_1\sqrt{R_2}}{E_2\sqrt{R_1}} \qquad (R_1 \text{ and } R_2 \text{ not equal})$$

The factor 20 arises from the fact that when a number is squared the logarithm is doubled. For power ratios, the decibel is 10 times the logarithm; for current or voltage ratios, the decibel is 20 times the logarithm of the ratio.

* Note that log (P_1/P_2) is used instead of $\log_{10} (P_1/P_2)$. The base "10" is understood, and the meaning is the same.

Voltage or current ratios can be translated into decibels only when the resistances into which the currents flow, or across which the voltages exist, are known. If these resistances are equal for both currents or for both voltages, they cancel out, one being in the numerator and one in the denominator; but, in general, they do not cancel out and must be considered.

13.28. Power levels. *The decibel is always an expression for a ratio.* To say that an amplifier has a power output of so many decibels has no significance unless the power to which the output is compared is given. If, however, it is said that the power output is, for example, 20 db above 1 mw, the statement then gives specific information—20 db represents a power ratio of 100 and the power output of the amplifier is 100×10^{-3} or 0.1 watt. Thus, *when the power output of an amplifier or any other device is rated in decibels, the reference level of power must be given.*

The two common reference levels used by sound engineers are 1 mw and 6 mw. When voltage or current ratios are expressed in decibels the reference resistance must also be stated. Common reference resistance values are 500 and 600 ohms.

13.29. Decibel meters. Practically all so-called decibel or power-level meters are in reality only a-c voltmeters calibrated to read in terms of decibels when placed across a certain resistance. When used across any other resistance the indicated number of decibels is in error and a correction factor must be applied. This correction factor can be determined from the voltage ratio expression for decibels:

$$N_{\text{db}} = 20 \log \frac{E_1\sqrt{R_2}}{E_2\sqrt{R_1}}$$

which may also be written

$$N_{\text{db}} = 20 \log \frac{E_1}{E_2} + 10 \log \frac{R_2}{R_1}$$

where E_1 and R_1 belong to the power being measured and E_2 and R_2 are reference values. The first term of the expression gives the value the decibel meter will indicate. The second term is the correction factor which must be added to the meter indication to obtain the correct number of decibels. If R_1 is greater than R_2 the resistance ratio in the correction factor is turned over and a negative sign attached to the factor; that is, if R_1 is greater than R_2 the correction factor is minus and must be subtracted from the meter indication.

Example 1. An amplifier has 1 volt applied to its input resistance of 10,000 ohms. Across its output resistance of 4000 ohms appears a voltage of 40 volts. What is the power gain in decibels? Would it be worth while to increase the output voltage from 40 to 50 volts?

Solution:

$$\text{Power input } P_i = \frac{E_i^2}{R_i} = \frac{1}{10{,}000} = 10^{-4} \text{ watt}$$

$$\text{Power output } P_o = \frac{E_o^2}{R_o} = \frac{40^2}{4000} = 0.4 \text{ watt}$$

$$\frac{P_o}{P_i} = \frac{0.4}{10^{-4}} = 4 \times 10^3 = 4000$$

Power gain = 10 log 4000 = 36 db (The log of 4 is 0.6 and the log of 1000 is 3. Then, the log of 4000 is 3.6.)

If E_o becomes 50 volts,

$$P_0 = \frac{50^2}{R_0} = \frac{2500}{4000} = 0.625$$

The gain due to this increased output is

$$10 \log \frac{0.625}{0.400} = 10 \log 1.56$$

$$= 2.0 \text{ db} \quad \text{(approx.)}$$

Thus the gain due to increasing the output from 40 to 50 volts, or from 400 to 625 mw, will be audible to the ear, but the difference is not worth a great deal of effort to attain.

Example 2. A certain amplifier has a characteristic such that at 100 cycles its amplification in voltage is 8, at 1000 cycles it is 80, and at 6000 cycles it is 200. Are these differences appreciable to the ear?

Take the amplification at 1000 cycles as a zero level. At 100 cycles the voltage ratio is 80/8 or 10. At 6000 cycles the voltage ratio is 200/80 or 2.5. At 100 cycles the amplification is *below* zero level; at 6000 cycles the amplification is *above* zero level. Thus:

$$\text{At 100 cycles } N_{db} = 20 \log \tfrac{80}{8} = 20 \log 10 = 20 \text{ db} \quad \text{(down)}$$

$$\text{At 6000 cycles } N_{db} = 20 \log \tfrac{200}{80} = 20 \log 2.5 = 8 \text{ db} \quad \text{(up)}$$

Such a characteristic indicates a poor amplifier. The low notes would be totally lost, and the high ones are out of balance.

Example 3. In a certain circuit there is a loss of 25 db. What power ratio corresponds to this loss?

Power ratios of 10 = 10 db, 100 = 20 db, and 1000 = 30 db. Therefore the power ratio of 25 db lies somewhere between 100 and 1000. The figure 2 of 25 db indicates that the loss is somewhere between 100 and 1000 times. The figure 5 of 25 db is 10 times the logarithm of 3.1, and so 25 db corresponds to a power ratio of 310.

The solution of such a problem is as follows:

$$25 \text{ db} = 10 \log \frac{P_1}{P_2}$$

$$2.5 = \log \frac{P_1}{P_2} \quad \text{(dividing both sides by 10)}$$

$$\frac{P_1}{P_2} = \text{antilog } 2.5$$

$$= \text{antilog } 2.0 \times \text{antilog } 0.5$$

$$= 100 \times 3.10 = 310$$

If the loss were a voltage loss of 25 db the solution would be:

$$25 \text{ db} = 20 \log \frac{E_1}{E_2}$$

$$1.25 = \log \frac{E_1}{E_2} \quad \text{(dividing both sides by 20)}$$

$$\frac{E_1}{E_2} = \text{antilog } 1.25 = \text{antilog } 1.0 \times \text{antilog } 0.25$$

$$= 10 \times 1.78 = 17.8$$

13.30. The use of the decibel. The decibel may be used to express any ratio of power, voltage, current, mechanical power loss or gain, etc. It may be said that a symphony orchestra has a range of 60 db in power; that is, when it is playing very loudly, *fortissimo*, it is 60 db louder than when playing very softly, *pianissimo*. This corresponds to a power range of one million to one. In wire circuits which carry the amplified microphone currents from the symphony hall to the broadcast station, the weakest of the desired signals must be 40 db above the noise in the line. The very weak passages of the orchestra are built up by local amplifiers until the currents are greater than the noise currents in the line. The limit of the louder passages is fixed by overloading in the amplifiers, and by "cross talk" from one wire circuit to another. And so the stronger passages are cut down deliberately by the operator at the studio console.

Whenever a circuit suffers a loss in power or voltage or current, that loss may be expressed in decibels. The frequency characteristic of an amplifier, of a loud speaker, or of a telephone line may be expressed in decibels by plotting a curve on which zero level is the amplification or power output at some arbitrarily chosen frequency. Thus, if 1000 cycles is chosen as a reference frequency, the powers (or voltages or currents) at all other frequencies are either "up," "down," or "flat" with respect to the level at 1000 cycles.

13.31. The volume unit. The inclusion of the reference power and resistance along with a statement of the number of decibels output of a piece of apparatus is rather cumbersome and has led many people to neglect to give all the needed information. The lack of the reference values leads to no end of confusion, for without the reference levels the number of decibels gives very little useful information. A power-level unit is in use which avoids this difficulty. It is the **volume unit** (abbreviated **VU**) which is merely a decibel which always has a reference power level of 1 mw and a reference resistance value of 600 ohms. The use of the VU avoids the necessity of also giving the reference values since they are a part of the definition of the unit.

Problem 15. What number of decibels correspond to a voltage ratio of 100? a power ratio of 100? What voltage ratio corresponds to 100 db? what power ratio?

Problem 16. An amplifier has its power output reduced by 25 per cent. Is this change detected by the normal ear?

Problem 17. An amplifier has a normal output of 10 watts. A switch is provided with which the output can be reduced in 5-db steps. What is the output in watts when the output is reduced by 5, 10, 20, and 25 db?

Problem 18. The final load resistance of an amplifier is 500 ohms and the output voltage is 100 volts. A volume control reduces the output voltage to 10 volts. How much is the power reduced in decibels?

Problem 19. The noise is 40 db down from the broadcast signals on a certain telephone line. What is their power ratio? If the broadcasting voltages are of the order of 100 mv, what are the noise voltages?

Problem 20. An amplifier has four stages which have power gains of 25, 15, 15, and 30 db. The output transformer has a 5-db loss. What is the total gain of the amplifier in decibels? What is the power output if the input is 10^{-6} watt?

Problem 21. A metal shield around a certain transformer reduces the hum pickup 60 db below the unshielded value. What is the power ratio?

Problem 22. A certain decibel meter is calibrated on a 6-mw, 500-ohm base. When placed across a 5000-ohm load resistance the meter indicates

20 db. What is the corrected reading? What power is in the load resistor?

Problem 23. A volume unit meter reads 10 VU when placed across a 300-ohm load resistor. What is the true VU value? What power is in the load resistor?

Problem 24. The sensitivity of a velocity microphone is rated at —65 db where 0 db corresponds to 1 volt per dyne of force exerted by a sound wave impinging on each square centimeter of the microphone element. A carbon button microphone has a sensitivity of —45 db. What must be the voltage gain of an amplifier which will bring the level of the velocity microphone up to that of the carbon button microphone?

13.32. Low-frequency compensation of amplifiers. In several important applications the frequency-response curve of

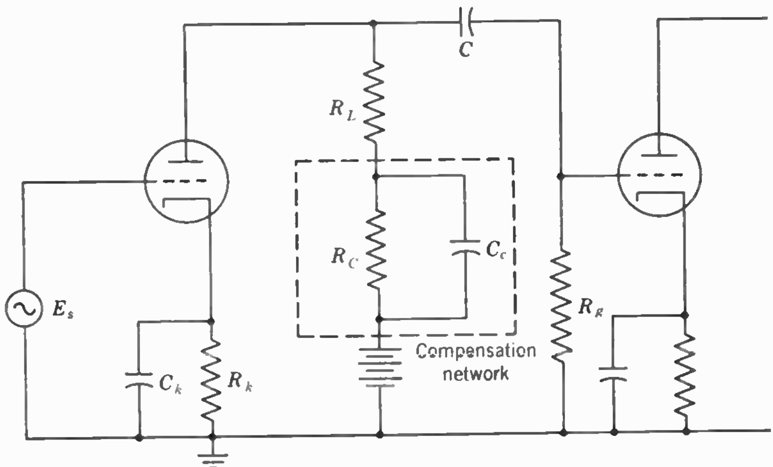


FIG. 23. An amplifier with low-frequency compensation.

an amplifier must be flat down to very low frequencies. A typical application is a video amplifier which handles television signals and must have a gain curve flat to below 30 cycles. In still other applications it is desirable to have the low-frequency gain greater than that at higher frequencies. An example is a sound amplifier which is equipped with a small baffle for the speaker. Such a small baffle does not radiate the low tones properly. This deficiency can be corrected if the gain of the amplifier is made higher at the low-frequency end.

A simple system for **low-frequency compensation** (called "bass-boosting") is shown in Fig. 23. At high frequencies C_c effectively shunts the compensation resistor R_c and the net load on the tube is R_L. As the frequency of the signal is lowered, the shunting effect of C_c becomes less and less until the effective load on the tube is $R_L + R_C$. With this higher load, the amplification becomes greater and compensates the usual voltage loss across the coupling capacitor C. This compensation network could be described as one which reduces the normal gain at higher frequencies while retaining full amplification at lower frequencies. Such operation is typical of many compensation networks. As in this case, the gain of compensated amplifiers over most of their frequency range is usually lower than that of uncompensated amplifiers using the same tubes.

13.33. High-frequency compensation of amplifiers. Just as a condenser shunting part of the load resistance can be used to pull up the response of an amplifier at low frequencies, an inductor in series with the plate load resistor can be used to extend the high-frequency response. The circuit of an amplifier

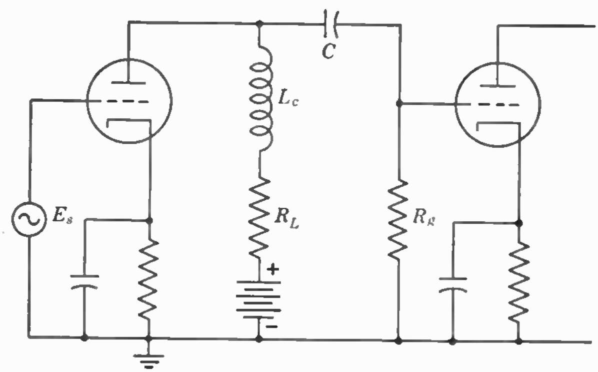


FIG. 24. A high-frequency compensated amplifier.

compensated in this manner is shown in Fig. 24. As the frequency is raised the reactance of the inductor increases and the total load impedance becomes greater, resulting in a larger gain. The cause of loss of gain at the higher frequencies is the shunting effect of tube and wiring capacitances. The increasing impedance of the load in the above amplifier can be used to just counteract the increasing effect of these shunt capacitances up to a certain maximum frequency, or it may slightly over- or undercompensate. In any event, the inductor and shunt capacitance go into resonance at a frequency near the maximum compensated value. At higher frequencies the gain falls very rapidly.

13.34. Direct-coupled amplifiers. A direct-coupled amplifier is one in which the plate of one tube is connected directly to the grid of the following tube with no intervening coupling capacitor. Such a circuit is capable of amplifying direct as well as alternating voltages. If there are no condensers elsewhere in the circuit, such as in the screen grid or cathode circuits, the direct-coupled amplifier will produce uniform amplification down to zero frequency.

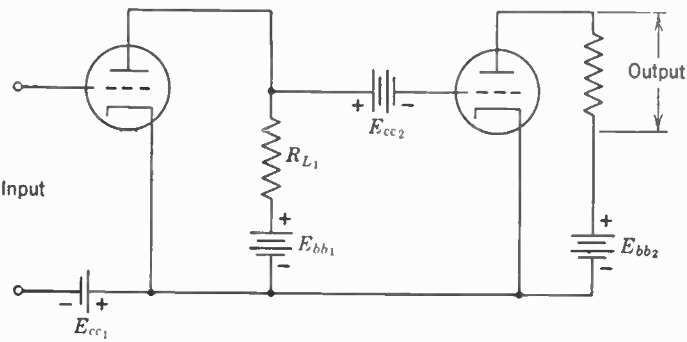


FIG. 25. Simple direct-coupled amplifier.

A simple direct-coupled circuit is shown in Fig. 25. The battery E_{cc_2} in the grid lead of the second tube is necessary to counteract the positive voltage of the plate supply battery E_{bb_1}, and to establish the proper negative bias on this grid. The amplifier in this circuit is very sensitive to changes of battery voltages since these changes are amplified along with the input signal. A wide variety of circuits has been devised in an effort to minimize this dependency upon the constancy of the battery voltages and to give high amplification at the same time. A discussion of these more complicated circuits is beyond the scope of this book.

13.35. Noise in amplifiers. Noise may be produced in amplifiers by a variety of mechanisms. The important sources of noise generated within an amplifier may be grouped as follows: (1) thermal agitation, (2) tube noises, (3) hum, and (4) microphonics.

Thermal agitation is caused by random motion of free electrons in a conductor which produce small voltages at the terminals. The rms voltage produced by thermal agitation is proportional to the temperature and size of the resistor and to the frequency band being amplified. If the electrical element is an impedance, only the effective series resistance is considered. This type of noise is so-called "white noise"; it is uniformly distributed over the entire frequency spectrum to well above any frequency used in present-day communication equipment. A 500,000-ohm resistor at a temperature of 80° F produces a noise voltage of approximately 13 μv in a 10,000-cycle frequency band.

Resistors which are made up of granular particles, such as the common carbon resistor, generate noise voltages far greater than that predicted from considerations of thermal agitation. These additional voltages arise from relatively high-resistance contacts between adjacent particles of conducting material in the resistors. This additional noise is normally bothersome only when some direct current is passed through the resistor. For this reason, carbon resistors should not be used in the grid and plate circuits of the low-level stages of high-gain amplifiers.

Noise voltages are generated in tubes as a result of several phenomena. **Shot effect** is a result of the random emission of electrons from the cathode. These electrons upon arriving at the plate produce a characteristic hiss which is troublesome in extremely high-gain amplifiers. **Random current division** of current between the plate and other positive electrodes such as the screen is another source of noise in multigrid tubes. **Grid current variations** resulting from the flow of positive ions to the grid also contribute to the total noise generated by a tube.

Hum is a general term which describes the alternating voltages appearing in an amplifier and originating from the a-c power source. These noises have frequencies of 60 and 120 cycles and possibly higher harmonics of the power-line frequency. Inadequate filtering in the rectifier-filter is a common source of hum. Transformers and chokes may pick up hum from electromagnetic induction, that is, from fields set up by alternating currents flowing in adjacent power transformers, filament leads, and other leads carrying a-c currents. The grid leads of the low-level stages of amplifiers are subject to **electrostatic pickup.** This noise voltage is greatest when the grid-leak resistor is large and the leads are long and unshielded. Alternating-current hum can arise in a-c-operated filaments of directly heated tubes. Even though the grid-return lead is connected to the mid-tap of the filament transformer some residual hum may still be present. Indirectly heated tubes, operated with a-c voltage on the heater, sometimes are subject to hum originating in the flow of a few electrons from the heater to the cathode and from magnetic fields set up by the heater current.

The hum originating from inadequate filtering can be eliminated by the use of larger filter chokes and condensers and by the placement of the filter choke so that no voltages are induced in it by the fields of the power transformer. All coupling transformers and chokes, particularly those in low-level stages, should be located as far as practicable from the power transformer. They should be of the shielded variety. Sometimes the proper orientation, that is, rotation on the chassis, will reduce pickup in transformers and chokes. All a-c power leads should be twisted together rather than run at random through the circuit since twisted leads produce very little external field. The grid leads of low-level stages should be as short as possible and should be well shielded. Sometimes the hum originating in an a-c-operated heater can be minimized by properly biasing the heater in respect to the cathode. In very high-gain amplifiers the heater may be operated from a battery.

Microphonics are noise voltages set up by relative movement of tube elements or of circuit components, usually as a result of vibration. Some tubes are more subject to trouble from microphonics than others, variations existing even among tubes of a given type. Therefore, tubes must be carefully selected for their low microphonic characteristics when they are to be used in critical applications. The coils and variable capacitors used in r-f amplifiers may also produce microphonic noise. Much of the trouble from microphonics can be eliminated by mounting tubes and other critical components in shock-mountings. In some cases it may be necessary to shock-mount the entire apparatus as well as the individual critical components.

The noise generated in low-level stages may easily be as large as the signal to be amplified, and therefore the noise may set the lower limit to the value of voltage that may be amplified without its being masked by noise voltages. For this reason, the design of the low-level stages of high-gain amplifiers requires a great deal of care.

14 · Detection of Amplitude-Modulation Signals

Before the output of a microphone or other audio-frequency (a-f) apparatus can be sent through space from a radio transmitter to a distant receiver, the audio-frequency voltages must be combined with a radio-frequency (r-f) voltage in some manner. The reason for this necessity is that a-f voltages (and currents) do not radiate effectively. Hence they are combined with higher-frequency voltages which are radiated more easily and which act as "carriers" for the a-f signals. The lowest "carrier frequencies" which are used are around 15,000 cycles, and transmission at this low frequency is very inefficient. Commercial broadcast stations operate in the frequency range from 550 to 1600 kilocycles.

14.1. Modulation. *Modulation* is the term applied to the process of changing some characteristic of a high-frequency carrier voltage (and current) wave by an a-f wave so that the resulting wave retains the ability of the r-f wave to radiate easily from an antenna but contains the audio frequencies. The process of removing the a-f modulation from a modulated wave at the receiver is called **demodulation** or **detection.** The general nature of modulation must be considered before various methods of detection may be studied. The details of modulation are left for later chapters.

A r-f voltage, or any alternating voltage for that matter, may be represented by the equation

$$e = E_{max} \sin (\theta + \phi)$$

where e is the instantaneous value of the voltage, E_{max} is the maximum value of the voltage during a cycle, θ is the part of the phase angle that varies with time, and ϕ is the part of the phase angle normally called phase-shift angle (Sect. 7.7). ϕ does not vary with time; it is a constant in the case of a simple sine wave of voltage.

The value of θ depends both on the frequency f and on time t. In a complete cycle θ must rotate through 360° or 2π radians. Thus, θ may be written $\theta = 2\pi \times f \times t$, and the entire expression for the voltage becomes

$$e = E_{\max} \sin (2\pi f t + \phi)$$

Consider the case in which the frequency is 1000 cps and the time is 1/4000 sec. Assume that the phase-shift angle ϕ is zero. Then

$$\begin{aligned} e &= E_{\max} \sin \left(2\pi \times 1000 \times \frac{1}{4000} + 0° \right) \\ &= E_{\max} \sin \left(2\pi \frac{1000}{4000} \right) = E_{\max} \sin \left(\frac{\pi}{2} \right) \\ &= E_{\max} \sin 90° = E_{\max} \times 1 \end{aligned}$$

If $E_{\max} = 10$ volts, then $e = 10 \times 1 = 10$ volts.

The expression for a r-f voltage is usually written $e_c = E_{\max} \sin (2\pi f_c t + \phi)$, where the subscript $_c$ refers to the carrier (radio) frequency. An a-f wave may be caused to change this r-f wave in four different ways:

1. The a-f wave may cause the amplitude, $E_{\max}$, to vary at an a-f rate. This is called **amplitude modulation (a-m).**
2. The a-f wave may cause the frequency, f_c, to vary above and below its normal value. This is called **frequency modulation (f-m).**
3. The a-f wave may cause the phase angle, ϕ, to vary above and below its normal value. This is called **phase modulation (p-m).**
4. The carrier wave may be sent out in a series of spaced pulses. The height or spacing of these pulses can be controlled by the a-f wave. This type of modulation is variously called **pulse** or **code modulation.** It is used for special services and will not be considered further in this book.

Another type of amplitude modulation which does not fit into the above classification is **keying.** This is the familiar "dot-dash" system in which the r-f wave is turned on or off to send messages, as with the Morse code. This type of modulation, of course, carries no a-f signal.

The detection of a-m waves will be treated in this chapter. Therefore, it is necessary to consider the nature of an a-m wave in a little more detail. Figure 1 shows the waveform of a carrier wave before and during modulation. The process of modulation has caused the maximum value (the amplitude) of the carrier wave to change in accordance with the variations of the a-f signal. The amount the amplitude of the carrier varies above and below its unmodulated value is a measure of the degree of modulation. This degree or **percentage modulation** is defined

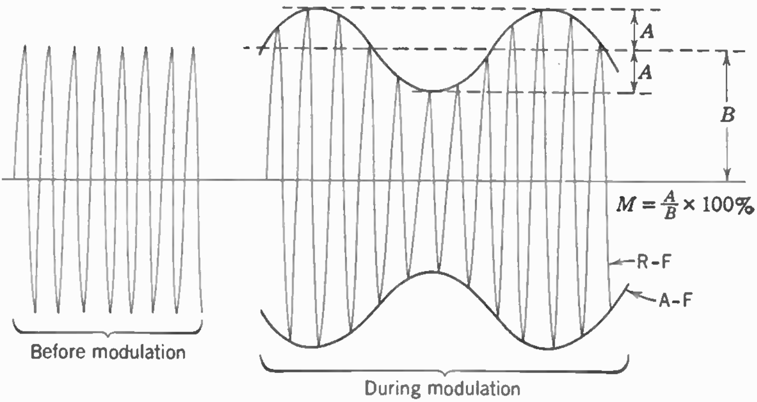


FIG. 1. Unmodulated and amplitude-modulated wave. The number of r-f cycles per a-f cycle is usually greater than shown in the figure.

by $M = A/B \times 100$ per cent, where A and B are measured as shown in Fig. 1.

The modulated carrier wave actually consists of three frequency components: one which corresponds to the original carrier frequency, another whose frequency is equal to the sum of the carrier and audio frequencies and is called the **upper sideband,** and a third which is equal to the difference of the carrier and audio frequencies and is called the **lower sideband.** These three frequency components together form the modulated wave. Commercial broadcast stations transmit the entire modulated wave. Transmitters for special services sometimes eliminate one of the sidebands and even the carrier. The reception of these waves requires special receivers.

14.2. Demodulation. At the receiving point, some process opposite to modulation must take place to secure, from the r-f energy, the modulating low frequencies which carry the desired intelligence. In this process, called demodulation or, more generally, detection, the a-f signal is separated from the modulated carrier wave.

There are three basic steps in the detection of an a-m wave. The first is rectification, by which one-half of the modulated wave is cut off. The amplitude of the remaining half cycles changes in accordance with the modulation. Sometimes only partial rectification is effected, and alternate half cycles are smaller than the other half cycles. The second step is to filter out the remaining r-f voltages but at the same time retain the amplitude variations. The third step is to remove any direct current produced in the complete or partial rectification process. Detection, therefore, consists essentially of passing modulated r-f voltages or currents through a rectifier. The output of the rectifier and its accessory apparatus consists of (1) a voltage (or current) whose amplitude varies in accordance with the modulation, and (2) some direct current.

14.3. Types of a-m detectors. The various detectors for a-m waves may be grouped into two broad classes: **linear** and **square-law.** These are often called large-signal and small-signal detectors, respectively. There can be no hard-and-fast classification because detector operation depends to a considerable extent upon the magnitude of the applied modulated signal and on other operating conditions. A given detector may operate either as a linear detector or as a square-law detector, depending upon the signal strength. It will be helpful, however, to separate detectors into these two groups for purposes of analysis.

14.4. Linear detectors. Linear detectors employ complete rectification. This rectification may be secured by means of various tube arrangements or by the use of crystal, copper oxide, or other rectifiers. An ideal linear detector characteristic * is shown in Fig. 2. The ideal characteristic is a straight line rising at an angle from the zero-current axis. For voltages less than the value at which the characteristic intersects the axis no current flows. In some types of linear detectors the point P is at some negative value of voltage and a fixed bias voltage is required. In others it is at zero voltage.

* The term "linear" here refers to the linear or straight portion of the characteristic for voltages at which current flows. Since there is a sharp bend at point P in Fig. 2, the curve is non-linear in a strict sense.

If an unmodulated wave is applied to the terminals of a linear detector the output current is made up of rectified pulses as shown in Fig. 2. These pulses have an average (d-c) value, and this average value would be read on a d-c milliammeter. If, on the other hand, a modulated wave is applied to the detector the output current consists of a series of pulses as shown in Fig. 3. The peaks of these pulses correspond to the variations in the

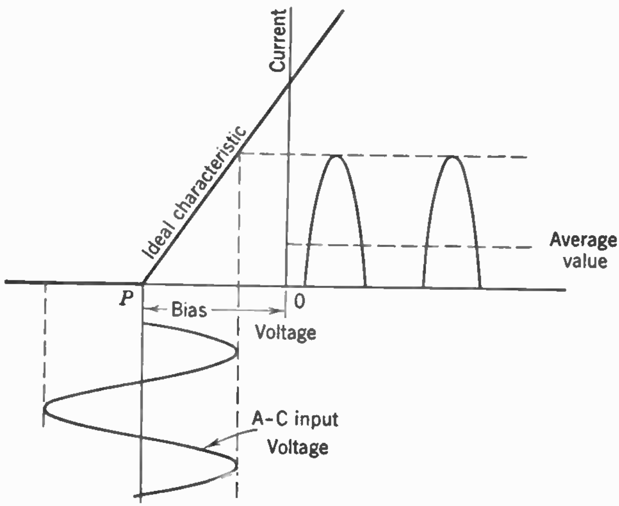


Fig. 2. Ideal linear detector characteristic.

modulation. The average heights of the peaks, shown by the dashed line in the figure, also correspond to the variations in the modulation. A detector responds either to the peaks of the pulses or to the average height of the peaks.

Now, if, at the transmitter, modulation consists in merely turning on and off the transmitter, then a plate-current meter will show a flow of current when the transmitter is on but will indicate zero current when the transmitter is off. Instead of a meter, a telegraph sounder could be used to indicate the message which the transmitting operator wishes to convey.

The peak value of the input voltage rises and falls in accordance with the modulation. The peak values of the rectified current rise and fall in unison with these input variations. A line connecting these peaks is called the **envelope** of the wave; and, if there is no distortion in the detection process, the envelope of the output current will resemble, exactly, the envelope of the input voltage. The "average" value of the current also follows the envelope.

If the carrier frequency is 1000 kc, for example, one million of the individual cycles will be impressed on the detector input

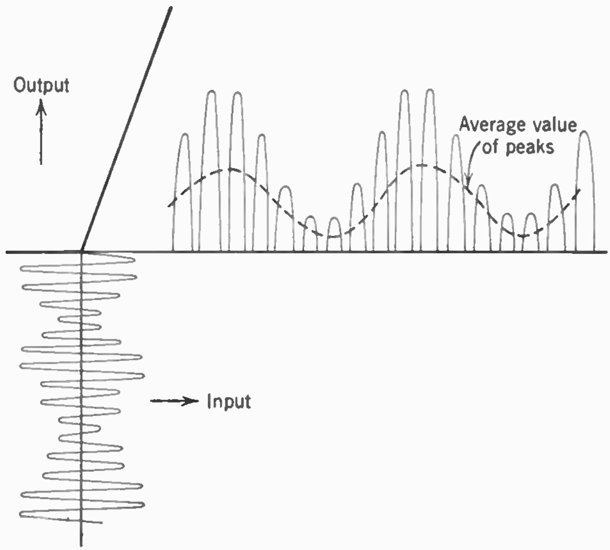


FIG. 3. Rectification or detection taking place about the sharp bend of a "linear" characteristic.

each second; for simplicity only a few are shown in Fig. 3. The peak value of these individual cycles varies at the modulation rate. Thus it can be said that each audio or modulation cycle consists of thousands of r-f cycles. For example, if a steady 1000-cycle tone modulates the 1000-kc carrier, each cycle of modulation will consist of 1000 cycles of carrier. Stated in another way, each cycle of modulation affects 1000 cycles of the carrier.

14.5. Square-law detectors. **Square-law detectors** employ partial rectification obtained through operation on a non-linear part of an *E–I* curve of a tube or other electrical device. A typical characteristic for square-law detection is shown in Fig. 4. Here the modulated wave is applied at a point on the characteristic where there is a rather sharp bend. The operating point P is fixed by a bias voltage.

When a modulated wave is applied to a square-law detector more current is passed on the positive half cycles than on the negative half cycles, as shown in the figure, and so partial recti-

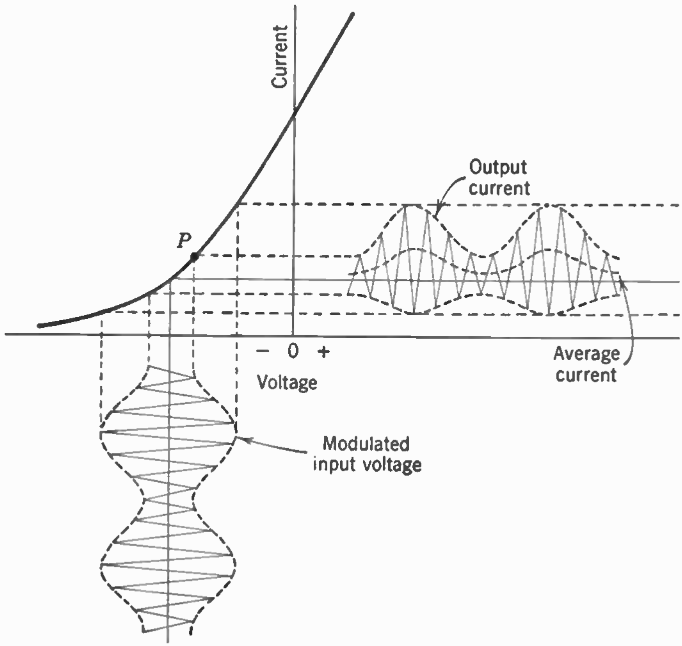


FIG. 4. Square-law detection. The output current is only partially rectified.

fication occurs. The "average" value of the output current contains the original a-f signal. It also contains harmonics of the a-f signal, and the percentage of these harmonics increases with the percentage modulation of the input wave. If the input signal is only 25 per cent modulated, the distortion will be about 6 per cent. However, if the input wave is completely modulated, the distortion rises to 25 per cent and is much too great to be tolerable for entertainment purposes. Square-law detectors were used in most of the earlier radio receivers since amplification as well as detection can be obtained in some square-law circuits. Modern tubes have made it easier to secure adequate amplification, and so present-day receivers rarely use these detectors. Linear detectors are used instead because of the low distortion, even with large degrees of modulation, which may be realized.

14.6. Filtering the detector output. The first step in detection is to rectify the modulated wave. In linear detection only the positive half cycles of the r-f wave appear in the output current. The peaks of these half cycles vary with the modulation.

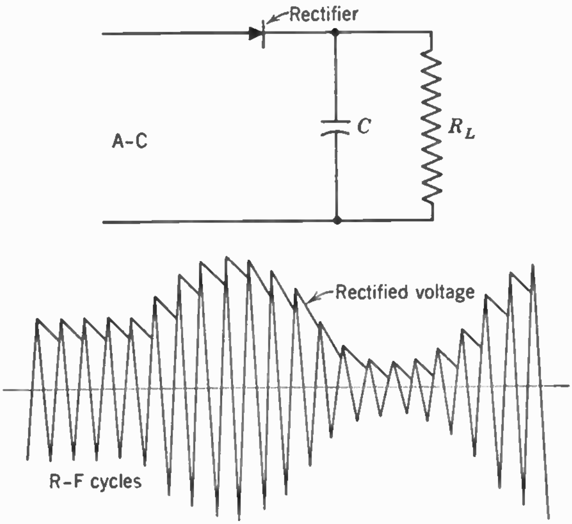


Fig. 5. Rectification process in which a condenser is used to maintain the envelope of the input voltage across the load.

Now that rectification has occurred, the r-f components are no longer of use and must be eliminated to prevent trouble in succeeding stages of a-f amplification. These r-f currents may be eliminated simply by providing an easy path to ground for them. without, at the same time, short-circuiting the modulating voltages to ground. Figure 5 shows a simple detector employing a rectifier which may be a diode, crystal rectifier, or other rectifying device. The load R_L may be, for example, a pair of headphones or the input to an audio amplifier. A condenser is placed across R_L.

When the first positive half cycle of current comes along from the rectifier it charges the condenser. When the charging voltage begins to fall the condenser discharges through the load resistance, holding up the voltage across the resistance in spite of the fact that the rectifier output is falling. If the condenser is large enough so that the time required to discharge it is greater than the time interval between positive half r-f cycles, it will completely smooth out the r-f variations across the load; if the condenser is too large it will also smooth out the low-frequency modulation. If the percentage modulation is high the condenser must discharge faster since then the sides of the envelope are steeper. The trick, then, is to choose a value of C (along with R) that filters out the r-f voltages without at the same time doing away with part or all of the modulation and introducing unwanted distortion. A good compromise is given by

$$R_L C = \frac{1}{\omega_m M}$$

where $\omega_m = 2\pi f_m$, and f_m is the maximum audio-modulating frequency, and M is the highest percentage modulation of the signal to be detected.

The jagged edges of the wave in Fig. 5 are overemphasized. In the usual case the number of r-f cycles per a-f cycle is many hundreds or thousands, and the envelope then has a very large number of small serrations.

The output of a detector must usually be amplified before being fed to a loud speaker. In this case it is necessary to eliminate the d-c component, leaving only the modulation itself, that is, the envelope of the wave. This can be done simply with a series condenser as shown in Fig. 6. The series condenser will pass the a-f currents but not the direct currents.

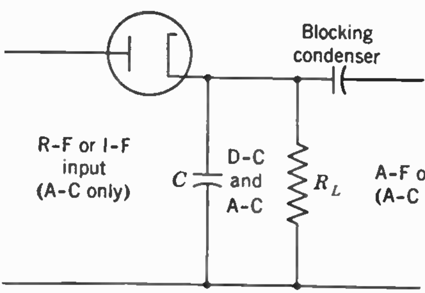


FIG. 6. Use of a blocking condenser to pass the a-f envelope but not the direct current produced by the detector.

14.7. Crystal detectors. Crystal diodes are frequently used as detectors, particularly in apparatus operating at very high frequencies, where the low interelectrode capacitance of the crystal unit is a distinct advantage. The small size of the crystal diode and the fact that no filament current is required are other advantages. A crystal diode is used in circuits such as are shown in Figs. 5 and 6. Even though there is some current through the crystal in the reverse direction, this current is usually small enough to give little trouble.

Crystal diodes may be used either as linear or square-law devices, depending upon the applied voltage and the size of the load resistance. However, they are generally employed as square-law detectors for which a rather low voltage and low load resistance are used. A description of crystal diodes and their characteristics is found in Chapter 10.

14.8. Diode detection. A large majority of present-day detector applications employ simple two-element high-vacuum tubes. The diode is about as sensitive as a crystal and is somewhat more stable. It delivers an a-f output with very low distortion if properly operated.

The diode may be made especially for the purpose of detection, like the 6H6. However, any rectifier diode may be used as a detector, although the relatively high interelectrode capacitances may give trouble at high frequencies and the filament power requirements are usually quite high. A triode or multigrid tube may be made into a diode by connecting two or more elements together. Thus, a triode can be operated as a diode if its grid and plate are connected together.

A typical diode detector circuit is shown in Fig. 7. The a-f

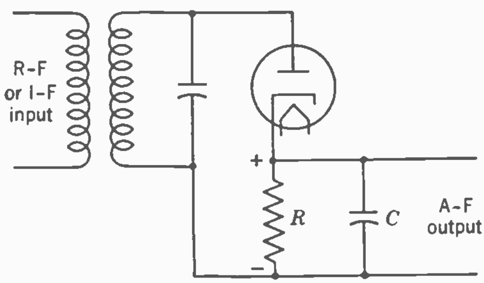


Fig. 7. Simple diode detector. Rectified voltages appear across R.

voltage appears across the load resistor R; C filters out the r-f components as discussed above. In practically all circuits R is made many times the plate resistance of the tube, which is of the order of a few thousand ohms. The high load resistance tends to decrease the effect of the curvature of the diode characteristic at low values of applied voltages and so decreases distortion in the output. The diode draws current from the input circuit during positive half cycles and therefore puts a load upon it, drawing power from it. This decreases the selectivity of the input circuit. A high load resistance helps reduce the power drawn from the input circuit in addition to decreasing the distortion.

Example 1. A diode detector employs a 0.5-megohm resistor shunted with a 100-$\mu\mu$f condenser as its load circuit. The audio frequency is 1000 cps and the carrier frequency is 1.0 Mc. What is the impedance of the diode load at 1000 cps and at 1.0 Mc?

Solution: At 1000 cps the reactance of the condenser is

$$X_c = \frac{1}{2\pi \times 1000 \times 100 \times 10^{-12}} = 1{,}590{,}000 \text{ ohms}$$

At 1.0 Mc the reactance of C is

$$X_c = \frac{1}{2\pi \times 1.0 \times 10^6 \times 100 \times 10^{-12}} = 1590 \text{ ohms}$$

Hence, X_c is about three times R at 1000 cps and will have little shunting effect at this frequency. However, at 1.0 Mc X_c is $\frac{1}{318} \times R$, and practically all the current will pass through C—very little will go through R.

At 1.0 Mc the total impedance of R and C in parallel is almost exactly equal to that of C, or 1590 ohms. At 1000 cps the total impedance is slightly less than the value of R (actually 478.000 ohms). The effective load impedance to a-f currents is about 300 times the impedance to r-f currents. As a result, r-f currents build up very little voltage across the load. This is just another way of stating that the r-f currents are by-passed by C.

In the circuit of Fig. 7, R is ordinarily about 0.5 to 1.0 megohm; C is about 150$\mu\mu$f at broadcast frequencies and higher than this value at intermediate frequencies. Except at low input voltages, R is large compared with the internal resistance of the tube, and for this reason most of the voltage produced in the rectification process appears across R. The actual value of the rectified voltage is almost equal to the peak value of the a-c voltage across the tuned input circuit. Values of C are chosen so that this condenser offers little reactance to the r-f currents but has high reactance to the a-f currents.

If an unmodulated r-f voltage is applied across the tuned circuit of Fig. 7 and the direct voltage is read across resistor R, a curve like that in Fig. 8 will be obtained. This curve shows, for

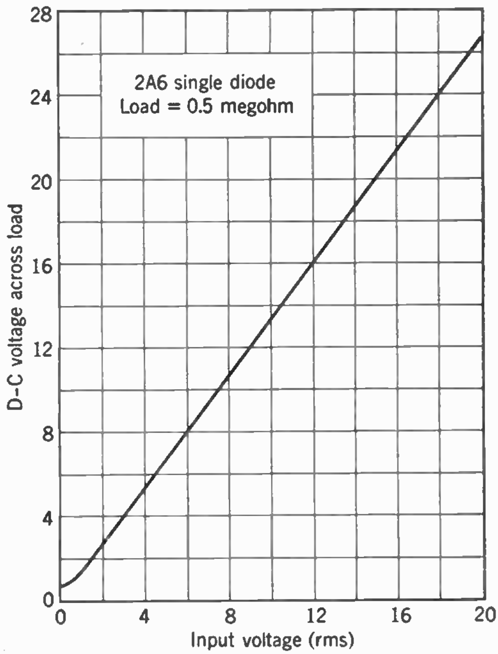


FIG. 8. Input-output characteristics of a diode detector. Curve also applies for diode unit of 2B7, 6B7, 6SQ7, and 75 tubes.

example, that 10 volts (rms) (a peak voltage of 14.14 volts) across the input produces a direct voltage across the load resistor R of approximately 13.5 volts. Now, if this input voltage is completely modulated, at some instants it will be zero and therefore no (or very little) rectified voltage will be produced, and at some other instants the input voltage will be 20 volts and approximately 27 volts will appear across R. In other words, the voltage across R varies from zero to 27 volts and has an average value of approximately one-half of 27 or 13.5 volts. A voltmeter responding only to a-c voltages, when placed across R, would measure a voltage having a peak value of 13.5 volts—an rms value of 9.55 volts if the modulation was sinusoidal.

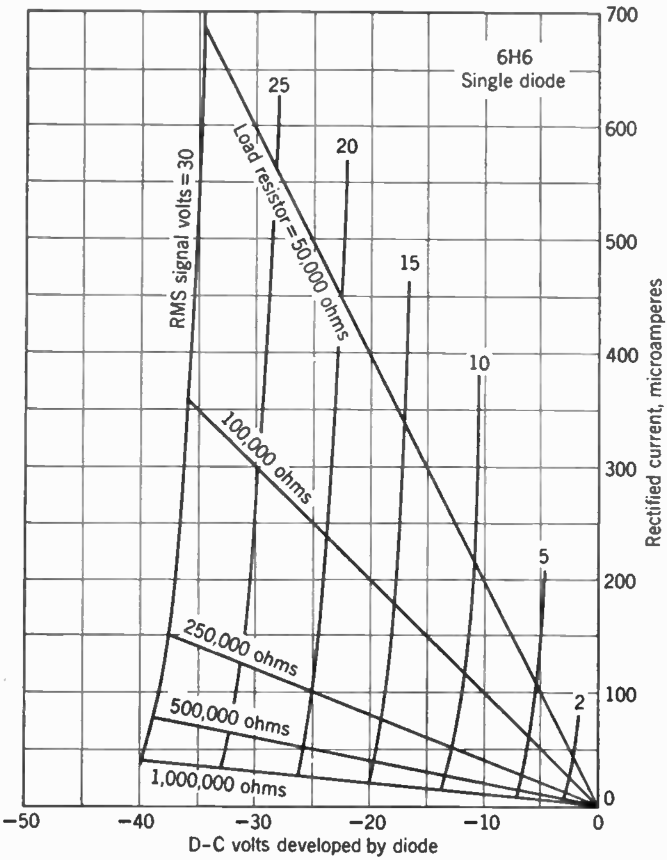


Fig. 9. Rectification diagram, showing d-c output as a function of a-c input for various load resistances.

Another method of showing the characteristics of a diode detector is the **rectification diagram** of Fig. 9. Several curves are shown for different rms values of signal voltage, and for several values of load resistance.

Example 2. A 6H6 detector employs a 100,000-ohm load resistor. A 15-volt (rms) carrier (unmodulated) is applied to the input. What d-c voltage is developed across the load resistor? What is the a-f output voltage if the carrier is 67 per cent modulated?

Solution: From Fig. 9, the intersection of the 100,000-ohm load line and the 15-volt input voltage line indicates that the d-c voltage developed across the load is 17.5 volts.

For 67 per cent modulation the minimum value of the envelope is, from Fig. 1,

$$A = B \times M = 15 \times 0.67 = 10 \text{ volts (rms)}$$

Then, the maximum rms signal voltage is

$$15 + 10 = 25 \text{ volts}$$

and the minimum rms signal voltage is

$$15 - 10 = 5 \text{ volts}$$

The corresponding d-c voltages developed by the diode are -29 and -6 volts, respectively. Then, the a-f output voltage is

$$E_{af} = \frac{29 - 6}{2\sqrt{2}} = \frac{23}{2\sqrt{2}} = 8.1 \text{ volts (rms)}$$

Resistance-capacitance coupling is almost always used to couple the output of a diode detector to the grid of a following amplifier stage. Since the load resistance of a detector is high it is probable that the grid-leak resistance of the amplifier tube will not be more than a few times greater than the diode load resistance. When the r-f voltage is unmodulated the diode load resistance acts as the total load resistance. However, when the input signal is modulated, the voltage across the filter condenser C is of a varying nature, the coupling condenser passes current to the grid-leak resistor, and the effective load on the diode is made up of the parallel combination of the actual diode load resistance R and the grid-leak resistance R_g. The lower effective load resistance on the detector results in lower sensitivity and may result in considerable distortion if the input signal is completely modulated. For example, Fig. 10 shows the rectification characteristics of a diode detector with a 100,000-ohm d-c load. The value of R_g is 200,000 ohms, which makes the effective a-c load of the diode only 67,000 ohms. Now suppose that a 15-volt completely modulated signal is applied to the detector input. On the maximum of the modulation cycle the carrier rises to 30 volts and the d-c voltage developed across the load is -35 volts. On the minimum part of the cycle the carrier drops to 0 volt and the d-c voltage is also zero. Actually, the d-c voltage goes to zero when the carrier drops below about 4 volts. The d-c voltage when the carrier is at its unmodulated or center value of 15

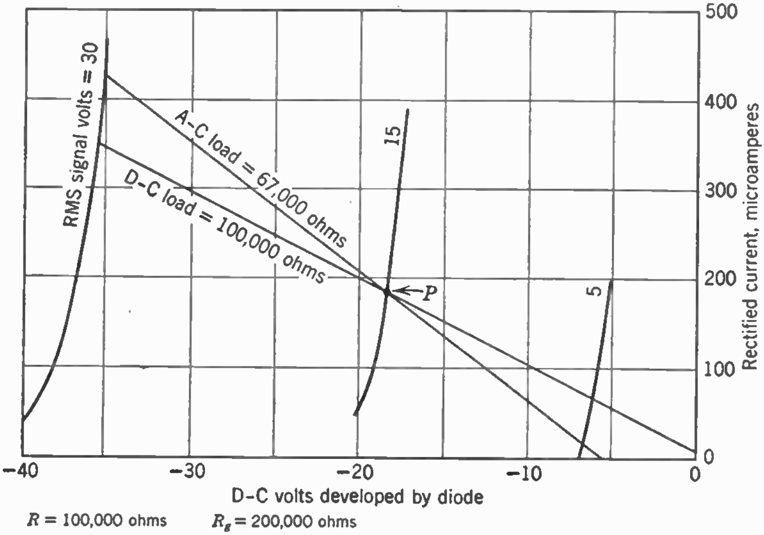


FIG. 10. D-c and a-c load lines on a diode-detector rectification characteristic.

volts is -18 volts. Then the envelope, which corresponds to the a-f output voltage, changes from -18 to -35 volts, a net change of 17 volts, on the positive half cycle of modulation. On the negative half cycle the voltage changes from -18 volts to -5 volts (the cut-off value for the a-c load line), a net change of 13 volts. Thus, the positive loop of a-f voltage is 17 volts, whereas the negative loop is 13 volts, and the peak of the negative half cycle is 4 volts less than the peak of the positive half cycle. This difference represents considerable distortion, mostly a second harmonic. The "bottom" of the a-f wave is flattened. This type of distortion is called **negative peak clipping.**

If the diode load resistance could be made nearly the same for a modulated as for an unmodulated wave, the distortion in the output could be reduced. One approach is to make the grid leak very large. However, this has the practical limitation that most tubes do not operate properly if the grid leak is too large. For most amplifier tubes, the grid-leak resistance should not be greater than 0.5 to 1.0 megohm. Another method of reducing the distortion is shown in Fig. 11. Here the input to the amplifier is taken from a point part way up on the diode load resistance. This effectively reduces the loading caused by the grid-leak resistance. This method also reduces the voltage supplied to the amplifier, and so the amplifier must be designed to have higher gain than would otherwise be necessary.

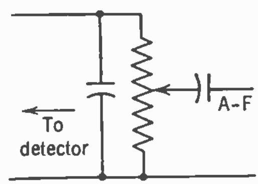


Fig. 11. Tapping the output of the detector puts stray capacitances across only part of the detector load, thus increasing the impedance of this load.

14.9. Plate circuit detector (strong signals). In a plate circuit detector an amplifier is biased for approximately Class B operation and rectification of the input signal occurs much as in a diode detector. For strong signals the circuit operates as a linear detector. There is somewhat greater distortion than in a diode detector as a result of the curvature at the lower end of the triode characteristic. The circuit has the advantage of providing amplification of the input signal and so has rather high sensitivity.

A typical **plate circuit detector** (also called grid-bias or linear plate detector) is shown in Fig. 12. The condenser C by-passes r-f voltages in the plate circuit to the cathode and, in combination with the r-f choke, keeps these currents out of the primary of the transformer. There is considerable curvature near the low end of the characteristic curve. This curvature will produce square-law detection if the signal level is low, and the distortion will be rather high as is the case with all square-law detectors. Even for strong signals and approximately linear operation, there will be considerable distortion if the modulation is great enough to force the envelope to low points on the characteristic curve during the "troughs" of the modulation cycle.

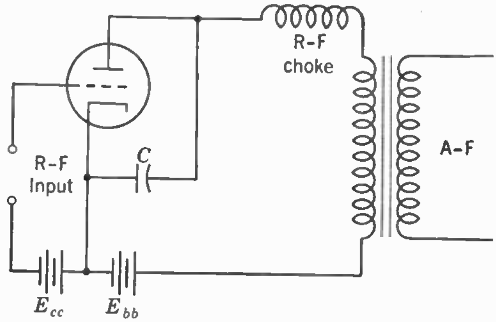


FIG. 12. Plate circuit detector. The combination of the choke and condenser C keeps r-f currents from the transformer.

14.10. Infinite input impedance detector. The circuit of an infinite input impedance detector is shown in Fig. 13. The circuit is essentially the same as that of a cathode follower. One important difference is that the cathode resistance is made very large so that, with no input signal, the plate current is approximately zero. The envelope of the rectified voltage in the cathode is nearly as large as the envelope of the input signal. The circuit will operate satisfactorily for widely varying levels of input signals, the peak value of the input signal that will be detected without distortion being approximately half the plate supply voltage. The input signal should, however, always be a few volts at least, in order to prevent distortion resulting from square-law detection. The detector is subject to negative peak clipping, much as in a diode detector, when the degree of modulation exceeds the a-c to d-c impedance ratio of the load. The chief advantage of the infinite impedance detector compared with a diode detector is that its input impedance is very high (almost infinite compared to the diode) and so the detector imposes very little load on the input circuit.

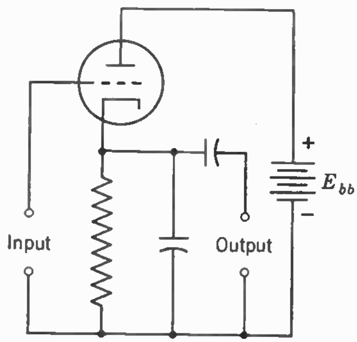


FIG. 13. Infinite input impedance detector.

14.11. Grid-leak and condenser detector. The signal in a plate circuit detector can be considered as first being amplified and then passed on to the plate circuit to be detected. There is rather small amplification in such a tube at radio frequencies because the load impedance is deliberately made low to get rid of the r-f variations in plate current. In Fig. 12 C practically short-circuits the r-f currents in the plate circuit.

The grid-leak and condenser detector shown in Fig. 14 is more

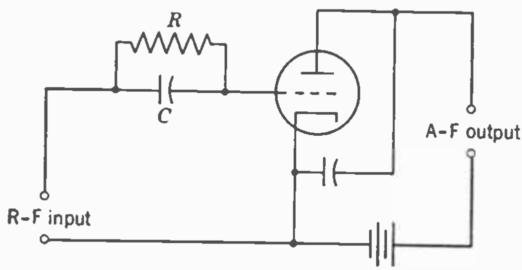


FIG. 14. Grid-leak and condenser detector.

sensitive than the plate circuit detector but has limited power-handling ability. In this circuit detection is accomplished in the grid circuit; then the amplifying properties of the tube are available for producing a rather high a-f output voltage.

The **grid-leak detector** utilizes the curvature of the grid-voltage–grid-current curve, and it detects by square-law action. The plate circuit should operate on the linear part of its characteristic. This requires that the plate voltage be low, usually 25 to 75 volts. The grid-leak resistance R is usually made rather high so that the flow of grid current will produce adequate bias for operation of the tube and so that the input impedance will be high. The condenser C filters out the r-f variations in the voltage. Typical values of the grid-leak resistance range from 1 to 3 megohms, with C about 250 $\mu\mu$f.

Grid-leak detectors were widely used in early-day receivers because of their high sensitivity. However, because of their high distortion in comparison to that of a diode detector and because of the ease with which adequate amplification can be secured with modern tubes, grid-leak detectors are rarely used now except in special applications.

A grid-leak detector can be operated as a linear detector if the resistance of the grid leak is reduced and the input signal is made fairly large. In this type of operation, the circuit operates roughly the same as a diode detector followed by a stage of amplification. The distortion is higher than in a diode detector, and the circuit cannot handle extremely high input voltages without rectification in the plate circuit, which produces additional distortion.

14.12. Automatic volume control. A notable advance in radio receiver design took place when **automatic-volume-control (a-v-c)** circuits were developed and placed in active use. By means of avc the effects of wide variations of voltage at the antenna terminals (fading) are overcome to a great extent. Furthermore, blasts of loud signals are prevented when the receiver is tuned from a weak station for which the manual volume control might be turned up high to a powerful station for which the volume control must be turned down.

As the name implies, a-v-c circuits control the gain automatically. This is accomplished by varying the amplification of the r-f and intermediate-frequency (i-f) circuits inversely as the strength of the incoming signal. When this incoming signal is high, avc reduces the gain to a low value, and when the incoming signal is weak avc increases the gain to its maximum value. The circuits are designed to maintain a nearly constant voltage at the detector terminals and are able to do this over a wide range of incoming signal strengths.

Although a separate tube may be employed to produce the voltages necessary for avc, it is common practice to utilize the d-c voltages developed in the diode detector. In Fig. 7 the lower end of the diode load resistor R is at a negative potential with respect to the cathode of the diode tube. This negative potential may be utilized for avc by impressing it on the grids of r-f and i-f tubes as bias. The stronger the signal impressed on the input of the receiver, the stronger will be the signal impressed upon the detector input (in the absence of avc). The stronger the incoming signal, the more direct current is developed in the diode by the process of rectification, and this current, flowing through the diode load resistor, produces a voltage drop along it. A greater voltage is developed when a greater current flows through this resistor, and this greater voltage applied as bias to the previous tubes will reduce their gain.

The r-f and i-f amplifier tubes are usually super-control (variable-mu) tubes when a-v-c bias is to be applied to them. These tubes are designed so that their mutual conductance is reduced gradually and more or less linearly as the bias is made more negative. Then, the stronger the incoming signal, the greater the voltage applied as bias to the amplifier tubes which reduces their

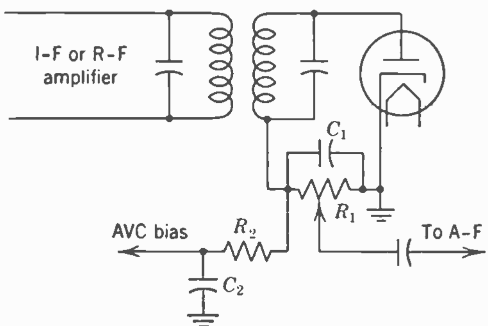


Fig. 15. Typical a-v-c circuit.

amplification. Obviously, since a-v-c action depends upon an increase (or decrease) of voltage at the detector terminals to initiate its action, the detector voltage cannot be held constant. However, the variations of detector voltage as a receiver is tuned from a strong to a weak station may be made rather small by proper design of the a-v-c circuit.

Figure 15 shows a typical a-v-c circuit. Note that a filter made up of a series resistor R_2 and a shunt condenser C_2 is placed across the a-v-c lead as it is taken from the diode. This is necessary because the voltage across the diode load R_1 varies with the modulation of the incoming signal. If this voltage were applied to the grids of the amplifier tubes, the bias would go up and down with the modulation, again in the inverse sense, so that the output of the detector would tend to be the same strength for all levels of modulation. There would be no variations in audio volume. This is not the action that is desired. Rather, avc should depend on the strength of the unmodulated carrier. With the a-v-c filter in place, the current from the diode charges C_2 through R_2. Condenser C_2 is forced to charge slowly by making R_2 very high, and therefore the voltage applied to the amplifier grids as bias cannot follow the modulation of the incoming signals, although it may follow the variations in the signals due to fading or tuning from one station to another. These latter variations are much slower than the a-f modulation. The design of the a-v-c filter is important. The values of R_2 and C_2 must be chosen so that the time taken by the condenser to discharge (the time constant R_2C_2) is such that the lowest modulation frequencies will not cause any variation in the grid bias of the amplifier tubes, but the discharge time must not be so large that a delay occurs when the system is recovering from a crash of static. A time constant ($R_2 \times C_2$) of $\frac{1}{10}$ to $\frac{1}{5}$ second is usually employed.

There are many variations of this simple a-v-c circuit. Sometimes the a-v-c voltages are amplified before being applied as bias; sometimes more a-v-c voltage is applied to one tube than to another; sometimes none is applied until the incoming signal reaches a certain value. The latter variation is known as "delayed" avc since the action is delayed until signals of a certain strength are attained. In this way the amplifiers have maximum amplification for very weak signals.

14.13. Manual volume control. An a-v-c system supplies nearly constant voltage to the detector over a very wide range of incoming signals. Control of the loud-speaker volume to suit individual requirements must, therefore, be made by a manual adjustment, by taking more or less of the audio output of the detector and applying it to the a-f amplifier. This is the purpose of the tap on resistor R_1 in Fig. 15.

In receivers which do not employ avc the volume may be controlled in several ways besides the tap on the detector load resistor. One way is to decrease the gain of the r-f or i-f amplifiers by manually adjusting the bias. Since more bias adjustment is generally required for the amplifying stages nearest the detectors than for those near the antenna input, some means is usually provided for proportioning the volume-control potential so that the proper bias is applied to each amplifier tube. Sometimes the bias of the oscillator tube of a superheterodyne is adjusted at the same time the amplifiers are controlled, either manually or by a-v-c circuits.

Potentiometers are not often used in r-f and i-f circuits as a means of volume control since they tend to introduce noise. Furthermore, their size and construction usually introduce more shunt capacitance into the circuit than can be tolerated.

15 · Amplitude-Modulation Receiver Systems

In an amplitude-modulated system, the carrier power of the transmitter varies at an audio-frequency rate, the magnitude of the variation representing the amplitude of the a-f modulation. A receiver must convert these variations in signal strength into audible sounds.

15.1. Basic performance characteristics of a radio receiver. Three basic performance characteristics of a radio receiver must be weighed not only by the designer but also by the ultimate user: **sensitivity, selectivity,** and **tone fidelity.** The sensitivity of a receiver is an indication of the over-all amplification from the antenna input to the loud-speaker output. A receiver that is very sensitive requires only a small input voltage to deliver considerable output power. The selectivity of a receiver is an indication of its ability to discriminate between wanted and unwanted signals. An infinitely selective receiver would be one that would respond only to a given station and not at all to another, no matter how powerful this undesired station might be, or how close in frequency it was to the desired station. That there is no such receiver goes without saying. The tone fidelity of a receiver indicates how well it reproduces the a-f modulation on the carrier which is picked up at the input terminals. A receiver that delivers a high-fidelity signal is one which has a flat frequency-response curve over a wide a-f range and which is free from noise and distortion.

15.2. Field strength. The voltage that is produced across a receiving antenna is proportional to the field strength of the transmitter at that particular point on the earth's surface.* This

* **Field strength** is the same as field intensity and is measured in volts per meter. Then, if a receiving antenna is 1 meter long and the field strength along the antenna wire is 1 volt per meter, the total voltage impressed across the antenna is 1 volt.

voltage is usually of the order of millivolts or microvolts. The greater the field strength at a given point the greater the power output of a receiver for a fixed amount of amplification. Similarly, the greater the field strength, the less receiver amplification is necessary to obtain a certain amount of loud-speaker power.

15.3. Desirable signal strengths. Experience has indicated that signal strengths of three general classifications are necessary to provide good service to listeners. In city business areas 10 to 25 mv per meter is required to override high interfering electrical noise and compensate for the electrical "shadows" cast by large buildings; in the residential areas of large cities, field strengths of 2 to 5 mv per meter are required; and in rural areas where man-made noise is low, field strengths as low as 0.1 to 0.5 mv per meter will usually provide satisfactory service. It must be remembered that the absolute value of the signal is seldom the important quantity; it is how much louder the signal is than the noise; and that in a locality where noise is low (as in the country) the signal need not be so strong.

The signal strength at a remote receiving point is proportional to the square root of the power at the transmitter; it depends also upon the frequency employed, the heights of the receiving and transmitting antennas, and the kind of soil or terrain between transmitter and receiver. Also, of course, the signal strength decreases as the distance increases. There is, in addition, some loss of signal strength resulting from absorption of the r-f power in the signal by foliage and by absorption and reflection from buildings and other structures.

The purpose of the transmitting station is to provide a good lusty signal that will override static and other disturbances; the purpose of the r-f amplifier and associated equipment in the receiver is to provide the listener with good loud signals from the field strengths which the transmitting stations produce at the receiving point.

15.4. Advantage of high power. Whatever voltage is produced across a receiving antenna, whether noise or desired signal, is amplified by the r-f amplifier; there is, therefore, a distinct advantage, as far as the receiver is concerned, in using large amounts of power at the transmitting station. The greater the ratio of signal to noise at the receiver terminals the better will be the reception. No matter how great the gain of the amplifiers in a receiver, this gain cannot bring a weak signal out of the noise and give satisfactory reception. That is, if the desired signal and the noise are about the same strength, no amount of amplification will make the signal any more intelligible. The signal should be about 40 db * above the noise level on the input of an a-m receiver to provide entertainment free from excessive background noise. Whenever the noise increases, as on a warm summer day or during a thunderstorm, and the transmitter power remains constant, reception suffers, and it suffers more the farther the receiver is removed from the station. The noise picked up by a receiver is more or less constant under a given set of conditions, whereas the field strength due to a transmitter decreases with distance from the transmitter.

If a receiver is situated in an electrically quiet locality, where the noise level made up of stray voltages from street cars, elevators, arc lamps, corona on high-tension wires, x-ray machines, electric razors, etc., is weak, more amplification can be employed to bring in weaker stations; or, with the same amplification, stations further away can be received. So long as the signal is about 40 db stronger than the noise, it can be amplified and its program enjoyed.

One of the virtues of the high frequencies is the fact that there is less natural static than on lower frequencies; however, man-made static and electrical noise tend to increase with frequency until very high frequencies are reached.

If the transmitter power is increased by 10 db (10 times increase in power), the signal-to-noise ratio, expressed in terms of power, is likewise increased 10 times. This desirable effect can be achieved by an actual increase in the power output of the transmitter, or by the use of directive antennas at the transmitter or receiver, or both.

15.5. Amount of amplification necessary. On the surface it would appear to be possible to add as many stages of amplification to a receiver as desired. They could operate at the frequency of the incoming signals, or at the modulation (audio) frequencies secured by the detection process. If the receiver is a superheterodyne, some of the amplification could be at a frequency

* Signal voltage should be 100 times the noise voltage.

intermediate between the incoming frequencies and the audio frequencies in the output. The amount of amplification which can be used successfully is, however, limited. For one reason, amplifiers tend to become unstable if the gain is too high, owing to the difficulty, especially at high frequencies, of preventing some of the output from getting back into the input. Secondly, some noise is always created in the amplifying and detecting processes. Additional noise is present at the input terminals. This noise may be amplified along with the desired signal, and, if the signal is weak, the noise may seriously interfere with or completely mask the desired signal.

In practical receivers, some of the amplification takes place at the frequency of the incoming signals and some at audio frequencies. In superheterodynes the incoming signals are converted to signals at a lower (intermediate) frequency, ordinarily in the range of 150 to 500 kc, and some amplification takes place at this intermediate frequency. In such a receiver, there may be amplifiers operating at radio frequency, at intermediate frequency, and at audio frequency.

If a diode detector is employed, not less than about 10 volts (rms) must be available at the detector terminals if distortion is to be kept low. Between the antenna input terminals and the detector input terminals must be sufficient amplification so that 10 volts is produced from the incoming signal whatever its voltage may be. Under ideal conditions, inherent noise at the antenna input terminals is no lower than 1 μv, and is nearly always higher. If the receiver is to pick up signals no greater than the noise level, the total required amplification will be $10 \div (1 \times 10^{-6})$ or 10 million times up to the detector input terminals. A receiver is usually designed to have less amplification than this for the simple reason that a signal no greater than the noise level would be scarcely intelligible after detection.

There is usually a voltage gain of 2 to 5 times in the input system to the first tube, accomplished through the use of resonant circuits. A stage of r-f amplification preceding the frequency-converter stage may have a gain of about 30 times. Another gain of about 20,000 will occur from the grid of the frequency converter to the detector input, if a two-stage i-f amplifier is employed. Some amplification occurs between the detector output and the loud speaker. The total voltage amplification amounts to about 3 million, which is about the limit found in a good receiver.

15.6. Receiver selectivity. The selectivity of a receiver depends largely upon the use for which it is intended, and, more specifically, upon the width of the band of frequencies which the r-f and i-f amplifiers must pass. A good-quality receiver for broadcast reception must have amplifiers with pass bands of 10,000 cycles or somewhat more, so that modulation ranging up to above 5000 cycles will be passed. Communications receivers require band widths of around 5000 cycles since only voice signals ranging up to 2500 cycles need be passed. Code receivers employing quartz-crystal filters are able to separate signals only a few hundred cycles apart; they have a very narrow band width. Of course such receivers would be unsatisfactory for voice or music reception.

It is rather difficult to design an amplifier which has a wide band width, gives fairly constant amplification over the band, and yet rejects unwanted signals whose frequencies are near those passed by the amplifier. The normal amplifier response curve slopes off rather gradually outside the range of frequencies to be passed and, therefore, gives some amplification for signals both above and below the desired band. For this reason, broadcast receivers generally do not have nearly as good selectivity as do communications receivers.

15.7. Tone fidelity required. If a receiver is to be used on a code circuit, only a very narrow band width is required. If a general communications receiver is used for both code and voice, better tone fidelity is required. If the receiver is to be a high-fidelity broadcast set, then very good tone fidelity is absolutely necessary.

Tone fidelity requires not only wide-band response in the r-f and i-f amplifiers, but also flat response in the a-f stages. The response of the speaker is also of great importance. A loud speaker which is relatively free from distortion and which transmits frequencies from 60 to 10,000 cycles may cost as much as a small table-model receiver. Thus the cost of a high-fidelity receiver may be several times that of the average radio.

15.8. Types of receiver circuits. Several types of circuits are used for radio receivers:

1. Crystal detector or single-tube receiver.

2. Regenerative detector plus audio amplifier.
3. Tuned-radio-frequency (t-r-f) amplifiers.
4. Superheterodyne.
5. Superregenerator.

15.9. Crystal detector receiver. In Fig. 1 is a simple circuit which was used by wireless operators for many years before

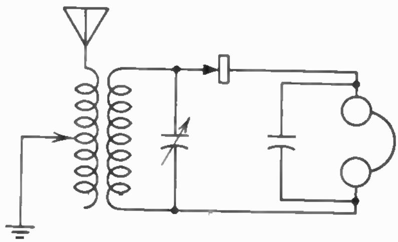


FIG. 1. Detector using a crystal as rectifier.

vacuum-tube equipment was available. The single tuned circuit provided relatively little selectivity, but in those days problems due to the multiplicity of radio stations were not so great; signals were desired only over fairly short distances (ship to shore); and headphone reception was all that was required. Nowadays, a crystal detector would be used only where sufficiently strong signals were available, far enough removed from interfering signals in distance or in frequency to cause no trouble. In place of the crystal, a grid-leak detector tube could be utilized. This would have about 10 times the sensitivity of the crystal, but the problem of selectivity would remain unsolved.

15.10. Regenerative detector. If, as in the circuit of Fig. 2, part of the output of the tube is fed back into the input in phase, regeneration takes place, the output signals will be increased, and the sensitivity and selectivity of the circuit will be improved. If the amount of regeneration is increased sufficiently, as by moving the feedback coil close to the input coils, the tube will oscillate, that is, it will generate alternating currents of its own. If the receiver is made to oscillate at a frequency slightly different from the incoming carrier, say 1000 cycles off, the combination of the incoming signal and the locally generated signal creates a third signal whose frequency is the difference between the two combining frequencies. In this example it is 1000 cycles; this

1000-cycle note will go on or off if the distant transmitter is turned on or off (keyed), and thus the circuit may be employed for code reception.

Up to the point where the receiver actually breaks into oscillation, voice or music can be heard, but after oscillations begin, only unintelligible squeals are heard unless both the carrier frequency of the incoming signal and that generated by the tube

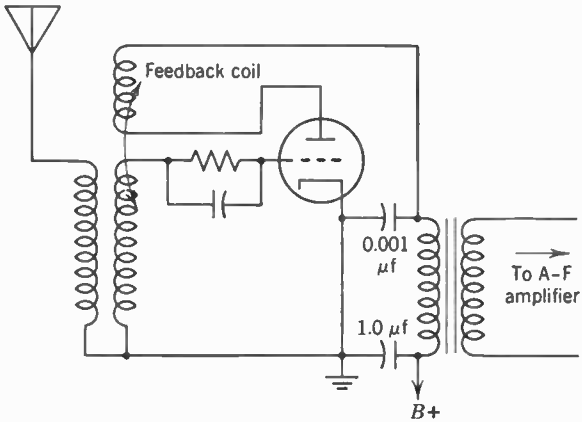


Fig. 2. Regenerative detector. Voltage fed back from plate to grid circuit increases sensitivity and selectivity.

itself are very stable. Then the operator may tune the receiver to exact resonance with the incoming signals and can hear music or voice. This procedure is not generally very practicable.

A regenerative receiver made up of a single regenerative detector and one stage of audio amplification is remarkably sensitive, considering the simplicity of the circuit and the small amount of apparatus required. It is most sensitive at the point where the detector is on the verge of breaking into oscillation. Here the pass band accepted by the tuned circuit of the receiver is quit narrow, since the effect of regeneration is to reduce the resistance of the tuned circuit so that its response curve becomes very sharp. At this point, the tone fidelity is not at all good. However, the regenerative circuit has adequate tone fidelity for many purposes.

At the point of greatest sensitivity, the detector is unstable and a change in d-c tube voltages or a mechanical jar or vibration may set the circuit into oscillation. Then the tuned circuit and its accompanying tube become a generator of signals (oscillator); and, if the coupling between the tuned circuit and the antenna is close, the signals produced by the receiver may radiate from the antenna, causing a disturbance which may be heard in other receivers over a distance that may be several miles. This is one serious disadvantage of such detectors. In wartime, such radiating receivers are not tolerated, and great care is taken to prevent any signals generated within a receiver from radiating.

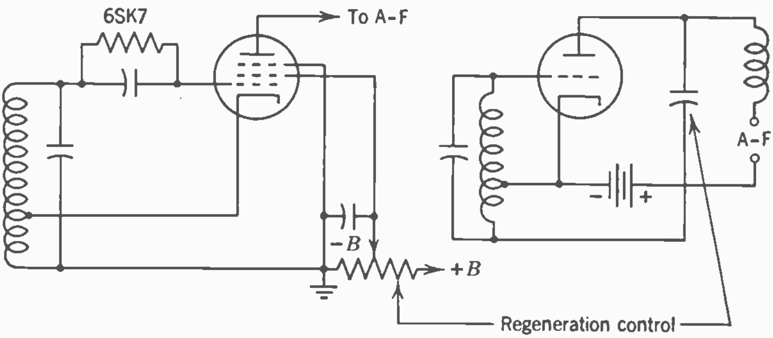


Fig. 3. Two feedback circuits for improving sensitivity of a detector.

Unwanted radiation from the antenna of a receiver employing a regenerative detector can be reduced materially by using a stage of r-f amplification ahead of the regenerative detector. The r-f stage not only acts as a buffer between the source of oscillation and the antenna but also produces a certain amount of amplification of the incoming signals and increases the selectivity of the circuit. A simple regenerative receiver which employs no preliminary r-f amplification is very broad in its tuning unless the regeneration applied is great enough that the detector is just on the verge of instability. The preliminary r-f stage helps the selectivity problem materially.

Feedback to produce regeneration may be secured in several ways. Two typical circuits, along with methods for the control of the amount of feedback, are shown in Fig. 3. A typical receiver employing regeneration is shown in Fig. 4. Regenerative receivers have two tuning controls—one for tuning the receiver to the incoming signals, and one for adjusting the amount of re-

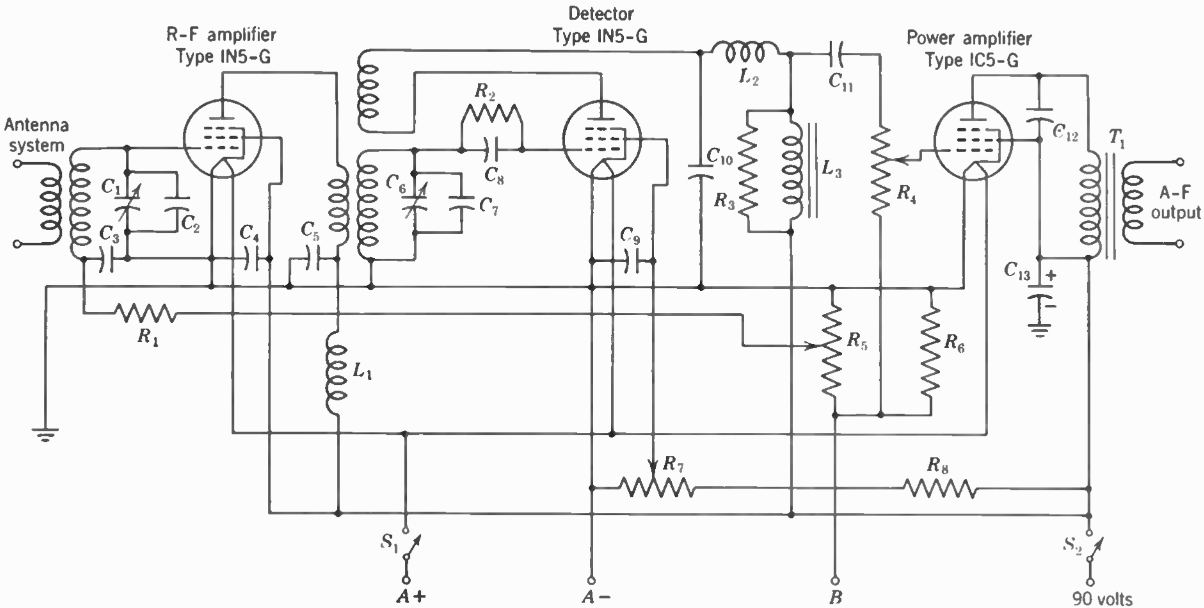


FIG. 4. Typical short-wave receiver using one stage of r-f amplification and a regenerative detector.

generation. A receiver of this type can be extremely sensitive since the application of regeneration may increase the response several hundred times compared to that of the same circuit without regeneration.

15.11. Superregenerative receiver. As discussed above, a simple regenerative circuit is most sensitive at the point at which it is on the verge of breaking into oscillation. In the superregenerative receiver, invented by Major Edwin H. Armstrong, means are provided for allowing the circuit to oscillate for a small fraction of a second, and then for breaking off the oscillations. That is, the circuit is allowed periodically to reach the oscillation point and is then forced to recede from it. This function is performed by an additional "quench" frequency, usually around 50 kc, impressed on the tube so that some characteristic, which tends to make the tube oscillate, is periodically brought into and out of the oscillating region.

A one- or two-tube receiver of this type is extremely sensitive. Motion of electrons in the input circuits produces a characteristic "rush" noise voltage until a carrier signal stronger than these voltages appears. Then the noise disappears and the desired signal may be heard. The "quench" frequency is kept out of the headphones or speaker by appropriate filtering networks.

The superregenerative circuit is most useful in the very high-frequency bands (5 to 10 meters), where it is difficult to get high amplification in any other way. The circuit is not selective, but this lack can be supplied by the use of one or more stages of r-f amplification ahead of the superregenerative detector.

15.12. Tuned-radio-frequency receivers. If a detector is preceded by several stages of r-f amplification, the receiver is called a tuned-radio-frequency or simply a t-r-f set. The tuned circuits, of which there may be as many as six, are all tuned simultaneously by means of a "ganged" condenser. This is a variable condenser with as many sets of plates as there are circuits to tune. Pentodes are usually used, coupled together by means of transformers. The primary of these transformers is usually of rather high inductance—4 mh is typical for the broadcast band—and the secondary coil inductance depends upon the tuning condenser size. For 550 to 1600 kc a secondary with an inductance of 270 μh is satisfactory when employed with a condenser having a capacitance range of 15 to 300 $\mu\mu$f. If the coil has a Q of 100 the voltage gain will be approximately 30 per stage. Much higher amplification is possible, but not with adequate stability.

The selectivity of a t-r-f amplifier depends to a large extent upon the characteristics of the interstage coupling transformers employed. The resistance of the coils and associated apparatus governs the selectivity; and since the resistance is different at different frequencies, the selectivity of a t-r-f receiver is different at different frequencies, being relatively high at the lower carrier frequencies (600 kc) and low at higher frequencies (1600 kc).

A t-r-f receiver is difficult to align and to keep lined up so that each of the individual stages tunes to resonance at the same setting of the variable condenser. Another difficulty is undesired feedback from output to input, which leads to instability and often to outright oscillation.

Before the introduction of screen-grid and pentode tubes, much trouble was experienced with triodes in t-r-f receivers because of the feedback through the relatively high grid-plate capacitance of the tubes. For this reason, t-r-f circuits employing triode tubes require special neutralizing circuits which balance out the effect of feedback by providing a second feedback voltage equal and opposite to that resulting from the grid-plate capacitance. These neutralizing circuits are not generally required in t-r-f receivers utilizing screen-grid or pentode tubes, and so the use of triodes in these circuits has been almost entirely discontinued. However, triodes are widely employed in the tuned circuits of r-f amplifiers in transmitters. The neutralizing circuits used in connection with these amplifiers are discussed in Chapter 17.

15.13. Superheterodyne receivers. By far the vast majority of present-day receivers employ the superheterodyne principle. In such a receiver, the incoming r-f signals are changed to a new frequency which is almost always lower than the carrier frequency. It is called the **intermediate frequency.** All incoming r-f signals are converted to this same intermediate frequency, and most of the amplification in the receiver can then be carried out at a single frequency. These amplifier circuits can be adjusted for the optimum conditions as regards both amplification and selectivity. The i-f amplifier can be made very selective or rather broad, depending upon the service for which the receiver is designed. After amplification in the i-f amplifiers, demodulation takes place in a conventional detector; the a-f output of the detector may be further amplified if desired.

A block diagram of a typical superheterodyne receiver is shown in Fig. 5. The modulated carrier is given some amplification and selection in the r-f amplifier and then passed on to the first detector or mixer stage. Also feeding a signal into the mixer is an oscillator which is an integral part of the receiver. The carrier and local oscillator frequencies f_c and f_o are "mixed" in the first detector so that the output frequency of this stage is the differ-

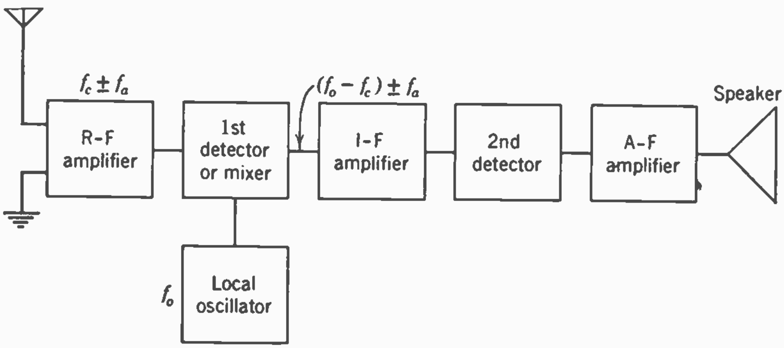


FIG. 5. Block diagram of a superheterodyne receiver.

ence $f_o - f_c$. This difference frequency is the intermediate frequency. If the carrier is modulated, the modulation is transferred to the intermediate frequency and the new frequency becomes $(f_o - f_c) \pm f_a$. Actually the carrier frequency may be higher rather than lower than the oscillator frequency, and this is often the case in short-wave receivers. In broadcast receivers, on the other hand, the oscillator frequency is almost invariably the higher for reasons which are discussed below.

The advantages of a superheterodyne compared to a t-r-f receiver follow: (1) All incoming signals are changed to a single fixed frequency and are amplified at this new frequency. In a t-r-f receiver, all amplification takes place at the carrier frequency. This raises the difficulty of securing uniform amplification and selectivity in an amplifier which tunes over a broad range of frequencies. (2) There is less danger of feedback troubles at the lower frequency for which the i-f amplifier is normally designed. Therefore, somewhat greater amplification can be secured with the same degree of stability, compared to a t-r-f receiver. (3) The selectivity of a stage of i-f amplification in a superheterodyne receiver can be made much better than that of a stage of r-f amplification in a t-r-f set.

This third advantage can be illustrated as follows: Suppose that a desired signal has a frequency of 1000 kc and that an unwanted signal is at 1010 kc. The unwanted signal is 10 kc away from the 1000-kc signal, or off resonance by 1 per cent. If both signals are converted to a lower frequency, for example to 456 kc for the desired 1000-kc signal, the undesired signal will be converted to 466 kc or about 2.2 per cent off resonance. If both the 1000-kc and the 456-kc resonance curves are equally sharp on a percentage basis, the 466-kc signal will be farther down its resonance curve than will be the 1010-kc signal, and will therefore produce less interference. This is an example of how superheterodyne action produces increased selectivity.

15.14. Heterodyne action. In a superheterodyne the frequency is changed by mixing the incoming signal with a locally generated signal that may be higher or lower in frequency than the incoming signal by the amount of the intermediate frequency. Consider an incoming 1000-kc carrier. If a locally generated signal of 1456 kc is mixed with the 1000-kc signal, new frequencies representing the sum and difference of the two frequencies will be produced. In this example, the difference frequency, 456 kc, will be utilized. This is the intermediate frequency. The local oscillator, which produces the locally generated signal, is tuned to a frequency higher than that of the incoming signal, and of course it must be tuned to a new frequency each time a carrier of different frequency is to be received. The problem arises, then, of maintaining a constant frequency difference of 456 kc between the carrier and the local oscillator frequency as the receiver is tuned to various stations. This is done, usually, by a system of series **(padder)** and shunt **(trimmer)** capacitors connected to the oscillator tuning condenser. Actually, the exact difference frequency of 456 kc is produced at only three points between 550 and 1600 kc, but the variation over the entire band is not great enough to give serious trouble. The sum frequency produced by heterodyne action is filtered out by the tuned i-f transformers.

15.15. Preselection. A stage or two of r-f amplification ahead of the frequency-conversion stage, although not an absolute necessity in a superherterodyne receiver, has several advantages. The advantages are as follow: (1) Coupling between the oscillator and the antenna is made very small, thus reducing the danger of interference in near-by receivers resulting from radiation of the local oscillator signal from the antenna. (2) Interference created by signals coming into the input of the receiver from a transmitter operating on a frequency equal to the intermediate frequency is reduced. (3) **Image-frequency** troubles, which come about for the following reasons, are reduced. If the intermediate frequency is f_i, and the oscillator frequency is f_o, then incoming signals having frequencies of either $f_o + f_i$ or $f_o - f_i$ will be heterodyned and will be amplified by the i-f system. For example, if the local oscillator is tuned to 1455 kc and if the intermediate frequency is 455 kc, then incoming signals of 1000 kc (1455 − 455 kc) or 1910 kc (1455 + 455 kc) will produce an i-f signal and be sent on through the system. Preselection reduces the voltage of the 1910-kc signal because the circuits of the preselector, being tuned to 1000 kc, will have little response to voltages of a frequency of 1910 kc. (4) A higher signal voltage will be introduced into the frequency-converter system and will increase the signal-to-noise ratio of the converter output, since the noise voltage in a converter is inherently greater than in an amplifier. The receiver, therefore, will be quieter in operation.

A tube used in the i-f system will provide greater amplification of the desired signal and greater selectivity to signals on channels adjacent (10 kc away) to the desired signals than if used in an r-f stage, and will cost less. The temptation, therefore, is to neglect the preselector in favor of added i-f amplification. A good superheterodyne should, however, have at least one stage of r-f amplification.

15.16. Frequency conversion. The simplest type of frequency converter (mixer) is one in which the carrier and local oscillator signals are together fed to the grid of a triode or pentode amplifier. The tube is biased so that the operation is on a curved portion of the transfer (E_c–I_b) characteristic curve—mixing of the signals occurs only when a non-linear portion of the curve is used. The output current of the tube then contains, in addition to the original frequencies put into the tube, new frequencies which are the sum and the difference of the carrier and oscillator frequencies. The plate tuned circuit selects either the sum or difference frequency and rejects all other frequencies. The tube used in this service is called a **mixer.**

The use of a single grid for both the carrier and local oscillator signals sometimes gives rise to difficulties resulting from coupling between the carrier input circuit and the local oscillator circuit. Several tubes have been developed for frequency conversion which make this coupling very small. In a **pentagrid converter**

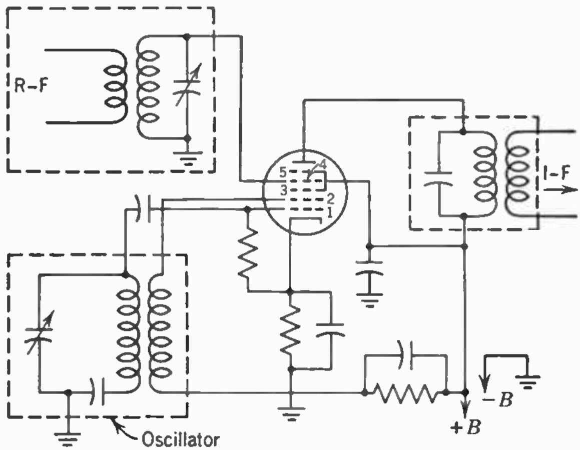


Fig. 6. Pentagrid converter circuit for frequency changing.

tube (such as a 6A8) there are five grids, a cathode, and an anode. With these seven electrodes, all the functions of oscillation and mixing take place. A typical pentagrid converter circuit is shown in Fig. 6. Grids 1 and 2 along with an external tuned circuit and the cathode form a triode oscillator. Grid 2 (actually a pair of side rods) acts as the anode of the oscillator. Thus, the first two grids and the cathode can be looked upon as a "virtual" cathode supplying for the remaining electrodes of the tube an electron stream which varies at the oscillator frequency.

Grids 3 and 5 are connected together and are used in the same way as a screen grid in any tube. These grids have the further function of shielding grid 4 from grids 1 and 2. Grid 4 is connected to the signal input, and its voltage varies at the frequency of the carrier. As the electron stream flows past grid 4 (the electron stream is already modulated at the oscillator frequency), it gets additional modulation, this time at the carrier frequency impressed on grid 4. The anode finally collects the doubly modulated electron stream. The output current contains frequency components at the carrier frequency, at the oscillator frequency, and at sum and difference frequencies just as in the simple mixer described above.

In a converter circuit employing a 6A8, good frequency conversion takes place up to moderately high frequencies, but, as the frequency is increased, the performance decreases for two

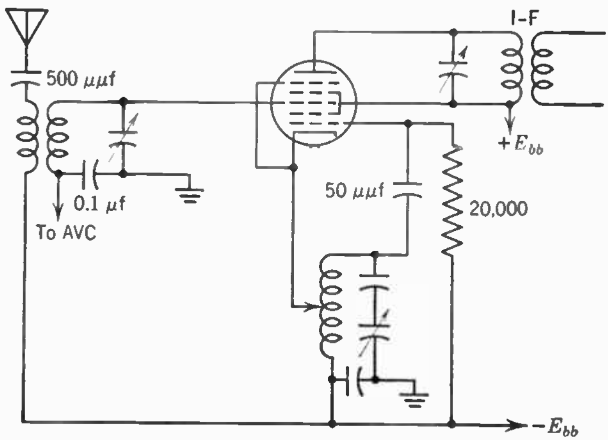


Fig. 7. Use of 6SA7 type of tube as frequency converter.

principal reasons. In the first place, the output of the oscillator falls off at the higher frequencies; in the second place, there are certain unavoidable couplings between the oscillator and the carrier frequency sections. The effect of these couplings increases with frequency.

The coupling between the oscillator and carrier input sections of pentagrid converter tubes is decreased in tubes such as the 1R5 and 6SA7 (Fig. 7) by a different arrangement of the grids. In these tubes grid 1 is the oscillator * grid. Grids 2 and 4 are connected together inside the tube and act as the oscillator anode and also as the shield for grid 3, the signal input grid. Grid 5 is a suppressor grid. Grids 2 and 4 are by-passed to the cathode. The coupling between the signal and the oscillator grid is very slight. In addition, changes of voltage of the signal grid, such

* This is a Hartley oscillator circuit. See Chapter 16.

as with a-v-c bias, have little effect on the transconductance of the oscillator grid or on the interelectrode capacitances associated with grid 1.

In the 6K8, a **triode-hexode converter,** still another tube structure is employed. This tube includes a triode oscillator section and a separate hexode mixer section. The triode grid and grid 1 of the hexode are connected together so that the oscillator frequency is injected into the hexode section. The functioning of the tube is not critical with respect to changes in oscillator

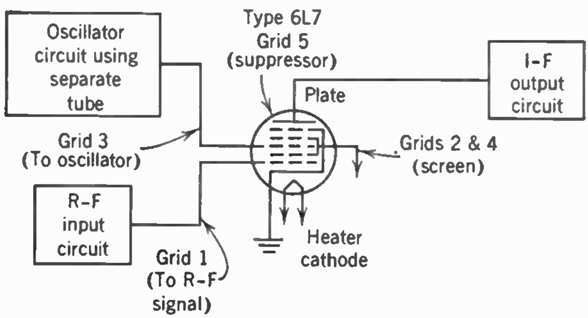


FIG. 8. Mixer tube and circuit useful at high frequencies.

plate voltage or signal grid bias. It is used in "all-wave" receivers in which it is important to reduce frequency shift of the oscillator at the higher signal frequencies.

The 6L7 tube, which requires a separate oscillator, as shown in Fig. 8, is a mixer tube especially adapted for high-frequency use. It has two control grids, 1 and 3. The first is a remote cut-off type and the second a sharp cut-off type. The incoming signal voltages are applied to grid 1 and oscillator voltages to grid 3. Grids 2 and 4 are connected together and act as a shield for grid 3 and at the same time accelerate the electrons. The final grid 5 is a suppressor. The 6L7 is known as a **pentagrid mixer** tube.

The **conversion transconductance** (g_c) of a converter tube is the ratio of the i-f current in the output to the r-f signal voltage input to the circuit. It is used in computing the i-f voltage output in a converter. The ratio of the i-f voltage output to the r-f voltage input is called the **translation gain.** The translation gain ranges from 30 to 80 for typical pentagrid converter circuits.

The frequency-converter tube is often called the "first" detector, and the entire system is sometimes called a "double detection" system since there are two detectors. A signal which has a frequency equal to that of the carrier plus twice the intermediate frequency also combines with the oscillator signal to produce an i-f signal and is called an "image" frequency; the i-f signal is called the "beat" frequency, and the process of frequency conversion by heterodyne action is known as "beating" one frequency with another.

15.17. Choice of intermediate frequency. When there is some preselection, the chance of image-frequency trouble decreases as the intermediate frequency is increased. On the other hand, with a low intermediate frequency the chances of harmonic voltages from the second detector being fed back into the input are lessened because only the higher and weaker harmonics would fall in the broadcast band. Also, lower frequencies are more easily amplified in the i-f amplifiers. Finally, more selectivity against adjacent channels can be obtained if the intermediate frequency is low.

Many intermediate frequencies are used, from 175 to several hundred kilocycles, depending upon the manufacturer, the type of set, the frequencies to be received, etc. It has, however, become almost standard practice to use intermediate frequencies in the vicinity of 456 kc for home radio receivers. In other types of low-frequency receivers the intermediate frequency may be as low as 60 kc. The image-frequency ratio of a high-class receiver is about 20,000 to 1. That is, the voltage of a signal differing from the desired signal by twice the intermediate frequency must be 20,000 times as strong as the desired signal to produce the same output.

Problem 1. A broadcast receiver is to cover the band from 550 to 1600 kc. If the local oscillator is tuned higher than the signal input circuit so that an intermediate frequency of 456 kc is produced, what must be the frequency range over which the oscillator tunes? Suppose that identical inductances are used in the tuned circuits and that the maximum capacitance of the input circuit tuning condenser is 300 $\mu\mu$f. What will be the approximate maximum value of the oscillator tuning condenser?

Problem 2. The intermediate frequency is 456 kc in an all-wave receiver. What frequency band must the oscillator cover if one of the short-wave bands tunes from 4.5 to 12 Mc?

15.18. Frequency converters. A frequency converter may be used in addition to a regular broadcast receiver so that high-frequency signals may be received. In such a system the receiver is tuned to a convenient frequency, and the separate frequency converter is designed so that its output frequency (for example, the difference frequency of the converter input and its local oscillator) is equal to that to which the receiver is tuned. Thereafter, all tuning is accomplished by varying the frequency of the oscillator along with the input circuit of the separate converter. The output frequency of the converter could be termed the "first" intermediate frequency. Occasionally a receiver intended for the reception of very high-frequency signals will incorporate two stages of frequency conversion operating at different intermediate frequencies. These are called **double-heterodyne** sets.

15.19. Band-pass amplifier. The ideal frequency-response characteristic of the r-f and i-f sections of a receiver would have a flat top and steep sides. The flat top should be 10,000 cycles wide for broadcast receivers, and the sharper the sides the more selective will be the receiver, without, at the same time, cutting off the higher modulation frequencies. An amplifier which has this characteristic is termed a "band-pass" amplifier since it passes and amplifies a band of frequencies and rejects frequencies both above and below this band. This ideal characteristic can only be approximated in practice. A typical characteristic is shown along with the ideal characteristic in Fig. 9.

The r-f sections of a receiver must tune over a wide range of frequencies and so do not have a sharp frequency response. However, the i-f amplifiers operate at a fixed frequency, and these amplifiers can be designed to have a response approaching the ideal. The usual coupling between i-f stages consists of a transformer with tuned primaries and secondaries. If the two coils are individually tuned to the operating frequency and then coupled, as by moving them close together, the resultant response curve may be a singly peaked curve, like that of a single tuned circuit, or a flat-topped curve, or a curve with two more or less widely separated peaks with a substantial dip in between. The shape of the resultant curve is controlled largely by the degree of coupling.

The result of coupling two such coils together is shown in the curves of Fig. 10. It will be seen that very close coupling ($M = 14\ \mu$h) produces widely separated peaks, less coupling gives a comparatively flat-topped characteristic, and rather loose coupling gives a sharply tuned circuit. In a multistage i-f amplifier various combinations of these characteristics may be used

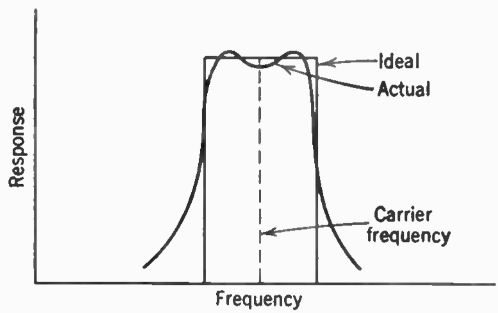


FIG. 9. Ideal and actual response of band-pass amplifier.

to produce an over-all response approaching the ideal. For example, one stage might utilize a characteristic such as 3 and the following stage one such as 2. The sharply tuned characteristic of 2 could fill in the dip in 3, and the over-all characteristic could be made rather flat-topped with steep sides.

Multistage i-f amplifiers sometimes employ "stagger tuning"; that is, the coils in successive i-f stages are tuned to slightly different frequencies. Then, since the peaks occur at slightly different frequencies, the amplifier can be adjusted to have a good over-all response. Both double- and triple-stagger-tuned amplifiers are used.

In a t-r-f receiver, in which all the high-frequency amplifiers work at the same frequency and are tuned together, the width of the band passed may differ at each frequency to be received. If the interstage coupling is by transformer the band width will be broad at the high frequencies; if the coupling is capacitive the band width will be broad at low frequencies. A combination of transformer and capacitive coupling may be used to keep the

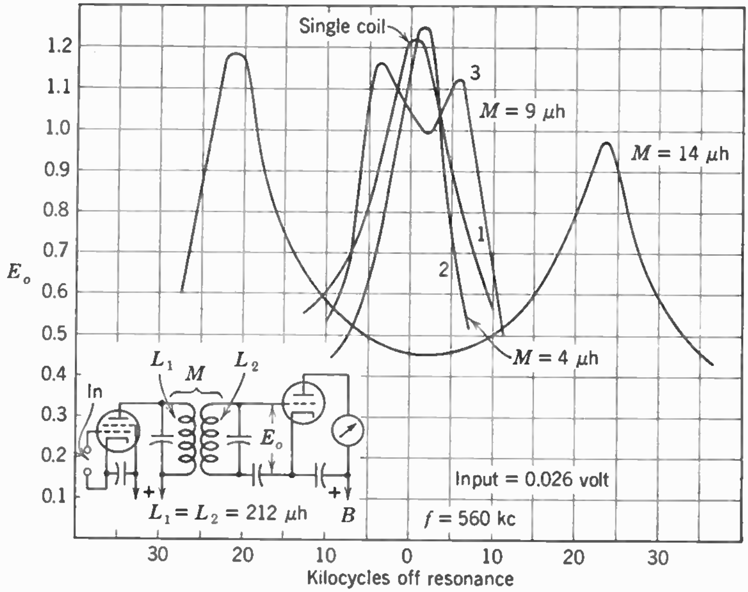


FIG. 10. Experimental determination of effect of close coupling in band-pass filter.

band width nearly constant over the tuning range. However, the selectivity of a t-r-f receiver is usually poorer than that of one employing the superheterodyne principle.

15.20. Second detector. After sufficient amplification of the carrier signal at radio and intermediate frequencies, any type of detector discussed in Chapter 14 may be used to remove the modulation and thus produce an a-f output signal. Most modern sets have a diode detector followed by one or two stages of a-f amplification. The direct currents produced by the detection process are utilized in an a-v-c circuit to maintain the i-f voltage supplied to the detector nearly constant in spite of rather large variations in the voltages picked up by the antenna.

15.21. Tone control. Manual tone control is incorporated in many receivers. These tone-control circuits are arranged so that frequencies at either end of the audio range may be accentuated or decreased with respect to each other or to the middle frequency region. Often a tone control is nothing more than a device for reducing the high-frequency response of the receiver in order to reduce the noise generated within the set or coming in on the antenna. Such a circuit consists of a resistance in series with a capacitance, the combination being shunted across the input to the audio amplifier. The capacitance shunts out the higher audio tones; the resistance, which is the variable feature, increases or decreases the shunting effect.

15.22. Automatic tone control. When static is bothersome, as when receiving a distant or weak station, cutting out the higher audio notes is advantageous. Often a program that is hopelessly lost in noise can be received with a certain amount of pleasure if the control is advanced to the point where little beyond 2000 cycles is passed through the audio amplifiers.

Circuits have been devised which give some measure of automatic tone control. When the receiver r-f gain is low, as it would be if receiving a strong local station, the tone-control circuit is not in operation. However, for weak stations the tone-control circuit, which is connected to the a-v-c system, cuts off more and more of the high audio frequencies. In some circuits this action is inherent; in others, an additional tube acting as a variable resistance or reactance is employed. Automatic-tone-control circuits are not widely used on broadcast receivers.

The use of **automatic-selectivity-control (a-s-c)** circuits is another approach to the problem of reducing the effect of noise when weak stations are being received. The elements of a simple a-s-c circuit are shown in Fig. 11. The additional inductance in the transformer secondary is placed in and out of the circuit by the switch. With this inductance added to the circuit the i-f transformer has a typical overcoupled characteristic. With the inductance switched out of the circuit, the response of the i-f transformer is sharper and the higher frequency-modulation tones are cut off. The switch shown in the figure may be operated by circuits connected to the a-v-c line, or it may be manually operated.

15.23. Noise-suppression systems. In receivers employing avc the sensitivity is at a maximum when no signal is being received, for example when the receiver is being tuned between two stations. Therefore, any static or other noise will be amplified to the limit of the circuit and be passed to the loud speaker. Circuits have been developed which shut off this noise. They operate in various ways. For example, a tube may be so connected that its plate current will be high in the absence of a carrier signal. This plate current can be used to supply bias voltage to the audio amplifiers. Absence of carrier signal can then be used to overbias the a-f amplifiers and reduce their amplification to practically zero.

These are known as "quiet a-v-c" or "squelch" systems. They

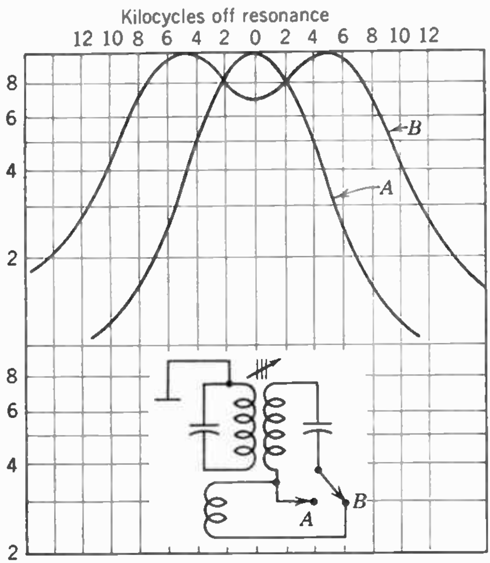


Fig. 11. Selectivity curve for two-position i-f transformer.

are not used to any extent on home receivers, especially since push-button sets, with which interchannel noise is not a problem, came into vogue. On the other hand, receivers used at fixed frequencies, such as police or taxi receivers, almost always have squelch systems so that the receiver is quiet until the transmitter comes on the air.

15.24. Automatic frequency control. Another automatic feature which has been incorporated into some receivers is a circuit for maintaining the oscillator frequency accurately at the desired value. The frequency at which an oscillator operates is determined by several factors such as the tuning capacitance, the inductance, the tube capacitances, and the tube parameters. If, after a receiver is tuned to a desired station, any of these values should change because of temperature or line voltage variations, or for any other cause, the oscillator would drift from its correct value and would fail to produce the proper intermediate frequency. Furthermore, the difficulty of accurately tuning a highly selective superheterodyne receiver with several i-f stages is very real. Bad distortion results if the intermediate frequency is not the correct value.

15.25. Automatic-frequency-control (a-f-c) circuit. A basic circuit for an a-f-c system is shown in Fig. 12. A tube feeds

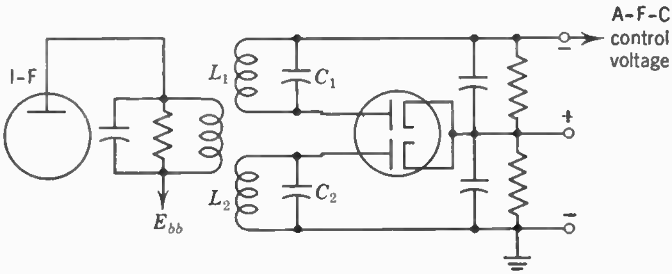


FIG. 12. Basic circuit for automatic frequency control.

the i-f signal to a transformer with two tuned secondary windings. The upper secondary is tuned slightly higher and the other slightly lower than the exact intermediate frequency. If the intermediate frequency is exactly half way between the two frequencies to which the secondaries are tuned, the secondaries have the same magnitude of voltage impressed across them and the rectified output voltages of the diodes are equal. If, however, the intermediate frequency is slightly higher than the correct value, then the upper tube has somewhat greater voltage impressed upon it than does the lower tube, the rectified voltages of the diodes will be somewhat different (that of the upper diode being greater), and a negative voltage will be produced at the a-f-c terminal. If, on the other hand, the intermediate frequency is slightly low, the voltage on the lower diode is greatest and the resulting a-f-c voltage is positive. The capacitors shunting the resistors in the cathode circuits of the diodes serve to filter the rectified output so that only direct voltage appears at the a-f-c terminals.

The d-c voltage appearing at the a-f-c terminals is applied to the grid of a tube which acts as a variable reactance. This variable reactance is connected into the tuned circuit which controls the frequency of the oscillator. This variable reactance may be either inductive or capacitive. For example, if it is inductive, a negative a-f-c voltage (indicating that the intermediate frequency is too high) would be used to increase the inductive reactance and thus decrease the oscillator frequency. A positive control voltage would decrease the inductive reactance and thus increase the oscillator frequency. The tube circuits which provide these variable reactances are called "reactance tubes." They are discussed in connection with frequency-modulation systems.

Automatic-frequency-control circuits require so many tubes and other components that they are not widely used in broadcast receivers. Their development and success have, however, led to investigations of other means of maintaining the oscillator frequency at the correct value. As a result manufacturers have greatly improved the design of coils, condensers, and tubes so that they are less sensitive to variations in temperature, vibration, or changes in voltage.

15.26. Push-button receivers. Three types of push-button tuning systems have been applied to receivers, one or the other of which is now found on most automobile sets and on many home receivers as well. The advantages are speed of tuning from one station to another and low interchannel noise.

In one system, individually tuned circuits are provided for each push button. These circuits are tuned to the desired stations by variations in capacitance or inductance. The push buttons are mechanically connected to switches which select the required tuned circuits. The number of stations that can be tuned in is limited by the number of push buttons provided. Changes in the setting of a tuned circuit are made by a simple screwdriver adjustment.

In a second system, the gang-tuning condenser may be rotated throughout its range by means of the conventional dial. In addition, cams are attached to the condenser shaft so that when a push button is depressed the shaft is rotated to the proper position to tune a desired station. The positions of these cams on the shaft may be changed easily.

In a third system an electric motor is utilized to turn the condenser shaft. This method is more expensive and more likely to get out of order, but it permits remote control and has the advantage that any desired channel may be selected.

None of these systems would work well without the development of stable components such as condensers and coils which do not vary in their electrical characteristics with time, temperature, humidity, or vibration.

15.27. Tuning indicators. Highly selective receivers must be correctly tuned; otherwise severe distortion results. A variety of tuning indicators has been developed. Most of them are actuated by the direct current in diode detector load resistance or by the a-v-c voltage.

Since the strength of the d-c current in a detector load is an indication of the signal being applied to its terminals, a sensitive d-c meter placed in series with the load resistor will indicate when the receiver is accurately tuned. The meter, in this service, is usually called an **S meter.**

The a-v-c voltage itself is also a measure of the signal strength at the detector terminals. This voltage may be used to actuate an **electron-ray tube** (also called a "magic eye"). Such a tube indicates visually the condition of correct tuning by means of a fluorescent target which is bombarded by electrons and which glows when so bombarded. There are several types of such tubes, but each depends upon the same fundamental principle. Between the cathode (source of electrons) and the fluorescent target is a ray-control electrode or deflecting vane. The voltage difference between the deflecting vane and target determines how large a sector of the fluorescent screen will glow.

In some electron-ray tubes, such as the 6AB5, there is a triode in addition to the ray tube. The triode acts as a d-c amplifier. The grid of this tube is supplied with voltage from the a-v-c system as shown in Fig. 13(*a*). The flow of plate current, under the control of the voltage supplied to the grid of the triode through the 1-megohm resistor, fixes the voltage of the deflecting vane of the indicator section of the tube. When more plate current flows into the triode through R, the voltage drop in R increases and there is less voltage on the triode plate and deflecting vane of the indicator section, which are electrically connected together. The target is held at a fixed voltage E_{bb}, and so the shadow on the target increases. When the triode grid is more negative, less plate current flows, the deflecting vane becomes more positive, and a larger sector of the target is illuminated.

In another type, the 6AF6, there is no triode amplifier, but there are two control electrodes, one on either side of the cathode

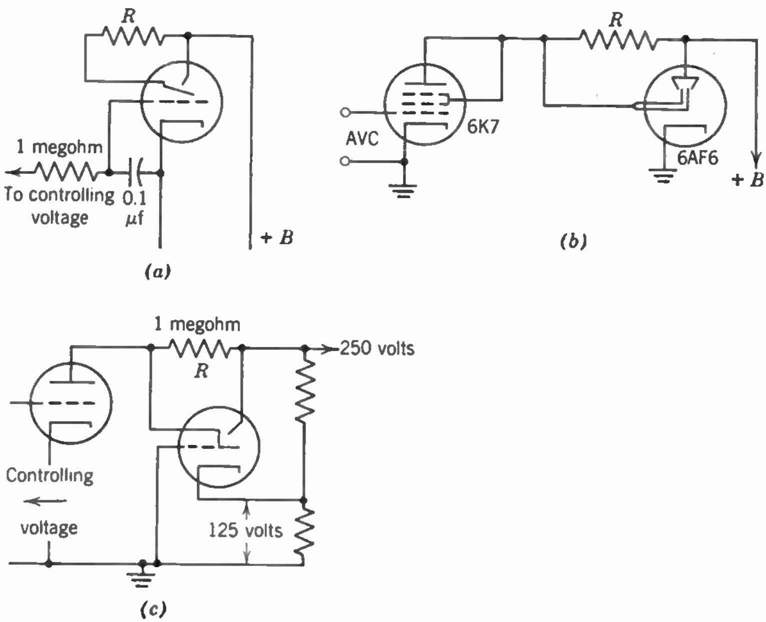


FIG. 13. Electron-ray tuning indicator tube circuits. In (b) and (c) separate d-c amplifiers are used, whereas in (a) the triode amplifier is in the same envelope with the indicator element.

and affecting different portions of the target. Thus, two symmetrically opposed patterns or two unlike patterns may be obtained, depending upon how the ray-control electrodes (deflecting vanes) are connected. This tube requires an external d-c amplifier.

Electron-ray tubes are available with either sharp or remote cut-off triode sections.

In operation, electron-ray tubes get their controlling voltages from the a-v-c system. Since the maximum negative voltage is produced in the a-v-c system when the receiver is properly tuned to a station, a maximum negative voltage is supplied to the control grid of the triode acting as a d-c amplifier for the indicator tube. Under these conditions the least plate current will flow through R, the least voltage difference will appear between the deflecting vane and target, and there will be the smallest shadow angle on the target.

Electron-ray indicator tubes are found in many measuring instruments, taking the place of indicators of the milliammeter type. The sensitivity of the indication may be increased by using a separate d-c amplifier to control the action of the deflecting vane as shown in Fig. 13(c).

15.28. Loud speakers. The loud speaker is usually the final link in a radio system, and because of its position with respect to the listener it is frequently blamed for bad reproduction that really originates elsewhere in the receiver.

A loud speaker converts the electrical energy output of the final power tube into sound energy. It should do this as effectively and faithfully as is necessary for the particular job at hand. It is useless to design and operate a high-class amplifier with a poor loud speaker. The wide range of tones coming from the amplifier is lost in the loud speaker and does not get to the listener. Likewise, it is absurd to install a high-grade speaker in a poorly engineered receiver in the hope that the fidelity of reproduction will be bettered. The full benefit of a wide-range speaker cannot be attained unless the complete chain of apparatus is also properly designed—the amplifiers, detectors, plate voltage supply apparatus, etc.

15.29. Elements of a loud speaker. A loud speaker, in general, has three essential elements. First, a **motor** converts electrical into mechanical energy. The mechanical energy is transferred to the second element, the **diaphragm,** which vibrates and sets air in motion, creating sound waves. Third, a **horn** couples the diaphragm to the "mechanical impedance" of the air load. Thus the horn is a sort of mechanical transformer to couple the mechanical impedance of the diaphragm to the load, which is the air that is to be set in motion. Many speakers have no horn, as such, but the diaphragm is extended in size and accomplishes the needed coupling by itself.

Each of these elements may take one of several forms, and one assembly may be considerably more efficient than another, that is, more of the electrical energy may be converted into acoustical energy in one type of speaker than in another. Some speakers will handle quite a lot of electrical energy with relative freedom from distortion. Other types are limited in the volume they can handle. Some types will reproduce a wide range of frequencies, and others will convert electrical into acoustical energy only over narrow frequency bands.

The engineering of speakers is quite as complex as the engineering of an audio system or the designing of a radio receiver. The terms and quantities used are different, but there are certain analogies between the electrical and mechanical-acoustical systems which can be used to advantage. Each design involves coupling a generator, the motor, to a load, the air, at as great an efficiency as possible and over as wide a band of frequency as necessary.

15.30. The horn speaker. In a horn speaker the motor and disk-like diaphragm are small but capable of intense mechanical vibrations. In appearance and operation the motor and diaphragm are similar to those of the well-known telephone receiver. The sound waves generated by the diaphragm are coupled to the air through the horn. The horn must permit the air vibrations at the diaphragm to expand in such a manner that they will leave the horn without reflecting back into it. The small end of the horn is called its **throat;** the large end, its **mouth.** The rate and manner in which the horn changes its shape are known as the **taper.** The relative dimensions of the mouth and throat determine the load on the diaphragm. The taper determines the amount of reflections back into the horn.

The lowest frequency that can be effectively radiated into space is controlled by the size of the mouth. The larger the mouth, the lower the frequency which the horn will radiate without excessive distortion.

The taper controls the efficiency of sound transmission along the length of the horn. The nature of the taper also has a marked effect on the characteristics of the speaker. An exponential taper has been widely used since it gives good transmission over a wide frequency range, but has the disadvantage that the mouth must be extremely large if low frequencies are to be properly radiated. In recent years, several new tapers have been worked out which give appreciably better performance at low frequencies and yet do not require that the mouth be excessively large.

Horn speakers are not suited for home receivers and similar applications because of their size. They are widely used in conjunction with auditorium and outdoor public-address systems.

15.31. Diaphragm speakers. A diaphragm speaker uses a light paper or metal cone to couple to the air load. To transmit from a localized source of sound to an acoustic air load spread over a large area, the cone diaphragm must be light (to avoid loading up the motor with a resistance load) and at the same time it must be rigid so that it does not break up into all sorts of unwanted vibrations when driven by the motor. The highest frequency that will be transmitted satisfactorily will depend upon the size and weight of the cone, a small and light cone transmitting high frequencies better than a large and heavy one. On the other hand, to handle low frequencies the cone must be large to move a large mass of air. These two requirements are contradictory and so the designer must make some sort of compromise. If a wide range of frequencies is to be reproduced, two speakers are often used: a large speaker to transmit the low frequencies, and a small speaker to transmit the high frequencies. These two speakers may be built as an integral unit on the same axis, or may even be part of the same diaphragm. Or they may be distinct speakers, physically displaced from each other.

Cone speakers can be made which artificially raise the low-frequency response. These speakers generally distort badly, but are preferred for some applications. Their quality of reproduction is poor, as can easily be determined by comparison with the low-frequency reproduction from a high-quality speaker.

15.32. The moving-coil speaker. The construction of the moving-coil or dynamic type of speaker is shown in Fig. 14. A strong permanent magnet, or an electromagnet energized by direct current from the plate supply system or from a 110-volt a-c line by means of a rectifier, furnishes a steady field. The voice-frequency currents from the final power tubes in the amplifier are passed through a few turns of wire, the small movable coil, to which is attached the diaphragm. When alternating currents flow through the coil, the coil tends to move at right angles to the lines of force across the air gap. The motion of the coil is imparted to the cone and thence to the air.

The impedance of the movable coil is very low, of the order of 2 to 15 ohms, and is fairly constant over the audio range, rising at high frequencies because of the inductance of the coil. Because of this low impedance, the power from the output tubes must be fed to the speaker coil through a step-down transformer. The coil can move through a considerable distance, and good

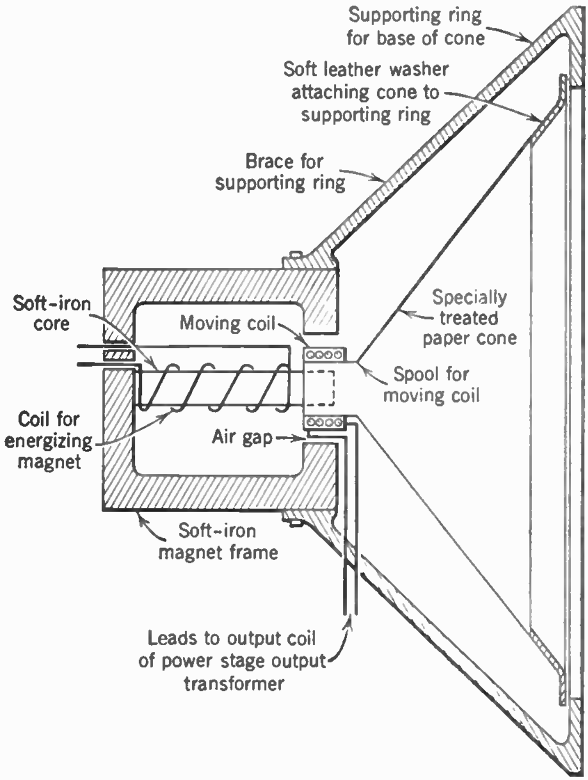


Fig. 14. Modern "dynamic" or moving-coil loud speaker.

low-frequency response of such a speaker is possible. Considerable sound energy can be delivered by a moving-coil speaker. The resonant frequency of the moving coil and cone is usually lower than the lowest audio tone to be reproduced.

The impedance of the speaker tends to rise at high audio frequencies, owing to the inductance of the moving coil, so that distortion is caused when the speaker is operated from a high-impedance source. Therefore condensers are shunted from plate to cathode of pentodes and beam power tubes, used in the final power amplifier, to decrease the distortion which results from the high-frequency impedance.

15.33. Baffles for dynamic speakers. It is necessary to install a dynamic speaker in the center of a rather large and heavy baffle if the low notes are to be properly reproduced. Otherwise the wave set up from the back of the cone can interfere with the wave set up from the front, and little or no sound will get to the listener. The baffle increases the air path between front and back and should be great enough so that the shortest

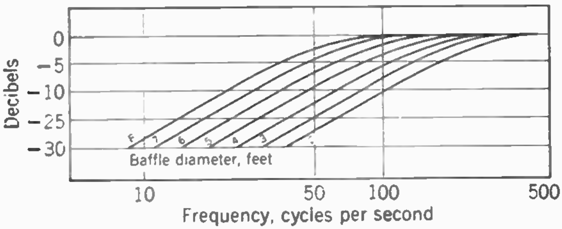


FIG. 15. Dynamic speaker response as controlled by baffle size.

mechanical length between front and back edges is at least one-quarter wavelength for the lowest note to be reproduced. Since the wavelength of sound, like that of radio waves, is equal to the velocity at which it travels (1100 ft per sec in air) divided by the frequency, it is not difficult to prove that a baffle at least 32 in. square is necessary for notes as low as 100 cycles, and a 110-in.-square baffle is required for notes as low as 30 cycles (see Fig. 15). When the unit is mounted in a box, peculiar resonances are set up which may spoil the good qualities of a dynamic speaker. These resonances are sometimes smoothed out by the introduction of resonating chambers or diaphragms which absorb energy at the offending frequency.

15.34. Improving loud-speaker performance. Several methods have been developed for increasing the bass response of speakers (for it is at the low frequencies that good coupling between the diaphragm and the air load is difficult to obtain) and for eliminating troublesome resonances. One method of reducing resonance in the cabinet is to place cones (not driven by motors) in front of the cabinet. These cones are of such dimensions that they resonate to the frequencies at which the cabinet itself "booms" and thus absorb energy and act as dissipators of the unwanted resonant energy.

To eliminate the loss of low frequencies because the wave from the back of the diaphragm, at low frequencies, cancels some of the energy radiated from the front of the diaphragm, several schemes have been applied. One method makes use of

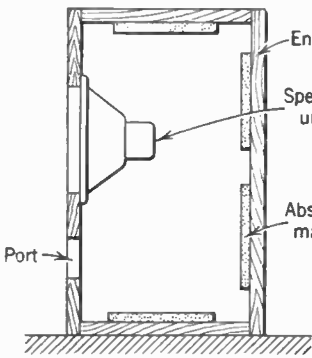


FIG 16. Bass reflex housing for improving low-frequency response of dynamic speaker.

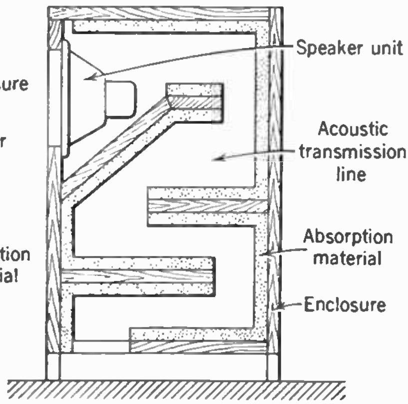


FIG. 17. Construction of the labyrinth enclosure for dynamic speaker.

a total enclosure like a box closed on all sides. Ordinarily with such a cabinet the system would boom at some frequencies and at low frequencies would be very inefficient. If, however, vents or holes of the right size are placed in the front of the enclosed cabinet correctly with respect to the speaker diaphragm as shown in Fig. 16, the low-frequency response will be increased appreciably and at the same time the distortion at low frequencies will be much less.

The "bass reflex" principle just described makes it possible to improve the low-frequency response by as much as 10 db compared with that obtained from a total enclosure without the vent. The distortion may be reduced by as much as one-third. The vent acts like a second speaker which furnishes acoustic power only at frequencies near the cut-off frequency of the speaker. This power comes from the back side of the speaker cone. Higher-frequency acoustical power is absorbed within the cabinet by lining it with material which is highly absorptive at all but low frequencies.

Another method of improving low-frequency response is to connect an acoustical transmission line, or labyrinth, to the diaphragm in such a manner that it brings the back-side radiation of the diaphragm in phase with the front-side radiation and thereby improves the response. Such an arrangement is

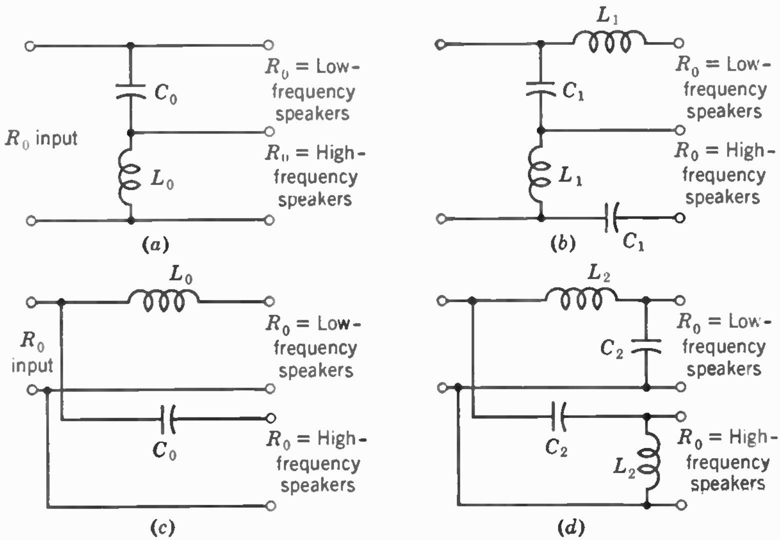


FIG. 18. Dividing networks to keep low frequencies out of high-frequency speaker, and vice versa.

shown in Fig. 17. Distortion at low frequencies is reduced compared with that of a cone speaker without the labyrinth.

Two speakers are often used when a wide range of frequencies must be reproduced. In this system a small speaker, or "tweeter," is used for the high frequencies and a large speaker, called a "woofer," for the low frequencies. By means of frequency-dividing networks, examples of which are shown in Fig. 18, low-frequency currents are kept out of the small speaker and high-frequency currents out of the large speaker. Thus each speaker is required to handle only that power which will produce useful acoustical output. Each speaker can then be selected with its own restricted frequency range in mind, and an excellent over-all frequency response secured.

16 · Oscillators

Not only will an amplifier tube increase the amplitude of an alternating voltage placed upon its grid but it will also generate alternating currents in the absence of any exciting voltages from an external source. The frequency of these currents may be anything from a few cycles per hour to hundreds of millions of cycles per second; the efficiency of generating a-c power from d-c sources may be as great as 70 per cent or even more; the stability and constancy of frequency and amplitude may be very great; and the power output may range up to many kilowatts. Tubes and the associated circuits which perform this function of converting d-c power to a-c power are called **oscillators.**

16.1. Types of oscillators. There are several possible ways of classifying the various types of oscillators, according to frequency range, stability, power output, waveform, etc. For the purposes of this chapter it will be most convenient to divide oscillators into three groups according to the general method used for obtaining oscillation: (1) feedback oscillators, (2) negative-resistance oscillators, and (3) relaxation oscillators.

In feedback oscillators a part of the output power is fed back into the grid circuit to reinforce the signal already there. These oscillators generally employ a tuned circuit to establish and maintain the oscillating frequency. The output waveform is essentially sinusoidal. Feedback oscillators may be used to generate frequencies ranging up to hundreds or even thousands of megacycles. Most of the oscillator circuits found in radio receivers and transmitters are of this type.

Negative-resistance oscillators utilize the negative-resistance characteristic of a tube or combination of tubes to produce oscillation. Their frequency of oscillation is normally fixed by an *LC* circuit, and the waveform is usually sinusoidal.

Relaxation oscillators are generally used to generate non-sinusoidal waves such as triangular or square waves. Their operation is generally based on the "triggering" action of a tube and its associated circuit, as, for example, the firing voltage of a gas tube. Only in special cases can a true sinusoidal waveform be obtained from relaxation oscillators. These oscillators are used to provide sweep voltages for cathode-ray oscillographs, to provide the signal from which the complex waveforms for television or radar are derived, and in many other applications for which shaped waves are required.

16.2. Feedback-oscillator circuits. The possibility of oscillations occurring in an amplifier as a result of positive feedback from the plate to the grid circuit has already been discussed briefly. Oscillations occurring in this manner are unwanted and, generally speaking, uncontrolled. The amplitude of the voltage and the frequency are determined by whatever the circuit arrangement happens to be. Now, however, we must look more deeply into this phenomenon and investigate methods for producing oscillations at a desired frequency, and of the required strength, both under the control of the designer or operator.

Consider first a simple amplifier (Fig. 1, Chapter 11). Suppose that 1 volt impressed on the grid-cathode circuit appears as 10 volts across a load resistance in the plate circuit. The gain of the circuit is, therefore, 10. Now suppose that by some sort of network a part of the output voltage, say 1 volt, is inserted in series in the grid circuit and in phase with the original 1-volt input. The grid-cathode circuit now has 2 volts across it, and the output voltage will be 20 instead of 10 provided that the gain remains the same. Now the original 1 volt of exciting voltage can be removed and the output will drop to 10 volts, with the tube now furnishing its own excitation. It has become a self-excited oscillator. It is producing a-c power from the d-c power furnished by the d-c plate supply system.

It is essential, of course, that the network feeding the voltage back into the grid circuit insert this fed-back voltage in the proper phase. That is, the feedback voltage must be increasing, say in a positive direction, when the input or grid voltage is increasing in a positive direction. In this manner, the feedback voltage increases the grid-cathode voltage. If the phase is reversed, a negative feedback amplifier rather than an oscillator results.

Not only must the voltage fed back be of the proper phase, but the power fed back must be large enough to supply all the power losses in the input circuit; otherwise oscillations will not result. The amplifier will merely become a regenerative or positive feedback amplifier with more gain than before.

A feedback oscillator can be likened to many mechanical oscillatory systems, as, for example, a pendulum clock. The essential elements of an oscillatory system of this type are (1) a mechanism capable of alternately storing and releasing energy at a definite rate, (2) a synchronous tripping mechanism which will supply enough energy to the system at the proper time to make up for power losses in the system, and (3) a source of unidirectional energy or power. In the clock the pendulum serves as the storage element. The frequency at which it swings to and fro is determined by the length of the pendulum and the pull of gravity. At the extreme ends of its swings energy is stored as potential energy; that is, the bob is acted upon by the force of gravity attempting to pull it down to its lowest position. As the bob swings through its lowest position it loses its potential energy but has its maximum kinetic energy due to its motion. The synchronous tripping mechanism in the clock is the escapement. This device serves to give the pendulum a little push once each cycle and makes up for losses suffered as a result of friction and the like. The source of energy is found in the weights. These are always pulling downward, and so this source of energy is unidirectional—always downward.

In a feedback oscillator these three essential elements may also be identified. The circuit which may store and release energy at a resonant rate is a parallel *LC* circuit. Maximum energy is stored alternately in the coil and then in the condenser. The rate at which the alternations occur is determined primarily by the resonant frequency of this *LC* circuit. The synchronous tripping device is made up of the feedback loop and the grid of the tube. The source of energy is the d-c plate supply.

The same essential elements are found in a rope swing, a pocket watch, and many other common oscillatory systems.

The question that almost invariably arises in regard to vacuum-tube oscillators is, How does an oscillator get started if no initial exciting voltage is supplied from an external source? The answer is relatively simple. Because of the feedback loop any voltage, however small, will be amplified in the tube and recirculated in the circuit if the voltage is of the proper frequency. Almost any transient disturbance of the circuit will produce a minute amount of voltage of the proper frequency to start the oscillations. This disturbance may come from closing the switch to the d-c power supply. It may result from the random motion of the electrons as they move to the anode.

Certain losses occur in an oscillatory circuit because of the power that is wasted in the circuit resistance. This resistance resides in the coil, the condenser, the lead wires, the tube itself, or in any external circuit which may be coupled to the circuit. For example, consider only the resistance of the coil. Suppose that it is 5 ohms. If 1 amp flows through this coil, 5 watts of power is consumed in the resistance. The feedback from the output of the oscillator must supply this 5 watts. There is no other place it can come from. If the feedback supplies only 4 watts, oscillations will cease; and if the circuit could feed back, say 6 watts, then the oscillations would continue to increase and the system would soon burn up. Actually, the circuit tends to stabilize at the point at which the power losses are just balanced by the power fed back.

A circuit in which the losses are high is more difficult to force into oscillation than one in which there is little power loss. Oscillators, therefore, are made up of high-Q tuned circuits. In this way oscillation is insured, and only a small part of the power generated is lost in the oscillator itself. More can be taken out for useful purposes, driving power into an antenna, for example.

An oscillator can be looked at in another way. In the case given above where the coil had a resistance of 5 ohms, let it be connected to a circuit which acts as though it had 5 ohms of "negative" resistance. Now the two resistances are added together, and, because one is the opposite of the other, they cancel out and the circuit now has no resistance. In other words a supplier of power may be thought of as a device which has negative resistance; instead of consuming power in a positive resistance it supplies power.

The tuned LC circuit may be placed in either the grid or the plate circuit of the oscillator tube or separate tuned circuits may be placed in each; or a single tuned circuit may be shared by the grid and plate. Thus, there are several possible combinations which will produce oscillations.

A simple oscillator circuit is shown in Fig. 1. This is called a **tuned-grid** oscillator since the tuned circuit is in the grid circuit. A feedback coil in the plate circuit couples energy from the output into the input circuit. The *RC* circuit connected directly to the grid is to provide grid-leak bias, to be described later.

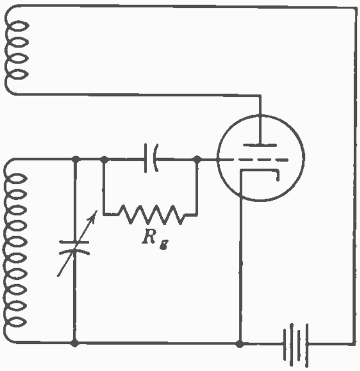


FIG. 1. Tuned-grid oscillator.

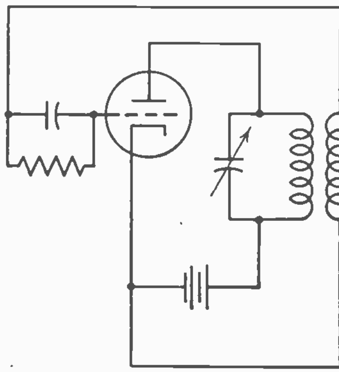


FIG. 2. Tuned-plate oscillator.

A **tuned-plate** oscillator is shown in Fig. 2. Except that the positions of the feedback coil and the tuned circuit are reversed, the circuit is much the same as that in Fig. 1.

In Fig. 3 is shown a **tuned-grid tuned-plate** oscillator. For proper operation the tuned circuits in the grid and plate leads are tuned to approximately the same frequency but there is no necessity for them to be physically near each other. The coupling between them is through the grid-plate capacitance of the tube.

In the three circuits discussed above the plate supply battery occupies its normal position between the lower end of the plate load circuit and the cathode. These are called **series-fed** oscillators.

Two circuits in which the tuned circuit is shared by the grid and plate are shown in Figs. 4 and 5. In the **Hartley** oscillator (Fig. 4) the cathode connection is made to a tap on the coil of the resonant circuit. The amount of grid excitation and the efficiency are determined by the position of this tap. Moving the tap towards B increases the grid voltage because a greater share of the total voltage across the coil now appears between grid and cathode. Neither side of the tuning condenser is at ground po-

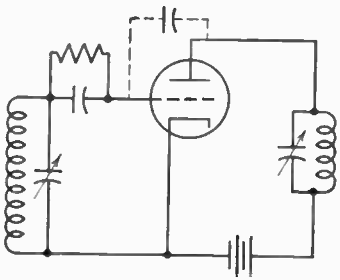


FIG. 3. Tuned-grid tuned-plate oscillator. Feedback is through grid-plate capacitance.

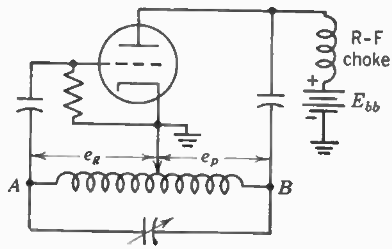


FIG. 4. Hartley oscillator. This is a "shunt-fed" circuit.

tential in this circuit, and the circuit is, therefore, subject to "hand capacitance" effects. If the condenser is adjusted with a long insulated shaft, and then if the hand is brought near the plates of the condenser, the frequency will change because the condenser capacitance has been effectively increased by the capacitance of the operator's body with respect to ground.

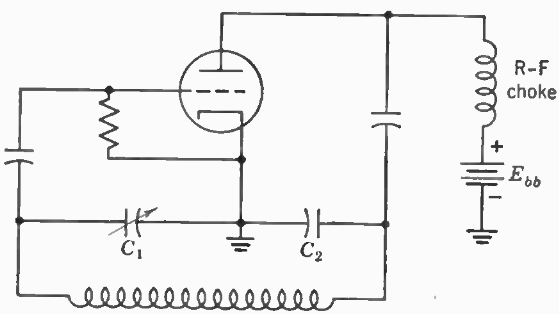


FIG. 5. Colpitts oscillator.

In the **Colpitts** oscillator circuit of Fig. 5 the division of voltage between grid and plate is secured by a split-stator capacitor. This circuit is often used as the local oscillator for superheterodynes. In this service C_1 is the tuning capacitor and C_2 serves as a "padder" to make the circuit tune to the correct frequency over the frequency band of the receiver. The amount of feedback is determined by the relative sizes of C_1 and C_2. Condenser C_2 is commonly made much greater than C_1.

Note that in both the Hartley and Colpitts circuits the plate voltage is supplied to the tube in shunt to the tuned circuit by means of a r-f choke coil. The virtue of this connection is that no d-c voltages appear on the tuning coils or condensers. These are called **shunt-fed** oscillators. The same scheme may be applied to the three oscillators discussed earlier.

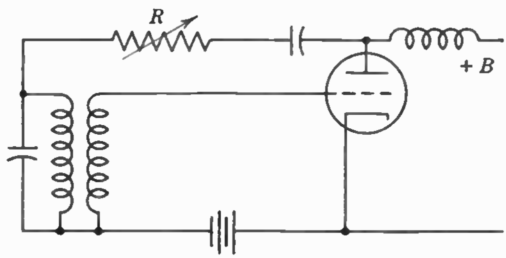


FIG. 6. Circuit often used in low- (audio-) frequency oscillators. Feedback is provided by resistance, which also loads up the plate circuit and straightens out the E_g-I_p characteristic.

There are several other combinations which will also serve for feedback oscillators. They operate on much the same principles as those discussed above.

A **resistance feedback** circuit sometimes employed in laboratory oscillators is shown in Fig. 6. This is basically a shunt-fed tuned-plate oscillator. However, a series resistor in the plate lead is provided for adjusting the amount of feedback. In this type of oscillator the most nearly sinusoidal waveform will be obtained when the feedback is reduced to a value just great enough to insure stable oscillations. This resistance, being in series with the plate, adds resistance to the plate circuit and reduces the effect which changes in plate resistance have upon the frequency and amplitude of the oscillations. The dynamic plate characteristics are also made more nearly linear so that harmonic generation is lessened. Shunt feed is used in this circuit to keep direct currents out of the iron-cored transformer since direct currents would cause saturation of the core and excessive distortion would result. It should be noted that direct current should be kept out of iron-cored coils or transformers in all audio oscillators if the distortion is to be kept to its minimum value.

As mentioned earlier in this section, an oscillator can be analyzed as a circuit in which the input impedance has a "negative"-resistance component just equal to the positive resistance of the tuned circuit. This method can be illustrated by reference to Fig. 7. This shows an oscillator in which an LC circuit is connected to the input terminals of a tube circuit made up of the two

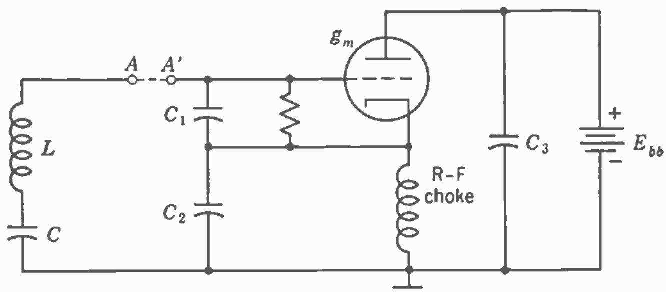


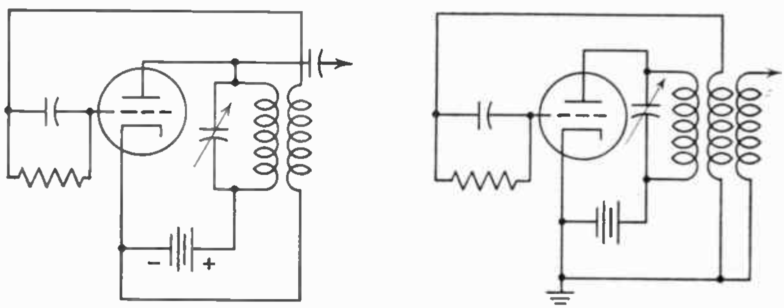
Fig. 7. A modified Colpitts oscillator. The impedance at A–A' has a negative resistance component equal to $-X_{C1}X_{C2}g_m$.

condensers C_1 and C_2 in series, C_1 being connected from grid to cathode and C_2 from cathode to plate through a blocking condenser C_3 as in an ordinary Colpitts circuit. What impedance does this tube circuit look like to the tuned circuit? By juggling some mathematics it can be proved that the impedance looking into the tube circuit at A–A′ is made up of the two reactances in series (X_L and X_C) plus a resistive component equal to the product of the two reactances X_{C_1} and X_{C_2} and the transconductance of the tube ($X_{C_1}X_{C_2}g_m$). But the mathematics will also show that this resistive component has a negative sign—it looks to the tuned circuit like a *negative* resistance. This circuit will oscillate if this negative-resistance term is equal to the positive resistance of the LC circuit.

Note in this circuit that the grid and plate are connected to opposite ends of the two condensers C_1 and C_2 in series. Grid and cathode are out of phase by 90°, and cathode and plate are also out of phase by 90°. Therefore, grid and plate are out of phase by 180°, the proper condition for oscillation. The voltage across C_1 caused by the oscillatory current flowing through C_1 is the grid driving voltage. The output voltage is taken across C_2 and is equal to the alternating current through it multiplied by the reactance of C_2.

In practice, C_1 and C_2 are made very large so that the grid-cathode and plate-cathode capacitances of the tube are shunted by very large condensers. Any change in the interelectrode capacitances of the tube, therefore, will have little effect upon the frequency at which the system oscillates. If the tuned circuit has a high Q and if L is as large and C as small as possible, the frequency generated will be remarkably stable.

16.3. Coupling circuits. The power may be coupled from an oscillator to the next stage by several methods. The two most common are capacitance and transformer coupling, both of which are illustrated in Fig. 8. Capacitance coupling is generally used when the following stage is a voltage amplifier which does not



(a) Capacitance Coupling (b) Transformer Coupling

FIG. 8. Oscillator-to-load coupling circuits.

require any appreciable amount of power, and for which the impedance matching problem is not severe. Transformer coupling is generally used when the following stage is an amplifier requiring considerable grid power, or in cases in which the oscillator works directly into the load. The amount of coupling is more or less fixed in the capacitance network, but it can be varied at will in the second case by making the coupling coil movable with respect to the coil of the tuned circuit of the oscillator.

16.4. Feedback-oscillator design and adjustment. To attain the best efficiency and usable power output from an oscillator, the various parts, such as coils and condensers, should be selected or designed to have low inherent resistance. Thus, the loss in the circuit components will be held to a minimum, and the maximum power will be available for delivery to the load.

Few adjustments can be made with a tuned-grid tuned-plate oscillator once it is designed except to tune it to frequency and vary the coupling to the load. The feedback is more or less fixed by the grid-plate capacitance of the tube. In both the tuned-grid and the tuned-plate circuits, which have feedback loops, the amount of feedback can be adjusted by varying the coupling, so that the best operation can be obtained for a particular load resistance. In the Hartley oscillator, the amount of feedback and the power output can be adjusted, to a certain extent, by moving the position of the cathode tap on the oscillator coil. It is generally not practical to make any feedback adjustment in a Colpitts oscillator other than that which occurs as the condenser is tuned to the desired operating frequency.

The direct plate current in most oscillator circuits, because of the grid-leak bias arrangement, is very high when the circuit is not oscillating, and decreases sharply when the circuit goes into oscillation. It rises gradually as more power is taken from the circuit. A d-c milliammeter in the plate lead provides a convenient means of indicating when the oscillator is working properly. As plate voltage is supplied to the oscillator and the tuning capacitor is tuned, the current should drop sharply when the circuit goes into oscillation. As the coupling to the load is increased so that more power is supplied to the load, the direct current will rise gradually. If a feedback adjustment is available, further adjustment of it may reduce the direct current to the tube without adversely affecting the power output. The ideal condition to be achieved is the required power output with minimum direct plate current and, of course, at the correct frequency. Under these conditions the oscillator will be operating at its best efficiency. Should the oscillator drop out of oscillation during the time adjustments are being made, the plate current will rise sharply and may become excessive. For this reason the feedback control is initially set to near its maximum value and the load coupling to its minimum value. Then, as oscillations are indicated by the meter, the proper adjustments can be made on both controls to obtain the desired operating conditions.

Although the basic frequency of the oscillators described above is determined by the resonant frequency of the *LC* circuits, several other factors may change the operating frequency to some extent. These include the amount of load coupled into the circuit, the inherent resistance of the circuit, the tube parameters, and the supply voltage. In general, if these oscillators are to maintain their frequency, the amount of power taken from the circuit should be as small as possible and the d-c voltage supplied to the plate circuit should be well regulated.

16.5. Electron-coupled oscillators. A circuit employing a tetrode or pentode has been devised in which the oscillating frequency is practically independent of the plate voltage and load

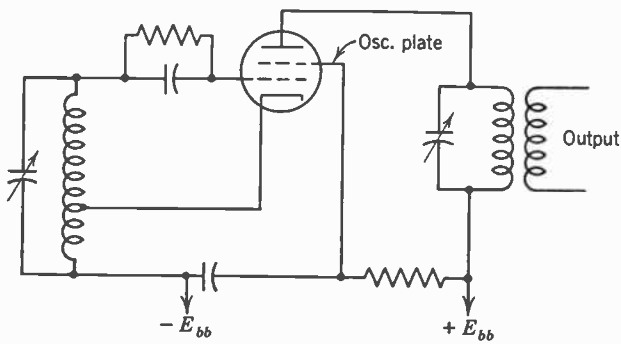


Fig. 9. Electron-coupled oscillator.

variations. This simple circuit, shown in Fig. 9, is called an **electron-coupled** oscillator because the output is coupled to the oscillator only through the electron stream. The actual oscillator shown uses the two grids and the cathode in a Hartley circuit. The screen grid serves as the oscillator plate. The plate current will be modulated by the oscillations in the grid circuit, and an alternating voltage will be produced across the tuned-plate circuit when it is tuned to the frequency of the oscillator. At the same time, however, the oscillator is effectively shielded from what occurs in the output or load circuit because the screen grid is at a-c ground potential. As long as the ratio between screen and plate voltages remains constant (the screen voltage may be taken from a tap of a voltage divider across the plate supply terminals) the frequency will be almost entirely independent of the actual voltage on the plate. A resistor or r-f choke can be substituted for the tuned load in the plate circuit at the expense of some efficiency.

Electron-coupled oscillators are widely used in the frequency-converter stages of superheterodynes, and in the tunable "master" oscillators of transmitters which operate over a wide frequency range.

16.6. Grid bias for oscillators. Oscillators are generally operated with some sort of self-bias, rather than with fixed (battery) bias. Self-bias is preferred since oscillators usually operate under Class C conditions, and under these conditions fixed bias prevents the initial build-up of the oscillations. Audio-frequency oscillators, however, must employ Class A bias to keep distortion low.

Bias may be secured from a grid-leak condenser combination in the grid circuit. The amount of bias is determined by the amount of grid current and the size of the grid-leak resistor. Two circuit arrangements are used: a series and a shunt arrangement. In the series arrangement the grid leak and condenser are connected in parallel and the combination placed in series with the grid and the rest of the grid circuit (Figs. 1, 2, and 3). In the shunt arrangement the grid-leak resistor is connected between grid and cathode. A series condenser is connected to the remainder of the grid circuit. This circuit is illustrated in Figs. 4 and 5.

Self-bias in oscillators has the advantage that it tends to increase the stability of the oscillations. For example, if the output voltage tends to increase, the grid excitation also increases and this in turn increases the bias voltage. The increased bias tends to decrease the output voltage and hold it to its original value. There is one disadvantage: if the tube stops oscillating for any reason, there is no bias on the grid, and the plate current may rise high enough to damage the tube. For this reason, power oscillators usually obtain some or all of their bias from a cathode resistor.

The power dissipated in the grid-leak or cathode resistor must be supplied from the plate supply system, and this lowers the over-all efficiency of the oscillator.

The manner in which the direct voltage builds up in a grid-leak system is shown in Fig. 10. Note that the average plate current decreases as the oscillations build up since an increase in the strength of the oscillations also increases the bias as a result of the flow of a larger grid current. The purpose of the condenser is to smooth the ripple on the bias voltage so that a non-fluctuating bias voltage is obtained.

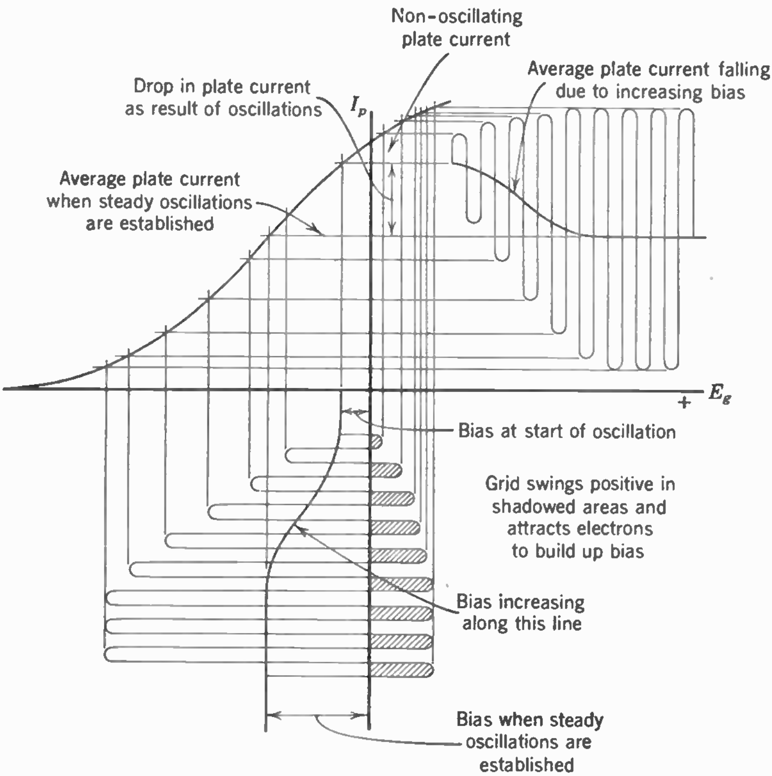


FIG. 10. Gradual build-up of oscillations with resultant increase in bias and decrease in steady plate current.

16.7. Oscillator efficiency. By adjusting the grid bias of oscillators so that plate current flows only during a small part of the cycle the power ($E \times I$) dissipated on the plate may be kept very low. Under these conditions the plate voltage may be increased beyond the value which would be safe if plate current flowed all the time. The circuit efficiency is high, but the alternating plate current will be non-sinusoidal; that is, it will contain many harmonics. Since, however, the tank circuit and the load coupled to the tank are tuned to the fundamental frequency the harmonics will not create much voltage in the load because the load will have a low impedance to frequencies higher than the fundamental. Such high-efficiency circuits may, therefore, be employed at radio frequencies for which tuned circuits with high values of Q may be utilized. Such circuits, then, perform like self-excited Class C amplifiers (see Sect. 17.5). It is difficult to secure high-Q a-f tuned circuits, and for this reason Class C bias cannot be used in a-f oscillators if the output waveform must be comparatively free of harmonics.

16.8. Frequency stability. Oscillators which must furnish a constant frequency must be carefully designed and built. Several factors must be taken into consideration if a high degree of frequency stability is to be achieved. Circuit components (tubes, coils, condensers, resistors, etc.) must be selected that do not change their characteristics appreciably with temperature, and must be mounted so that they are not subject to mechanical vibrations. The d-c plate supply voltage of the oscillator should be well regulated. Another factor which must be considered is the input capacitance of the tube. This capacitance is in shunt with the capacitance of the tuned circuit of many oscillators and therefore affects the frequency. The equation for the input capacitance C_g was given in Sect. 13.11 as

$$C_g = C_{gk} + C_{gp}(A_v + 1)$$

where A_v, for the oscillator circuits shown, is

$$A_v = \frac{\mu R_L}{R_L + r_p}$$

In this equation, R_L is the effective resistance of the plate load circuit. Anything which changes the dynamic plate resistance, r_p, of the tube, or the load resistance, R_L, produces a change in the input capacitance and may change the frequency of oscillation. Changes in filament temperature, in C bias, or in plate voltage will affect r_p and cause a shift in frequency.

One way to lessen this difficulty is to shunt a fairly large capacitance directly across the grid and cathode to increase the total effective input capacitance so that small changes in C_g will have little effect on the frequency. In many circuits this effect may be accomplished by merely using a small coil and large condenser (low L/C ratio) in the tuned circuit. High-capacitance circuits, however, have large circulating currents in them and may be quite inefficient as a result of high power losses.

16.9. Crystal oscillators. Often it is desirable to achieve a higher degree of stability than can be obtained in any of the circuits described above. For example, the oscillators for broadcast transmitters must hold their frequency within very close tolerances. An oscillator, employing a quartz crystal as the frequency-determining element, provides a simple means for achieving a high degree of frequency stability.

Quartz, in the crystalline form, exhibits an interesting property called **piezoelectricity.** If two surfaces of a slab cut out of a crystal are compressed mechanically, an electrical voltage appears across these faces. Conversely, when the voltage across the faces is changed, the crystal tends to change in size. The crystal is, in effect, a mechanical resonator or mechanical tuned circuit, and when inserted into a vacuum-tube oscillator it will serve the same purpose as a tuned circuit made up of a coil and condenser.

A quartz crystal cut for oscillator service has a value of Q as high as 10,000 if intended to operate at 100 kc. The crystal acts like a very large inductance shunted with a very small capacitance. The effective resistance is extremely low. In Fig. 11 the crystal replaces a tuned circuit in the grid of a tuned-grid tuned-plate oscillator. When the plate circuit is tuned slightly higher in frequency than the resonant frequency of the crystal, the entire circuit oscillates at the crystal frequency. For a considerable variation in the tuning capacitance in the plate circuit, as long as the resonant frequency of the plate tank is higher than the crystal frequency, the oscillator will produce an output at the crystal frequency. Minor variations in the capacitance of the circuit, such as occasioned by changing tubes, variations in plate voltage, etc., have no controlling effect upon the frequency of the oscillator. In another typical oscillator circuit, the crystal is inserted between the grid and plate as shown in Fig. 12.

In applications where extreme frequency stability is required the crystal is often mounted in an oven the temperature of which is thermostatically controlled.

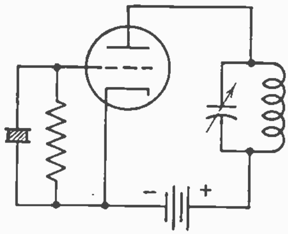


Fig. 11. Quartz crystal acting as a tuned circuit of very high Q and high stability.

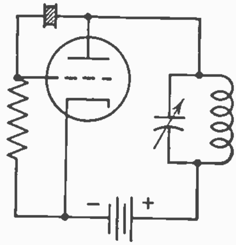


Fig. 12. Circuit in which crystal is connected from grid to plate.

16.10. Crystal cuts. There are many ways in which crystal plates may be sawed from the mother crystal. These are known as "cuts," and some of them are good for one purpose and some for other purposes. The resonant frequency of a crystal depends to some extent upon the temperature, and this dependence on temperature is different for crystals cut in different directions in respect to the optical and electrical axes of the mother crystal. Much research has gone into determining the kind of cuts which will produce the smallest temperature coefficient, that is, cuts that result in crystals which will not change their frequency even though the temperature may change over a considerable range. The type of cut also determines, to some extent, the "activity" of the crystal.

A quartz crystal may vibrate in several modes. That is, the basic frequency may be determined by the thickness, length, or breadth of the plate. In some cuts the temperature coefficient is controlled by one dimension and the frequency by another; thus it is possible to control these two important factors more or less independently.

In crystal cuts in which the thickness determines the frequency, thinner crystals oscillate at higher frequencies, the relation being (1960 ÷ millimeters of thickness) kilocycles for the widely used *AT* cut. For this reason the higher frequencies require very thin crystals; and, since in service these plates vibrate with an amplitude dependent upon the amount of power they are controlling, oscillators working at very high frequencies are limited to small power outputs because of the danger of crystal breakage. The output of the crystal oscillator can be used to drive additional amplifiers to secure a large final power output. Often, however, a lower-frequency crystal capable of more power output is used. The frequency is doubled or tripled or multiplied still more in subsequent amplifier circuits.

It is generally considered wise to keep the r-f current flowing through the crystal lower than 100 ma, not to attempt to control more than 5 or 10 watts directly with the crystal, and not to control directly a circuit whose output is higher in frequency than about 12 Mc.* With care in design, a crystal-controlled oscillator should not vary more than a few parts per million over an appreciable time. That is, an oscillator having an output frequency of 1000 kc should not vary more than 10 or 15 cycles from this frequency. With great care stabilities as high as 1 part in 100 million may be obtained.

16.11. Negative-resistance oscillators. If, in any circuit, a *decrease* of current is produced by an *increase* of voltage, the dynamic resistance of that circuit is negative and this negative

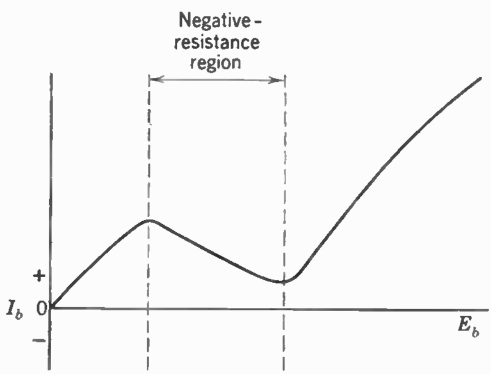


FIG. 13. Negative-resistance region of tetrode.

resistance can be utilized in an oscillator circuit. Such a negative-resistance characteristic is found in a tetrode tube when the plate voltage is lower than the screen voltage. The negative-resistance region is shown in Fig. 13.

A circuit for a negative-resistance oscillator is shown in Fig. 14. The tuned circuit is placed in the plate lead, and the plate

* Special "harmonic" crystals operate satisfactorily up to about 30 Mc.

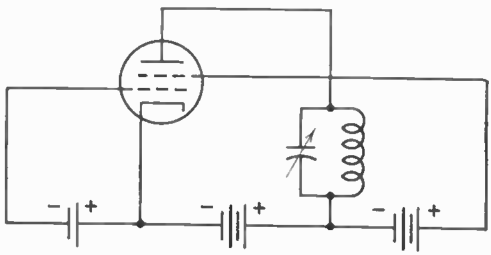


FIG. 14. Negative resistance produced by operating screen at higher potential than plate causing secondary emission.

and screen direct voltages are adjusted so that operation is near the middle of the negative-resistance region of the tetrode characteristic. If this negative resistance is large enough to cancel the positive resistance of the tuned circuit, oscillations will result. Such an oscillator may produce an excellent waveform. It has the disadvantage that some instability results from changes in the secondary emission from the plate, which produces the negative-resistance characteristics.

Many circuits have been devised, most of which use two or more tubes in various combinations, which produce the effect of negative resistance and can be used in oscillator circuits. Many of these circuits could be analyzed either as feedback or as negative-resistance oscillators.

16.12. Relaxation oscillators. A gas tube can be used in a relaxation-oscillator circuit as shown in Fig. 15. When the plate

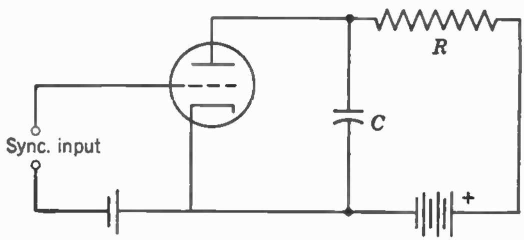


FIG. 15. *RC* oscillator using gaseous triode and with synchronizing voltages placed on grid to control time of condenser discharge.

voltage is applied to the circuit the condenser C charges through the resistor R and its voltage rises as shown in Fig. 16. When the voltage rises to the breakdown (firing) potential of the tube, the tube (heretofore acting as an open circuit) suddenly ionizes and begins to conduct current. Its internal resistance is quite low, and so it discharges the condenser almost immediately to the voltage at which the tube again deionizes (the extinction voltage). At this point the tube resistance again increases and the tube no longer passes current. The condenser again begins to charge and the cycle repeats itself. The output has a sawtooth waveform. The frequency, or more correctly the repetition rate, is controlled by R and C and by the applied voltage. The exact

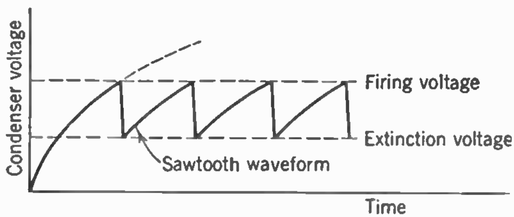


FIG. 16. Voltage waveform of relaxation oscillator of Fig. 15.

timing of the pulses can be controlled by an external synchronizing voltage applied to the grid.

Relaxation oscillators are widely used in cathode-ray oscilloscopes and in television circuits, and are discussed more completely in later chapters.

16.13. Phase-shift oscillators. A phase-shift oscillator employs RC networks to obtain the required 180° phase shift between plate and grid circuits. The circuit of such an oscillator is shown in Fig. 17. Normally each phase-shifting section, consisting of one C and one R, provides a phase shift of 60°, so that three such sections are required. Operation is Class C, but the waveform may be made nearly sinusoidal if the bias is adjusted so that oscillations are barely maintained. The frequency stability of a phase-shift oscillator is quite good. The frequency of oscillation may be changed by adjusting the resistors or capacitors. However, the required 180° phase shift must be maintained.

Problem 1. An amateur employs a 6C4 tube as an oscillator. When this is working properly the d-c grid current will be 5 ma. It is desired that the grid bias be -10 volts. If the grid current flows through a resistor as in Fig. 1, what must be the value of this resistor?

Problem 2. If, in Prob. 1, half the bias is to be provided by the grid resistor and half by a cathode resistor, what values will the two resistors have? The cathode current is 10 ma and the grid current remains at 5 ma.

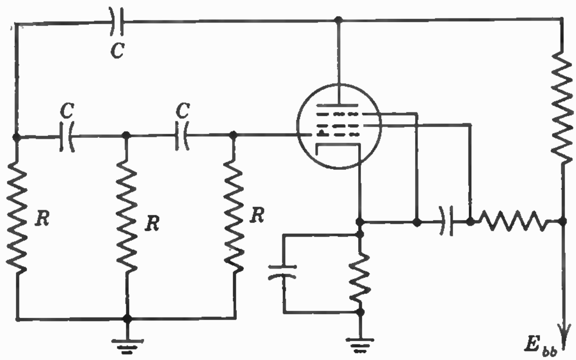


Fig. 17. Phase-shift oscillator.

Problem 3. An amateur wishes to buy a quartz crystal to control the frequency of his transmitter. He wishes to operate as near as possible to the lower-frequency edge of the 7-Mc band, that is, as close to 7 Mc as possible without danger of getting out of the band. If the crystal frequency is guaranteed to within 0.01 per cent of the marked value, and if 1 kc additional is allowed for temperature variations, what is the lowest frequency to which he should ask the manufacturer to grind his crystal?

Problem 4. During operation a certain oscillator may vary 50 cycles up and down from its designed frequency of 3505 kc. If the frequency output of the oscillator is tripled twice in the following amplifiers and then applied to an antenna, between what frequencies will the antenna signal vary?

17 · Amplitude-Modulation Transmitters

A radio transmitter is a combination of an oscillator, voltage and power amplifiers, modulating equipment, and power supply apparatus all designed to put r-f power into an antenna from which it is radiated in the form of electromagnetic energy. The oscillator stage is of low power, perhaps 5 watts or less, is usually crystal controlled, and in high-frequency stations operates at a submultiple of the final radiated frequency. It is followed by amplifiers which build up the voltage and power of the crystal stage to a value high enough to drive the final power output stage. These intermediate-frequency amplifiers also act as "buffers" to isolate the crystal stage from the power stage so that the crystal can operate without any reaction back upon it by the power stages. These amplifiers may also act as frequency multipliers, increasing the frequency to some integral multiple of the oscillator frequency. The frequency multiplication may take place in several stages. Connected to one of the stages is a modulating system which gets its excitation from a microphone or a telegraph key.

17.1. Types of r-f power amplifiers. When a considerable quantity of r-f power is generated and transmitted the efficiency of the equipment becomes important since the electrical energy purchased from a public utility, or generated locally by the transmitting station, must be paid for according to the amount consumed and not according to the amount usefully radiated. A 50-kw broadcast station that consumes 250 kw from the power line is less efficient and will have a higher annual bill for electrical energy than one which delivers the same power to the antenna but draws only 105 kw from the power lines.

Since the power output stage of a transmitter feeds the antenna through a circuit tuned to the operating frequency, the harmonics generated within the amplifiers do not get on the air

—at least, they are greatly attenuated. For this reason it is not necessary to use Class A amplifiers in the r-f sections of a radio transmitter. Much higher efficiency can be secured from Class C amplifiers, which, it is true, produce many harmonics of the fundamental radio frequency, but these harmonics can be prevented from reaching the antenna.

If the modulation takes place in a low-power stage, then all subsequent stages must be linear in order that the modulation (which is the important message-bearing portion of the transmitted wave) will not be distorted. Linear stages are operated Class B. If modulation takes place in the final stage, the modulated stage can be operated Class C. If the transmitter is designed to transmit code rather than voice or music, all stages may be operated Class C.

The principal difference between Class B and C amplifiers is in the amount of *C* bias used and the portion of the a-c cycle during which plate current is permitted to flow. In a Class B amplifier the bias is adjusted so that the plate current is almost zero when the grid excitation is removed. If sine-wave voltage is applied to the grid, plate current consists of a series of half sine waves similar to the output of a half-wave rectifier. The load impedance is adjusted so that the relation between plate current and grid voltage is linear. The grid swings positive on excitation peaks and grid current flows. During peaks of plate current the plate voltage reaches a minimum value because of the high voltage drop across the load. The tube power, then, is the product of a high current and a low voltage, but the load power is the product of a high current and a high voltage. This means simply that, of the total power taken from the d-c plate supply, most is supplied to the load, and relatively little is used up in heating the plate of the tube. Power does not flow continuously from the plate supply source. Rather, it flows in a series of pulses. The efficiency of a Class B amplifier is rather high, approaching 78.5 per cent as a maximum.

Two tubes may be operated in push-pull in a Class B r-f amplifier if desired, and about twice the power output of a single tube will be secured. Tubes in Class B r-f amplifiers can be operated to produce greater power output than in Class A service since the plate efficiency is much higher.

A Class C amplifier has its bias adjusted so that plate current is definitely zero for a portion of the excitation cycle. The grid excitation is high and may produce saturation current from the cathode of the tube on the positive portions of the excitation cycle. Other conditions are quite similar to those of a Class B amplifer. The plate efficiency of Class C amplifiers may be quite high, approaching 90 per cent in amplifiers using large tubes.

17.2. Neutralization. Buffer or intermediate power amplifiers may be triodes or, better, screen-grid or pentode tubes. Triodes must be "neutralized" to keep them from oscillating and to keep variations in their output circuit from reacting upon the crystal-oscillator stage. Neutralization consists of placing the triode in a bridge circuit in which a part of the output voltage, equal to that fed to the input by the grid-plate capacitance, is fed to the input. This additional voltage is in such a direction that it opposes (neutralizes) the effect of grid-plate capacitance voltage.

The several types of neutralizing circuits may be roughly broken down into three general groups: (1) plate neutralization, (2) plate-to-grid neutralization, and (3) push-pull cross neutralization. Diagrams illustrating these methods are shown in Fig. 1.

An equivalent circuit of the cross-neutralized push-pull amplifier of Fig. 1(*d*) is shown in Fig. 2. Here the various elements taking part in the neutralizing process are shown as parts of an a-c bridge. The inductances, L_g, are the inherent inductances of the grid leads. The bridge circuit is symmetrical; that is, opposite arms contain the same kind and size of reactance. If the tubes are exactly identical, and the value of L_g for each tube is negligibly small, the condition for neutralization is that the neutralizing condensers shall be adjusted to have the same value as the grid-plate capacitance C_{gp} of the tubes. Some slight change from this value is usually required to compensate for the actual value of L_g and for wiring and other distributed capacitances not shown on the diagram.

Equivalent bridge circuits, similar to that in Fig. 2, may be drawn for the other neutralizing schemes shown in Fig. 1.

Triodes are widely used in the power stages of r-f amplifiers in preference to tetrodes and pentodes in spite of the necessity for neutralization. Some of the reasons follow: (1) plate-modulated stages must use triodes since varying the plate voltage of a multi-

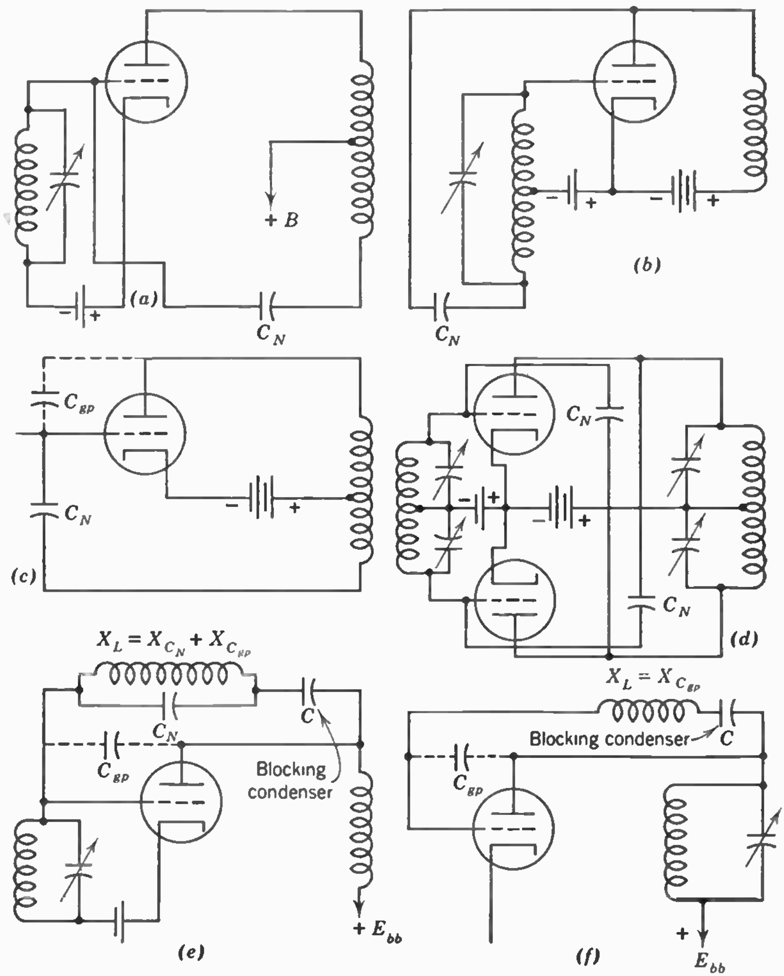


FIG. 1. Neutralizing circuits. (*a*) Use of split plate transformer. (*b*) Split transformer in grid circuit. (*c*) Equivalent bridge circuit of (*a*). (*d*) Neutralized push-pull amplifier. (*e*) Network connected from plate to grid. (*f*) Resonant circuit for neutralization.

grid tube has little effect on plate current; (2) high-power screen-grid tubes have not become available; and (3) at the high frequencies (1 to 2 meters wavelength) the effect of the input and output capacitances of screen-grid tubes is so great that simpler tubes of the triode type are preferred.

At the present time considerable effort is being expended in developing medium-power screen-grid tubes for high-frequency r-f

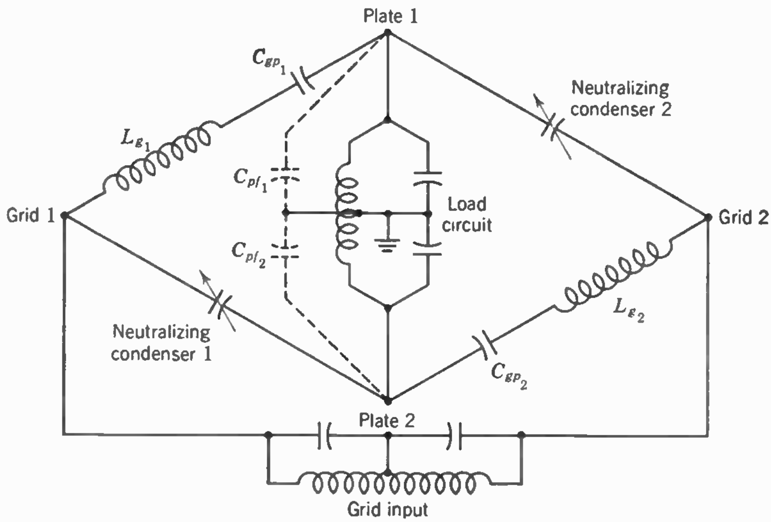


FIG. 2. Bridge circuit made up of neutralizing and tube grid-plate capacitances in a cross-neutralized push-pull amplifier.

power-amplifier use. These tubes must, however, be neutralized if oscillations are to be prevented and the neutralizing circuits are rather complicated. The big advantage of these tubes over triodes is that they have a higher power sensitivity; that is, more output power is obtained for a given grid power.

17.3. Power-amplifier design. Since the grids of Class B and C amplifiers are driven positive, the grid circuits pass current and therefore require power. This power must be provided by the stage that precedes and drives the Class B or C stage.

Actual design of a Class C stage is a matter both of calculation and of cut and try. Since the plate current does not flow continuously, the internal resistance of the tube varies over the cycle and a numerical value which can be employed in calculating the load resistance, power output, etc., cannot be supplied by the manufacturer (or measured by the user).

Relations between the various voltages and currents in a Class C amplifier are seen in Fig. 3. The figure is drawn for a unity power factor load; that is, the plate circuit is tuned to resonance with the frequency of the grid signal. Note that the minimum value of plate voltage and the maximum value of grid voltage

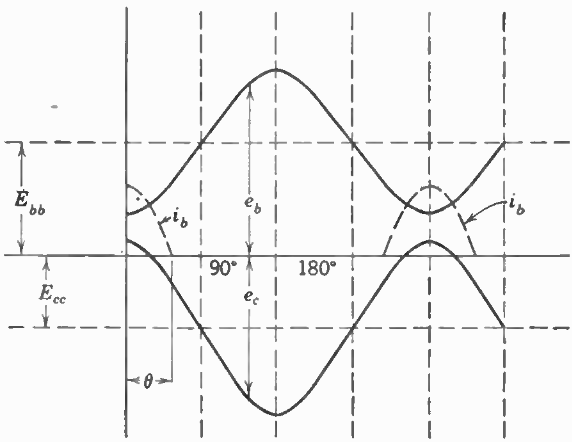


Fig. 3. Relations existing in Class C amplifier. The angle during which plate current flows is 2θ. The grid voltage magnitude is exaggerated here.

occur at the same instant. At this same instant, the maximum plate current flows. The plate current is seen to flow for substantially less than 180° during a cycle. Grid current is not shown on the diagram, but it flows only while the grid is positive.

Even though the complete design and calculation of a Class C amplifier are much too complex to be considered here, several important factors can be discussed which will give the basic requirements for a good design.

1. Decide upon the maximum value of the grid excitation voltage and the minimum value of the plate voltage. Ordinarily it is desired that the excitation power be kept small. This means that excessive grid current should not flow, and, therefore, that the maximum excitation voltage should not be high enough to drive the grid highly positive. Furthermore, the maximum plate current must be kept from being too high. These requirements

can generally be met by taking the maximum or peak value of the instantaneous grid voltage (the algebraic sum of the a-c exciting voltage and the d-c grid-bias voltage) as 80 per cent of the minimum instantaneous plate voltage.

2. Decide upon the proportion of the excitation cycle in which plate current flows. If current flows for only a small portion of the cycle, the efficiency (power output divided by power input) will be high but the power output (watts) will be low. If the current flows for a longer period, efficiency will be low but the power output will be higher. The actual time during which plate current flows is expressed as an angle in Fig. 3; it is equal to 2θ. Ordinarily the angle of flow will be of the order of 120° in a voice-modulated stage. This requires bias voltages between 3 and 6 times the cut-off voltage for most modern transmitting triodes which have rather high amplification factors. Suitable values can be found in various tube manuals.

3. Determine the cut-off bias. This will be approximately equal to

$$E_{co} = E_{bb} \div \mu$$

4. Determine the required peak value of the excitation voltage. This will ordinarily be somewhere between 1.5 and 4 times the cut-off bias. The larger the value of grid excitation voltage, the greater the power output, but more driving power is required. The total instantaneous grid voltage should never exceed the instantaneous plate voltage.

5. Calculate L, C, and the resistance of the plate tank circuit. This circuit must tune to the operating frequency and present the proper load to the tube. The tank circuit is discussed in more detail in the following sections.

17.4. Power relations in a Class C amplifier. The power supplied by the d-c plate supply is

$$P_{\text{input}} = E_{bb}I_b$$

The output to the tank circuit is

$$P_{\text{tank}} = \frac{E_oI_p}{2}$$

where E_o is the peak a-c voltage across the tank.
I_p is the peak a-c plate current.

The peak voltage across the tank is

$$E_o = E_{bb} - E_{p\min}$$

Thus, if the minimum plate voltage, $E_{p\min}$, is low, the peak tank voltage is almost equal to the d-c plate supply voltage and the total instantaneous voltage across the tank may rise to nearly twice the d-c voltage. The tube and the tank coil and condenser must be able to withstand this high voltage.

The power lost in the tube itself is the difference between the total power supplied by the d-c supply and that supplied to the tank. The plate efficiency is the ratio of the a-c power supplied to the tank to the d-c power taken from the power supply.

Since the load of a Class C amplifier is tuned to resonance at the frequency which is to be transmitted, this load to the tube represents a pure resistance with a value nearly equal to L/CR, where L, C, and R are the constants of the tank circuit. This value of load should be approximately equal to the effective resistance of the tube for greatest power transfer (in this case the efficiency will be about 50 per cent) and higher to attain greater efficiency. The value of R is the effective a-c resistance of the tank coil and includes both the inherent resistance of the coil itself and the resistance coupled from the load.

17.5. Flywheel effect of tank. The student may wonder why tank current can flow continuously if the tube passes current only at periodic intervals, and why the tank current may be nearly sinusoidal even though the plate current is in the form of pulses.

It must be remembered that the only power used in the tank is employed in the inherent resistance of that circuit and in the resistance reflected into it by the process of supplying power to a load such as an antenna. Aside from these losses of energy, energy supplied by the tube to the tank is transferred back and forth from condenser to coil, just as, in a swinging pendulum, energy is transformed from that of position (potential) to that of motion (kinetic). To maintain current flowing in the tuned tank circuit, all that is required is that the power losses be furnished from an external source (the tube), not necessarily continuously but at intervals. The pendulum of a clock keeps swinging, although it gets an impulse from the source of energy, the spring, only at the proper points in its cycle, by the escapement mechanism, and similarly the tube acts as an escapement mechanism in the Class C amplifier, supplying energy to the circuit at the proper point in the cycle and of the proper amount to overcome the losses inherent in the circuit plus energy transferred to the load.

If the effective Q of the tank is 10 or greater, the current in it will have a nearly sinusoidal waveform. However, if the Q is made lower than 10, as by drawing more power from the circuit, the tank current becomes non-sinusoidal and considerable harmonic current may be transferred to the antenna.

17.6. Frequency multipliers. Crystal oscillators are usually rather loosely coupled to the following stage and are not required to furnish any considerable amount of power, and reaction from the succeeding amplifiers has little effect on the crystal-oscillator frequency. The upper frequency at which crystal oscillators may operate is limited by the fact that higher-frequency crystals are thinner and the crystals more fragile. The highest frequency at which crystals may be operated satisfactorily is about 30 Mc, and special crystals are required above 10 Mc. It is generally better to use a fairly low-frequency crystal and frequency multipliers than to use a fragile high-frequency crystal.

A **frequency multiplier** is an amplifier so biased that it produces high-amplitude harmonic currents and voltages in its output; it is a Class C amplifier. A crystal may, for example, control a 1-Mc oscillator. The oscillator may be used to drive an amplifier the plate circuit of which is tuned to the second harmonic of the crystal frequency (this amplifier would be called a **frequency doubler**), and this amplifier may be used to drive a **frequency tripler.** Then, the output frequency of the last amplifier is 2×3 or 6 times the crystal frequency, or 6 Mc. Several stages of frequency multiplication may be used to yield a final frequency of hundreds of megacycles.

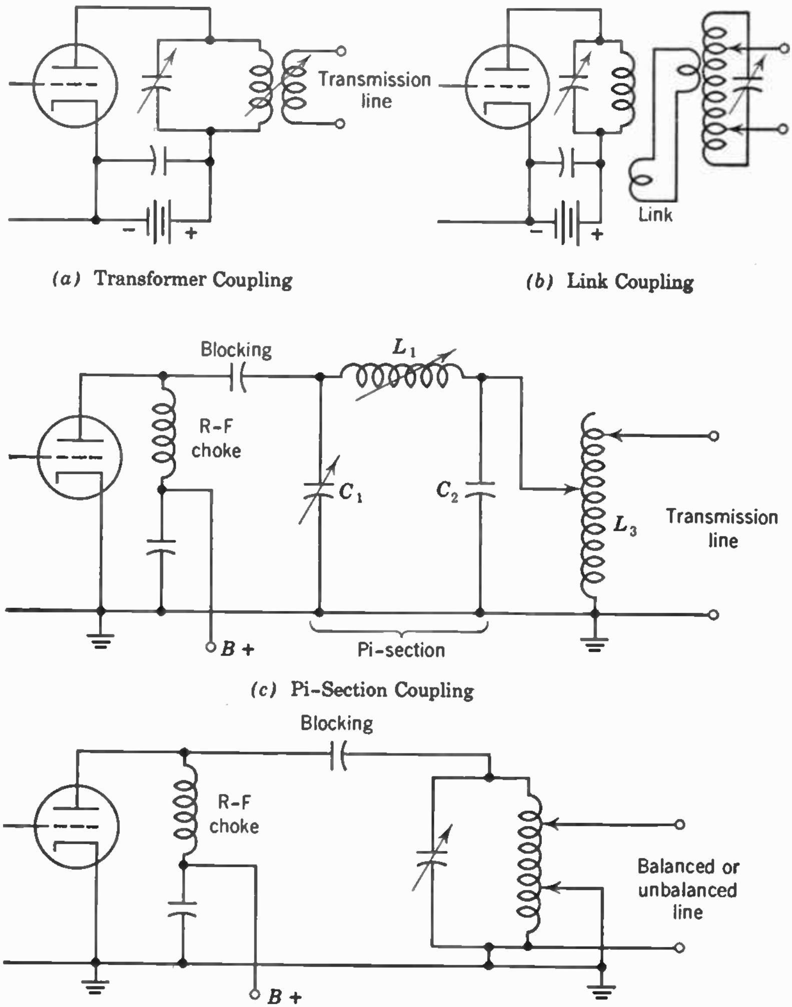
The plate circuit of a frequency multiplier is tuned to the desired harmonic. This is possible since the plate current of a Class C amplifier contains many harmonic components. A frequency doubler may furnish about half as much power output as a straight amplifier. Frequency triplers, quadruplers, etc., become progressively less efficient. A frequency multiplication of 4 or 5 times in a single stage is about all that can be achieved by the methods described here.

Any tube circuit which has a non-linear relation between grid voltage and plate current will have harmonics in its output current. Overbias of the grid is one method of securing a high harmonic content. Special circuits have been devised for the purpose of frequency multiplication.

17.7. Coupling the transmitter to the antenna. A transmitting antenna operates best if mounted well above surrounding objects and well separated from buildings, vegetation, etc., all of which can cause unwanted reflections and absorption of power. In addition, high-frequency waves are received consistently only at "line-of-sight" points from the transmitting antenna, and so the antenna is mounted as high as practicable to increase the service area. For these reasons the antenna is usually a considerable distance from the transmitter and means must be provided for properly connecting the antenna to the transmitter.

Several factors must be considered in determining the nature of the transmitter-to-antenna coupling network. The entire network as "seen" by the tube must be resonant to the operating frequency and must present the proper resistive load to the final power-amplifier tube. Not only must the network properly match the impedance of the final tube of the amplifier; it must also match the impedance of the antenna for maximum power transfer. For the best efficiency, coils, condensers, and transmission lines must be provided which have the least possible resistance. Often it is desirable to select a coupling network which inherently rejects harmonic frequencies. The transmission line may be operated either balanced or unbalanced. A **balanced line** consists of two wires both insulated from ground and operated so that both wires have equal capacitance to ground. An **unbalanced line** is either a single-wire line with a ground return, or a coaxial line. Either type requires a coupling network that permits one side to be grounded.

Typical networks for coupling the transmitter to the transmission line are shown in Fig. 4. In Fig. 4(a) the proper impedance match and power transfer are achieved by varying the coupling. This is a simple network but does not give very high attenuation to harmonics. Transformer coupling may be used to feed either a balanced or an unbalanced transmission line. For unbalanced operation one side of the secondary is grounded. The link coupling shown in Fig. 4(b) allows the tank circuit and the



(a) Transformer Coupling

(b) Link Coupling

(c) Pi-Section Coupling

(d) Coupling for Balanced or Unbalanced Line

FIG. 4. Typical coupling networks for transmitters.

circuit connected to the transmission line and antenna to be separated physically. This is sometimes advantageous. The coils of the link should be placed near the grounded ends of the coils to which they are magnetically coupled to avoid excessive capacitive coupling which will allow harmonics to be transferred to the transmission line. The pi-section network of Fig. 4(*c*) provides adjustments for both tube loading (L_1 and C_1) and impedance match to transmission line (L_3). This is also a type of low-pass filter, and the harmonics are greatly attenuated in this network. Still another type of coupling is shown in Fig. 4(*d*). Figures 4(*c*) and 4(*d*) employ shunt feed to the final amplifier tube. That is, d-c plate voltage is fed through a r-f choke selected to have a high impedance at the operating frequency. The blocking condenser keeps direct voltages out of the tank circuit. This is advantageous from a safety standpoint.

The transmission line must be selected to have proper characteristics. The discussion of transmission lines, along with formulas and charts, is in Chapter 18.

17.8. Keying a code transmitter. If the radio station is to transmit code only, its modulation system can be very simple. A key is merely inserted in some portion of the power supply system so that when the key is up the oscillator does not generate r-f power, or the amplifier does not operate, thus preventing any of the oscillator output from reaching the antenna. Keying is usually accomplished in a buffer stage between the crystal oscillator and the final power-amplifier stage. The grid of this keying stage may be overbiased so that no plate current is drawn until the key is pressed when the bias on the tube is reduced. If the power taken by this tube is not too great, keying may be accomplished in the cathode lead as in Fig. 5(*a*). The tube side of the key is "hot," and it is dangerous to operate a hand key in this position. It is better to use a mechanical key (relay) here, which, in turn, is operated by the hand key. In Fig. 5(*b*) a keying tube is shown. When the key is down, the keying tube draws no plate current since it is biased to cut-off. When the key is open, the grid of the keying tube is somewhat positive, and a high current flows, producing a high voltage drop across R and reducing the voltage to the power tube.

17.9. Audio portion of broadcast transmitter. Any station which transmits speech or music must necessarily be more

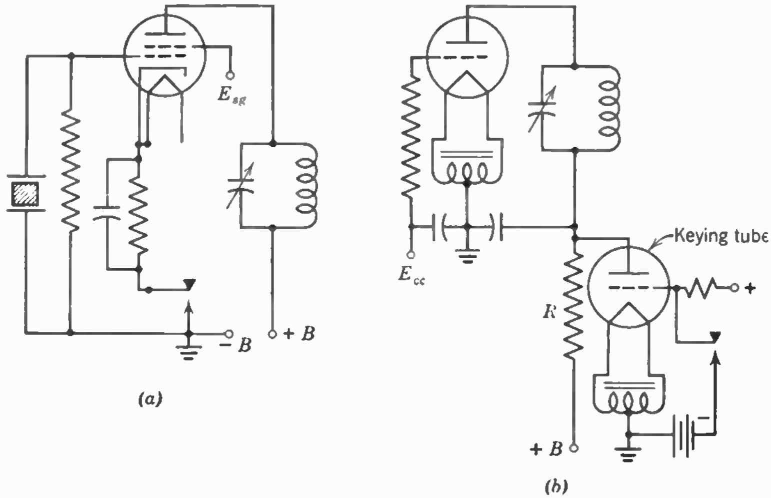


Fig. 5. Methods of keying an oscillator or an amplifier. At (a) the plate current is keyed directly; at (b) a keying tube is employed. Voltages across the key in (b) are much lower than in (a).

complex than one which transmits code. Between the microphone and the final power stage which feeds the antenna must be high-quality audio amplifiers having sufficient voltage amplification and power output to raise the output of the microphone to the level required to modulate the r-f power of the transmitter. All these amplifiers and the modulating equipment must be electrically quiet. The modulator must not introduce noise, hum, or distortion into the signals that come from the microphone.

17.10. Audio-frequency range. The wider the a-f range to be transmitted, the more difficult is the problem of keeping noise and distortion out of the system. A voice-frequency station serving a police or fire department need not have as wide a frequency range as a high-fidelity broadcast station. Speech can be transmitted with good intelligibility if a frequency range of 250 to 2500 cycles is transmitted. If higher frequencies are transmitted, the intelligibility will be better. If lower frequencies are transmitted, the sounds will be louder and the sender will not have to talk so loud. Such a frequency range would be entirely inadequate if an accurate reproduction of a musical program were to be transmitted.

It is recognized that accurate reproduction of speech and music requires the transmission, without distortion, of a frequency band from about 30 to 15,000 cycles. The broadcast audience has never had reproduction of this quality. Transmitters are placed in the standard broadcast band so close together in frequency, to accommodate all of them, that frequencies higher than 5000 cycles conflict with stations on adjacent channels. Manufacturers of receivers have spent the energies and capabilities of their engineering staffs in reducing the cost of receivers for the sake of mass production of low-priced units instead of in improving the tone fidelity of reproduction. So much distortion and noise are inherent in the average broadcast receiver that, if the frequency-response band were widened out to reproduce what is available from the average broadcast transmitter, the receiver would be almost unsalable. The average receiver has very little response below 100 cycles or above 5000 cycles.

Nevertheless broadcast stations transmit wide bands of audio tones, free from distortion and noise. Their output is vastly beyond the capabilities of the average receiver to reproduce with fidelity.

17.11. Microphones. The input to the whole broadcast station is the microphone. If it has limited tone response, or if it is noisy, the entire transmission system will be faulty. All microphones have one chief purpose—to translate a mechanical motion into an electrical voltage. Sound waves are variations of air pressure; they produce mechanical changes in the microphone which in turn produce electrical changes that are representative of the sounds affecting the microphone.

The earliest microphones, used at present in the common telephone, contained carbon grains between two metallic plates across which was placed a voltage. Sound waves impinging upon the diaphragm caused the carbon grains to pack closer together or to be released from pressure, and these variations in pressure produced variations in the resistance of the carbon. The current through the microphone changed in response to sounds in the vicinity. The electric current through the device was "modulated" by the sound waves, and the variations of current were characteristic of the air-pressure variations.

A later microphone was made up of two plates of a condenser, one plate being movable with respect to the other. When sound waves hit the variable plate, acting as the diaphragm, the spacing between the two plates changed and the voltage across the condenser changed in response to the changes in capacitance produced by the varying spacing between the two electrodes.

A more modern microphone is the dynamic loud speaker in reverse. When a coil placed in a steady magnetic field moves in response to sound waves impinging upon the diaphragm attached to it, voltages are produced across the coil. These varying voltages may be amplified and finally used to modulate the broadcast transmitter. The ribbon microphone has a light metallic ribbon suspended in a magnetic field. The crystal microphone has a piezoelectric crystal of Rochelle salt. When sound waves strike the diaphragm of a crystal microphone, mechanical pressures are exerted across the two faces of the crystal, and corresponding voltages are produced across them.

No matter what kind of microphone is employed, the voltages produced may be amplified and used to modulate the radio transmitter. The voltages are quite low, in general being about 60 to 80 db below 1 volt (roughly 1 to 0.1 mv) when the microphone is picking up sound of average intensity, such as a normal speaking voice. This is a small fraction of a volt. Considerable amplification is needed to bring the microphone output up to the level at which it may be placed upon studio or telephone lines without being masked by the line noise.

17.12. Pre-amplifiers. The first amplifier which the microphone output goes through is called a **pre-amplifier** or preliminary amplifier; it generally has a voltage gain of approximately 50 db or about 300 times.

Additional amplifiers are needed after the signals come in from the line if the pickup is at a distant point, and often amplifiers are used for the sole purpose of isolating the line from whatever apparatus the isolating amplifier feeds into. Since these amplifiers produce voltage gain, losses are often introduced deliberately with resistance "pads" so that the level out of the amplifier is not greater than desired.

17.13. Volume range and compression. A violin playing very softly has an output of about 4 mw, and a full orchestra at its peak has an output of about 70 watts. This is a volume range of 43 db. Before the weakest sounds picked up by the microphone can be placed upon lines or used by the transmitter, they must be raised in level above any possible line or transmitter noise. Before the loudest sounds picked up are utilized, they must be reduced in level to the point where they will not cause "cross talk" from one line to another or will not overload the transmitter. Thus the original 43-db volume range must be compressed, usually to a volume range of about 25 to 30 db, either automatically or manually, by operators who control the gain of some amplifier, so that the output falls within these limits.

If the average level of modulation is raised, listeners will get louder signals from a given broadcast station. If peak sounds are not to overload the transmitter, however, the average modulation must be kept quite low, perhaps 30 per cent. Some amplifiers now employed act as **volume compressors,** limiting the peak voltages automatically by reducing the amplifier gain and then allowing it to return to normal after the peaks have passed. In this manner, overloading of the transmitter is avoided and at the same time the average modulation level can be raised. Some distortion is, however, introduced by this process.

17.14. Modulation. If the r-f power radiated from the antenna is to convey intelligence to the distant receiver, some means must be provided for modulating the power with the intelligence. Simply keying the power on and off in accordance with a prearranged code is a means of so modulating r-f power. In this system the amplitude of the transmitting antenna current is varied from zero to maximum value.

If voice or music is to be used to modulate the transmitter, some means must be provided for varying the r-f power in the transmitter antenna according to both the strength and the frequency (tone) of the voice or music. Several methods are available: amplitude modulation, frequency modulation, phase modulation, and pulse or code modulation. These various methods are described in Sect. 14.1. Of these various schemes, amplitude and frequency modulation are the most widely used.

Amplitude modulation is used in standard broadcast transmitters. In this system the peak amplitude of the antenna current is varied in accordance with the strength of the microphone voltages as shown in Fig. 6. Thus a loud, audible sound produces a correspondingly great change in the antenna current. The *rate* at which the antenna current (not its amplitude) is varied depends upon the frequency of the audio sound.

In newer systems of police radio, and in the frequency-modulation broadcasting band, the frequency of the transmitter is varied in accordance with the audio tones. When the microphone is quiet (no modulation) the station sends out a fixed-frequency carrier wave just like an amplitude-modulated station. When the microphone picks up a sound, its voltage output is used to vary the frequency of the transmitter, the rate (that is, the number of times per second) at which the frequency is varied

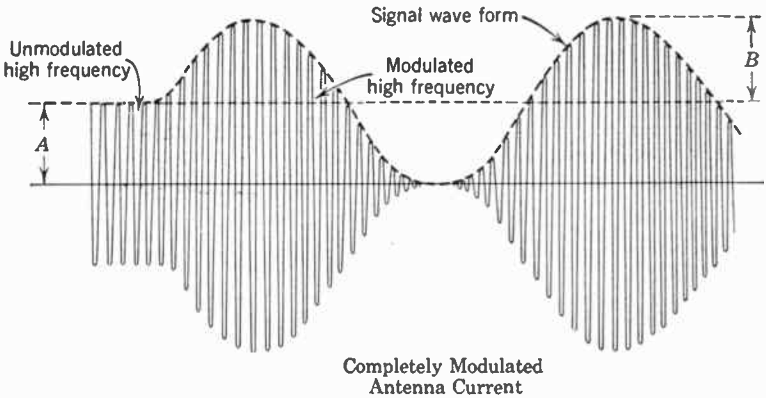


FIG. 6. Radio-frequency wave before and after modulation.

depending upon the frequency of the audio tone, and the extent of the variation being a measure of the strength of the microphone input.

Various methods of producing amplitude modulation will be discussed in this chapter. The details of frequency modulation are left for Chapter 19.

17.15. Low-level vs. high-level modulation. Amplitude modulation can be produced practically anywhere along the line-up of transmitter tubes from the oscillator to the final output stage. If it is performed early, say just after the crystal oscillator, modulation is said to take place at a *low level.* All succeeding stages must amplify this modulated r-f voltage. If modulation takes place in the final stage, where the power level is already high, it is said to take place at a *high level.* Each method has its advantages and disadvantages.

Audio-frequency power is required to modulate a r-f carrier wave, and more a-f power is required for high degrees of modulation. The amount of a-f power required to modulate a carrier wave is determined both by the power in the unmodulated carrier wave and by the method used to produce modulation. However, in general, it is true that the higher the power level in the unmodulated carrier the greater the a-f power required to effect modulation.

If modulation takes place at a low level, each of the succeeding amplifiers must handle both the carrier power and the modulating power; that is, they must handle both the power in the carrier and that produced in the sidebands by the modulation process. These succeeding amplifiers must be linear and, consequently, low-efficiency amplifiers. They must also be designed with sufficient band width to accommodate the sidebands, or somewhat more than 10,000 cycles for broadcast service. The amount of a-f power required to modulate a low-level stage, however, is small; in fact, all that is required is a good a-f amplifier with a few watts output.

If modulation is produced in the final stage, at high power levels, the amount of a-f modulation power must be quite large, but this disadvantage is partly overcome by the fact that the modulated and preceding stages can all be efficient Class C stages. Suppose that a 10-kw transmitter is to be plate-modulated (a common type of high-level modulation). The sideband power for 100 per cent modulation will, in this case, be 5 kw, and this power must be furnished by the a-f system. If the modulated Class C amplifier has a plate efficiency of 60 per cent the undistorted output of the modulator must be $5 \div 0.60$ or 8.33 kw. An a-f amplifier capable of supplying this power would be large and costly.

17.16. Amplitude-modulation systems. Modulation is actually the reverse of detection. In the modulation process r-f and a-f waves are combined, whereas in detection these waves are separated. In general, modulation can take place in the plate, control grid, suppressor grid, or cathode circuit of a tube, or it may take place in a copper oxide rectifier. Since there is such a wide range of modulating circuits, only a few of the most important will be discussed here.

A general requirement for a system which will produce a satisfactorily modulated wave is that the output current is proportional to the voltage of the tube electrode being modulated. For example, if the output current of a Class C r-f amplifier is proportional to the plate voltage, then, when the plate voltage is varied in accordance with the microphone voltages, the output current will be modulated in amplitude, and the antenna current will vary accordingly. Voltages picked up at the distant receiving station will bear these modulations, which can be used by the detector.

Modulation can be accomplished in the oscillator stage, but this is rarely done in commercial equipment because of the reaction upon the oscillator frequency. Practically all commercial transmitters introduce the modulation somewhere along the line of r-f amplifiers, preferably after at least one "buffer" stage.

17.17. Plate modulation. In the plate-modulated Class C amplifier (Fig. 7) r-f voltages of constant frequency and ampli-

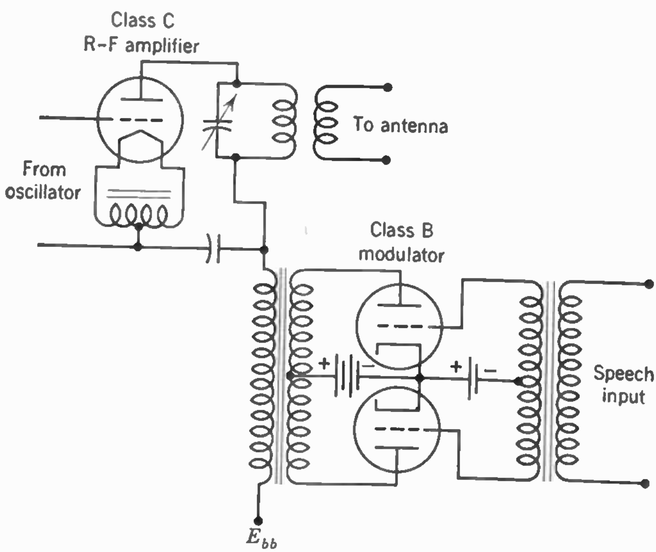


FIG. 7. Plate-modulated r-f amplifier.

tude are fed to the grid circuit, producing in the plate circuit r-f power at the frequency of the exciting voltage. In the absence of modulation this tube acts merely as a powerful r-f amplifier, feeding its output to an antenna where the r-f power is radiated at a constant frequency and unvarying amplitude. When the microphone in the studio picks up sound waves and converts them to voltages, and when these voltages are amplified sufficiently, they may be combined with the plate voltage (E_{bb}) of the Class C amplifier, at some instants adding to the voltage, and at other instants subtracting from it. Current and power transferred to the antenna vary in accordance with the variations in plate voltage. All that is necessary is to arrange the circuit and voltages properly so that the envelope of the modulated antenna currents is exactly like the microphone currents. This condition will be met if the r-f output of the tube varies linearly with respect to the plate supply voltage.

If the Class C tube has 1000 volts peak supplied to it from the plate voltage supply system, then 1000 volts peak of audio voltage must be supplied to modulate it completely. At these instants of 100 per cent modulation, the peak power output of the Class C tube is 4 times its unmodulated output. (Twice the voltage increases the power by 4 times.) Part of this additional power is supplied by the modulator, and the remainder comes from the plate voltage supply system. To modulate the amplifier completely, the modulator must supply 50 per cent ($E_{bb}I_b/2$) of the d-c plate power ($E_{bb}I_b$) that is required under non-modulated conditions. The modulator, therefore, must be capable of considerable power output. It is usually a highly efficient Class B audio amplifier, or one of the newer forms of amplifiers developed within recent years.

Pentodes and beam power tubes can be operated as plate-modulated Class C amplifiers provided modulation voltages are applied to both the plate and screen circuits as shown in Fig. 8.

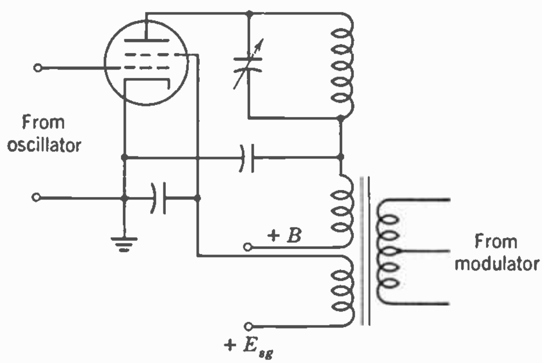


Fig. 8. Circuit in which both plate and screen voltages are modulated.

17.18. Design of plate-modulated amplifier. The relation between modulator and modulated amplifier may be understood from the following analysis for a plate-modulated system:

Power taken from d-c supply by Class C amplifier:

$$P_{bc} = E_{bb}I_b$$

Power required from Class B modulator:

$$P_m = \frac{m^2}{2} E_{bb}I_b = \frac{m^2}{2} P_{bc}$$

where E_{bb} and I_b are values for the Class C amplifier with no modulation and m is the degree of modulation (Sect. 14.1).

Total power supplied to Class C amplifier:

$$\begin{aligned} P_b &= P_{bc} + P_m \\ &= \left(1 + \frac{m^2}{2}\right) P_{bc} = \left(1 + \frac{m^2}{2}\right) E_{bb}I_b \end{aligned}$$

If the plate efficiency of the Class C amplifier is η_c, assumed constant over the modulation cycle, then the modulated r-f power output to the tank circuit is

$$P_{ac} = \eta_c P_b = \eta_c \left(1 + \frac{m^2}{2}\right) E_{bb}I_b$$

The power loss (plate dissipation) in the Class C tubes is the difference between the total power supplied, P_b, and the power output, P_{ac}, or

$$\begin{aligned} P_{pc} &= P_b - P_{ac} \\ &= \left(1 + \frac{m^2}{2}\right) E_{bb}I_b - \eta_c \left(1 + \frac{m^2}{2}\right) E_{bb}I_b \\ &= (1 - \eta_c)\left(1 + \frac{m^2}{2}\right) E_{bb}I_b \end{aligned}$$

The effective resistance of the Class C amplifier is $R_b = E_{bb}/I_b$, and the modulation transformer must match the modulator tubes to this resistance.

The modulated power output, P_{ac}, is the power into the tank circuit. The power supplied to the antenna will be the tank circuit power multiplied by the efficiency of the coupling network and transmission line. P_m is the modulator power output measured across the secondary of the modulation transformer. The power output of the modulator tubes must be somewhat larger than this to compensate for losses in the transformer.

Example 1. A transmitter uses a type 805 triode for the Class C amplifier and two 1623 triodes for the Class B modulator. The manufacturer's ratings for these tubes for one set of conditions follow:

TYPE 805 TRIODE—*as plate-modulated r-f power amplifier, Class C telephony. (Values are for a single tube with no speech input.)*

D-c plate voltage	1000 volts
D-c grid voltage	−155 volts
Peak r-f grid voltage	295 volts
D-c plate current	160 ma
Grid driving power	16 watts (approx.)
Power output	110 watts

TYPE 1623 TRIODE—*as a-f power amplifier and modulator, Class B operation. (Values for two tubes in push-pull.)*

D-c plate voltage	750 volts
D-c grid voltage	−25 volts
Peak a-f grid-to-grid voltage	200 volts
D-c plate current	
Zero signal	35 ma
Maximum signal	200 ma
Effective load resistance	
Per tube	2100 ohms
Plate-to-plate	8400 ohms
Maximum-signal driving power	4 watts (approx.)
Maximum-signal power output	100 watts (approx.)

The d-c power to the 805 amplifier is

$$P_{bc} = 1000 \times 160 \times 10^{-3} = 160 \text{ watts}$$

The a-f power required on the secondary of the modulation transformer for 100 per cent modulation is

$$P_m = \tfrac{1}{2} \times 160 = 80 \text{ watts}$$

The total power to the 805 plate circuit is

$$P_b = 160 + 80 = 240 \text{ watts}$$

The plate efficiency of the 805 amplifier is

$$\eta_c = \tfrac{110}{160} = 0.688 \text{ or } 68.8 \text{ per cent}$$

Note that η_c is calculated from values given for no modulation. The plate efficiency is assumed to remain constant during the modulation cycle.

The r-f power output for 100 per cent modulation is

$$P_{ac} = 0.688 \times 240 = 165 \text{ watts}$$

The plate dissipation of the 805 tube is

$$P_{pc} = 240 - 165 = 75 \text{ watts}$$

The modulation transformer must match 8400 ohms to $R_b = 1000 \div 160 \times 10^{-3}$ or 6250 ohms, and the turns ratio must be $\sqrt{8400/6250}$ or 1.16 (total primary to total secondary turns).

If the efficiency of the modulation transformer is assumed to be 90 per cent, then the power required from the Class B modulator must be $P_m \div 0.90 = 80 \div 0.90 = 89$ watts, which is well within the capabilities of the 1623 tubes.

If the coupling-circuit and transmission-line efficiency is 90 per cent, the antenna power for 100 per cent modulation is

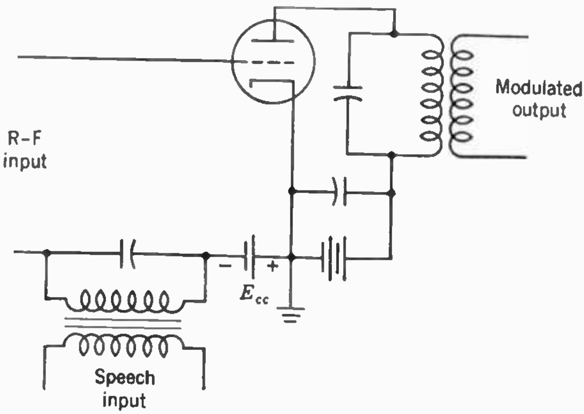
$$\text{Antenna power} = 0.90P_{ac} = 0.90 \times 165 = 148.5 \text{ watts}$$

17.19. Grid-bias modulation. A Class C amplifier may be modulated by introducing the modulating voltage in series with the r-f voltage in the grid circuit as shown in Fig. 9(*a*). In this circuit arrangement, the tube acts as though the bias were varied at an audio rate, and so the r-f output current is modulated and has much the same appearance as that obtained by plate modulation. The waveforms of grid voltage and plate current are shown in Fig. 9(*b*).

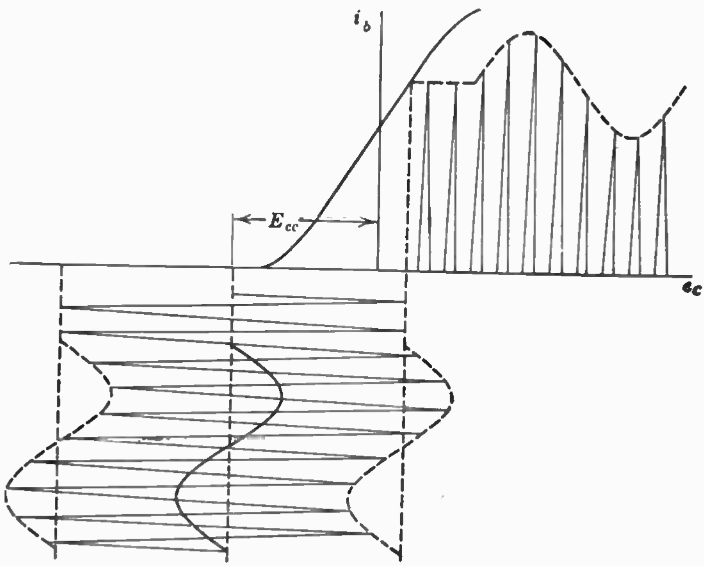
A condition for proper operation of a grid-bias-modulated amplifier is that the plate current and grid voltage be linearly related. This relation is rather difficult to achieve, and consequently there is somewhat more distortion than in a plate-modulated amplifier. Also the plate efficiency is somewhat lower. These disadvantages are counteracted by the smaller amount of audio-modulating power required.

The amount of harmonic generation in a grid-bias-modulated amplifier can be lessened by operating the modulated amplifier in a push-pull circuit. This scheme has been used for some low-powered broadcast transmitters.

In still another method, the modulating voltage can be introduced in series with the cathode. This is called **cathode modulation** and is a combination of plate and grid-bias modulation. The plate efficiency is intermediate between that obtained by plate and that obtained by grid-bias modulation.



(a) Circuit



(b) Voltages and Currents

FIG. 9. Grid-bias modulation.

17.20. Van der Bijl modulation. A circuit arrangement similar to that of Fig. 9 can be operated Class A and still produce modulation if the grid-bias and signal voltages are selected so that operation takes place on a non-linear portion of the I_b–E_b curve. That is, the operating point is on the lower, curved part of the curve. This scheme is called **Van der Bijl modulation.**

In common with all tube circuits operating on non-linear portions of their characteristic curves, this circuit produces many undesired harmonics. These harmonics must be removed with suitable tuned circuits or filters so that only the desired frequency terms are transmitted. Since the operation is Class A, only small amounts of carrier and audio power are required. Modulating systems of this type have been widely used in carrier-current telephony. They are not used in broadcast transmitters.

17.21. Analysis of modulation. In the antenna of a radio station flow alternating currents of the frequency of the carrier of the station. As long as the transmitter is not modulated, the peak amplitude of these alternating currents is constant. When, however, the microphone is spoken into or the station is otherwise modulated, the peak amplitude of the antenna current changes from instant to instant. This is the essence of modulation—a variation of the carrier-current peaks caused by the microphone voltages. These peaks increase and decrease in accordance with the frequency of the modulating tones.

The modulation factor, or percentage modulation, was defined in Sect. 14.1. It is the ratio $B/A \times 100$ per cent in Fig. 1, Chapter 14. Figure 6 (this chapter) shows the case for complete, or 100 per cent, modulation. The peak antenna currents in this latter case have doubled over the unmodulated values; the peak power is quadrupled; and the average power in the antenna is 50 per cent larger. More power is radiated from the antenna when the carrier is modulated.

A modulated waveform like that of Fig. 6 is the equivalent of a single carrier frequency plus two "sideband" frequencies displaced from the carrier by the audio frequency. Thus, if the carrier is 1000 kc and if the carrier is modulated with a pure 1000-cycle tone, the antenna will radiate three frequencies, 1000 kc minus 1 kc, 1000 kc, and 1000 kc plus 1 kc, and a sensitive and selective detector will pick up these radiated signals.

If a station is a broadcast station accepting all frequencies up to 10,000 cycles, for example, the sidebands of the station occupy the region between the carrier minus 10,000 cycles and the carrier plus 10,000 cycles. This is why the carriers of adjacent broadcast stations cannot be placed closer together than the highest modulating audio frequency. Otherwise their sidebands will overlap and create interference and distortion in the listeners' receivers.

The antenna current in this discussion has been described as having a symmetrical waveform as shown in Fig. 6, whereas the plate current in typical modulated amplifiers is unsymmetrical as shown in Fig. 9. The student may wonder how this can be. The reason again is that the pulses of plate current make the tank current (and the antenna current) symmetrical because of the storage action of the tank coil and condenser (Sect. 17.5).

17.22. Increase of antenna current with modulation. The antenna current of a transmitter will increase when modulation occurs because the power into the antenna increases by the amount of the modulation. Therefore the increase in antenna current can be taken as a measure of the percentage of modulation. This additional power is contained in the sidebands.

The average power in the carrier is $I^2R/2$, where I is the peak amplitude of the carrier current and R is the effective resistance of the tank circuit that is presented to the tube. The amplitude of the current in each sideband is the current in the carrier multiplied by $m/2$, where m is the modulation factor. Therefore the power in each sideband is $m^2I^2R/8$ or $(I^2R/2) \times (m/2)^2$, and the total power in carrier and sidebands, on the basis that each sideband is a replica of the other, is

$$\text{Total power} = \frac{I^2R}{2}\left(1 + \frac{m^2}{2}\right)$$

Since the current is proportional to the square root of the power, the current at any degree of modulation is proportional to

$$\sqrt{1 + \frac{m^2}{2}}$$

Thus the following table can be calculated. The same data are shown in Fig. 10.

Percentage Modulation	Percentage Increase in Antenna Current	Percentage Modulation	Percentage Increase in Antenna Current
0	0.0	50	6.0
10	0.25	60	8.6
20	1.00	70	11.6
25	1.6	80	15.0
30	2.2	90	18.5
40	3.9	100	22.5

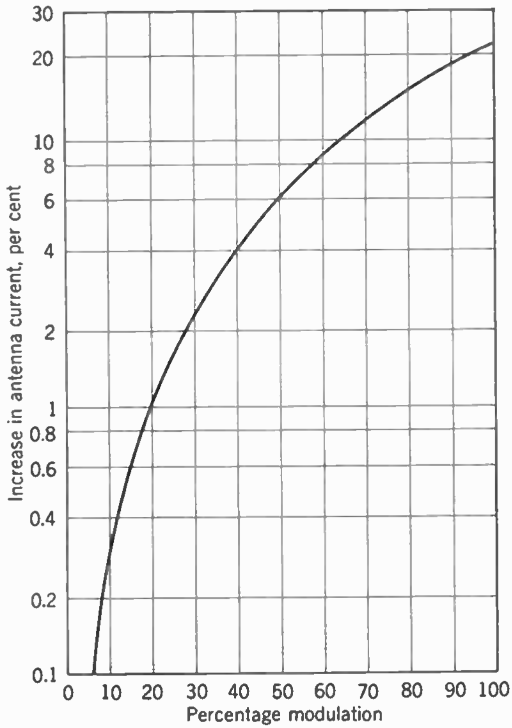


Fig. 10. Increase of antenna current with increasing modulation percentage.

18 · Transmission Lines, Antennas, and Electromagnetic Radiation

The methods of producing and modulating high-frequency energy have been discussed. The next step is to make this energy convey intelligence to a distant receiving station. Several questions arise. How is the energy conducted from the transmitter to the antenna and in turn radiated off into space? What is the nature of the process called radiation? How can the greatest amount of energy be radiated at the transmitter and the greatest amount picked up at the receiver?

18.1. Transmisson lines. Transmission lines have, heretofore, been considered merely as connecting links between pieces of electrical apparatus. No particular consideration has been given to their characteristics except that their inherent resistance may cause some loss of voltage and power. This concept is satisfactory as long as the transmission lines are electrically short; that is, as long as the length of the line is a small fraction of a wavelength. In practice, if the length of a transmission line is not greater than $\frac{1}{16}$ wavelength, the special characteristics of a transmission line, to be described later, need not be considered.

A wavelength * at 60 cycles is given by

$$\lambda = \frac{300 \times 10^6}{60} = 5 \times 10^6 \text{ meters}$$

$$= 3107 \text{ miles}$$

An ordinary 60-cycle power-transmission line is a small fraction of a wavelength long, and the special characteristics associated with transmission lines need not be taken into consideration except for very accurate calculations. At 1.5 Mc, however, a wavelength is only 200 meters or 965 ft, and $\frac{1}{16}$ wavelength is about

* See Sect. 7.23.

60 ft. Many transmission lines which connect transmitters to their antenna systems are considerably longer than this.

The L, C, and R of a transmission line cannot be considered as lumped in a particular place in the line. On the other hand, these circuit elements must be considered as distributed uniformly along the line. It is the distributed nature of these qualities that gives rise to the interesting and useful properties of transmission lines.

A rough picture of how the various elements of transmission lines are distributed is seen in Fig. 1. Each element shown must

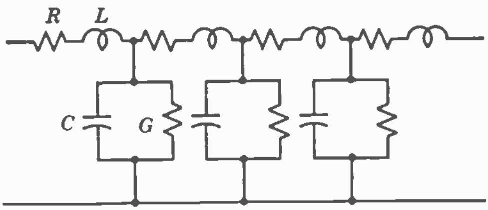


FIG. 1. Approximate equivalent circuit of a transmission line.

be taken as exceedingly small; a very large number of elements are required to represent even a short length of line. The series inductance L of the line arises from the fact that even a straight wire has inductance. At the same time it has resistance R. The shunt capacitance C arises since the two wires serve as the plates of a condenser with air or other dielectric material between the wires serving as insulation. The shunt conductance G (conductance = 1/resistance) is due to the imperfect insulation of solid dielectric coaxial lines, and to leakage across insulators on open-wire lines. Since all these elements are distributed evenly along the ideal line, their values are expressed for a unit length of line. For example, L might be 1 mh per meter, and C, 0.06 μf per meter.

Any circuit which has L, C, and R has a certain impedance, and a transmission line, having L, C, and R (also G), has an impedance called its **characteristic impedance** (sometimes called the surge impedance). This is the impedance with which the line should be terminated for the maximum transfer of power through the line. The characteristic impedance is determined by the size and spacing of the conductors, the kind of dielectric, and the height of the line above the earth if it is a single-wire line. These are also the factors which determine R, L, G, and C. Because of the distributed nature of R, L, G, and C, the length of the line has no effect on its characteristic impedance. Single-wire or two-wire lines in air have characteristic impedances in the range of 200 to 800 ohms. The characteristic impedance of coaxial lines is usually in the range of 50 to 100 ohms. The characteristic impedance of lines is almost entirely resistive, except at low audio frequencies, for which it is slightly capacitive.

The characteristic impedance of two-wire lines can be computed from the formula

$$Z_0 = 276 \log_{10} \left(\frac{2S}{d}\right)$$

where Z_0 = characteristic impedance in ohms.
S = spacing between wire centers in inches.
d = wire diameter in inches.

For example, the characteristic impedance of a pair of No. 14 wires, spaced 5 in. apart, is approximately 600 ohms.

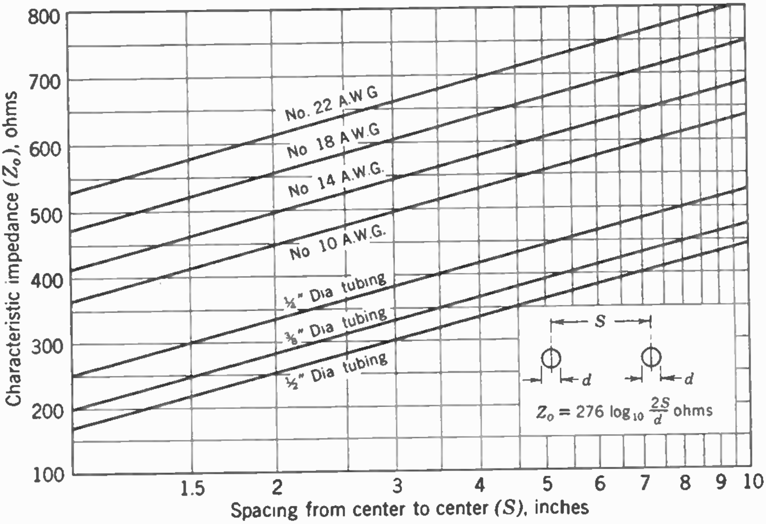
A coaxial (concentric) line is made up of two conductors, one located within the other and maintained in its position by insulating spacers or by solid insulating material. The characteristic impedance of such a line may be found from the formula

$$Z_0 = \frac{138}{\sqrt{k}} \log_{10} \left(\frac{D}{d}\right)$$

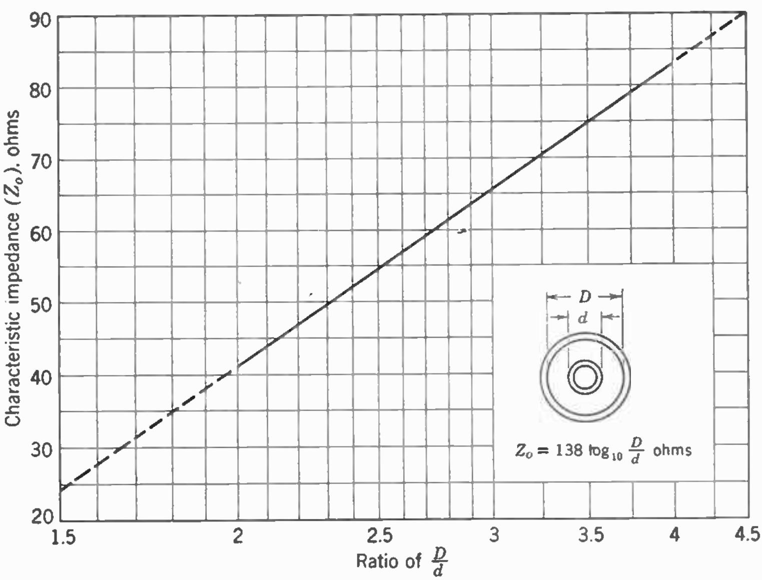
where Z_0 = characteristic impedance in ohms.
k = dielectric constant of the internal insulation.
D = inside diameter of outer conductor in inches.
d = outside diameter of inner conductor in inches.

For example, D/d is equal to 3.22 for a 70-ohm line. If the two conductors have diameters of ¾ and ¼ in., the characteristic impedance is 66 ohms.

In general, the characteristic impedance of either a two-wire or coaxial line is equal to $\sqrt{L/C}$, where L and C are henries per unit length and farads per unit length, respectively. The length unit is a matter of choice. It may be inches, meters, miles, etc. Charts showing the characteristic impedances of open-wire and coaxial lines are shown in Fig. 2.



Impedance of Two-Wire Transmission Lines



Impedance of Air-Filled Coaxial Line

FIG. 2.

The two wires of an open-wire line are usually arranged so that they are at equal distances from the earth or surrounding objects. When so arranged they are called **balanced** lines; the capacitance from each wire to earth is the same. Coaxial lines are inherently **unbalanced** lines; their outer conductor is normally operated at ground potential, whereas the inner conductor has considerable capacitance to ground. Coaxial lines cannot be employed in applications requiring balanced lines except through the use of auxiliary balancing circuits. Coaxial lines have the distinct advantage that they are self-shielding, the outer conductor effectively shielding the line from external fields. For the same reason the radiation from coaxial lines is less than from open-wire lines, and this is particularly important at the very high frequencies.

18.2. Voltage and current distribution along a line. When a transmission line is terminated in a load equal to its characteristic impedance (more briefly, "terminated in its characteristic impedance"), the voltage and current decrease gradually from the generator (sending) end to the load (receiving) end as shown in Fig. 3. The greater the attenuation, or inherent loss of power in the line, the more rapidly the voltage and current drop, but they always drop in a smooth curve as long as the load is equal to the characteristic impedance. A line so terminated is called a non-resonant or "flat" line.

If, however, a line is terminated in a load not equal to its characteristic impedance, an important phenomenon is observed. Neither the voltage nor current varies smoothly along the line. Instead the profile of each is characterized by peaks and valleys, and the greater the departure from a characteristic impedance load, the more pronounced are the peaks and valleys. Consider, for example, a line short-circuited at the load end. The voltage at the load end must be zero because of the short circuit. The current will have a definite and finite value because of the impedance of the line. The general appearance of the voltage and current along such a line is shown in Fig. 4 for two cases. In Fig 4(a), the case for a perfect line with no attenuation, the voltage returns to zero every half wavelength * along the line,

* A wavelength along a transmission line is equal to the velocity with which the wave moves divided by the frequency, or $\lambda = v/f$. On open-air

measured away from the load. In a similar manner the current has a value equal to the load value at regular intervals equal to a half wavelength. The case for a line with attenuation is shown in Fig. 4(*b*). The effect of attenuation is to keep the voltage and current from making as wide excursions up and down as in the first case.

The undulations of voltage and current discussed above are caused by what are known as **reflections.** When the load is other than the characteristic impedance, a part of the power that

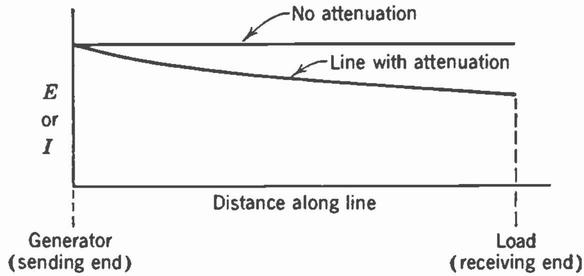


FIG. 3. Voltage and current distribution along a line terminated in its characteristic impedance.

would be absorbed in a characteristic impedance load is reflected back toward the generator. The phase of the oncoming voltage wave, for example, at certain points along the line, is the same as the phase of the reflected wave. These add to produce the maxima. At other points along the line the two waves are out of phase and cancel, producing the minima.

Figures 4(*a*) and 4(*b*) apply for the case of a line open-circuited at the load end if E and I are interchanged. Similar diagrams can be drawn for any load; however, for loads other than open- or short-circuited there will be no zeros for either voltage or current even for lines with no attenuation. Some of these cases are discussed in a later section.

The voltages discussed above are the a-c values such as would be read by a voltmeter connected across the line and slid along

lines the velocity is approximately equal to the speed of light, 3×10^8 meters per second or 186,000 miles per second. On lines with solid dielectric the velocity of light must be divided by $\sqrt{k}$, where k is the dielectric constant. See also Sects. 7.22 and 7.23.

from point to point. At any given point the voltage goes through its alternations with time, and these could be observed with the aid of an oscillograph. Likewise, the currents are those which would be measured by an a-c ammeter inserted in series with the line at various points.

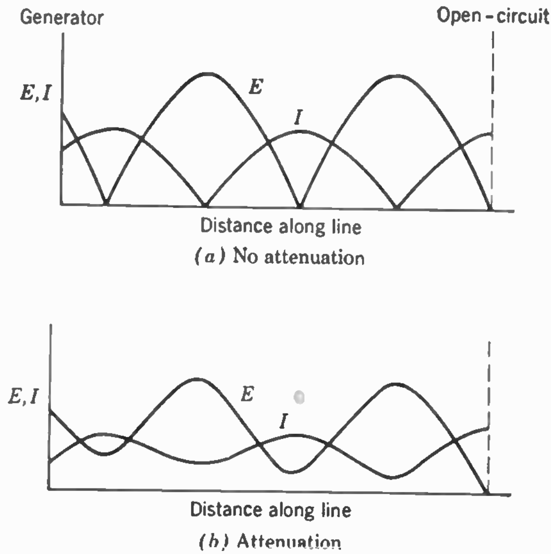


Fig. 4. Voltage and current along a short-circuited line.

18.3. Quarter- and half-wavelength lines. Transmission lines of a particular length for the frequency for which they are designed have very interesting and useful characteristics. A quarter-wavelength line has such a physical length (in feet, for example) that it is a quarter wavelength long at the operating frequency. Such a line can be used as a matching transformer to connect a generator of one impedance to a load of another impedance with the purpose of transferring the maximum power. For example, let Z_i be the impedance of the generator to which the input end of the line is connected; Z_r be the impedance of the load to which the other end of the line is connected; Z_0 be the characteristic impedance of the line. Now, if Z_0 is chosen so that it is equal to $\sqrt{Z_iZ_r}$, the maximum power will be transferred from the generator to the load. For example, if a 600-ohm line is to be connected to a 70-ohm antenna, this can be accomplished by using a line one-quarter wavelength long and with a characteristic impedance of $\sqrt{600 \times 70}$ or 205 ohms.

The relation between these three impedances may also be expressed as

$$Z_i = \frac{Z_0^2}{Z_r}$$

That is, the impedance looking into the generator end of the line is equal to the square of the characteristic impedance divided by the load impedance. Now what happens if Z_r is very low (a short circuit, for example)? In this case Z_i will be very high; and if Z_r is very high (an open circuit), then Z_i will be very low. Such a line is indeed an impedance transformer like a matching transformer used in a-f work.

A half-wavelength line, on the other hand, has a quite different quality. The impedance looking into such a line is equal to the impedance terminating the other end of the line. This is true regardless of the impedance of the line itself. Thus a half-wavelength line acts like a one-to-one transformer and may be used as such.

A half-wavelength line open-circuited at the far end has a high input impedance. On the other hand, if a quarter-wavelength line is short-circuited at the far end, the impedance looking into it is high. Either line resembles a parallel circuit tuned to resonance. If for some reason it is desirable to connect an impedance across a circuit, say an antenna, having a high value at the operating frequency but a low value at all other frequencies, a short-circuited quarter-wavelength line may be used. The analogy to this usage is a parallel tuned circuit shunted across another circuit to drain off power at unwanted frequencies; currents lower in frequency than the resonant frequency will find a low-impedance path through the inductance, and higher-frequency currents will find a low-impedance path through the condenser.

Lines slightly longer or shorter than a quarter or half wavelength also possess interesting properties. For example, an open-circuited line shorter than a quarter wavelength has a capacitive input impedance. All or part of such a line could be replaced by a condenser with the proper reactance. This is sometimes done when it is required that the effective electrical length of the

TRANSMISSION-LINE CHARACTERISTICS

Length of Line	Impedance at Far End	Z Looking into Line	Equivalent Circuit
$\frac{1}{4}\lambda$	High Zero Z_r	Low High $\sqrt{\frac{Z_0^2}{Z_R}}$	Series tuned circuit Parallel tuned circuit Transformer
$\frac{1}{2}\lambda$	High Zero Z_r	High Zero (low) Z_r *	Parallel tuned circuit Series tuned circuit 1–1 transformer

* Independent of line impedance.

line be variable, as when a line is used as the tuned circuit of a r-f amplifier. Figure 5 shows the equivalent "lumped" impedances of several combinations.

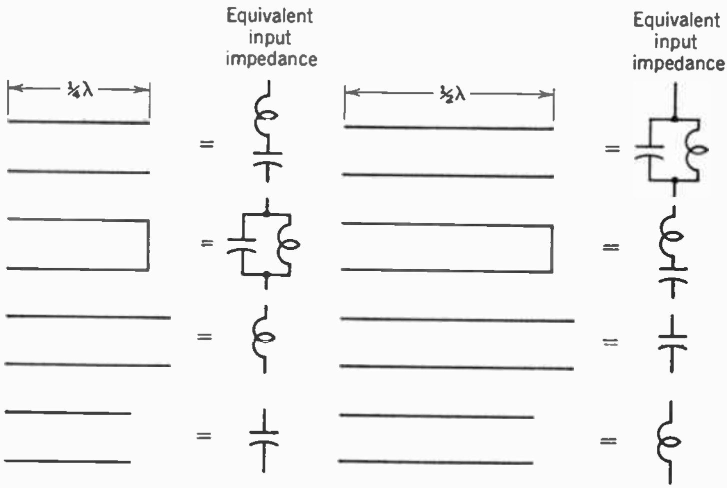
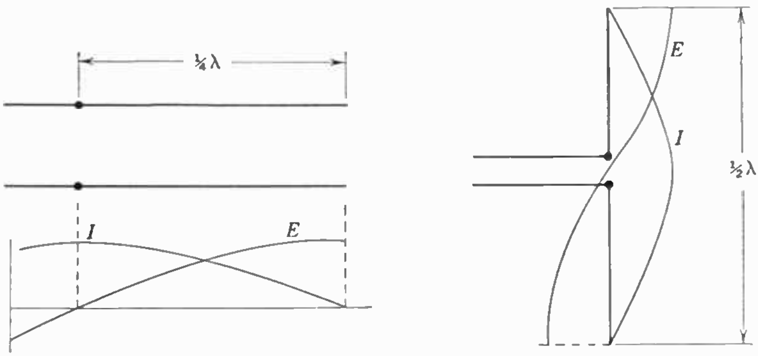


FIG. 5. Equivalent input impedance of several transmission line configurations.

18.4. Dipole antennas. One of the simplest types of antennas is the **dipole** or **half-wave doublet,** also called a Hertz half-wave radiator. It can be constructed by connecting two



(a) Quarter-Wavelength Line (b) Half-Wavelength Dipole

Fig. 6. Voltage and current distributions along a quarter-wave open-circuited line and a half-wave dipole are similar.

wires, each a quarter wavelength long, at right angles to a transmission line. Bending the wires at right angles in this manner makes it relatively easy for them to radiate. Transmission lines, with wires close together, do not radiate appreciable amounts of energy. In Fig. 6(a) is shown a transmission line connected to an open-circuited quarter-wavelength section of line. The voltage and current distribution is shown below the line. Clearly the current at the open-circuited end of the line must be zero—current cannot flow into an open circuit. The voltage at this point is maximum. At the point one-quarter wavelength back along the line, where the short section connects to the main line, the current is maximum and the voltage zero. Now suppose that this short section of line is bent at right angles to the main line to form a dipole antenna as shown in Fig. 6(b). The voltage and current distribution is as shown. Again, the current at the outer ends must be zero. The voltage and current do not quite follow a sinusoidal pattern as they do in Fig. 6(a) because the distributed capacitance and inductance vary along the length of the wires which have been bent at right angles. For example, the capacitance between short sections near the center of the antenna is greater than that between sections near the outer ends. However, the pattern of the current and voltage distribution does not depart markedly from sinusoidal.

At the center of the dipole in Fig. 6(b), the voltage E is low and the current I is high in respect to the voltage and current at the outer ends. Their ratio is the input impedance to the antenna and is very nearly 72 ohms, and this impedance is resistive in nature. Practically all this resistance is made up of the radiation resistance of the dipole; a very small part, a few ohms at most, is made up of the inherent resistance of the antenna wires. Since the dipole is fed its power at a point of high current and low voltage, it is said to be "current fed."

The **radiation resistance** of an antenna represents the effective resistance into which the power is fed that is radiated into space. It has different values for different antennas. The antenna can be thought of as a sort of transformer which couples energy from the transmission line into space. In doing so, a certain amount of effective resistance is reflected back to the transmission-line output terminals. This radiation resistance is a real thing so far as the transmission line is concerned, for it is the resistance to which the transmission line must feed power and the resistance to which it must be matched.

Certain antennas, notably those which are not integral quarter wavelengths long, possess input impedances which have a reactive component, either inductive or capacitive. Usually a suitable coupling device is used at the end of the transmission line which cancels this reactance, leaving only the resistive portion (radiation resistance) to be supplied power by the transmission line.

A dipole acts very much like a resonant circuit or like the open-circuited quarter-wavelength line of Fig. 5. A dipole may be vertical or horizontal with respect to the earth, and energy may be fed to it in several ways, some of which are discussed in later sections.

18.5. Voltage-fed dipole. It is not necessary to feed a half-wave dipole at its center. Instead, the antenna may be fed at one end as shown in Fig. 7. The voltage at the end will be high and the current low, and consequently the input impedance is high. The antenna is said to be "voltage fed."

The ordinary coaxial or open-wire line will not operate very efficiently when coupled directly to the end of a half-wave dipole because of the wide difference between the characteristic impedance of the line and the input impedance to the antenna at the feed point. Therefore, impedance-matching devices must be used. Methods of coupling such an antenna to both a coaxial and an open-wire line are discussed in Sect. 18.11.

The most important advantage of a vertical voltage-fed dipole is that it is of simple construction, particularly when fed from a coaxial line. The transmission line may consist of an upright metal pipe containing an inner conductor extending the proper distance from the upper end. In a measure, such an antenna is self-protected from lightning strokes since the outer conductor can be grounded and appropriate lightning arrestors can be

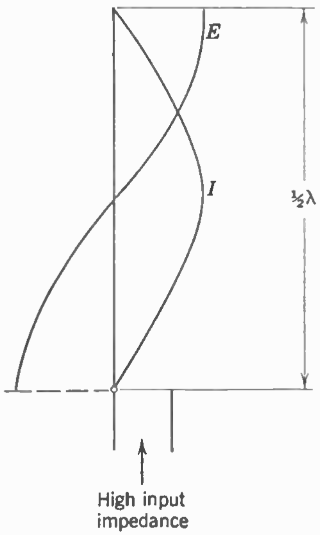


FIG. 7. Voltage-fed half-wave dipole.

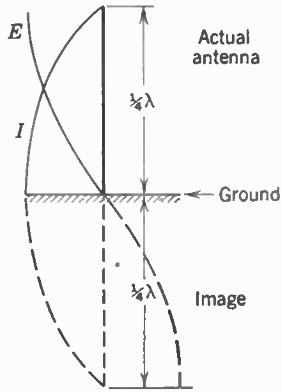


FIG. 8. Quarter-wave grounded antenna.

placed between the outer and inner conductors. Because of its appearance, such an antenna is often called a "flagpole antenna."

18.6. Grounded antennas. Dipoles are usually operated at a considerable distance above the earth. It is feasible, however, to operate a vertical antenna with one end grounded. In fact, if a quarter-wavelength wire or structure is grounded and excited near the grounded point, the operation is much the same as that of a dipole. The distribution of the current and voltage is virtually the same as for one half of a dipole, as is seen in Fig. 8. The "image" shown below ground level represents a section of wire which would have an effect equivalent to that of the currents induced in the earth by the actual antenna. The radiation resistance is approximately half that of an ungrounded dipole, or

35 to 37 ohms. Since the current at the base of such a grounded antenna is high, it is important that the ground resistance be low to avoid large power (I^2R) losses. Several radial wires, like spokes of a wheel, are often buried in the earth under a grounded antenna to insure that the effective ground resistance will be low. Sometimes the grounding system is mounted on insulated stakes a short distance above the earth. The grounding system is then called a **counterpoise.**

A grounded antenna may be more or less than a quarter wavelength long to obtain a specific radiation pattern in the vertical plane. The input resistance varies with the length, and the total input impedance usually has a reactive component.

The physical length required for an antenna which is a specified number of electrical wavelengths long depends upon the frequency of the desired radiation, the cross section of the antenna structure if it is other than a simple wire, and the nearness of the antenna to surrounding objects. A thin wire in space, isolated from surrounding objects, will have approximately the physical lengths shown below.

Half-wave dipole:

$$L = \frac{480}{f} \text{ ft} \quad \text{or} \quad \frac{146}{f} \text{ meters}$$

where f is the frequency of operation in megacycles.

Quarter-wavelength antenna:

$$L = \frac{240}{f} \text{ ft} \quad \text{or} \quad \frac{73}{f} \text{ meters}$$

The above formulas give approximate lengths only, even though they are corrected for "end effects." In constructing an antenna it is usually wise to cut it slightly longer than the calculated value, then trim it to the proper length as it is being adjusted.

18.7. Antenna characteristics. Antennas have inductance, capacitance, and resistance. The only useful part of the resistance is the radiation resistance; the rest of the resistance—d-c resistance, the increase of resistance due to skin effect, and that representing loss of energy in the dielectric material near the antenna—represents energy wasted. The radiation resistance may be represented by a resistor which, inserted in place of the antenna, would absorb as much power as that radiated by the antenna. If this radiated power could be measured, and if the current at the center of a half-wave dipole or at the base of a grounded quarter-wavelength antenna could be measured, then the radiation resistance would be

$$R = \frac{P}{I^2}$$

where P is the radiated power and I is the current to the antenna.

The longer the wavelength to be transmitted, the larger must be the physical structure of the antenna to resonate to this wavelength. On the very long wavelengths (low frequencies) it is practically impossible to make the antenna big enough to be resonant to the transmitting frequency. Therefore, so far as the transmission line feeding the antenna is concerned, resonance is achieved by "loading" the antenna with inductance. Any power losses in the loading coil are wasted so far as radiation is concerned. Another method of increasing the effective length of a vertical antenna is "top loading"; that is, an "umbrella" of wire or metal is mounted at the top of the antenna. This increases the effective length of the antenna.

An antenna operated at a resonant wavelength (a half-wave antenna, for example) is efficient because its radiation resistance is high, but an antenna which is shorter than a half wavelength is inefficient because its radiation resistance is low and may not be much greater than the inherent resistance of the wires making up the antenna. Therefore most of the power put into such an antenna is used in heating the antenna and the surrounding physical objects, and the amount which is radiated is correspondingly less.

The general manner in which the input impedance of a center-fed dipole varies as the frequency is varied below and above the value at which it is resonant (half wavelength) is shown in Fig. 9. At resonance the radiation resistance is maximum and the reactance is zero. Below resonance the radiation resistance falls off and there is a capacitive component of the input impedance. A similar thing occurs above resonance except that the reactance is inductive. The resistance r in Fig. 9 is the inherent resistance of the antenna. In comparison with the radiation resistance, the resistance r is small at resonance but it becomes significant at other frequencies.

The sharpness of the curve of radiation resistance in Fig. 9 depends largely upon the size of the wire of which the antenna is made. If the antenna wire is quite small, No. 14, for example, but low in inherent resistance, the curve is quite sharp. If, on the other hand, the antenna structure is quite thick, the curve

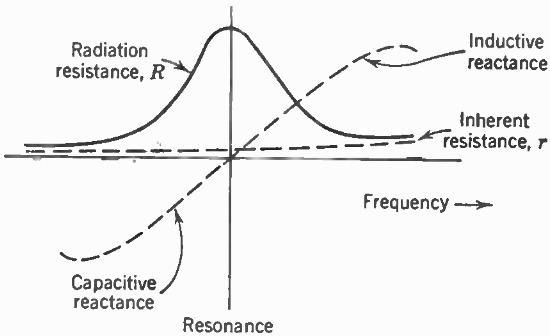


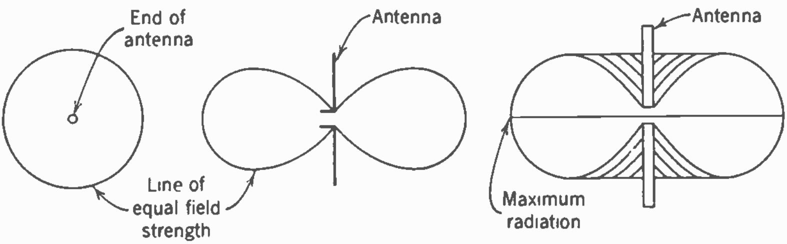
FIG. 9. Variation of impedance of center-fed dipole with frequency.

rises and falls rather gradually, the antenna is said to be a low-Q antenna, and its efficiency is impaired. However, low-Q antennas are useful for television service for which the band width of the signal is quite broad.

Even though the curves of Fig. 9 are for a particular type of antenna, a dipole, the statements made above apply in general to most types of antennas.

18.8. Horizontal and vertical patterns. If a person were to measure the field strength of an antenna at some distance from the base of the antenna, say 10 miles, and then were to move around the antenna a short distance, search for a point where the same field strength existed, continue this process until the antenna had been circled, and finally plot the results, he would have a **horizontal field pattern** for the antenna. If the same field strength was found to occur at equal distances from the antenna, the pattern would be a circle and the antenna would be **non-directional.** If the resulting pattern had pronounced lobes, the antenna would be called a **directional** antenna.

If, in a similar manner, the points at which equal field strengths occurred in an arc through the air and over the top of the antenna and down to the earth on the other side could be found and the results plotted, the plot would be the **vertical field pattern.** Obviously the vertical pattern cannot be measured in most cases; it can, however, be calculated for a good many types of antennas. Sometimes it is possible to turn a small antenna on its "side" and measure the vertical pattern in the same way as the horizontal pattern was measured.



(*a*) Horizontal Pattern (*b*) Vertical Pattern (*c*) Composite Pattern

FIG. 10. Field patterns of an isolated half-wave dipole.

A simple dipole, if mounted vertically and well above the earth, has a circular horizontal pattern as shown in Fig. 10(*a*). This should be expected since there is nothing to cause more radiation in one direction in a horizontal plane than in any other direction. The vertical pattern, on the other hand, is quite directional. The strongest radiation is in a direction at right angles to the center of the antenna. There is less radiation at angles oblique to the antenna, and there is no radiation off the ends of the antenna. A quarter-wavelength grounded antenna has horizontal and vertical patterns which look much the same as those in Fig. 10, except that the vertical pattern below the center-line is absent.

Earth or surrounding objects may change the field patterns of an antenna markedly. For example, if a vertical half-wave dipole is operated one-quarter wavelength above the earth, the vertical pattern is as shown in Fig. 11. The maximum radiation is still in the horizontal direction, but there are minor lobes at a high angle, and, of course, no radiation below the horizontal.

The horizontal pattern is circular as in the case of the isolated dipole.

The vertical field patterns of vertical grounded antennas may contain many lobes when these antennas are operated at other than their resonant frequency. At some frequencies most of the energy is radiated in an almost horizontal direction, and operation of an antenna in this fashion would be desirable for broadcast station use. At other frequencies most of the energy is directed at a high angle toward the sky. A vertical grounded

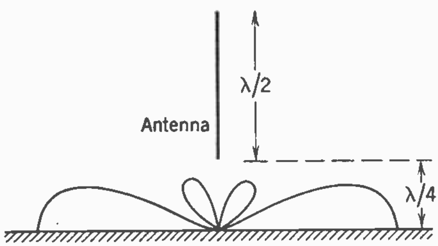


Fig. 11. Radiation pattern of a vertical doublet ¼λ above ground.

antenna that is 0.625 wavelength long radiates the maximum possible energy in the horizontal direction, while a similar antenna 1 wavelength long radiates its maximum energy into the sky at an angle of 60° above the horizontal.

The field patterns of an antenna apply equally well whether they are to be used for transmitting or receiving purposes. That is, they will radiate the maximum power or pick up the most power in the directions for which the maximum values of the lobes of the patterns occur and will radiate little power and be insensitive in the directions in which the minimum points of the patterns occur.

Except for fixed-frequency receivers there is generally no effort made to use a receiving antenna which is resonant to a particular frequency. On the contrary, the antenna system is usually operated at a point far removed from resonance, and the consequent loss of sensitivity made up by amplification within the receiver.

18.9. Antenna arrays. When it is desirable to concentrate more radiation in a given direction than can be done with simple antennas, combinations of antennas are used in such a manner that radiation in one or more directions is increased and that in other directions decreased. These combinations of antennas are called **directional arrays.** Arrays cannot, of course, increase the total radiated power, but they can redirect the available energy and concentrate it in a desired direction.

There are several reasons why arrays are desirable. Broadcast stations must often suppress their radiation in certain directions to avoid interference with other stations operating on the same frequency, or a station may be located on the seacoast or in front of a range of mountains and radiation toward the sea or mountains would be wasted. Directional combinations may also be used to reduce the amount of energy radiated toward the sky and concentrate the available energy in the horizontal wave. Highly directional arrays are also used for aircraft beacons, radar, and similar purposes.

A simple array consists of two vertical antennas spaced one-fourth wavelength apart and excited 90° out of phase. The horizontal pattern of such an array is shown in Fig. 12. Practically all the energy is directed in the "forward" direction with very

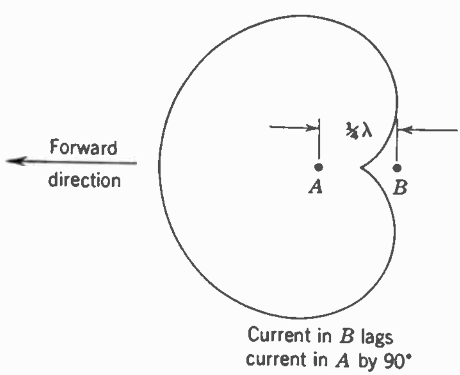


Fig. 12. Horizontal field pattern of two similar vertical antennas spaced ¼λ apart and excited 90° out of phase.

little being radiated rearward. Other antenna spacings and excitations will produce different field patterns for the array.

Sometimes two or more antennas are used in an array but only one antenna is energized from the transmitter, the others being excited by induced currents resulting from a kind of transformer action. The antenna connected to the transmitter is called a "driven" antenna, while the others are called "parasitic" antennas.

18.10. Loop antennas. Loop antennas are simply coils of wire and may have almost any shape: square, rectangular, or circular. The dimensions of the ordinary loop are small in comparison with the wavelength. A loop is a directional antenna. It receives better in the direction toward which the narrow dimen-

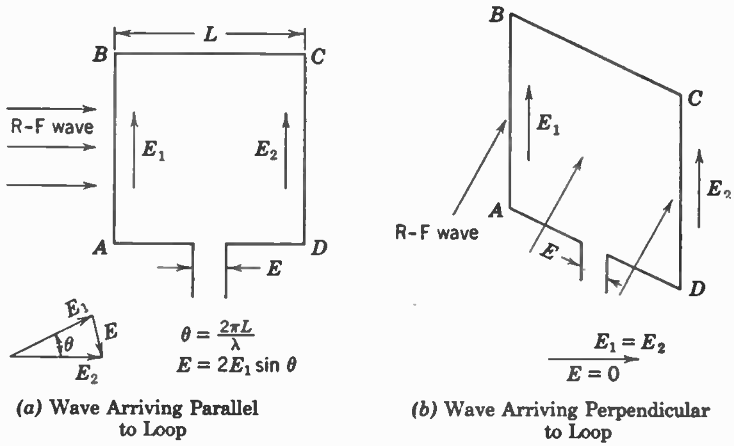


Fig. 13. Loop antenna. Induced voltage E is greatest when wave arrives in a direction parallel to the plane of the loop; it is zero when the direction is perpendicular.

sion points. Its pattern has a zero for a wave which arrives in a direction perpendicular to the loop.

The directional characteristics of a loop antenna can be explained as follows. Suppose that a wave arrives in the direction of the plane of the loop as in Fig. 13(*a*). When the wave passes the vertical section AB of the loop, a voltage E_1 is induced. A short time later the wave passes the vertical section DC and induces a voltage E_2 in this section. No voltage is induced in the top and bottom portions, which lie in the direction of travel of the wave. The two voltages E_1 and E_2 are equal in magnitude but slightly different in phase. Voltage E_2 lags E_1 because the wave passes DC at a slightly later time. The two voltages oppose each other, and the resultant voltage E is the vector sum of E_1 and E_2 as shown in the figure.

Suppose, now, that the wave arrives in a direction *perpendicular* to the plane of the loop as in Fig. 13(*b*). The voltages E_1 and E_2 are again equal in magnitude; they are also in phase since the wave passes both *AB* and *DC* at the same time. The net induced voltage is zero.

Loop antennas are usually a rather small fraction of a wavelength "wide," and the resultant induced voltage is small. Several turns are usually employed to increase the output voltage. The above analysis was made for a square loop, but the same principles apply for a loop of any shape.

In connection with a sensitive receiver, a loop antenna may be used to determine the direction from which signals come. The loop antenna is the heart of the radio compass and **direction-finding** stations which are situated along the seacoasts of the world. When a ship wants bearings, its signals are picked up by the coastal station, which determines the position of its receiving loop which gives the least signal—a more accurate indication than the position which gives a maximum signal. A compass is attached to the base of the loop, and the indicator then points to the bearing of the vessel. A receiving operator in another location swings his loop on the vessel, and in this way two bearings are obtained. From them the master of the ship can determine his position. This method is illustrated in Fig. 14.

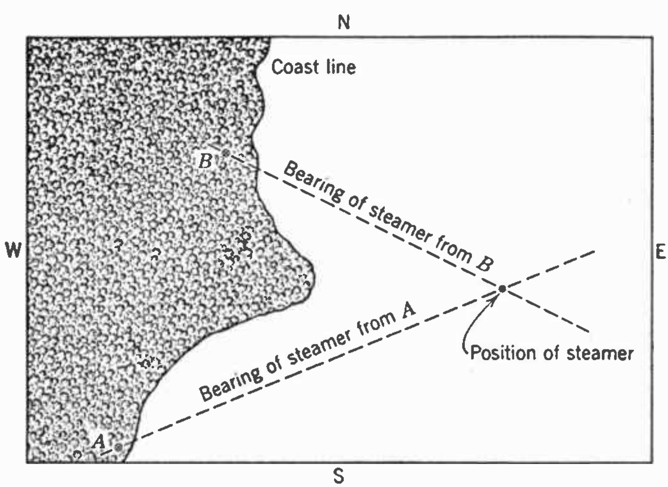


Fig. 14. Method of plotting a ship's position by obtaining bearings from two land stations.

There is one disadvantage of the loop when used alone. Its directivity pattern is composed of two equal loops, so that the operator can tell that a signal is coming, for example, from a north-south direction, but he cannot determine whether the signal is coming from the north or from the south. The addition of a single vertical antenna whose output is properly added to that of the loop solves this problem. The combination may be

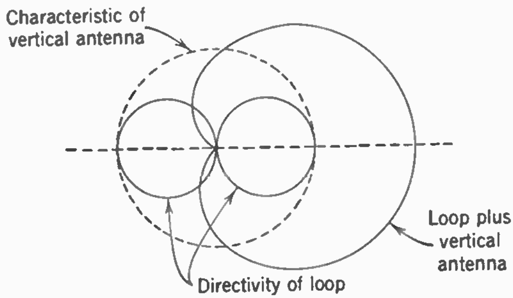


FIG. 15. Directional patterns of loop and vertical antenna, and the two combined.

adjusted to produce a pattern as shown in Fig. 15. In use, the loop alone is rotated until the least signal is received. Then the antenna is rotated exactly 90° and the vertical antenna switched into the system. If the signal is now much greater than it was, then it must have come, say, from the north. If the signal does not increase, then it must have come from the south.

18.11. Coupling transmission lines to antennas. A wide variety of methods is available for coupling transmission lines to antennas to secure the proper impedance match and the maximum transfer of power.

The input impedance of a half-wave dipole is approximately 72 ohms, and this is much lower than the characteristic impedance of most open-wire lines (200 to 800 ohms). However, it will be seen in Fig. 6 that at points away from the center of the antenna the voltage increases and the current decreases. This means, in effect, that the impedance is higher away from the center. This suggests that the transmission line may be flared out so that connections are made at some distance from the center of the antenna to obtain an impedance match. Two methods are shown in Fig. 16. In the upper diagram a long insulator is used to insulate the center of the antenna and the flared ends of a twisted-wire line are connected to the antenna. It is not necessary, however, to insulate the center of the antenna. The transmission line may merely be flared out and connected as shown in the lower diagram. The tapered flare connecting the transmission line to the

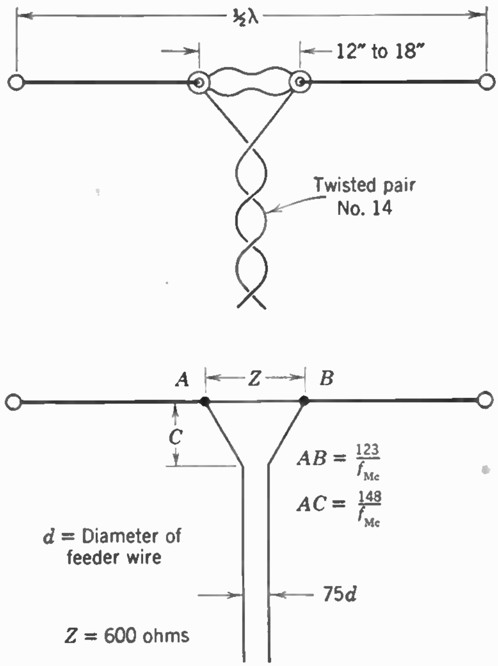


FIG. 16. Manner of connecting transmission lines to antennas to produce proper impedance match.

antenna must be long enough so that the spacing between the wires changes gradually. Typical dimensions are shown on the two diagrams.

Since the impedance of an antenna is always the ratio of the voltage to the current, the impedance along a length of the wire varies from point to point. For example, if a half-wave wire for use at 7 Mc is tapped in the center, it may be fed with a 72-ohm line; but, if it is tapped about 22 ft from one end as seen in Fig. 17, it may be fed with a 300-ohm lead since at this point the ratio of E to I is approximately 300. Such an antenna will work well on both 7 and 14 Mc. If made twice as long and fed at 44 ft from end, it will resonate on 3.5, 7, and 14 Mc.

A single-wire feeder can be used to connect to an antenna as shown in Fig. 18. Here the return path for the current is through the capacitance of the antenna to ground.

It was previously mentioned that the input impedance to the

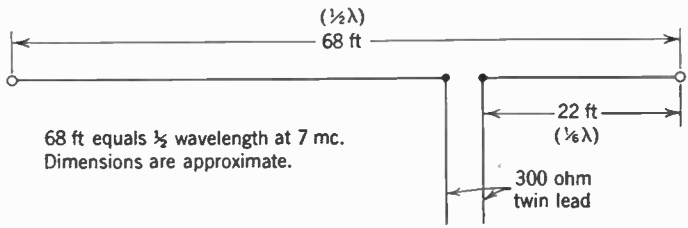


FIG. 17. Method of feeding a half-wave antenna with a 300-ohm line at 7 and 14 Mc.

voltage-fed half-wave dipole in Fig. 7 is quite high and that coupling sections are required between the transmission line and the antenna. Two methods of matching the impedance in such a system are shown in Figs. 19 and 20. In Fig. 19 a wire is extended one-quarter wavelength from the outer conductor of a coaxial line and the center conductor is extended a half wave-

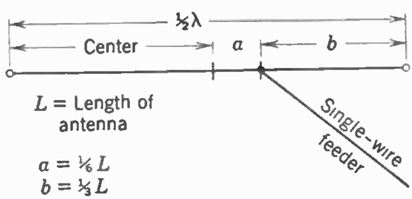


FIG. 18. Proper place to connect a single-wire feeder to an antenna for most efficient energy transfer.

length beyond the end of this wire. At the lower end of the half-wavelength section the voltage and impedance are high; the quarter-wavelength section is then effectively terminated in a high impedance, and the impedance at its lower end (the end to which the coaxial line connects) is low. In practice, the quarter-wavelength matching section may be trimmed slightly to secure the proper impedance match. There is little radiation from the quarter-wavelength section because of the relatively close spacing of the two wires; most of the radiation occurs from the half-wavelength dipole.

In Fig. 20 is shown a "J" antenna. Here the half-wavelength dipole is connected to a short-circuited quarter-wavelength open-wire line. The transmission line is then connected to a point

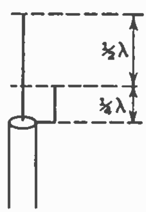


Fig. 19

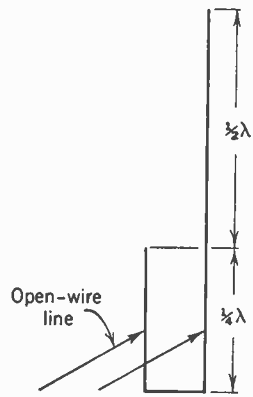


Fig. 20

Fig. 19. Flagpole antenna using ¼λ coupling line to connect coaxial cable to ½λ antenna.

Fig. 20. "J" antenna. Position of open-wire line is adjusted for an impedance match.

along the quarter-wavelength matching section where an impedance match is achieved.

A resonant line may be used to connect a transmitter to an antenna, the advantage being that no impedance matching at the antenna end is needed. The line really acts like an extension of the antenna, but, because its two conductors are close together, it does not radiate much energy. However, resonant lines may have considerably more loss of power than non-resonant lines, and this is frequently a serious disadvantage.

Problem 1. In Fig. 21 is shown a two-wire quarter-wavelength line connecting a 500-ohm transmission line to a dipole (72 ohms). If the frequency of operation is 30 Mc, determine: (*a*) the length of the antenna; (*b*) the characteristic impedance of the quarter-wavelength line; and (*c*) the dimensions of the quarter-wavelength line if the conductors are ½-in.-diameter tubing. (Use Fig. 2.)

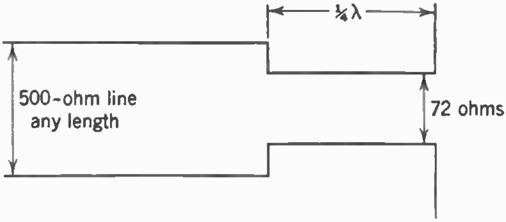


FIG. 21. Use of quarter-wave line as matching transformer between two unequal impedances.

Problem 2. By the use of the charts in Fig. 2 determine the dimensions of three two-wire transmission lines which will have a characteristic impedance of (*a*) 300 ohms; (*b*) 500 ohms.

Problem 3. Determine, by the use of Fig. 2, the dimensions of three coaxial lines which have a characteristic impedance of 50 ohms. Assume that the effect of the spacers is negligible.

Problem 4. A coaxial line has an inner conductor with a diameter of 0.2 in. The dielectric between conductors is polyethylene ($k = 2.25$). What must be the inner diameter of the outer conductor if the characteristic impedance is to be 50 ohms?

18.12. Radiation. A wire or coil carrying a direct current will set up a magnetic field in the vicinity. This field can be detected and explored by means of a compass. Energy is stored in the magnetic field of the coil, and this energy can be put to work in various ways.

Not only can energy be stored in the magnetic field of a coil; it can also be stored in the electric field in a charged condenser. If the ends of two wires connected to the two terminals of the charged condenser are brought near each other a spark will jump across, showing the presence of stored electrical energy.

If a resonant circuit is excited by a voltage of the resonant frequency, energy can be stored in the circuit, at one instant all of it residing in the magnetic field of the coil, and at another instant being shifted to the electric field of the condenser. If the coil and condenser have no resistance, there will be no loss of energy in the circuit and no power will be consumed by it.

Suppose that a resonant circuit is excited by a voltage of the proper frequency and then the voltage source is removed. If the circuit has no resistance, the energy will continue to shift back and forth from the magnetic field of the coil to the electric field of the condenser. Under these idealized conditions, the oscillations in the circuit will continue indefinitely. If, however, there is a small resistance in the circuit, the oscillations will gradually die out. In either the resistanceless case or the case with a small resistance, the effect is as though the electric field, in collapsing, induced a magnetic field which, in turn, induced an electric field when it collapsed. This is in reality a more general form of Faraday's law (Sect. 4.14). The general statement of the law follows: Whenever a magnetic field (or current) *changes*, an electric field (or voltage) is produced. Similarly, a changing electric field produces a magnetic field.

It was stated above that a resistanceless resonant circuit would continue to oscillate indefinitely once it had been excited. This statement was not wholly true. Not all the energy in the magnetic field of the coil, for example, is used in establishing an electric field in the condenser. A very small part of it leaves the vicinity of the circuit and does not return. This constitutes a "radiated" magnetic field, and, since the strength of this field is changing, there is an associated electric field. The fields that are so freed from the circuit constitute a radiated electromagnetic wave. This radiated wave carries some of the energy that was originally in the circuit, and so, even in the resistanceless case, the oscillations in the circuit gradually die down.

If the physical dimensions of the resonant circuit elements are small in comparison with the wavelength corresponding to their oscillatory frequency, the amount of electromagnetic energy radiated into space and lost to the circuit will also be small. On the other hand, if the physical dimensions of the circuit elements are not small in comparison with the wavelength, the radiated energy may be a significant portion of the total energy.

An elementary concept of how radiation from a circuit occurs may be built up by considering the simple half-wave dipole in Fig. 22 which is driven at its center by an a-c generator. Suppose that the generator voltage is zero but changing so that the upper terminal A will soon become positive. As terminal A becomes slightly positive, electrons will be drawn from the upper section AB of the dipole and others will flow into the lower section CD. A small section of the upper part of the dipole will be positively charged, and a small section of the lower part of the dipole will be negatively charged. There will be an electric field between these charged sections. This electric field is shown as $\mathcal{E}$ in Fig. 22(*a*). The charges on these sections are circled so that they may be identified in the following discussion.

As the generator voltage continues to increase, more and more electrons flow out of the upper and into the lower part of the dipole. The upper part becomes more positive and the lower section more negative. As the voltage becomes greater more charges

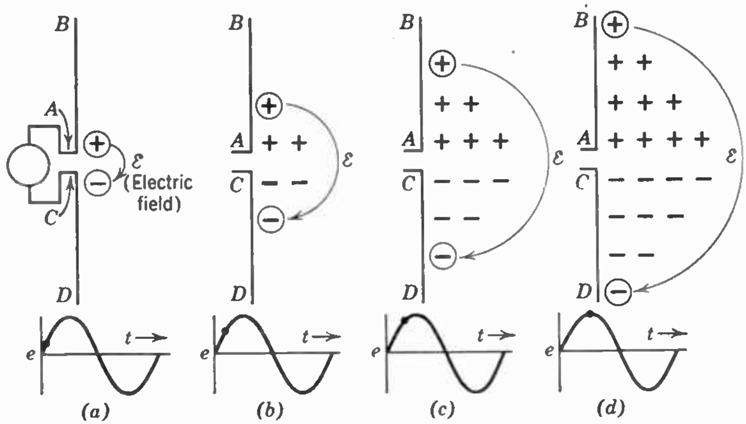


FIG. 22. Charges move out along a half-wave dipole antenna as the voltage rises.

move into the antenna in a given period of time. The initial (circled) charges move farther out along the antenna as shown in the successive pictures in (*b*), (*c*), and (*d*). The electric field $\mathcal{E}$ of the charges continues to expand and move away from the antenna as shown. If the dipole is one-half wavelength long * (one-fourth wavelength per section), the circled groups of charges which started moving away from the generator when the voltage started to go positive will have reached the ends of the antenna when the generator voltage is at its positive peak value, since then the generator has completed one-fourth of a cycle. This is the basic meaning of a quarter wavelength as measured along any conductor.

* A half-wave dipole is considered since this gives the simplest picture. The general behavior of a dipole of any length would be similar, but more complex.

The "charges" travel outward along the upper and lower sections of the dipole with a speed approaching that of light (186,000 miles per second). This does not mean that electrons move this fast—they do not. However, the *effect* is the same as though their charge moved this fast. This is similar to what happens when the engineer takes up the slack on a long freight train. The engine moves slowly, as do the cars. The last car need not move at all. Yet the slack is taken up very rapidly.

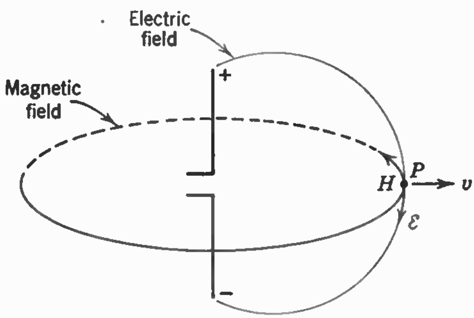


FIG. 23. Direction of electric and magnetic fields around a dipole antenna.

As the generator voltage begins to decrease from its peak positive value, the charges begin to move away from the ends of the antenna and back towards the generator and by the time the voltage is again zero all the charges have returned to the generator; no charge remains on the dipole. During the negative half cycle of voltage the process is repeated except that the lower half of the dipole now becomes positive and the upper half becomes negative.

During the time the charges are moving out along the antenna a current flows, for current is a movement of charge. The current is never high near the ends of the antenna because only a small amount of charge ever gets out this far. Near the center, however, the current may be fairly high.

The current flowing in the antenna sets up a magnetic field H. This field is directed in a circle around the antenna as shown in Fig. 23, and is at right angles to the electric field $\mathcal{E}$. Thus, both an electric field and a magnetic field (an **electromagnetic field**) are produced. This is in agreement with Faraday's law.

Consider now the fields at a point P (Fig. 23) at some distance from the antenna, and suppose that these are the fields which started moving away from the antenna when the voltage started on its positive half cycle. These fields continue to move away from the antenna until the generator voltage starts to decrease. Then the fields start moving back towards the antenna; they begin to "collapse." It would seem that all the fields (and the energy associated with them) would have time to move back to the antenna before the generator reverses the direction of voltage. However, some energy is unable to get back to the antenna before the voltage has changed. The reversal of the voltage causes new fields to move out from the antenna, this time with the directions of both the electric and magnetic fields reversed. The energy that failed to return to the antenna before the voltage reversed moves away from the antenna. It is "free" or radiated energy.*

It appears that the fields which had time to move out to P would also have time enough to return to the antenna by the time the voltage reverses. This apparent contradiction may be explained by considering in more detail just what happens during the positive half cycle. When the voltage starts going positive the fields start moving away from the antenna. The fields produced by the charges which move into the antenna at a slightly later time in the cycle act to "push" these initial fields away with a little added velocity. (The velocity varies from instant to instant during the cycle.) The fields set up at each instant are similarly given an added boost by the fields set up the next instant. When the initial fields have traveled out to a point such as P, the generator voltage starts to decrease. The electrical information regarding this voltage decrease requires a short time to be conveyed to P. Because of the "push" given these fields as they left the vicinity of the antenna, and because of the time required for the effect of the change of voltage at the antenna to be felt at P, these fields at P are too far from the antenna to completely collapse upon the antenna before the voltage

* These elementary concepts may appear to be artificial and inaccurate to the careful reader. The mechanism of radiation is too complex to explain completely with a simple picture.

reverses. The portion of the fields that does not collapse back upon the antenna constitutes the radiated wave.

The radiated wave is made up of both an electric and a magnetic field. The two fields are in *phase;* that is, they reach their maximum intensities and their zero intensities at the same time. The fields are perpendicular to each other in space.* With the dipole in Fig. 22, the electric field is vertical and the magnetic field is horizontal. The direction of the motion of the fields (and their energy) is at right angles to the fields and away from the antenna.

The **polarization** of an electromagnetic wave refers to the direction of the fields. The electric field is usually chosen as the reference. The wave from the dipole discussed above is **vertically polarized.**

An antenna is designed for the specific purpose of radiating an electromagnetic wave into space. Antennas, then, must have dimensions which are comparable to the wavelength of the energy being radiated. It was for this reason that nearly all the antennas discussed earlier in this chapter were a quarter wavelength long or longer. Only by making the antennas of such a length can they be made to be efficient radiators.

The electromagnetic field surrounding an antenna or any other radiating element may be broken up into two components: the **induction field** and the **radiation field.** The induction field is the portion of the entire electromagnetic field which is not radiated. The radiated field is the portion of the total electromagnetic field which continues traveling on through space and does not return to the radiating element.

The induction field is much stronger than the radiation field when the point of measurement is close to the antenna. The strength of this field varies inversely with the **square** of the distance from the source (the antenna, in this case) and does not depend upon the frequency.

At a few wavelengths away from the source the radiation field becomes much larger than the induction field because the strength of this component of the total field is inversely proportional to the **first** power of the distance from the source. It is also

* This statement is not entirely accurate except at points several wavelengths away from the antenna.

directly proportional to the frequency. The manner in which the strengths of these fields vary with distance from the source is shown in Fig. 24.

Since the strength of the radiated field goes up directly with frequency, everything else being equal, it follows that much less power is required to deliver a high-frequency signal of given strength to a distant receiver than would be required for a lower-frequency signal.

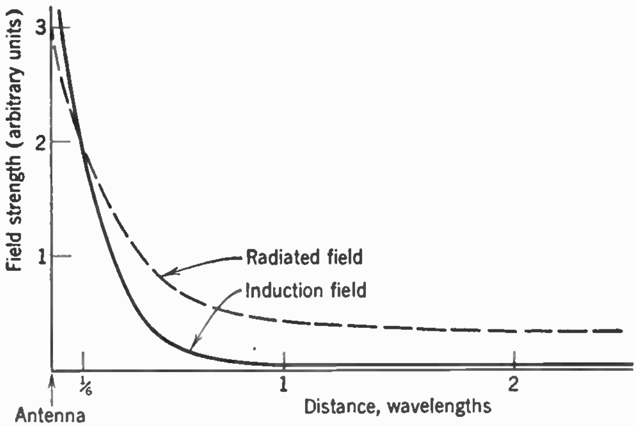


Fig. 24. Relative strengths of radiation and induction fields vs. distance from antenna.

The strength of the induction and radiation fields are equal at a distance of approximately ⅙ wavelength from the antenna. At a distance of 16 wavelengths from the antenna the induction field is only 1 per cent of the radiation field. For this reason field-intensity measurements should be carried out at distances at least 16 wavelengths from the antenna to minimize errors introduced by the induction field.

18.13. Ground wave. Energy is radiated from an antenna in all directions, along the earth (in a horizontal plane) and toward the sky (in a vertical plane). The radiation in a horizontal direction tends to follow the earth much as electrical currents follow a copper conductor, for the earth is a conductor. The **ground wave,*** as the energy radiated along the earth is called,

* The ground wave can be further separated into two components, the **direct** wave and the **ground-reflected** wave. These are shown in Fig. 25.

produces electric and magnetic fields in the earth and other objects. These fields cause currents to flow in the earth and these other objects. Because these objects have some resistance, the currents produced in them by the ground wave cause heating, and the energy used to heat these resistances is lost to the ground wave. The ground wave, therefore, does not extend very far from the transmitter.

The ground wave from a particular station at a particular point is steady. It is the useful part of the total radiation from a broadcast station. It extends out 50 to 100 miles from the antenna, the distance depending upon the power of the transmitter, the type and height of the antenna, the resistance of the earth, the frequency, and the nature of the terrain.

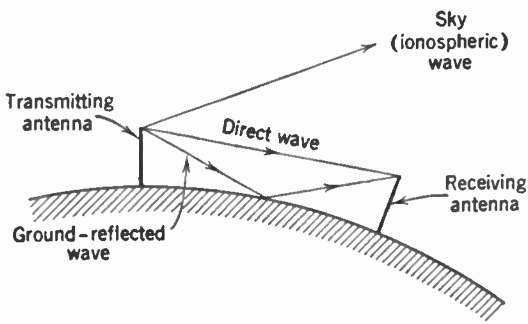


FIG. 25. Various waves radiated from an antenna. Only one "ray" of the sky and ground-reflected waves is shown.

18.14. Sky (ionospheric) wave. Energy radiated into the sky would shoot off into space and be lost except for a natural phenomenon which was discovered through radio. Up in the sky, from about 50 to 250 miles above the earth, are several layers of ionized particles apparently produced from the air by ultraviolet radiation from the sun. These layers, acting as reflectors and refractors of radiation, reflect or bend downward some of the energy of the sky wave. Other parts of the total energy may be merely deflected. When, therefore, radio waves shoot off into the upper atmosphere and encounter these ionized particles, part of the energy is reflected back toward the earth, part is absorbed by the ionized layers, and part penetrates the ionized layers to escape into outer space. The region above the earth in which these ionized layers are effective in causing bending of the radiated wave is called the **ionosphere.** For this reason, the sky wave is often called the **ionospheric wave.**

18.15. The ionosphere. Several ionized layers have been observed at various levels in the ionosphere. The term "layers" is somewhat inaccurate since there is some ionization between the layers. The location of the layers is specified by the regions of densest ionization. The most permanent of these layers are the E and F. The region in which the E layer exists is between about 50 and 90 miles above the earth, and the region in which the F layer exists extends from about 90 to 250 miles above the earth. During the daytime, in a particular locality, two layers exist in the F region. The lower is called the F_1 layer and the upper the F_2 layer. At night these two merge to form the F layer.

Other layers have been observed from time to time. They fade in and out in an irregular fashion. Although these layers may have an important effect upon radio communication while they are present, their average effect, day by day, is not as important as that of the E and F layers.

The mechanism by which radio waves are returned to the earth is rather complicated. It involves the action of both the positively ionized particles and, more importantly, of the negatively charged electrons in the various ionized layers. Other factors such as the frequency of the radio wave, the earth's magnetic field, sunspot activity, solar radiation, and the latitude are also involved.

An elementary picture of how a radio wave may be "reflected" back to the earth may be gained by reference to Fig. 26. A wave that leaves the transmitter antenna at a high angle is bent slightly as shown, but it penetrates the ionized layer and continues on into outer space. Waves that are radiated at a somewhat lower angle may be bent enough that they are returned to the earth. The highest angle at which the waves are returned to the earth is called the **critical angle.** Waves that are radiated at a still lower angle may also be returned to the earth, but at a greater distance from the transmitter. Waves that are radiated at an even smaller angle may be reflected, but at such an angle that they miss the earth. Some energy is lost in the ionized layer, and some penetrates into outer space whatever the angle of radiation. Therefore the wave which is returned to the earth contains less energy than would be calculated taking the length of the path alone into consideration.

The amount of bending suffered by a radiated wave leaving the earth at a given angle depends both on the density of ionization of the ionized layer and on the frequency of the wave. As the

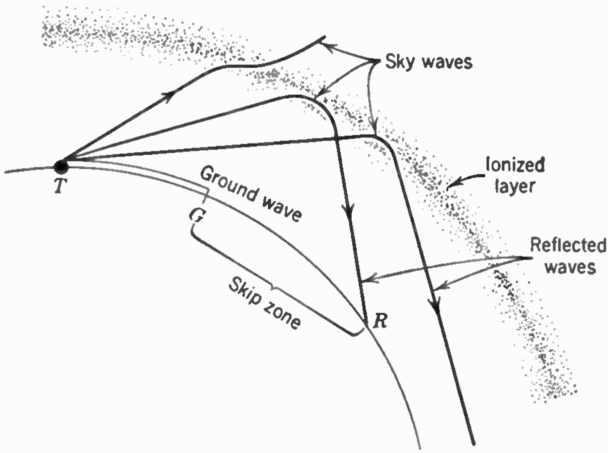


Fig. 26. Radiation is bent downward to earth by ionosphere so that it may be detected at great distances from its source. Between the outer limits of the ground wave and the reflected or refracted radiation, no signals can be heard.

density of the ionized particles increases, the amount of bending increases. High-frequency waves are not bent as much as low-frequency waves. Since the upper layers tend to be more highly ionized, it is possible for a wave to penetrate the E layer and then be reflected by the F layer.

The highest frequency useful for radio transmission between two points on the earth at a specified time, employing waves reflected from the ionized layers, is called the **maximum usable frequency** (MUF). The MUF varies from hour to hour, and from daytime to nighttime. It is different for various times of the year and for different degrees of sunspot activity. A great deal of time has been expended in research in an effort to correlate all these factors which affect the propagation of radio waves.

In the daytime, on broadcast frequencies (550 to 1600 kc), the sky wave is normally so completely absorbed by the E layer that it is of little use in communication, and only the ground wave is available. At nighttime, beginning just before sunset, the amount of absorption of the sky wave decreases rapidly, and a considerable portion of the sky wave may be reflected to the earth. The nighttime absorption is less in the winter than in the summer, and broadcast stations may be heard at great distances during winter evenings via the sky wave.

The losses suffered by the ground-reflected wave increase with frequency to an extent that above about 2 Mc the ground wave is not usable for communication except within a few miles of the transmitter. The direct wave can be used for communication up to line-of-sight distances (Sect. 18.17). Only the sky wave is useful for communication for distances greater than about 50 miles. A general rule is that the highest frequency that can be used (a frequency nearly equal to the maximum usable frequency) will normally give the best signal strength at the receiver. Since the MUF is different in the daytime than at night (it is usually less at night than in the daytime), two frequencies, sometimes more, are employed in commercial installations which must maintain dependable communication between two widely separated points. Another factor that must be taken into consideration is whether the entire transmission path lies in darkness, or whether part of the path is in daylight. Different frequencies would be used for these two cases.

Ordinarily frequencies above about 30 Mc are not reflected from the ionosphere, and reliable communication at these frequencies must be carried on by means of the direct wave.

18.16. Fading. Suppose that a receiving station is within range of the ground wave of a 1000-kc transmitter, say at the edge of it. Signals come through in the daytime clear and steady. At night, however, when reflections occur, the receiver begins to pick up the reflected sky wave from the 1000-kc transmitter as well as the ground wave. The two waves may be in phase and so reinforce each other, or they may be out of phase and tend to nullify each other. If the ionized layer shifts in height or density, the sky wave will vary in intensity so that the alternate additions and subtractions to and from the ground wave by the sky wave cause the voltages at the receiving antenna to vary from time to time. This variation is known as **fading** and may be very severe.

The a-v-c system of a receiver does not help nighttime reception from a station 50 to 100 miles away from which both sky and ground waves are received. Signals from these stations, received clearly in the daytime, may be badly garbled at night because of the variations in magnitude and phase of the sky wave.

On the broadcast band, the sky wave in the daytime is absorbed so that it has little effect. Receivers 50 to 100 miles away from the average transmitter, therefore, get clear signals. Receivers farther away will be unable to detect the transmitted signal because the ground wave is too weak at these frequencies. At night, the sky waves from distant broadcast stations are reflected downward to the receiving antennas and long-distance reception is possible. If the ionized layers are comparatively stable, as during times of weak sunspot activity, reception from a distant station can be fairly stable.

Sometimes the sidebands carrying the modulation fade more or less than the carrier of the station, resulting in **selective fading.** Neither an a-v-c system nor an increase of transmitter power is of help in overcoming this trouble.

Another type of fading may occur when two ionospheric waves arrive at the receiver along slightly different paths. The lengths of the transmission paths will be different and also the phase angles of the arriving signals. The two paths may change from time to time and independently of each other. As a result the two signals may at times be in phase and so reinforce each other. At other times they may be out of phase and tend to cancel each other.

Still another type of fading may occur with ultra-high-frequency signals. The dielectric constant of the earth's atmosphere varies with height above the earth, and the received signal depends to some extent upon this dielectric constant. The atmosphere varies from time to time and thus may cause a variation in the received signal strength. The fading from this cause is usually not too severe and occurs at a rather slow rate. Local weather conditions may also cause fading at these frequencies.

18.17. Short-wave transmission. In the broadcast band the ground wave is quite usable; the sky wave is undependable. Above about 1500 kc the attenuation (reduction) of the ground wave increases rapidly because of increased ground losses, but at the same time the effectiveness of the sky wave increases. For this reason short-wave communication takes place mainly by means of reflected waves. Antennas are designed so that their radiation is aimed up into the sky at the correct angle to be re-

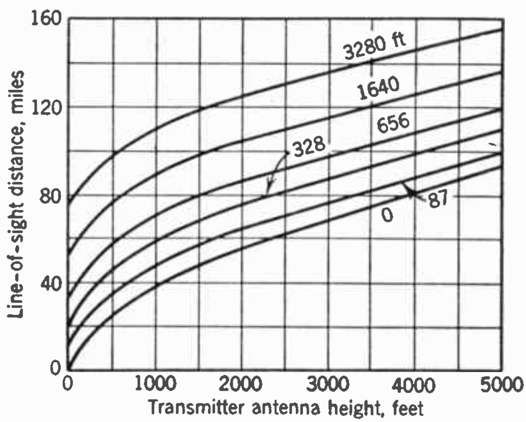


Fig. 27. Length of line-of-sight paths for receiving and transmitting antennas of various heights.

flected downward to the desired receiving station. Frequencies greater than about 30 Mc are not so easily reflected, tending to shoot on through the ionized layers; the energy that is reflected is reflected at such an angle that it does not return to the earth except under unusual circumstances. Communication at these frequencies depends upon a wave received in a direct line from the transmitter (like light from a lighthouse), and, at frequencies of the order of 100 Mc and greater, very little usable energy is radiated beyond the distance the transmitter antenna can be seen (line-of-sight transmission).

On the very high frequencies, therefore, it becomes necessary to erect the transmitting and receiving antennas as high as possible above the surface of the earth since radiated waves at these frequencies act much like light waves. The transmitter resembles a lighthouse, sending its radiation out from the antenna at various angles, to be picked up as far as the antenna can be seen. Actually, effective communication can be carried out at greater distances than line-of-sight would indicate, sometimes two or more times this distance. Figure 27 shows line-of-sight distances from transmitting to receiving antennas at various heights and over average terrain.

19 · Frequency Modulation and Detection

Previous chapters have discussed the various features of amplitude-modulated systems. In such systems the *strength* of the carrier wave at any instant is a function of the *strength* of the audio modulation at that instant. The frequency of the carrier remains constant. In a f-m system, on the other hand, the *strength* of the carrier is constant regardless of the strength of the modulation. The *frequency* of the carrier, however, is a function of the modulation *strength.*

In an a-m system, the rate at which the carrier strength varies is a function of the frequency of the modulation. Thus, if the modulation is a fixed tone of, say, 1000 cycles, the carrier strength goes up and down at a 1000-cycle rate. However, in the f-m system, the carrier-frequency *change* is greatest for a loud modulating signal; the rate at which the frequency varies depends upon the frequency of the modulating signal. Thus, if the modulation is a 1000-cycle tone, the carrier frequency will vary at a 1000-cycle rate.

In an a-m system, modulation causes a change in the *strength* of the carrier; in a f-m system, modulation causes a change in the *frequency* of the carrier. In either system, the rate at which the strength (a-m) or frequency (f-m) changes is a function of the frequency of the modulation; the degree or amount of the change depends upon the strength of the modulation.

19.1. F-m vs. a-m system. A f-m system has several advantages when compared to an a-m system. Among these advantages are (1) a gain in effective signal strength because the transmitter can radiate at full power at all times regardless of the strength of the audio modulation; (2) a reduction in noise with a resultant effective increase in signal-to-noise ratio, the extent of this gain being dependent upon the width of the frequency band that is employed; (3) a reduction in interference from other stations on the same channel by a factor of 30 to 1 compared to amplitude modulation.

Frequency-modulation systems not only have advantages when compared to a-m systems; they also have some disadvantages. The most important are (1) a much wider frequency band is required for transmission, perhaps 7 to 15 times greater than for an a-m system; (2) transmitting and receiving equipment tends to be somewhat complicated; and (3) *because of the high-frequency band* used by most f-m stations to obtain the necessary band width, reliable reception is limited to slightly more than line-of-sight distances.

The comparisons between f-m and a-m systems as outlined above can be discussed in more detail.

1. Constant transmitter power. In a-m transmission, the strength of the signal at the detector output terminals in a receiver is dependent upon the power in the sidebands of the transmitted wave. This power varies from instant to instant, depending upon the strength of the signal picked up by the microphone. The power in the sidebands is added directly to the power in the unmodulated carrier. The average modulation of the average broadcast station is, perhaps, 50 per cent. With this degree of modulation there is a 6 per cent increase in transmitter antenna current compared to the unmodulated condition. This increase in current results from the production of sideband power. If the transmitter could be 100 per cent modulated all the time, there would be a 22.5 per cent increase in antenna current and the detector voltage output in the receiver would be more than three times greater than for "average" modulation. At the same time the total power output of the transmitter would almost double.

There is another fact, important from the standpoint of the transmitter owner. A 50-kw a-m transmitter must have the ability to handle without distortion as much power as 200 kw during the occasional peaks of 100 per cent modulation because, at full modulation, the peak plate current flowing in the final power amplifier will be doubled, and the power, being proportional to the square of the current, will have increased four times. It is true that this is peak power and is radiated for very short intervals of time. Nevertheless, a 50-kw station must be designed to handle this 200 kw during periods of peak modulation.

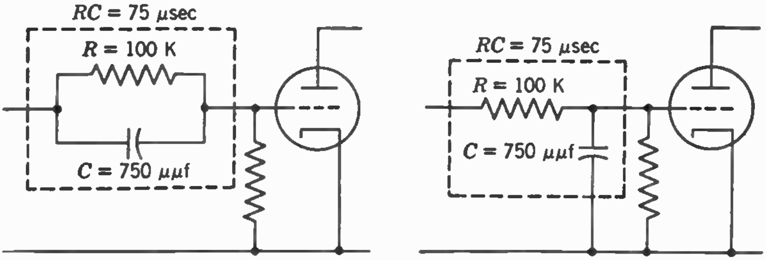
On the other hand, a f-m station is modulated by changing the frequency and not the amplitude of the antenna current. The total transmitted power does not increase with higher degrees of modulation. Rather, modulation causes a shift of power from the carrier to the sidebands so as to maintain the total power essentially constant. At some instants the power transmitted at the center carrier frequency may be zero. A f-m transmitter can transmit its full power at all times with no design allowance being required to take care of modulation peaks.

2. Greater freedom from interference. In an a-m system, interfering signals of only one-hundredth of the strength of a desired signal produce interference of 1 per cent. In a f-m system, the interference is down to about 1 per cent when the interfering signal is one-third the strength of the desired signal. Thus, much stronger signals from interfering stations can be tolerated in a f-m system than in an a-m system. This means that stations on the same f-m channels can be put much closer together geographically than would be feasible with a-m stations.

3. Reduction in noise. In an a-m system, the amount of noise that is permitted to enter the system is a function of the band width. In a narrow-band a-m system, such as a radio-telephone circuit utilizing frequencies from 250 to 2500 cycles, much less noise both from natural and from man-made static will enter the system than in a high-fidelity system passing all frequencies from 30 to 15,000 cycles. In an a-m system the only effective way to increase the signal-to-noise ratio is to increase the power of the signal.

The band width of a f-m system is roughly seven times that of even a high-fidelity a-m system, and one would think that much more noise would be admitted to such a system. However, it is characteristic of nearly all noise that it is mostly amplitude-modulated and that very little noise is frequency-modulated. Moreover, the degree of modulation of frequency-modulated noise is rather small. Frequency-modulation receivers are equipped with amplitude-limiting circuits (limiters) which remove all amplitude modulation before the signal is passed onto the detector. Therefore the receiver is insensitive to this type of noise. The signal radiated by the transmitter is intentionally heavily modulated; that is, the audio signal is made to cause large changes in the carrier frequency. In this manner, the effect of frequency-modulated noise in causing trouble in the receiver is made almost negligible.

Frequency-modulation transmitters normally employ **pre-emphasis** circuits designed to provide greater amplification of the high audio frequencies than of the lows. They are applied in the circuit ahead of the stage in which frequency modulation occurs. The high-frequency signals represented in the radiated wave are then stronger than they would have otherwise been. These high-frequency audio components are passed through a **de-emphasis**



(*a*) Pre-emphasis (*b*) De-emphasis

Fig. 1. Typical pre-emphasis and de-emphasis networks.

circuit in the receiver to return them to their proper relation to the other components of the composite audio signal. This process has the advantage of increasing the signal-to-noise ratio of the higher audio frequencies. Since most of the power in an audio signal is contained in the lower-frequency components, the total audio power is barely increased. On the other hand, the amount of reduction of noise is great enough to be of importance to the listener. This same system could be applied to a-m broadcast systems, but a change-over would require that all receivers be adapted to the new system. Typical pre-emphasis and de-emphasis circuits are shown in Fig. 1.

19.2. Wide- and narrow-band f-m systems. Frequency-modulation systems require considerably wider band widths than do a-m systems because many sidebands, rather than two as in a-m systems, are required to carry the modulated wave. However, f-m and a-m systems are alike in that wider band widths are required if high-frequency audio signals are to be transmitted without distortion.

In f-m systems a wider band width not only permits a greater range of a-f tones to be transmitted; it also effects an increase in signal-to-noise ratio. Frequency-modulation transmission could take place on the standard broadcast band, but since the band width required for each station is so great there would be room for only seven or eight stations. For frequency modulation it is necessary, therefore, to go to the short-wave regions, where band widths can be relatively much wider.

Frequency-modulation broadcast stations are now assigned to 1 of 100 channels, each 200 kc wide, in the band from 88 to 108 Mc. Each station actually uses a band width only 150 kc wide, 75 kc on each side of the carrier frequency; and the remainder of the separation between stations is a guard band to reduce interference between stations on adjacent channels. These stations utilizing a band width of 150 kc employ **wide-band** f-m systems. They are high-fidelity stations capable of transmitting a-f tones up to 15,000 cycles with little distortion and almost no noise.

Frequency-modulation stations designed for general communications use, such as, for example, police radio, commonly employ **narrow-band** systems and utilize a band width of around 20 kc. Such systems do not, of course, pass the higher a-f tones. The receiver output of narrow-band systems is, however, perfectly intelligible. The noise is somewhat higher than in wide-band systems for strong signals. However, on weak ("threshold") signals a narrow-band receiver will admit less noise. For this reason, the service range of narrow-band stations is somewhat greater than that of wide-band stations.

On the frequency bands employed by practically all f-m stations there is less natural static than on the 550- to 1600-kc band allotted to a-m broadcasting; the range of each station is more or less limited to two or three times line-of-sight distance; little interference is created by stations on the same channel provided that they are a few hundred miles apart; there is no fading; the day and night range is the same; and, on f-m broadcast stations, the band width is sufficient so that all audio frequencies between 30 and 15,000 cycles can be transmitted, with the result that broadcast listeners have the chance to hear music in the home as it is produced at the transmitter and not restricted in audio range by the limitations of the 10-kc separation necessary in the standard a-m broadcast band.

19.3. Production of f-m waves. A general requirement for systems which produce frequency modulation is that the audio signal from the microphone must vary the carrier frequency linearly in accordance with the strength of the audio signal. A simple system, impractical for technical reasons, will be considered first.

Suppose that a condenser microphone is connected across the tank circuit of a simple oscillator. As speech waves strike the microphone, the diaphragm moves and changes the capacitance of the microphone. This change of capacitance also produces a change in the oscillator frequency, and so the oscillator output is frequency-modulated. The *amount* of the frequency change depends upon the loudness of the sound striking the microphone. The *rate* at which the frequency is changed, that is, the number of times per second it is shifted, depends upon the frequency of the sound that is picked up by the microphone.

The amplitude of a f-m wave need not change at all during the modulation process. If the microphone in a particular transmitter picks up a 1000-cycle tone which, when amplified, has a peak voltage of 1 volt, the carrier frequency may be shifted, for example, from 90 Mc to 90 Mc plus or minus 2 kc, and this frequency will shift back and forth from the average, unmodulated value of 90 Mc to the new value 1000 times per second. If the output voltage of the microphone and amplifier changes to 2 volts, the carrier will shift from 90 Mc to 90 Mc plus or minus 4 kc, but the amplitude of the carrier will be the same as before. If the microphone picks up a 2000-cycle tone which, when amplified, has a strength of 2 volts, the carrier again shifts from 90 Mc to plus or minus 4 kc, but the shift occurs 2000 times per second instead of at the former rate of 1000 cycles per second. The general appearance of a frequency-modulated wave is shown in Fig. 2. Compare with that of the amplitude-modulated wave of Fig. 1, Chapter 14.

Major Edwin H. Armstrong must be credited with the present interest in and appreciation of the benefits of wide-band f-m transmission. In spite of the greatest skepticism of those well established in a-m broadcasting, he persisted in his belief that frequency modulation had great advantages.

19.4. Reactance tube modulator. A simple method of producing frequency modulation is to vary the output frequency

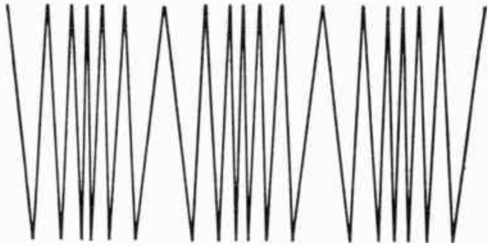


Fig. 2. Constant-amplitude frequency-modulated wave.

of an oscillator. How this could be done with a condenser microphone was described above, but the method was impractical for various technical reasons. However, the oscillator frequency is varied in a similar manner in one widely used system for producing f-m signals. The system uses a **reactance tube,** a most ingenious application of electron-tube principles.

Consider the circuit in Fig. 3, in which two impedances, Z_1 and Z_2, are connected from plate to cathode of a tube. The grid is

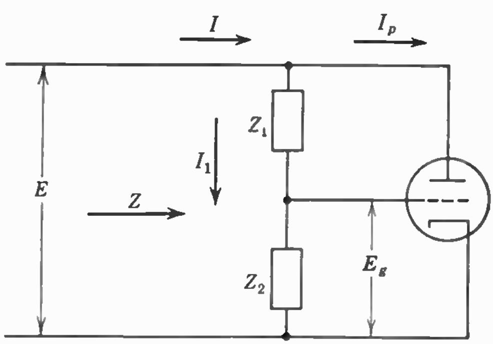


Fig. 3. Simple reactance tube.

connected to the common junction of the two impedances. Whatever the source of the grid voltage E_g the a-c plate current I_p will be approximately

$$I_p = g_m E_g$$

if the tube is a pentode or other multigrid tube with a high dynamic plate resistance.

The impedance Z looking into the circuit is

$$Z = \frac{E}{I} = \frac{E}{I_1 + I_p} = \frac{E}{I_1 + g_m E_g}$$

But $E_g = I_1 Z_2$. Therefore,

$$Z = \frac{E}{I_1 + g_m I_1 Z_2} = \frac{E}{I_1(1 + g_m Z_2)}$$

If $g_m Z_2$ is made much larger than unity, as can easily be done by choosing a tube with a g_m around 3000 micromhos and choosing Z_2 greater than about 5000 ohms at the operating frequency,

$$Z = \frac{E}{I_1 g_m Z_2} \quad \text{(approx.)}$$

The total voltage across the two impedances is equal to the current through them times their total impedance, or

$$E = I_1(Z_1 + Z_2)$$

If Z_1 is made much larger than Z_2, then

$$E = I_1 Z_1 \quad \text{(approx.)}$$

and

$$Z = \frac{I_1 Z_1}{I_1 g_m Z_2} = \frac{Z_1}{g_m Z_2} \quad \text{(approx.)}$$

Now suppose that Z_1 is a condenser of reactance $1/(2\pi fC)$ and Z_2 is a resistance R. Then

$$Z = \frac{1/(2\pi fC)}{g_m R} = \frac{1}{2\pi fC \times g_m R} = \frac{1}{2\pi f(g_m RC)}$$

Hence, the impedance Z is equal to the impedance of a capacitance of $g_m RC$ farads, and the tube and its associated circuit provide an input impedance which is essentially capacitive in nature. Suppose the g_m of the tube is varied in some manner. Then the effective value of the input capacitance is also varied. This, then, is a circuit which has the characteristics of a variable capacitance. The g_m of the tube can be varied simply by inserting a variable voltage in series with Z_2, or by varying the voltage on the screen or other grid in a multigrid tube. The second method is generally preferred since it serves to isolate the r-f and a-f voltages. Thus the capacitance of the circuit can be made to vary with the amplitude of the audio modulation and at a rate which depends upon the frequency of the modulation. This circuit can be connected across the tank circuit of an

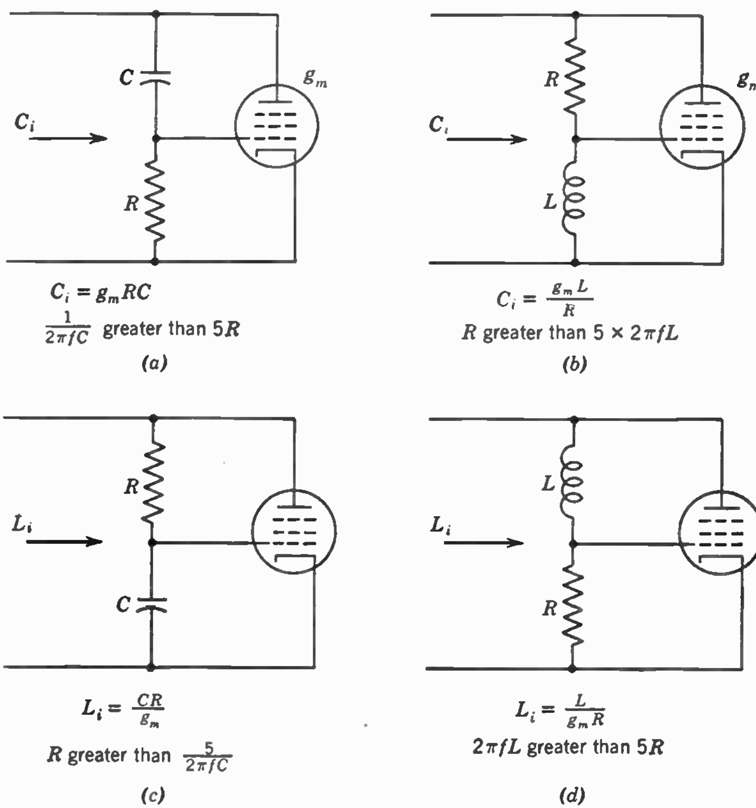


FIG. 4. Reactance tubes which provide variable capacitance or inductance.

oscillator to cause the frequency to vary in accordance with a-f modulating voltages.

The choice of elements given for Z_1 and Z_2 is only one of several which will produce a variable reactance at the input terminals of the circuit. It is possible to produce the effect of either a variable capacitance or inductance. Four typical arrangements are shown in Fig. 4 along with design equations. See also Fig. 5.

If the frequency of the oscillator is increased in a series of doublers and triplers, as it will be in the usual f-m transmitter,

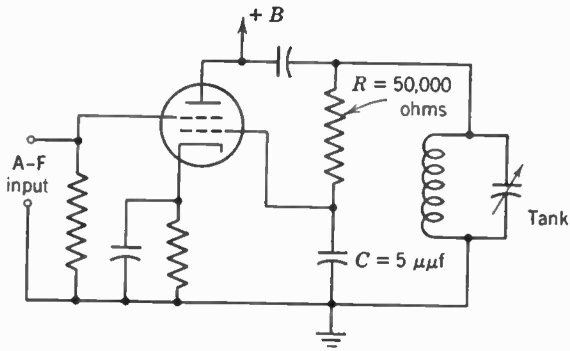


FIG. 5. A workable reactance-tube modulator which affects the resonant frequency of the tank as though a variable inductance were shunted across the tank.

the deviation in frequency due to modulation is multiplied along with the center carrier frequency.

Example 1. A f-m transmitter is to operate on 90.1 Mc. The maximum deviation is to be 75 kc. The oscillator center frequency is 7.508 Mc. How much is this frequency to be shifted to obtain the desired 75-kc deviation in the output?

Solution: At 90.1 Mc the frequency change is 75,000 in 90,100,000. At 7.508 Mc the shift will be one-twelfth (90.1 ÷ 7.508) as much, or 6250 cycles (75,000 ÷ 12).

19.5. F-m terminology. Several new terms will be found in f-m literature. The **frequency deviation** is the maximum change of the instantaneous frequency from the average (unmodulated) carrier frequency. Thus a 100-Mc carrier shifted to plus or minus 75 kc on modulation peaks has a frequency deviation of 75 kc. The **deviation ratio** is the frequency deviation divided by the highest audio frequency to be transmitted. The greater this ratio, the greater will be the signal-to-noise ratio at the receiver (except on weak signals as discussed above). A broadcast transmitter having a deviation ratio of 5 to 1 (75 kc ÷ 15 kc) will present the same effectiveness against noise as a police transmitter in which the highest audio tone is 3000 cycles and which is deviated a maximum of 15,000 cycles each side of the average carrier frequency.

In a f-m system the **modulation index** is the ratio of the frequency deviation to the actual modulating (audio) frequency.

It is proportional to the ratio of the amplitude of the modulating signal to its frequency. In contrast, in a p-m (phase-modulation) system the **phase deviation** is proportional to the amplitude of the modulating signal and is independent of the frequency.

19.6. The Armstrong system. Phase modulation and frequency modulation are closely related. In both types of modulation some portion of the angular term * of the carrier frequency is affected in the modulation process. There is one significant difference between the two types of modulation. The phase deviation in a p-m system depends only upon the amplitude of the a-f modulating signal. However, the modulation index in a f-m system (comparable to the p-m phase deviation) varies directly as the amplitude and inversely as the frequency of the modulating signal. Thus, in a f-m system, a low-frequency signal produces a greater modulation index than does a high-frequency signal, provided the amplitudes of the low- and high-frequency signals are the same.

The nature of the difference between p-m and f-m waves suggests that, if the a-f modulating signal was passed through a network which reduced its amplitude in proportion to the frequency and if the resulting signal was employed in the p-m system, the resulting modulation would actually be frequency modulation. This is the method used in the Armstrong system. The network through which the a-f signal is passed is called a **pre-distorter.**

A block diagram of the essential elements of the Armstrong system is shown in Fig. 6. The output of a crystal-controlled oscillator, usually operating around 200 kc, is divided and fed into two different channels. One channel shifts the phase of the 200-Mc carrier by 90°. The other channel is a balanced modulator which receives both the carrier and the a-f signal (which has been passed through the pre-distorter). Amplitude modulation occurs in this circuit. The output contains only the sidebands produced in the modulation process; the carrier itself is suppressed in the balanced modulator circuit. These sidebands are mixed in a combining amplifier with the carrier which has received a 90° phase shift in the other channel. The result is a

* This angular term is the "$2\pi f_c t + \phi$" of $e = E_{max} \sin (2\pi f_c t + \phi)$. See Sect. 14.1.

p-m wave, but because of the pre-distortion of the a-f signal, it also represents a f-m wave in terms of the original a-f signal.

The frequency deviation which can be produced in this process without distortion is quite small, only a few cycles per second, and the carrier frequency is much lower than that of the f-m broadcast channels (88 to 108 Mc), so the signal is passed through a series of frequency multipliers which increase both the frequency deviation and the carrier frequency to the desired

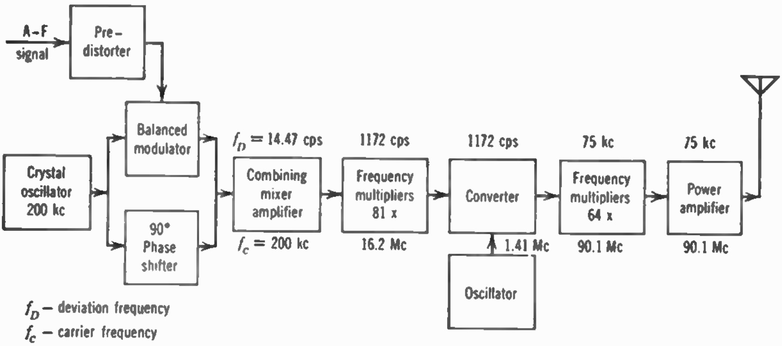


FIG. 6. Block diagram of Armstrong system transmitter.

values. A converter stage is used to reduce the carrier frequency without affecting the frequency deviation. Such an arrangement is required since a much greater increase of the frequency deviation than of the carrier frequency is required. Frequency multiplication ordinarily occurs both before and after the converter stage. Typical frequency values of both the carrier and deviation frequencies are shown at various points on the block diagram.

All the functions described above can be performed in small receiving-type tubes and at low power. Large power tubes are required only in the final power amplifier.

19.7. Frequency stabilization. Since the instantaneous frequency in a f-m system is varied, some means must be provided so that the carrier frequency comes back to its original value when modulation ceases. Otherwise, the average or nonmodulated frequency might drift and a receiver could not locate it at the same place on the tuning dial, and so interference would be created by one station straying into the frequency band of another. Federal regulations also require that the center frequency be maintained within specified limits.

In the reactance tube modulator, a crystal-controlled oscillator cannot be used since the frequency of this oscillator is actually varied during modulation. Therefore, an additional oscillator is usually provided which furnishes a reference frequency that is accurately controlled by a quartz crystal. The frequency of the crystal oscillator is compared to the center frequency of the f-m output of the reactance tube modulator by passing both signals through a mixer that produces a low-frequency beat-note in the output, say 1000 cycles when the modulator frequency has the proper value. This beat-note is passed through a circuit, such as a discriminator, which produces a direct voltage whose strength is dependent upon the beat-note frequency. This direct voltage is utilized as a control bias on the reactance tube. Suppose that the f-m carrier frequency should slowly drift so that the beat-note rose to 2000 cycles. The d-c bias voltage would rise and tend to cause the f-m carrier frequency to return to its proper value. To achieve greater control sensitivity, the frequency comparison circuit is usually applied after the f-m carrier has passed through several stages of frequency multiplication. A more complete discussion of the application of such a-f-c circuits is found in Sect. 15.25. In some frequency-stabilization schemes, the beat-note is used to produce an "error" signal which in turn operates a small motor that makes slight adjustments on the variable capacitor associated with the reactance tube modulator.

Frequency-modulation transmitters employing the Armstrong system do not require additional frequency stabilization since their carrier frequency is derived from a crystal oscillator.

19.8. F-m receivers. The circuit of the ordinary f-m receiver is similar to that of an a-m superheterodyne with two important exceptions: one or more amplifier stages immediately preceding the detector are operated as limiting amplifiers **(limiters)**; and a special kind of detector, called a **discriminator,** is required. Block diagrams of typical f-m and a-m receivers are shown in Fig. 7. The band width of the tuned circuits must be considerably broader than that of an a-m receiver. In addition, it is customary to decrease the higher-frequency audio re-

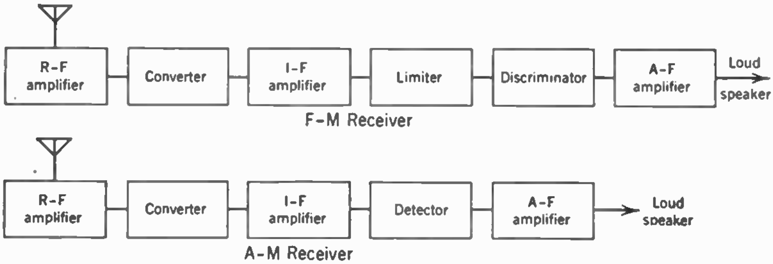
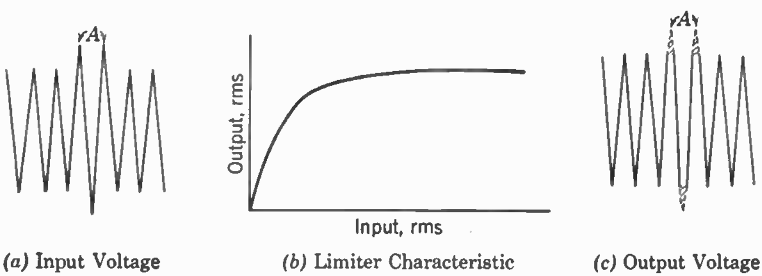


Fig. 7. Comparison of f-m and a-m receivers.

sponse at the same rate at which it was increased at the transmitter. This is done with a simple *RC* **de-emphasis** circuit such as shown in Fig. 1.

19.9. The limiter. All incoming signals to a f-m receiver, often after preliminary amplification, are changed to the same intermediate frequency by the converter stage. The i-f amplifiers must pass a broad frequency band, about 200 kc, and are operated at frequencies between about 3 and 12 Mc. The signals, after passing through the i-f amplifiers, go to the limiter stage. A limiter is a tube circuit which produces a constant output voltage even though the peak amplitude of the input voltage varies. The purpose of the limiter is to remove any amplitude



(a) Input Voltage *(b)* Limiter Characteristic *(c)* Output Voltage

Fig. 8. A limiter eliminates variations in amplitude.

variations that may have been impressed on the f-m signals by any cause. Most limiter circuits are able to perform this function satisfactorily except for very weak signals.

The appearance of an unmodulated f-m signal is shown in Fig. 8(*a*). The individual r-f cycles are equidistant and of constant amplitude except in the region of *A*, where the amplitude varies. This variation in amplitude may be due to noise, static, or some other disturbance not in the original signal. If amplitude variations are permitted to go through the f-m receiver, they will produce audible noise If, however, the signal is passed through a circuit with the characteristic shown in Fig. 8(*b*), the variations of amplitude will not be passed to the detector and noise will be reduced. The output of the limiter will not look like the input—there will be distortion [Fig. 8(*c*)]—but as long as the distance apart of the individual cycles of the incoming

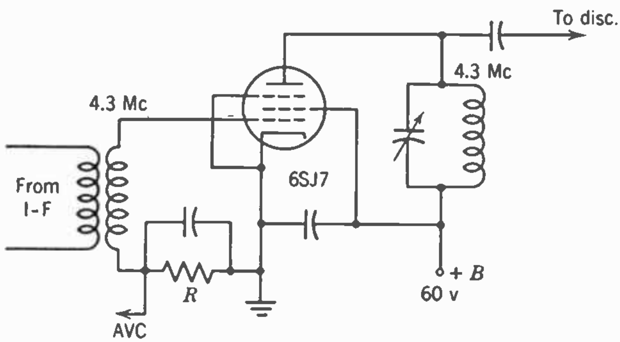


Fig. 9. Circuit of simple limiter using low plate and screen voltages.

f-m wave does not change, the receiver will be quiet. The distortion of the waveform does not result in objectionable noise in a f-m system as it does in an a-m system.

The purpose of the limiter, therefore, is to chop off any amplitude modulation that may exist in the incoming signal or that may be created within the input circuit, the converter stage, or the i-f amplifier.

Limiter action may be secured in the pentode tube circuit of Fig. 9, when operated with low plate and screen voltages. Because of the low supply voltages a weak signal saturates the tube, which then passes about the same plate current for various amplitudes of grid voltage. Therefore, increases in input grid voltage, supplied by the i-f amplifier, produce very little increase in output voltage. Only when the grid signal is very weak does the circuit fail to reach the saturated condition. It is for this reason that the limiter characteristic in Fig. 8 drops for low input voltages.

If the antenna signals are so weak that, after amplification, they do not reach the saturated portion of the limiter characteristic, they will still produce audio tones, but there will be noise in the output since the amplitude variations have not been removed. A f-m receiver, therefore, will work without a limiter, but all the advantages of frequency modulation with respect to low noise will not be realized. A f-m receiver, properly designed, manufactured, and operated, and receiving strong incoming signals, will be almost totally without sound (tube hiss, static, etc.) during moments when the f-m transmitter is not being modulated.

The output of the limiter is a constant-amplitude, variable-frequency signal, the amount the frequency varies from the unmodulated frequency being proportional to the microphone voltages at the transmitter, and the rate at which the frequency varies being proportional to the *frequency* (rate) at which the microphone voltage changes.

19.10. The discriminator. A discriminator in a f-m receiver is the counterpart of the detector in an a-m receiver. It must convert the frequency variations in the f-m wave to audio voltages at the output of the discriminator circuit, the *amplitude* of the audio voltage being proportional to the variation of frequency and the audio *frequency* being proportional to the rate of variation of frequency of the f-m wave.

One simple method (called "slope detection") of extracting the audio signal from a f-m wave is to use a tuned circuit with a characteristic shown in Fig. 10. Suppose that the output of the limiter is impressed on the tuned circuit at a frequency somewhat below the resonant value, such as at point A. When the frequency of the f-m wave increases, the output of the circuit will increase; when the frequency decreases, the output of the circuit also decreases. Thus a variation in frequency to such a circuit produces an audio output signal. This signal may be amplified and fed to the loud-speaker system. Once an audio signal is extracted from a f-m wave, the remainder of the receiver does not differ from an a-m receiver except that the a-f amplifier and speaker should have a wider frequency response and be freer from noise and distortion.

It is difficult to design a simple resonant circuit as shown in Fig. 10 with a linear characteristic suitable for use with wide-band f-m signals. In practice, a more complex discriminator

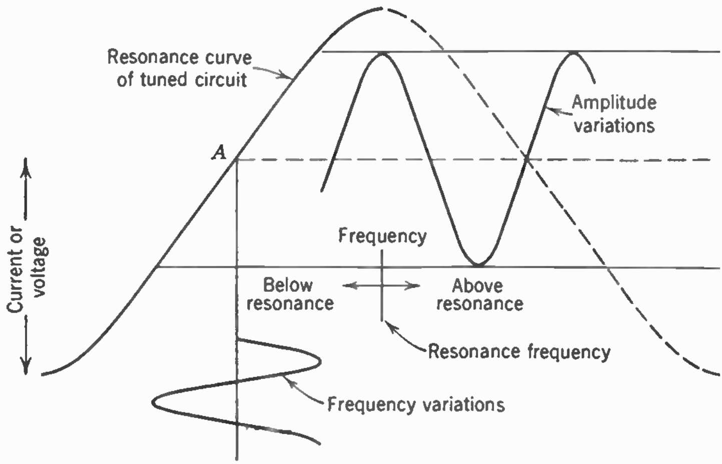


Fig. 10. Use of simple resonant circuit to convert frequency variations into amplitude variations.

such as the one shown in Fig. 11 is employed. The signals from the limiter pass through an i-f transformer with a split secondary. Each half of the secondary winding is connected to a diode recti-

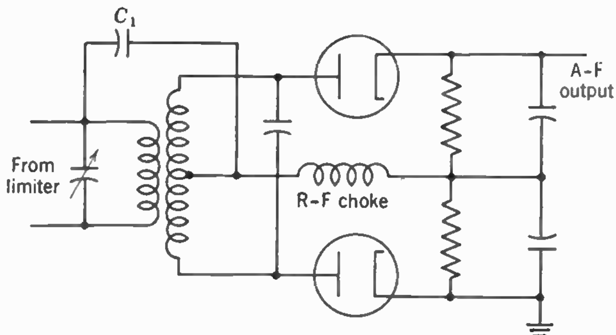


Fig. 11. Discriminator circuit for producing audio frequencies from f-m voltages.

fier. Radio-frequency current is also fed to the secondary from the limiter through condenser C_1. This produces a voltage across the r-f choke. Primary and secondary circuits are tuned to the center of the i-f pass band.

When the incoming signal is unmodulated, that is, when the incoming signal is at the frequency for which the primary and secondary circuits are tuned, the diodes are supplied with equal voltages which are opposite in polarity since they are connected to opposite ends of the center-tapped secondary. If, however, the output of the limiter is higher (or lower) in frequency than the unmodulated signal frequency, voltages induced across the secondary are added to (or subtracted from) the voltage secured from the C_1 path. Thus, one diode will have more voltage on

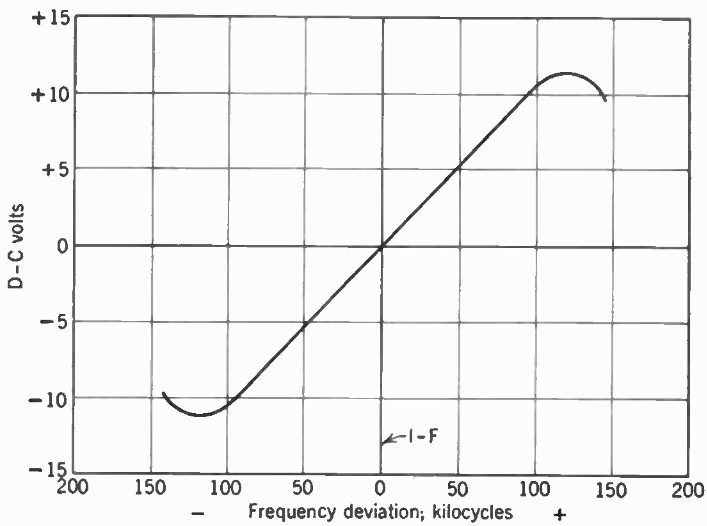


FIG. 12. Discriminator characteristics.

it than the other. This is true because of the phase relations existing between the two voltages in the secondary circuit—the induced voltage and the C_1 voltage. The voltage across the diode load is equal to the difference between the rectified voltages of the two diodes so that, under the unmodulated condition, no audio signal results; but, when one diode gets more input voltage than the other, the output voltage will increase. On the other hand, if the opposite diode gets more input voltage, less audio output will be produced. The condensers across the diode load resistors remove any remaining r-f signal. This is the circuit used for automatic frequency control (Sect. 15.25) with slight modifications.

The audio amplitude variations will be proportional to the frequency of the f-m input wave, and the frequency of the audio output voltages will depend upon the rate at which the f-m limiter voltage varies in frequency. The over-all characteristic of the discriminator may be seen in Fig. 12, which shows that, over a considerable frequency range, there is a linear relation between frequency variation and output voltage.

The discriminator, therefore, is a converter which changes frequency variations into voltage variations. These voltage changes (the audio signal) may be amplified in typical audio systems and applied to a loud speaker.

20 · Ultra-High-Frequency Phenomena

As one proceeds to generate, transmit, and receive waves of higher and higher frequencies, certain effects which were more or less negligible at the lower frequencies begin to assume new importance.

20.1. Transmission. Let us first consider the actual radiation of ultra-high-frequency waves. The first experiments by Hertz and Marconi were concerned with waves of very short wavelength. Hertz demonstrated with centimeter waves that wireless and other forms of electromagnetic radiation—light and heat—had identical characteristics. That is, "radio" energy could be *transmitted* through space; *reflected* just as light waves are reflected; *refracted* as light rays are refracted (bent) by a lens; and *diffracted* as light is broken up into interference patterns by a grating.

Marconi, however, wanted to send wireless messages great distances, across the ocean, in fact. He reasoned naturally, and correctly, that the antennas must be high in the air. A high antenna meant a long wavelength because transmission lines to connect a short-wave antenna, high in the air, to a transmitter on the ground had not been developed. Marconi's antenna and the ground lead both contributed to the radiation, and the total length governed the resonant wavelength.

Commercial practice was to use long waves for the above reasons and for other reasons as well. It was not until much later in the development of the art, after the first World War, that shorter wavelengths were employed commercially. Even then, it was assumed that there was a lower limit to the wavelength (or upper limit to frequency) beyond which it was useless to go, because of the rapid dissipation of the ground wave in short-wave radiation. Then it was discovered that short waves could be sent long distances, contrary to the generally accepted theory of the time. This discovery, in which American amateur wireless operators played an important part, followed the theoretical work of Kennelly and Heaviside and led to the discovery of the ionosphere, which makes long-distance, short-wave radio communication possible. All the time, however, communication could have taken place at wavelengths vastly shorter, but no one wanted to communicate for only short distances, or do the many other things now done with ultra-short waves.

As one increases the frequency (decreases the wavelength) to the neighborhood of 30 Mc (10 meters), the sky wave no longer is a useful part of the communication system because it no longer is returned to the earth. The ground wave is very quickly dissipated by ground losses. At these frequencies, and higher, only the direct rays between transmitter and receiver antennas are useful.

20.2. Direct ray. Why are not the very high frequencies reflected by the ionosphere? To be useful as a reflector, the individual particles which make up the reflecting surface must be close together compared to the wavelength of the radiation to be reflected. For example, an antenna reflector made of woven wire with rather coarse spacing ("chicken wire") will work satisfactorily for 50-cm (600-Mc) waves; it will not serve as an effective radiator for 10-cm (3000-Mc) waves. Similarly, the reflecting coating on a mirror must be continuous because the wavelength of visible light is so short—around 50×10^{-6} cm.

Now consider the ionosphere. It is made up of ionized particles with some sort of irregular distribution. But, whatever that distribution is, there is some wavelength of radiation small enough that most of the radiation escapes through the spaces between the particles, and only part is reflected.* At still shorter wavelengths (higher frequencies) very little or none of the radiation is reflected back to the earth.

On the very high frequencies, 30 Mc and above, there is no sky wave which can be depended upon for communication like that used at lower frequencies of the order of 2 to 30 Mc. Even though a part of the high-frequency waves may be reflected, the angle of reflection is such that the returning wave misses the

* This elementary concept is not intended to be rigorous. Many factors affect the frequency at which a radiated wave fails to be reflected back to the earth.

earth. Since the ground wave is highly attenuated, or decreased in strength, as it moves out from the antenna, only the energy which moves in a straight line through space from the transmitting to the receiving antenna can be depended upon for communication. This is the **direct wave.**

Since only the direct wave is useful at high frequencies, both transmitting and receiving antennas must be mounted well above the earth to achieve a large service area. This is similar to the ability of a man in the crow's nest of a ship to see farther than a man on the deck. Actually, the direct waves have a slight tendency to follow the curvature of the earth and reliable communication is possible over somewhat greater than the line-of-sight distances shown in Fig. 27, Chapter 18.

20.3. Radiation field at high frequencies. The strength of the radiated wave increases with frequency. This fact has already been mentioned in Sect. 18.12, but it is of such great importance that further discussion is warranted. Such a characteristic of radiated waves means that a small amount of radiated energy at a very high frequency can be just as effective as a large amount of radiated energy at a low frequency in producing a signal at a given distance from a transmitter.

Suppose that a receiver requires a certain signal voltage on its input circuit to cause satisfactory operation, and suppose that a transmitting station 30 miles away can transmit either 30- or 60-Mc signals. If the receiver is equally sensitive to either frequency signal, how much greater must be the power output of the transmitter at 30 Mc than at 60 Mc to produce the same power at the receiver?

The distance is great enough that only the direct wave need be considered. The voltage (or electric field) at the receiver, for a given voltage (or electric field) at the transmitter, is directly proportional to frequency. Thus it would take only half as much radiated electric field at the transmitter to produce a given electric field (and voltage) at the receiver at a frequency of 60 Mc as it would take at 30 Mc. The power output of the transmitter at 30 Mc would have to be four times as great as the power at 60 Mc since the power varies as the square of voltage.

20.4. Band width. There are many reasons why very short waves are useful. Consider the matter of band width. In the standard a-m broadcast region, each station requires a 10-kc band, and stations can be placed 10 kc apart. Between 550 and 1600 kc (the total band assigned to such stations) there are 1050 kc. Room exists in any one large geographical region for only 105 stations. The usable number is actually much less than this since modern broadcast receivers will not separate stations on adjacent channels if the strengths of the two signals are approximately equal. Now consider the region between 55 and 160 Mc, in which the ratio of the highest to the lowest frequency is the same as in the broadcast band. The total band is 105 Mc, or 105,000 kc, and if each station still requires only 10 kc, 1050 of them can be assigned to this band.

There are many services, however, which require wide bands: television and frequency modulation, for example. The only feasible place for these services is in the high-frequency regions. Even these regions are becoming extremely congested because of the demands of a wide variety of services: commercial, military, police, television, frequency-modulation, amateur, and the like.

Band-width Requirements

Service	Band Width
Continuous-wave telegraphy	Equals telegraph speed in bauds. (1 baud = 0.8 word per minute for a telegraph code having 8 dots or blanks per letter.)
Modulated continuous-wave telegraphy	Add twice the modulation frequency to the above.
Commercial telephony	6 to 8 Mc
A-m broadcasting	10 to 30 kc
Wide-band frequency modulation	200 kc
Black-and-white television	6 Mc
Facsimile	Number of picture elements ÷ time of transmission in seconds

20.5. Antennas. Consider now the half-wave dipole antenna, one of the most efficient radiators. At 15,000 meters (20 kc), about the longest wavelength used commercially, a half-wave antenna would be 7500 meters or approximately 25,000 ft long.

At 1000 kc (middle of the a-m broadcast band), or 300 meters, the antenna need be only 500 ft long (approximately); at 100 Mc the antenna need be only 5 ft long.

The fact that an antenna for use at high frequencies can be much shorter than one for use at low frequencies gives rise to incidental, but important, advantages. A short antenna requires not only less space but also less expensive supporting structures. Furthermore, it is much easier to construct a highly directive system at high frequencies by the simple expedient of using several half-wave dipoles properly spaced from each other and with the currents in them properly phased with respect to each other. Such directive systems at low frequencies would be out of the question because of space and cost factors.

It is no wonder, then, that short-wave radio is so effective on airplanes, in automobiles, or in portable transmitters.

20.6. Difficulties at high frequencies. The advantages of the high frequencies are commensurate with the difficulties in the methods of generating and receiving them. Some of these difficulties are discussed below.

The frequency to which a tuned circuit responds is proportional to $1/\sqrt{LC}$. To cause a circuit to respond to higher and higher frequencies, the product of L and C must be made smaller and smaller. Soon the point is reached where the individual values of L and C are merely those which are inherent in the circuit or apparatus—the inductance of the connecting wires shunted by the stray capacitance between the tube elements, in the sockets, etc. For example, 0.1 μh shunted by 10 $\mu\mu$f will be resonant at approximately 160 Mc. Now 10 $\mu\mu$f is the approximate value of the input or output capacitance of the average radio tube. Therefore, if the grid and cathode of a tube are connected together with a wire having an inductance of 0.1 μh, a parallel-resonant circuit is produced which is resonant to 160 Mc. This amount of inductance is equal to the self-inductance of a single No. 10 copper wire about 5 in. long.

In the above example, the inductance of the grid, cathode, or plate leads inside the tube (Fig. 1), and the capacitance between the socket or base terminals have not been taken into account. Thus, with the circuit as described above, the maximum frequency has been reached. How can higher frequencies be generated and amplified?

One simple way is to reduce the dimensions of all the tube elements, making grid and plate smaller. In this way the capacitances inherent in the tube may be decreased by a factor of about 10. But the power output of the tube is correspondingly reduced. Fortunately, high power is not so necessary at these high frequencies, but lack of power is a real limitation in many applications.

The upper frequency at which certain tubes designed for u-h-f use will operate is given in some tube manuals. For example, a 6J4 triode will give usable power output up to about 500 Mc, and a 6F4 "acorn" triode will work up to about 1200 Mc. The resonant frequency of these tubes, with the shortest possible connections between their terminals, is slightly greater than the frequency values cited. It is not practical, therefore, to connect

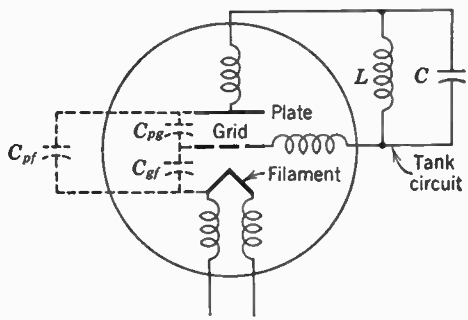


FIG. 1. Interelectrode capacitances and inductances inherent in tube.

these tubes to external circuits that are resonant to higher frequencies, for the tubes are unable to amplify these higher frequencies.

In working with short waves, every precaution must be taken to see that unnecessary inductance and capacitance are eliminated from the circuit.

20.7. Impedances at high frequencies. The reactance of a capacitor of given size decreases directly as the frequency increases, and the reactance of an inductor increases directly as frequency increases. Consider a capacitance as small as 10 $\mu\mu$f. the input capacitance of the average radio receiving tube. It has a reactance of only 160 ohms at 100 Mc, and may represent a very low-impedance path for alternating currents.

On the other hand, suppose that an inductance of only 1 μh exists in a part of the circuit where it is not wanted. The reactance, at 100 Mc, will be 628 ohms, and appreciable feedback voltage might be produced in such a wire by permitting currents from two points in an amplifier to flow through this common reactance.

Even small capacitances have low reactance at high frequencies. Trouble is experienced in making a choke coil, for example. Even though it looks like a coil, and at low frequencies behaves as such, its distributed capacitance may have less reactance than its inductance at the frequency at which it is to be used, and the net effective reactance is actually capacitive. If the inductive and capacitive reactances of the "coil" are equal, and they will be at some frequency, then the "coil" turns out to be a parallel-resonant circuit presenting a very high impedance to currents of the resonant frequency.

The behavior of capacitors at high frequencies must also be considered very carefully. The leads have some inductance; there is distributed capacitance existing between the leads as well as the normal capacitance of the capacitor. The inductive and capacitive reactances again may combine to form a parallel-resonant circuit. If, then, a condenser is employed to by-pass currents of the resonant frequency, the usefulness of the component is completely vitiated by the fact that it has both inductance and capacitance. A further inductive reactance may exist in a capacitor as a result of current flowing in the foil plates, particularly if the capacitor is spirally wound.

20.8. Distributed constants. In low-frequency circuits little trouble is experienced in making good coils and condensers to supply required values of inductive and capacitive reactance, but at the higher frequencies these components must be so small that real trouble is experienced in producing them. At 100 Mc or so it becomes impractical to produce coils and condensers of the proper size, and other methods of producing needed reactances must be investigated.

One widely used method of producing capacitance or inductance at high frequencies is to take advantage of the characteristics of transmission lines. For example, suppose that two parallel wires have a certain capacitance per unit length, and say that it is 1 $\mu\mu$f per in. Then, if 10 $\mu\mu$f is required, a 10-in. length of the two wires could be used. This would be a good solution for frequencies at which the wavelength was very much longer than 10 in., but at wavelengths comparable to the length of the two wires the inductance of the wires must be taken into account. In fact, if the length of the wires is one-quarter wavelength, the input impedance is nearly zero; if the wires are one-half wavelength long, the input impedance is very high. In the first case, the two wires represent a series circuit at resonance, and in the second case, a parallel circuit at resonance (see Sect. 18.3). If the wires are less than one-quarter wavelength long, they will behave as a capacitance and can be used as such. If longer than a quarter but less than one-half wavelength, they behave as an inductance. Because of these characteristics, transmission lines (either open-wire or coaxial) are widely used as circuit elements at higher frequencies. They may replace condensers, coils, or tuned circuits.

Another device which is used at high frequencies is the butterfly condenser. It is so made that current flowing in the plates causes the effect of an inductance shunting the capacitance, and so the unit may be used as a parallel tuned circuit.

20.9. *Q* at high frequencies. A designer has no trouble securing high impedances to work with at, say, 1 Mc. He simply shunts a capacitor by a coil. The lower the inherent resistance of the circuit, the higher the Q, and the greater the impedance of the circuit at resonance.

Suppose that a condenser of 300 $\mu\mu$f is shunted by a coil and the combination shunted across the output of an amplifier tube. The condenser, at 1 Mc, will have a reactance of 530 ohms. (The output capacitance of the tube is neglected since it is very small in comparison to 300 $\mu\mu$f.) The impedance of this circuit at resonance will be equal to the reactance of the condenser (or of the coil, since the two reactances are equal at resonance) multiplied by the Q of the circuit. Thus

$$\text{Impedance at resonance} = QX_c = QX_L$$

If $Q = 100$, the impedance is

$$100 \times 530 = 53{,}000 \text{ ohms}$$

This is a very respectable impedance, and considerable amplification could be obtained by using it as an amplifier load.

Now, suppose that the same impedance is required at 150 Mc. To secure the same reactance as above, the tuning capacitance would have to be reduced to 2 $\mu\mu$f, which is no greater than the input or output capacitance of even specially designed triodes or pentodes. Can this trouble be avoided? Suppose that, to achieve stability or to have some tuning ability, an effort is made to swamp the input capacitance of the tube by using a large tuning capacitor, which is the solution at 1 Mc. Let the tuning capacitance be increased by 30 times, to 60 $\mu\mu$f. Now the reactance is only 18 ohms at 150 Mc. To get an impedance of 53,000 ohms requires that Q be approximately 3000 ($R_{ar} = Q2\pi fL$). This value of Q is roughly 10 times that which is obtainable with a coil-condenser combination.

It is a fortunate fact that transmission lines, used as tuning elements, can be made to have Q values much higher than those of lumped inductances and capacitances. Coaxial lines have higher values of Q than open-wire lines, since the small radiation of energy from open-wire lines effectively reduces their Q. Certain types of resonant cavities, described later, can be made to have even higher Q values than coaxial lines.

And so the situation is not hopeless; it is only different, and new ideas must be developed, new agencies must be brought into use.

Problem 1. Calculate the r-f resistance of a tuned circuit which will have a Q of 3000 at 150 Mc if L and C each have a reactance of 18 ohms. (See Sect. 8.9.) Is this value of resistance reasonable?

20.10. Transit time. The upper frequency limit at which tubes will work is not only affected by the various capacitances and inductances within the tube which were described above; it is also affected by the **transit time** of the electrons. Electrons require a certain amount of time to travel from the cathode to the plate, and this transit time depends both on the spacing between the cathode and plate and on the speed of the electrons. The spacing between the cathode and plate can be made smaller by the tube designer; the speed of the electron can be increased by raising the plate voltage. There are, of course, definite limitations on how far one can go in both these directions. As the cathode-plate spacing is made smaller and smaller it becomes increasingly difficult to hold mechanical tolerances and to keep the grid from shorting to cathode or plate. With smaller spacings, too, the plate voltage must be kept low to prevent an arc-over within the tube.

If the transit time is small compared to the period ($1/f$) of the signal to be amplified by a tube, then no trouble develops. If, however, before an electron has reached the plate, the plate has become negative with respect to the cathode owing to the reversal of voltages in the plate load circuit, then the electron will not strike the plate. It will turn about and start back toward the cathode. This results in a loss of current to the plate circuit, and at some high frequency the tube becomes inoperative as an amplifier.

The most important effect of transit time, is, however, that the cathode-to-grid impedance of a tube is reduced materially when the frequency is raised so that the transit time is comparable to the period. A tube may have an input resistance of 10 megohms at 20 or 30 Mc, and only a few thousand ohms at 100 Mc.

The effects of transit time and of lead inductance and distributed capacitances within the tube all combine to set an upper frequency limit for the operation of triode and other similar amplifying tubes. Even the most carefully designed triodes will not operate satisfactorily at frequencies much above 2000 Mc. Receiving-type tubes are generally limited to frequencies below 100 Mc.

The time of transit can be calculated as follows: If the electron starts from rest at the cathode (no initial velocity), and if the voltage across the tube is E_b volts, then the final velocity with which the electron hits the plate is given by

$$v = 0.595 \times 10^8 \sqrt{E_b}$$

in which E_b is in volts and v is the velocity in centimeters per second. The average velocity is $v/3$.*

The time required for an electron to travel a given distance is the distance divided by the average velocity or

$$\text{Transit time} = \frac{\text{Distance}}{\text{Average velocity}} = \frac{D}{v/3} = \frac{3D}{v}$$

Suppose that the distance from cathode to plate in a tube (D) is 0.3 cm and that the voltage E_b is 300 volts. What is the transit time?

* Because of space charge in the tube, the voltage does not vary uniformly from cathode to plate. This causes the average velocity to be $v/3$ rather than $v/2$, which would be the case in the complete absence of space charge.

$$\text{Transit time} = \frac{3D}{v} = \frac{0.9}{v}$$

$$v = 0.595 \times 10^8\sqrt{300} = 0.595 \times 10^8 \times 17.3$$
$$= 10.3 \times 10^8 \text{ cm per sec}$$

$$\text{Transit time} = \frac{0.9}{10.3 \times 10^8} = \frac{1}{1145}\ \mu\text{sec}$$

This is the time required to complete one cycle of current at a frequency of 1145 Mc. Since transit-time effects are serious even when the period of the signal is 4 or 5 times as long as the transit time, the upper limit of operation for the tube in the example would be of the order of 200 Mc.

20.11. New concepts at high frequencies. Many concepts that were perfectly good at low frequencies must be examined critically when the frequency is 100 Mc and above. One of the most important of these concepts is that of current flow. Up to this point, an electric current has been considered as a completed motion of electrons from one place to another. Thus, within a tube, electrons go from cathode to plate and no current is considered to flow until an electron arrives at the plate. This concept is good only when the period of the signal is long compared with the transit time.

In a simple condenser it has heretofore been implied that alternating current flows through the condenser, and that electrons pass from one electrode through the dielectric and to the other plate. Actually this is not so at all. Electrons, in spite of their small dimensions and other properties, are unable to go through the dielectric of a condenser. What actually happens is that a *motion* of electrons to one condenser plate from the outside circuit and away from the other plate to the outside circuit causes an effect which is similar to the actual passage of electrons through the dielectric. The shift in electrons on the two plates of the condenser creates a strain within the dielectric (and may break it down), but there is no necessity for a single electron actually to move across the gap between the electrodes, any more than there is a necessity for electrons to pass between primary and secondary of a transformer for a voltage to appear across the secondary when the primary current is changed.

Any motion of electrons, then, may produce a motion of electrons in a near-by circuit. If, therefore, an electron motion is produced, say a surging back and forth, then a surging or rhythmic current is produced in the external circuit.

20.12. Lighthouse triodes. One of the special triodes developed for high-frequency use is the **lighthouse** triode or **planar-grid** tube. The mechanical construction of this tube is quite

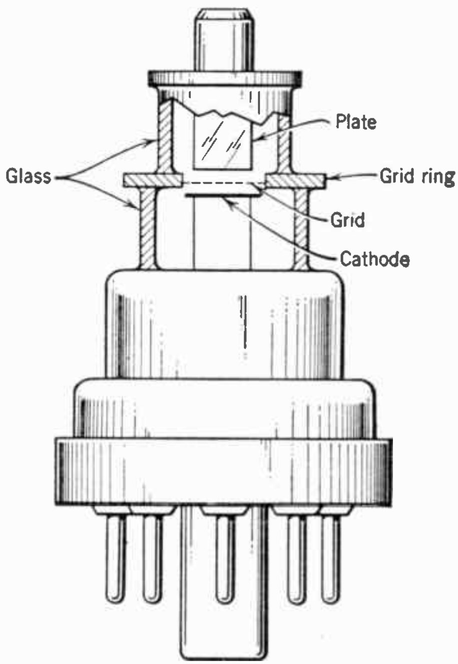


Fig. 2. Lighthouse tube, partial cross section.

different from that of the conventional triode since the grid is a fine mesh mounted in a circular metal ring. The grid ring is supported in place by being molded into the glass envelope of the tube. The emitting surface of the cathode is a disk on top of a cylinder, which contains the heater, and is about ¼ in. in diameter. The anode (plate) is a metal cylinder of the same diameter as the cathode disk and is sealed into the glass envelope above the grid. The general construction is shown in Fig. 2.

The metal ring supporting the grid expands along with the grid when heated in operation and thus keeps the grid taut and prevents it from sagging and shorting to the cathode or plate. This feature permits the grid-cathode and grid-plate spacing to be much smaller than in the conventional triode structure. Also, the construction results in very low inherent inductance. The lighthouse tube is designed as part of a coaxial system so that sections of transmission lines may be used as the tuned circuit elements. The upper frequency limit is about 3500 Mc.

20.13. Transmission systems. At very high frequencies, the problem of connecting the generator to the antenna occurs as at

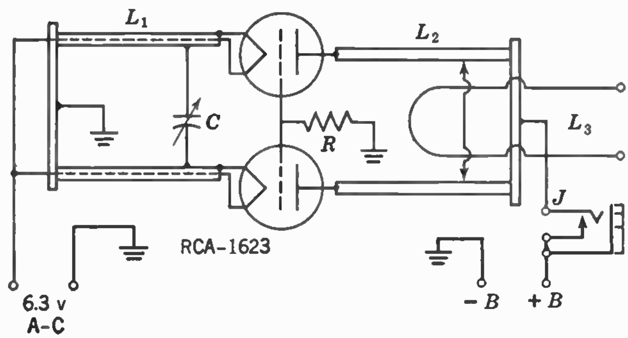


Fig. 3. A-c push-pull 2½-meter, 45-watt oscillator, RCA-1623 tubes; $C = 15$ $\mu\mu$f; $R = 1000$ ohms, 5 watts; L_1 and $L_2 = 12$-in. lengths of ½- and ⅝-in. metal tubing, spaced 1 in. and 1⅛ in. between centers, respectively; plate voltage = 500 for CW; plate current = 200 ma.

lower frequencies. A simple two-wire line may serve both as the tuned circuit and as the line to connect the generator to the antenna. If the wires are fairly close together, say 0.1 wavelength, they will not radiate much energy. At the same time they will have a high Q and so the circuit can be fairly efficient. Greater efficiency may be secured by using a coaxial line. Then no radiation occurs until the energy arrives at the antenna, because the outer conductor of the coaxial line completely shields the inner wire. As is true at lower frequencies, a quarter-wave line can be used as a matching transformer, coupling the generator to its load. A 2½-meter (120-Mc) oscillator employing sections of transmission lines as the tuned elements is shown in Fig. 3.

20.14. Cavity resonators. If a two-wire transmission line of the proper length (one-quarter wavelength) is short-circuited at the far end, it acts like a parallel-resonant circuit. To the generator it presents a high impedance; the line is resonant. Now suppose that a movable piston is inserted in a hollow pipe and the other end closed as shown in Fig. 4, and that r-f energy

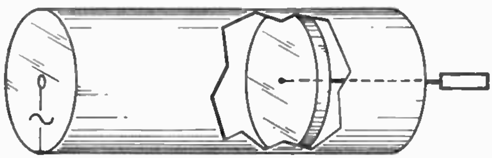


FIG. 4. Elements of cavity resonator. If the frequency of the energy introduced and the position of the plunger are correctly related, standing waves will be set up in the chamber, which then acts like an anti-resonant circuit of very high Q.

of the proper frequency is admitted through a hole in the closed end of the pipe. If, now, the movable piston is adjusted properly, it will be found that the pipe acts as a parallel-resonant circuit just like the shorted quarter-wave transmission line. It presents a high impedance to the generator which supplies the r-f energy. Such a device is known as a **cavity resonator.** It acts like an anti-resonant circuit with an extremely high Q.

Another way of looking at a cavity resonator is as follows. Suppose that a condenser is made up of two flat metallic plates and these plates are connected together with a coil as shown in the left-hand diagram of Fig. 5. The result is a circuit whose

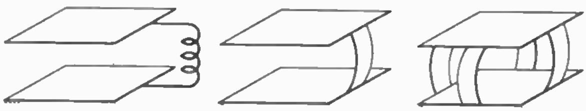


FIG. 5. How a cavity resonator may be built up by completely closing in the sides of a two-plate condenser.

resonant frequency is a function of L and C. Suppose, however, that the two plates are connected by a flat strip as shown in the center diagram. Since the strip has inductance, this arrangement is also a resonant circuit. Now, suppose that several strips are connected between the upper and lower plates. The resulting combination is still a resonant circuit. As a matter of fact, the strips may be made continuous around the edges of the condenser plates so that a closed box results, and the box is still a resonant circuit. It is a cavity resonator.

Cavity resonators are made in a variety of shapes to fit particular requirements. They may be closed cylinders, spheres, rectangular boxes, or doughnut-shaped devices. Their shape is a factor in determining their effective Q, and, along with their physical dimensions, the frequency at which they resonate. The resonant frequency is also affected by the manner of excitation.

A copper sphere having a radius of 17.5 cm resonates to a wavelength of 40 cm (750 Mc) and has a Q of about 50,000 and a shunt resistance of about 4 megohms. Since the radius of spherical resonators is proportional to the wavelength, it is a simple matter to design a whole series of resonators, once one has been designed. The value of Q and the shunt resistance vary approximately as the square root of the wavelength.

Cavity resonators are used as the tuned elements in high-frequency wavemeters, as essential elements of certain tubes such as klystrons, and for many other purposes.

20.15. Velocity-modulated tubes. Suppose that a beam of electrons of uniform velocity is caused to pass through two grid-like structures **(buncher)** upon which an alternating potential

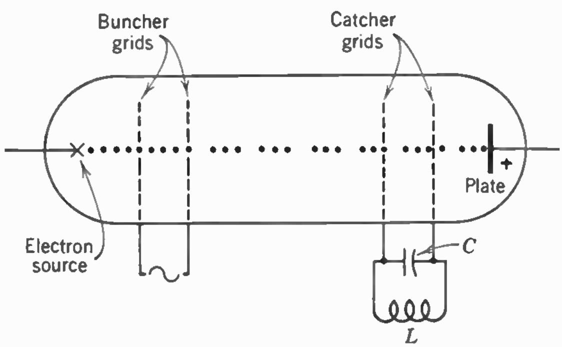


FIG. 6. Elements of velocity-modulation tube. Note that electrons flow in bunches after passing the field of buncher grids.

is placed as in Fig. 6. If the spacing of the buncher grids, the frequency of the alternating potential, and the transit time of the electrons between the grids are properly related, the electrons which pass into the space between the buncher grids on the half cycle when the right grid is positive will be speeded up. Those which enter when the right grid is negative will be slowed down. In this manner the uniformity of the velocity of the electrons is destroyed, and, at certain points further along the tube, some of the electrons are bunched together and others are scattered out. If successive groups of the bunched electrons pass through another pair of grids **(catcher)** on their way toward the final positive electrode or **collector,** alternating voltages will be produced upon the catcher electrodes. This follows from the discussion in Sect. 20.11, where it was found that electrons do not have to actually strike an electrode to induce voltages and currents in it.

This is the basis of several types of tubes very useful in generating centimeter waves, of which the **klystron** is typical. In one tube of this general type, 110 watts may be produced at 450 Mc.

In practice, the tuned circuits for the buncher and catcher are replaced by cavity resonators. Klystrons using such resonators are shown in Figs. 7 and 8.

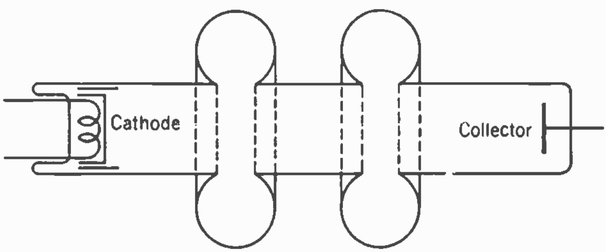


Fig. 7. Elements of the klystron. The circular devices along the tube are cavity resonators.

To make such tubes self-generating (oscillate) it is necessary only to use a feedback system by which some of the energy in the catcher circuit is fed back to the buncher circuit. Output power is taken from the catcher cavity.

Another type of velocity-modulated tube, the **reflex klystron,** uses a single cavity. A **repeller** electrode replaces the collector of the double-cavity tube. This repeller is held at a negative voltage and so repels electrons which come through the buncher grids, causing these electrons to pass back through these grids. If the repeller voltage is of the proper value, the electrons will be returned to the buncher in bunches and at the proper time to give up their energy. Tubes of this type may be made which will oscillate above 9000 Mc. Their power output is usually of the order of 0.1 watt. They are widely used as the local oscillators in superheterodyne microwave receivers.

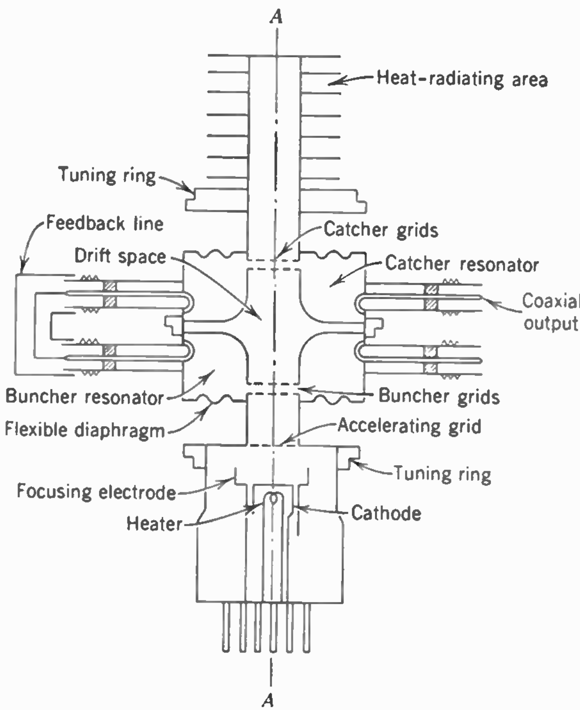


FIG. 8. Sectional view of the 410-R/2K30 klystron of Sperry, a tube of the velocity-modulation type which makes possible generation, amplification, and conversion of frequencies in the centimeter-wavelength region. The tube is 6 in. long overall.

20.16. The magnetron. The **magnetron** is a high-frequency oscillator utilizing both electric and magnetic fields to produce a high-frequency output. There are several types of magnetrons, of which the multicavity type is most widely used.

Diagrams of a multicavity magnetron are shown in Fig. 9. The anode is either machined from a solid block of copper or assembled from thin copper punchings. The circular holes in the anode, with slots into the anode cavity, are cavity resonators, and determine the operating frequency. Different numbers of cavities are used. Six are shown in the figure. The cathode is a metallic cylinder coated with an emitting material and containing an internal heater. Power is taken from the tube by means of a loop in one of the cavities. In operation, the tube is placed between the poles of a powerful permanent magnet.

An electron moving in a magnetic field, and perpendicular to it as in the magnetron, is caused to move in a circle. If, in addition, there is an electric field as is the case here, the circular

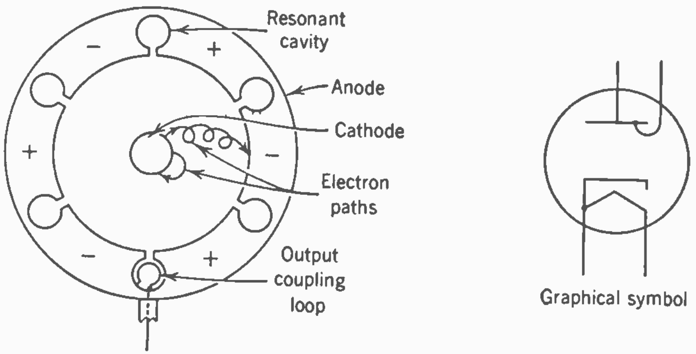


Fig. 9. Multicavity magnetron.

motion is distorted but the electron moves in a looping path as long as the circular motion is predominant. The magnetron is operated with its anode positive with respect to its cathode. Thus, electrons are attracted toward the anode, but move in paths such as shown in the figure.

Consider the case in which the magnetron is oscillating. Then the cavities in the anode, acting as resonant circuits, will produce alternating voltages, and there will be voltage differences across the slots. The polarities of these alternating voltages (in addition to the direct voltage supplied to the anode) at a particular instant of time are shown in the figure. The instantaneous voltage across adjacent slots is such that the intervening segments of the anode are alternately positive and negative. Now, if an electron, moving in its looping path, passes the slots to adjacent cavities in a time equal to the period of the oscillations, the electron will give up energy to the cavities and so produce r-f power which is available at the output terminals of the tube.

Magnetrons have found their widest application in radar equipment. In this service they are usually pulsed. That is, high direct voltages are applied to the tubes for short periods of time, then turned off. The procedure is then repeated. Each time the voltage is applied, the tube generates a pulse or burst of high-frequency energy. Magnetrons are available which will produce more than a million watts of r-f energy during these short pulses, even though the average power output is not more than 500 or 600 watts. The frequency range of magnetrons is from about 150 to 20,000 Mc.

20.17. Wave guides. One of the interesting devices used in very high-frequency apparatus is the **wave guide.** A wave guide is nothing more than a hollow pipe which may have a circular,

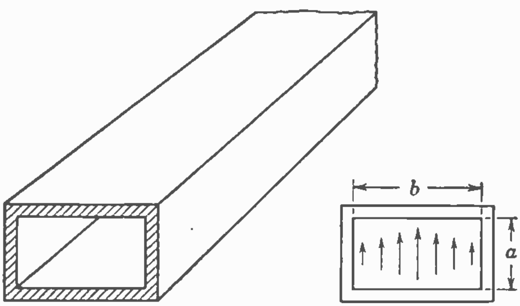


FIG. 10. Rectangular wave guide. The lengths of the arrows are proportional to the electric field strength.

rectangular, or some other cross section. Under certain conditions a wave guide will transmit high-frequency energy with high efficiency.

The ability of a hollow pipe to guide r-f energy may seem strange since there is only one conductor—the inner surface of the pipe. And yet a transmission line requires two conductors. It must be remembered, however, that radio waves are radiated through space without any conductors. Waves on transmission lines or in wave guides are "guided" from one point to another. Radiated waves in space are essentially unguided.

The question that naturally arises in regard to wave guides is, Why are not the top and bottom sections short-circuited since they are all part of a continuous conductor? This question can be answered by reference to Fig. 10, which shows a rectangular wave guide. If it were possible to inject a wave into such a guide so that the electric field intensity (the volts per meter) was maximum at the center and zero at the edges of the top and bottom sections of the guide as shown in the figure, then there would be no voltage along the edges and the top and bottom sections would not be short-circuited. There would still be an electric field in the wave guide (also a magnetic field which is not shown) and energy could be transmitted down the guide. The dimension b in the diagram must be at least one-half wavelength since it must accommodate a wave with a field intensity that

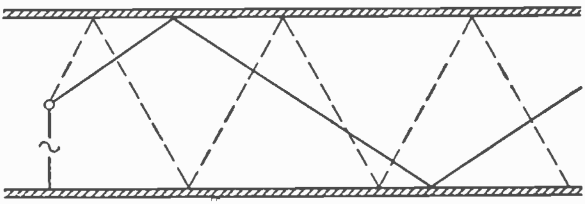


FIG. 11. How radiation is reflected by inner walls of a wave guide just as light may be sent down a glass pipe or tube.

starts at zero at one side, rises to a maximum at the center, then falls to zero at the other side. This is equivalent to a half wavelength. For the wave shown in the diagram, the dimension a has no effect on the operating frequency. It must merely be large enough to prevent voltage breakdown between the upper and lower sections of the guide. There are, however, other manners in which a wave guide may be excited—the manners of excitation are called "modes"—and with some of these both dimensions of the wave guide affect the lowest frequency that can be transmitted.

The lowest frequency (longest wavelength) that can be transmitted down a wave guide is called the **cut-off frequency (cut-off wavelength)**.

As energy is transmitted down a wave guide it is bounced back and forth from upper to lower walls as seen in Fig. 11. The solid line shows the path for a wave whose frequency is considerably above the cut-off frequency, and the dashed line is for a wave near the cut-off frequency. As the cut-off frequency is reached, waves bounce up and down and make no progress down the guide. Even for waves near the cut-off frequency, which do have forward motion, there is considerable attenuation of the power in the wave. For this reason wave guides are generally not used too near the cut-off frequency.

Wave guides are useful only at very high frequencies because the physical dimensions must be comparable to the wavelength of the energy to be guided. At long wavelengths, the dimensions would have to be so great that it would be impractical to construct, support, or maintain the guide. But at high frequencies a wave guide can be a small and extremely efficient conductor. The losses in metallic pipes of circular or rectangular cross section are of the order of a few decibels per mile. In a 5-in. cylindrical copper wave guide, for example, the attenuation for 5000-Mc signals is about 10 db per mile, and over a band several thousand megacycles wide the attenuation does not differ much from this value. The attenuation of even the best coaxial cable at these frequencies is considerably higher because the currents are concentrated in the rather small surface area of the center conductor and therefore produce a relatively higher loss than in the outer conductor, which has a large surface area. There is, of course, no inner conductor in a wave guide.

Since wave guides will not pass r-f energy below the cut-off frequency, they behave in this respect like high-pass filters and are sometimes used as such. The cut-off wavelength for the rectangular guide shown in Fig. 10, and for the mode of excitation indicated, is

$$\lambda_c = 2b$$

where b is the long dimension of the wave guide.

For a cylindrical tube the longest wave that will pass down the pipe varies from 1.64 to 3.46 times the radius, depending upon the mode of excitation.

A wave guide has a characteristic impedance much as a transmission line. And, as with a transmission line, the generator and the load device should be matched to the impedance of the guide for the most efficient transfer of power. The lowest practical characteristic impedance of a round pipe is about 250 ohms, but, by varying the small dimension of a rectangular pipe for a given value of the large dimension, the impedance may be made practically any value desired up to 465 ohms. By varying both the dimensions correctly the impedance may be made to have practically any value.

It is not necessary that a wave guide be made of metal. A solid rod of dielectric material will also guide waves, and polystyrene is sometimes used in this service. Like metal tubes, these dielectric wave guides have a cut-off frequency. The attenuation in such guides is ordinarily higher than in metal guides.

20.18. Practical aspects of wave guides. In a coaxial line, the attenuation or loss of energy per unit length depends upon the high-frequency resistance of the inner surface of the outer conductor and the outer surface of the inner conductor. Since the inner conductor has the smaller surface area, most of the losses occur in the central wire or pipe. In a wave guide there is no inner conductor, and all the energy dissipation takes place as a result of currents flowing along the inner surface of the tube. The losses, in general, are lower than in a coaxial line. If the lowest losses are to be experienced, low-resistance material (copper) should be used. Silver plating will still further reduce the losses. It must be remembered that, at the frequencies under consideration, currents flow on the surface or barely penetrate the metal, owing to skin effect.

For 3000-Mc waves, brass pipe having an outside diameter of 3 in. and a $\frac{1}{16}$-in. wall may be used. A rectangular pipe for this frequency may be made of 0.081-in. material. Dimensions of $1\frac{1}{4}$ by $2\frac{1}{2}$ in. are suitable.

20.19. Wave-guide couplers. In many high-frequency systems employing wave guides, provisions must be made for coupling energy into or out of a wave guide from a coaxial line. This is usually done by terminating the outer conductor of the coaxial line at the wall of the wave guide and extending the center conductor into the guide as an antenna. Sometimes the center conductor of the coaxial line is run on through the wave guide and into a pipe extending from the other side as shown in Fig. 12. A piston in this pipe short-circuits the coaxial conductor and gives a means of matching the coaxial line to the guide. The movable plunger shown in the end of the wave guide is also for impedance matching. The face of the plunger which is inside the wave guide is located roughly one-quarter wavelength from the center conductor of the coaxial line.

The wave-guide couplers described above are only two of many that are available. Some employ loops rather than a straight

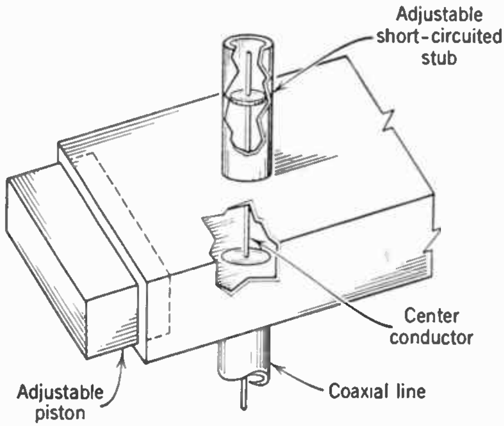


FIG. 12. Wave-guide coupler.

wire. The type used and its placement within the guide determine the mode of excitation.

20.20. Wave-guide horns. Many of the antennas described in Chapter 18 can be adapted for use at very high frequencies. All that is necessary is to scale them down to the proper size. However, these are all for use with either a two-wire or coaxial line. Another problem arises when it is desired to radiate energy from the open end of a wave guide.

It would be possible, of course, to permit the energy to be radiated from the end of a wave guide with no special provisions for directing it properly or securing an "impedance match" between the wave guide and free space. However, such an arrangement would be inefficient and the resultant field pattern would not be highly directive. What is needed is a gradual transition from the wave guide to space. This can be conveniently achieved by merely flaring the end of the wave guide much like the flare of a horn speaker and for much the same reason. The resultant field pattern is a highly directive elongated "oval" such as shown in Fig. 13. By the proper choice of the angle of flare and length of the wave-guide horn, the pattern can be made to be very directional.

20.21. Parabolic reflectors. Everyone is familiar with the fact that a source of light placed at the focus of a parabolic reflector can be made to produce a highly directive beam. This

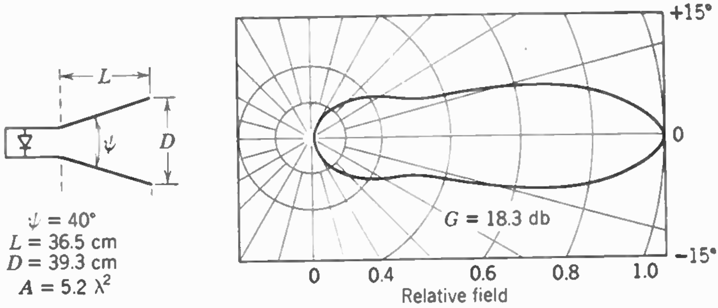
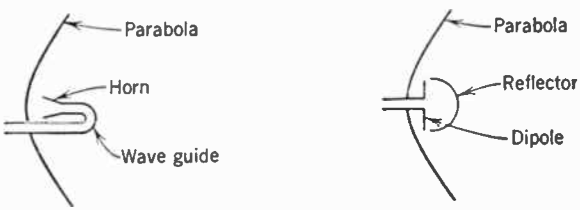


Fig. 13. Directional characteristics of a metal horn-type antenna operating on 15.3-cm wavelength.

scheme is used in most automobile headlamps. Similarly, a source of r-f energy can be properly located with respect to a metallic parabolic reflector with the result that a highly directive beam of r-f energy is produced. At short wavelengths the entire structure, antenna, reflector, and generator, can easily be made portable. The angle of radiation may be made extremely nar-



(a) Wave Guide Feed (b) Dipole Feed

Fig. 14. Parabolic reflectors.

row, and a surveyor's transit is often used to line up a transmitter with a receiver utilizing such a system.

A parabolic reflector may be fed from either a wave guide or a dipole antenna. In the case of a dipole antenna, a disk or half cylinder of metal is usually mounted on the side of the dipole away from the reflector to keep the r-f energy that would otherwise be radiated forward from spoiling the directivity of the beam. This energy is reflected back to the reflector. Typical arrangements for both wave-guide and dipole feeds are shown in Fig. 14. To achieve good directivity patterns the outer diameter of the parabolic reflector should be several wavelengths.

Much simpler to construct than a parabolic reflector, although not quite so effective, is a "corner" reflector made up of two sheets of metal placed at an angle to each other and with the antenna at the apex. The angle between the two plates and the location of the antenna control the width of the beam produced.

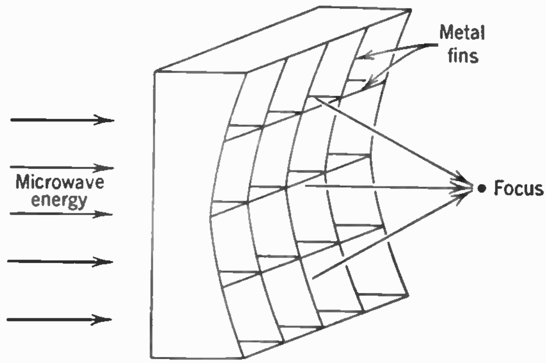


Fig. 15. Metal-lens antenna.

20.22. Metal-lens antennas. Parabolic reflectors used for short wavelengths concentrate or focus radio energy much the same as similar reflectors concentrate light energy. In a similar manner, it is possible to construct metallic "lens" systems which will concentrate r-f energy or radiate it in highly directive patterns. These are usually assembled from a series of metal strips in an "egg-crate" form as shown in Fig. 15. The contours of the two faces depend upon the use to which the metal-lens antenna is to be put. The faces may be plane, concave, or convex, the same as those of glass-lens systems for optical use. Lens antenna systems are finding considerable use in the microwave systems which relay television programs from city to city.

21 · Electronic Instruments

Recent years have seen the development of many instruments designed specifically for testing electronic circuits and instruments for many purposes which utilize vacuum tubes as essential elements. An understanding of the theory, operation, and use of these instruments is rapidly becoming a "must" for anyone who is working in the electronic field. Vacuum tubes make possible the design of instruments having the desirable features of high input impedance, extremely wide frequency range, high sensitivity, and ruggedness.

VACUUM-TUBE VOLTMETERS

21.1. Characteristics of vacuum-tube voltmeters. In the study of vacuum-tube detectors in Chapter 14, it was learned that a d-c component of current was always present in the output of the detector. It was also found that the magnitude of this d-c component depended upon the magnitude of the voltage impressed on the detector. Therefore, a d-c milliammeter inserted in the plate circuit of a detector will give an indication of the a-c voltage impressed on the grid circuit. If known voltages are impressed on the grid and the milliammeter scale is marked correspondingly in volts, the device will indicate voltage; in other words, the composite instrument will be a voltmeter. Also, since it uses a vacuum tube in its operation, it is called a **vacuum-tube voltmeter.**

A simple vacuum-tube voltmeter (VTVM) as described above operates on the principle that a small voltage applied to the grid of a triode controls a relatively large amount of plate current without drawing any appreciable power from the source of voltage. This type of voltmeter has three distinct advantages over the ordinary d'Arsonval type of d-c instrument, or the iron-vane or electrodynamometer types of instruments used in a-c measurements.

1. The tube voltmeter has high sensitivity: a relatively small change in input voltage may produce a relatively large change in the current flowing through the indicating meter. The high sensitivity is due primarily to the power amplification taking place in the tube. This is a very desirable feature when low voltages are to be measured.

2. The meter movement in a tube voltmeter can be much more rugged than the movements in, for example, a d'Arsonval meter having the same sensitivity. Because the voltage being read is in effect amplified in the vacuum-tube voltmeter, more current is available for actuating the indicating meter. For example, a d-c voltmeter having a sensitivity of 20,000 ohms per volt requires a movement which reads full-scale with a current of only 50 μa, whereas a tube voltmeter of similar or greater sensitivity would probably have a 0- to 1-ma (1000-μa) movement for the indicating instrument. The 0- to 1-ma movement is much more rugged than the 0- to 50-μa movement.

3. The tube voltmeter has a much higher input impedance than the ordinary type of voltmeter, particularly when the a-c ranges are compared. The advantages of this characteristic are great. With a high-impedance voltmeter, it is possible to measure voltages while drawing only a negligible amount of power from the voltage source. Thus it is possible to measure voltages accurately where the use of an ordinary voltmeter would so seriously disturb the circuit by lowering the voltage as to render any measurements unreliable. (This type of circuit disturbance by a measuring instrument is called loading. See Sect. 3.9.) For example, it is possible to measure the a-c voltage appearing on the grid of a detector tube with a tube voltmeter. Such a voltage measurement would generally be very difficult with other types of a-c instruments.

Other advantages will soon become evident. For example, a well-designed tube voltmeter may be calibrated at a low frequency and this calibration will be accurate at much higher frequencies.

The question may be asked, Why are any but vacuum-tube voltmeters used if they are so superior? The answer is that they are not superior in every respect. In general, they are not so accurate as the standard types of voltmeters. They may require more frequent calibration, owing to changes in circuit or tube characteristics. They may be sensitive to line-voltage variations if the power to operate them comes from the 115-volt a-c lines. They are generally bulkier than other types of a-c meters. However, these disadvantages are gradually being overcome, and the tube voltmeter of today is an accurate and reliable instrument.

21.2. Plate-circuit voltmeters. Vacuum-tube voltmeters using plate rectification to convert the a-c input into direct cur-

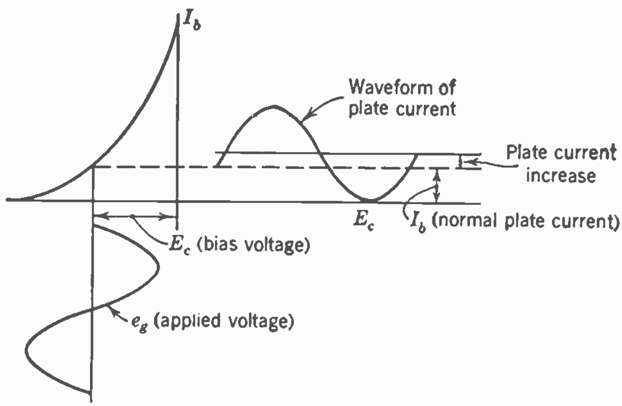


FIG. 1. Operating point for a full-wave square-law vacuum-tube voltmeter.

rent may be divided into three general classes, depending on the initial grid bias used: (1) the full-wave square-law type is biased midway in the lower curved portion of the I_b–E_c characteristic curve; (2) the half-wave square-law type is biased approximately at cut-off; and (3) the peak type is biased past cut-off.

In the full-wave square-law type (Fig. 1), plate current flows during most or all of the cycle of input voltage. The change in plate current produced by the rectifying action when an a-c voltage is impressed on the grid is almost exactly proportional to the square of the rms value of the applied voltage. Both halves of the applied voltage wave affect the plate current.

If the grid is biased to cut-off (Fig. 2) the negative alternations of the applied voltage have no effect on the plate current, and the change in plate current will be very nearly proportional to the square of the positive alternations. If this type of voltmeter is

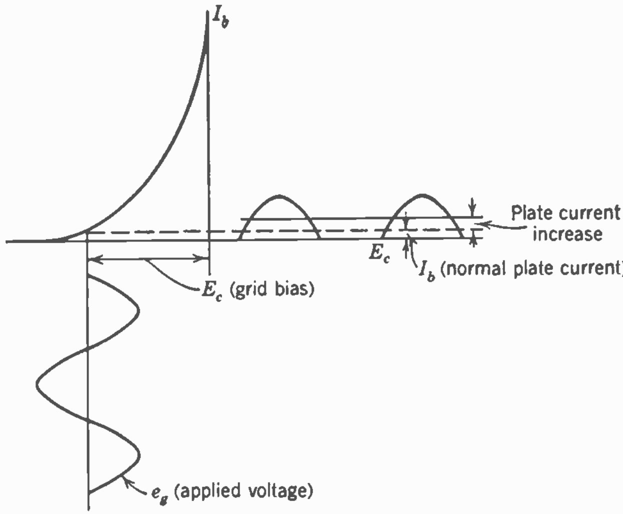


FIG. 2. Grid biased at or near cut-off in plate-circuit detector used as a voltmeter.

used to measure a non-symmetrical voltage wave, it may be necessary to reverse the input terminals for a second reading, then average the two readings to obtain a more nearly correct value of voltage.

If the grid is biased well past cut-off (Fig. 3) the change in plate current is determined primarily by the peaks of the positive alternations, and the meter becomes a **peak-reading voltmeter.** The wave form of either alternation has little effect on the reading since the voltage reading depends mostly on the peak value. A condenser connected in parallel with the meter in the plate circuit makes the instrument even less dependent on the waveform. However, in either case, if the peak values of the negative alternations are different from the positive peaks, a different reading would be obtained by reversing the input terminals.

21.3. Balancing circuit. Some plate current may flow in a tube voltmeter even when no signal is applied, unless the grid is biased at or beyond cut-off. This initial current is of no value and, when flowing through the plate meter, reduces the amount of the scale which may be used for indicating voltages. This disadvantage may be overcome by balancing out the initial plate

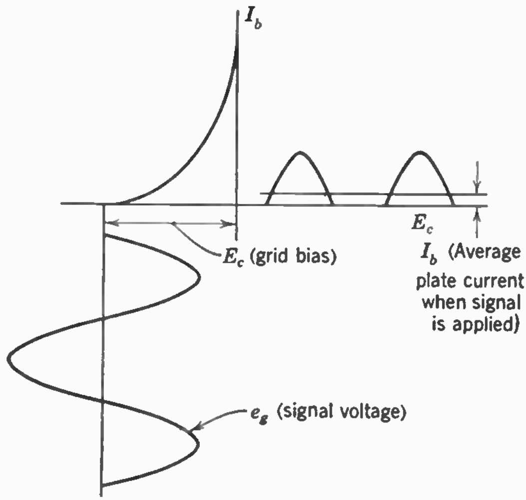


FIG. 3. When bias is well beyond cut-off, peaks of input voltage determine plate current.

current so that the plate milliammeter reads zero when zero voltage is applied. A typical circuit for accomplishing this result is shown in Fig. 4. The variable resistor R_2 is adjusted so

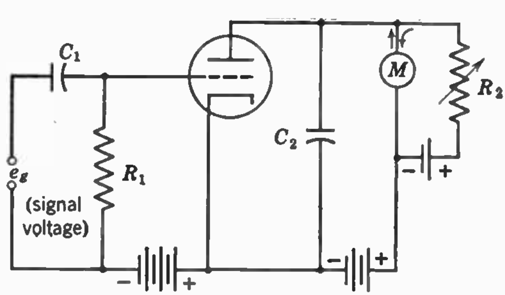


FIG. 4. Use of battery and rheostat for balancing out the steady no-signal current.

that the current from the auxiliary battery is exactly equal to that flowing from the main battery, with no signal applied to the grid. Both these currents flow through the meter, and since they flow in opposite directions the meter reading will be zero. Various other circuit arrangements may be made to accomplish this same purpose, some not requiring an auxiliary battery.

21.4. Slide-back voltmeter. Another voltmeter circuit which indicates the positive peaks of the a-c input wave is shown in Fig. 5. In this circuit the grid bias is adjusted so that a small current flows in the plate circuit. The unknown voltage is then applied and the bias readjusted so that the plate current is the same as before. The peak value of the unknown voltage is then equal to the change in grid bias required. The necessity of re-

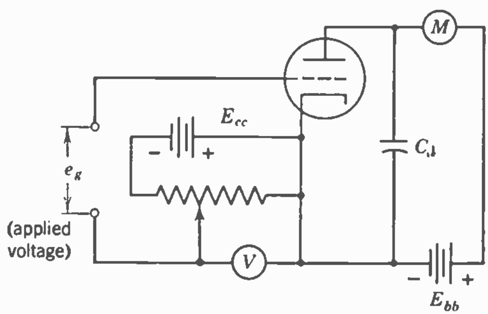


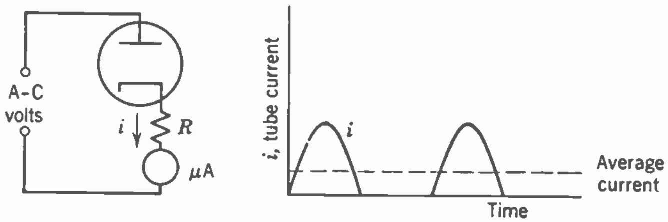
FIG. 5. Slide-back voltmeter. In operation, the grid bias, V, is adjusted to give a chosen value of plate current. The signal voltage which is to be measured is then applied. This increases the plate current. Finally, the bias voltage is increased to return the plate current to its original value.

adjusting the bias to a more negative value gives rise to the name **slide-back voltmeter.**

For low input voltages, the voltmeter is not very accurate and a correction must be applied. However, for higher input voltages, this meter gives very accurate readings. The accuracy depends primarily on the accuracy of the voltmeter employed in the grid circuit and not on changes in circuit constants or tube characteristics.

21.5. Diode voltmeters. The discussion so far has dealt only with voltmeters in which the unknown voltage is impressed on the grid of a tube, implying the use of a triode or multigrid tube. Diodes may also be used in vacuum-tube voltmeters. The purpose of the diode is to rectify the a-c voltage so that a d-c meter may be used as the indicating instrument. Since the diode does not amplify, the diode voltmeter cannot be more sensitive than the meter employed. However, a d-c meter and rectifier arrangement for measuring a-c voltages is desirable, since d-c meters in general have much higher sensitivities and thus higher input impedances than the corresponding a-c voltmeters.

Diode voltmeters may be divided into two types: (1) average-reading, and (2) peak-reading. Either type may employ half- or full-wave rectification. A simple average-reading diode voltmeter and the waveform of the rectified current flowing through the meter are shown in Fig. 6. Even though both a-c and d-c



(*a*) Typical Circuit (*b*) Rectified Current Meter Reads Average Current

FIG. 6. Simple average-reading diode voltmeter.

components of current flow through the meter, the meter responds only to the average or direct value of the current. Hence, the name **average-reading.** Except for low-voltage ranges, the sensitivity of the meter is determined largely by the full-scale current of the d-c meter and the series resistor R. For low-voltage ranges R may be small enough that the internal resistance of the diode has an appreciable effect on the sensitivity. A typical instrument incorporates a 0- to 50-μa meter with an appropriate multiplying resistor for each voltage range. The meter indicates the average value of the positive half cycle of the input voltage.

The diode voltmeter may be made into a **peak-reading** instrument by shunting the multiplying resistor and meter with a condenser as in Fig. 7. During the positive half cycle the current divides and part flows through the meter and part into the condenser. Thus the condenser is charged during this portion of the cycle and may discharge through the meter during the negative half cycle when the tube is not conducting. This discharge current passes through the meter and causes the instrument to indicate the peak of the input voltage. This device is more sensitive than the average-reading meter described above, but the

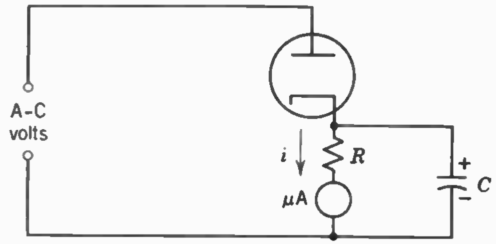


FIG. 7. Simple peak-reading diode voltmeter.

readings are subject to larger errors if the waveform of the voltage being measured is not sinusoidal.

21.6. Voltmeter amplifiers. Diode rectifiers are often combined with amplifiers to make tube voltmeters respond to lower voltages, to permit the use of a more rugged indicating instrument, and to secure other desirable characteristics. Two combinations of the rectifier and amplifier are in common use. The first combination places the diode rectifier at the input terminals. The resulting direct voltage output is amplified in a d-c amplifier. The General Radio type 727-A is an example of this circuit. Figure 8 shows the basic operating parts of this meter. A tube

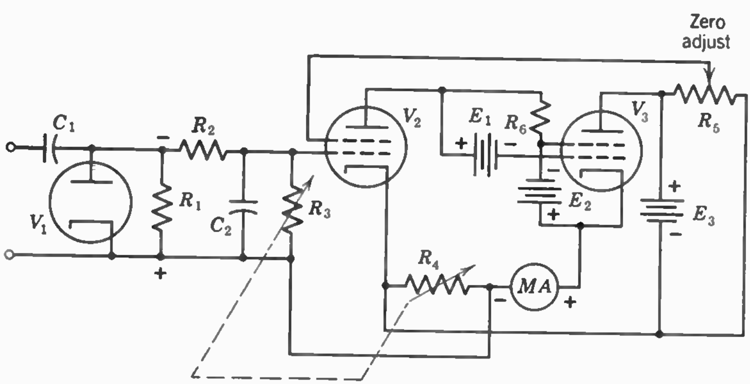


FIG. 8. Diode rectifier with d-c amplifier to enable the use of a more rugged indicating instrument.

V_1 having a very low input capacitance is used as the rectifier. Because of this low capacitance the calibration is essentially independent of frequency up to about 30 Mc and is still usable up to 100 Mc.

The second type of meter places a broad-band amplifier at the input, followed by the detector. Thus the rectifier is always supplied with a comparatively large voltage, making its operation nearly linear. The amplifier normally employs a large amount of inverse feedback. The Hewlett-Packard type 400A meter is a typical example. Figure 9 shows the essential features of this meter.

21.7. Calibration and use of vacuum-tube voltmeters. A well-designed VTVM may be calibrated at 60 cycles, and this calibration will be accurate up to 100 kc or several megacycles, depending upon the particular instrument.

Most tube voltmeters incorporate a series condenser in the input leads as shown in Fig. 9. This condenser blocks any d-c

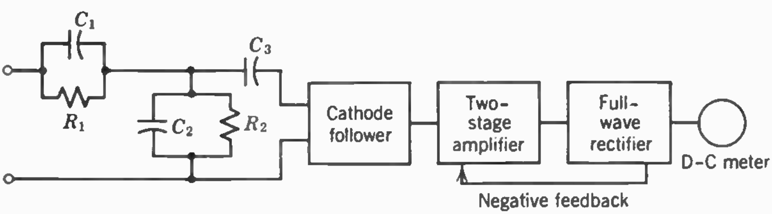


Fig. 9. Amplifier followed by diode rectifier.

voltages in the input signal and allows the meter to respond only to a-c voltages.

There are a few important considerations if the proper use is to be made of tube voltmeters. These include the factors inherent in the tubes and meter circuits which may cause erroneous readings, as well as factors over which the user has direct control. Some of the more important factors follow:

1. Series resonance
2. Transit time error
3. Input impedance
4. Waveform error
5. Turnover
6. Grounding
7. Zero drift

Series resonance occurs between the inductance of the input leads and the input capacitance of the first tube and associated parts. This series resonance usually causes the reading to be high near the upper usable frequency range of the instrument, and to decrease rapidly as the frequency is raised still further.

The **transit time** of the electrons may also enter the picture, since some of the electrons leaving the cathode near the end of the cycle may not reach the plate before the cycle reverses. This error depends both on the voltage being measured and on the frequency, and causes the readings to be lower than the true value.

The **input impedance** usually becomes lower as the frequency is raised. Because the input capacity remains essentially constant regardless of the input frequency, but its reactance decreases as frequency increases, the input impedance is reduced.

An erroneous reading may result if the **waveform** of the signal voltage is not sinusoidal since practically all tube voltmeters are calibrated with sinusoidal voltages. The amount of error depends both on the waveform and on the type of meter circuit, that is, whether it is full-wave, half-wave, or square-law, and whether it is peak- or average-reading.

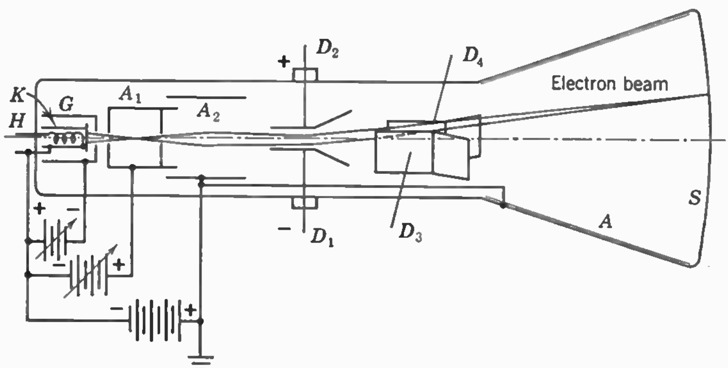
Reversing the input terminals sometimes changes the reading. This effect is known as **turnover** and is greatest in peak voltmeters. It is present to a certain extent in the half-wave square-law type but is non-existent in true full-wave square-law and linear instruments. When turnover is present, averaging the direct and reversed-polarity readings will give approximately, but not necessarily, the correct value. The presence of turnover indicates that the amplitude of the positive voltage peaks is different from the amplitude of the negative peaks.

The **ground** terminal must be connected to the ground or low potential point of the circuit whose voltage is being measured. Unless the ground terminals are properly connected, the capacitance between the VTVM case and the test circuit may cause considerable error, particularly at frequencies above a few thousand cycles. This kind of error is different from the turnover effect discussed in the preceding paragraph.

Sufficient time should be allowed after a VTVM is turned on before the adjustment is made which balances out the residual current in the indicating meter. Otherwise the residual current may change as the instrument continues to warm up to the final operating temperature. The residual current may also change because of aging of tubes and condensers, but this is a long-time effect. Both these changes in the residual current are called **zero drift.**

CATHODE-RAY OSCILLOGRAPHS

21.8. Cathode-ray tubes. The cathode-ray oscillograph is probably the most versatile of all electronic instruments. It has



A	Aquadag coating
A_1	Focusing electrode (300 to 500 volts)
A_2	Accelerating electrode (1500 to 2000 volts)
D_1, D_2	Vertical deflecting plates
D_3, D_4	Horizontal deflecting plates
G	Control or intensity grid (0 to −500 volts)
H	Heater
K	Cathode
S	Fluorescent screen

FIG. 10. An electrostatically deflected cathode-ray tube.

many uses, and new ones are being found constantly. The heart of the cathode-ray oscillograph is the cathode-ray tube.

Within this tube, a beam of electrons is focused on a fluorescent screen at the end of the tube. When the beam strikes the fluorescent material, visible light is produced. The beam may be deflected by causing the electrons to pass between parallel plates on which potentials are placed or by causing the beam to pass through a magnetic field. These are called, respectively, **electrostatic** and **magnetic** deflection.

A diagram of an electrostatically deflected cathode-ray tube is shown in Fig. 10. Electrons are produced at the cathode and are accelerated along the axis of the tube by the high voltages on electrodes A_1 and A_2. The electrode G is called the control or intensity grid and acts in a manner similar to the grid of a triode in that it controls the number of electrons passing on to the screen. Varying the voltage on this grid varies the intensity of the light emitted from the screen. The small holes in G and A_1 are for the purpose of forming the electron stream into a beam and for stopping any electrons which are too far off the axis of the tube. Besides accelerating the electrons A_1 and A_2 together form an electron lens system which focuses the electron stream into a thin pencil beam at the screen. The potential of A_1 is made variable so that the best focus can be obtained at the screen.

Deflection of the electron beam is accomplished by the vertical deflecting plates D_1 and D_2, and by the horizontal deflecting plates D_3 and D_4. When a voltage is placed on D_1 and D_2 with the polarity shown in Fig. 10, the beam will be deflected upward.

The screen at the end of the tube is coated on the inside with a fluorescent phosphor which produces visible light when the electron beam strikes it. The fluorescent material most widely used for tubes in cathode-ray oscillographs has a green color (when excited by the electron beam), while a white material is generally used in tubes designed for television. Other phosphors are available for special services, including a blue one for photographic use and some which glow for an appreciable time after the beam of electrons has moved on. The latter type is called a **long-persistence phosphor.**

The Aquadag coating on the inside of the large part of the tube is a conducting material which serves to return excess electrons from the screen area to ground.

In cathode-ray tubes which employ magnetic deflection the deflecting plates shown in Fig. 10 are replaced by deflecting coils as shown in Fig. 11. Only one pair of deflecting coils is shown in the diagram. Current in these coils sets up a magnetic field across the tube, causing the electron beam to be deflected. For the set of deflecting coils shown, the deflection will be perpendicular to the plane of the paper. The electron beam may be focused with a coil placed around the neck of the tube. Such a focusing coil is shown in Fig. 11. Some tubes employ combinations of electrostatic focusing and magnetic deflection.

Magnetically deflected tubes suffer from one distinct disadvantage. Negatively charged ions in the tubes are directed toward the screen in the same manner as electrons. However, the magnetic field gives them comparatively little deflection, and all the ions, which are about 1800 times as heavy as electrons, strike the screen at or near the center, causing a black spot by burning the phosphor. To avoid this trouble, many tubes have built-in ion traps which eliminate the difficulty but add to the

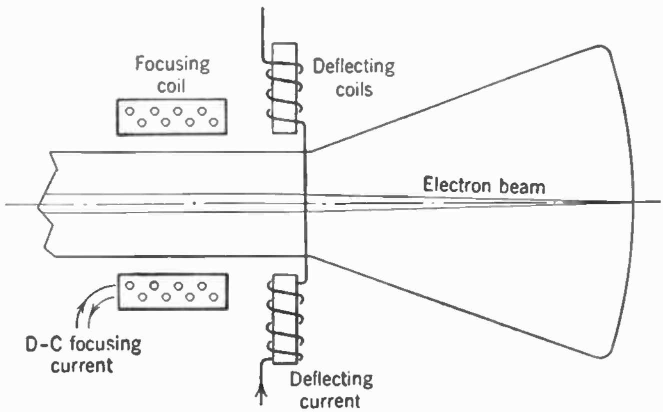


Fig. 11. Cathode-ray tube with magnetic focus and deflection.

complexity of the tube. Magnetically deflected tubes are widely used in television receivers since the tubes may be made somewhat shorter for the same screen size, resulting in a reduction of the required depth of cabinets for television receivers.

21.9. Cathode-ray oscillographs. Cathode-ray oscillographs have very wide use in observing waveforms, comparing phase relations, checking modulation, and for numerous other purposes.

A typical cathode-ray oscillograph has two amplifier channels connected to the horizontal and vertical deflection plates, respectively. These amplify the signals which are to be observed so that small signals will cause adequate deflection on the screen. Actually, for the observation of waveforms, the horizontal amplifier channel is normally used to amplify a sawtooth "time base" voltage, while the signal to be observed is connected to the vertical amplifier channel.

21.10. Time bases. If a sine wave of voltage is applied to the vertical set of plates, the beam will be deflected up and down from its normal position. A straight vertical line will be seen on the screen. If, however, the beam is moved to the right at the same time that the sine wave of voltage tends to make it move upward, then a pattern which is a sine wave will be traced on the screen. If, after the beam has reached a point near the right edge of the screen, the horizontal deflecting voltage suddenly

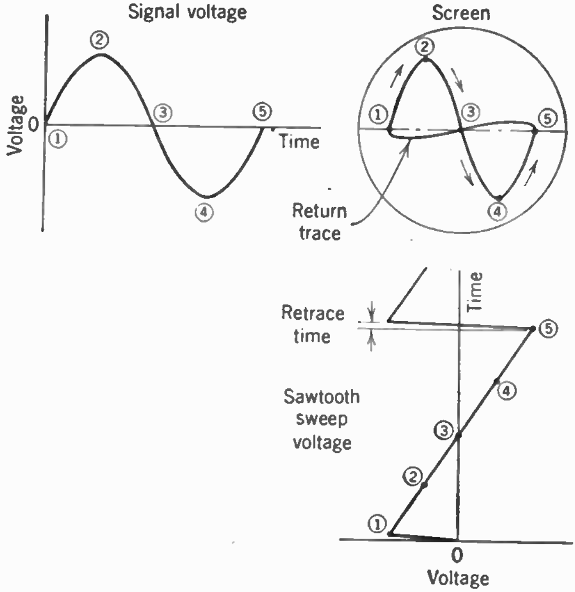


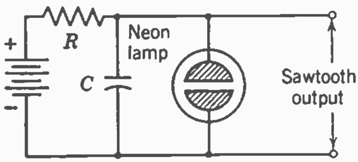
Fig. 12. Use of sawtooth time base for observing waveforms of a signal voltage.

changes so that it returns the beam to its starting point, successive traces of the sine wave will appear on the screen. The desired form of horizontal deflecting voltage, then, is one which will move the beam horizontally at a uniform speed so that, at any instant of time, say a particular point on an a-c cycle, the beam will be at a particular spot on the screen. The sine wave of voltage on the vertical plate will be "spread out" by the horizontal deflecting voltage. This voltage is called a **time base** since it enables the operator to get the time relations of the pattern he is seeing on the screen. This action is shown in Fig. 12.

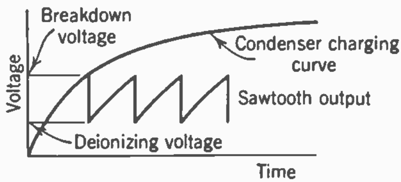
There are many forms of time bases and time-base circuits. In all of them, the beam is swept across the screen (sometimes around the screen in a circle or spiral), and then the beam is brought back to its starting point. Sometimes the return trace is blanked out by driving the intensity grid highly negative during the return period.

21.11. Sawtooth generators. The sawtooth waveform is widely used for time bases since it is linear; that is, it sweeps the beam across the screen at a uniform rate, then returns the beam to the starting point very quickly.

The ideal sawtooth waveform can be approached quite closely by means of a simple relaxation oscillator, the circuit of which is shown in Fig. 13. Direct current flows through R, charging up



(a) Simple Circuit for Producing Sawtooth Time Base



(b) Sawtooth Waveform

Fig. 13. Sawtooth generator.

the condenser until the voltage across the neon lamp is sufficient to cause the gas to break down, discharging the condenser. The charge and discharge cycle is then repeated at a rate depending on the size of R, C, and the d-c voltage.

As the condenser in the above oscillator charges exponentially rather than linearly, the rate of charging is not quite constant. It can be made very nearly linear, however, by careful design. In special applications, the charging rate may be made constant by replacing the resistor with a constant-current device, such as a pentode vacuum tube. Then, since the charging current must be constant, the rate of charging of the condenser must be uniform, and the voltage across the condenser will increase at a uniform rate. A type 884 or similar grid-controlled gaseous discharge tube (thyratron) is generally used in place of the neon bulb, as the frequency of the oscillations may be controlled to a certain extent by the application of a small voltage of the proper frequency to the grid. The frequency of the oscillator employing the neon bulb cannot be controlled in this manner.

A circuit of a sweep-frequency oscillator using a constant-current pentode is shown in Fig. 14. Such a circuit is capable of producing a waveform very close to that of an ideal sawtooth wave.

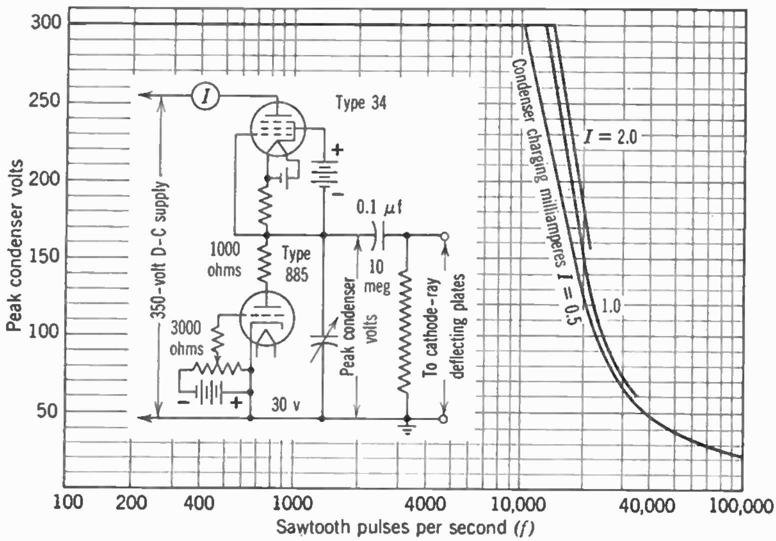


FIG. 14. Complete sweep-voltage oscillator circuit and its characteristic. (From RCA.)

21.12. Oscillograph controls. A typical front panel of an oscillograph is shown in Fig. 15 and a block diagram of the internal connections in Fig. 16. Although there is some variation from one oscillograph to another, the basic controls are much the same. Some of the controls which are common to nearly all oscillographs will be described. Instruction manuals should be referred to for the operation of a particular instrument.

After a preliminary warm-up time, after the instrument is turned on, the spot is positioned on the screen with the VERTICAL POSITION and the HORIZONTAL POSITION controls. (An intense spot should not be allowed to remain on one part of the screen for more than a few seconds since there is danger of burning the phosphor coating.) The signal is then connected between the VERTICAL INPUT terminals and GROUND. The INTENSITY and FOCUS knobs are adjusted to give a clean-cut trace. Along with this operation, the COARSE FREQUENCY switch must be set to the approximate sweep frequency and the FINE FREQUENCY knob adjusted to cause the trace on the screen to remain approximately

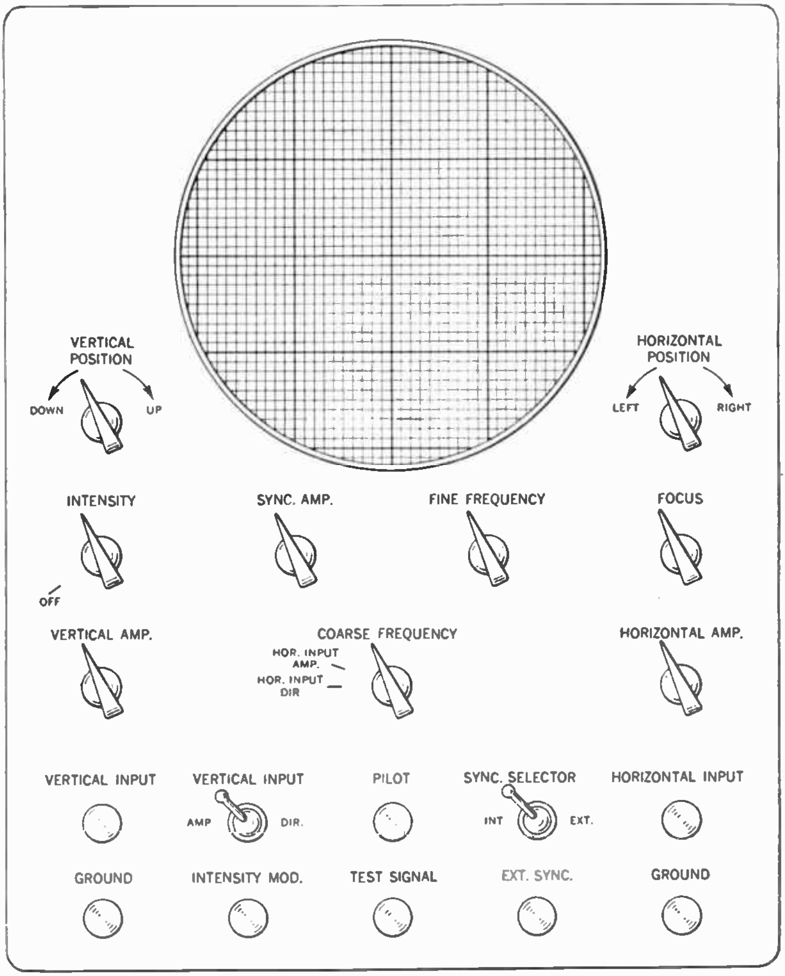


FIG. 15. Front panel of a typical cathode-ray oscillograph.

stationary. The SYNC SELECTOR switch is set to INT and the SYNC AMP control adjusted to hold the trace stationary. It is important that the FINE FREQUENCY control be used to make the major adjustment which keeps the pattern stationary and that the

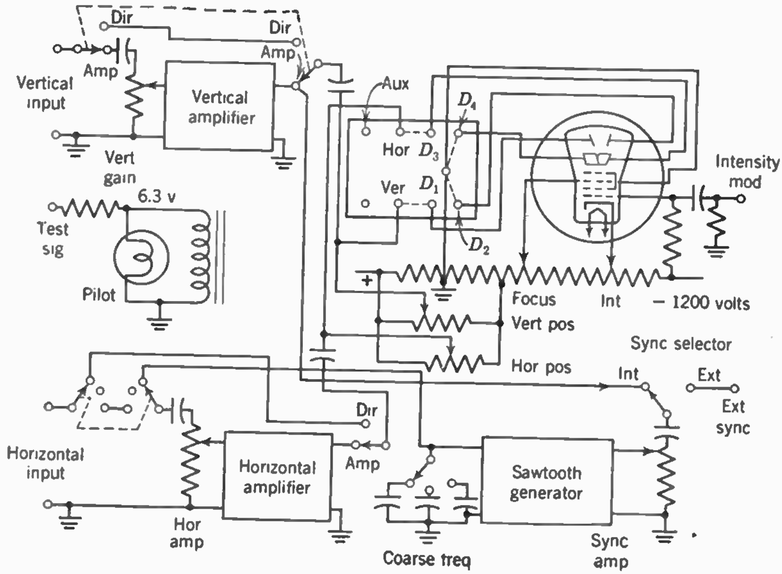


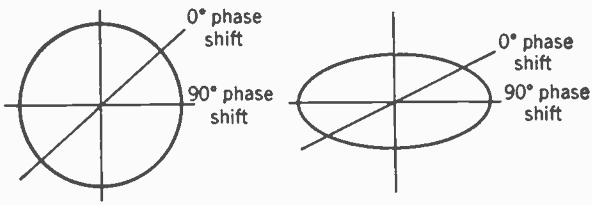
FIG. 16. Circuit of a typical cathode-ray oscillograph.

SYNC AMP control not be turned too high. Otherwise, distortion of the signal waveform is likely to result.

The horizontal width of the pattern may be adjusted by the HORIZONTAL AMP control. If the sweep generator is to be synchronized by some signal other than the vertical input signal, the other signal is connected between the EXT SYNC terminal post and a GROUND terminal, and the SYNC SELECTOR switch is set to EXT. Other adjustments are made as before.

An external signal may be applied to the horizontal deflection plates by connecting the signal to the HORIZONTAL INPUT terminals, switching the COARSE FREQUENCY control to HOR INPUT AMP (or to HOR INPUT DIR if no amplification is desired), and adjusting the amount of amplification of the signal with the HORIZONTAL AMP control knob. For this service the sweep generator is not used.

21.13. Measuring phase shift. The most common use of a cathode-ray oscillograph is for the display of the waveform of a voltage as a function of time. This use was discussed in Sect. 21.10 and illustrated in Fig. 12. Other important results are obtained if the sawtooth voltage on the horizontal deflecting plates is replaced with a sine wave of voltage of the same frequency as the signal applied to the vertical plates. For example, the vertical plates might be connected across a coil and resistor in series, and the horizontal plates across the resistor alone. The voltage across the resistor has the same waveform and phase as the current through the resistor, so we have a means of comparing the phase shift between the voltage across and the current through the coil and resistor. All that is necessary is to observe the resulting pattern.



(a) Horizontal and Vertical Voltages Equal

(b) Horizontal Voltage Twice as Large as Vertical Voltage

FIG. 17. Patterns on oscillograph screen when sine waves are applied to both vertical and horizontal plates.

If the voltages applied to the vertical and horizontal plates are in phase, the resulting pattern will be a straight line as shown in Fig. 17. A circle results if the phase shift is 90° and the gain of the horizontal and vertical amplifiers is such that the magnitude of the voltages applied to the two sets of deflecting plates is the same. If the signals applied to the deflecting plates are not the same size the pattern is an ellipse. Patterns obtained for other phase-shift angles are ellipses with tilted axes.

The patterns obtained for various phase-shift angles are shown in Fig. 18. The angle of the phase shift may be calculated from the following equation:

$$\text{Sine of angle} = \sin\theta = \frac{Y\text{-intercept}}{\text{Maximum } Y\text{-position of pattern}}$$

Although the above equation is correct even though the voltage applied to the two sets of deflecting plates is different in magnitude, the gain controls are usually adjusted to give approximately equal horizontal and vertical deflections. It is extremely important that the spot be accurately centered on the screen before measurements are made. The value of the Y-intercept and the maximum Y-position of the trace may be read from the calibrated screen on the face of the oscillograph.

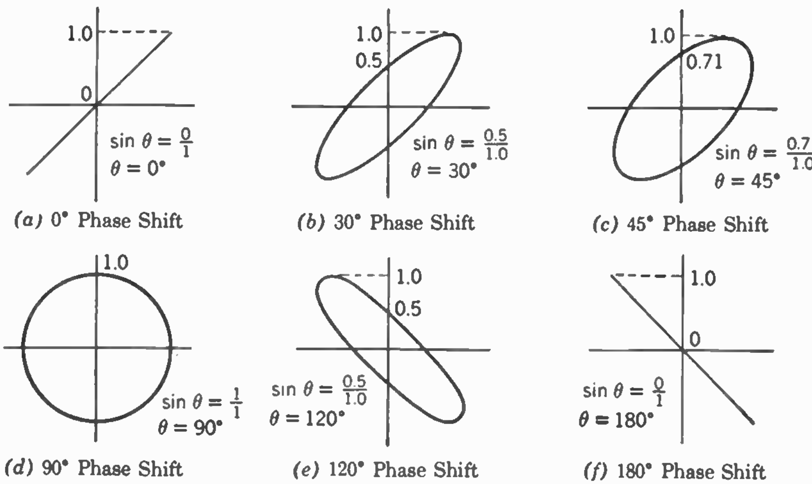


FIG. 18. Patterns produced for various phase differences between horizontal and vertical voltages.

21.14. Comparing frequencies. Two frequencies may be compared on an oscillograph by placing one signal on the vertical plates and one on the horizontal plates. The pattern which is observed consists of a number of loops, depending on the frequencies of the two signals. The pattern is known as a **Lissajous pattern.**

Suppose that a signal of known frequency is applied to the horizontal input terminals and a signal of double frequency is applied to the vertical input terminals. The resulting pattern on the screen is shown in Fig. 19(b). Other frequency combinations are also shown in this figure.

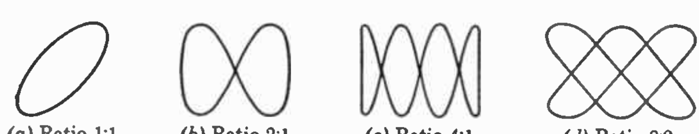


FIG. 19. Lissajous patterns resulting from application of signals of different frequencies on horizontal and vertical plates.

The general method of calculating frequency ratios by the use of Lissajous patterns is shown in Fig. 20. The relation is given by

$$\frac{\text{Frequency on horizontal input}}{\text{Frequency on vertical input}} = \frac{\text{Number of loops touching } AB}{\text{Number of loops touching } BC}$$

For the pattern shown the ratio is 5:2.

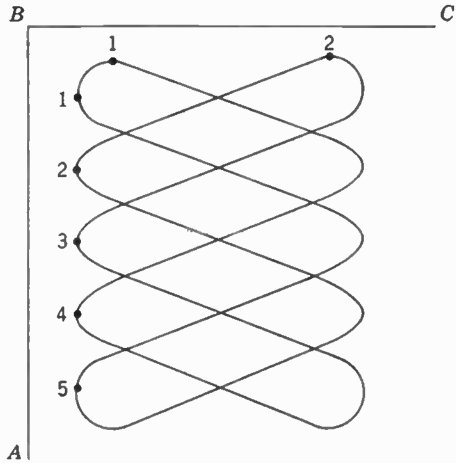


FIG. 20. Method of calculating frequency from Lissajous pattern.

Lissajous patterns give a method of calibrating an oscillator against a source of known frequency. For example, if a 1000-cycle signal source is available, the dial of an oscillator might be calibrated at a number of points from about 100 to 10,000 cycles. It is usually difficult to count the loops if the ratio is greater than about 10:1.

Experiment 1. Connect a 60-cycle test signal to the vertical input terminals of an oscillograph, and an a-f oscillator to the horizontal input terminals. Adjust the oscillator frequency to obtain Lissajous patterns for frequency ratios of 1:2, 1:1, 2:1, 3:1, etc., up to a ratio of about 10:1. For each point at which a pattern is obtained, note the frequency marked on the oscillator dial. Plot a calibration curve showing actual frequency against the frequency indicated on the oscillator dial. Even though this experiment assumes that the line frequency is accurately 60 cycles, which may not be quite true, it illustrates a very practical use of Lissajous patterns. A typical calibration chart is shown in Fig. 21.

21.15. Modulation monitor. The modulated output waveform of an amplitude-modulated transmitter may be observed

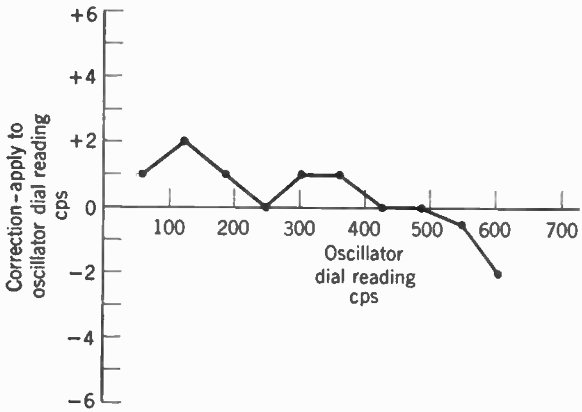


FIG. 21. Calibration chart for an oscillator. Unlike most graphs, calibration curves are often drawn as a series of straight lines connecting experimental points.

on an oscillograph by coupling a small coil very loosely to the output tank circuit and connecting the terminals of the coil to the vertical input terminals of the oscilloscope. The sweep generator is used in the same manner as for the normal observation

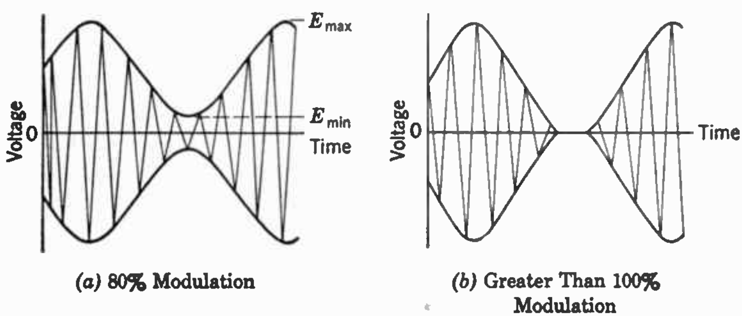


FIG. 22. Amplitude-modulated carrier as seen on a cathode-ray oscillograph.

of voltage waveforms, with the sweep frequency set to the frequency of the a-f modulating voltage. The pattern obtained is shown in Fig. 22. The percentage modulation is given by

$$\frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \times 100\%$$

OTHER ELECTRONIC INSTRUMENTS

21.16. Electronic switch. A device frequently used with cathode-ray oscillographs is the **electronic switch,** which makes possible the simultaneous observation of two voltages on a single screen. A block diagram of an electronic switch is shown in Fig. 23. It consists essentially of two amplifiers, alternately

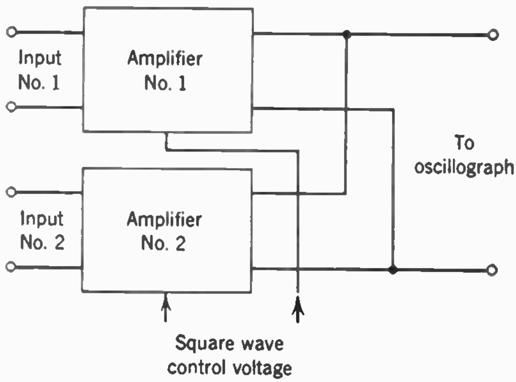


Fig. 23. Electronic switch for oscillograph.

biased at cut-off by the application of a square wave of voltage. Both amplifiers are connected to one set of oscillograph plates. The two voltages to be compared are impressed on the amplifiers, and the trace on the screen corresponds first to one voltage and then the other, depending on which amplifier is in operation. If each amplifier operates half the time, then each voltage will appear on the screen half the time.

In addition to presenting two voltage waveforms for simultaneous observation, the electronic switch may also be used to determine the phase shift between two voltages. One or both of the voltages may be derived from currents if a small resistance is placed in series with the lead in which the current flows and the voltage across this resistor is applied to one of the amplifiers of the electronic switch. This series resistance must be much smaller than the resistance of the circuit in which it is placed so as not to disturb the normal circuit operation.

Recently developed double-trace oscillographs are taking the place of electronic switches in some applications. These instruments incorporate a tube which has two separate electron guns and deflecting systems. Thus, two traces may be placed on the screen and the waveform of the two traces compared, or the phase shift between the two signals from which the traces were derived may be measured. Cathode-ray tubes with more than two electron guns have been built, but they are not in general use.

21.17. Signal generators. The r-f signal generator is an instrument which provides a convenient source of modulated or unmodulated r-f voltage. The basic elements of an a-m signal generator are shown in Fig. 24. The frequency and amplitude

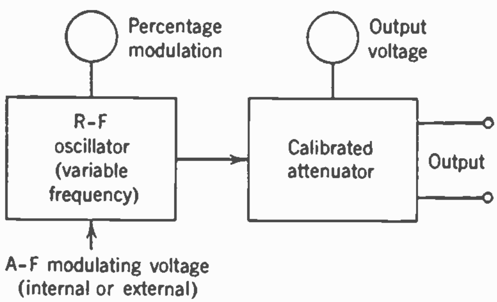


FIG. 24. Basic elements of an a-m r-f signal generator.

are variable over very wide limits, adding to the usefulness of the device. It is, in fact, a miniature radio transmitter with an accurately measured output voltage at a known frequency.

The signal generator consists essentially of a well-shielded oscillator which can be modulated, together with an attenuator for varying the output. Many of the more expensive signal generators are provided with output meters and calibrated attenuators so that the magnitude of the output voltage may be accurately known. The output may be continuously variable from as low as 1 μv to as high as 1 volt. Usually the percentage modulation is also variable over wide limits.

Radio-frequency signal generators may be frequency-modulated rather than amplitude-modulated. The center frequency of these signal generators is usually variable over a rather limited range, such as over the f-m band of frequencies. The frequency deviation is usually variable, and provisions are often made for using different modulating frequencies. In most other respects these generators are similar to a-m signal generators.

21.18. A-f oscillator. The a-f oscillator provides a convenient source of a-f voltage for many test purposes. As in the r-f signal generator, the frequency and output are continuously variable over rather wide limits. In a good instrument the output voltage stays nearly constant for a particular setting of the output control, even though the frequency is varied. The output is very nearly a sine wave.

A **beat-frequency oscillator** consists of two r-f oscillators operating at different frequencies as shown in the block diagram

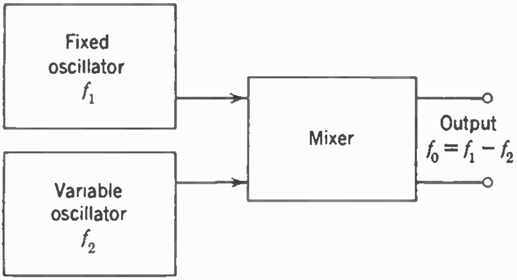


FIG. 25. Block diagram of a beat-frequency oscillator.

of Fig. 25. When these two frequencies are mixed, a difference or "beat" frequency is produced. In this instrument the frequency of one r-f oscillator is fixed and that of the other is variable. The usual practice is to make the frequency of the fixed oscillator about five times the maximum beat frequency desired. The frequency range may be rather wide and usually extends from around 20 cycles to 20 or 30 kc. Since beat-frequency oscillators tend to drift during the warm-up period, provisions are usually made so that the frequency may be calibrated against the 60-cycle line voltage or other source of known frequency.

21.19. *Q*-Meter. The *Q* of coils (Sect. 8.7) is often a controlling factor in the design and operation of r-f circuits. A high-*Q* circuit is required if good waveform, high efficiency, high gain, and good selectivity are to be attained. While the *Q* of a circuit element may be determined in several ways, most methods require considerable calculation after the necessary measurements are made. The *Q*-meter is an instrument which measures *Q* directly. It may also be used to determine the resistance of a circuit and the capacitance of an unknown condenser.

The fundamental circuit of a Q-meter is shown in Fig. 26. The oscillator, set to the desired frequency, furnishes a current I_1 to the measuring circuit. The coil whose Q is to be measured is connected to terminals 1 and 2. Condenser C tunes the coil to series resonance. The voltage E_c across the condenser is measured by a vacuum-tube voltmeter. At resonance, E_c is proportional to the Q of the coil, and the scale of the meter is calibrated

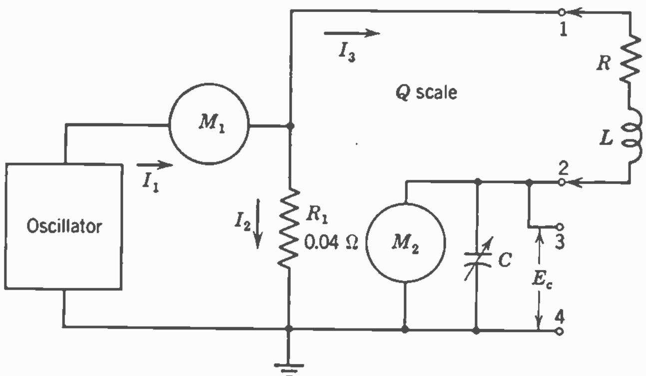


FIG. 26. Q-meter.

in units of Q. Since E_c is also directly proportional to I_1, this current is made adjustable to extend the range of Q values which can be measured.

The capacitance of an unknown condenser may be measured by tuning a coil to resonance as before, then connecting the unknown condenser to terminals 3 and 4, thus placing it in parallel with condenser C. The value of C is then reduced until the circuit is again in resonance as indicated by a maximum reading of the Q-meter. The value of the unknown capacitance is calculated by taking the difference in the capacitances of C before and after the unknown condenser was connected to the circuit. This method is limited to condensers whose capacitance is smaller than that of C, but may be extended to larger condensers by using a small condenser of known value in series with the unknown condenser and then using the methods of Sect. 6.12 to calculate the capacitance of the unknown condenser. The Q-meter is a very versatile instrument and may be adapted to many other measurements.

Example 1. Suppose a coil is measured by the use of a Q-meter and the following values are obtained:

Frequency	2.4 Mc
Q	240
Condenser, C	120 $\mu\mu$f

The inductance of the coil may be calculated from the equation

$$2\pi fL = \frac{1}{2\pi fC}\left(\text{or } L = \frac{1}{(2\pi f)^2C}\right) \text{ at resonance}$$

$$\begin{aligned} L &= \frac{1}{(2\pi \times 2.4 \times 10^6)^2 \times 120 \times 10^{-12}} \\ &= 36.7 \times 10^{-6} \text{ henry} \\ &= 36.7\ \mu\text{h} \end{aligned}$$

The r-f resistance of the coil is, from $Q = 2\pi fL/R$,

$$R = \frac{2\pi fL}{Q} = \frac{2\pi \times 2.4 \times 10^6 \times 36.7 \times 10^{-6}}{240} = 2.31 \text{ ohms}$$

Problem 1. The Q of a coil is found to be 150 when the frequency is 1.0 Mc and the tuning capacitance is 200 $\mu\mu$f. Find (*a*) the inductance of the coil, and (*b*) the r-f resistance of the coil. If the coil is used with a 100-$\mu\mu$f condenser, find the resonant frequency.

Problem 2. Two condensers are connected in series and their combined capacitance measured with a Q-meter and found to be 200 $\mu\mu$f. If the capacitance of one of the condensers is known to be 220 $\mu\mu$f, what is the capacitance of the other?

22 · Transients and Wave Shaping Circuits

In many radio, television, and radar circuits advantage is taken of the fact that the final value of current (or voltage) in a circuit is not attained instantaneously after voltage (or current) is applied. The time required for the final values to be established is used for time measurements, selective switching, wave shaping, and many other purposes. The voltage and current conditions in a circuit during the time in which final values are being established are called **transient** conditions. The final values of current and voltage are called **steady-state** values.*

22.1. The nature of transients. Thus far in this book the analysis of circuit conditions has assumed that the exciting voltages, whether direct or alternating, have been applied to the circuit long enough so that the final conditions have been reached and so that all transient effects have disappeared or become entirely negligible. It is important now to look at the period in which transients are the important feature of the circuit phenomena.

Only in a circuit containing resistance alone are the final currents and voltages attained immediately when a voltage is applied to the circuit. If a circuit contains a condenser, then the condenser tends to prevent the voltage across itself from rising to its final value and in so doing passes a rather large, rapidly decaying, transient current. On the other hand, a circuit containing resistance and inductance tends to keep the current from reaching its final value and produces a large transient voltage (counter-emf) in the process.

* The term "steady state" should actually be applied only to d-c circuits in which voltage and current have reached their final values. However, by common usage, the term is also applied to circuits in which sine waves of voltage and current are used. The latter condition is sometimes called "quasi-steady state."

Transients may appear in circuits energized with either direct or alternating voltages. The general nature of the transient phenomena can, however, be studied on a d-c basis. This has the advantage of being much less complicated than a study based on a-c voltages and currents. Furthermore, most of the tube circuits of interest here can be analyzed on a d-c basis.

22.2. Transients in *RC* circuits. Consider a simple case of R and C in series with a battery and switch. Let a fast-acting

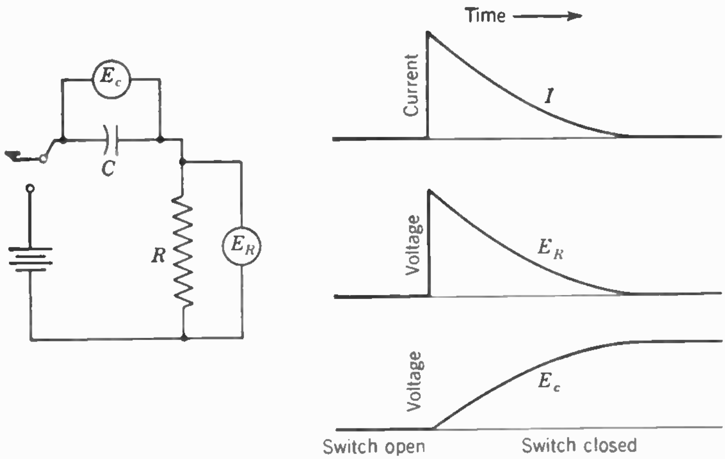


FIG. 1. When the key is closed, current rushes into the capacitor, building up a voltage there which opposes the battery voltage and ultimately equals it. Then current around the circuit ceases to flow, the condenser is charged, and, since there is no current, there is no voltage across R. This chain of events may occur very quickly (in a microsecond or less) or may require much longer time if C and R are large.

voltmeter be placed across R, and another across C. This circuit is shown in Fig. 1. What happens when the switch is closed?

Since there is no initial charge on C, there is no voltage across it. At the instant of closing the switch, the full battery voltage must, therefore, be impressed across R. Current flows through R, and must also flow into C and charge it. The charge which the current carries to the condenser builds up a voltage across the condenser which opposes the battery voltage, and the current drops off gradually. When the condenser voltage becomes equal to the battery voltage, current flow ceases, and all the voltage drop appears across the condenser; none appears across the resistor. The voltage across the resistor falls off at the same rate as the condenser voltage rises.

Consider, as a second case, the circuit of Fig. 2. This is the same as the circuit of Fig. 1 except that the battery has been removed. Suppose that the condenser has been charged to some voltage (say the battery voltage of the first case), then the switch is closed. What are the transient conditions as the condenser

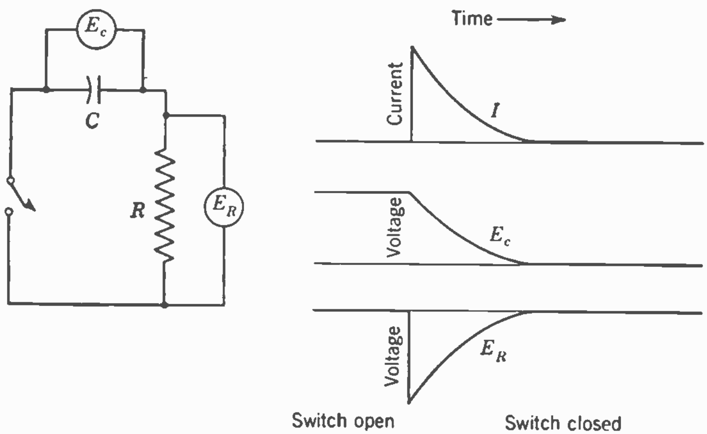


FIG. 2. In this figure the condenser is charged. When the key is closed, current flows through the circuit, building up a voltage across R which opposes E_c. When the charge in C has been dissipated, current ceases and E_c and E_R both become zero.

discharges through the resistor? At the instant the switch is thrown, there is but one voltage in the series circuit, that existing across the condenser. An instant later current begins to flow and a voltage drop appears across R. This voltage drop must be opposed to the condenser voltage according to Kirchhoff's law. The current flowing through the resistance produces a power loss in this resistance. The energy that was stored in the condenser in the form of potential energy is dissipated and soon current flow ceases. Maximum current flows at the instant the switch is closed; maximum voltages exist across condenser and resistor at this instant. The sum of the two voltages must equal zero at all times; that is, $E_C - E_R = 0$ and at all times the voltages are related by Ohm's law.

22.3. Transients in *RL* circuits. When voltage is impressed on a series circuit made up of resistance and inductance, the voltage across the inductance rises to the full battery voltage instantly, but the current through the circuit builds up slowly. Therefore, the voltage drop across the resistance builds up slowly (at the same rate as that at which current increases). The voltage across the inductance decreases at the same rate as that at which the current increases since there is induced voltage across an inductance only when the current through the inductance changes and it is proportional to the rate at which the current changes.

If the *RL* circuit is now shorted, the impressed voltage is reduced instantly to zero; the current decreases at a rate depending upon R and L; the voltage across the resistance decreases at the same rate; and the instantaneous voltage across the inductance, which has a polarity opposite that of the original impressed voltage, decreases at the same rate as the current decreases.

22.4. Calculation of transients. The rapidity with which voltages and currents change from their initial to their final values under transient conditions is determined by the value of the circuit components. In an *RC* circuit, the smaller the value of R for a fixed C, the faster the transient condition passes. The reason is that a large R limits the current and the condenser charges (or discharges) at a slower rate. In an *RL* circuit, on the other hand, the larger the value of R for a fixed value of L, the faster the transient condition passes. The nature of inductance is to keep current from changing, and thus the transient passes slowly if R is small. However, if R is large, the action of the inductance is overcome to some extent, and, if R is made very large, the transient is very short.

The most convenient way in which to make calculations involving transients is to use the concept of **time constants.** A time constant in either an *RC* or *RL* circuit is the time required for current or voltage to make 63.2 per cent of the total change from initial to final conditions. That is, there is only a change of 36.8 per cent remaining. The formulas for the two cases are

$$T = RC \quad \text{(series } RC \text{ circuit)}$$

$$T = R/L \quad \text{(series } RL \text{ circuit)}$$

where T is in seconds, R is in ohms, C is in farads, and L is in henries.

A transient disturbance is considered as ended after 7 time constants, for in this time it is 99.91 per cent completed.

Values of voltage or current at any time during a transient disturbance may be calculated by the use of the curves in Fig. 3.

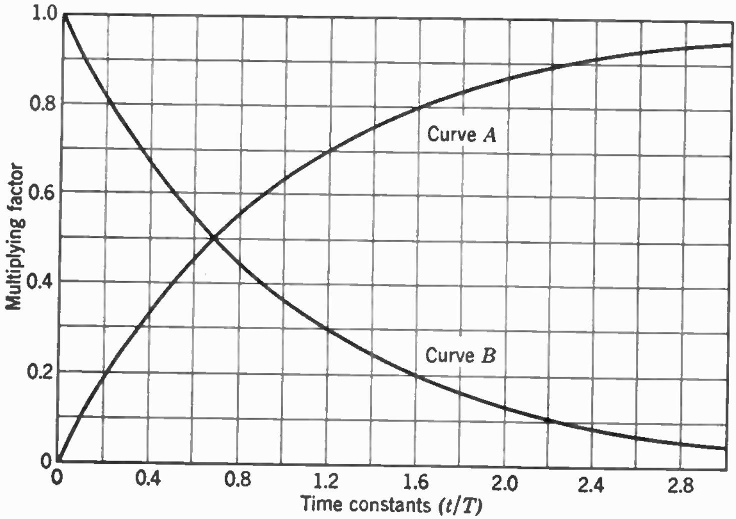


FIG. 3. Curves for use in calculating transients. (See text.)

These curves are specifically for the RC circuits of Figs. 1 and 2 and for the RL circuits discussed in Sect. 22.3. They may, however, be readily adapted to many other circuit arrangements. Along the horizontal axis are plotted time constants. Thus, if the time constant for a particular circuit is 0.01 sec, and circuit conditions after 0.005 sec are required, the number of time constants is 0.005 ÷ 0.01 or 0.5 time constant. Along the vertical axis are plotted "multiplying factors." These are applied to known initial or final values of voltage or current to determine their values at intermediate times. The use of the curves will be illustrated in the following examples:

Example 1. A 0.1-μf condenser, 10,000-ohm resistor, and switch are connected in series and across a 100-volt battery (see Fig. 1). What voltages are across the condenser and resistor 0.002 sec after the switch is closed? What is the current at this time? After how long may the transient be considered to be complete? The condenser is initially uncharged.

Solution: The time constant for this circuit is

$$T = RC = 10{,}000 \times 0.1 \times 10^{-6} = 10^{-3} \text{ sec}$$

The initial value of the current is

$$I_0 = \frac{E}{R} = \frac{100}{10{,}000} = 0.01 \text{ amp or } 10 \text{ ma}$$

This current will fall gradually from this initial value. The initial voltage across R is 100 volts. Like the current, this voltage decreases gradually. Both are computed by the use of Curve B.

For 0.002 sec $t/T = 0.002/0.001 = 2$ time constants. From Curve B, the multiplying factor is 0.13. Then

$$E_R = 100 \times 0.13 = 13 \text{ volts}$$

$$I = 10 \times 0.13 = 1.3 \text{ ma}$$

The voltage across the condenser increases from zero; it may be computed at any time from Curve A. At 2 time constants the multiplying factor is 0.86, and the voltage across the condenser at this time is

$$E_C = 100 \times 0.86 = 86 \text{ volts}$$

The transient is considered complete in 7 time constants, or in $7 \times 0.001 = 0.007$ sec.

Example 2. A 0.1-henry coil and 10,000-ohm resistor are connected in series with a switch and 200-volt battery. What final value of current flows? How long is required for the current to reach 50 per cent of its final value?

Solution: The time constant is

$$T = \frac{L}{R} = \frac{0.1}{10{,}000} = 10^{-5} \text{ sec or } 10\ \mu\text{sec}$$

The final current is $200 \div 10{,}000 = 0.02$ amp or 20 ma. From Curve A (the current rises from zero to its final value) the current reaches 50 per cent of its final value in 0.7 time constant, or in $0.7 \times 10 = 7\ \mu$sec.

Problem 1. A 0.002-μf capacitor is charged to a potential of 250 volts, then discharged through a 100,000-ohm resistor. What is the time constant of the combination? What is the initial current? What is the voltage across the capacitor after 40 μsec?

Problem 2. A 10-henry coil and 100-ohm resistor are connected in series with a 100-volt battery. After the current has reached a steady value, the battery is suddenly removed and the RL combination is short-circuited. What is the current after 0.16 sec? What is the voltage across the coil at the same time?

22.5. Differentiating and integrating circuits. In many applications it is necessary to produce a wave which has a specified shape. A common starting point is a square wave, a wave which has very steep sides and a flat top. Square waves can be passed through various circuits to obtain many other waveforms.

Suppose that a square wave is applied to the *RC* circuit of Fig. 4. If the time constant of the *RC* circuit is *short* compared to the period of the square wave,* the condenser will charge very rapidly as the square wave rises to its maximum positive voltage. In charging, the condenser passes current for a short period of

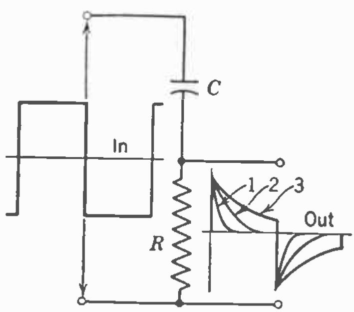


FIG. 4. *RC* differentiating circuit.

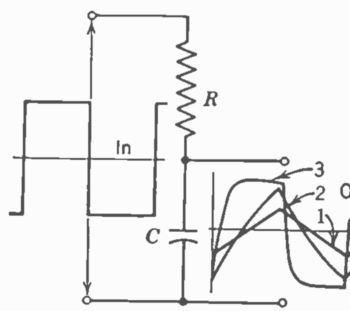


FIG. 5. *RC* integrating circuit.

time, and during this time produces a voltage drop across the resistance. During the next half cycle the voltage across the condenser changes to the opposite polarity. Thus the output voltage consists of a series of short positive and negative pulses. These are labeled "1" in the figure. The positive pulses correspond to the time during which the square wave of voltage is increasing in a positive direction, and the negative pulses occur when the square wave of voltage is going in a negative direction. If, however, the time constant is made somewhat longer by increasing the size of *C*, the current flows for a longer time and the length of the pulses increases as in "2." Finally, if the condenser is made still larger, the condenser does not have time to charge during the positive half cycle of the square wave and the output voltage pulses begin to take on the appearance of the applied square wave as in "3." This, and other circuits which produce sharp pulses from a square wave, are called **differentiating** or **peaker** circuits.

* The period of any periodic wave is the time required for the wave to complete one entire cycle.

Now, suppose that the output voltage from an *RC* circuit, excited with a square wave, is taken from across the condenser, as in Fig. 5. If the time constant of the *RC* circuit is *long* compared to that of the square wave, the condenser will not have time enough to charge to more than a fraction of its final value before the end of the half cycle of square-wave voltage. When

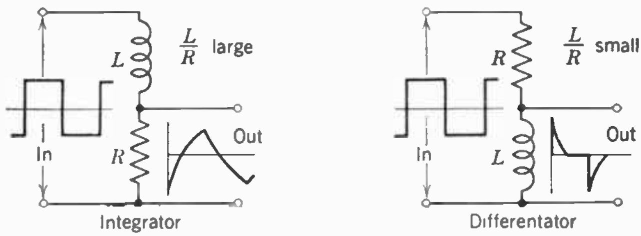


Fig. 6. *RL* integrating and differentiating circuits.

the square wave reverses its polarity, the condenser is forced to begin charging in the opposite direction. The resulting output wave is a triangular wave shown as "1" in the figure. Now, if the time constant of the *RC* circuit is reduced, the condenser is charged to a larger fraction of its final voltage during a half cycle and the output wave is as seen in "2." If the time constant

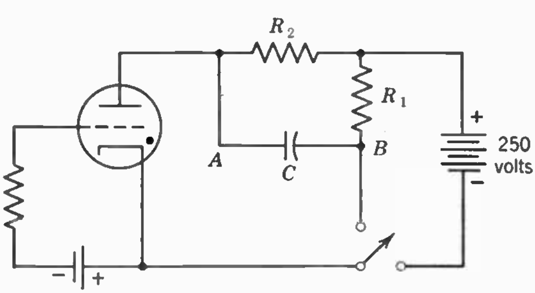


Fig. 7. An *RC* circuit for switching a thyratron.

is made still shorter, the condenser rapidly reaches its final voltage and is held at this value for the remainder of a half cycle. This condition is shown in "3." This circuit is called an **integrating** circuit.

Other differentiating and integrating circuits use *R* and *L* in series. These are seen in Fig. 6.

There are, then, simple methods of producing peaked and triangular waves, and waves with shapes intermediate between these two extremes. If sine waves are impressed on any of these circuits, sine waves of voltage will appear across the individual circuit components, the only change occurring being a shift in phase between the voltage across the resistance compared to that across the capacitor or the inductor.

There are many uses of RC and RL circuits. Whenever a sudden, and temporary, change in voltage or current is needed, an RC or RL circuit may be employed. As one example consider a thyratron tube which is conducting current. Some means of stopping this conduction is desired. If the plate voltage of the tube is removed, it will cease passing current. But there is a more elegant way by using an RC circuit as shown in Fig. 7.

When the tube conducts, the voltage drop across it is about 15 volts. Thus terminal A of the condenser is +15 volts with respect to ground. At the same time terminal B is at +250 volts because the condenser is being charged through R_1 to line voltage. Now what happens if the switch is closed?

Terminal B goes to zero or ground voltage instantly. Since the voltage across C cannot change instantly, the same voltage (250 − 15 or 235 volts) must still appear across the condenser terminals. But if one terminal is at 0 potential and the other is 235 below it, this terminal must have suddenly gone to −235 volts when the switch was closed. This terminal is connected to the plate of the tube, and the plate, therefore, must suddenly become −235 with respect to the cathode, and the tube ceases to pass current.

Of course the condenser soon changes its voltage, the rate of change depending upon the time constant of the circuit, but the plate terminal has been at a high negative voltage long enough to shut off the tube. The tube will remain off if E_{cc} is made negative enough to keep the tube from firing with 250 volts on its plate.

22.6. Peak-clipping circuits. Often, in the process of forming a wave of a specified shape, it is necessary that a wave be "clipped"; that is, a part of the peak must be removed. A simple method for doing this is shown in Fig. 8. The cathode of the diode is held positive ("biased") by the battery E. As the peaked exciting voltage rises, the voltage across the diode and battery also rises in step until the exciting voltage is equal to the battery voltage. As the exciting voltage rises still more the diode plate becomes positive with respect to its cathode and the diode passes current. Up to the point at which the diode passes current its resistance is practically infinite. Now, however, its resistance drops to a few thousand ohms. If R is much larger than the resistance of the diode when conducting, say R is 0.1 megohm, then current flow in R (and the resulting voltage drop across R) prevents the output voltage from rising above the

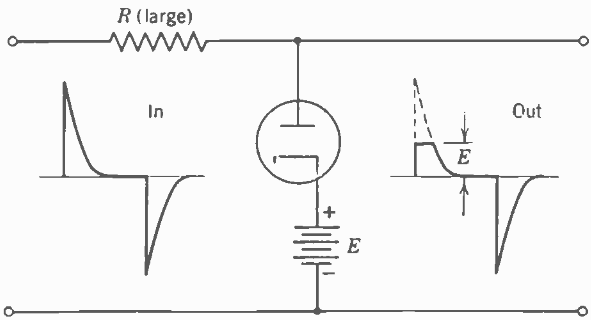


FIG. 8. Diode positive-peak clipper.

battery voltage. The positive half cycle of output voltage is, then, prevented from rising above the voltage of the battery. On the negative half cycle of the exciting wave, the plate of the diode is always negative with respect to its cathode and no current flows. All the voltage appears across the output terminals, and this half cycle is unchanged in form. This is a **positive peak-clipper** circuit. The negative peaks could be clipped by the simple expedient of reversing the diode and battery so that the negative battery terminal is attached to the top output terminal and the diode plate attached to the bottom terminal.

The circuit of Fig. 8 can be modified so that both the positive and negative peaks of a wave are clipped. The circuit is shown in Fig. 9. Let the exciting wave to this circuit be a sine wave with a peak amplitude several times the voltage of the batteries. The output voltage is then only the portions of the sine wave near the axis, and has a shape closely approximating a square wave. The sides are somewhat sloping, but this can be largely corrected by using a very large input signal, or by amplifying the output signal and applying it to another similar clipping circuit. This is a common method for producing square waves.

The circuit of Fig. 9 could be used for clipping the positive and negative half cycles at different levels by using different biasing voltages on the two diodes.

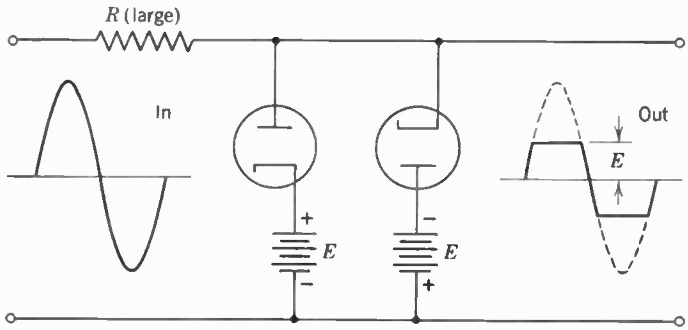


FIG. 9. Diode circuit which clips both peaks.

Peak clipping can be accomplished in a triode by driving the grid positive with the exciting wave. A high resistance in the grid circuit prevents the grid from rising above zero volts, just as in the diode clippers. If the bias voltage is set at half the cut-off value, the negative peaks may also be clipped. The

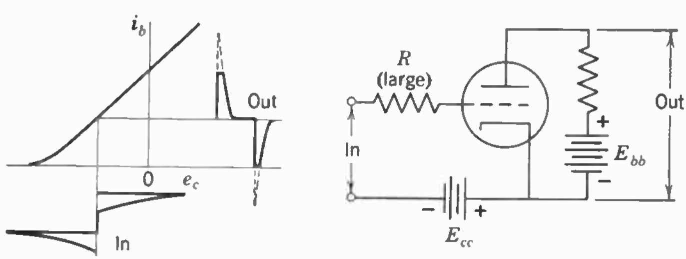


FIG. 10. Triode circuit which clips both peaks.

general operation of such a circuit is shown in Fig. 10. The adjustments for equal positive and negative peak clipping are rather critical, and the curvature of the I_b–E_c characteristic near cut-off may cause some distortion of the negative half cycle of the wave.

22.7. Multivibrators. A circuit widely used for the production of waves with special shapes, and for many other purposes, is the **multivibrator.** In its simplest form, a multivibrator consists of two similar resistance-coupled amplifiers, the output of each feeding the input of the other. Such a circuit can oscillate if the components are properly chosen. When a multivibrator is permitted to oscillate, and no external driving voltage is applied to the circuit, it is called a free-running multivibrator. When

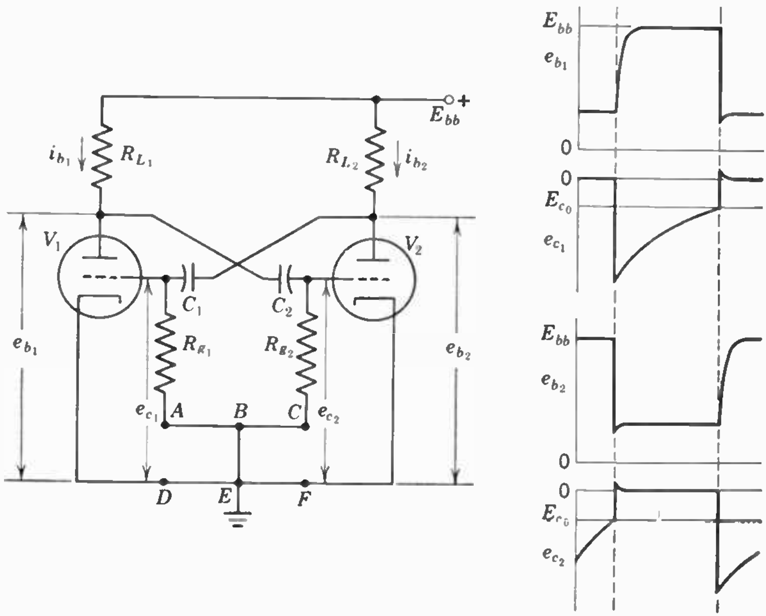


FIG. 11. Plate-coupled multivibrator.

excited from an external voltage source, it is called a driven multivibrator.

The circuit arrangement of a multivibrator is shown in Fig. 11. When the plate voltage, E_{bb}, is first applied, each tube begins to draw plate current. If the tubes and their associated circuits were exactly identical, the circuit would soon reach a steady-state condition with each tube drawing the same current. However, in practice, some differences are bound to exist that cause one tube to draw slightly more current than the other. This slight increase in current causes the current in the other tube to decrease; the effect is cumulative and the current in one tube rises to its maximum value as the current in the other tube falls to zero. Thereafter, conditions reverse themselves, and current starts to flow in the tube which was cut off, current in the other tube is reduced to zero, and the cycle repeats itself. How these effects are produced in the circuit can be described as follows:

Suppose that the current in V_1 starts to increase. This causes a greater voltage drop across R_{L_1}, and e_{b_1} decreases. This decrease in e_{b_1} reflects itself through C_2 as a decrease (more negative) voltage e_{c_2} on the grid of V_2. Thus the current to V_2 decreases, and the voltage drop across its load resistor R_{L_2} decreases, or the voltage e_{b_2} increases. This is reflected as a "positive-going" voltage on the grid of V_1, and thus the current i_{b_1} continues to rise. The current i_{b_1} continues to rise and the current i_{b_2} continues to drop until V_1 is carrying all the current and V_2 is cut off. Condenser C_2 now begins to discharge through R_{g_2} and V_1, and when its voltage drops low enough so that V_2 starts to conduct, the cycle repeats itself except that now V_1 is cut off and V_2 passes all the current. The switching rate is determined by the sizes of the various circuit components.

A more detailed explanation of the operation of a multivibrator can be given by the use of numerical values for the circuit components and various currents and voltages. Suppose, for example, that the tubes in Fig. 12 pass 8 ma when fully conducting, and the cut-off voltage of the tubes is -15 volts. Suppose that V_1 has just cut off and that V_2 is conducting its full 8 ma. The voltage across V_1 (e_{b_1}) has just risen from 90 to 250 volts. The voltage across V_2 has just dropped from 250 to 90 volts ($e_{b_2} = E_{bb} - i_{b_2}R_{L_2} = 250 - 20{,}000 \times 8 \times 10^{-3} = 250 - 160 = 90$ volts). Several other events now occur which are excellent examples of the transient behavior of circuits.

Condenser C_2 begins to charge toward 250 volts. The charging current flows through R_{L_1} and then through R_{g_2} and the grid-cathode resistance of V_2 (r_{c_2}) in parallel. Since r_{c_2} is ordinarily much smaller than R_{g_2} (r_{c_2} is of the order of a few thousand ohms) practically all the charging current passes through the tube; very little passes through R_{g_2} except at the beginning of the period. During this short interval of time at the beginning of the period the grid of V_2 is driven somewhat positive. The rest of the time that V_2 is conducting, its grid is just barely positive—practically zero. Since r_{c_2} has a low value, the time constant is determined largely by R_{L_1} and C_2 and is approximately $20{,}000 \times 0.001 \times 10^{-6}$ or 20 μsec.

At the same time that C_2 is charging, C_1 is discharging through the plate resistance of V_2 and through R_{g_1} in series. R_{g_1} is ordinarily much larger than the plate resistance of V_2 so that the time constant of this circuit is approximately $1{,}000{,}000 \times 0.001 \times 10^{-6}$

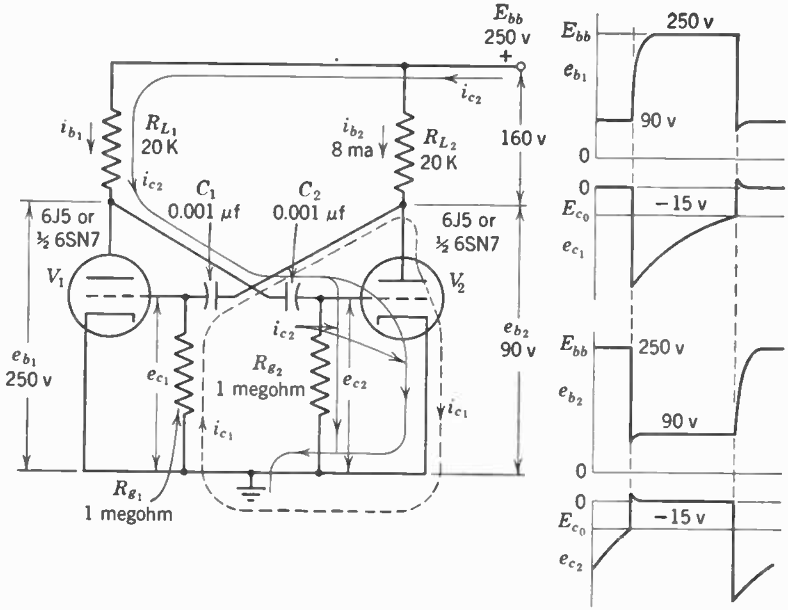


FIG. 12. Waveforms in a plate-coupled multivibrator.

or 1000 μsec. The initial voltage across C_2 is approximately 250 volts. It discharges to the voltage across V_2 (90 volts) plus the cut-off voltage of V_1 (15 volts) or to about 105 volts. The discharge current, i_{c_1}, flowing through R_{g_1}, biases the grid of V_1 negative. Initially, the discharge current is high and the bias is much beyond cut-off, but as the current decreases, the bias voltage becomes less and less negative until V_1 starts to conduct. Thereupon, the cycle is repeated except that now V_1 is conducting and V_2 is cut off.

The waveforms of the various voltages are shown in Fig. 12. By proper choice of components, e_{b_1} and e_{b_2} can be made essentially rectangular in shape, and a multivibrator can be used to produce rectangular or square waves. The curves for e_{c_1} and e_{c_2} are typical capacitor discharge curves. Since the waveforms in the multivibrator are so irregular, the circuit can be used for the production of harmonics up to as high as the hundredth harmonic. The base frequency is readily controlled by some external source.

One of the interesting uses of multivibrators is for frequency division. That is, the base frequency (repetition rate) of the multivibrator may be controlled by an external signal of higher frequency than that of the multivibrator. The control frequency must be an integral multiple of the free-running frequency of the multivibrator. The control frequency may be applied to the circuit so that either even or odd multiples are favored. Figure 13 shows typical injection methods.

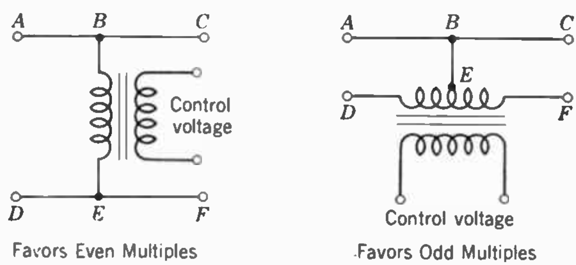


FIG. 13. Methods of injecting controlling voltages in multivibrator of Fig. 11.

22.8. Square-wave testing of amplifiers. The behavior of square waves when passed through simple *RL* and *RC* circuits suggests that square waves could be used to determine how well amplifiers and other circuits reproduce an input wave. A square wave contains many harmonics in addition to the base frequency.* Thus, in effect, an amplifier may be tested at many frequencies with a single setting of the base frequency of the square wave. Phase shift of some of the frequency components with respect to others, or unequal amplification of the various frequency components, will cause characteristic changes in the

* A square wave contains harmonics in the following proportions:

Fundamental,	f	100	$15f$	6.7
	$3f$	33.3	$17f$	5.9
	$5f$	20	$19f$	5.3
	$7f$	14.3	$21f$	4.8
	$9f$	11.1	$23f$	4.3
	$11f$	9.1	$25f$	4.0
	$13f$	7.7	$27f$	3.7

resulting output wave. The method is much faster than a point-by-point method using sine waves. It is usual procedure to test an amplifier for low-frequency response with a square wave having a frequency (repetition rate) of the lowest frequency of interest, say 60 cycles, and to test the high-frequency response

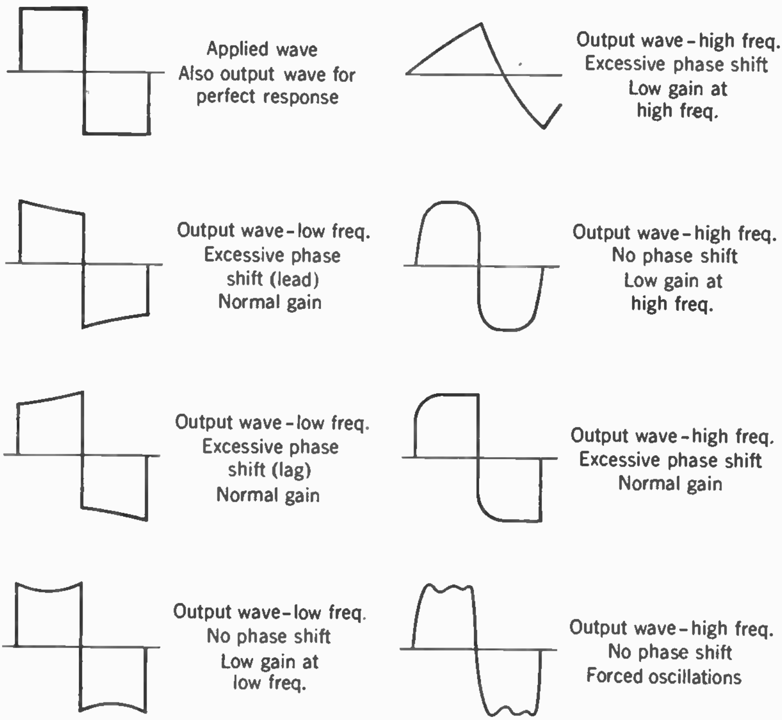


FIG. 14. Typical square-wave responses of a-f amplifiers.

with a square-wave frequency of around 1000 cycles. Typical output waves for amplifiers with typical defects are shown in Fig. 14.

Radio-frequency amplifiers may be tested by employing an r-f carrier, modulated with a square wave, and observing the amplified wave. The method is limited to rather low radio frequencies, or to intermediate frequencies, because of the inability of most oscilloscopes to reproduce high-frequency signals.

Sawtooth waves are sometimes used for amplifier testing. However, generators with a variable-frequency sawtooth output are not often available in radio laboratories.

23 · Television

One of man's constant endeavors has been to improve and extend his communications systems. The newspaper, the post office, the telegraph are all results of a desire to convey information to persons at great distances. So, too, is radio, which permits almost instantaneous communication of audible messages. Television, like radio, permits rapid communication over great distances, but it adds a feature missing in radio—a picture.

23.1. Elements of a television system. A television system requires a device for systematically exploring a scene, converting differences in light intensity into an electrical signal, and combining this signal with a r-f carrier. At the receiving end must be a device capable of amplifying this modulated signal, extracting (detecting) the picture signals, and converting them into an image on a receiving screen. Along with this visual **(video)** system, there are accompanying devices which carry the sound associated with the scene. Auxiliary apparatus must be provided for synchronizing the elements of the over-all television system so that they function in the proper predetermined sequence.

23.2. Scanning. A modern television transmitter employs a camera tube upon which the image of the scene to be televised is focused by a camera lens. Within the tube the image falls on a rectangular surface on which are millions of small silver globules which have been made light-sensitive. That is, when illuminated they emit electrons like tiny photocells. The number of electrons emitted per unit of time depends upon the intensity of the illumination, brighter portions of the scene producing more electrons.

The emitted electrons are collected by a plate which is at a positive potential with respect to the silver globules. When an electron leaves the photo surface it leaves the spot positively charged; one might say it leaves a hole in the surface.

Scanning is accomplished by moving a beam of electrons across the surface on which the image is focused. The change in potential between the surface and the collector plate, from point to point, corresponds to the amount of light in the scene. When the beam passes a spot with a positive charge, electrons from the beam fall into the "holes" and neutralize the positive charge.

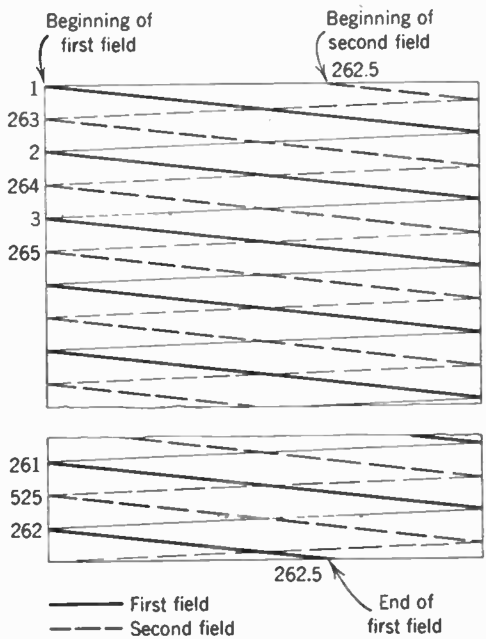


Fig. 1. Interlaced scanning sequence. Spacing between scanning lines is actually much closer than shown.

In this manner variations in illumination are translated into variations of electric current which can be transmitted to distant television receivers where the changes in voltage caused by the changes in current at the transmitter are reconverted into changes in illumination.

23.3. Interlaced scanning. The beam can be moved across the image in many ways but in this country **interlaced** scanning is employed. In this method the beam starts at the upper left-hand corner (Fig. 1), moves across horizontally to the right, is then whipped back quickly to the left, where it is dropped down a bit, and then is caused to "scan" across to the right again.

After 262.5 lines have been scanned the beam is moved quickly up to the top again and a new line is scanned. This line is placed between the first two lines scanned, and all subsequent lines down to the bottom of the scene are placed between two previously scanned.

Thus the visual image is translated into 525 lines, line 263 falling between lines 1 and 2, line 264 between lines 2 and 3, and so on. When the 525 lines have been completed a single **frame** made up of two 262.5-line **fields** has been produced. In each second of time, 30 frames are produced.

During the time the beam is moved back to the right or from bottom to top it is extinguished (blanked) so that no electrons hit the photoelectric surface. In this manner no spot on the surface is scanned more than once during each frame.

23.4. Picture tubes. At the receiver the reverse action must take place in exact synchronism with the events at the transmitter. That is, each change in the strength of the electrical signal must be converted into a change in illumination. This is accomplished in a cathode-ray tube (Sect. 21.8) in which a narrow beam of electrons is sent down the neck of the tube and allowed to fall upon a surface so coated that it will glow with a spot of light when the electrons strike it. This electron beam is moved across the face of the tube—the screen—in exact synchronism with the motion of the electron beam at the transmitter. Furthermore, the intensity of the beam, that is, the number of electrons allowed to fall upon the screen, is modulated or controlled by the number of electrons that fall into the "holes" in the photo surface at the transmitter. When a high positive charge is neutralized at the camera tube, many electrons hit the screen of the picture tube and a bright spot occurs.

In this manner the original scene focused upon the camera tube is reconstructed at the receiver picture tube. When the scanning beam at the transmitter is extinguished, the beam at the receiver is also extinguished; when the beam at the transmitter moves from bottom to top, the receiver beam moves from bottom to top, and so on.

The picture tubes in most television receivers are cathode-ray tubes employing electromagnetic deflection and focusing. These tubes produce more uniform focusing over the picture area than do tubes employing electrostatic methods or a combination of the two methods, and may be made considerably shorter in physical length. The anode voltage of these tubes is of the order of 6000 to 15,000 volts. The fluorescent screen of the tube is sometimes coated with a very thin film of aluminum. This metallic film serves both to conduct electrons rapidly away from the fluorescent screen and to act as a mirror-like reflector to direct practically all the light produced at the screen through the front of the tube toward the viewer. The metallic film also filters out the large and heavy positive ions, keeping them from burning the fluorescent screen. Picture tubes for direct viewing range in size from 3 to 20 in. in diameter, popular sizes being 12 in. and above. Some of the newer types have rectangular rather than round screens to correspond more closely to the outline of the picture area. The cone of some tubes, that is, the flared portion of the tube between the neck and the screen, is made of metal. This makes the tube both lighter and stronger.

Some television receivers employ a projection system. In these receivers a rather small picture tube is used which produces a very bright picture. The picture is enlarged in a Schmidt optical system and finally projected upon a viewing screen which may be several square feet in area. Projection systems usually suffer from loss of detail and brightness due to imperfections in the optical system and unavoidable losses of light.

23.5. Camera tubes. The tubes upon which the image of the subject being televised is focused, and which translate changes of illumination of the subject into electrical signals, are called camera or pickup tubes. Several types of camera tubes are available: iconoscopes, orthicons, image dissectors, and image orthicons.

A simplified diagram of an **iconoscope** is shown in Fig. 2. The image is focused on a mosaic which is made up of a large number of small silver globules deposited on one side of a thin flat mica plate. The globules are photosensitive and are insulated from one another. The other side of the mica plate is metallized to form a **signal plate.** The electron beam from the electron gun is deflected by magnetic or electrostatic means so that it scans the mosaic surface. The **collector** electrode is a volt or two positive with respect to the mosaic surface, and electrons emitted from the mosaic are attracted to the collector. Because of the low potential difference, however, the electrons do not move to the collector very rapidly.

When the scanning beam moves across the mosaic surface, it leaves each globule with a slight negative charge. In the absence of an optical image on the mosaic surface, the charge on all the globules would be the same. If, however, an image is focused on the mosaic, the globules with high illumination eject photoelectrons and lose some of their negative charge. The intensity

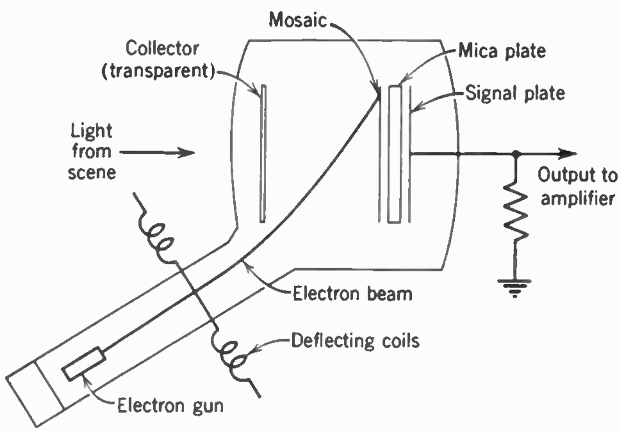


FIG. 2. Simplified drawing of an iconoscope.

of the light determines how much of this charge is lost. The next time the scanning beam strikes these globules their negative charge is restored and a current flows. This current appears as an output signal.

The **image orthicon** differs from the iconoscope in that the charge pattern which the electron beam explores is formed by secondary emission rather than by photoelectrons. In addition, an electron multiplier * is incorporated within the tube. The sensitivity of this tube is about 100 times that of an isonoscope. A schematic diagram of an image orthicon is shown in Fig. 3. Electrons are accelerated in the electron gun, formed into a beam,

* An electron multiplier is a tube or arrangement of electrodes which produces very high current amplification through the use of secondary emission effects. A nine-stage multiplier may produce a current amplification as high as one million times.

and projected toward the target. The electron beam is focused by an axial focusing coil and is deflected by coils whose axes are perpendicular to the axis of the tube. The electron beam moves at a rather low velocity and, after reaching the target, is reflected back toward the cathode where it is focused on an adjacent secondary-emission cathode **(dynode).** This dynode is the first element of a five-stage electron multiplier. The current of the electron beam is multiplied many times in the electron multiplier, and then passed through the output load resistor.

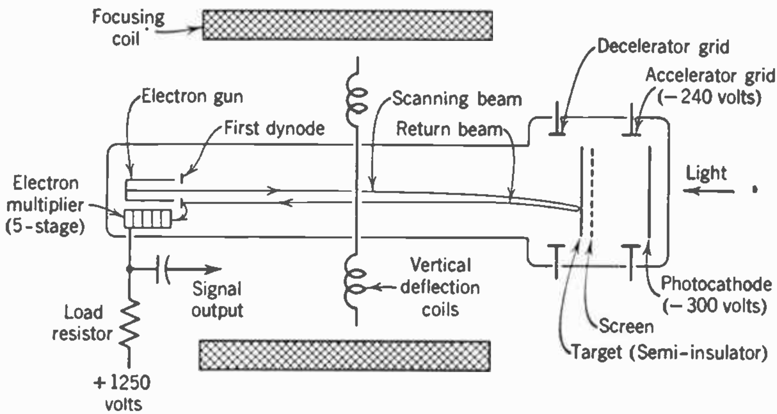


FIG. 3. Image orthicon, simplified diagram.

Light, upon striking the semitransparent photocathode, causes the emission of photoelectrons. These electrons are attracted to the target where they produce secondary emission. The target is made of a thin sheet of semiconducting material, often low-resistivity glass. The electrons emitted by secondary emission do not immediately redistribute themselves to other portions of the target because of its high resistance. Thus, the charges which these electrons produce remain fixed at the point on the target where they were produced for a long enough time for the scanning beam to explore them; the amount of charge is proportional to the number of photoelectrons emitted by the photocathode, or to the brightness of the portion of the image which is being viewed and which they represent. A fine-mesh screen is placed very close to and in front of the target. The secondary electrons are drawn to this screen as soon as they are emitted. Thus the charge on a small area of the target is more positive for highlights than for shadows in the image.

As the scanning beam explores the back side of the target, it releases just enough electrons to each small target area to counteract the positive charge residing there. Then the beam returning to the first dynode has fewer electrons by the number released to the target. The output current of the tube is, therefore, less for areas of the target representing high levels of image illumination than for areas representing low illumination levels.

23.6. Film on television. Television programs originate from 16-mm moving-picture films as well as from live subjects. Films for television are of two types: movie films and kinescope recordings. Movie films are prepared by moving-picture producers or by the broadcasters themselves. Kinescope recordings are moving pictures of the images on the face of the kinescope (picture tube) of a television receiver. Most kinescope-recording systems are "wired" systems. That is, the television transmitter is connected directly to the receiver without intervening antennas.

The image from a movie projector using either type of film is focused directly on the camera tube of the television transmitter and translated into electrical signals much the same as any other kind of television transmission. The quality of television programs originating from films is inherently poorer than the quality of "live" programs because of the added steps, electrical, optical, and photographic, in producing a film program. This loss of quality is not troublesome when high-quality movie films are employed since the "noise" and distortion arising from a direct photographic process can be kept low. Kinescope recordings, however, are noticeably poorer than live programs because of the added noise and distortion picked up in the television circuits which are a part of the kinescope-recording process.

Film projectors for television are designed to accommodate standard 16-mm movie films. Although this practice makes available a large reservoir of standard moving pictures it gives rise to a rather troublesome problem. Standard movie equipment projects at the rate of 24 frames per second, whereas the scanning rate for television is 30 frames per second. If a standard movie projector were used for television, the film image would be flashed on the camera tube at a non-synchronous rate and the image on the camera tube would change during a scanning sequence, producing a poor-quality picture. Special projectors are, therefore, required for television.

Standard movie projectors are built so that each frame (individual picture) is flashed on the viewing screen twice; then the next frame is moved into position and is also flashed on the screen two times. Thus, even though there are only 24 frames per second, the audience sees 48 images per second and there is no noticeable flicker as long as the viewing screen is in a darkened room.

Television film projectors are built so that one frame is flashed on the camera tube twice, the next frame 3 times, the following frame twice, etc., in a 2-3-2-3 sequence. This permits the film to be run through the projector at the normal rate of 24 frames per second and still produce 60 images per second on the camera tube, or a separate image for each field (2 images per frame) of the television scanning sequence. This procedure makes it possible to synchronize the projector with the television scanning signal. In addition, the "flicker frequency" is high enough (60 images per second) that the television image can be viewed in an undarkened room without troublesome flicker effects.

It is usual practice to flash the image from the film projector on the camera tube during the vertical blanking pulses (Sect. 23.7) and depend upon the storage action of the camera tube to preserve the image long enough for it to be scanned. The film "pull down" (movement of the next film frame into place) occurs during the relatively long time intervals between vertical blanking pulses. Even with this arrangement the mechanical pull-down system must operate very rapidly. Because of the requirement for storage action, only iconoscopes and image orthicons are suitable for film projection. The screens of image orthicons, however, are subject to burning if a stationary image, such as a slide, is allowed to remain more than a few seconds. The storage action of other types of camera tubes currently available is not long enough that a satisfactory television signal can be produced.

Slides may be projected on all types of camera tubes with a standard slide projector, provided the lens system will project a clear image of the required size on the sensitive screen of the camera tube. Care must be exercised to prevent burning the screen of the camera tube. Usually the same pickup tube serves for both slide and movie projectors. Various projectors are mounted so that any of them may be used with the pickup tube. A movable mirror allows any of two or three projectors to be used by merely rotating the mirror. In this manner the projectors need not be moved and they can be permanently focused.

23.7. The television signal. The picture signal transmitted by a television transmitter, and picked up by the receiver, is necessarily rather complex. It must not only contain the electrical information concerning the scene being viewed; it must also carry along signals for blanking the receiver at the proper times. In addition, both vertical and horizontal synchronizing signals are required. The blanking signals turn off the beam of electrons in the cathode-ray tube (kinescope) in the receiver during the retrace time (both horizontal and vertical) of the beam. The synchronizing signals keep the scanning beam at the receiver in step with that at the transmitter. Another series of signals, equalizing pulses, keeps the interlace pattern at the receiver properly adjusted.

The picture signal is *amplitude-modulated.* Variations in the degree of modulation are used to carry the various types of information that must be transmitted to the receiver. Figure 4, illustrating various functions, shows only the *envelope* of the transmitted wave; the individual alternations of voltage are not shown.

In the United States television transmitters utilize negative modulation. That is, an increase in picture brightness causes a decrease in the amplitude of the envelope and a resulting decrease in power output. The reference "white" level is set at not more than 15 per cent of the maximum amplitude of the carrier envelope. "Black" is represented by an amplitude of the carrier envelope of 75 per cent of the maximum value. This black level serves as a reference point from which the background brightness of the picture is set. Blanking pulses have an amplitude equal to that of the black level, and the scanning beam in the receiver is cut off whenever this level of modulation is received. Synchronizing and equalizing pulses rise above the black level to 100 per cent of the maximum amplitude of the carrier. Thus, these pulses are transmitted as "blacker-than-black" signals. The scanning beam at the receiver is, therefore, cut off whenever a blanking, synchronizing, or equalizing pulse is transmitted.

Details of the various signals making up the composite transmitted envelope are shown in Fig. 4. The picture signal occupies

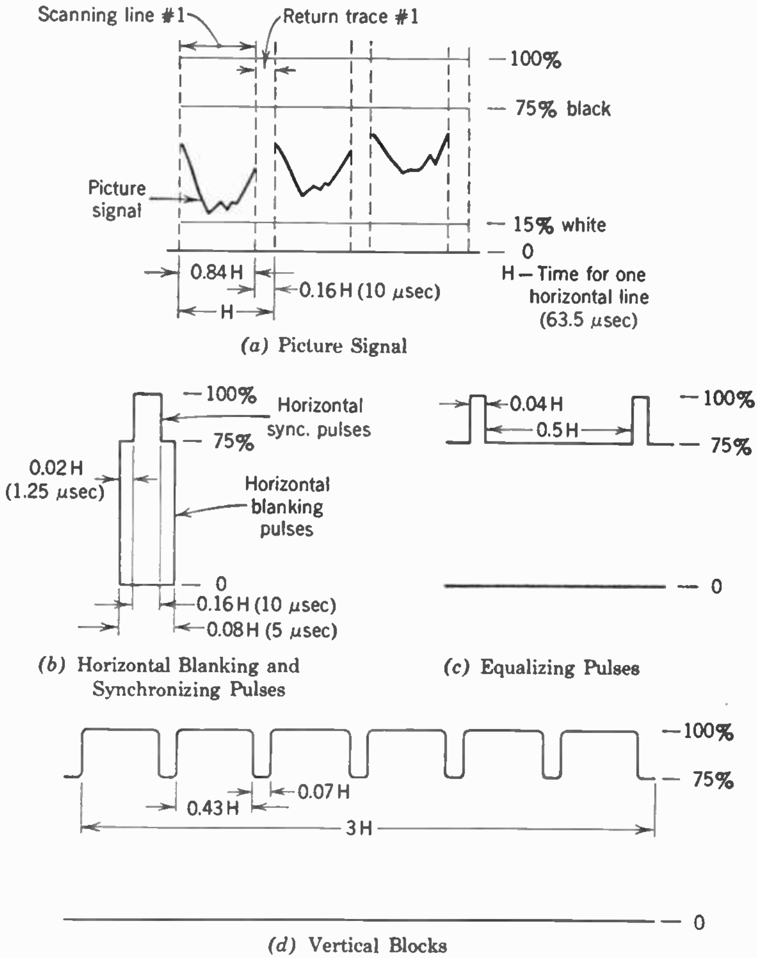


Fig. 4. Details of a television signal. Only the envelopes of the pulses are shown.

about 84 per cent of the time required to scan one horizontal line. The remaining 16 per cent of the time is occupied by a blanking pulse upon which is superimposed a horizontal synchronizing pulse. The blanking pulse turns off the scanning beam during the retrace time; the synchronizing pulse causes the scanning beam in the receiver to begin scanning the next line in synchronism with the beam at the transmitter. Details of these pulses are seen in Fig. 4.

At the end of each field (262.5 horizontal lines) several pulses are transmitted which perform various functions. The first are a group of six **equalizing pulses** to keep the received picture properly interlaced. These equalizing pulses are followed by a **vertical synchronizing pulse.** This pulse is three times as long as the time required to scan one horizontal line and is much longer than the horizontal synchronizing pulses so that the two types of pulses may be separated by the circuits in the receiver. However, to maintain horizontal synchronization, the vertical synchronizing pulses are "serrated"; that is, the pulse is broken up into six blocks as shown in Fig. 4(*d*). These blocks start at the proper times to keep the receiver in horizontal synchronization. At the end of the vertical synchronizing pulse another series of six equalizing pulses is transmitted. These pulses are followed by nine to thirteen horizontal synchronizing pulses. During this entire time the electron beam is cut off and no picture is transmitted. This time, about 1250 μsec, represents the vertical retrace time of the beam.

The time the vertical blanking pulses occupy is 7 to 8 per cent of the total time occupied by a frame. As a result, only about 485 lines out of a total of 525 are active in producing a picture.

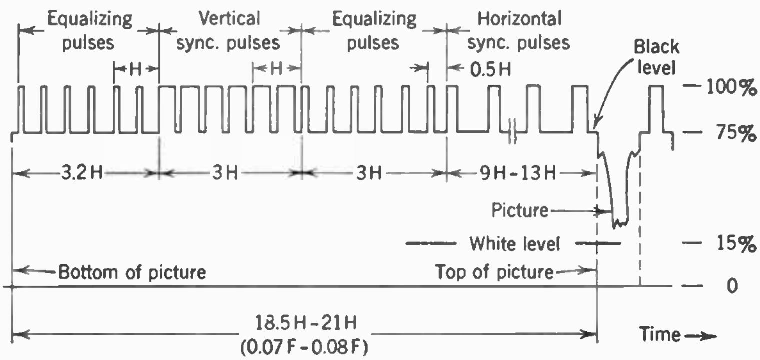
A typical composite envelope of the transmitted signal is shown in Fig. 5 together with the relation of the picture pulses to the various synchronizing, blanking, and equalizing pulses. The figure is for the time just before, during, and after the end of a field in order to include the details during the vertical retrace time.

23.8. Television transmitters. In the United States black-and-white television stations are assigned channels 6 Mc wide. The assigned carrier frequencies range from 54 Mc to 216 Mc.*

Television transmitters employ **vestigial sideband** transmission. In this system, one sideband is largely eliminated at the

* The present FCC channel allocations are in the following bands: 54 to 72 Mc, 76 to 88 Mc, and 174 to 216 Mc. There is a total of 12 bands. The channels are numbered consecutively from 2 to 13. Channel 1 has been discontinued.

transmitter by a filter. The other sideband is transmitted without appreciable suppression. By this means most of the channel space can be utilized by the single sideband. The result is that a



H – Time for one horizontal line (63.5 μsec)
F – Time for one field (16.667 μsec)

Fig. 5. Envelope of a composite television signal.

clearer picture, with more detail, can be transmitted in a given channel width.

The normal response of a transmitter is shown in Fig. 6. Note

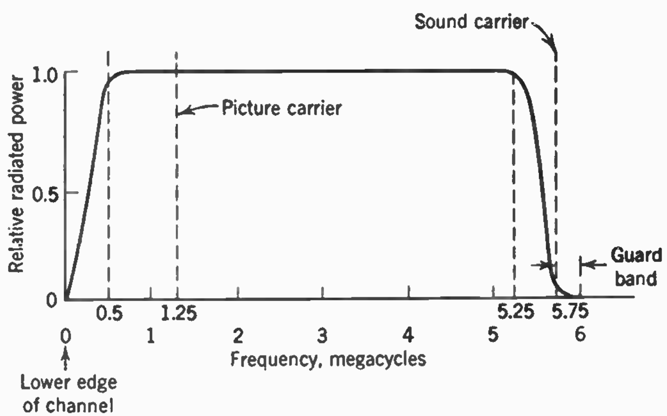


Fig. 6. Television transmitter channel with vestigial sideband transmission.

that the picture carrier (center carrier frequency) is located 1.25 Mc above the lower edge of the 6-Mc channel. The vestigial sideband filters are designed to have a gradual cut-off beginning about 0.75 Mc above the lower end of the channel, the cut-off being complete at the lower end of the channel. The transmitter response is approximately flat for a 4-Mc band above the picture carrier frequency. It then drops gradually to zero response at 5.75 Mc above the lower end of the channel. The remaining 0.25 Mc serves as a guard band to separate adjacent channels. The sound carrier is centered on a frequency 5.75 Mc above the lower end of the channel.

23.9. Television receivers. Television receivers are usually superheterodynes, often with preliminary r-f amplification. A block diagram of a typical receiver is shown in Fig. 7. The

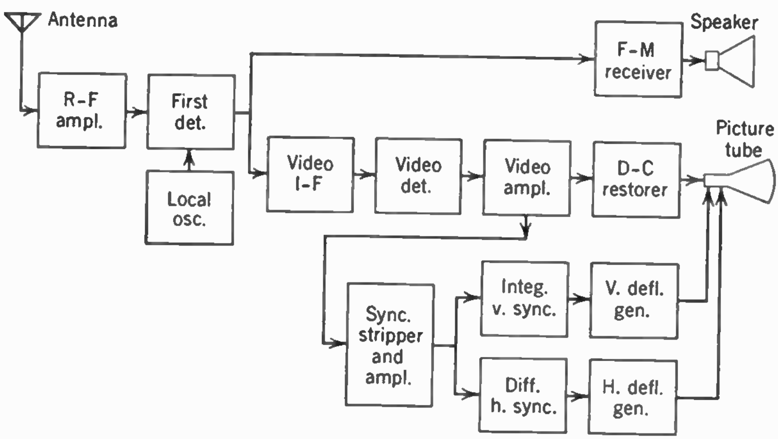


FIG. 7. Simplified block diagram of a typical television receiver.

picture and sound carriers are passed through the r-f amplifier and mixer stage. After the mixer (after one or two stages of i-f amplication in some receivers), the sound carrier is separated from the picture carrier by means of a high-Q "trap" circuit. The sound carrier which is frequency-modulated is then passed through a typical f-m receiver.

The ideal i-f system of the picture amplifiers should have a selectivity curve as shown in Fig. 8. In practice, the picture i-f amplifier is considered to be satisfactory if the band width between half-power points (0.7 times maximum response) is 3.5 to 4.0 Mc. The amplifiers must be compensated to have good gain and time-delay (phase-shift) characteristics over this frequency range. Because of the wide band width required, the tuned circuits of the i-f amplifiers must have a rather low Q; often a resistor of low value, a few thousand ohms at most, is shunted across the tuned circuit. This arrangement reduces the available gain per stage, and several stages of amplification are generally necessary to obtain the required amount of gain. Special "television pentodes" such as the 6AC7 have been developed which feature a large g_m. These tubes permit the designer to obtain a greater gain per stage than would otherwise be possible.

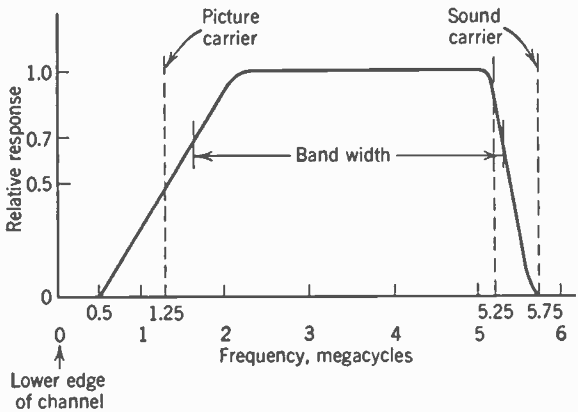


FIG. 8. Ideal selectivity curve of video intermediate-frequency amplifier of a television receiver. Local oscillator frequency is lower than picture carrier frequency.

The intermediate frequency in most television receivers is of the order of 21 Mc. Additional sound-carrier "traps" are incorporated in one or more of the i-f amplifier stages to insure that the sound carrier is completely eliminated from the picture signal.

The video detector is usually a diode employing a circuit much the same as that discussed in Sect. 14.8. The diode load impedance is often provided with some form of high-frequency compensation, such as a small coil in series with a 2000- or 3000-ohm load resistance. The output signal level is ordinarily somewhat greater than 1 volt peak-to-peak.

After detection, the video signal passes into a video amplifier. This must be a high-quality amplifier capable of passing frequencies from a few cycles per second up to about 4 Mc. This signal appears much the same as that shown in Fig. 5. It would appear that the synchronizing and equalizing signals would cause interference on the picture-tube screen. This is not the case, however, since these signals represent picture elements that are "blacker-than-black" and are not seen.

The video signal in most receivers, after detection, passes through video amplifiers and their associated blocking and coupling condensers. Since a condenser cannot pass a d-c compo-

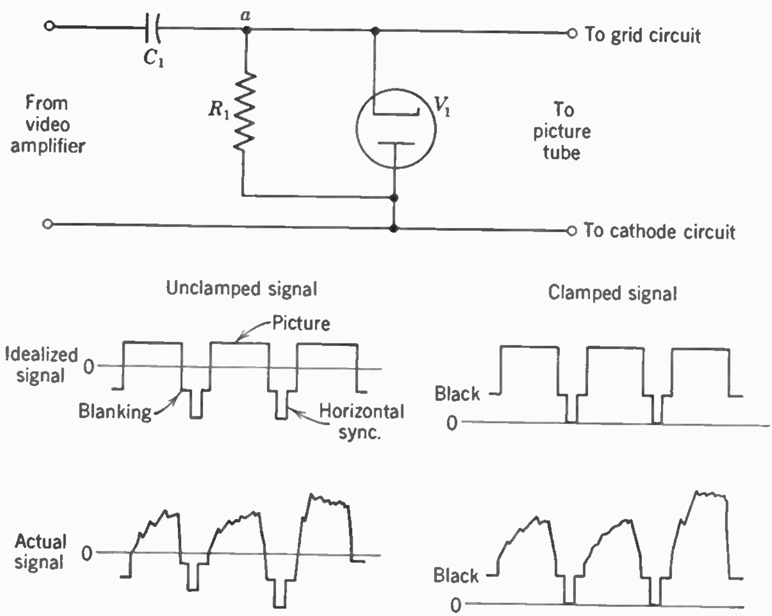


FIG. 9. Simple diode clamping circuit for restoring "reference" black level to video signal.

nent, the resulting a-c signal adjusts itself so that the areas in the positive and negative portions of the wave are equal. The heights of the successive synchronizing and blanking pulses are no longer the same. That is, these pulses extend different distances above the zero-voltage axis. The electrical information regarding the picture brightness is lost since the brightness is related to the synchronizing and blanking pulses. The blanking and pulses must be realigned so that the picture signal can be passed onto the picture tube with the correct brightness information. The circuits which restore this reference level are called **clamping** or **d-c restorer** circuits.

A simple clamping circuit is shown in Fig. 9 along with an idealized video signal. Note that the picture signal is a "positive-going" signal in this case, as contrasted to the "negative-going" picture signal in Fig. 5. This phase reversal is necessary so that the grid of the picture tube will be driven in a positive direction for the brighter portions of the televised scene.

Consider what happens when the video signal begins to go positive. Current flows through the C_1–R_1 path, but no current flows through V_1 since its cathode is positive with respect to its plate. Condenser C_1 charges rather slowly since the time constant of C_1–R_1 is ordinarily much greater than the time required for the scanning of one line. When the video signal begins to go negative the plate of V_1 becomes positive and V_1 begins to conduct. Condenser C_1 now begins to charge in the opposite direction. This charging action is rapid because of the low resistance of V_1, usually between 200 and 2000 ohms, and terminal *a* follows the changing negative potential with very little time delay. The resistance of R_1 is made so much greater than that of V_1 that very little current flows through R_1 during the negative portion of the cycle. The clamping circuit, in effect, "clamps" the synchronizing pulses at zero, and the rest of the signal rises above zero. In this manner, the reference axis for brightness information is restored and a steady level is provided for the picture tube. A germanium crystal diode, a type 1N34, for example, may be used in place of V_1. The clamping action on a typical video signal is shown in Fig. 9.

A portion of the video signal, after detection and amplification, is diverted into a "sync stripper" which removes the vertical and horizontal synchronizing signals and rejects other portions of the composite wave. The horizontal synchronizing signals are of much shorter duration than the vertical synchronizing signals. This fact permits the two signals to be separated in differentiating and integrating circuits.

The synchronizing signals which pass through the integrator act together to produce an output pulse. However, since a vertical synchronizing pulse is very much longer than the other pulses, only this pulse is effective in producing a large amplitude in the output signal. The other pulses produce a signal of rather low amplitude. A multisection integrator is generally employed. The action of such a circuit is to discriminate against the shorter horizontal synchronizing and the equalizing pulses, and to produce an output signal dependent only upon the vertical synchronizing pulses. This output voltage is used to synchronize the vertical deflection generator.

The synchronizing signals passing through a differentiator produce sharp positive pulses corresponding to the leading edges of the synchronizing pulses, and sharp negative pulses corresponding to the trailing edges. The negative pulses are removed by diode clippers or other means, and the positive pulses are used to trigger the horizontal deflection generator. It should be noted that these triggering pulses are produced not only by the leading edge of the horizontal synchronizing pulses, but also by the leading edges of the equalizing pulses and by the leading edges of the serrations of the vertical synchronizing pulses.

23.10. Television sound circuits. The sound associated with a television picture is ordinarily transmitted separately from the picture, and the sound carrier is *frequency-modulated.* The sound channel is located at the upper edge of the allocated picture channel. The sound transmitter is, in effect, entirely separate from the picture transmitter, although the same antenna is ordinarily used, the output of the two transmitters being coupled to the antenna through a special network.

The separation of the picture and sound carriers at the receiver has already been discussed. Once the sound carrier has been isolated, it is fed to a conventional f-m receiver.

23.11. Television power supplies. The requirements for the power supplies for the amplifiers and other low-voltage circuits of a television receiver are no different from those for any other high-quality receiver. They must furnish the required d-c power, must have good voltage regulation, and must be free of ripple. The high-voltage supply for the picture tube, on the other hand, poses special problems for the receiver designer.

It has already been mentioned that the picture tube requires direct voltages ranging from 6000 to 15,000 volts. The current requirements are low, about 200 μa. The high-voltage power supply should be of the limited-energy type, that is, one which is not capable of passing more than a few milliamperes on short circuit. This condition is imposed for safety reasons so that the shock hazard to persons who inadvertently come in contact with this high voltage is lessened. In all cases, the high-voltage supply should be turned off and the terminals short-circuited before working on a television receiver, unless it is absolutely necessary that the high-voltage portions of the receiver be operative.

High-voltage television power supplies do not usually operate at power-line frequencies since such operation would require rather heavy and bulky components. Since only a few milliamperes of current at most are required by the picture tubes, special power supply circuits have been devised for television use. The most common are flyback, radio-frequency, and pulse supplies.

A **flyback** supply is found only in receivers employing magnetic deflection. A typical flyback supply is shown in Fig. 10.

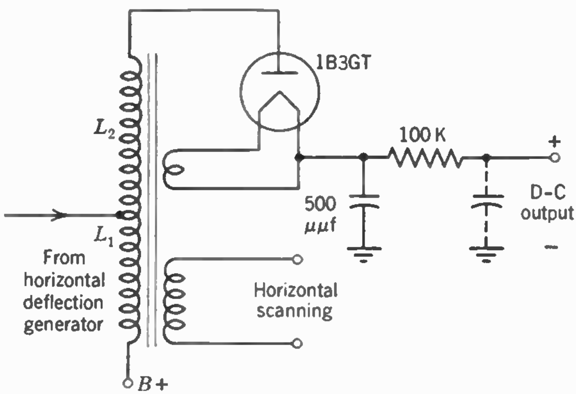


FIG. 10. Flyback power supply.

The circuit gets its name from the fact that it is operated by voltage produced in a coil during the flyback time of the horizontal scanning wave, that is, during the time the electron beam is being returned from the right to the left side of the viewing screen. This flyback or retrace time is quite short, about 10 μsec. The current in the coil L_1 changes rapidly, producing a large counter-voltage. The voltage is made still greater by connecting L_2 as a step-up autotransformer. This voltage is then rectified and filtered. The frequency of the flyback pulses is 15,750 cps, and small filter components may be used.

Radio-frequency power supplies utilize a r-f oscillator. The output voltage of the oscillator is fed to a rectifier circuit. Because of the high frequency, 50 to 500 kc, the filter elements may be quite small. The second filter condenser is often the capacitance between inner and outer conductive coatings found on many picture tubes. Radio-frequency power supplies must be carefully shielded to prevent interference with other portions of the receiver. A typical supply of this type is shown in Fig. 11.

A pulse-type power supply is useful in receivers employing electrostatic deflection. A pulse derived from the horizontal scanning circuit is used to trigger a single-shot blocking oscil-

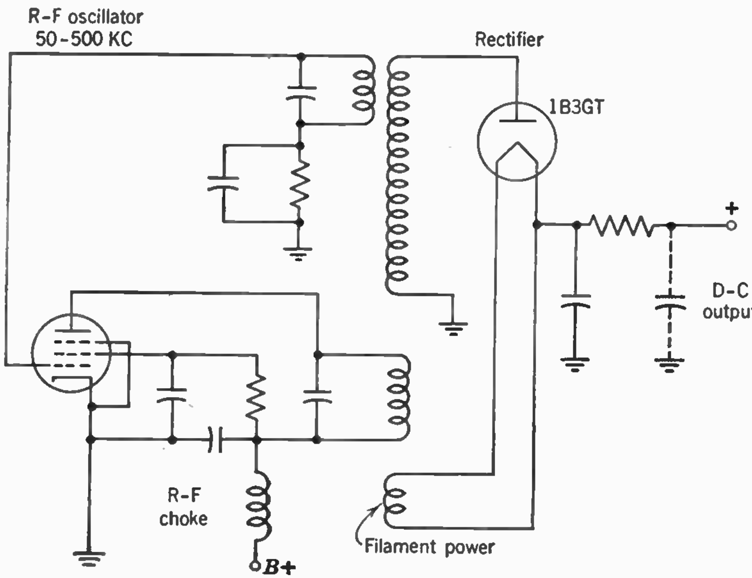


FIG. 11. Radio-frequency high-voltage power supply.

lator. The output of the oscillator is amplified and then rectified as in the two previous circuits. The pulses are triggered by the horizontal scanning circuit and, if the scanning circuit is of the "driven" type, the supply is operative only when a picture is present. Thus, no high voltage is applied to the tube in the absence of a picture. This is advantageous in that it prevents burning of the picture-tube screen.

23.12. Television antennas. Television antennas for both transmission and reception must be capable of transmitting or receiving a broad band of frequencies without discrimination. The transmitting antenna is operated at a fixed carrier frequency and so can be designed to have optimum performance at this frequency. Several types of transmitter antennas are in use.

They have two things in common: (1) they have relatively low gain (because of their broad band width) in comparison to standard broadcast and f-m antennas, and (2) they are mounted as high in the air as practicable to achieve a large service area.

Receiving antennas for television signals must be capable of satisfactorily responding to an extremely wide band of frequencies if all channels are to be received. The antenna is often a dipole made of rather large-diameter rods and tuned to the middle of the lower-frequency band or about 65 Mc. In many cases the dipole is a part of a directive system including both directors and reflectors. A highly directive system is usually mounted so that it can be rotated to pick up a desired station with the maximum sensitivity. It is important that the antenna be properly matched to the transmission line and the receiver. An impedance mismatch may cause serious loss of signal and may, at the same time, produce troublesome "ghosts."

23.13. Reflections, ghosts, and interference. Television reception is seriously affected by multiple-path transmission. A signal from a transmitter may not only travel directly to the receiving antenna; it may also be reflected back and forth between buildings and other objects before being picked up by the receiving antenna. Signals arriving by these longer paths are delayed in time and may produce a delayed image or "ghost" on the receiver screen. A highly directional receiver antenna helps minimize trouble from these multiple paths. Sometimes the simple expedient of relocating the antenna at a more favorable spot will eliminate most of the trouble.

Natural static does not ordinarily cause much interference in a television picture. Much more serious is interference caused by man-made static. This static arises from the ignition systems of automobiles, from unshielded diathermy and r-f heating devices, from streetcar and elevated systems, and from many other electrical devices. Directive antennas help eliminate some of this static. A receiving antenna should be located as far as possible from the street.

23.14. Color television.* Ever since the development of black-and-white television, there has been considerable interest

* An excellent discussion of color television systems will be found in the article, "The Present Status of Color Television," *Proc. I.R.E.*, Vol. 38, No. 9 (September, 1950), pp. 980–1002.

in the development of systems which would reproduce a picture in its true colors. Several color television systems have been devised. Many research workers and designers are presently working on improvements and searching for new ways of producing a color television picture.

Color television systems are based on the fact that all colors can be produced (approximately) by combinations or mixtures of the primary colors, blue, green, and red. A basic color television system would, therefore, separate the image into these three basic colors in the pickup device. The electrical signals corresponding to these three colors would be combined in appropriate circuits at the transmitter and then be radiated by the antenna. At the receiver, additional circuits would separate the signals corresponding to the different colors, and then recombine them in a viewing device so that the original picture was reproduced in its true color values.

A scanning pattern in any system, black and white or color, is made up of a multitude of *dots*, each dot represented by a separate globule of the photosensitive camera tube. These dots, then, are the basic building blocks of a television image. The sensitive surface is scanned in horizontal *lines;* these lines are the second step in the construction of the electrical representation of the image. A series of lines covering the area of the sensitive surface constitutes a *field*, the third step in the construction of an image. In an interlaced system, two fields (one frame) are required to cover the complete image. In spite of this modification, however, the basic elements of a television picture are (1) dots, (2) lines, and (3) fields.

Two different systems of color television have been proposed, the **sequential** and the **simultaneous** systems. In the sequential system, the three primary colors are transmitted one at a time in succession. A single picture transmitter is required. A simultaneous system, on the other hand, transmits all three primary colors at the same time. This system ordinarily requires three separate transmitters, each with a band width roughly the same as for a black-and-white transmitter (6 Mc) or a total of 18 Mc. The simultaneous system, because of its requirement for an extremely large band width, is not actively proposed for commercial use and will not be discussed further. Sequential systems are, however, receiving much attention, and their features will be discussed in some detail.

Sequential systems can be divided into three types, depending upon which "building block" of the picture is utilized: **dot-sequential** (RCA), **line-sequential** (CTI), and **field-sequential** (CBS). In a dot-sequential system the color is assigned to successive picture elements or dots on the sensitive surface of the camera tube. In a line-sequential system the color values are assigned to successive lines in the scanning pattern. In the field-sequential system color values are assigned to successive fields of the image.

In a dot-sequential color television system the scanning pattern is the same as in a conventional black-and-white system. In any one field the lines of the image consist of successive dots of the three primary colors. In the RCA system the dots are arranged in the sequence red, blue, green, and the space between successive dots of the same color is equal to the width of a dot. Thus, there is some overlap of the different colored dots. (The "colored" dots are actually black and white. They merely represent the colors specified.) On successive fields the positions of the dots are shifted so that the green dots, for example, fall at different points from on the preceding field. In addition, the dots on successive lines are shifted so that in any given field the dots of one color do not appear directly above or below the dots of the same color in adjacent lines. Because of the successive arrangement of the dots, and the arrangement on adjacent lines, this is called a dot-interlaced dot-sequential system.

Two types of dot-sequential systems have been designed and tested. In one, three separate camera tubes are employed. The image is focused simultaneously on all three tubes, one primary color per tube, and the tubes are scanned in synchronism. The scanning beams are switched on and off so that the tubes are scanned in succession, a small section of a line (dot) in one tube being scanned, then a small section of a line in the second tube, and so on. Appropriate signals transmitted to the receiver switch the electron beams of three different color tubes. The color tubes are picture tubes which glow red, blue, and green, respectively. The images on the three color tubes at the receiver are combined by an optical system. This color television system imposes extremely strict requirements on the over-all operation of the various component parts. The three images must be accurately positioned with respect to each other (kept in register), and the relative color values must be maintained.

In a second type of dot-sequential system the same type of transmitter is used, but the receiver uses a single picture tube. The viewing screen of the picture tube is composed of a very large number of tiny clusters of phosphors, red, green, and blue. Three electron guns are employed. The three electron beams are caused to strike the different phosphors in order, much as the different colors are produced on the three screens in the receiver described above. Registration of the image at the receiver is automatic, if the image is properly registered at the transmitter, and if the three electron beams are accurately deflected.

In a line-sequential system, any given line is scanned entirely in a single color. A line-interlaced pattern is employed. For example, the odd-numbered lines * are scanned first, line 1 being scanned in green, line 3 in blue, line 5 in red, and so on until one field is scanned. The second field also scans the odd lines, except that line 1 is scanned in red, line 3 in green, and lie 5 in blue. The third field also scans the odd lines, the first line this time being scanned in blue. After three fields the image has been scanned in all three colors, but only the odd-numbered lines have been scanned.

The next three fields scan the even-numbered lines. In successive fields line 2 is scanned in green, then blue, then red. At the completion of six fields the image has been scanned by all lines and by all colors, and a complete color picture has been produced.

A line-sequential system may employ a single camera tube. Color-selective filters are used to produce three images, side by side, on the sensitive screen. The scanning beam is moved entirely across the sensitive screen, thus scanning the images corresponding to the three primary colors one after another. The beam-deflecting circuits are arranged so that the proper sequence of scanning is obtained.

The fluorescent screen of the viewing tube at the receiver is composed of rectangular areas of phosphor, side by side, which glow in the three primary colors. The scanning beam moves

* These are lines 1, 2, 3, etc., in Fig. 1. Here, the lines are numbered consecutively regardless of the field to which they belong.

across these three colored areas much as at the transmitter. Thus a copy of the images on the camera tube is reproduced on the receiving screen, this time in three colors (the actual images on the camera tube are in black and white). The three images on the viewing tube are combined in an optical system and a color picture is produced.

The line-sequential system may employ the same number of lines per field and the same number of fields per second as a black-and-white system and still utilize the same band width. Therefore, a line-sequential transmission may be picked up on a conventional black-and-white television receiver as a black-and-white picture. The system is subject to trouble arising from the difficulties of accurately registering the images both at the transmitter and at the receiver.

A field-sequential system scans an entire field in a single color. The second field is scanned in another color, and the third field in the third color. In an interlaced system, the first three fields include only the odd-numbered lines, and three additional fields are required for the even-numbered lines. A total of six fields then comprises the complete color picture. In a modification of this system, each line is broken up into dots, all of the same color. Blank spaces between these dots are filled with dots of another primary color on the next scanning of that line. In this manner, both lines and dots are interlaced, and a clearer picture results.

The CBS field-sequential system (1950) employs a motor-driven color wheel both at the transmitter and at the receiver. The color wheel is a disk containing six segments. Two segments are red filters, two are green filters, and two are blue filters. The wheels are rotated in synchronism with the switching and scanning of the electron beams so that the scanning pattern proceeds as described above. The fields are reproduced on the viewing screen in succession and with the proper colored segment of the color wheel in front of the screen. In this manner a colored picture is produced. This system is capable of producing an excellent color picture. It is, however, subject to two limitations. In the first place, to achieve a good picture and still stay within a 6-Mc band width, the number of lines and field rate must both be different from those in a conventional black-and-white system. A field-sequential system is, for these reasons, incompatible with present black-and-white systems. In the second place, the viewing tube is restricted in diameter to about 12 in. Otherwise, the color wheel and its driving motor become unreasonably large.

A modification of the field-sequential system employs a single viewing tube in which the different colored fields are produced one above the other and in sequence, the three areas of the screen being coated with different phosphors. The scene is viewed through an optical system which combines the three images. This system eliminates the necessity for a color wheel at the receiver. It does, however, suffer from registration difficulties the same as the two previously described systems.

24 · Radar

Radar was one of the supreme contributions of electronics engineers to World War II. By its means it became possible accurately and quickly to determine the direction of an enemy airplane and at the same time to measure its elevation and its distance. Since the war, radar has found many peacetime uses, the most important of which is guiding ships in fog and at night. By its use, ship's pilots can maintain "visual" contact with other ships and obstructions.

The word "radar" was coined from the term "*ra*dio *d*irection *a*nd *r*anging," and it means simply the determination of direction and range of any object which will reflect back to the transmitter a portion of the energy it receives from the transmitter.

Basically, radar depends upon three simple phenomena, all well known for many years. These are (1) the ability to aim a narrow band or cone of radio energy by means of directive antennas; (2) the fact that radio waves travel at a known velocity in space, namely 300,000,000 meters or 186,000 miles per second; and (3) the fact that radio energy striking an object such as an airplane induces currents in the object. The currents cause some of the energy to be reradiated or reflected. These facts were known to Hertz, the first and third facts being demonstrated by him and the second being a part of the basic theory developed by Maxwell. Thus the basis for radar existed and was known before 1900. Only in recent years, however, have means been available for measuring the time required for a radio wave to go out to a reflecting object and be returned to the sender. This portion of a radar system had to wait for modern techniques.

24.1. Elements of a radar system. A simple radar system consists of a transmitter which sends out periodic short bursts of energy, an antenna array designed to concentrate the energy into as narrow a beam as possible, a simple receiving antenna, and a means of measuring the time of transit of the pulses of radio energy. The transmitting and receiving antennas are rotated in synchronism and are often the same antenna. The direction of the antennas when a return echo is picked up indicates the direction of the reflecting object, or target. The elements of a radar system are shown in Fig. 1.

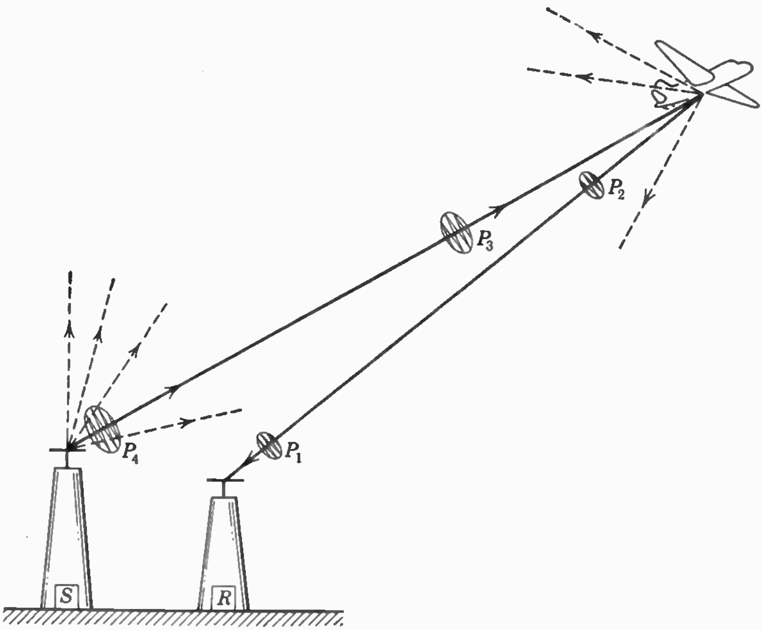


FIG. 1. Energy sent out in pulses from transmitter S is reflected in pulses from plane to receiver R. (From *Wireless World*.)

The transmitter sends out radio energy in pulses so that there may be time between pulses for the burst of energy to travel to the target and for some of the energy to be reflected back to the receiver. The receiver measures the elapsed time from the instant a pulse is radiated by the transmitter until it is picked up by the receiver. The time between pulses is comparatively long with respect to the length of the pulses. It is not practicable to send out a continuous stream of radio energy since then it would be difficult for the receiver to measure the transit time of any particular portion of the transmitted signal. Furthermore, the returned energy is many times weaker than the transmitted energy.

If the transmitter were on all the time, the receiver would be overwhelmed by the local signal and would be unable to pick up the weak echo.

Since the transmitter is energized at intervals and only for a very short time (a microsecond or so), the peak transmitted power may be very great without the average power being above the rating of the transmitting tubes. This peak power may be hundreds or even thousands of kilowatts; the average power is usually less than a hundred watts.

The distance or range of the target is measured by "timing" circuits, and the information is displayed on cathode-ray oscillographs. When the pulse is sent out, a signal starts a sweep voltage which moves the electron beam across the screen. When the echo is returned, a signal is sent to the vertical plates of the tube so that a slight kick (called a "pip") upwards from the horizontal trace on the screen is produced. If the electron beam is swept across the screen at a constant and known speed, the place on the screen where the pip occurs is a measure of the time required for the signal to go out from the transmitter to the target and to be returned by the target to the receiver. Since the speed with which the waves travel is accurately known, the transit time is also a measure of the distance to the target. The relation between transit time and target distance is shown in Fig. 2.

For example, radio waves travel 186,000 miles per second or 0.186 mile per microsecond, out and back, or a one-way trip of 0.093 mile (491 ft) per microsecond. Suppose, now, that the electron beam is swept horizontally across the cathode-ray-tube screen a distance of 4 in. in 400 μsec, or 1 in. for every 100 μsec. If, therefore, a pip occurs 1 in. away from the starting point of the horizontal trace, the elapsed time between transmission of the signal and reception of the echo is 100 μsec. In 100 μsec the signal went out to and returned from a distance of 100×0.093 or 9.3 miles. Thus the reflecting target is 9.3 miles away.

If echoes are obtained when the antenna is pointing at a bearing of 185° with respect to true north, then the target lies 9.3 miles away on the 185° line. Furthermore the elevation above earth can be determined by tilting the transmitting and receiving antennas away from the horizontal until echoes are received. The angle of the antennas with the horizon, then, is a measure of the height of the target, which can be determined by elementary geometry.

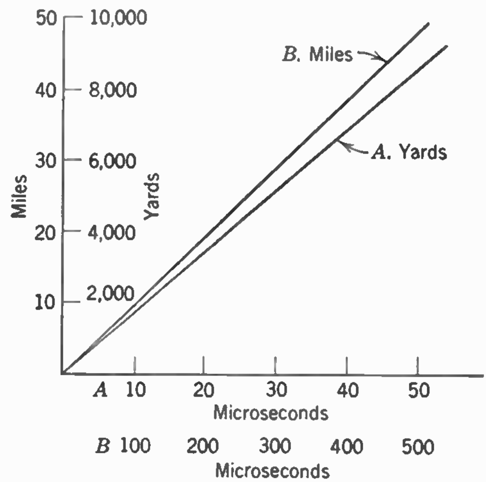


Fig. 2. Relation between transit time from radar to plane to radar and distance between plane and radar.

24.2. The radar transmitter. Since highly directional antennas with reasonable physical dimensions can be constructed when the wavelength of the transmitted radio energy is short, radar systems commonly operate at very high frequencies, ranging from hundreds to thousands of megacycles. The transmitter, therefore, is a very short-wave transmitter. It is not continually on the air like a broadcast transmitter. It resembles a code transmitter more than a voice transmitter, and a short-wave code station could be employed as a radar transmitter by merely making the dots occur at a regular rate and also making them very short in duration. Special directive antennas would, of course, have to be provided.

A simplified block diagram of a single-antenna radar is seen in Fig. 3. The timer controls the time at which the transmitter sends out a pulse of energy and also the length of the pulse. Whenever the timer causes a pulse of energy to be transmitted, it also produces a time marker signal for the cathode-ray-tube indicator screen. The *TR* (transmit-receive) switch, often a gas-filled tube, connects the transmitter to the antenna during the time of a pulse. At all other times this switch connects the antenna to the receiver.

Means are provided for furnishing requisite instantaneous power during the pulses, and there must be some sort of keying system which allows the transmitter to produce high-frequency energy at regular intervals, and for the required length of time during the individual pulses. This keying system must be accurate, for if the circuits timing the pulses get out of order, the

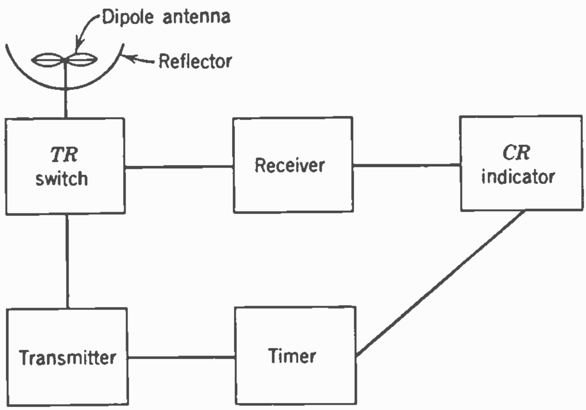


Fig. 3. Block diagram of typical radar. *TR* switch throws antenna from transmitter to receiver.

returning echoes may be obscured by the more powerful transmitter pulses. The timing mechanism must be synchronized with the indicator in the receiver so that the trace of the cathode-ray tube starts across the screen when the pulse leaves the transmitter.

Since the transmitter is essentially a short-wave station, any of the well-known methods of producing short-wave signals may be employed. The most widely used tube for producing high-power high-frequency signals, however, is the magnetron. Special triodes can be used in radars operating up to a few hundred megacycles. Klystrons can be used in low-power radars operating at frequencies up to several thousands of megacycles.

24.3. Transmitter power. The average power required from a radar transmitter depends upon the fraction of time in each second that the power is on and the amount of power taken during this time. In this situation, a radar transmitter resembles an electron-tube-controlled welding machine which supplies high-powered pulses of energy to a weld at regular intervals, each pulse lasting only a fraction of a second.

As an example, suppose that a 200-watt transmitter is to be keyed with pulses lasting 2 μsec each, there being 500 pulses per second. What fraction of a second is the transmitter actually taking power from the power source? How much power may be packed into each pulse without overloading the transmitter or power supply?

If there are 500 pulses per second, each 2 μsec long, the total time per second the transmitter is on is $500 \times 2 \times 10^{-6}$ or 1000 μsec; and, since there are 1 million μsec in each second, the transmitter is on only one-thousandth of the time. During the pulses, therefore, the power may be as high as 200 kw (200×1000 watts). This simple analysis assumes that it is the average power (200 watts) that is the limiting factor. Of course, limitations on current and voltage in the transmitter must also be observed and may, in some cases, limit the peak power to less than 200 kw.

As in welding technique, the factors mentioned above can be related as follows: Let the term "duty cycle," taken from welding terminology, be the relation between the width of a pulse and the time between pulses:

$$\text{Duty cycle} = \frac{\text{Pulse width}}{\text{Time between pulses}}$$

These relations are shown in Fig. 4.

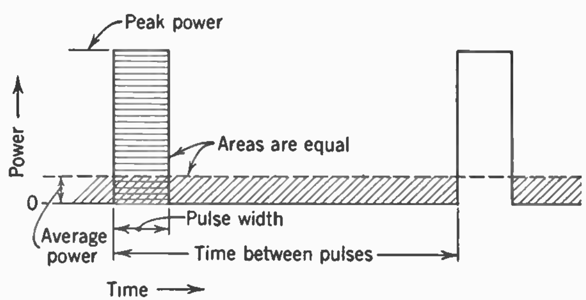


FIG. 4. High peak power does not mean high average power. In radar transmitters the time between pulses may be quite long compared to pulse time.

If there are 500 regularly spaced pulses per second, the time between successive pulses is $\frac{1}{500}$ sec or 2000 μsec. If the pulse is on for 2 μsec, the duty cycle is $2 \div 2000 = 0.001$.

The duty cycle is also the ratio between the average and the peak power. Therefore,

$$\frac{\text{Average power}}{\text{Peak power}} = \frac{\text{Pulse width}}{\text{Time between pulses}}$$

In the example above, in which the transmitter had an average power rating of 200 watts, 200/peak power = 0.001, or the peak power is 200,000 watts or 200 kw.

Since the amount of energy reflected by a small object at some distance from the transmitter is very small in comparison with the transmitted power, it is important that the power of the transmitted pulses be quite high so that the pulses returned to the receiver will be strong enough to be successfully amplified and used.

24.4. Pulse repetition rate. How many pulses per second should be sent out? Consider the example above, in which 500 pulses are sent out per second. How far away can the target be and still reflect one pulse back to the receiver before the next pulse is emitted by the transmitter? The pulses are 2000 (actually 1998) μsec apart in time. The energy will travel a one-way distance of 0.093×2000 or 186 miles. This will be the *maximum* range of a radar transmitter using a pulse repetition rate of 500 per second.

How long should the pulse last? If it lasts too long, the returning signal will arrive before the transmitter is turned off. If the pulse lasts 2 μsec, the signal will travel (one way distance) 2×0.093 or 0.186 mile, or 328 yards. This is the *minimum* distance over which a radar using a 2-μsec pulse would be useful.

The pulse *rate* determines the maximum effective distance, a higher rate giving the signals less time to go out and return and thus lowering the maximum distance; the pulse *width* governs the minimum distance over which echoes can be received successfully. A wide pulse may return more energy from a distant target but will increase the minimum distance over which the radar is useful. Numerous pulses per second increase the possibility that several pulses may hit the target and be returned, with the result that a brighter trace on the cathode-ray tube will be secured.

Figure 5 shows how minimum and maximum useful distances depend upon pulse rate and width.

In practice, the pulse repetition rate is much greater than the rate at which the antenna is to rotate about the horizon or about the sector the radar is exploring. In this manner, there is sufficient time for several pulses to go out to the target and to

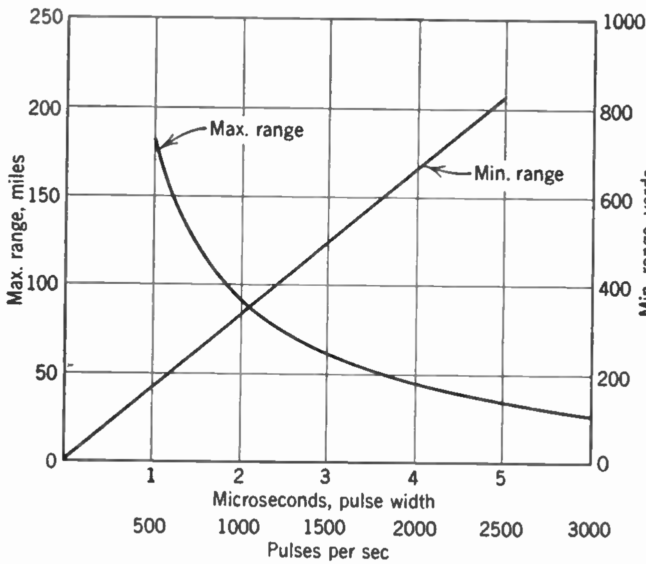


Fig. 5. The minimum range increases as pulse width in microseconds increases; the maximum range, however, decreases with wider pulses.

return to the radar site before the antenna has moved on to a new location. If proper pulse control is attained, the received echoes will be superimposed upon one another so that a visible "pip" may be produced from the sum of several weak signals.

The lowest effective pulse rate, therefore, is determined by the maximum range desired, by the rate at which the antenna is rotated, and by the persistence characteristics of the cathode-ray-tube screen.

Problem 1. A radar employs pulses 3 μsec long. The repetition rate is 1000 per second. What are the minimum and the maximum ranges? What is the duty cycle? If the transmitter tube is limited to 100 watts average power, what is the peak pulse power?

Problem 2. A 2J51 magnetron operates in the frequency range of 8500 to 9600 Mc. It has a maximum rated average power output of 50 watts.

(*a*) If the pulse power output is to be 80 kw, what is the maximum permissible duty cycle?

(*b*) What must be the pulse repetition rate if 1-μsec pulses are used?

(*c*) What will be the minimum and maximum ranges?

24.5. Modulator. Since a radar transmitter is turned on and off at regular intervals, some means must be provided for modulating the carrier wave, in this case modulating it completely and abruptly. This modulation may be accomplished in the oscillator tube itself, or additional tubes may be employed for the purpose.

24.6. Radar receiver. Since the energy returned to the receiver antenna by a target may be exceedingly weak, the receiver must have the maximum possible sensitivity. The incoming wave is first lowered in frequency by superheterodyne techniques (crystal diodes are used as mixers in microwave radars), and the lower-frequency wave is amplified as much as necessary before the detection process in which the pulse envelope is secured free from the r-f and i-f voltages. After detection the pulse voltage may be increased further in a video or wide-band amplifier before being applied to the vertical plates of the cathode-ray tube.

The various circuits of the receiver through which the signal passes must be carefully designed and constructed so that the time it takes a pulse to travel through the receiver is accurately known. Only in this manner can the time required for the pulse to travel to the target and be reflected back to the receiver be accurately measured.

24.7. The indicator. As is true of many modern applications of electronics, the cathode-ray tube is an important element in a radar system. It serves as a visual indicator of range (calibrated in yards or miles), or height (calibrated in degrees from the horizontal), or bearing with respect to north (calibrated in degrees). In each case, the movement of the electron beam across the cathode-ray-tube screen in a horizontal direction is synchronized with the movement of the antenna in a horizontal direction for bearing or in a vertical direction for height. The screen can be used to indicate bearing or vertical angle, or merely to indicate when a response is received from the distant target, a compass card synchronized with the antenna movements indicating the actual bearing.

If the radar on shipboard is to serve only to locate ships at sea, the dimension of height may be eliminated since all that is needed is the distance and the bearing of the target. If the radar is to be used for gun control, much greater accuracy is required than if the system is merely to indicate the vicinity and range of a

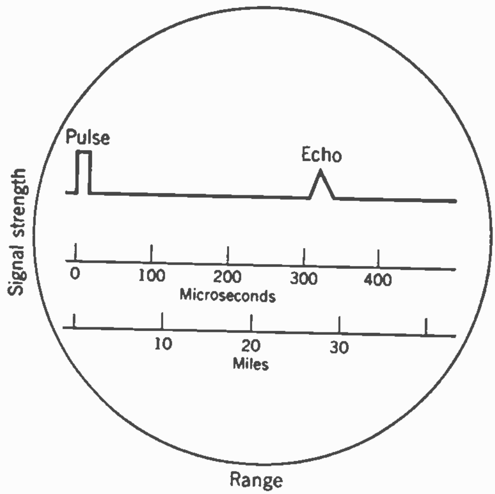
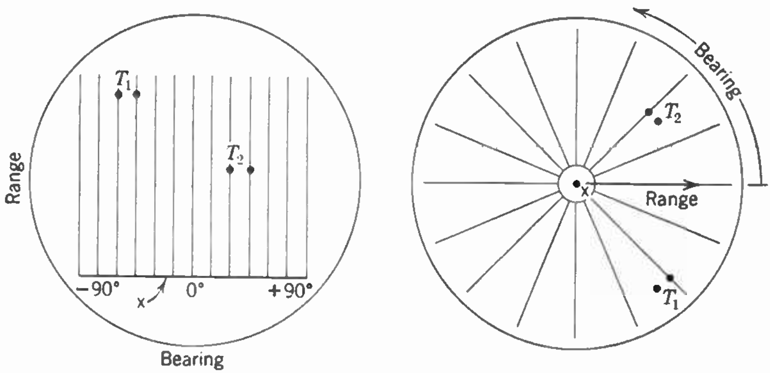


Fig. 6. Type A radar indicator. The "range" base line may be calibrated in units of time or distance.

target. When the radar is employed for gunfire control, the movements of the antennas are synchronized with the movements of the guns or searchlights.

The picture obtained on the cathode-ray-tube screen described above, and shown in Fig. 6, is called a type A presentation. This system shows only the distance or range of the target. The direction is obtained by synchronization with auxiliary devices as described.

In type B presentation [Fig. 7(*a*)], both range and azimuth (bearing) data are shown on the screen. A highly directive antenna is rotated so that the transmitted beam of energy sweeps in a horizontal circle. The horizontal axis (base line) of the cathode-ray screen is calibrated in degrees of bearing. Vertical distances on the screen correspond to range. The electron beam is caused to make a light vertical trace on the screen at evenly spaced intervals (every few degrees of bearing). Signals reflected from a target are used to intensify the trace. Thus, if a bright spot appears on the trace, the range of the target is indicated by the distance above the base line, and the bearing is indicated by the particular vertical line upon which the bright spot appears. The transmitted pulses are synchronized with the sweep pattern so that they always fall at a time corresponding to the base line.



(a) Type B *(b)* PPI

FIG. 7. Types of radar indication. X, transmitted pulses. T_1, T_2, received pulses from targets.

A Plan Position Indication (PPI) shown in Fig. 7(*b*) presents *bearing* information as angles and *range* information as radial distances from the center of the screen. This type of presentation is used for applications in which height of the target is not of importance. The picture on the screen may be described as a simple map with the radar transmitter at the center. The information on the map is incomplete in the sense that, if two targets are directly in line and are at the same height, only the nearest target will be indicated on the screen. The antenna, for this type of indication, is rotated on a vertical axis; that is, the beam is swept in a horizontal circle. Many pulses per antenna revolution are sent out. This type of radar usually employs a cathode-ray-tube screen with long-persistence characteristics so that the information is retained for several seconds after a particular reflecting target has been scanned. PPI presentation is widely used in shipboard radars designed for navigational purposes.

24.8. Radar antennas. Radars utilize frequencies ranging from 100 Mc (3 meters) to 10,000 Mc (3 cm) or higher. The trend is to the higher frequencies as tubes capable of high-power operation at the higher frequencies become available. The types of antennas used for radar can roughly be divided into two groups: (1) those for the lower frequencies, and (2) those for the higher frequencies. No boundary frequency can be specified since there is considerable overlapping of antenna types in the region from about 600 to 3000 Mc. The general requirement for a radar antenna is that it must produce a highly directive beam. In special cases, the beam must not only be highly directive; it must also fulfill predetermined vertical and horizontal field pattern specifications.

Antennas for use at lower frequencies are usually made up of several simple antennas, such as half-wave dipoles, made up into an "array." These arrays usually consist of a number of "driven" antennas to which power is supplied by the transmitter and an equal number of reflectors placed the proper distance behind the driven elements. Sometimes a Yagi antenna is employed. This is a driven element with a reflector and one or several directors. The directors are dipoles, usually shorter than the driven element, placed parallel to and in front of the driven element.

A common antenna array used at the lower radar frequencies consists of 16 half-wave dipoles with a similar set of 16 dipole reflectors, all properly spaced, and connected so that the required phase relations are obtained. Such an array will produce a power gain of about 400 times. (Power gain is defined as the power radiated in the desired direction divided by the power which would be radiated in the same direction with a single dipole. The total power supplied to the array and to the dipole must, of course, be the same.) Thus, a 100-kw pulse fed to the 16-dipole array becomes in effect a 100×400 or 40,000-kw signal in the desired direction. The directive property of such an antenna may be indicated by stating that the radiated beam is 30° wide at a point where the radiated energy is one-half that radiated along the axis of the beam. Ordinary transmission lines conduct the power from transmitter to antenna.

In the mircowave region, the antenna may be very small and is often located at the focus of a parabolic reflector or in a spherical reflector. Such an arrangement produces high power gain and high directivity. Wave guides serve as connecting links between antenna and transmitter or receiver.

24.9. Shoran. Shoran (*sho*rt *ra*nge *n*avigation) is an example of the use of radar for aircraft navigational purposes. A simple shoran system consists of two beacon receivers on the ground plus special radar apparatus in the plane. A radar beacon is merely a low-powered radar transmitter and an associated receiver. Pulses sent out by the radar apparatus on the plane cause a beacon to be triggered and to send out a succession of pulses which are in turn picked up by the plane. The beacons are usually arranged so that their operation is practically instantaneous and the time required for the radar to trigger the beacon and for the beacon pulses to return to the plane is practically the same as though the beacon were merely a reflecting target. Then, by methods already described, a navigator may determine how far the plane is from a beacon. The beacon receiver circuits are usually arranged so that they will respond only to pulses of a particular length and repetition rate. Their return pulses are often coded for identification purposes.

The manner in which the navigator may fix his position is shown in Fig. 8. By contacting two beacons within the range of the plane's radar, the distance from each may be determined. Knowing the geographical location of the beacons, the navigator can draw arcs of circles on his navigation map, the radius of each circle corresponding to the distance from the associated beacon. The intersection of the two circular arcs fixes the position of the plane.

Because of the high frequencies involved, shoran is limited to line-of-sight distances, and the beacons must be somewhat less than twice line-of-sight distances apart. Even though these distances are somewhat greater from an airplane than at ground level, a large number of beacons are required to cover any large area.

24.10. Loran. Loran (*lo*ng *ra*nge *n*avigation) apparatus normally operates in the frequency range just above the broadcast band, from about 1.7 to 2 Mc. Two pairs of ground stations are required in the simplest system. One station, the **master** station, sends out high-powered pulses. Not only are these pulses transmitted to a loran receiver on an airplane; they are also picked up by a second ground station, some 200 to 300 miles from the first station, and used to trigger the pulse transmitter in this second, or **slave** station. A constant and known time difference between the pulses of the two stations is thus maintained. The *difference* of time of the arrival of these pulses at the receiver on

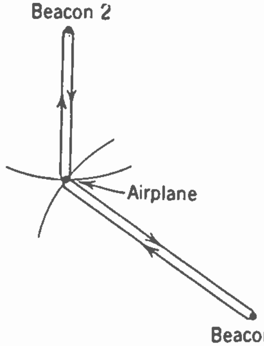


FIG. 8

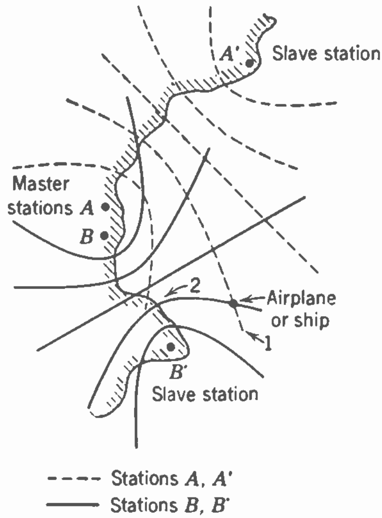


FIG. 9

FIG. 8. Shoran pulses from transmitter in plane trigger beacons. Transit times of pulses from two beacons enable one to obtain a "fix."

FIG. 9. Loran. The curves are lines of constant difference in distance from station pairs.

the plane is a measure of the *difference* of the distance of the receiver from the two stations, although not a measure of the actual distance from either station. A second pair of stations (master and slave) is required in order that the actual position of the plane may be fixed.

The manner in which the receiver location is fixed is shown in Fig. 9. The time difference in the reception of signals from stations A and A' fixes the position somewhere along line 1, and the time difference of signals from stations B and B' fixes the location somewhere along line 2. The intersection of these two curves gives the location of the receiver. The curves ("1" and "2" in the figure) are hyperbolas, and navigation by loran is sometimes termed "hyperbolic navigation."

The daytime range of loran is normally not greater than about 200 miles. At nighttime, however, the range is increased to greater than 1000 miles. This increase comes about because of the sky wave, which is useful only at nighttime. The accuracy with which positions can be determined by loran is about 1 mile in 700 miles with ground-wave propagation (daytime) and 4 miles in 700 miles with sky-wave propagation (nighttime).

Sometimes loran systems are operated in the low r-f range around 180 kc. The ground-wave range is increased by the use of low frequencies. Such systems are particularly useful in polar regions. In addition to the increased ground-wave range, these low-frequency signals are not as subject to interference from magnetic storms as are high-frequency signals.

Loran principles are also used in equipment operating at microwave frequencies. Such apparatus is called "Gee."

Index