FUNDAMENTALS of RADIO

Fundamentals
of
RADIO

by

EDWARD C. JORDAN

INSTRUCTOR IN ELECTRICAL ENGINEERING
OHIO STATE UNIVERSITY; ASSOCIATE,
INSTITUTE OF RADIO ENGINEERS

PAUL H. NELSON

ASSISTANT PROFESSOR OF ELECTRICAL ENGINEERING
THE UNIVERSITY OF CONNECTICUT

WILLIAM CARL OSTERBROCK

PROFESSOR OF ELECTRICAL ENGINEERING, UNIVERSITY
OF CINCINNATI; MEMBER OF A.I.E.E. AND I.R.E.

FRED H. PUMPHREY

PROFESSOR OF ELECTRICAL ENGINEERING
RUTGERS UNIVERSITY

LYNNE C. SMEBY

DIRECTOR OF ENGINEERING FOR THE NATIONAL ASSOCIATION
OF BROADCASTERS; MEMBER OF THE BOARD OF EDITORS OF
THE INSTITUTE OF RADIO ENGINEERS

EDITOR

W. L. EVERITT

PROFESSOR OF ELECTRICAL ENGINEERING, OHIO STATE
UNIVERSITY; FELLOW OF A.I.E.E. AND I.R.E.; MEMBER OF
THE BOARD OF DIRECTORS AND BOARD OF EDITORS OF THE
INSTITUTE OF RADIO ENGINEERS

PRENTICE-HALL, INC.

New York : 1942

First printing, June, 1942
Second printing, July, 1942
Third printing, July, 1942

PRINTED IN THE UNITED STATES OF AMERICA

Preface

THE purpose of this volume is to present the basic material of radio required for all types of radio work, both civil and military. The authors have covered each topic in such a way as to make clear the functioning of a complete radio system, and they have also laid the foundation for a more advanced study of the subject.

The reader need have only an elementary knowledge of algebra, which is reviewed briefly in Chapter 1. This chapter and the two following on direct and alternating current may be omitted if desired and used for later reference when the need arises. The remainder of the book has been written to give basic physical descriptions with a minimum of mathematics.

Technical radio work may be divided into operation, maintenance, development, and manufacture. An individual can operate some machines without knowing how they work, but if he wishes to maintain them, he must know the fundamental principles on which their operation is based. In the operation of the more complex radio systems, a knowledge of the equipment and the behavior of radio waves is essential in order to obtain the best results. A still greater understanding is needed on the part of the person who is to contribute to development and research.

W. L. Everitt

Table of Contents

CHAPTER 1

Mathematics of Radio

Need for Mathematics in Radio. Radio, like most of life itself, asks the questions, "How much?" or "How many?" How tall should the tower of a broadcasting station be? How long should a radio antenna be? What size of wire is needed on a radio receiver coil? Answers to these and many other questions are given (at least in part) by mathematics.

What is mathematics? Mathematics may be defined as a shorthand system which uses easy words (actually letters or other symbols) to simplify difficult ideas. In addition, rules are set up so that everyone may use these words or symbols in the correct way. One result is that the action of a certain piece of radio equipment may be predicted under given conditions because the way it acted under similar conditions at another time is known.

Understanding radio is easier when some of the rules of mathematics are known. The starting point of a system of mathematics is a set of numbers.

Our number system uses ten symbols or *digits*, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The symbol 3 always means three: three men, three radio sets, or three of whatever is being counted. The symbol 9 indicates nine objects, and so forth. If more than nine objects are to be counted it is necessary either to use other symbols or else to agree that the *place* of the symbol or digit in the number shall have a special meaning. The fact that place value is used makes our system better than most other systems man has tried. Thus the number 1,492 means 1 thousand + 4 hundreds + 9 tens + 2 units. The place of the digit in the number shows whether it means thousands, hundreds, tens, units, or something else. We say we "take" 9 tens, 4 hundreds, 2 units and 1 thousand. There is no limit to the size—large or small—of numbers that can be written in this way; distances between the stars in the sky are examples of the possible range in one direction.

The digit 0 (zero) is a number like the rest and plays an important part in the system. For example, in the number five hundred and two, no tens are to be taken; this fact is indicated by a zero in the tens place: 502.

No more than three digits are needed to write any number from 1 to 999; each digit shows how many hundreds, tens, or units to take by its place in the number; zero indicates that none are to be taken where it occurs.

Besides the saving of time in writing numbers in such a manner, the operations which may be performed with numbers are made much easier or, indeed, possible.

Positive and *negative* numbers also are needed. Of course, positive numbers alone are enough to count persons or objects, but everyone knows of temperatures above and below zero, distances above and below sea level, profits and losses, and even bank balances which people sometimes try to overdraw, that is, to get below zero. *Positive* numbers are those which extend on one side of zero; they are greater than zero. *Negative* numbers are those on the other side of zero; they are less than zero. A diagram, Fig. 1–1, will help to make this clear. Distances along a line are marked

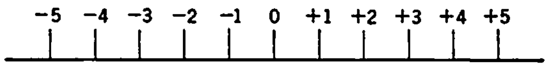


Fig. 1–1. Graph Showing Positive and Negative Numbers.

with numbers that correspond to the length of the line from zero. For each number there is a certain length. Distances measured to the right of a convenient point, marked 0, are called positive; distances measured to the left are negative. Thus, + 1, + 2, + 3, + 4, + 5, + 6, and so on, are positive numbers, and − 1, − 2, − 3, − 4, − 5, − 6, and so on, are negative numbers. If no sign is written before a number it is understood to be positive. The *absolute value* of a positive or a negative number is the value of the number without the positive or negative sign.

The Four Fundamental Operations. As long as whole objects only are to be counted, the natural numbers or *integers* are enough; that is, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and so on. These numbers and others are used in the four fundamental operations of addition, subtraction, multiplication, and division.

Addition is needed when two groups of things are to be combined. Of course, each group might be counted and then the combined group counted again, but a trial will show that instead of the objects themselves being counted, the numbers representing them may be added in a manner that never varies. That is, 20 apples added to 15 apples, 20 radios and 15 radios, or 20 objects of any kind combined with 15 objects of the same kind always amount to 35 objects in the whole group. Two conclusions follow from this process: (1) operations with numbers themselves, rather than with the objects they represent, may be performed; (2) if the operations are correctly carried out, the results are always right. Taking advantage of the properties of the system of numbers, it is necessary only to memorize sums of digits like 1 plus 1 equals 2, 1 plus 2 equals 3, and other basic combinations in order to add any numbers whatsoever.

Subtraction is the opposite or inverse of addition. If a television set has 35 tubes and a deluxe radio receiver has 24, the difference between the numbers of tubes is 11, which is obtained by subtracting 24 from 35. To check the correctness of the work, add the difference, 11, to the smaller number, 24, and the larger number, 35, is obtained.

Multiplication may be thought of as continued addition. If troops march past four abreast they may be counted as $4 + 4 + 4 + 4 = 16$ if there happen to be four ranks. This is the same as counting the four ranks only and multiplying by the number in each rank: that is, 4 times 4 equals 16.

Division is the inverse of multiplication, just as subtraction is the inverse of addition. Thus, if 60 is divided by 12 the result is 5; the inverse operation, 12 multiplied by 5, gives 60. Division by zero (0) is not possible because there is no number which, when multiplied by zero, will give any number except zero itself.

All of the ideas discussed are well known and are easy to understand; but they must be stated here as a foundation for other ideas, no more difficult to learn, ideas needed for an understanding of radio.

To save time in writing mathematics, certain *symbols of operation* are used. Among these are:

For addition: the plus sign (+). $21 + 12 = 33$.
For subtraction: the minus sign (−). $18 - 8 = 10$.
For multiplication: the multiplication sign (×), sometimes the dot (·); sometimes the quantities to be multiplied are simply written side by side. $5 \times 6 = 30$ or $5 \cdot 6 = 30$; $4 \times a = 4a$.

For division: the division sign ÷ , the bar ——, or the mark /. Thus $28 \div 7 = 4$, $\frac{28}{7} = 4$, $28/7 = 4$.

Sometimes, to avoid confusion, a complicated expression must be enclosed in parentheses ( ), brackets [ ], or braces { }. For example, consider $15 \times 5 - 2$. Does it mean $75 - 2$ or 15×3? By using parentheses confusion is avoided: $(15 \times 5) - 2 = 75 - 2 = 73$, $15 \times (5 - 2) = 15 \times 3 = 45$. When we use parentheses in a multiplication we can omit the × sign: $15(5 - 2)$ means the same as $15 \times (5 - 2)$, $(5)(6)$ means the same as 5×6 or $5 \cdot 6$.

Exercise 1–1. Perform the indicated operations.

1. $18 + 6 + 4 =$	7. $35 \div 7 =$	12. $\dfrac{3 \times 5 \times 8}{2 \times 3 \times 4} =$
2. $37 - 5 =$	8. $42/6 =$	
3. $30 + 6 - 10 =$	9. $7/0 =$	13. $(200)(0)(5) =$
4. $25 - 5 + 15 - 3 =$	10. $\frac{18}{9} =$	
5. $5 \times 9 \times 2 =$	11. $\dfrac{20 + 5}{3 + 2} =$	14. $(4 + 3)(2 + 5) =$
6. $3 \times 0 \times 7 =$	(*Note:* Add first.)	15. $(8 + 0)(3 + 1) =$

Fractions. As long as the objects to be counted are whole units, like tubes in a radio set, the natural numbers will do for counting; but if the problem is to divide one apple pie among three boys, the pie must be cut into slices. These slices will be *fractions* of the pie; each will be

$$\frac{\text{the whole pie}}{\text{three slices}} = \frac{1}{3}.$$

Thus, if one integer (natural number), called the *numerator*, is divided by another integer, called the *denominator*, the result is called a fraction. If the numerator is smaller than the denominator, the fraction is said to be a *proper fraction;* if the numerator is the larger, the fraction is said to be an *improper fraction.* Thus, $\frac{1}{2}$, $\frac{2}{3}$, $\frac{5}{8}$, $\frac{81}{82}$ are proper fractions; $\frac{3}{2}$, $\frac{5}{3}$, $\frac{11}{6}$, $\frac{201}{82}$ are improper fractions, The latter are frequently reduced to the sum of an integer and a proper fraction: $\frac{3}{2} = \frac{2}{2} + \frac{1}{2} = 1\frac{1}{2}$; $\frac{11}{6} = \frac{6}{6} + \frac{5}{6} = 1\frac{5}{6}$; $\frac{201}{82} = \frac{164}{82} + \frac{37}{82} = 2\frac{37}{82}$.

The four fundamental operations may be applied to fractions as well as to integers, but some care is necessary. For example, to add $\frac{3}{7}$ and $\frac{2}{5}$ they must be reduced to a *common denominator,* thus:

$$\frac{3}{7} + \frac{2}{5} = \frac{3 \times 5}{7 \times 5} + \frac{2 \times 7}{7 \times 5} = \frac{15 + 14}{35} = \frac{29}{35}.$$

The same process is needed in subtracting fractions; to subtract $\frac{4}{9}$ from $\frac{7}{8}$, proceed as follows:

$$\frac{7}{8} - \frac{4}{9} = \frac{7 \times 9}{8 \times 9} - \frac{4 \times 8}{8 \times 9} = \frac{63 - 32}{72} = \frac{31}{72}.$$

Fractions may be written in several forms which mean the same thing; thus $\frac{1}{2}$ is the same as $\frac{2}{4}$, since upon dividing the numerator and denominator of the latter by 2 the original $\frac{1}{2}$ is obtained.

Multiplication of fractions may be done without reducing to a common denominator; for example, $\frac{3}{4}$ times $\frac{5}{8}$ becomes

$$\frac{3}{4} \times \frac{5}{8} = \frac{3 \times 5}{4 \times 8} = \frac{15}{32}.$$

Division of fractions is easy if one rule is used: invert the *divisor* (fraction divided into another) and then multiply. Thus, dividing $\frac{3}{4}$ by $\frac{5}{8}$,

$$\frac{3}{4} \div \frac{5}{8} = \frac{3}{4} \times \frac{8}{5} = \frac{3 \times 8}{4 \times 5} = \frac{24}{20} = \frac{6}{5} = 1\frac{1}{5}.$$

Exercise 1-2. Perform the indicated operations.

1. $\frac{3}{7} + \frac{5}{8} =$
2. $\frac{5}{6} + \frac{2}{3} =$
3. $\frac{1}{3} + \frac{1}{4} + \frac{1}{5} =$
4. $\frac{5}{8} - \frac{4}{7} =$
5. $\frac{7}{8} - \frac{3}{4} =$
6. $\frac{35}{43} - \frac{1}{6} =$
7. $\frac{5}{9} \times \frac{4}{5} =$
8. $\frac{17}{64} \times \frac{19}{49} =$
9. $\frac{1}{3} \times \frac{2}{5} \times \frac{3}{7} =$
10. $\frac{5}{8} \div \frac{9}{16} =$
11. $\frac{13}{32} \div \frac{13}{15} =$
12. $\dfrac{\frac{5}{8} + \frac{3}{7}}{\frac{4}{5} + \frac{7}{16}} =$

Note: In No. 12, perform addition, then division.

The above are called *common fractions.* Another type of fraction is the *decimal fraction,* which utilizes the place values of the number system to better advantage.

If the numerator and denominator of $\frac{1}{4}$ be multiplied by 25, the fraction becomes $\frac{25}{100}$, which also equals $\frac{2}{10} + \frac{5}{100}$. This may be written 0.25, by extending the idea of place values discussed on p. 1. Any common fraction may be converted to a decimal fraction by dividing the numerator by the denominator. For example, $\frac{7}{8} = 0.875$; $\frac{11}{64} = 0.171875$; $\frac{1}{3} = 0.333 \cdots$, the dots indicating that no matter how long the division is continued, there will be still more digits.

As soon as common fractions have been converted to decimal fractions, they may be added, subtracted, multiplied, and divided just like integers. The following examples show the process.

$$0.875 + 0.125 = 1.000 \qquad 0.21 \times 0.3 = 0.063$$
$$0.625 - 0.0625 = 0.5625 \qquad 1.5 \div 0.5 = 3.0$$

It will be seen that in adding or subtracting, the periods or *decimal points* are always lined up with one another; in multiplying, the number of *decimal places* (digits to the right of the decimal point) in the result or product is the sum of the number of decimal places in the numbers multiplied together; in dividing, the decimal point may be located by setting the decimal point to the right in both the divisor and the *dividend* (number divided into) until the divisor is no longer a fraction, and locating the decimal point in the result at this place. An example will show how this is done.

To divide 1.728 by 0.12, write the figures either

```
0.12)1.728(____      or      0.12)1.728
```

Move the decimal point to the right in both numbers until the divisor (0.12) is no longer a fraction, thus:

```
                                 14.4
12)172.8(14.4       or       12)172.8
   12                           12
    52                           52
    48                           48
     48                           48
     48                           48
```

Exercise 1–3.

1. Change to decimal fractions: $\frac{3}{4}$, $\frac{5}{8}$, $\frac{9}{16}$, $\frac{17}{64}$, $\frac{72}{80}$, $\frac{2}{3}$, $\frac{13}{17}$.

2. Prepare a table of decimal equivalents of fractions of an inch in $\frac{1}{64}$-inch steps; that is, $\frac{1}{64}$, $\frac{1}{32}$, $\frac{3}{64}$, $\frac{1}{16}$, and so on.

3. Add: 0.125, 0.791, 0.345, 1.403, 7.142.

4. $0.784 - 0.038 =$ 5. $3.142 \times 7.553 =$ 6. $1.173 \div 1.42 =$

Accuracy and Significant Figures. Most of the numbers used in radio work are obtained by measurement of one sort or another. It is necessary, therefore, to look at the amount of accuracy which may be expected from these measurements, and especially how many decimal places should be used in the calculations based upon these measurements.

Many small instruments, or meters, have an accuracy stated as "Maximum error 3% of full-scale reading." This means that with a full-scale reading of 100 v*, for example, the actual voltage when the meter reads full scale is between 97 v and 103 v, 3 v being 3% of the full-scale reading. Unfortunately, the same 3-v error may also happen at lower scale readings, and if this instrument reads 10 v, the actual voltage may be between 7 v and 13 v.

To see what else this may lead to, *relative error* and *percentage of error* must be considered. Suppose that a meter reads 16.0 v and the reading should be 16.2 v. The *error* is 16.2 − 16.0 = 0.2 v, and the *relative error* is

$$\frac{\text{error}}{\text{true value}} = \frac{0.2}{16.2} = \frac{1}{81} = 0.0123.$$

Multiplying the last figure by 100 to convert it to percentage gives a *percentage of error* which is about $1\frac{1}{4}\%$. If another meter, used on a power line, read 16,000 volts when the correct reading was 16,200 volts, the error would be 16,200 − 16,000 = 200 v. The relative error would be

$$\frac{200}{16{,}200} = \frac{1}{81} = 0.0123,$$

or about $1\frac{1}{4}\%$ again. The position of the decimal point or the size of the quantities involved have nothing to do with the *relative error*.

The diameter of the earth has been carefully measured by accurate instruments. However, its average diameter is often stated as 8,000 miles, which is accurate only to the nearest thousand miles. A more accurate figure would be 7,900 miles, accurate to the nearest hundred miles; or 7,930 miles, accurate to the nearest ten miles; or even, 7,927, accurate to the nearest mile. Yet 8,000 miles is a useful expression if it is remembered that only the 8 in the thousands place means very much, that the zeros are just there to keep the 8 in its place. Another way of saying this is that 8 is the significant figure, the zeros not being significant. It is agreed that the significant figures shall be only those digits determined by measurement. In the expression 7,930 miles for the earth's diameter, only the figures 7,93 are significant; in 7,927, all figures are significant. When we are paying attention to significant figures, zeros following the last of the other digits do not count unless it is so stated. For example, the following numbers have two significant figures: 17,000; 0.00057; 95; 23,000,000. With certain instruments measurements may be made which are accurate to five significant figures, such as 60,103. If the instrument had read

* *Volts* is abbreviated *v*.

60,100, the last two zeros would not count as significant figures unless the fact was expressly so stated. Measurements and calculations in radio seldom are carried to more than three significant figures.

Now if two numbers which represent quantities obtained by measurement, each with three significant figures, are multiplied together, the product may consist of as many as six digits, such as $32.4 \times 41.4 = 1341.36$. There is nothing in the process of multiplying which will increase the accuracy of the original measurements; therefore, the product is accurate only to three significant figures and should be written 1340.

In general, the result of a calculation should be *rounded off* to the number of significant figures in the least accurate of the measured values used in the calculation. In rounding off numbers it is usual to take the next larger number if the last digit is greater than 5; to take the next smaller number if the last digit is less than 5; and if the last digit is 5 the preceding digit is increased by 1 if it is odd and left unaltered if it is even. Thus, 124 would be rounded off to 120 with two significant figures; 127 would be rounded off to 130; and 125 to 120; but 135 would be rounded off to 140.

Exercise 1–4. State the number of significant figures in the following.

1. 24,000.
2. 5,280.
3. 186,230.
4. 3.141597.
5. 0.0000543.
6. 0.00000006.
7. 0.08735.
8. 50,000.21.
9. 0.0809007500.
10. 0.142857.

Scientific Notation. Many very large and very small numbers are used in radio. Two schemes are used to get around the difficulty of writing the large number of digits required to express such numbers in ordinary place notation.

The first plan is to write the figure as a number less than 10 and then multiply by ten as many times as necessary to locate the decimal point correctly. Instead of writing out the tens, an exponent shows how many times ten is to be taken. For example, 10^2 means ten taken twice or $10 \times 10 = 100$; 10^3 means ten taken three times or $10 \times 10 \times 10 = 1{,}000$; the (^{2}) and the (^{3}) are called *exponents.* A table showing *powers of ten* and their *exponents* follows.

TABLE 1–1

Large Numbers			*Small Numbers*
$10^1 =$	10	$10^0 = 1$	$10^{-1} = 0.1$
$10^2 =$	100		$10^{-2} = 0.01$
$10^3 =$	1,000		$10^{-3} = 0.001$
$10^4 =$	10,000		$10^{-4} = 0.0001$
$10^5 =$	100,000		$10^{-5} = 0.00001$
$10^6 =$	1,000,000		$10^{-6} = 0.000001$

It is easy to see that when 1 is added to the exponent of a number to the left of an equal sign in this table, the number to the right of the sign is multiplied by 10; when 1 is subtracted from the exponent of a number to the left, the number to the right is divided by ten. Any number on the right of an equal sign might be represented by the corresponding number on the left; that is, for 100 write 10^2, for 1,000 write 10^3, and so forth. When used this way, the 10 is called the *base*.

To write 2,500 in this shorthand, consider that 2,500 is 2.5 times 1,000 or 2.5×10^3. Likewise, 2,500,000 is 2.5 times 1,000,000. From the table $1{,}000{,}000 = 10^6$. So 2,500,000 is written 2.5×10^6, which avoids writing a lot of zeros.

To write 0.0025 in a similar way, remember that 0.0025 is 2.5 times 0.001. From the table $0.001 = 10^{-3}$. Therefore 0.0025 is written 2.5×10^{-3} in scientific notation. For 0.0000025 the system works even better to reduce the number of zeros. This number is 2.5 times 0.000001; from the table $0.000001 = 10^{-6}$; therefore $0.0000025 = 2.5 \times 10^{-6}$. Now for a rule use the following:

RULE FOR WRITING NUMBERS IN SCIENTIFIC NOTATION. To write a given number in scientific notation, (1) move the decimal point to the right of the first digit which is not 0 (counting from the left of the given number) which will give a new number between 1 and 10; (2) multiply this new number by 10 with an exponent numerically equal to the number of places the decimal point has been moved; (3) make the exponent positive if the decimal point was moved to the left and negative if the decimal point was moved to the right.

The rule operates in reverse to change from scientific notation to ordinary place notation: Move the decimal point (to the right if the exponent is positive, to the left if negative) as many places numerically as indicated by the exponent of 10, supplying zeros as necessary. For example, $2.56 \times 10^2 = 256$; $1.86 \times 10^5 = 186{,}000$; $7.853 \times 10^{-3} = 0.007853$; $2.4 \times 10^{-5} = 0.000024$.

Exercise 1–5. Express in scientific notation.

1. 605,000,000,000,000,000,000,000.
2. The age of the earth, estimated as about 694,000,000,000 days.
3. One light-year, about 5,870,000,000,000 miles.
4. The distance from earth to the sun, about 93,000,000 miles.
5. The thickness of an oil film on water, 0.0000002 inch.
6. 0.000000000000000000000003.

Exercise 1–6. Express in ordinary place notation.

1. The diameter of the sun, about 8.6×10^5 miles.
2. The mass of the earth, about 6.6×10^{21} tons.
3. Diameter of red corpuscle in blood, about 3×10^{-5} inch.
4. 6.4×10^{-2}. 5. 1.20×10^9. 6. 6.28×10^{-6}.

Besides being easily and quickly written, numbers in scientific notation are handy when multiplying and dividing. Suppose that 2.2×10^3 is to be multiplied by 3×10^2. First multiply 2.2 by 3, which gives 6.6. For the 10's, consider that $10^3 = 1{,}000$ and $10^2 = 100$. $1{,}000 \times 100 = 100{,}000 = 10^5$ (from the table). But 10^5 equals 10^{2+3}. Rewriting the solution gives

$$
\begin{aligned}
(2.2 \times 10^3)(3 \times 10^2) &= (2.2 \times 3)(10^3 \times 10^2) \\
&= 6.6 \times 10^{2+3} = 6.6 \times 10^5.
\end{aligned}
$$

Rules of operation with exponents. For any given base, such as 10, the following rules apply:

To multiply, add exponents: $10^4 \times 10^2 = 10^{4+2} = 10^6$.

To divide, subtract exponents: $10^4 \div 10^2 = 10^{4-2} = 10^2$.

Exercise 1-7. Perform the indicated operations.

1. $(1.86 \times 10^5)(3.6 \times 10^2) =$
2. $(9.3 \times 10^7) \div (1.86 \times 10^5) =$
3. If a true value is 4.3225×10^5, what is the relative error in using the value 4.33×10^5?

The second method of writing very large or small quantities is to use large or small units. These units are named from the ordinary units by adding prefixes. A table of such prefixes follows.

TABLE 1-2

Multiply the known number of →	mega-	kilo-	unit	centi-	milli-	micro-	milli-micro-	micro-micro-
to obtain the unknown number of ↓	by ↓	by ↓	by ↓	by ↓	by ↓	by ↓	by ↓	by ↓
mega-	1	10^{-3}	10^{-6}	10^{-8}	10^{-9}	10^{-12}	10^{-15}	10^{-18}
kilo-	10^3	1	10^{-3}	10^{-5}	10^{-6}	10^{-9}	10^{-12}	10^{-15}
unit	10^6	10^3	1	10^{-2}	10^{-3}	10^{-6}	10^{-9}	10^{-12}
centi-	10^8	10^5	10^2	1	10^{-1}	10^{-4}	10^{-7}	10^{-10}
milli-	10^9	10^6	10^3	10	1	10^{-3}	10^{-6}	10^{-9}
micro-	10^{12}	10^9	10^6	10^4	10^3	1	10^{-3}	10^{-6}
millimicro-	10^{15}	10^{12}	10^9	10^7	10^6	10^3	1	10^{-3}
micromicro-	10^{18}	10^{15}	10^{12}	10^{10}	10^9	10^6	10^3	1

For example, a kilocycle is one thousand cycles; a megacycle is a million cycles; a microfarad is one millionth of a farad; a millihenry is one thousandth of a henry; and so on. The meanings of cycles, farads, and henries will be explained later.

To use the table, find the standard unit in the top row; proceed down this column to the line containing the desired unit; the number at this point multiplied by the standard unit gives the desired unit. For example, to convert millihenries into microhenries, find milli- at the top of the table, go down this column to the line marked micro-, and find 10^3. Therefore, multiply millihenries by 1,000 to get microhenries.

Symbols of Algebra. *Algebra* is a special form of language, which uses *symbols* in place of words to express ideas. Just as the word "tree" brings to mind certain objects, the letter x in algebra represents certain ideas. There are many sorts of trees: large or small in size, slender or bushy in shape, light green or dark green in color; but the idea of a tree will not be confused with the idea of a building, or a boat. In the same way, an algebraic symbol may represent different numbers at various times but will generally be used for the same sort of numbers, subject to the same rules of operation.

Symbols are used in mathematics so that ideas may be put down on paper and talked about easily; using words for the same purpose would often cause confusion. What method shall be used to choose these symbols? First, symbols must be as brief as possible; second, their meaning must be generally agreed upon by those who use them. One choice is to use the initial letter of a word, like R for resistance, C for capacitance, w for width, P for pressure. Sometimes symbols for certain things have been used so long they are accepted almost universally; for example, I for current, E for voltage, L for inductance, and so on. (The meaning of these terms will be explained later.) Other *symbols* generally indicate *numbers*; such as x, y, z (last few letters of the alphabet), which usually mean quantities *unknown* or *variable;* a, b, c (first letters of alphabet) often stand for *known* or *unvarying* quantities. Mathematics packs much meaning into one symbol to make the symbol useful where words or numbers would not serve.

The symbols used in mathematics are precise in meaning; words also are used with meanings more precise and more limited than they may have in everyday conversation, and sometimes these meanings are different from the common ones. Some definitions of such words follow.

An *algebraic expression* is any combination of numbers, letters used for numbers, and signs of operation written according to the rules of algebra; like $10x$, $t + t$, $mx + b$. Note that $10x$ means 10 times x; this is a product and the 10 and the x are the two factors. *Factors* of a product are the numbers which, when multiplied together, form the product. In algebra numbers and letters or combinations of letters written together without signs of operation are to be multiplied; xy means x times y, $10a$ means 10 times a. In the product $10a$, 10 is the *numerical coefficient* or factor and a is the *literal* (letter) *coefficient* or factor. When x is written alone, the coefficient 1 is understood; x means 1 times x. A *term* is any expression like $10x$, t, mx, b, $5r$, and so forth. *Similar terms* are those with the same letter or letters, such as $10x$ and mx. If terms do not have the same letters (or literal coefficients) they are *dissimilar terms*. Similar terms may be combined by *adding* the *numerical coefficients;* remember that subtracting is the same as adding a term with a minus sign ahead of it. For example,

$$3x + 4x = (3 + 4)x = 7x; \qquad 3a - 2a = (3 - 2)a = 1a = a.$$

Exercise 1–8. Combine the following terms.

1. $10m + 8m + 6m =$
2. $40v - 30v =$
3. $90q + 50q - 40q =$
4. $\frac{1}{4}b + \frac{3}{4}b =$
5. $\frac{5}{8}e - \frac{1}{3}e =$
6. $9.5x - 7.2x =$
7. $19k - 1.73k =$
8. $11h + 12h - h =$

Some Laws of Algebra. In *addition* terms may be combined or grouped in any order. That is, $a + b = b + a$ and $a + b + c = (a + b) + c = a + (b + c)$. If dissimilar terms occur, they may be combined in groups of similar terms; for example, $50a + 98b + 5a + 3b + a = (50 + 5 + 1)a + (98 + 3)b = 56a + 101b$.

Multiplication and *division* are carried out with the coefficients of similar terms, like $7 \times 2a = 14a$; $\frac{1}{2} \times 10w = 5w$; $\frac{10x}{2} = 5x$; $\frac{2(6x - 3x)}{3} = 2x$; $\frac{18x}{6x} = 3$ (a number only).

Exercise 1–9. Perform the indicated operations.

1. $4a + 6b + 5a + 7b + a =$
2. $6.2e + 8.9d + 7.8c + 10.2d =$
3. $18 \times 2f =$
4. $20 \times \frac{4}{5}x =$
5. $10 \times 9.2y$
6. $2.5 \times 6z =$
7. $\frac{20a}{4} =$
8. $\frac{3.68r}{4} =$
9. $4.5d \div 1.5d =$
10. $\frac{3a + 4x + 2y}{2} =$

Just as $10 \times 10 = 10^2$, $a \times a = a^2$, and $a \times a \times a = a^3$. Similar terms with the same exponents may be combined; that is, $3a^2 + 4a^2 = 7a^2$. Expressions containing several terms are called *polynomials,* for example $3x^3 + 4x^2 - 5x + 2$, which is a polynomial in *descending powers* of x. Polynomials may be added by combining similar terms as follows:

$$\begin{array}{r} 3x^3 + 4x^2 - 5x + 2 \\ -\,2x^3 - x^2 + 8x - 1 \\ \hline x^3 + 3x^2 + 3x + 1 \end{array}$$

Exercise 1–10.

1. Add $2a^2 - 3ab + 4b^2$ to $6a^2 - 2ab - 4b^2$.
2. Add $x^2 - 4x + 10$ to $2 - 6x - 2x^2$.
3. Add $r^2 + 3rs - 5s^2$, $2r^2 + 7s^2$, and $-5r^2 - rs - 5s^2$.
4. Subtract $-2a^2 - 8a + 14$ from $a^2 - 4a + 4$.

To make the step from arithmetic to algebra easier, here is another example worked out. Suppose 2,565 is to be added to 5,331. In the positional and scientific notation these numbers can now be written

$$\begin{array}{r} 2 \times 10^3 + 5 \times 10^2 + 6 \times 10 + 5 \\ 5 \times 10^3 + 3 \times 10^2 + 3 \times 10 + 1 \\ \hline 7 \times 10^3 + 8 \times 10^2 + 9 \times 10 + 6 \end{array}$$

Incidentally, numbers are usually read in this form: seven thousand, eight hundred, and ninety-six. The additions of the digits are done from memory and the 10's with exponents are carried through without change. Now if x be substituted for 10, the resulting *algebraic expression* may be added in the same way.

$$\begin{array}{r} 2x^3 + 5x^2 + 6x + 5 \\ 5x^3 + 3x^2 + 3x + 1 \\ \hline 7x^3 + 8x^2 + 9x + 6 \end{array}$$

Now any other number, such as 2 or 3, may be put in place of x and the result will still be correct. The x stands for any number.

The same rules for exponents apply as stated earlier for the base 10; that is, $a^2 \times a^3 = a^{2+3} = a^5$; $\dfrac{a^2}{a^3} = a^{-1}$; $(a^2)^3 = a^2 \times a^2 \times a^2 = a^6$; $a^{\frac{1}{2}} \times a^{\frac{1}{2}} = a = \sqrt{a} \times \sqrt{a}.$ $a^{\frac{1}{2}} = \sqrt{a}$

Rule of signs. In multiplying or dividing, if the terms have *like signs* the result is *positive* (has plus sign); if the terms have *unlike signs* the result is *negative* (minus sign).

$$\begin{array}{ll} (+a) \times (+b) = +ab & (+a) \div (+b) = +\dfrac{a}{b} \\ (-a) \times (-b) = +ab & (-a) \div (-b) = +\dfrac{a}{b} \\ (+a) \times (-b) = -ab & (+a) \div (-b) = -\dfrac{a}{b} \\ (-a) \times (+b) = -ab & (-a) \div (+b) = -\dfrac{a}{b} \end{array}$$

Example.

$(+3a^2b) \times (-2a^3b^2c) = (+3)(-2)(a^2)(a^3)(b)(b^2)(c) = -6a^5b^3c.$

Exercise 1–11. Perform the indicated operations.

1. $(2x^2)(3x^5) =$
2. $h^4(-h^3) =$
3. $(-5d)^3 =$
4. $(5E)(-3E^3)(-2E^2) =$
5. $-(2t)^3 =$
6. $5y(-3x^2y) =$
7. $(-3a^3t)^2 =$
8. $-(3a^3t)^2 =$
9. $e(8e^3)(-\frac{3}{4}e^5) =$
10. $(-2nd)(-3n)(-4d^2) =$
11. $(-x^2y^6)^3 =$
12. $(2dc^5)^2 =$

Removal of parentheses (and of other signs of grouping) from expressions is easy if certain rules are followed. These signs of grouping are often used to indicate multiplication of terms, and in the course of solving a problem it may be necessary that they be removed systematically. Therefore:

(1) Apply the rule of signs (above).

Example:

$$2(a + b - c) = 2a + 2b - 2c.$$

Also,

$$-3a(a^2 - 3a + 5) = -3a^3 + 9a^2 - 15a.$$

(2) If no coefficient is indicated for the group, consider the coefficient to be 1.

Example:

$$(a + b) = 1(a + b) = a + b.$$

Also,

$$-(x - 2y) = -1(x - 2y) = -x + 2y.$$

(3) Perform indicated multiplications or divisions first; then addition or subtraction.

Example:

$$5 - 3(a + 2b) = 5 - 3a - 6b.$$

This does not equal $2(a + 2b)$.

(4) Remove signs of grouping one set at a time, starting with the innermost set.

Example:

$$-3c[a + 2b(b - c)] = -3c[a + 2b^2 - 2bc] = -3ac - 6b^2c + 6bc^2.$$

Exercise 1–12. Remove parentheses and signs of grouping and combine similar terms (that is, simplify).

1. $4(2b + 3) - 7 =$
2. $4n - 4(n + 8) =$
3. $-3(5d - 4) + 12 =$
4. $4a(2a + b - c) + 7ac =$
5. $6s - 3(r + s + t) + (2r - s + 3t) =$
6. $-(8d - 3) + (-3d + 4) =$
7. $6n^2 - (3n + 2) + 2n(n - 5) =$
8. $5[2n - 6(n + 2)] - 3n + 8 =$
9. $3a[6 + 2(a - 3)] + 10a^2 =$
10. $s\{s - 2(3 - 5s) + 7\} - 2s^2 =$
11. $2x - 3[2x - 3(x - 5)] =$
12. $4y + 2y[3y - (5 - y) + 6] =$
13. $8r - [7 - 2(3r - 5)] =$
14. $9x^2 + 3x + 2[2x - 5x(x - 1) - 10] - (2x^2 + 7x) =$
15. $19x - 3[x - 3(x - 2\{4y + 3\} - 2y) + 3x] - y^2 =$

In the multiplication of algebraic expressions certain combinations occur so often that they need special study, so that they may be known and the answers obtained quickly. These combinations are often called *special products*. If the special product is recognized, its factors may often be found by inspection. *Factoring* is defined as the process of finding two or more expressions whose product is a given expression.

The simplest sort of factoring is dividing out a *common factor*. For example, to factor

$$3x^4 - 9x^3 + 12x^2$$

it is possible to divide out a 3, an x^2, or a $3x^2$. The latter, being the largest, is usually the factor desired; thus,

$$\begin{aligned} 3x^4 - 9x^3 + 12x^2 &= 3(x^4 - 3x^3 + 4x^2) \\ &= x^2(3x^2 - 9x + 12) \\ &= 3x^2(x^2 - 3x + 4). \end{aligned}$$

Exercise 1–13. Factor the following.

1. $7xy - 14xz =$
2. $5a^2b - 25ab^2 =$
3. $2\pi r^2 + 2\pi rh =$
4. $8xy^2 - 16xy + 12x =$
5. $a^2b^2 - 3ab^2 + 4a^3b - 12ab =$
6. $24m^2n - 6mn^2 + 36m^2n^2 - 42m^3 =$

Another product which occurs frequently is the product of two *binomials* (expressions having two terms). By multiplication, like that used with numbers, it is found that

$$\begin{array}{rrr} 2x + & 6 & \\ x - & 5 & \\ \hline 2x^2 + & 6x & \\ & - 10x & - 30 \\ \hline 2x^2 - & 4x & - 30 \end{array}$$

or

$$\underset{\text{A}}{\overset{\uparrow}{(}}2x + \underset{\text{B}}{\overset{\uparrow}{6}})\underset{\text{C}}{\overset{\uparrow}{(x}} - \underset{\text{D}}{\overset{\uparrow}{5}}) = 2x^2 - 4x - 30$$

From this example it may be seen that (1) the product usually will have three terms; (2) the first term is the product of the first terms in the binomials; (3) the second term is the sum of the products of the two outer terms and the two inner terms, that is, (A)(D) + (B)(C); (4) the last term is the product of the last terms of the binomials. The rule of signs must be applied at all times.

Exercise 1–14. Write out the following products according to the four steps above.

1. $(n + 3)(n + 2) =$
2. $(a - 5)(a + 2) =$
3. $(p - 4)(p - 5) =$
4. $(a + b)(a + 2b) =$
5. $(x^2 - 4)(x^2 - 12) =$
6. $(th + 7)(th - 12) =$
7. $(3x + 1)(3x - 3) =$
8. $(x^2 - 11y^2)(x^2 + 5y^2) =$

This method may be used in reverse; that is, given the product the factors may be found. For example, to factor $9x^2 + 18x + 5$, two binomials are needed having their first terms factors of $9x^2$, having their second terms factors of 5, and having $18x$ as the sum of the product of the two outer terms plus the product of the two inner terms of the trial factors (A)(D) + (B)(C). The correct factors are found by trying out different combinations until the right ones are obtained. Several possible combinations may be set down, but only one meets the third requirement.

	Trial factors	Second term	Conclusion
First attempt...........	$(x + 5)(9x + 1)$	$x + 45x = 46x$	No good.
Second attempt.........	$(x + 1)(9x + 5)$	$5x + 9x = 14x$	No good.
Third attempt..........	$(3x + 1)(3x + 5)$	$15x + 3x = 18x$	Correct.

Practice, of course, will help in finding the right combination.

Exercise 1–15. Factor the following.

1. $n^2 + 5n + 6 =$
2. $x^2 - 6x + 5 =$
3. $a^2 - 9a + 20 =$
4. $6n^2 + n - 2 =$
5. $x^2 + 3xy + 2y^2 =$
6. $15a^2 + 22a + 8 =$
7. $2 + 7x + 5x^2 =$
8. $a^4 - 6a^2b^2 - 55b^4 =$

Certain binomials occur often as equal factors of squares. These are of the type $(a + b)$ or $(a - b)$. The square of either of these consists of

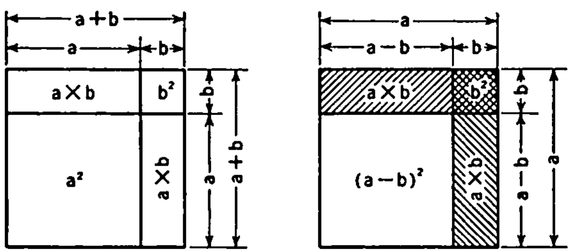


Fig. 1–2. Illustrating Binomial Squares by Areas.

the sum of three terms: (1) the square of the first term in the binomial; (2) twice the product of the two terms of the binomial; (3) the square of the second term of the binomial. That is,

$$(a + b)^2 = (a + b)(a + b) = a^2 + 2ab + b^2;$$
$$(a - b)^2 = (a - b)(a - b) = a^2 - 2ab + b^2.$$

This process is sometimes used with numbers as a short-cut method to find their squares. For instance,

$$(23)^2 = (20 + 3)^2 = (20)^2 + 2(20)(3) + 3^2$$
$$= 400 + 120 + 9 = 529;$$
$$(29)^2 = (30 - 1)^2 = (30)^2 - 2(30)(1) + 1^2$$
$$= 900 - 60 + 1 = 841.$$

Using the formulas for these squares is easier than multiplying out the usual way. Diagrams illustrating the binomial squares are shown in Fig. 1–2.

Exercise 1–16. Perform the following operations.

1. $(x + 3)^2 =$
2. $(x - 5)^2 =$
3. $(x + 6y)^2 =$
4. $(a - c)^2 =$
5. $(2x - 3)^2 =$
6. $(3a + 5)^2 =$
7. $(a - 3b)^2 =$
8. $(3k^2 - 1)^2 =$
9. $(5r^2 + s^2)^2 =$
10. $(a^3 + 12)^2 =$
11. $(h - \frac{1}{4})^2 =$
12. $(\frac{2}{3}a + \frac{1}{2}b)^2 =$

It is easy to factor expressions of this sort; simply take the square root of the first and the last terms and write the sign of the middle term between them. For example,

$$a^2 + 6a + 9 = (a + 3)^2,$$
$$4x^2 - 12x + 9 = (2x - 3)^2.$$

Note: Be sure that the middle term is twice the product of the square roots of the first and last terms.

Exercise 1–17. Factor the following expressions.

1. $x^2 - 12x + 36 =$
2. $16x^2 - 40xy + 25y^2 =$
3. $25a^4 + 120a^2b^2 + 144b^4 =$

Another useful combination is the result of multiplying the sum and difference of the same two terms. Thus,

$$(a + b)(a - b) = a^2 - b^2.$$

Factoring the difference of two squares is done by taking the square root of each square and making the factors the sum and difference of these roots.

$$4n^2 - 36 = (2n + 6)(2n - 6)$$

This special product may also be used in arithmetical problems. If the product of 31 and 29 be desired, it may be considered as the product $(30 + 1) \times (30 - 1) = 900 - 1 = 899$.

Exercise 1–18. Perform the operations indicated.

1. $(x + y)(x - y) =$
2. $(10 - a)(10 + a) =$
3. $(4a + 3b)(4a - 3b) =$
4. $36r^2 - s^2 =$
5. $-81v^2 + 25u^2 =$
6. $12c^3 - 3c =$
7. $63 \times 57 =$
8. $75 \times 45 =$

Solving Equations. It is very fortunate that the behavior of the equipment and devices used in radio is rather uniform; that is, if a circuit behaves in a certain fashion at one time it will behave the same way at some other time. This uniformity permits the writing of *equations* or *formulas* to forecast the behavior of radio circuits. Some formulas are simple and others look complicated; but after the equation has been arranged in a useful form, values (numbers) may be placed in it and the equation *solved* or *evaluated*. The usual result is information about the action of the circuit under particular conditions. So it is necessary to get some practice in working with equations.

As usual, there are definitions and rules. The two parts of the equation separated by the equality sign are usually called the *members;* either *left-hand* and *right-hand* member or *first* and *second* member, respectively. Any operation of algebra may be performed on one member of an equation as long as the same operation is performed on the other member.* That is, the *laws of operation* state that if *equals* are added to, or subtracted from, or multiplied by, or divided by *equals*, the results are *equal*. For example, suppose an equation

$$\lambda = \frac{300{,}000}{f}$$

is given and λ (read lambda) is given as 600; how may f be found?

Solution: First multiply by f on both sides of the equal sign,

$$f\lambda = \frac{300{,}000f}{f},$$

and then divide both members of the equation by λ:

$$f\frac{\lambda}{\lambda} = (300{,}000)\frac{f}{f\lambda}\cdot$$

but $\frac{\lambda}{\lambda} = 1$ and $\frac{f}{f} = 1$, so

$$f = \frac{300{,}000}{\lambda}\cdot$$

Substitute the value of $\lambda = 600$ in this equation and obtain

$$f = \frac{300{,}000}{600} = 500.$$

An important formula in radio is Ohm's Law, usually written in symbols as $E = IR$. The meaning of the symbols will be explained later, but this equation will serve as another example of the operations which may be performed on an equation.

Following the laws of operation, let both sides be divided by I. That is,

$$\frac{E}{I} = \frac{IR}{I} = R, \quad \text{or} \quad R = \frac{E}{I}\cdot$$

Again, let both sides be divided by R, and obtain

$$\frac{E}{R} = \frac{IR}{R} = I, \quad \text{or} \quad I = \frac{E}{R}\cdot$$

* Division by 0 is not possible.

Still another change may be made; suppose both sides of the equation are multiplied by I; then

$$EI = (I)(IR) = I^2R.$$

Just above, it is shown that $I = \frac{E}{R}$; substituting this value for I in the last equation makes it

$$E\frac{E}{R} = \frac{E^2}{R} = I^2R = EI.$$

The rule for the last operation is that *things equal to the same thing are equal to each other.*

Exercise 1–19.

1. Given $R = 20$, $I = 5$, find E.
2. Given $E = 246$, $R = 600$, find I.
3. Given $E = 1$, $I = 2 \times 10^{-2}$, find R.
4. Given $E = 115$, $I = 2.8$, find I^2R.

Sometimes it is not possible to evaluate an equation exactly because some factors in it are not exact; for example, a formula containing $\pi = 3.1416$ (approximately), such as

$$X_L = 2\pi fL \qquad \text{or} \qquad X_C = \frac{1}{2\pi fC}.$$

In this case, the number used for π should have the same number of significant figures as the rest of the data; 6.28 is often used for 2π.

Exercise 1–20. Perform the following calculations, using the appropriate formula above.

1. Given $f = 60$, $L = 1$, find X_L.
2. Given $f = 1.5 \times 10^6$, $L = 5.5 \times 10^{-6}$, find X_L.
3. Given $f = 500$, $C = 2 \times 10^{-6}$, find X_C.
4. Given $f = 3.105 \times 10^6$, $C = 0.85 \times 10^{-8}$, find X_C.
5. Given $X = X_L - X_C$, $f = 10^6$, $L = 1.4 \times 10^{-4}$, $C = 1.8 \times 10^{-10}$, find X.

Clearing fractions is often necessary in solving equations. Consider the equation

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

The first step is to place the denominator over its own common denominator:

$$R = \frac{1}{\frac{R_2 + R_1}{R_1R_2}}.$$

Finally, the denominator is inverted and multiplied by the numerator, which in this case is 1:

$$R = \frac{R_1R_2}{R_2 + R_1}(1) = \frac{R_1R_2}{R_1 + R_2}.$$

Exercise 1–21. Perform the following operations, based on the formula above.

1. Given $R_1 = 175$, $R_2 = 362$, find R.
2. Work out a formula for R_1, given R and R_2.

Quadratic Equations. An equation of the type

$$ax^2 + bx + c = 0$$

is called a *quadratic equation*. The coefficients a, b, c are any numbers, positive or negative, and x is a variable whose value is to be found. Special equations of this type have already been considered on pp. 14–15, but sometimes it is difficult to find the necessary factors. *To solve the quadratic equation,* that is, to find a formula which will always give the correct value of x, first divide the entire equation by a, the coefficient of x^2, to give

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Now subtract the term $\frac{c}{a}$ from both sides of the equation:

$$x^2 + \frac{b}{a}x = -\frac{c}{a}.$$

From the discussion on p. 15, it will be seen that the third term in a perfect square is the coefficient of x divided by 2 and then squared. Performing this operation, and adding this term to both sides of the equation, we get

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}.$$

The square root of both sides may now be taken. The left-hand side has now been made into a perfect square, so it may be factored; on the right-hand side the operation can only be indicated. Since both + and − quantities have the same sign when squared (rule of signs), both signs must be used before the square-root sign or *radical* ($\sqrt{\ }$) on the right. So

$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}},$$

or

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

The process used to work out this formula is known as "completing the square." Now an example of its use will be given.

Consider the equation

$$4x^2 + 12x + 9 = 0.$$

This is a quadratic form, with $a = 4$, $b = 12$, and $c = 9$. Putting these figures in the formula.

$$x = \frac{-12 \pm \sqrt{(12)^2 - 4(4)(9)}}{2(4)}$$
$$= \frac{-12 \pm \sqrt{144 - 144}}{8} = -\frac{12}{8} = -\frac{3}{2}.$$

Since equations do not always yield a solution as readily as the one above, consider another example. Given the equation

$$2x^2 + 5x - 12 = 0,$$

it is seen that $a = 2$, $b = 5$, $c = -12$. Substituting these values in the formula

$$x = \frac{-5 \pm \sqrt{5^2 - 2(4)(-12)}}{2(2)} = \frac{-5 \pm \sqrt{25 + 96}}{4}.$$

With the plus sign,

$$x = \frac{-5 + 11}{4} = \frac{6}{4} = \frac{3}{2}.$$

With the minus sign,

$$x = \frac{-5 - 11}{4} = \frac{-16}{4} = -4.$$

As a check on the result, substitute each of the values of x in the given equation:

$x = -4$	$x = 1\frac{1}{2} = \frac{3}{2}$
$2(-4)^2 + 5(-4) - 12 = 0$	$2(\frac{3}{2})^2 + 5(\frac{3}{2}) - 12 = 0$
$2(16) - 20 - 12 = 0$	$2(\frac{9}{4}) + \frac{15}{4} - 12 = 0$
$32 - 32 = 0$	$12 - 12 = 0$
$0 = 0$	$0 = 0$

Exercise 1–22. Solve the following quadratic equations.

1. $x^2 + 2x - 3 = 0$
2. $x^2 - 8x + 15 = 0$
3. $x^2 - 9x - 22 = 0$
4. $2x^2 - 9x + 4 = 0$
5. $6x^2 - 5x + 1 = 0$
6. $5x^2 - 11x + 2 = 0$
7. $12y^2 + 25y + 12 = 0$
8. $x^2 + 19 = 20x$
9. $x^2 + 2x = 48$
10. $x^2 + 2ax - 3a^2 = 0$

Trigonometry. *Trigonometry* is that branch of mathematics which deals with the properties of triangles. Actually, the ideas of trigonometry have been extended to solve many problems in radio, as will be seen later.

A *right triangle* is a triangle in which one of the angles is a *right angle* or *90°*, like the triangle in Fig. 1–3. For convenience, the angles are marked with the capital letters A, B, C (with C at the right angle) and the side opposite each angle is marked with the corresponding small letter a, b, c.

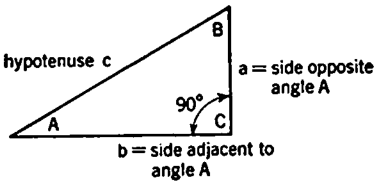


FIG. 1–3. Right Triangle with Parts Labeled.

In geometry it is shown that the sum of the angles in any triangle is 180°; that is, $A + B + C = 180°$. But since $C = 90°$, $A + B = 180° - C = 180° - 90° = 90°$. Therefore A and B each must be less than 90° (*acute angles*) and the following relations must be true:

$$A + B = 90°; \qquad A = 90° - B; \qquad B = 90° - A.$$

Now consider two triangles like ABC and $AB'C'$ in Fig. 1–4, one of which is larger than the other. In geometry it is shown that these two triangles, ABC and $AB'C'$, are *similar*, that is, their angles are equal and their sides are proportional to one another. Therefore, it may be said that

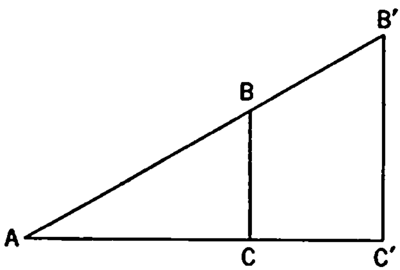


FIG. 1–4. Triangles ABC and $AB'C'$ are Similar.

$$\frac{BC}{AC} = \frac{B'C'}{AC'} = \text{a constant.}$$

No matter what the size of triangle $AB'C'$, as long as the shape is the same as that of the triangle ABC this ratio will be true. If the angles were changed, the ratio would change also. There is a definite relation between an acute angle of a right triangle and the ratios of the sides.

To talk about these ratios, it is handy to name them according to the following definitions, which refer to Fig. 1–3.

If A is an acute angle in any right triangle, then by definition:

$$\text{sine of } A = \frac{\text{length of side opposite angle } A}{\text{length of hypotenuse}} = \sin A = \frac{a}{c};$$

$$\text{cosine of } A = \frac{\text{length of side adjacent to angle } A}{\text{length of hypotenuse}} = \cos A = \frac{b}{c};$$

$$\text{tangent of } A = \frac{\text{length of side opposite angle } A}{\text{length of side adjacent to angle } A} = \tan A = \frac{a}{b}.$$

The abbreviations sin, cos, and tan are commonly used in place of the corresponding complete names sine, cosine and tangent. Other ratios may be formed but are not needed here. It is important to remember that these ratios depend only upon the angle in the right triangle and not upon the size of the triangle; the ratios are numbers.

It may be shown that in any *right triangle* the square on the *hypotenuse* is equal to the *sum of the squares* on the *two sides;* this generalization is called the *hypotenuse rule* or the *Theorem of Pythagoras.* That is,

$$a^2 + b^2 = c^2.$$

From which,

$$a = \sqrt{c^2 - b^2} \quad \text{and} \quad b = \sqrt{c^2 - a^2}.$$

Values for the ratios or *trigonometric functions* may be worked out for certain angles quite easily. Consider a triangle with three equal sides, each one unit in length. Since the sides are equal the angles opposite them must be equal. The sum of the three equal angles is 180° and so each angle must be 60°, as shown in Fig. 1–5. A line drawn from the top of the triangle to the center of the lower side will divide the lower side (or base) into two portions, each ½ unit long, and will divide the triangle into two portions, each of which is a right triangle. Taking one portion of the

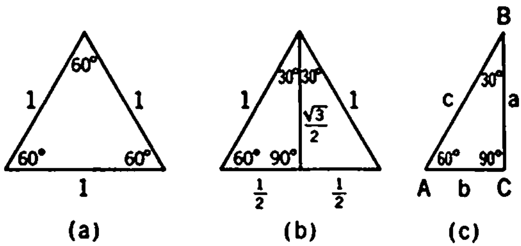


FIG. 1-5. Equilateral Triangle and 30°–60° Triangle.

triangle (Fig. 1–5c) and lettering the angles and sides as stated at the first of this section, the ratios can be evaluated for the angles in this triangle. From the hypotenuse rule the side a is

$$a = \sqrt{c^2 - b^2} = \sqrt{1^2 - (\tfrac{1}{2})^2} = \sqrt{1 - \tfrac{1}{4}} = \sqrt{\tfrac{3}{4}} = \frac{\sqrt{3}}{2} = 0.87.$$

From the definition,

$$\sin A = \frac{a}{c} = \frac{0.87}{1} = 0.87 = \frac{\sqrt{3}}{2} = \sin 60°.$$

From the definition,

$$\cos A = \frac{b}{c} = \frac{1}{2} \div 1 = 0.50 = \cos 60°.$$

Likewise,

$$\tan A = \frac{a}{b} = \frac{\sqrt{3}}{2} \div \frac{1}{2} = \sqrt{3} = 1.73 = \tan 60°.$$

From the fact that angle $A = 60°$ it is found that $\sin 60° = 0.87$, $\cos 60° = 0.50$, and $\tan 60° = 1.73$.

The same triangle may be used to get the values for the trigonometric ratios or functions for 30°. The side b is opposite angle B and the side a is adjacent to angle B; angle B is 30°. From the definition,

$$\sin B = \sin 30° = \frac{b}{c} = 0.50;$$

also,

$$\cos B = \cos 30° = \frac{a}{c} = 0.87;$$

and

$$\tan B = \tan 30° = \frac{b}{a} = 0.58.$$

Exercise 1—23.

1. Draw a right triangle whose sides are 3 inches and 4 inches long; find the length of the hypotenuse.

2. The smaller acute angle in the right triangle of sides 3 inches and 4 inches is about 37°. From this triangle find sin 37°, cos 37°, and tan 37°, approximately.

3. What is the relation between sin 30° and cos 60°? Between sin 60° and cos 30°?

4. For the triangles studied, is this a true statement: $\sin (90° - A) = \cos A$?

It is not necessary to calculate all of the values needed to solve problems, because tables like Table 1–3 (p. 25) are available.

To use the table, proceed as follows.

(1) To find the functions of 37°, go down the left-hand column to 37° and on this line find $\sin 37° = 0.6018$, $\cos 37° = 0.7986$, and $\tan 37° = 0.7536$.

(2) To find the angle whose cosine is 0.8988: go down the column of cosines until 0.8988 is found and then proceed to the left on this line to find 26°; thus $\cos 26° = 0.8988$.

(3) To find cos 16.5° (16°30′). This cosine is not listed, but since 16.5° lies .5 or one half the interval between 16° and 17°, the cosine of 16.5° may be assumed to lie .5 of the interval between cos 16° and cos 17°. To calculate cos 16.5° it is necessary to subtract the proportional part of the difference from the value of cos 16° because the cosines decrease as the

angle increases. This operation is called *interpolation*, and is performed as follows:

(a)	$\cos 16° = 0.9613$	(d)	$\cos 16° = 0.9613$
(b)	$\cos 17° = \underline{0.9563}$	(e)	$0.5 \times 0.0050 = \underline{0.0025}$
(c)	Difference $= 0.0050$	(f)	$\cos 16.5° = 0.9588$

(4) To find tan 35.66° (35°40′). Since the tangent of an angle increases as the angle increases, the proportional part of the difference is added to the value of tan 35° to give tan 35.66°.

(a)	$\tan 36° = 0.7265$	(d)	$\tan 35° = 0.7002$
(b)	$\tan 35° = \underline{0.7002}$	(e)	$0.66 \times 0.0263 = \underline{0.0175}$
(c)	Difference $= 0.0263$	(f)	$\tan 35.66 = 0.7177$

(5) Find angle A if $\sin A = 0.6626$. From the table it is seen that A lies between 41° and 42°.

(a)	$\sin 42° = 0.6691$	(d)	$\sin A = 0.6626$
(b)	$\sin 41° = \underline{0.6561}$	(e)	$\sin 41° = \underline{0.6561}$
(c)	Difference $= 0.0130$	(f)	Difference $= 0.0065$

(g) Since A is between 41° and 42°, $A = 41° + \dfrac{0.0065}{0.0130} \times 1°$.

(h) $A = 41° + 0.5° = 41.5°$ or $41° + 30' = 41°30'$.

Exercise 1–24. Find the following functions and angles from the table.

1. sin 45°
2. cos 45°
3. tan 45°
4. sin 90°
5. cos 15.5°
6. tan 75°
7. cos 88.6°
8. sin 79.25°
9. Angle A, if $\sin A = 0.9135$
10. Angle A, if $\tan A = 0.2035$
11. Angle A, if $\cos A = 0.9310$
12. Angle A, if $\sin A = 0.9995$

These functions of trigonometry are used to solve many problems in radio work. First, however, an illustration based on the action of mechanical forces will be discussed.

Consider the ordinary boys' slingshot (Fig. 1–6a), which consists of a forked stick with a rubber band fastened to the ends. When a small stone is placed in the loop and pulled back, the stone flies straight ahead when released. If the angles shown in Fig. 1–6b are assumed, the forces on the stone and in the rubber bands may be found, since these forces are proportional to the lengths of the sides of the triangle ABD or BCD. Half the force on the stone will be supplied by each half of the rubber band. For

TABLE 1-3

SINES, COSINES, AND TANGENTS

Degrees	Sine $\left(\frac{opp.}{hyp.}\right)$	Cosine $\left(\frac{adj.}{hyp.}\right)$	Tangent $\left(\frac{opp.}{adj.}\right)$	Degrees	Sine $\left(\frac{opp.}{hyp.}\right)$	Cosine $\left(\frac{adj.}{hyp.}\right)$	Tangent $\left(\frac{opp.}{adj.}\right)$
0	.0000	1.0000	.0000	45	.7071	.7071	1.0000
1	.0175	.9998	.0175	46	.7193	.6947	1.0355
2	.0349	.9994	.0349	47	.7314	.6820	1.0724
3	.0523	.9986	.0524	48	.7431	.6691	1.1106
4	.0698	.9976	.0699	49	.7547	.6561	1.1504
5	.0872	.9962	.0875	50	.7660	.6428	1.1918
6	.1045	.9945	.1051	51	.7771	.6293	1.2349
7	.1219	.9925	.1228	52	.7880	.6157	1.2799
8	.1392	.9903	.1405	53	.7986	.6018	1.3270
9	.1564	.9877	.1584	54	.8090	.5878	1.3764
10	.1736	.9848	.1763	55	.8192	.5736	1.4281
11	.1908	.9816	.1944	56	.8290	.5592	1.4826
12	.2079	.9781	.2126	57	.8387	.5446	1.5399
13	.2250	.9744	.2309	58	.8480	.5299	1.6003
14	.2419	.9703	.2493	59	.8572	.5150	1.6643
15	.2588	.9659	.2679	60	.8660	.5000	1.7321
16	.2756	.9613	.2867	61	.8746	.4848	1.8040
17	.2924	.9563	.3057	62	.8829	.4695	1.8807
18	.3090	.9511	.3249	63	.8910	.4540	1.9626
19	.3256	.9455	.3443	64	.8988	.4384	2.0503
20	.3420	.9397	.3640	65	.9063	.4226	2.1445
21	.3584	.9336	.3839	66	.9135	.4067	2.2460
22	.3746	.9272	.4040	67	.9205	.3907	2.3559
23	.3907	.9205	.4245	68	.9272	.3746	2.4751
24	.4067	.9135	.4452	69	.9336	.3584	2.6051
25	.4226	.9063	.4663	70	.9397	.3420	2.7475
26	.4384	.8988	.4877	71	.9455	.3256	2.9042
27	.4540	.8910	.5095	72	.9511	.3090	3.0777
28	.4695	.8829	.5317	73	.9563	.2924	3.2709
29	.4848	.8746	.5543	74	.9613	.2756	3.4874
30	.5000	.8660	.5774	75	.9659	.2588	3.7321
31	.5150	.8572	.6009	76	.9703	.2419	4.0108
32	.5299	.8480	.6249	77	.9744	.2250	4.3315
33	.5446	.8387	.6494	78	.9781	.2079	4.7046
34	.5592	.8290	.6745	79	.9816	.1908	5.1446
35	.5736	.8192	.7002	80	.9848	.1736	5.6713
36	.5878	.8090	.7265	81	.9877	.1564	6.3138
37	.6018	.7986	.7536	82	.9903	.1392	7.1154
38	.6157	.7880	.7813	83	.9925	.1219	8.1443
39	.6293	.7771	.8098	84	.9945	.1045	9.5144
40	.6428	.7660	.8391	85	.9962	.0872	11.4301
41	.6561	.7547	.8693	86	.9976	.0698	14.3007
42	.6691	.7431	.9004	87	.9986	.0523	19.0811
43	.6820	.7314	.9325	88	.9994	.0349	28.6363
44	.6947	.7193	.9657	89	.9998	.0175	57.2900
45	.7071	.7071	1.0000	90	1.0000	.0000	.....

example, if the rubber band is stretched so that each half or each side exerts a force of 1 unit (which may be called F_R) in line with itself,

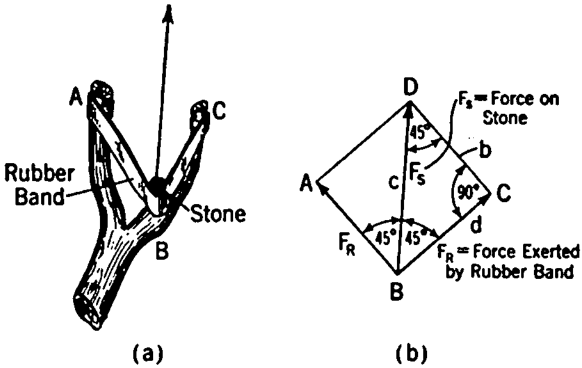


FIG. 1-6. Slingshot and Force Diagram.

what is the force exerted on the stone (represented as F_S)? From the triangle, $\cos 45° = \frac{1}{2}\frac{c}{d}$ or $\frac{1}{2}\ c = d \cos 45°$. Since each side of the band supplies half the force on the stone, then

$$\tfrac{1}{2}\ F_S = F_R \times \cos 45° = 1 \times 0.7071 = 0.7071.$$

The total force on the stone will be twice this, or 1.4142 units.

Vector Addition. Some relationships to be explained later are indicated in the triangle in Fig. 1-7.

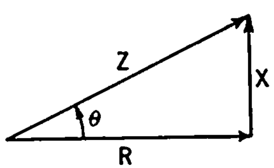


FIG. 1-7. Impedance Triangle with Parts Marked.

$$\frac{\mathbf{R}}{\mathbf{Z}} = \cos\theta \qquad \mathbf{R} = \mathbf{Z}\cos\theta \qquad \mathbf{Z} = \frac{\mathbf{R}}{\cos\theta}$$

$$\frac{\mathbf{X}}{\mathbf{Z}} = \sin\theta \qquad \mathbf{X} = \mathbf{Z}\sin\theta \qquad \mathbf{Z} = \frac{\mathbf{X}}{\sin\theta}$$

$$\mathbf{Z} = \sqrt{\mathbf{R}^2 + \mathbf{X}^2} \qquad \tan\theta = \frac{\mathbf{X}}{\mathbf{R}}$$

The symbol θ is the Greek letter theta.

Exercise 1-25.

1. Given **R** = 8, **X** = 6, find θ.
2. Given **Z** = 10, θ = 45°, find **R** and **X**.

Suppose it is stated that **R** = 8 and **X** = 6, which are to be combined to give **Z** = 10. Certainly these cannot be added by arithmetic because this would make the sum 8 + 6 = 14. Quantities which have a definite direction as well as size, such as **R**, **X**, **Z**, are called *vectors* and are added by *vector addition.* To add **R** to **X** in this fashion, place them as shown in the triangle (the arrowhead indicating the direction of each) and draw a line from the beginning of **R** to the end of **X**. This will be **Z**, the *vector* *sum* of **R** and **X**, and its length is $\mathbf{Z} = \sqrt{\mathbf{R}^2 + \mathbf{X}^2} = \sqrt{8^2 + 6^2} = 10$, as required.

It is customary to set vectors in boldface type. In writing, a short bar, as $\bar{x}$, over the vector distinguishes it from nonvector expressions, as x.

Another system for finding the values of the trigonometric functions is to draw a curve of sines. Suppose a circle is drawn whose radius is 1 unit long (Fig. 1-8), with center at A. Then draw a diameter AD and continue this line to E; draw another line at a right angle to the diameter and to the end of the radius (CB). The triangle ACB is a right triangle and $\sin\theta = \dfrac{CB}{AB}$.

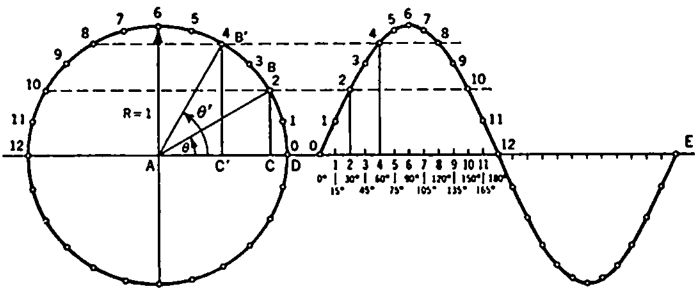


FIG. 1-8. Construction for Sine Curve.

But $AB = 1$ (unity in length) and so $\sin\theta = CB$. If the radius were drawn in another position (AB'), the line from B' to the diameter at C' will again be $\sin\theta' = C'B'$, and so on, no matter what the angle θ happens to be. If the entire circle be divided into parts (say 24) and the radius considered as if moving, in a direction opposite to that of the clock hands, it will occupy each of these positions in succession. Divisions are marked on the line DE with the numbers of degrees corresponding to the positions of the radius. By drawing lines horizontally from the end of the radius in the various positions to the vertical lines as shown, a curve of sines, or a *sine curve*, will be formed when the points are joined by a smooth curve.

Other information may be obtained from such a diagram. Following the lines used in making the diagram, it may be seen that $\sin 120° = \sin 60°$, $\sin 150° = \sin 30°$, and so on. This suggests a general statement that $\sin(180° - A) = \sin A$. Many similar rules have been worked out.

Exercise 1-26. Compare the height of several points on the curve with values in Table 1-3 for the same angles.

In Fig. 1-8, $\cos\theta = AC$. If the radius were again rotated, and the lengths AC, AC', and so forth, plotted vertically above the degree marks on DE, a *cosine curve* would be obtained. The two curves may be compared; the shape is exactly the same, but when the sine curve is zero the cosine curve has the value 1, and so on.

Exercise 1–27. Compare the values of several points on the sine curve with the values of cosines in Table 1–3. Could the relation sin $(90° - A) = \cos A$ be used to plot a curve of cosines?

For some purposes it is convenient to use a special unit to measure angles instead of the usual degrees. If the radius of a circle is bent around the circumference of the circle like a flexible rule, it will fit 2π times without overlapping ($\pi = 3.1416$, approximately). That is, 2π radii will be needed to make a curved line as long as the circumference. The relation is true no matter what the size of the circle. The radius, wrapped around the circle, will be the curved edge of a piece of pie, with an angle at the center (Fig. 1–9). This angle is called a *radian.* One radian is about 57.3°. There will be 2π radians in the whole circumference or 360°. To *convert degrees* to *radians,* divide the number of degrees by 57.3; to *convert radians* to *degrees,* multiply the number of radians by 57.3.

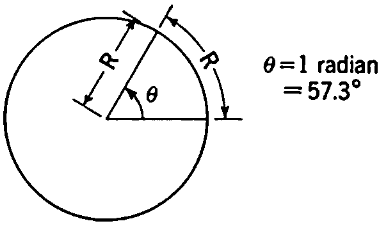


Fig. 1–9. Measure of Angle in Radians.

Exercise 1–28.

1. Convert the following angles to radians: 45°, 60°, 90°, 180°, 120°, 135°, 22.5°, 200°, 3,000°.

2. Convert the following angles to degrees: $\frac{\pi}{3}, \frac{3\pi}{4}, \frac{\pi}{72}, \frac{7\pi}{6}, \frac{20\pi}{3}, 0.98\pi$.

Graphs and Curves. A fairly reliable aid in predicting the future is the experience of the past. To this end information is collected about the number of automobiles built, number of babies born, temperatures in various cities, prices on the market, and many other things. This information may be listed in the form of tables or shown on graphs. A graph of the average daily temperature in some locality is shown in Fig. 1–10, which is plotted from Table 1–4.

TABLE 1–4

Date	*Temperature, degrees*
May 1	51
May 2	58
May 3	68
May 4	55
May 5	69
May 6	74

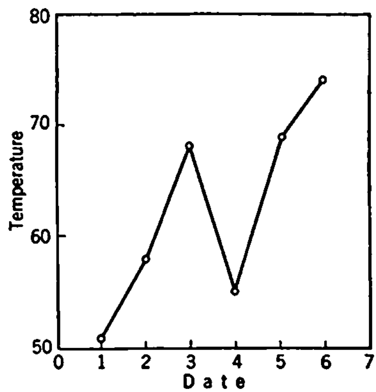


Fig. 1–10. Chart of Daily Temperature for May.

The gradual rise in temperature at this time of year and the drop of temperature on May 4 are easily seen on the graph.

Sometimes important decisions are made on the basis of information obtained from graphs. For example, if a power company plotted the power output of its plants over a long period of time the curve might be extended toward future dates by observing its shape, amount of bending, steepness, and so forth. From this extended curve the company might tell when the power required would be greater than their plants could furnish, and this would be the date when new generators should be put into operation (Fig. 1–11).

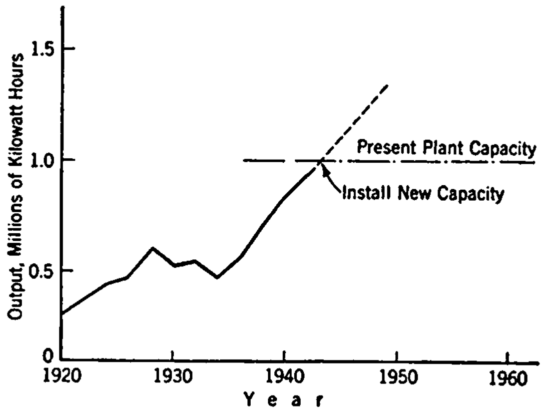


Fig. 1–11. Output of Plant, by Years.

Another use for graphs has already been shown—a curve of sines may be used as a table of sines of angles. The curves for cosines, tangents, and other trigonometric functions may be constructed and used in the same way.

Logarithms. It has already been shown (pp. 7–9) that very large or very small numbers are best written in scientific notation. A similar plan may be used to simplify multiplying, dividing, and other operations.

10^3	=	1000
10^2	=	100
10^1	=	10
10^0	=	1
10^{-1}	=	0.1
10^{-2}	=	0.01
10^{-3}	=	0.001
etc.		

Consider again the table showing powers of 10 and their equivalents in place notation. It is seen that $10^0 = 1$ and $10^1 = 10$. A number between 1 and 10 must be represented by 10 raised to some power between 0 and 1. Tables have been prepared to show what this power is, and Table 1–5 is this kind of a table for the base 10.

TABLE 1-5

LOGARITHMS

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

TABLE 1-5 (*Continued*)
LOGARITHMS

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

Before using the table consider that 2, for example, lies between 1 and 10; it may be represented by $10^{0.3010}$, in which the exponent lies between 0 and 1. Also it is possible to write 3 as $10^{0.4771}$. Since the 10 (the base) occurs in all cases, it is convenient to change the name of the exponent to *logarithm* and not to write the base at all. The logarithms of the numbers from 1 to 10 are the following (*logarithm* is abbreviated "log"):

$\log 1 = 0.0000$	$\log 4 = 0.6021$	$\log 7 = 0.8451$
$\log 2 = 0.3010$	$\log 5 = 0.6990$	$\log 8 = 0.9031$
$\log 3 = 0.4771$	$\log 6 = 0.7782$	$\log 9 = 0.9542$

Now from the table of powers of 10 the exponent (or logarithm) of numbers between 10 and 100 must lie between 1 and 2. But 20 is the same fraction of the distance from 10 to 100 that 2 is from 1 to 10. Therefore the logarithm of 20 is 1.3010 and similarly

$$\log 200 = 2.3010$$
$$\log 2{,}000 = 3.3010$$

The part of the logarithm to the right of the decimal point remains the same as long as the digits in the original number remain the same; the part to the left of the decimal point changes by 1 whenever the number is multiplied or divided by 10.

The decimal fraction or right-hand part of the logarithm is given in the table to four decimal places and is called the *mantissa*. The integral part to the left of the decimal point is called the *characteristic* and is found in the following way:

(a) Move the decimal point in the number until the number remaining is between 1 and 10.

(b) Count the number of places the decimal point has been moved and call this number the characteristic.

(c) Make the characteristic positive if the decimal point was moved to the left and negative if it was moved to the right.

(d) A negative characteristic is written as a positive one, with 10 subtracted from the entire log.

Example: Find log 4,570 in the table.

(a) Find 45 in the left-hand column of Table 1–5 and move across this line to the column headed by 7 to find 6599. Write this as 0.6599.

(b) Move the decimal point to the left three places to give 4.570, a number between 1 and 10. The characteristic is 3 and positive.

(c) The complete logarithm is $3 + 0.6599 = 3.6599$.

Example: Find log 0.00121.

(a) Find 12 in the left-hand column, go across this line to the column headed 1 and find 0.0828.

(b) Move the decimal point to the right three places, to give 1.21. The characteristic is 3 and negative, but is written 7.   − 10.

(c) $\log 0.00121 = -3 + .0828 = 7.0828-10.$

Exercise 1–29.

1. Check the following: (a) $\log 207 = 2.3160$; (b) $\log 8{,}500 = 3.9294$; (c) $\log 0.667 = 9.8241-10$; (d) $\log 91.9 = 1.9633$.

2. Find logs of these numbers: 206, 75,400, 8,300, 92.8, 0.00026.

The number corresponding to a given logarithm is called the *antilogarithm* and is found as follows:

(a) Find the mantissa of the logarithm in the body of the table. Move across to the left-hand column for the first two figures of the antilog and note the column heading which is the third figure; consider the result as a number between 1 and 10.

(b) Move the decimal point to the right as many places as the characteristic when the latter is positive and to the right when it is negative.

Example: Find antilog 2.5877.

(a) Find 0.5877 in line 38 and column 7; consider this as 3.87.

(b) Move the decimal point 2 places to the right, which gives 387, the required antilog.

Example: Find antilog 7.3243−10.

(a) Find 0.3243 in line 21, column 1; consider this as 2.11.

(b) The characteristic is $7 - 10 = -3$. Therefore, move the decimal point 3 places to the left to get 0.00211 as the required antilog.

If the exact value cannot be found in the table, take the value nearest to it; or interpolate as follows:

Example: Find antilog 0.4208.

$$\begin{aligned} \log 2.64 &= 0.4216 \\ \log 2.63 &= \underline{0.4200} \\ \text{Difference} &= 0.0016 \end{aligned} \qquad \begin{aligned} \log 2.64 &= 0.4216 \\ \text{given log} &= \underline{0.4208} \\ \text{Difference} &= 0.0008 \end{aligned}$$

Since $0.0008 \div 0.0016 = 0.5$, the required antilog must lie 0.5 of the way from 2.63 to 2.64, or 2.635. The characteristic is zero so the decimal point does not have to be moved.

Exercise 1–30.

1. Check the following: (a) antilog $1.8169 = 65.6$; (b) antilog $7.9325-10 = 0.00856$; (c) antilog $4.4814 = 30{,}300$; (d) antilog $3.5711 = 3725$; (e) antilog $8.8766-10 = 0.07527$.

2. Find the antilogs of the following: (a) 2.9274; (b) 9.9533−10; (c) 3.5441; (d) 0.6196; (e) 6.6169−10.

Logarithms may be used to make multiplication and division easier. The rules are the same as those stated for exponents. That is, *to multiply, add the logarithms of the numbers* and look up the antilogarithm of the result. The work of multiplying 479 by 89 may be arranged as follows:

$$\begin{aligned} \log 479 &= 2.6803 \\ \log \ \ 89 &= \underline{1.9494} \\ \text{log of product} &= 4.6297 \\ \text{antilog } 4.6297 &= 42{,}630. \end{aligned}$$

To divide, subtract logs; to divide 479 by 890, arrange thus:

$$\begin{aligned} \log 479 &= 12.6803 - 10 \\ \log 890 &= \underline{\ \ 2.9494 \qquad\ } \\ \text{log of quotient} &= \ \ 9.7309 - 10 \\ \text{antilog } 9.7309 - 10 &= \ \ 0.538. \end{aligned}$$

Since the difference of the logs is a negative number here, 10 is added at the left and subtracted to the right of log 479 (this does not alter its value) and the log of the quotient shown is obtained.

Exercise 1–31. Perform the operations indicated, using logs.

1. 3×7
2. 746×0.567
3. $5.55 \times 637.$
4. 0.0495×0.0267
5. $87.5 \div 37.7$
6. $0.685 \div 9.75$
7. $3.14 \div 2.72$
8. $0.0385 \div 0.00146$

Logs are very useful in finding higher powers and roots of numbers. Suppose that 24^5 is needed. This could be found exactly by multiplying $24 \times 24 \times 24 \times 24 \times 24$, but logs make the operation much easier. Thus:

$$\begin{aligned} \log 24 &= 1.3802 \\ \text{multiply by} &\ \ \ \underline{\qquad\ \ 5} \\ \log (5 \times 24) &= 6.9010 \\ \text{antilog } 6.9010 &= 7{,}962{,}000. \end{aligned}$$

Others powers are also easy: for example, $2^{0.5} = 2^{\frac{1}{2}} = \sqrt{2}$.

$$\begin{aligned} \log 2 &= 0.3010 \\ \text{multiply by} &\ \ \ \underline{\qquad 0.5} \\ \log 2^{0.5} &= 0.1505 \\ \text{antilog } 0.1505 &= 1.414 \end{aligned}$$

This is actually the process of extracting a square root, because

$$2^{0.5} \times 2^{0.5} = 2^{0.5+0.5} = 2^1 = 2.$$

Logarithms may be plotted graphically, and the resulting curve may be used as a table or otherwise. Instead of using equal divisions on the paper in plotting this graph, it is handier to use paper which is divided according to the logarithms of the numbers, as shown in Fig. 1–12. Entering the chart at 2, on the bottom, proceed upward to the line and then to the left to find 0.3, which is log 2. To find antilog 0.8, enter the chart at 0.8 on the left, proceed horizontally to the line, and then down to 6.4 as required.

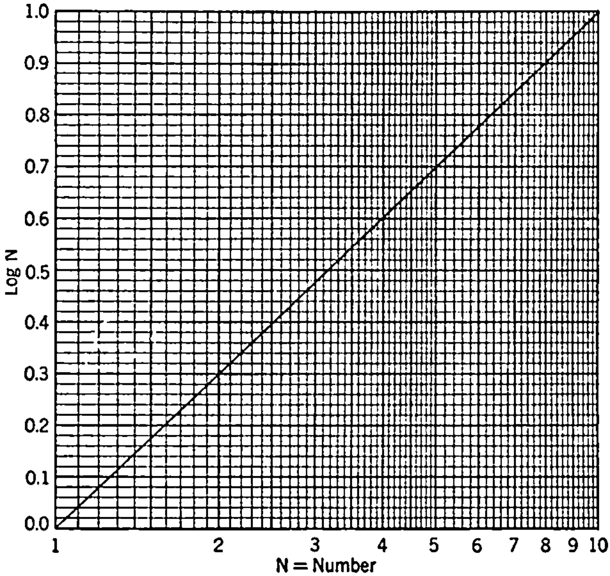


FIG. 1–12. Curve Showing Logarithms of Numbers.

Exercise 1–32. Compare the values of logs found on the chart with values from the table.

Besides the use of logarithms for calculating, there are other reasons for understanding their use. The human ear acts in a manner that is of logarithmic nature. What we hear depends upon where we are, among other things. As everyone knows, it is harder to make another person hear what is being said in a noisy location (such as city traffic) than in a quiet place. The reason is that the increase of sound necessary to give the impression of a change in sound intensity is governed not only by the actual change of sound intensity but also by the amount of other sound present. By selecting a unit of sound intensity which is of the same sort as the action of the ear, the communication engineers simplify discussion of this action and make it easier to set control dials to correct values.

The unit of sound level used is called the *decibel* (abbreviated db) and is defined by the equation

$$db = 10 \log \frac{P_1}{P_2},$$

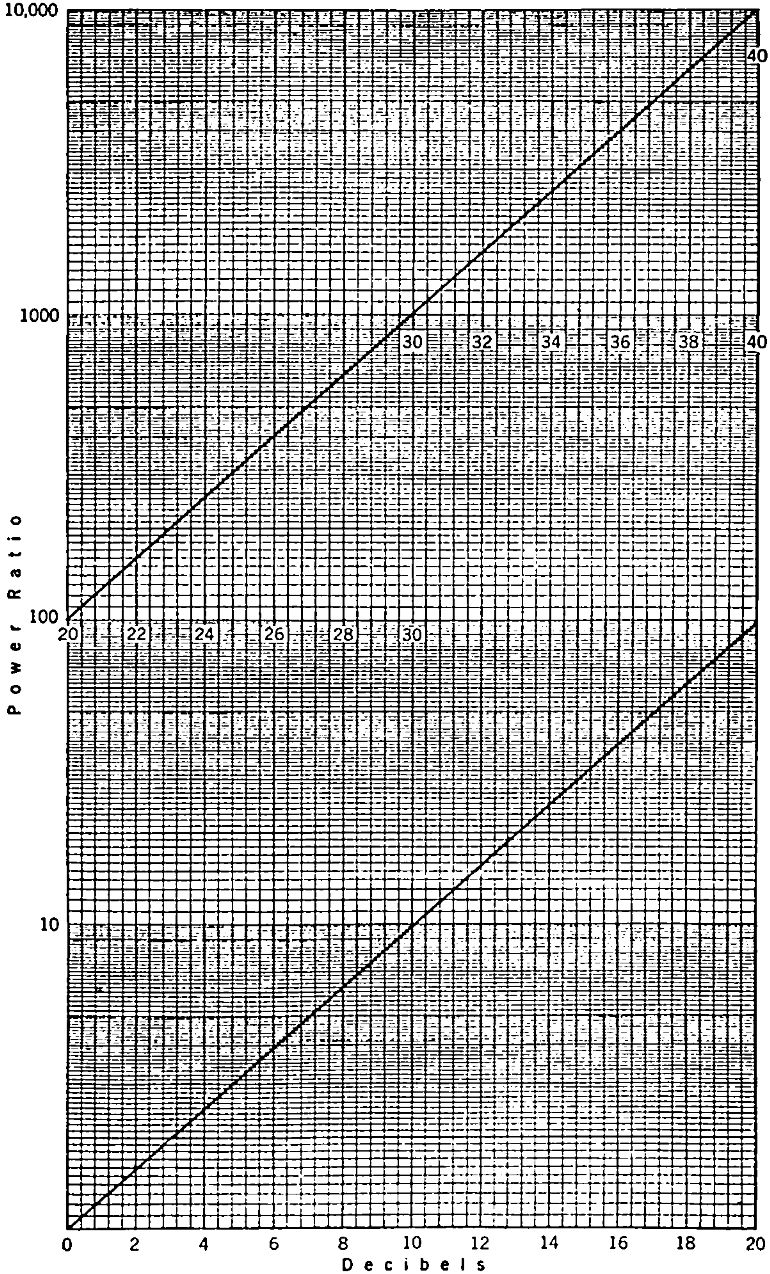


FIG. 1-13. Relation between Power Ratio and Decibels.

where db is the number of decibels, P_1 is the larger amount of power, and P_2 the smaller amount of power (or sound) being compared. If one sound is twice as loud as another (twice as much power), then the ratio $P_1/P_2 = 2$; $\log 2 = 0.3010$, and therefore the number of decibels is $10 \times 0.3010 = 3$. This may be seen from the chart showing the relation between power ratio and decibels, Fig. 1–13. If the output power in a given system is greater than the input power, the system is said to have a gain; if the output is less than the input, the system is said to have a loss, or the gain is expressed in negative decibels. By assigning a figure to the zero level, the power output of amplifiers and other equipment may be rated in decibels, as is often done.

Exercise 1–33. **1.** From the chart find the number of decibels corresponding to the following power ratios: 2, 4, 10, 100, 1,000, 3, 50, 600.

2. Find the power ratios corresponding to the following decibel gains (or losses): 2, 6, 35, 20, 10.

As has been shown, charts may be made with divisions which are logarithmic. For some purposes it is more convenient to use simply a scale which is divided into parts proportional to the logarithms of numbers, such as Fig. 1–14. It is seen that the point marked 2 is about 0.3 of the length of the scale from the left-hand end; 4 is located 0.6 of the length from the left, and so on. Other divisions are marked according to the following system. Considering the 1 at the left-hand end of the scale (called the *index*) to represent 100, the main divisions will be 1, 2, 3, · · ·, to represent 100, 200, 300, · · ·. Between 1 and 2 are other divisions marked from 1 to 9, which represent 110, 120, 130, · · ·. These are further divided with a third set of marks which do not have numbers but represent 101, 102, 103 · · ·, to 199. Between the 2 and 4 the subdivisions with the longest lines represent 210, 220, 230, · · ·, which are again divided to represent 202, 204, 206, · · · etc. The smallest divisions here represent numbers twice as large as those in the portion of the scale marked 1 to 2. From 4 to the right-hand end of the scale the major divisions are also by 10's (410, 420, 430 · · ·) but the smallest division represents only 5; that is, 405, 415, 425, · · · with 10's divisions between. In Fig. 1–15 is a scale with the following points marked: A 365, B 327, C 263, D 1,745, E 1,347, F 305, G 207, H 1,078, I 435, J 427.

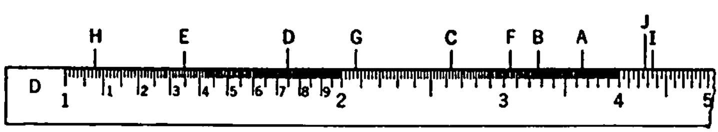


FIG. 1–15. Logarithmic Scale with Certain Points Marked.

Exercise 1–34. Verify the above readings on the scale and locate the following points: 445, 463, 772, 255, 279, 1,850, 1,763, 1,941, 1,005.

The numbers at either end of the scale (called the *left-hand* and *right-hand index*, respectively) may be multiplied or divided by any power of 10; that is, the left-hand index may represent 1.0, 100, 10^6, 0.001, and all other points on the scale will have the same multiplier. This property and other properties of logarithms permit two such scales as have been described to be used for multiplication, division, and other operations with numbers except addition and subtraction.

If the scale marked C (Fig. 1–14) be cut out on line xy, the C scale may be matched with the D scale or moved to any position along the latter. For example, if the 1 on the C scale (called C for brevity) be placed opposite 2 on D, 2 on C will be opposite 4 on D, 3 opposite 6, and so on. But 4 is 2 × 2, and actually log 2 on C has been added mechanically to log 2 on D and the result, quite properly, is log 4. This process (adding logarithms mechanically) is the basis of the *slide rule*, which makes tedious calculations much easier.

Division may be performed with equal ease. For example, to divide 9 by 3, set 3 on C over 9 on D and read 3, the answer, under the left-hand index (1) on C.

This same setting also gives the result of dividing 6 by 2, 7.5 by 2.5, 36 by 12, and many other combinations. This property may be used to work problems in proportion, such as

$$3:9 = x:42,$$

finding the answer 14 above 42 on the D scale.

The location of the decimal point may be found by rules like those used for logarithms, but is often located by inspection. Thus, in multiplying 195 by 24, an approximate calculation (made mentally) with 200 × 20 shows that the answer should be somewhere near 4,000, and actually is 4,680.

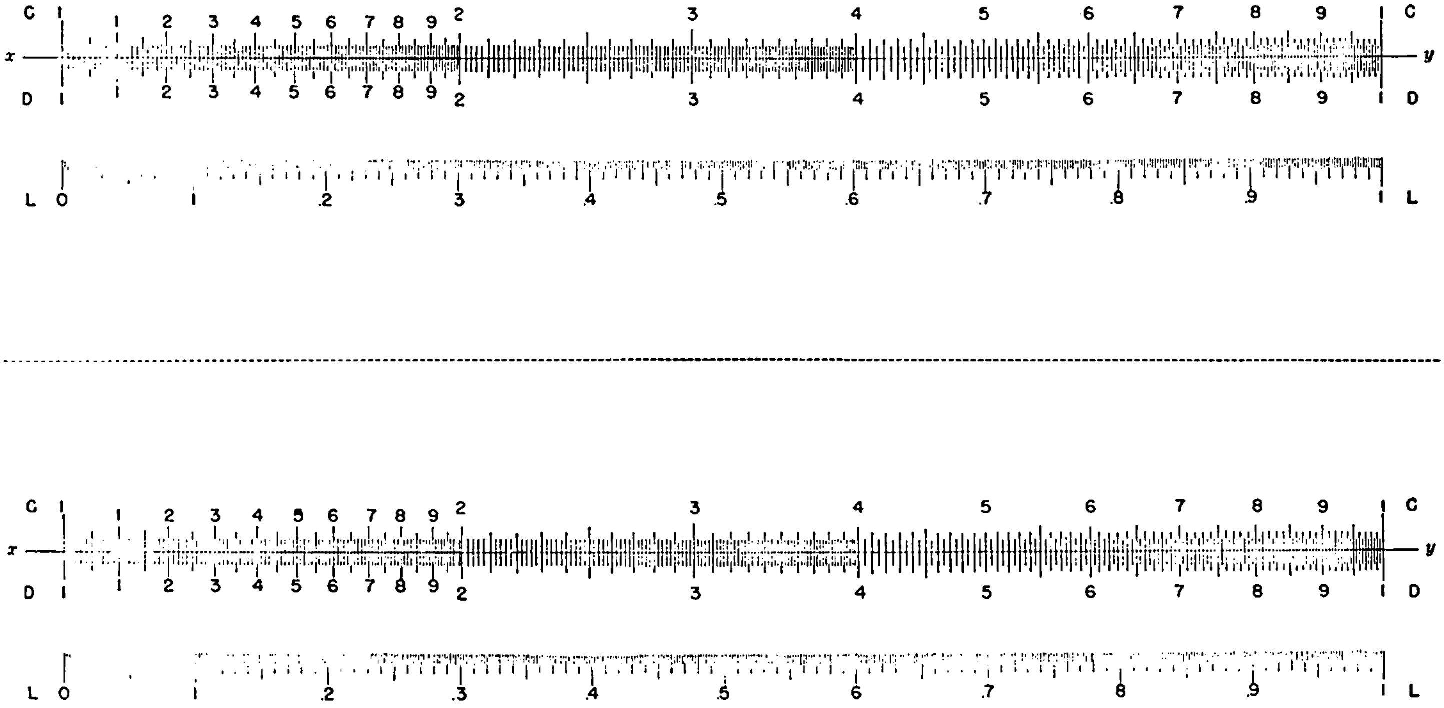
Common fractions are easily converted to decimal fractions on these scales, by dividing the numerator by the denominator. That is, to convert $\frac{1}{16}$ to a decimal fraction, set 16 on C over 1 on D and find 0.0625 on D under the right-hand index of C.

The relation between the divisions on the D scale and the logarithms of numbers may be seen by finding the logarithms on the scale of equal parts, marked L. For example, opposite 2 on D find 0.301 on L, opposite 3 on D find 0.477 on L, and so forth.

Exercise 1–35. Perform the following operations with the sliding scales.

1. Multiply: 3 × 5, 3.05 × 5.17, 5.56 × 634, 743 × 0.0567, 0.0495 × 0.0267.

FIG. 1-14. Typical Slide Rule Scales. (The second one may be cut out and pasted on cardboard or wood to make a model slide rule. Also cut along line x—y.)



2. Divide: 87.5 ÷ 37.7, 0.685 ÷ 8.93, 1,029 ÷ 9.70, 0.00377 ÷ 5.29.

3. Solve the proportion: $2 : 3 = x : 7.83$.

4. Convert to decimal fractions: $\frac{4}{5}, \frac{11}{16}, \frac{3}{32}, \frac{17}{348}, \frac{193}{1{,}095}$.

Compare values of logarithms found on the L scale with those in Table 1–5.

Answers to Exercises

1–1. (1) 28; (2) 32; (3) 26; (4) 32; (5) 90; (6) 0; (7) 5; (8) 7 (9) this operation not possible; (10) 2; (11) 5; (12) 5; (13) 0; (14) 49; (15) 32.

1–2. (1) $1\frac{3}{56}$; (2) $1\frac{1}{2}$; (3) $\frac{47}{60}$; (4) $\frac{3}{56}$; (5) $\frac{1}{8}$; (6) $\frac{167}{258}$; (7) $\frac{4}{9}$; (8) $\frac{323}{3136}$; (9) $\frac{2}{35}$; (10) $1\frac{1}{9}$; (11) $\frac{15}{32}$; (12) $\frac{590}{693}$.

1–3. (1) 0.75, 0.625, 0.5625, 0.265625, 0.9, 0.666 · · ·, 0.764 · · ·; (2) see machinist's or other handbook; (3) 9.806; (4) 0.746; (5) 23.731526; (6) 0.826.

1–4. (1) 2; (2) 3; (3) 5; (4) 7; (5) 3; (6) 1; (7) 4; (8) 7; (9) 7; (10) 6.

1–5. (1) 6.05×10^{23}; (2) 6.94×10^{11}; (3) 5.87×10^{12}; (4) 9.3×10^{7}; (5) 2×10^{-7}; (6) 3×10^{-23}.

1–6. (1) 860,000; (2) 6,600,000,000,000,000,000,000; (3) 0.00003; (4) 0.064; (5) 1,200,000,000; (6) 0.00000628.

1–7. (1) 6.7×10^{7}; (2) 5×10^{2}; (3) 0.00174.

1–8. (1) $24m$; (2) $10v$; (3) $100q$; (4) b; (5) $\frac{1}{2}c$; (6) $2.3x$; (7) $17.27k$; (8) $22h$.

1–9. (1) $10a + 13b$; (2) $7.8c + 19.1d + 6.2e$; (3) $36f$; (4) $16x$; (5) $92y$; (6) $15z$; (7) $5a$; (8) $.92r$; (9) 3; (10) $\frac{3}{2}a + 2x + y$.

1–10. (1) $8a^2 - 5ab$; (2) $-x^2 - 10x + 12$; (3) $-2r^2 + 2rs - 3s^2$; (4) $3a^2 + 4a - 10$.

1–11. (1) $6x^7$; (2) $-h^7$; (3) $-125d^3$; (4) $30E^6$; (5) $-8t^3$; (6) $-15x^2y^2$; (7) $9a^6t^2$; (8) $-9a^6t^2$; (9) $-6c^9$; (10) $-24n^2d^3$; (11) $-x^6y^{18}$; (12) $4d^2c^{10}$.

1–12. (1) $8b + 5$; (2) -32; (3) $-15d + 24$; (4) $8a^2 + 4ab + 3ac$; (5) $-r + 2s$; (6) $-11d + 7$; (7) $8n^2 - 13n - 2$; (8) $-23n - 52$; (9) $16a^2$; (10) $9s^2 + s$; (11) $5x - 45$; (12) $8y^2 + 6y$; (13) $14r - 17$; (14) $-3x^2 + 10x - 20$; (15) $16x - 90y - y^2 - 54$.

1–13. (1) $7x(y - 2z)$; (2) $5ab(a - 5b)$; (3) $2\pi r(r + h)$; (4) $4x(2y^2 - 4y + 3)$; (5) $ab(ab - 3b + 4a^2 - 12)$; (6) $6m(4mn - n^2 + 6mn^2 - 7m^2)$.

1–14. (1) $n^2 + 5n + 6$; (2) $a^2 - 3a - 10$; (3) $p^2 - 9p + 20$; (4) $a^2 + 3ab + 2b^2$; (5) $x^4 - 16x^2 + 48$; (6) $t^2h^2 - 5th - 84$; (7) $9x^2 - 6x - 3$; (8) $x^4 - 6x^2y^2 - 55y^4$.

1–15. (1) $(n + 3)(n + 2)$; (2) $(x - 5)(x - 1)$; (3) $(a - 4)(a - 5)$; (4) $(2n - 1)(3n + 2)$; (5) $(x + 2y)(x + y)$; (6) $(3a + 2)(5a + 4)$; (7) $(2 + 5x)(1 + x)$; (8) $(a^2 - 11b^2)(a^2 + 5b^2)$.

1–16. (1) $x^2 + 6x + 9$; (2) $x^2 - 10x + 25$; (3) $x^2 + 12xy + 36y^2$; (4) $a^2 - 2ac + c^2$; (5) $4x^2 - 12x + 9$; (6) $9a^2 + 30a + 25$; (7) $a^2 - 6ab + 9b^2$; (8) $9k^4 - 6k^2 + 1$; (9) $25r^4 + 10r^2s^2 + s^4$; (10) $a^6 + 24a^3 + 144$; (11) $h^2 - \frac{1}{2}h + \frac{1}{16}$; (12) $\frac{4}{9}a^2 + \frac{2}{3}ab + \frac{1}{4}b^2$.

1–17. (1) $(x - 6)^2$; (2) $(4x - 5y)^2$; (3) $(5a^2 + 12b^2)^2$.

1–18. (1) $x^2 - y^2$; (2) $100 - a^2$; (3) $16a^2 - 9b^2$; (4) $(6r + s)(6r - s)$; (5) $(-9v + 5u)(9v + 5u)$; (6) $3c(2c - 1)(2c + 1)$; (7) $(60 + 3)(60 - 3) = 3{,}600 - 9 = 3{,}591$; (8) $(60 + 15)(60 - 15) = 3{,}375$.

1–19. (1) 100; (2) 0.41; (3) 50; (4) 322.

1–20. (1) 377; (2) 51.9; (3) 159; (4) 0.603; (5) 0.

1–21. (1) 118; (2) $(RR_2)/(R_2 - R)$.

1–22. (1) $x = -3, +1$; (2) $x = +3, +5$; (3) $x = +11, -2$; (4) $x = 4, \frac{1}{2}$; (5) $x = \frac{1}{2}, \frac{1}{3}$; (6) $x = 2, \frac{1}{5}$; (7) $y = -\frac{3}{4}, -\frac{4}{3}$; (8) $x = 19, 1$; (9) $x = -8, +6$; (10) $x = -3a, +a$.

1–23. (1) 5 inches; (2) 0.6, 0.8, 0.75; (3) $\sin 30° = \cos 60°$, $\sin 60° = \cos 30°$; (4) yes.

1–24. (1) 0.7071; (2) 0.7071; (3) 1.0000; (4) 1.0000; (5) 0.9636; (6) 3.7321; (7) 0.0245; (8) 0.9824; (9) 66°; (10) 11.5°; (11) 21.5°; (12) 88.25°.

1–25. (1) 37°; (2) $\mathbf{R} = 7.07$, $\mathbf{X} = 7.07$.

1–27. Yes.

1–28. (1) 0.785, 1.047, 1.571, 3.142, 2.094, 2.356, 0.398, 3.491, 52.36; (2) 60°, 135°, 2.5°, 210°, 1,200°, 176.4°.

1–29. (2) 2.3139, 4.8774, 3.9206, 1.9675, 6.4150 − 10.

1–30. (2) (a) 846; (b) 0.898; (c) 3,500; (d) 4.16; (e) 4.14×10^{-4}.

1–31. (1) 21; (2) 422; (3) 3,530; (4) 1.32×10^{-3}; (5) 2.32; (6) .0702; (7) 1.155; (8) 26.4.

1–33. (1) 3, 6, 10, 20, 30, 4.7, 16.9, 27.8; (2) 1.6, 4, 3,200, 100, 10.

1–35. (1) 15, 15.8, 3,520, 42.1, 1.32×10^{-3}; (2) 2.32, 7.68×10^{-2}, 106, 7.13×10^{-4}; (3) 5.22; (4) 0.8, 0.688, 0.094, 4.88×10^{-2}, 0.176.

CHAPTER 2

D.C. Circuits

Introduction. *Radio* is the name given to the science of communication and control by means of electromagnetic vibrations in space. It is the problem of the designers and operators of radio equipment to set up vibrations at the proper frequencies, and of sufficient power so that they can be detected by electric circuits tuned to respond to these particular vibrations. The proper care and manipulation of the complicated radio apparatus of the present time requires an extensive knowledge of electric circuits, vacuum tubes, and other equipment. The basis of all work in radio is an understanding of the character of electric phenomena, particularly the manner in which electric circuits respond to electric impulses.

One hundred years ago electricity was thought to be some peculiar kind of fluid which flowed in wires very much as water or oil flows in pipes. This idea permitted scientists to explain many of their experiments with electricity. As more and more new experiments were tried, the fluid theory did not continue to explain all of the results that were obtained and the new theories that were used to explain the results became more and more complicated. At the present time so many different experiments have been tried, and the theories used to explain them have become so complicated, that much study involving higher mathematics is required to understand these advanced theories. Fortunately it is necessary for the electrical technician to understand only the more elementary of them in order to work effectively.

The old belief that electric current is a fluid flow caused by an electric pressure and opposed by a resistance in the wire is still very effective in solving many of the problems of the electrician and the radio operator. It is now known, however, that electric current is not a true fluid but that it consists of the drift of millions of negatively charged particles along a wire. These negatively charged particles are called *electrons* and are so extremely small that they flow through the spaces between the atoms of the conductor.

In most metals a few electrons in the outer portion of the atom are very loosely bound to the nucleus of the atom. As a result large numbers of these electrons are free to drift about in the interatomic space. Metals of this type are called *conductors.* If the electrons are attracted by connecting the conductor to a battery they will accelerate owing to the force of attraction but the speed attained by any individual electron will be relatively small because it will bump into one of the many atoms before it has gone very far and bounce off in another direction. The energy which the electron had absorbed by reason of its acceleration is given up to the atom and appears in the form of heat. Some substances have a molecular structure in which the electrons are so closely bound to the molecule that a very high electric pressure is required to tear them away. Substances of this type are called *insulators* and are used in electric circuits to keep the electron flow restricted to paths desired by the designer.

Although it is interesting to know that electrons float about inside metal conductors, a complete understanding is not essential to a workable knowledge of electric circuits. Occasional reference may be made to the above statements to give a clearer explanation of physical occurrences. The simple fluid theory, however, will be the basis for most of the development.

Electrical Quantities. In order to discuss electrical phenomena intelligently, it is necessary to define some electrical quantities and specify the units in which these quantities are measured. Four of these quantities and the units in which they are measured are given below. Other definitions will be added as the need arises.

THE COULOMB. The unit of electrical charge or the quantity of excess electrons is called the coulomb. It is the charge which would be obtained by collecting approximately 6,300,000,000,000,000,000 (6.3×10^{18}) free electrons on a single charged body. This is a large unit and is seldom used in elementary radio calculation. It is important, however, as a basis for other units.

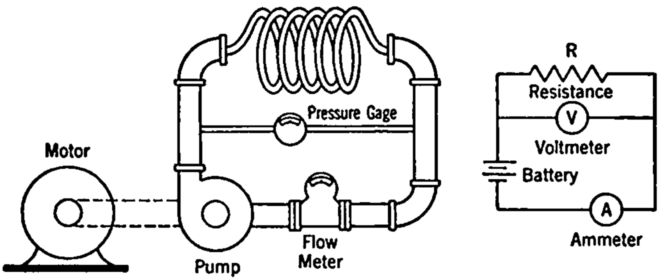
THE AMPERE. The unit of electric flow is called the ampere. If one coulomb of charge passes a given point on a wire in one second, then one ampere is said to flow. In other words, the ampere is a special name given to a *coulomb per second.* It is specified by international agreement as the constant current which will deposit silver at the rate of 0.0001118 g per second, as this definition gives a means by which a standard may be obtained anywhere, while the counting of the electrons would be difficult.

THE OHM. The unit of resistance to electric flow is called the ohm and it is specified by international agreement as the resistance which at a temperature of 0°C is offered to the flow of current by a column of mercury of uniform cross section, of a length of 106.3 cm, and having a mass of 14.45 g. The magnitude of the cross section so specified is essentially a square millimeter.

THE VOLT. The unit of electric pressure is the volt and is the pressure which will force one ampere to flow through a resistance of one ohm.

It is often difficult to visualize the magnitude of these units from the formal definitions so the following information may be of aid to the beginning student. Some idea of the size of the coulomb may be gained from the fact that if a charge of 6.8 millionths of a coulomb were placed one foot from a similar charge there would be a repelling force of one pound acting between them. The ampere is most easily visualized by referring to a 100-watt electric light which takes about one ampere of current. The favorite comparison for the volt is that an ordinary dry cell has 1.5 v. The normal pressure for domestic electric service is 120 v. The resistance of an electric toaster or flat iron is about 25 ohms. These statements may assist the novice in visualizing these quantities.

Ohm's Law. In order to obtain a better understanding of electrical circuits, use will be made of the old but still useful fluid analogy. In Fig. 2–1 a pump is shown driven by a motor. This pump is used to circulate



(a) (c)

(b)

Pressure (lb/in²)	Flow (gal./min.)	Pressure / Flow
20	1.6	12.5
35	2.8	12.5
75	6.0	12.5
100	8.0	12.5
150	12.0	12.5

(d)

Electric pressure (volts)	Electric flow (amperes)	Volts / Amperes
8.0	0.92	8.7
12.0	1.38	8.7
16.8	1.93	8.7
22.4	2.57	8.7
47.0	5.40	8.7

Fig. 2–1. Analogy Between Hydraulic and Electric Circuits.

oil* through a cooling coil of small copper tubing. A gauge is connected to the ends of the copper tubing to measure the difference in pressure across the coil and a flow meter is inserted in the pipe to measure the rate at which the oil flows through the tubing.

If the speed of the pump is changed and if readings are taken of the pressure gauge and flow meter at each pump speed, a set of data will be obtained as shown in the table in Fig. 2–1b. It is important to observe that at each pump speed the pressure divided by the flow gives the same

* Oil is used instead of water because it is a liquid of high viscosity and will obey the Ohm's Law of the hydraulic circuit.

result. In this case that value is 12.5. The flow at any pressure can, of course, be found by dividing the pressure by 12.5. If the flow is desired at some pressure other than those tested it might also be obtained by dividing the pressure by 12.5.

Example: What is the flow at 50 lb per square inch pressure?

According to the relation stated above,

$$\text{flow} = \frac{\text{pressure}}{12.5} = \frac{50}{12.5} = 4 \text{ gallons per minute,}$$

pressure being in pounds per square inch. The constant 12.5* is characteristic of this particular size and length of tubing and so can be called the resistance of the coil of tubing.

To the right of this simple hydraulic circuit is shown a similar electric circuit. A battery supplies the electric pressure that causes an electric current to flow through a coil of copper wire indicated diagrammatically as R. The meter used to measure the electric pressure in volts is called a *voltmeter*. The meter used to measure the current in amperes is called an *ammeter*. If taps are arranged on the battery so that different voltages may be applied to the coil of wire, then a set of readings of volts and corresponding amperes can be made. These readings would be comparable to the pressure and flow readings of the hydraulic circuit. In the electric circuit the voltmeter reading divided by the ammeter reading is always 8.7 and this constant is called the *resistance*. It is seen by this analogy that in the electric circuit also it is possible to predict the current flow with any given voltage. For instance, if the current corresponding to 65 v were desired, then

$$I = \frac{E}{8.7} = \frac{65}{8.7} = 7.5 \text{ amp,}$$

E being in volts. This value of 8.7 is a characteristic of the wire and is called the resistance. It is measured in the unit which has previously been defined as the ohm. The formal statement of the relationship observed above is as follows:

The current in amperes is equal to the pressure in volts divided by the resistance in ohms.

This statement is known as *Ohm's Law* and is the basis for a large portion of electrical circuit theory. It may be expressed mathematically in the three forms below:

$$I = \frac{E}{R}, \qquad R = \frac{E}{I}, \qquad E = RI.$$

A word of caution should be given at this time, for although this is the general rule of behavior of electrical circuits, there are many exceptions.

* This constant depends also upon the viscosity of the fluid. In electricity the variable corresponding to viscosity does not occur.

Many of these cases of unusual behavior are the basis of the operation of important commercial equipment. No great worry should, therefore, be caused the student when he later meets with these exceptions.

Series Circuits. Electrical conductors may be connected following one another so that any current flowing through one must flow through the other. This is shown in Fig. 2-2. When circuits are connected in this manner the resistances are said to be connected in series. The combined or equivalent resistance of R_1 and R_2 connected as in Fig. 2-2 is

$$R_{total} = R_1 + R_2.$$

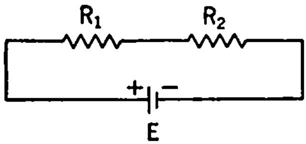


FIG. 2-2. Series Resistances.

The electrical pressure across R_1 when added to the electrical pressure across R_2 will equal the total battery pressure.

According to Ohm's Law the pressure or voltage across R_1 is R_1I and the voltage across R_2 is R_2I. Since the current is the same in both resistances, the voltage across the individual resistances will be proportional to their resistances. Also, the proportion of the total voltage across R_1 will be $R_1/(R_1 + R_2)$. This relationship is used many times in radio, and such a combination of resistances to give a reduced voltage is known as a *potentiometer* or *voltage divider.*

Example: If a resistance of 100,000 ohms is connected in series with a resistance of 5,000 ohms across a 90-v B battery, (a) what current will flow? (b) if the voltage across the 5,000-ohm resistor is used to control a vacuum tube, what would this voltage be?

Solution: The equivalent resistance of the two resistances in series is

$$R_{total} = 100{,}000 + 5{,}000 = 105{,}000 \text{ ohms.}$$

By Ohm's Law, the current will be,

$$I = \frac{E}{R} = \frac{90}{105{,}000} = 0.00086 \text{ amp.}$$

The voltage across the 5,000-ohm resistance is,

$$E = E_{battery} \frac{R_2}{R_1 + R_2} = 90 \frac{5{,}000}{105{,}000} = 4.28 \text{ v.}$$

Exercise 2-1. A vacuum-tube filament takes 0.9 amp at 6.3 v. (a) What is the resistance? (b) What additional resistance would be required if the filament were to be supplied from a 12-v battery?

Exercise 2-2. The heaters of five vacuum tubes are to be supplied in series. They require a current of 0.3 amp and each has a resistance of 21 ohms. What is the total resistance and what voltage must be used to supply the necessary current?

Exercise 2–3. How much resistance would be placed in series with a 50,000-ohm resistor to obtain a voltage of 8.4 across the 50,000-ohm resistor when a 45-v battery is the source of pressure?

Parallel Circuits. Resistances in electric circuits may be connected in parallel as shown in Fig. 2–3. When resistances are connected in this manner it will be noted that the battery voltage is impressed across each resistance just as if the other resistance were not there. The current in each resistance is determined by Ohm's Law as,

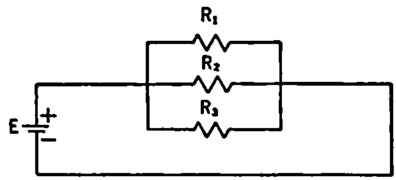


Fig. 2–3. Parallel Resistances.

$$I_1 = \frac{E}{R_1}, \quad I_2 = \frac{E}{R_2} \quad \text{and} \quad I_3 = \frac{E}{R_3}.$$

The total current in the circuit is the sum of the currents in the individual resistances, so

$$\begin{aligned} I_{total} &= I_1 + I_2 = I_3 \\ &= \frac{E}{R_1} + \frac{E}{R_2} + \frac{E}{R_3} \\ &= E\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right). \end{aligned}$$

The quantity $1/R$ is a constant called the *conductance* and is indicated by the symbol G. It is that characteristic of a resistance which when multiplied by the voltage gives the current. The unit of conductance is called the *mho.* This is recognized as the ohm spelled backwards and the name was chosen to be a reminder that the *mho* was the reciprocal of the ohm. Since the sum of the currents is equal to the total current, then the sum of the conductances in a parallel circuit is equal to the total conductance.

$$G_{total} = G_1 + G_2 + G_3 + \ldots$$

The proportion of the total current which is flowing in any one resistor is the conductance of that circuit element divided by the total conductance of the circuits which are in parallel. Expressed mathematically, this is

$$\frac{I_1}{I_{total}} = \frac{EG_1}{E(G_1 + G_2 + G_3 + \cdots)} = \frac{G_1}{G_1 + G_2 + G_3 + \cdots}$$

Example: Determine the equivalent resistance of the following four resistances connected in parallel. What proportion of the current will flow through the 8-ohm resistance?

$R_1 = 20$ ohms $\qquad R_2 = 25$ ohms

$R_3 = 12$ ohms $\qquad R_4 = 8$ ohms

Solution:

$$G_1 = \tfrac{1}{20} = .050 \qquad G_2 = \tfrac{1}{25} = .040$$
$$G_3 = \tfrac{1}{12} = .084 \qquad G_4 = \tfrac{1}{8} = .125$$

Total conductance is

$$G_{total} = G_1 + G_2 + G_3 + G_4 = .299$$

Equivalent resistance is

$$R_{eq} = \frac{1}{.299} = 3.34 \text{ ohms.}$$

Proportion of current flowing through the 8-ohm resistance is

$$I_{8\text{-ohm}} = \frac{G_4}{G_1 + G_2 + G_3 + G_4} = \frac{.125}{.299} = .418 = 41.8\%.$$

It is of interest to call attention to the special case where only two resistances are connected in parallel. This is so very common that it probably constitutes the majority of the problems in parallel circuits. In this special case,

$$I_{total} = I_1 + I_2 = \frac{E}{R_1} + \frac{E}{R_2}$$
$$= E\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = E\left(\frac{R_2 + R_1}{R_1R_2}\right)$$
$$= \frac{E}{\dfrac{R_1R_2}{R_1 + R_2}}.$$

Therefore,

$$R_{eq} = \frac{R_1R_2}{R_1 + R_2}.$$

This equivalent resistance will act in every way just as the two resistances in parallel, and so it is very common for engineers to refer to the equivalent resistance of two resistances in parallel as the product of the resistances divided by the sum of the resistances. It should be remembered, however, that this particular formula applies only to the case of two resistances in parallel.

Exercise 2–4. Two resistances, one of 50 ohms and one of 20 ohms, are connected in parallel across a 100-v line. What is the total current and the equivalent resistance?

Exercise 2–5. Two circuit elements are connected in parallel across a 240-v line. One has a conductance of .0063 mho and the other a conductance of .0172 mho. What current will be taken from the line? What proportion of this current will flow through the element having a conductance of .0172 mho?

Exercise 2–6. A 90-v battery supplies a total current of 0.134 amp to two parallel resistance elements. If one of the resistance elements takes 0.039 amp, what is the resistance of each element?

Exercise 2–7. Five resistances are connected in parallel. It is desired to know the total current and the proportion of this current going through R_3 when a potential of 24 v is applied. The resistances are: $R_1 = 4$ ohms, $R_2 = 7$ ohms, $R_3 = 22$ ohms, $R_4 = 10$ ohms, $R_5 = 65$ ohms.

Series-Parallel Circuits. Many times it is desirable to use combinations of series and parallel arrangements of resistances in radio equipment. The procedure used to solve circuits of this type is to combine the parallel resistances into equivalent series resistances and then determine the total current produced by the impressed voltage. This total current will divide in a parallel circuit in proportion to the conductances and so the current in any one of the resistances may be found.

Example: In the circuit shown in Fig. 2–4, determine the current in the 7-ohm resistance.

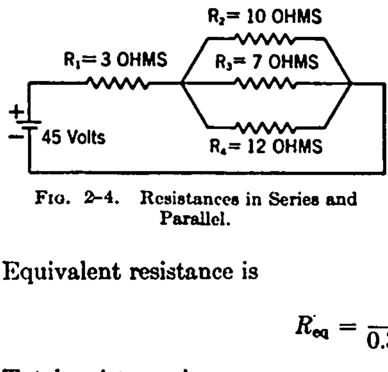


Fig. 2–4. Resistances in Series and Parallel.

Solution:

$$G_2 = \tfrac{1}{10} = 0.100$$
$$G_3 = \tfrac{1}{7} = 0.142$$
$$G_4 = \tfrac{1}{12} = 0.083$$

Equivalent conductance is

$$G_{eq} = 0.100 + 0.142 + 0.083 = 0.325.$$

Equivalent resistance is

$$R_{eq} = \frac{1}{0.325} = 3.08.$$

Total resistance is

$$R_t = 3.08 + 3 = 6.08.$$

Total current is

$$I_t = \frac{E}{R_t} = \frac{45}{6.08} = 7.40.$$

Current in 7-ohm resistance is

$$I_{7\text{-ohm}} = \frac{G_3}{G_{eq}} I_t = \frac{0.142}{0.325} \times 7.40 = 3.23 \text{ amp}.$$

An alternate method of obtaining the current in any branch of a parallel circuit is to determine the voltage across the circuit from the equivalent IR drop and then divide this voltage by the resistance to obtain the current in that circuit element.

Example: In the circuit shown in Fig. 2–5 determine the current in the 10-ohm resistance.

Solution: Equivalent resistance of parallel circuit is

$$R_{eq} = \frac{10 \times 15}{10 + 15} = \frac{150}{25} = 6 \text{ ohms.}$$

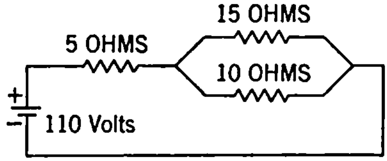


FIG. 2–5. A Series-Parallel Circuit.

Total resistance of circuit is

$$R_t = 5 + 6 = 11 \text{ ohms.}$$

Total current is

$$I_t = \frac{E}{R_t} = \frac{110}{11} = 10 \text{ amp.}$$

Volts across parallel resistances are

$$E_p = IR_{eq} = 10 \times 6 = 60$$

Current through 10-ohm resistance is

$$I_{10\text{-ohm}} = \frac{E_p}{R_{10\text{-ohm}}} = \frac{60}{10} = 6 \text{ amp.}$$

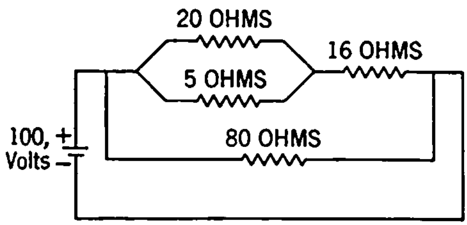


FIG. 2–6. Circuit for Exercise 2–8.

Exercise 2–8. Determine the total current and the current in the 5-ohm resistor of the circuit of Fig. 2–6.

Exercise 2–9. Determine the total current and the current in the 25-ohm resistance of the circuit shown in Fig. 2–7.

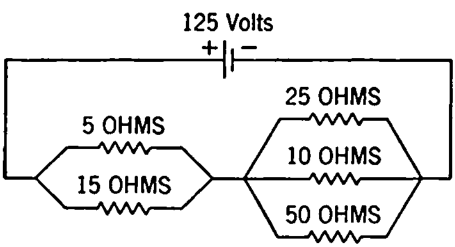


FIG. 2–7. Circuit for Exercise 2–9.

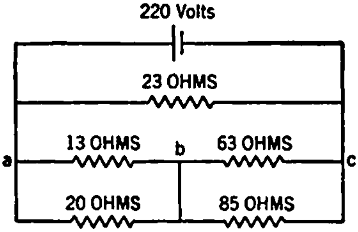


FIG. 2-8. Circuit for Exercise 2-10.

Exercise 2-10. What is the difference of potential across *ab* and across *bc* of the circuit shown in Fig. 2-8.

Determination of Resistance. Since copper wire is used so extensively in radio circuits it is important to know the resistance of various sizes of wires. In Table 2-1 not only the resistance but several other items of information are given for various sizes of copper magnet wire.

TABLE 2-1

PROPERTIES OF COPPER MAGNET WIRE

Size of wire (A.W.G.)	Diameter (mils*)	Ohms per 1000 ft at 20° C.	Pounds per 1000 ft	Diameter C† in mils	Diameter E‡ in mils	Diameter EC§ in mils
0000	460.0	.0490	640.5	468.		
000	409.6	.0618	507.9	418.		
00	364.8	.0779	402.8	375.		
0	324.9	.0983	319.5	333.		
1	289.3	.1239	253.3	297.		
2	257.6	.1563	200.9	266.		
3	229.4	.1970	159.3	237.		
4	204.3	.2485	126.4	212.		
6	162.0	.3951	79.46	170.		
8	128.5	.6282	49.97	134.	131.	136.
10	101.9	.9989	31.43	107.	104.	109.
12	80.81	1.588	19.77	85.8	83.0	88.0
14	64.08	2.525	12.43	69.1	66.1	71.1
16	50.82	4.016	7.818	55.8	52.6	57.6
18	40.30	6.385	4.917	45.3	42.0	47.0
20	31.96	10.15	3.092	37.0	33.5	38.0
22	25.35	16.14	1.542	29.4	26.8	31.3
24	20.10	25.67	1.223	24.1	21.3	25.8
26	15.94	40.81	.7692	19.9	17.0	21.5
28	12.64	64.90	.4837	16.6	13.6	17.6
30	10.03	103.2	.3042	14.0	10.8	14.8
32	7.950	164.1	.1913	12.0	8.75	12.8
34	6.305	260.9	.1203	10.3	7.01	11.0
36	5.000	414.8	.0757	9.00	5.60	9.60
38	3.965	659.6	.0476	7.97	4.47	8.47
40	3.145	1049.0	.0299	7.15	3.55	7.55

* 1 mil = 0.001 in.
† C means single cotton covered.
‡ E means enameled.
§ EC means enameled with single cotton covering.

Occasionally materials other than copper are used, and if the resistance relative to that of copper is known, the resistance of wires of these other materials may be found by multiplying the resistance of copper wire of the same size by the relative resistance. A list of relative resistances is given in Table 2-2.

TABLE 2-2

RESISTANCES OF METALS: RATIO OF RESISTANCE OF METAL TO RESISTANCE OF COPPER

Pure metals	Resistance relative to copper	Alloys	Resistance relative to copper
Iron	5.80	Radiohm	77
Zinc	3.43	Nichrome	65
Tungsten	3.20	Advance	28
Aluminum	1.55	High brass	4.8
Gold	1.40	Low brass	3.8
Silver	.943	Commercial bronze	2.4

The resistances here given are at normal or room temperature and are satisfactory for most use. It is well to remember that in most cases the resistance of metals increases with temperature and if the temperature is very high or if very accurate results are required, temperature effects must be considered.

Kirchhoff's Laws. Two rules or laws known as Kirchhoff's Laws are important in solving complicated electric circuits. These laws were implied in the solutions of series and parallel circuits but are stated explicitly as follows:

(1) *The current flowing into any junction of an electric circuit is equal to the current flowing out of that junction.*

(2) *The sum of the battery or generator voltages around any closed circuit is equal to the sum of the voltage drops in resistances around the same circuit.*

Example: The use of these laws is illustrated by determining the current flow in the 10-ohm resistance of Fig. 2-9. The currents are shown as I_{ab}, meaning that the current specified is the current flowing from a to b, I_{cb} the current flowing from c to b, and I_{bd} the current from b to d.

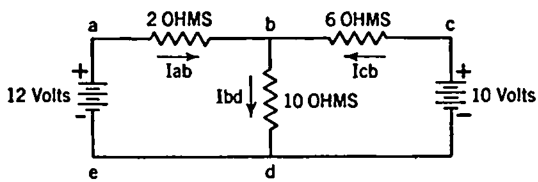


FIG. 2-9. Circuit Illustrating Use of Kirchhoff's Laws.

Using Kirchhoff's first law at the point b,

$$I_{ab} + I_{cb} = I_{bd}.$$

Using Kirchhoff's second law around circuit *eabd,*

$$2I_{ab} + 10I_{bd} = 12.$$

Using Kirchhoff's second law around the circuit *fcbd,*

$$6I_{cb} + 10I_{bd} = 10.$$

The above three equations have three unknown currents I_{ab}, I_{cb}, and I_{bd}. Since $I_{ab} =. I_{bd} - I_{cb}$, the second equation may be written as,

$$2I_{bd} - 2I_{cb} + 10I_{bd} = 12$$

or

$$12I_{bd} - 2I_{cb} = 12.$$

Multiplying both sides by 3 and adding to the third equation of the set of three above,

$$\begin{array}{rl} 36I_{bd} - 6I_{cb} &= 36 \\ 10I_{bd} + 6I_{cb} &= 10 \\ \hline 46I_{ab} \qquad\quad &= 46 \end{array}$$

$$I_{bd} = 1 \text{ amp.}$$

Method of Superposition. Another method of obtaining the solution for a circuit having several voltages is based on a principle often used in electrical theory. It states that *the current in a wire is the sum of the currents produced by each voltage acting by itself and with the other voltages shorted out.*

Example: In Fig. 2–9, I_{bd} can be obtained due to the 12-v battery with the 10-v battery shorted.

The equivalent resistance of the 5-ohm and 10-ohm resistances in parallel is $(5 \times 10)/15 = 3.33$. This is in series with the 2-ohm resistance, so that the total current is

$$I = \frac{12}{5.33} = 2.25 \text{ amp.}$$

One third of this current goes through the 10-ohm resistance, so the current contributed by the 12-v battery is $2.25/3 = .75$ amp. The current due to the 10-v battery can be found by shorting out the 12-v battery. The equivalent parallel resistance is

$$\frac{2 \times 10}{2 + 10} = \frac{20}{12} = 1.66 \text{ ohms.}$$

This is in series with the 5-ohm resistance, so that the total current is

$$\frac{10}{6.66} = 1.5 \text{ amp.}$$

Only one sixth of this current will go through the 10-ohm resistance Hence, the contribution of the 10-v battery is

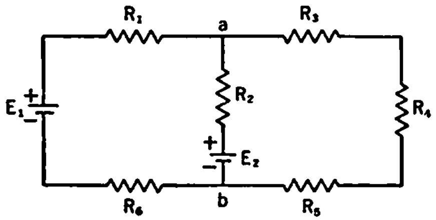
$$\frac{1.5}{6} = 0.25 \text{ amp.}$$

The total current is, then,

$$0.75 + 0.25 = 1.0 \text{ amp.}$$

The answer by this method is, of course, the same as that obtained by the use of Kirchhoff's Laws.

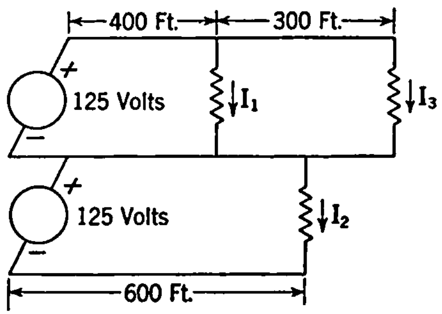
Exercise 2-11. Determine the voltage across *ab* of Fig. 2-10 by using both Kirchhoff's Laws and by the method of superposition.



E_1 = 120 volts	E_2 = 60 volts
R_1 = 10 ohms	R_2 = 20 ohms
R_3 = 5 ohms	R_4 = 7 ohms
R_5 = 8 ohms	R_6 = 2 ohms

FIG. 2-10. Circuit for Exercise 2-11.

Exercise 2-12. A three-wire distribution system is supplied with two 125-v generators connected in series as shown in Fig. 2-11. Determine the voltage across each load if the distribution wire is all No. 2 A.W.G. copper wire.



I_1 = 40 amp
I_2 = 65 amp
I_3 = 35 amp

FIG. 2-11. Circuit for Exercise 2-12.

Power and Energy. The manner of electron flow through conductors was discussed in the introduction of this chapter. The agitation or the increase in the random movement of the molecules owing to the fact that they are being continuously hit or bumped into by the drifting electrons was noted, and it was stated that this resulted in an increase in temperature. The passage of current through a conductor having resistance is always associated with such a generation of heat. The relation between the current, voltage, and resistance of the circuit and the conversion of electric energy into heat is an important element in the study of electric circuits.

Referring again to the hydraulic circuit of Fig. 2-1, it is known that for constant flow the energy put into the circuit by the pump will be doubled if the pressure is doubled. It will also be doubled if the pressure remains constant and the flow is doubled. A similar variation in power exists in the case of the electric circuit. The power converted into heat in a resistance may be said to be directly proportional to the product of the current and the voltage. Expressed mathematically, this is,

$$P = E \times I,$$

P being expressed in watts, E in volts, and I in amperes. This statement may then be used as a definition of a watt. *The watt is the rate at which electric energy is being supplied when a current of one ampere is flowing at a pressure of one volt.*

Several additional equations for power may be derived from the above

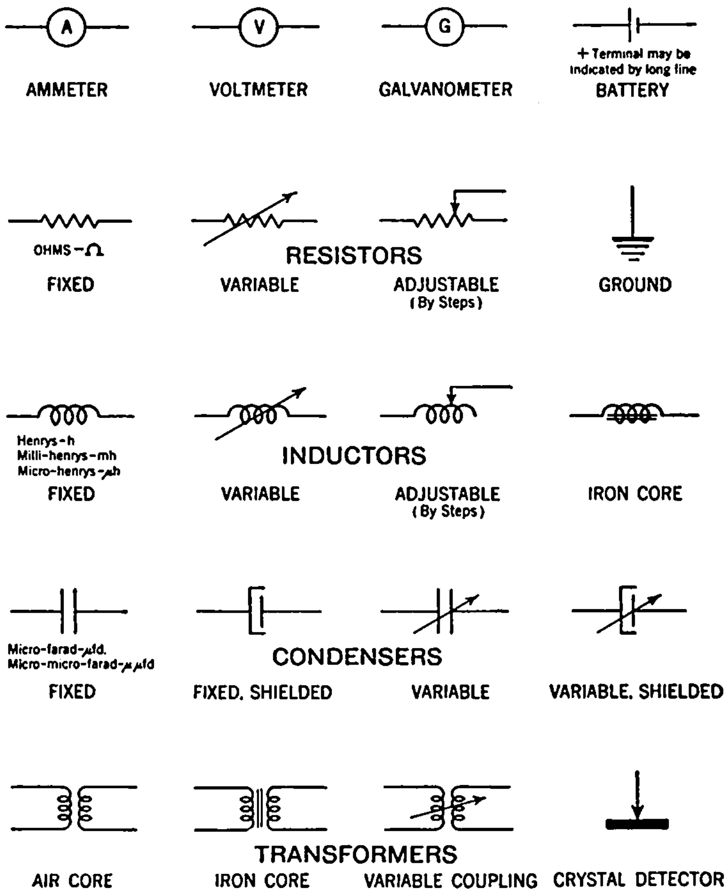


Fig. 2-12. Conventional Circuit Symbols.

statement by the use of Ohm's Law. These are very useful when the information supplied is not given in volts and amperes. These equations are:

$$P = E \times I = IR \times I = I^2R. \qquad (\text{since } E = IR)$$

$$P = E \times I = E \times \frac{E}{R} = \frac{E^2}{R}. \qquad \left(\text{since } I = \frac{E}{R}\right)$$

Since power is the rate at which energy is being transferred, the total energy is the product of the power and the time. Thus, a small unit of energy is the *watt-second* or *joule.* The more common unit, however, is a much larger one known as the *kilowatt-hour.* This unit specifies an

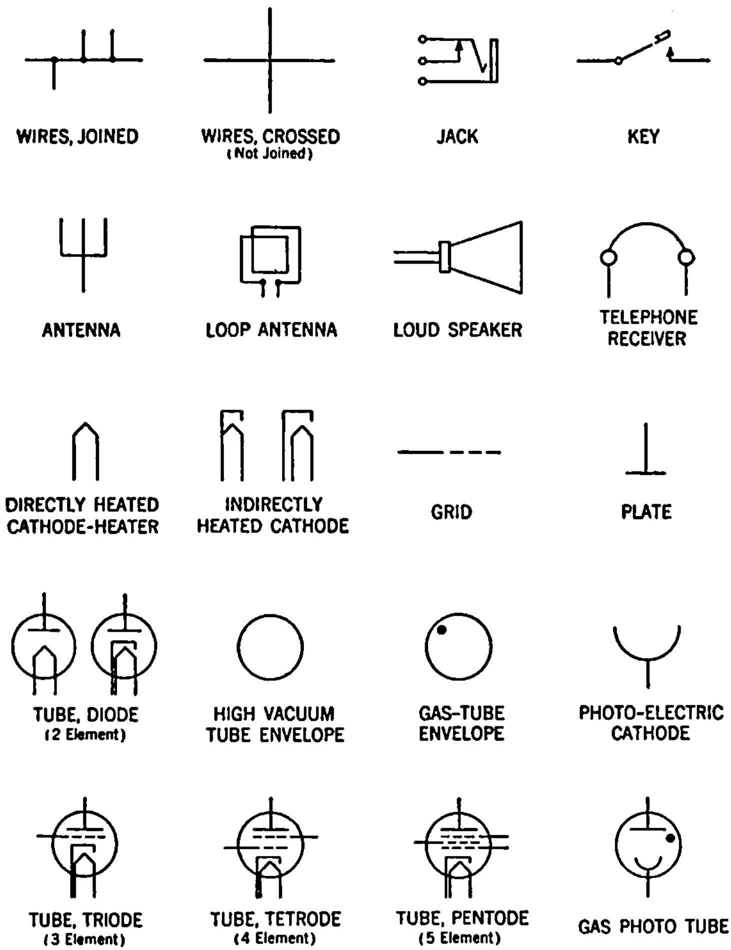


Fig. 2-13. Conventional Circuit Symbols.

energy equivalent to 1,000 watts, or one kilowatt continued over a period of one hour. It is this unit which is the basis of most of the bills for electric energy issued by the power companies.

Example: The cathode of a type 850 vacuum tube requires 3.25 amp at 10 v. What is the power requirement for the cathode heater?

Solution:
$$P = EI = 10 \times 3.25$$
$$= 32.5 \text{ watts.}$$

Fuses. One of the important ways in which the heating effect of the electric current is used is the insertion into the circuit of a resistance unit with small current-carrying capacity, so that when the current goes beyond a certain predetermined amount the resistance is burned up and opens the circuit. This resistance unit is called a *fuse* and is used to protect other and more expensive equipment from harmful effects when the current becomes too large. Fuses are of many different types, and range in size from a few milliamperes up to hundreds of amperes. Since they are placed in a circuit to protect the equipment, they should not be replaced by larger fuses or by heavy conductors because they were specifically designed for the purpose of opening the circuit under overload conditions. Oversize fuses or solid jumpers defeat the purpose of the fuses and permit operation at overload with consequent damage to equipment.

Symbols and Abbreviations. Many kinds of equipment are used in radio work and the circuits are often quite complicated. In order to simplify the appearance as much as possible a standard set of symbols is used to indicate the circuit elements. In order that they be available in one place they are assembled in Figs. 2–12 and 2–13. Below the symbol is given the name of the element and the name and abbreviations of the units where possible.

The very wide range of magnitudes of quantities used in radio has led to the adoption of many units which are decimal parts of the basic unit. Thus *milli-* placed before the name of a unit such as millivolt means that one millivolt is one thousandth of a volt. A list of prefixes and the size of the new unit in terms of the original unit was listed in Table 1–2. The use of such units saves many troublesome decimals. *It is necessary, however, to remember that the circuit laws are based on ohms, amperes, and volts, and other units must be converted to these before circuit problems can be solved.*

Batteries. An important source of electric pressure for radio purposes is the electric cell or battery of cells. Batteries are of two types. One of these, known as the *primary battery,* uses up the original materials and is thrown away after its useful life is accomplished. The common commercial form for this type is the *dry battery* or *dry cell,* which is used in portable sets and elsewhere when small amounts of power are required over a short period. The other type, known as the *secondary* or *storage battery*, may be recharged by forcing current through it in the reverse direction after it has supplied its normal amount of electric current. Usually storage batteries are of larger current capacity than dry batteries and are used as main sources of power on portable transmitters and receivers where it is possible to recharge the batteries conveniently, as in the case of sets that are located in automobiles.

DRY CELL. A diagram of the construction of a dry cell is shown in Fig. 2-14. The positive electrode of the dry cell is a fairly large carbon

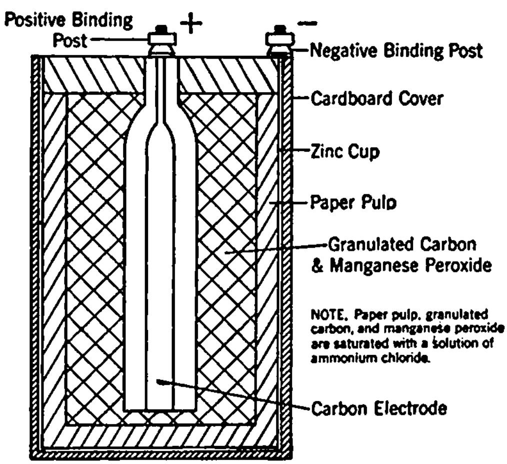
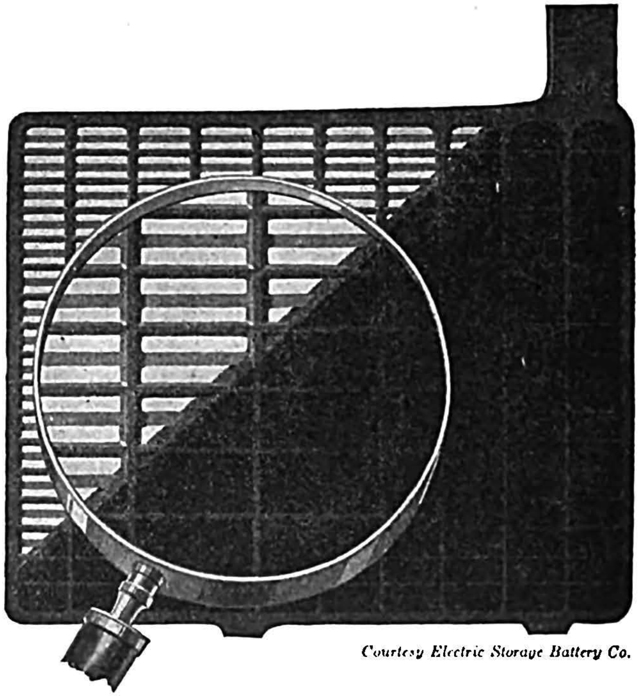


FIG. 2-14. A Dry Cell.

rod located in the center of a zinc container that forms the negative electrode. The electrolyte, which is a dilute solution of ammonium chloride (sal ammoniac), is mixed with some porous inert material to form a paste which is placed on the inside of the zinc container. Between this paste and the carbon rod located in the center is a porous mass of manganese dioxide and carbon granules. This material is called a depolarizer and its function is to absorb the ammonia and hydrogen gas that is given off at the positive electrode and tends to insulate the electrode. The top of the cell is sealed by an insulating compound so that only the terminals connected to the carbon rod and to the zinc container are visible. The zinc container is then placed in a cardboard carton, which acts as a protection and as an insulator. The larger dry cells, approximately $2\frac{1}{2}$ inches in diameter and 6 inches tall, are supplied as individual cells. Where larger voltages are desired and very small currents are sufficient, a number of small cells are connected in series and mounted in a common container. Several terminals will probably be brought out in order to supply different voltages. This unit is the well-known B battery used to supply voltage to the plates of vacuum tubes.

STORAGE BATTERIES. The *lead storage battery*, which is the most common type, consists of a positive plate of lead peroxide and a negative plate of porous or spongy lead in a dilute solution of sulphuric acid. The usual construction of batteries of the portable type is known as the pasted-plate construction. In this a lead and antimony alloy is used to make a grid of

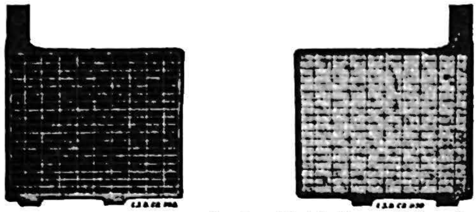


Courtesy Electric Storage Battery Co.

FIG. 2-15. Grid of Storage Battery Partly Filled with Active Material.

the type shown in Fig. 2-15. This grid is filled in with a paste which, after an electrochemical forming process, gives lead peroxide on the positive plate. This lead peroxide is the active chemical material, while the grid acts as physical support for the chemical and as electrical conductor for the current that is developed by chemical action. A similar grid is used for the negative plate but a different paste is used, which after forming consists mainly of pure lead in very porous condition.

In Fig. 2-16 is shown a completed positive and negative plate, while in Fig. 2-17 several of these positive and negative plates are assembled into groups. In Fig. 2-18, porous wood and fiber glass separators are shown. The plates of the positive group are nested between the negative plates, being separated from them by the thin and porous separators made

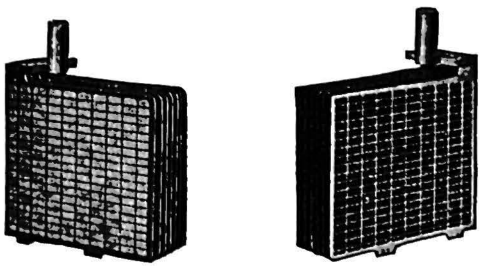


Courtesy Electric Storage Battery Co.

Positive Negative

FIG. 2-16. Storage-battery Plates.

of either wood, glass, rubber, or combinations of them. The whole unit is then set in a rubber jar with a cover to support the terminals and is sealed

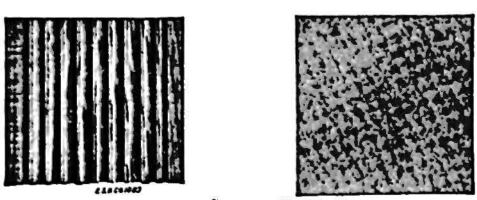


Courtesy Electric Storage Battery Co.

Negative Positive

FIG. 2-17. Storage-battery Plates, in Groups.

into the main jar as shown in Fig. 2-19. The sulphuric acid electrolyte is introduced through the filling hole in the cover.



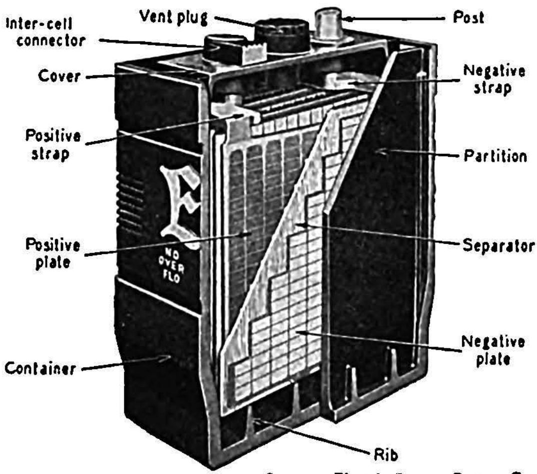
Courtesy Electric Storage Battery Co.

Wood Fiber glass

FIG. 2-18. Storage-battery Separators.

This construction permits maximum surface area to be exposed to the electrolyte with a minimum of distance between the positive and negative plates, this distance being only the thickness of the separators. The separators are sufficiently porous that they offer little hindrance to the flow of ions which are the carriers of electric charge, but they do maintain mechanical and electrical isolation of the electrodes.

The chemical reactions that take place are rather complicated but the final result is that both plates change their active material to lead sulphate as the battery is discharged. This lead sulphate has a tendency to harden, crystallize, and expand if permitted to stand. For this reason it is important to keep lead batteries in a charged condition. A battery which remains in a discharged condition soon will have buckled plates and reduced capacity, and will have to be discarded long before its normal life span.



Courtesy Electric Storage Battery Co.

FIG. 2-19. An Assembled Lead Storage Cell (cut away to show the construction).

INTERNAL RESISTANCE AND POLARIZATION. In both the dry cell and the storage battery the source of energy is chemical, and the products of chemical reaction may tend to insulate the electrodes or may tend to cause chemical or ion concentrations that reduce the terminal voltage when current is drawn from the battery. This action is known as polarization and in some cases may continue for some time after the current drain has been stopped. Usually the construction of the battery will permit these effects to be neutralized after a time. This entire process acts like, and is sometimes called, the internal resistance of the battery. Polarization is much more pronounced in the dry cell than in the storage battery.

Electromagnetism. A phenomenon which is closely associated with electricity is *magnetism*. The horseshoe magnet and its ability to pick up iron and steel objects are universally familiar. Apparently this attraction is caused by some form of disturbance in the space surrounding the magnet.

This disturbance is described as a magnetic field and is indicated by lines connecting the north and south poles, as shown in Fig. 2–20. The magnet shown here is not bent as is the horseshoe magnet, so is called a bar magnet.

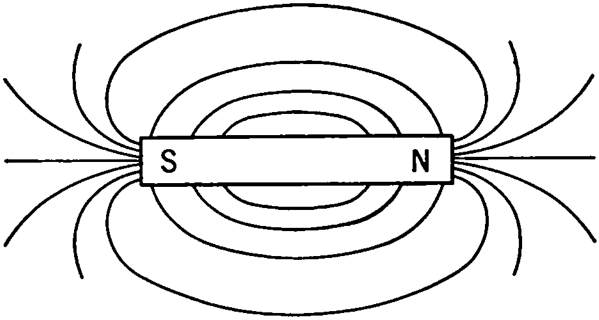


FIG. 2–20. The Magnetic Field of a Bar Magnet.

In Fig. 2–21 two magnets are placed with the north pole of one near the south pole of the other. The lines of magnetic flux flow from the north to the south pole, and since the poles attract one another we say with regard to this arrangement that the lines of magnetic flux tend to shorten themselves, much as stretched rubber bands might do. In Fig. 2–22 another arrangement of two bar magnets is shown, wherein the north pole of each magnet is near the north pole of the other. The lines of magnetic flux must leave both north poles and find their way back to the south poles. In this case they are squeezed sidewise against each other, and since the poles repel one another we say with regard to this arrangement that the lines of magnetic flux tend to exert a sidewise push or lateral force upon one another.

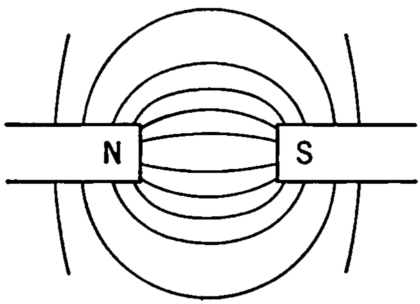


FIG. 2–21. The Magnetic Field about Two Unlike Poles.

Experiments indicate that wires carrying electric current have magnetic effects. In Fig. 2–23 a wire is shown that is carrying electric current from the right to the left. The north pole of a pocket compass located above the wire will point away from the observer and indicates that the arrows on the flux are in the same direction. This effect gives rise to a rule that is commonly used to remember the magnetic effect of an electric current. *If the right hand is placed around the wire with the thumb pointing in the direction of the current, then the fingers will indicate the direction of magnetic* *flux.* If the wire which is carrying electric current is wound around a cylindrical tube the magnetic effects of the various turns of wire aid each other and

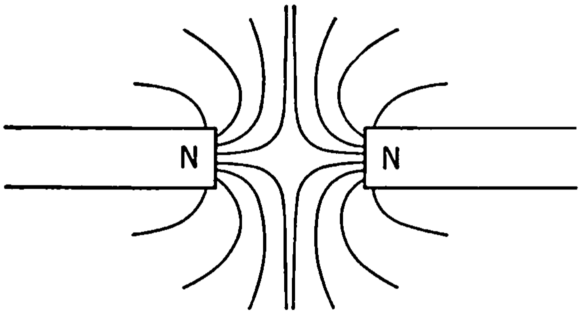


FIG. 2-22. The Magnetic Field about Two Like Poles.

produce a field very similar to that of a bar magnet, as is shown in Fig 2-24. The strength of the magnetic field depends upon the magnitude of the current and the number of turns in the coil. This relation has caused the *ampere-turn* to be adopted as a common unit of magnetic pressure or magnetomotive force. The ampere-turn may be defined as the magnetic pressure produced by a current of one ampere flowing in a loop of one turn. Thus a current of 3 amperes flowing in a coil of 5 turns would produce a magnetic pressure of 3 times 5 or 15 ampere-turns.

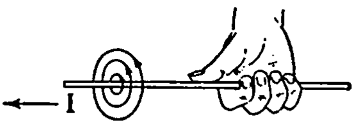


FIG. 2-23. The Direction of Magnetic Flux Is Indicated by the Right-hand Rule.

Iron as a Conductor of Magnetism. If an iron rod, not previously magnetized, is placed in a coil that is carrying current, it is observed that the strength of the magnetic field is greatly increased. This increase is due to the unusual ability of iron to act as a conductor of magnetic flux. The ratio of the flux produced in iron to that produced in air under similar magnetic pressure varies, but will range between several hundred and six or seven thousand for most commercial steel.

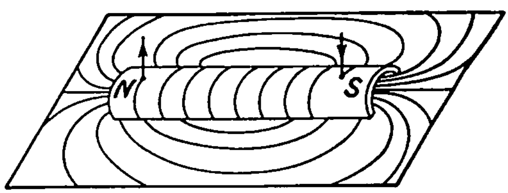
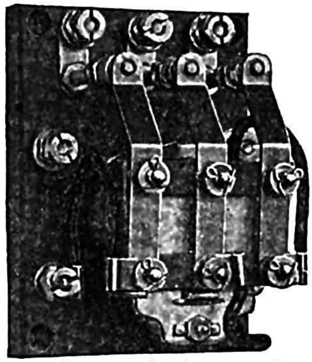


FIG. 2-24. The Magnetic Field about a Coil of Wire Carrying Current.

A simple illustration of the manner in which iron and steel are used to control magnetic flux in electric equipment is the electric contactor switch shown in Fig. 2–25. Here an electric coil is placed on a U-shaped iron assembly and an iron armature is hinged so that it normally will fall open due to gravity. When the coil is energized with electric current, the flux set up will cause the armature to be attracted and the contactor will be closed.



Courtesy of Struthers Dunn, Inc.

FIG. 2–25. A Magnetic Relay.

MAGNETIC CHARACTERISTICS OF IRON. Since iron and steel are of such importance in the construction and operation of electrical equipment it is of value to study the properties of iron and of various other types of magnetic materials having iron as their base.

Scientific investigation has shown that iron has the unusual characteristic of increasing the magnetic effect of an electromagnet because the iron molecule is itself a tiny electromagnet. In soft iron, these molecules are more or less free to assume a random arrangement and so no external effect is noticed. When the iron is placed in a coil of wire carrying electric current, the molecules tend to arrange themselves in line with the field and thus their magnetic pressure or magnetomotive force is added to that of the coil. When the current is stopped or when the iron bar is removed from the coil, the molecules in the iron again tend to take on a random arrangement. This process is not complete, however, since a type of molecular friction causes a small amount of magnetic effect to remain. This remainder is known as residual magnetism. If the soft iron is replaced by hardened carbon steel, then the alignment of molecules will not be as complete as in the case of the soft iron because of the vastly greater internal or molecular friction. When the hardened steel is removed from the coil it will, however, retain most of the magnetism and may be called a *permanent* magnet. The internal friction or resistance to magnetic alignment is called *hysteresis*, and a curve showing the variation of magnetism or magnetic flux density with magnetizing pressure or force is known as a hysteresis loop.

In Fig. 2–26 are shown several hysteresis loops that are typical of three distinctly different kinds of use. The one on the left represents the kind of steel used for motors, generators, and transformers. The one in the center is representative of permanent magnets, while the one on the right is used for coils and transformers in telephone circuits where the currents are very small. The hysteresis loop is determined by increasing the magnetizing force and measuring the magnetic flux until the maximum is reached at a in the center loop of Fig. 2–26. The magnetizing force is then reduced and the flux drops to b, after which the magnetizing force is reversed and increased to c and then on to d, which is the maximum in the opposite direction. The force is then reduced, reversed, and increased again to *a*. It will be seen that more magnetizing force is required for the same magnetic effect when the flux is increasing than when it is decreasing. The horizontal distance between the sides of the loop is then a measure of the friction or hysteresis. The actual energy loss for each magnetic reversal is proportional to the area of the loop. The residual magnetism is *ob*. The reversed magnetizing force necessary to bring the flux back to zero is *oc* and is called the *coercive force*. The average amount of flux produced by

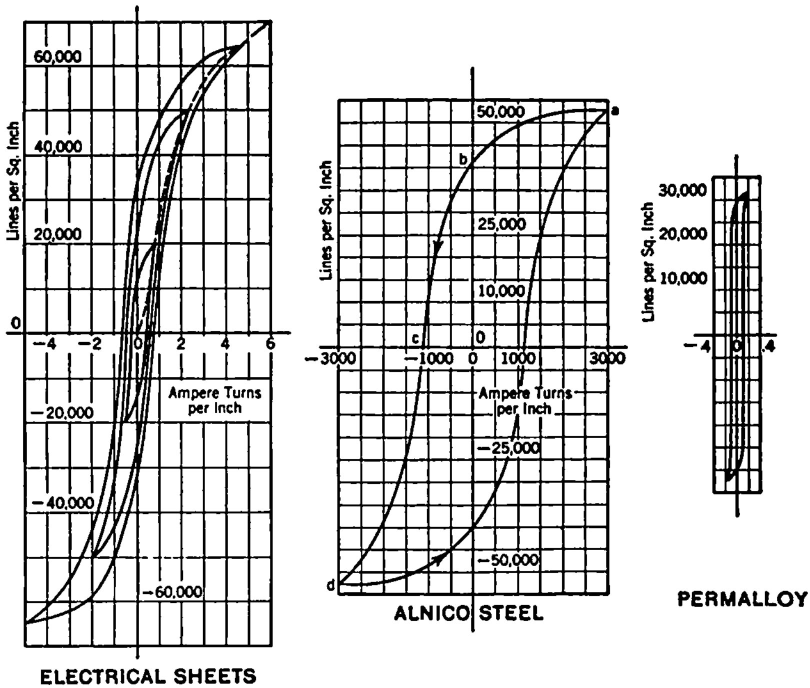


Fig. 2–26. Typical Hysteresis Loops.

a magnetizing force that is reversing will be a curve drawn through the points of the hysteresis loops as shown in Fig. 2–26a. Such a curve is called a *magnetization curve* and is important in determining the action of an iron core in a transformer. Magnetization curves for several different magnetic materials are given in Fig. 2–27.

Dynamo and transformer steel. Since the energy loss is proportional to the area of the loop, it is desirable that the steels used for A.C. machines have narrow hysteresis loops. It has been found that the electrical and magnetic characteristics are improved by adding some silicon to the steel. Most steel for dynamos and transformers is used in the form of sheet punchings which are stacked up to give an adequate magnetic path. Such construction is called a laminated type of magnetic structure and the material for it is supplied in large sheets known commercially as electrical sheets.

PERMANENT MAGNET STEELS. How hardened carbon steel may form permanent magnets has already been mentioned and explains the phenomenon of magnetizing the blades of pocket knives, hardened screwdrivers, and other tools. It has been found that an alloy of iron with aluminum, nickel, and cobalt, when properly heat-treated, gives very high-strength permanent magnets. These are known as *alnico* magnets and are widely used. Other

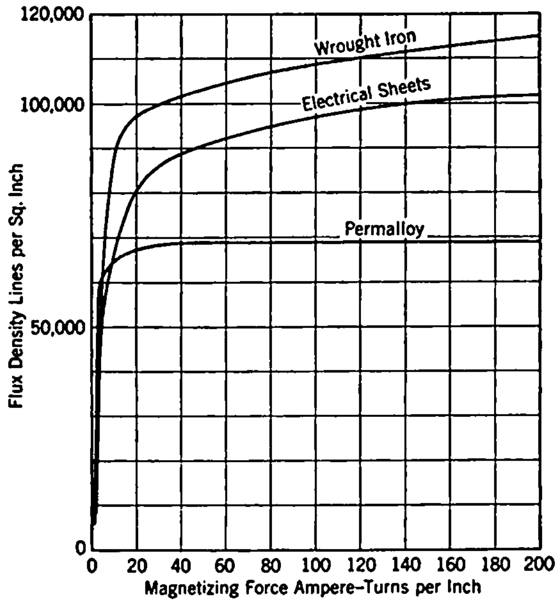


FIG. 2-27. Typical Magnetization Curves.

alloys including combinations of tungsten, chrome, and cobalt are also extensively used. They are characterized by high retentivity and high coercive force.

PERMALLOY. Still another type of magnetic material is known as permalloy and is an alloy of iron with nickel which is carefully heat-treated' to give high magnetic flux with very small magnetizing force and to have a very narrow hysteresis loop. This material is used in coils and transformers on telephone circuits where the currents are quite weak.

D.C. Generators. When a wire is moved across a magnetic field as shown in Fig. 2-28, an electric pressure or electromotive force is set up between the two ends of the wire. This action is the basis for the generation of nearly all electric power and so is very important. The amount of pressure set up is dependent upon the strength of the field, the length of the wire that is in the field, and upon the velocity of the wire across the field. Since the strength of the magnetic field is represented by the density of the lines, the voltage can then be said to be proportional to the rate of cutting flux.

Associated with this generating action is another action, based upon the fact that a force is produced on a wire located in a magnetic field and carrying electric current. If the current is flowing in the same direction as the voltage that is being generated, the force exerted on the wire will be opposite to the direction of movement. In other words, if the conductor is moving up through the magnetic field of Fig. 2–28, a voltage will be generated tending to send current out of the paper. If now a current is permitted to flow, this current will produce a force that opposes the motion. This condition is required by the law of conservation of energy and

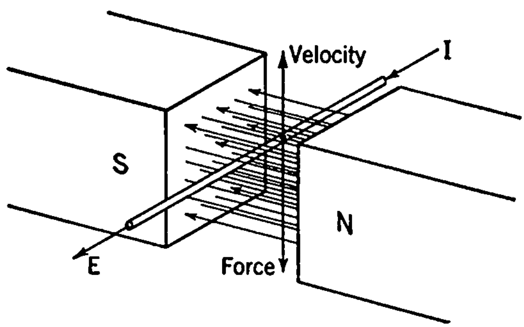


Fig. 2-28. Relation Between (a) Magnetic Field, Motion, and Voltage; (b) Magnetic Field, Current, and Force.

explains the change from mechanical to electrical energy in an electric generator. It should be observed that the voltage being generated is not influenced by the current flowing but is dependent only on the strength of the magnetic field and the speed of the generator. Likewise, the force or torque is independent of the speed of the generator, but does depend upon the magnitude of the field and of the current.

In Fig. 2–29 is shown an elementary form of direct-current generator. A cylindrical iron ring is mounted on bearings so that it can rotate between the two magnetic poles. The flux enters the ring on the side of the north pole and leaves at the south pole. A wire is wound around the ring as indicated in the diagram, and each turn is connected to an insulated copper bar in an assembly called the *commutator*. The commutator is composed of a number of tapered copper bars separated from each other by mica insulation, clamped together, and machined to form a smooth cylindrical surface. Carbon blocks or brushes are held in stationary supports and make contact with the copper bars on the rotating commutator.

When the armature is rotated the conductors on the outside surface cut the flux under the poles and electric pressure is produced. The conductors under the north pole produce a voltage out of the page. Each of these is added to the other so that the difference of pressure between the brushes is the sum of the voltages produced by all of the conductors on that side of the armature. The voltages generated by the conductors under the south pole are into the page, and these will also be additive, so the difference in pressure is the same as it was for the side under the north pole. The direction of current flow will thus be out of the bottom brush and into the top brush for both windings.

A careful study of the diagram shows that the rotation will change the relative position of the individual conductors on the surface of the armature but will not alter the voltage generated with respect to the brushes. This, then, is a direct-current generator.

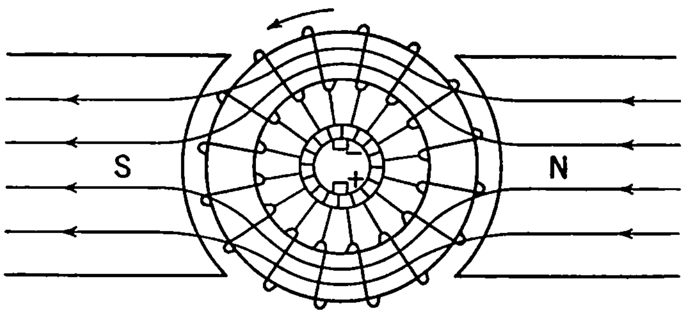


FIG. 2–29. An Elementary D.C. Generator.

When the brushes are connected to an external circuit, a current will flow through the external circuit from the positive to the negative brush, and this same current will flow from negative to positive within the machine, dividing equally between the two sides of the machine. The current flows in the same direction in the windings as the voltage generated, so the force produced by the conductors is opposed to the motion of the conductors or armature and tends to cause the generator to slow down. This set of forces must be overcome by some form of motor or engine, usually called the prime mover. The prime mover supplies the mechanical energy that is converted to electrical energy in the generator.

D.C. Motor. If the generator discussed above is connected to a direct-current power line and if the engine driving it is then disconnected, the generator will start to slow down. Since the voltage is directly proportional to the speed of the generator, as soon as it slows down the voltage produced in the windings is reduced and current will be forced through the windings by the larger external voltage, in a direction opposite to the direction of voltage previously generated. This change causes a force to be exerted on the conductors in the same direction as the motion, and so the machine continues to run. If a load of some type, such as a pump, is connected to the electric machine it slows down still more so that a larger current can flow and produce more force tending to continue the rotation. When an electric machine or dynamo is operating in this manner it is called a *motor*.

Dynamotors. In radio transmitting, quite high voltages are required, with a power demand beyond the capacity of small B batteries. In order to meet this need a high-voltage generator is driven by a low-voltage motor, which in turn is supplied from the storage battery. Since it is desirable to

Courtesy Pioneer Gen-E Motor Co.

Fig. 2-30. Two Commercial Dynamotors.

reduce weight as much as possible in portable equipment, it is customary to combine the two machines on one magnetic frame.

The motor winding is composed of a few turns of large wire connected to a commutator on one end. The generator winding consists of a large number of turns of small wire mounted on the same iron core but connected to a commutator on the other end of the machine. A motor-generator set of this type is called a dynamotor.

In commercial machines a drum type of armature core is used and the conductors are placed in slots. Figure 2-30 shows dynamotors used for portable radio power supply.

Electric Meters. The practical use of the circuit theories that have been discussed depends to a great extent upon an ability to measure the magnitudes of the currents, the voltages, and the resistances of equipment in actual operation. It is the student's ability to analyze and interpret correctly the variations of meter readings that is the ultimate justification for circuit analysis. A brief discussion of the theory and construction of direct-current meters is therefore of considerable interest.

Nearly all of today's direct-current meters are based on a design that was developed by Arsène d'Arsonval in 1881 and so are called *D'Arsonval-type* meters. The diagrammatic construction of this type of instrument is shown in Fig. 2–31, while in Fig. 2–33 is shown a partial assembly of a commercial instrument. The main parts of the instrument are the permanent magnet M to which the soft iron pole pieces P are fastened. The soft iron core C is held by nonmagnetic supports that are bolted to the pole pieces. The magnetic assembly is an integral unit with a uniform air gap between the core and the pole pieces. This design produces in the air gap a magnetic field that is constant in magnitude and radial in direction.

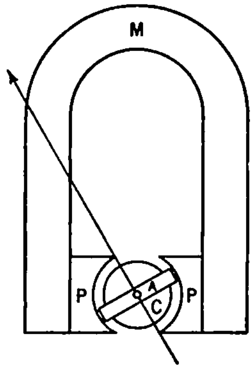


Fig. 2–31. A D'Arsonval Meter.

A coil assembly such as that shown in Fig. 2–32 is supported in jeweled bearings so that it is free to rotate back and forth in the air gap. The coil itself is composed of many turns of fine wire. The hairsprings located at the top and bottom hold the coil in the zero position when no current is flowing in the coil. They also act as insulated connections to the two ends of the coil. The counterbalanced pointer indicates the movement of the coil on a calibrated scale.

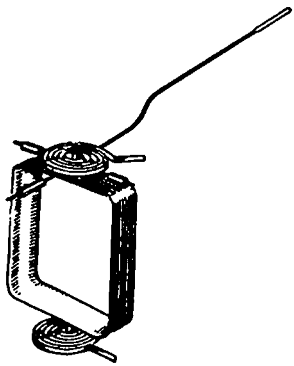


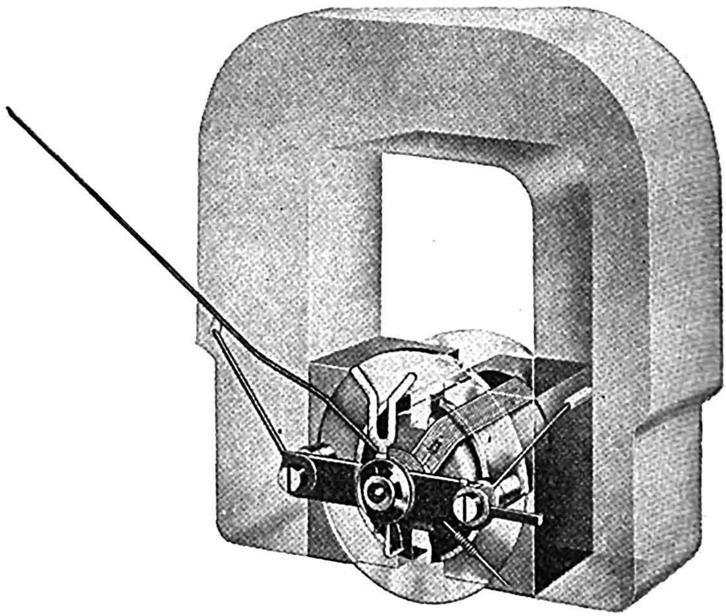
Fig. 2–32. The Moving-coil Assembly.

When a current flows in the coil, forces are exerted that cause the coil to rotate in the air gap. This rotation is limited by the tension in the hairsprings. The movement of the pointer is proportional to the force exerted on the coil and thus proportional to the current. The normal direct-current meter is, therefore, a current-measuring instrument. By means of auxiliary circuits, however, it may be made to act as either a voltmeter or ohmmeter.

Ammeter. Since only very small magnitudes of currents are necessary to cause the meter to read, it is customary to use very low resistances to carry the main portion of the current. The meter, being connected across this low resistance or shunt, takes a small but fixed proportion of the main current. By using shunts of varying resistances the same meter can be used for a number of different current ranges.

Voltmeter. If a resistance is placed in series with the meter, the voltage of a circuit can be determined by measuring the current flowing through the resistance. Additional voltage ranges may be obtained by simply adding resistances in series.

Ohmmeter. If a small battery is located in the meter case and a resistance is connected in series of sufficient magnitude to limit the meter current to full scale when the terminals are shorted, then when additional resistance is inserted between the ohmmeter terminals the current will be reduced and the pointer on the meter will move down the scale. These



Courtesy Weston Electrical Instrument Corp.

Fig. 2-33. A Phantom View of an Assembled D'Arsonval Type of Meter.

reduced readings may be calibrated in terms of external resistance; such a meter is called an ohmmeter.

Example: One of the common types of meters will give full-scale deflection with one milliampere flowing and with one volt impressed. This meter would have an internal resistance of 1,000 ohms. Without other resistance elements, it would be capable of acting as a 1-v voltmeter or a 1-ma ammeter.

Example: Added ammeter ranges. If a shunt of 111 ohms is connected across the meter it will now read full scale with 10 ma. Also, if a shunt of one ohm is connected across the meter it will act as an ammeter with one ampere full scale.

Example: Added voltage ranges. If a resistance of 9,000 ohms is connected in series with the meter the total resistance is 10,000 ohms and the instrument will act as a 10-v voltmeter. If the series resistance were increased to 99,000 ohms, then it would be a 100-v voltmeter.

Example: Ohmmeter connections. If an additional resistance of 2,000 ohms is used and a battery voltage of 3 v, then with the external leads shorted the meter will read full scale. If, however, a resistance of 3,000 ohms is connected to the external leads, the meter will read only half scale and that scale reading can be indicated as 3,000 ohms. Series and parallel resistances are added to change the scale of readings. Instruments of this general type are extensively used in testing radio equipment.

Practical Circuit Problems. Several practical aspects of direct-current circuits have been reserved until the last of the chapter so that advantage could be taken of information covered in the previous material. One of these is the matter of rating of resistors and the importance of ratings to their practical use. Many radio resistors of large magnitude so far as resistance is concerned will carry only very small amounts of current. In general, they are rated in the watts that they can dissipate without overheating. Since power is E^2/R, and since the voltage across a resistance is usually known or can be measured, it is not difficult to determine whether or not the actual power will be below the safe operating rating of the manufacturer. Many of the circuit elements of radio receivers can be of 1-watt rating or even below, but in other places considerable capacity is required.

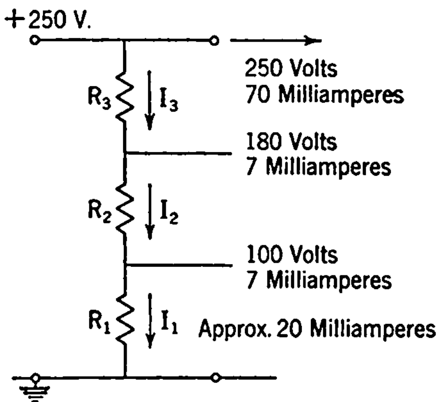


FIG. 2-34. A Voltage Divider.

Power ratings are particularly important on voltage dividers because here, if anywhere in the set, excessive heating of resistors may be anticipated. A rather typical circuit for a voltage divider is shown in Fig. 2-34 and analyzed in the illustrative example. The power requirements for the various tubes are specified and the required resistances are determined.

Example: For purposes of circuit stability a continuous "bleeder" current of about 20 ma is desirable and is assumed as I_1 flowing through R_1. The magnitude of the resistance and power rating of R_1 is then determined.

$$R_1 = \frac{E_1}{I_1} = \frac{100}{.020} = 5{,}000 \text{ ohms.} \qquad P = EI = 100 \times .02 = 2 \text{ watts.}$$

It is then required to determine R_2. Since the current in R_2 consists not only of the bleeder current but also of the 7 ma being supplied at 100 v, the current must be

$$I_2 = .020 + .07 = .027 \text{ amp;}$$
$$E_2 = 180 - 100 = 80 \text{ v;}$$
$$R_2 = \frac{E_2}{I_2} = \frac{80}{.027} = 3{,}000 \text{ ohms;}$$
$$P = EI = 80 \times .027 = 2.2 \text{ watts.}$$

Finally, R_3 is determined in a similar manner.

$$I_3 = .027 + .07 = .034 \text{ amp;}$$
$$E_3 = 250 - 180 = 70 \text{ v;}$$
$$R_3 = \frac{E_3}{I_3} = \frac{.70}{.034} = 2{,}000 \text{ ohms;}$$
$$P = 70 \times .034 = 2.5 \text{ watts.}$$

Attention should be called to the fact that accurate calculations were not made since standard resistances must be used.

Caution on Use of Voltmeter. Since many of the resistances used in a radio set are of very large values, it must be realized that placing a voltmeter in parallel with them may appreciably change the voltage across them. In many of the so-called universal meters the current required is very small, being about 50 μa for full-scale deflection. On one of these a scale of 500 will have a series resistance of 10 megohms. A meter of this type will probably give satisfactory results, but a meter that required 10 ma for full-scale deflection would cause erroneous results. If the resistance of the voltmeter is known it is usually possible, by means of an analysis of the circuit, to determine its effect.

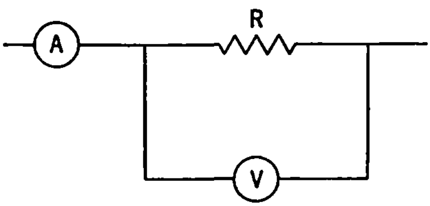


Fig. 2-35. Determination of Resistance by Ammeter and Voltmeter.

Example. If a resistance is to be determined by the use of an ammeter and voltmeter and it is known that the resistance of the voltmeter is 100,000 ohms, readings might be obtained with the connection of Fig. 2-35 as follows: ammeter reading, 1.4 ma; voltmeter reading, 100 v. The current in the voltmeter is, by Ohm's Law, 1 ma. Since the current in the milliammeter includes this current, the actual flow in the unknown resistance is 0.4 ma and the resistance is

$$R = \frac{E}{I} = \frac{100}{0.4} = 250{,}000 \text{ ohms.}$$

Answers to Exercises

2–1. (a) 7 ohms; (b) 6.33 ohms.

2–2. 105 ohms; 31.5 v.

2–3. 218,000 ohms.

2–4. 7 amp; 14.28 ohms.

2–5. 5.64 amp; 73.2%.

2–6. 2,308 ohms; 947 ohms.

2–7. 13.29 amp; 8.2%.

2–8. 6.25 amp; 4 amp.

2–9. 12.5 amp; 3.125 amp.

2–10. $E_{ab} = 39.3$ v; $E_{bc} = 180.7$ v.

2–11. $E_{ab} = 70.9$ v.

2–12. $E_1 = 119.7$ v; $E_2 = 118.6$ v; $E_3 = 118.45$ v.

CHAPTER 3

A.C. Circuits

Alternating Current and Voltage. In Chapter 2 consideration was given to electric circuits in which both the current and voltage were continuous. Although D.C. circuits are very important in radio equipment, an equally important and somewhat more complex system is that known as the alternating-current system. An alternating current or voltage is defined as a current or voltage in which the direction changes periodically. In other words, the current flow or electron drift is first in one direction in the circuit and then in the other, this reversal occurring at regular intervals.

The frequency with which a complete change occurs may be 60 times a second (as in the case of electric power supplied to most residences), from 20 to 10,000 times per second (for most voice and music waves in radio), or up to millions of times per second (as in the case of the radio signals that are now used for communications and other signal purposes).

The low frequencies used for power and the very high frequencies used for radio signals are almost entirely of sinusoidal character; sinusoidal variations of currents and voltages will therefore be the main topic of discussion in this chapter. The irregular currents used to reproduce music and speech must be given special treatment, reference to which will be made at the end of the chapter.

Representation of Sinusoidal Waves. Since most of the theory of radio circuits deals with currents that surge back and forth in a manner known as sine-wave variation, it is desirable to obtain as thorough an understanding of the sine wave as possible.

In Chapter 1 it was shown that the sine function could be obtained from a rotating radius similar to the one in Fig. 3–1. This concept is so important that it is repeated for emphasis. The line *OR*, of unit length, may be used to generate the curve on the right; in this curve the angle is laid off in terms of radians along the horizontal axis and the projection of the unit radius on the vertical axis is the sine of the angle. The arrow is placed on the end of the radius simply to mark it specifically as the end. Such a rotating line is sometimes called a radius vector. If the current in a wire pulsates back and forth so that its magnitude varies with time according to the sine wave, the diagram of Fig. 3–1 can be used to represent this current. If, for instance, it is known that the maximum value of the current is 10 amp and that it changes from zero to maximum positive to maximum negative and back to zero again in $\frac{1}{60}$ sec, then it is likewise known that the magnitude of the rotating radius is 10 and that it must complete one revolu-

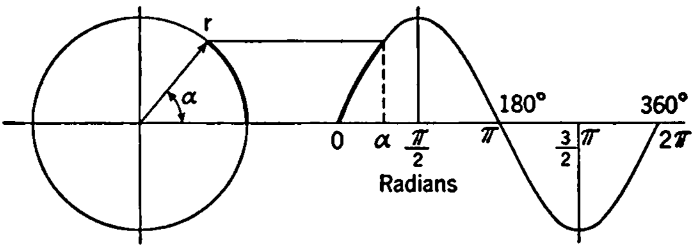


Fig. 3–1. A Rotating Radius Vector May Be Used to Produce a Sine Wave.

tion in $\frac{1}{60}$ sec. This is shown in Fig. 3–2, which will be recognized as a duplicate of Fig. 3–1 except that new scales are introduced. Here the rotating radius is labeled as 10 amp and it is rotating at a speed of $2\pi \times 60$ radians per second in order to complete the circle of 2π radians in $\frac{1}{60}$ sec. The sine-wave section of the diagram is plotted in seconds along the horizontal axis; it could just as well be labeled in radians, remembering that

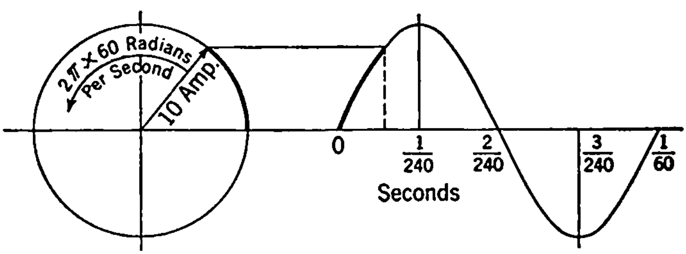


Fig. 3–2. A Rotating Radius Vector and Sine Wave Show the Variation of Current in a Conductor Carrying Alternating Current.

the 2π radians are completed in $\frac{1}{60}$ sec. The number of these complete reversals or cycles occurring in one second is called the *frequency*. This wave would be known as a 60-cycle wave or an alternating current with a frequency of 60. This is the standard commercial and residential power frequency in the United States.

Adding Alternating Currents. It is not possible to add alternating currents in the same way that direct currents are added because they do not always reach their maximum values at the same time in different branches of the circuit. To illustrate this difference, let it be assumed that an A.C. circuit exists such that a current of 5 amp maximum is added to the previous current of 10 amp maximum. The 5-amp current is also of 60 cycles but comes to its maximum 60° or $\pi/3$ radians after the 10-amp current. In Fig. 3–3 the radius vector for each current is shown, with the 5-amp radius lagging 60° behind the 10-amp radius. The instantaneous current at each instant and the instantaneous summation of the two currents are also shown by means of the sine waves. It is seen that this addition of instantaneous currents gives another sine wave. This additive behavior is one of the many interesting and convenient characteristics of sine waves. Another of these sine-wave characteristics is that if a parallelogram is completed using the original radius vectors as two sides, the diagonal will be a new radius vector which will generate the wave of the instantaneous sum of the currents.

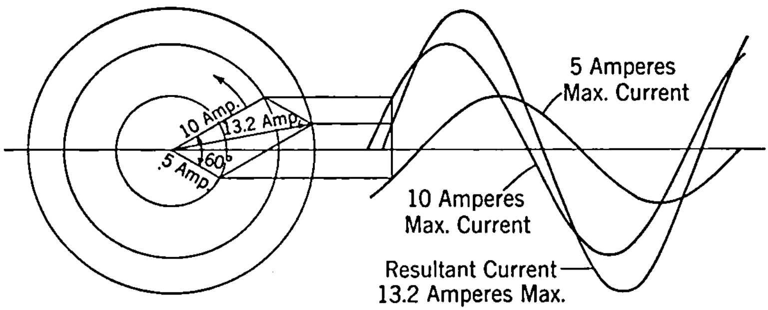


FIG. 3–3. Addition of Alternating Currents.

It is seen from the foregoing that either the sine wave or the rotating radius vector may be used to represent an alternating-current quantity. Sometimes one and sometimes the other is more convenient or more effective in showing the characteristics that are to be studied. Hereafter the method will be used which seems to be most effective, and only occasionally will both representations be used on the same problem. Both the radius-vector and wave type of diagrams may be used for A.C. voltages as well as for currents. Both types are particularly useful in showing the magnitude and phase relationships of currents and voltages when both are included in the same diagram. It is common practice in such a diagram to use one scale for voltages and a different scale for currents.

A.C. Circuit with Resistance. When a sinusoidal A.C. voltage is impressed on a circuit of resistance only, such as a lamp or a heater, the current will be sinusoidal and the current maximum will occur at the instant of maximum voltage. Thus it is observed that Ohm's Law holds true instantaneously in A C. as well as in D.C. circuits. The diagram in Fig. 3–4 shows this situation. In addition, a curve showing the instantaneous product of current and voltage is given. This product is the instantaneous power and is seen to vary from zero to maximum and back again twice in every cycle.

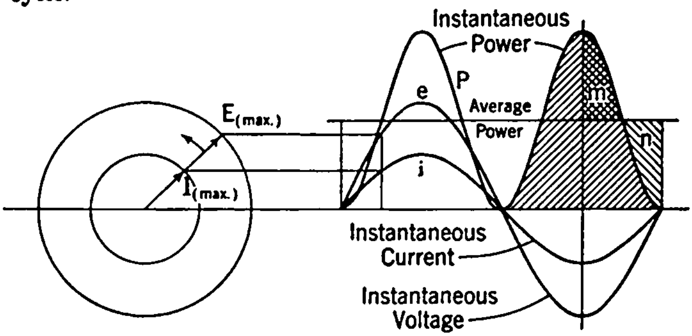


Fig. 3–4. Current, Voltage, and Power in a Resistive Circuit.

Peak, Average, and R.M.S. Values of A.C. Waves. In the discussion of alternating-current waves, so far, the magnitude was specified in terms of the maximum or peak value of the wave. This is one way of stating A.C. magnitudes. This method may be rather confusing or misleading, however, in dealing with average power output since, in a resistance circuit such as is shown above, it is only the peak power that is equal to the product of the maximum current and maximum voltage:

$$P_{peak} = E_{max} I_{max}.$$

The average power, or heating effect, is the important consideration in many studies of A.C. circuits. In Fig. 3–4 the instantaneous power is represented by a sine wave of double frequency. A horizontal line drawn through the center of this sine wave will show the average power. Since this axis of the wave is equally distant from the peaks it follows that the axis is one half of the peak power. This ratio is also verified graphically by the observation that the double cross-hatched section labeled *m* will just fit into the section labeled *n*. The area under the power wave is a measure of the energy, so the average power is a straight line which has the same area beneath it as the power wave. This is seen to be the horizontal line drawn at one half the peak power. Average power for a resistance circuit may, therefore, be specified as,

$$P_{average} = \frac{E_{max} I_{max}}{2} = \frac{E_{max}}{\sqrt{2}} \times \frac{I_{max}}{\sqrt{2}},$$

since $2 = \sqrt{2} \times \sqrt{2}$. The values represented by $\frac{E_{max}}{\sqrt{2}}$ and $\frac{I_{max}}{\sqrt{2}}$ are values of current and voltage that will give average power when multiplied together and are called, therefore, *effective* values of current or voltage. It will be observed that the current having 10 amp maximum would have an effective value of

$$I_{\text{effective}} = \frac{I_{\max}}{\sqrt{2}} = \frac{10}{\sqrt{2}} = 7.07 \text{ amp.}$$

Effective and R.M.S. are terms used interchangeably, and since R.M.S. is extensively used in the literature, the following explanation is given. R.M.S. means *root mean square;* more specifically, the square root of the average of the squares. This concept gives rise to an alternate development for effective values. Since it is desirable to have $P = I^2R$ in A.C. as well as D.C., the I for A.C. will have to be such as to give the same average power as the D.C. The instantaneous power is i^2R, where i is the instantaneous

TABLE 3–1

DETERMINATION OF AVERAGE AND R.M.S. VALUES OF A SINE CURRENT HAVING 10 AMPERES MAXIMUM VALUE

Degrees	i (amperes)	i^2
10	1.74	3.03
20	3.42	11.79
30	5.00	25.00
40	6.43	41.35
50	7.66	58.67
60	8.66	75.00
70	9.40	88.36
80	9.86	97.22
90	10.00	100.00
100	9.86	97.22
110	9.40	88.36
120	8.66	75.00
130	7.66	58.67
140	6.43	41.35
150	5.00	25.00
160	3.42	11.79
170	1.74	3.03
180	0.00	0.00
Sum........	114.34	900.8
Average.....	6.36	50.0

Equivalent direct current $= \sqrt{50.0} = 7.07$ amp.

current, so if the values of i^2 are averaged over a half period, the resulting magnitude must be equal to I^2. This average is given in Table 3–1 for the 10-amp maximum current referred to previously. It is found that the average of the i^2 is 50 and the square root is 7.07 amp. This conclusion confirms the previous analysis and indicates the reason for the use of the abbreviation R.M.S. The effective values are used so extensively in engineering work that unless otherwise specified it is customary to give the value of current or voltage in effective or R.M.S. amperes or volts. Electric meters are likewise calibrated in R.M.S. values unless otherwise marked.

In a few applications, in power supplies and rectifiers for example, the average electron drift or flow is the real measure of effectiveness of the current and for these the true average of the half wave is used. As is shown in Table 3–1, 6.36 amp D.C. will give a cumulative electron drift equivalent to a rectified alternating current having a maximum value of 10 amp. The average alternating current is, therefore, specified as 0.636 of the maximum value of a sine wave.

Rate of Change of Current in a Sine Wave. An important characteristic of A.C. circuits depends upon the rate at which the current is increasing or decreasing. Since the vertical change in the sine wave represents the change in current and the horizontal change represents the change in time, the current change divided by the corresponding time change is by definition the rate of change of the current. This ratio is also the slope or steepness of the sine curve. In Fig. 3–5 is shown the same current of

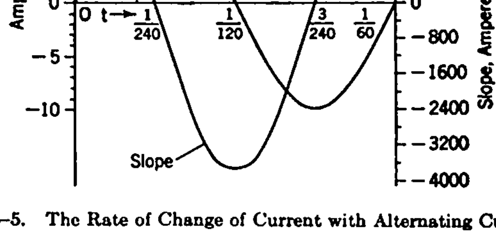


FIG. 3–5. The Rate of Change of Current with Alternating Current.

10-amp maximum that has been studied before. At the point *a* a tangent is drawn and the slope can be determined by dividing the distance i' in amperes by the distance t' in seconds. If this same procedure is followed for a number of points on the current curve and the results plotted against the corresponding time, a curve will be obtained for the slope. This curve will be a sine wave displaced by 90°. This displacement is another of the significant characteristics of the sine curve. It will be observed that the maximum value is at the time the current is passing through zero, which means that it is a sine wave that leads or precedes the current wave by a time-phase angle of 90° or $\pi/2$ radians. The maximum value of the curve of the rate of change of current is $2\pi f$ times the maximum of the current curve where f is the frequency. This maximum location may be explained by the fact that the rate of change when the current is going through zero is represented by the terminus of the radius vector moving at full velocity. Since its angular velocity is $2\pi f$ radians or f complete revolutions a second, its velocity or rate of change will be $2\pi f$ times its length, and the length is equal to the maximum value of the current. This relationship is very important, as it is the basis for the determination of the reactance of A.C. circuits.

Sine Waves in Nature. The fact that the sine wave has a rate of change which is another sine wave displaced by 90° is one of the chief reasons why it is found so extensively in natural phenomena. When a pebble is dropped in still water, the ripples are sine waves. A tuning fork gives out a sound or variation in the atmosphere that varies sinusoidally with time. The pendulum of a clock is shifting energy back and forth from kinetic to potential in its sinusoidal movement. If a hack-saw blade is clamped in a vise, a weight placed on the end of it oscillates with a sinusoidal variation that shifts energy back and forth from kinetic to spring or strain energy. Likewise, the oscillations in the tuned circuit of a radio set are sinusoidal in character. It is seen, then, that electric oscillations are just one of many "natural" sinusoidal variations. The manner in which these oscillations occur in electric circuits is the subject for study in the present chapter on A.C. circuits.

Inductance. A coil of wire has a property called inductance that is one of the foundations of all radio circuits. Inductance will be studied by

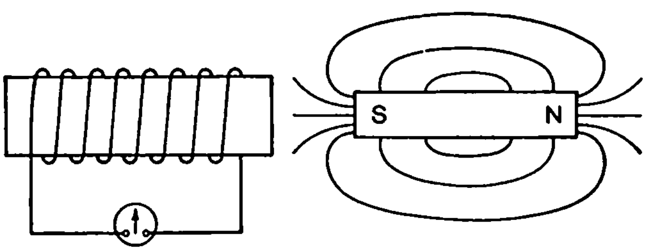


Fig. 3–6. A Coil and a Bar Magnet.

first observing, as did the scientists of a hundred years ago, some simple experiments. Suppose a wire is wound on a cardboard tube as shown in Fig. 3–6. The ends of the wire will be connected to a galvanometer or sensitive meter. A permanent bar magnet will also be needed as part of the equipment.

When the bar magnet is suddenly shoved into the coil as shown in Fig. 3-7, the galvanometer pointer is observed to swing to the right, which indicates that a voltage has been generated in the coil and a current forced around the circuit. When the magnet is permitted to remain at rest in the coil, as in Fig. 3-8, no current or voltage is observed on the galvanometer. When the bar magnet is then suddenly pulled out of the coil, as shown in Fig. 3-9, the galvanometer pointer swings to the left, which indicates that a voltage has again been generated. This time the direction of the current has been reversed, since the galvanometer is deflected in the opposite direction.

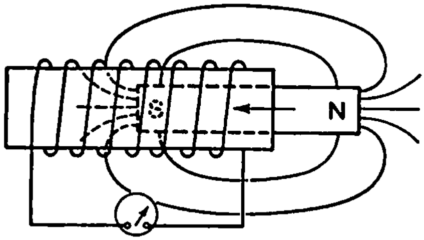


Fig. 3-7. The Bar Magnet Being Inserted in the Coil.

Lenz's Law. From these experiments several conclusions may be drawn. First, relative motion between a coil and a magnetic field produces a voltage. Also, the current reverses when the relative motion is reversed. If the direction of current flow in the coil was determined, it would be found that it was in such a direction as to produce a force which opposed the motion of the magnet. This observation is of such extensive use that it has the status of a physical law, known as *Lenz's Law.* It may be stated in its more generalized form as follows: *When a voltage is induced in a coil as a result of any variation of the magnetic field with respect to the coil, the voltage is in such a direction that the current which results tends to prevent the change which causes the voltage.*

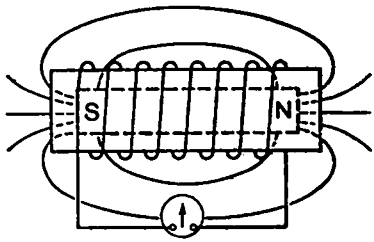


Fig. 3-8. The Bar Magnet at Rest in the Coil.

This same phenomenon was observed in Chapter 2 when the force on the generator conductor was found to be in opposition to the direction of the motion. Reference will be made to this law many times in the following pages.

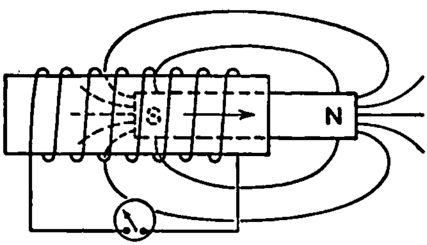


Fig. 3-9. The Bar Magnet Being Withdrawn from the Coil.

Magnitude of Induced Voltages. If the observations in connection with Figs. 3-6 to 3-9 were carefully made it would be noted that the swing of the galvanometer pointer was proportional to the rate at which the magnet was moved with respect to the coil. In other words, the voltage induced in the coil is proportional to the *rate of change* of the flux.

If two coils had been used, one having half the number of turns in the other, it would have been found that the galvanometer pointer would have moved only half as far for the coil with half the number of turns. This would occur, of course, only if the same rate of change of field were maintained. The voltage induced in a coil, then, is proportional to the *rate of change* of field flux and to the number of turns in the coil.

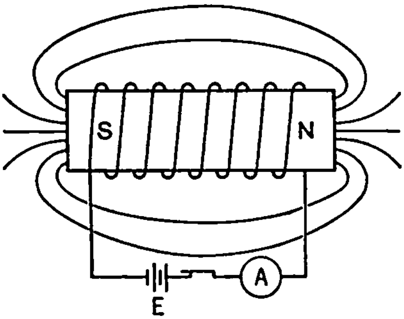


Fig. 3-10. The Magnetic Field of a Coil Carrying Current.

The galvanometer in series with the coil may now be replaced by a battery, an ammeter, and a switch. When the switch is closed, current flows and a magnetic field is set up as shown in Fig. 3-10. This is seen to be a field that is identical with that of Fig. 3-8 and it would be reasonable to expect that in setting up this magnetic field, opposing reactions would occur similar to those observed when the bar magnet was inserted into the coil. That they do occur may

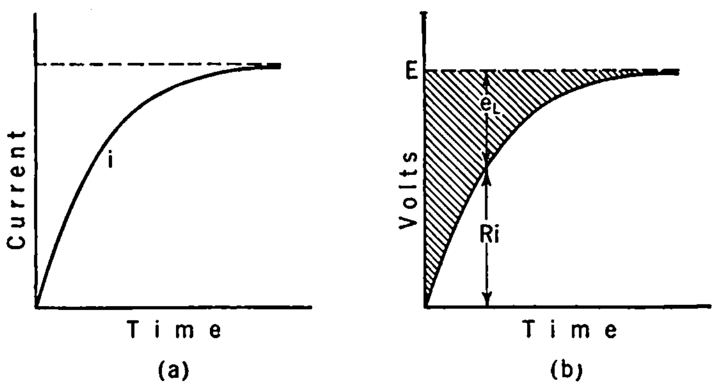


Fig. 3-11. (a) Current Rise in an Inductive Circuit. (b) Voltage Distribution in an Inductive Circuit.

be observed by inserting into the circuit an ammeter with very little inertia which will respond almost instantaneously to the current flow. Such an ammeter shows that the current does not suddenly jump to the value determined by the resistance but rises at a very definite rate and gradually approaches the final value determined by the resistance. This behavior is shown in Fig. 3–11a. In a coil of only a few turns, as indicated in the drawing, the total time indicated would be less than .01 sec, but in some large generators this time might be as much as several seconds. In any case the instruments confirm the expectations and show that a reaction does take place. If the *IR* drop is plotted against time it is found that there is a definite difference between this drop and the impressed voltage. This difference is indicated by the shaded area in Fig. 3–11b and is called the *voltage of self-induction.* It is labeled e_L because e is the symbol for instantaneous voltage and L is the symbol for inductance.

A further analysis of the results shown in Fig. 3–11 shows that the voltage of self-induction, or the *counter voltage* as it is sometimes called, is proportional to the *rate at which the current is increasing.* This relation should have been expected because it was learned in the experiment with the bar magnet that the voltage was proportional to the *rate* at which the magnet was inserted in the coil. Since the flux is directly dependent upon the current, it would be natural to expect, according to *Lenz's Law,* that the opposing voltage or voltage of self-induction would depend upon the *rate* of current increase. This conclusion is important, being the basis of all our ideas regarding the inductive reactance of coils carrying alternating current.

Energy Stored in the Magnetic Field. It may be well at this time to study some of the energy relations in a coil and its associated magnetic field. The rate at which energy was put into the coil in Fig. 3–11 is given by

$$P = Ei.$$

But since

$$E = iR + e_L,$$

then

$$P = i(iR + e_L) = e_Li + i^2R$$

and

$$e_Li = P - i^2R.$$

Since i^2R is the power that is lost in heat in the coil resistance, it follows that the remainder of the power, represented by e_Li, is not accounted for and must be presumed to be stored in the magnetic field.

Experiments show that this energy is stored in the magnetic field in very much the same way that kinetic energy is stored in a heavy flywheel that is rotating. A very common experiment used to verify this is shown in Fig. 3–12. A transformer with a large amount of inductance is connected across a battery and a light is connected in parallel with it. The light should be of somewhat higher voltage than the battery so that it will not light or will glow only slightly when the switch is closed. When the switch is opened, however, the light will flash up bright for a moment and then go out. The energy that caused the light to flash was supplied by the magnetic field as it endeavored to maintain itself, by forcing a current around the coil in the same direction that it has been flowing. The only path available for this current was through the lamp and it was, therefore, caused to burn brightly for a moment.

It may be said that a magnetic field tends to oppose any change in its condition by generating (due to its change) a voltage in the wires surrounding it that causes current (in a closed circuit) to flow in a direction opposing the change. This leads to the definition of unit inductance:

A coil of wire has an inductance of one henry if a current change of one ampere per second will cause a pressure of one volt to be set up in the coil.

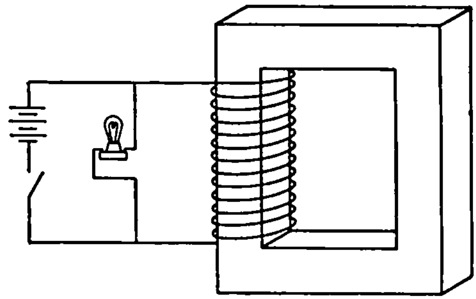


FIG. 3-12. Energy Is Stored in the Magnetic Field of a Transformer.

MAGNITUDE OF INDUCTANCE. The magnitude of the inductance of a coil is a physical characteristic of the coil and of the magnetic circuit, just as the inertia of a flywheel is a physical characteristic of the size, shape, and material used in its construction. If additional turns of wire are placed on the coil without appreciably changing any of the other dimensions, it will be observed that the magnetic field will be increased in proportion to the increased number of turns. The magnetic field also reacts on the increased number of turns. The induced voltage will, therefore, be due to an increased magnetic field acting on an increased number of turns, so the magnitude of the voltage will be proportional to the square of the number of turns on the coil. When it is desired to change the inductance of a coil having a known number of turns, a close approximation may be made by using the relationship that

$$L = KT^2$$

where K is a constant and T is the number of turns. If the number of turns is doubled, the inductance is increased to four times its original value, and if the number of turns is made three times as large, the inductance is increased to nine times.

Where very large values of inductance are desired, and where the frequency is not too great, it is customary to place the coil on an iron core. In Chapter 2 it was learned that iron would pass magnetic flux of several thousand times the amount that would flow in air. Therefore, inserting an iron core may increase the inductance a thousand times. Since the amount of flux in iron is not directly proportional to the current in the coil, as shown in the magnetizing curve of Fig. 2–27, it is common practice to insert an air gap in the magnetic circuit so that the flux may be maintained more nearly proportional to the current. This arrangement will permit the inductance to be quite constant over a reasonable variation of load current. Since there is a loss in the iron for every cycle of alternating current, this loss becomes excessive at high frequencies, and ordinary iron cannot be used. It is, therefore, unusual for ordinary iron cores to be used above 10,000 cycles, but iron-dust cores are often used at higher frequencies.

Inductive Reactance. If an alternating current is flowing through a coil of inductance L, the reactance voltage may be studied with the aid of the diagram in Fig. 3–13. Here the instantaneous current is represented by

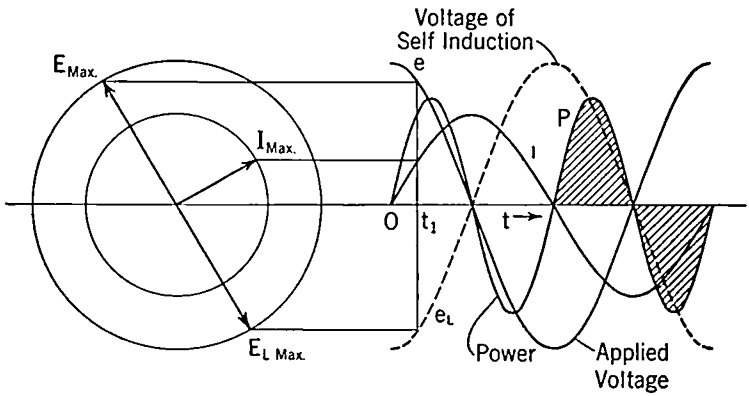


FIG. 3–13. Current, Voltage, and Power in an Inductive Circuit.

both sine wave and radius vector. From Fig. 3–5 it was found that the rate of change of a sine wave of current was another sine wave 90° ahead of the original current wave. When the henry was defined as the unit of inductance, the magnitude of the instantaneous voltage of self-induction was specified as the inductance of the coil in henries multiplied by the rate of change of current in amperes per second. The magnitude of the voltage of self-induction at any instant is obtained, therefore, by multiplying the inductance by the rate of change of current. In the analysis of the rate of change of current in Fig. 3–5, the maximum rate of change was $2\pi f I_{max}$. The maximum voltage of self-inductance will, therefore, be

$$E_{L(max)} = 2\pi f I_{max} L$$

and it will occur at the time the current is passing through zero. The direction in which this voltage will act is important. It was learned from the earlier analysis of coils and bar magnets that a voltage would be induced in such a direction as to cause a current that would oppose the change in flux. At the time indicated by the radius vectors of Fig. 3–13 the current is increasing, so a current that would oppose the change in flux would be a current in the negative direction and the voltage of self-inductance e_L is, therefore, negative. The instantaneous value of the voltage of self-induction is shown in Fig. 3–13 as e_L and is represented by a sine wave having its negative maximum, equal to $2\pi f I_{max} L$, at the time the current is passing through zero from negative to positive. Since this wave is just the opposite of the wave of the rate of change of current, it is customary to write the equation for the voltage of self-inductance as equal to the negative of the rate of change of current times the inductance.

$$E_{L(max)} = -(2\pi f I_{max})L$$

This equation is usually written

$$E_{L(max)} = -2\pi f L I_{max}$$

and the quantity $2\pi f L$ is called the *inductive reactance* or many times just the *reactance.* The usual symbol for this term is X or X_L, the subscript L being used to indicate that it is the reactance caused by an inductance coil. This reactance is expressed in ohms just as was resistance, because the product of the current and the reactance gives the magnitude of the reactance voltage.

Exercise 3–1. What is the rate of change of current in a coil carrying a maximum current of 100 ma at a frequency of 2,000 cycles? If the coil has an inductance of 50 millihenries, what is the voltage of self-induction?

Reactance Voltage. Although the voltage of self-induction is extremely important in understanding the mechanism of the voltage-current relationships in an inductance coil, it is itself seldom used in circuit analysis. Usually the designer is interested in how much current will flow when a certain voltage is impressed on the circuit, and the emphasis is therefore upon the impressed voltage and not upon the internal opposing voltage. If the resistance of the coil used in the present study is negligibly small and if an A.C. voltage is impressed upon the coil, the alternating current flowing will increase until the impressed voltage is neutralized by the voltage of self-induction. Under these circumstances the voltage of self-induction is equal at every instant to the impressed voltage, so that the impressed voltage wave e may be drawn on the diagram of Fig. 3–13 as equal and opposite to the e_L wave. The assumption that the resistance is negligibly small may seem somewhat idealistic to the beginning student, but many of the most effective methods of measurement at high frequency are based on this assumption.

Exercise 3–2. What voltage having a frequency of 750 kc would be required to force a maximum of 10 ma through an inductance of 0.80 millihenries?

Radius-vector Representation of Sine Waves. The study of the relationship between alternating current and voltage in an inductance coil has centered so far around the sine waves that may be used to represent them. Attention will now be directed to the simplicity with which the radius vectors at the left of the diagram indicate these same relationships. In the first place, it should be remembered that all diagrams of this type represent not stationary, but rotating radius vectors of specified length. In most cases, as in Fig. 3–13, the arrow showing rotation is omitted; but radius vectors should always be considered as rotating at a velocity equal in revolutions per second to the A.C. frequency in cycles per second. The current vector* marked I_{max} represents the current wave and the projection on the vertical axis represents the instantaneous value at time t_1. The vector E_{max} represents the wave of voltage impressed and is 90° ahead of the current. The voltage $E_{L(max)}$ represents the internal voltage of self-induction which opposes or neutralizes the impressed voltage and prevents a further increase in the magnitude of the current in the coil.

Power in an Inductance Coil. Another interesting and important observation can be made from a study of Fig. 3–13. The power curve has been obtained by multiplying the instantaneous values of the current and voltage. It is seen that this curve also is a sine wave, but of double frequency, and that the positive and negative loops are equal. The positive loops indicate that the power source is putting energy into the coil and the negative loops mean that the coil is returning energy to the power source. In other words, the power source is storing energy in the magnetic field during the time that the current is increasing, and the energy of the magnetic field is discharging back into the source during the time that the current is decreasing. This ability of the magnetic field to absorb and return energy gives the inductance its inertia characteristics.

Example: A coil having negligible resistance and an inductance of 15 millihenries is connected to a 120-v 60-cycle source. Determine the current flowing and draw the diagram of vectors.

First, determine the reactance. No mention is made to the contrary, so it must be assumed that 120 v is the effective value. The reactance of the coil is

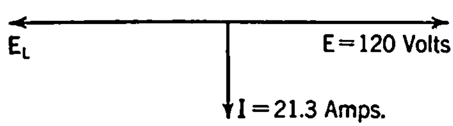


FIG. 3–14. Current and Voltage in an Inductive Circuit.

$$X_L = 2\pi f L$$
$$= 2\pi \times 60 \times 0.015$$
$$= 5.65 \text{ ohms.}$$

* For the sake of brevity, all radius vectors will be referred to, hereafter, as vectors.

Second, determine the current. Since the R.M.S. voltage is equal to the reactance multiplied by the R.M.S. current, it may be written

$$E = X_L I \qquad \text{or} \qquad I = \frac{E}{X_L};$$

$$I = \frac{120}{5.65} = 21.3 \text{ amp.}$$

Note. It is seen that, as yet, no properties have been given to the reactance X_L which specify that the current should lag 90° behind the voltage.

Characteristic of Reactance. The analyses of currents and voltages in an inductance have shown that the current lagged 90° behind the voltage, so that it remains only to set up some arbitrary plan or set of signals that will indicate the relative positions of the current and voltage. This is usually done by drawing the reactance vertically upward on a diagram which has the horizontal axis as the reference. This arrangement is shown in Fig. 3–15. In part (a) is shown the indication of the inductance X_L, 90° ahead of the reference line. In part (b) the current I, which is the reference, has been multiplied by X_L to obtain the IX_L which is the impressed voltage and leads the current by 90° as was demonstrated in the study of Fig. 3–13. In part (c) the quantity $1/X_L$ is shown and it is lagging 90° behind the reference. When the voltage is multiplied by $1/X_L$ the current is obtained, and it is known from the circuit analysis that the current must lag 90° behind the voltage. In part (d) of Fig. 3–15 the voltage is shown as the reference and the current is obtained by multiplying E by $1/X_L$.

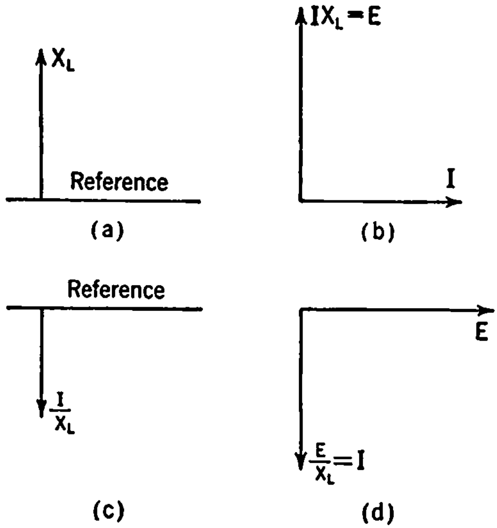


FIG. 3–15. (a) Inductive Reactance. (b) Voltage in an Inductive Circuit with Current as Reference. (c) Reciprocal of Inductive Reactance. (d) Current in an Inductive Circuit with Voltage as Reference.

The Effects of Frequency on Inductive Reactance. Since the reactance is equal to $2\pi fL$, it will increase in direct proportion to the frequency. If appreciable current is to be passed at high frequencies the inductance must be kept quite low; otherwise the reactance would be so high as to block out the current. This blocking is sometimes desirable, and inductances used for this purpose, usually of quite high values, are called "chokes" because they choke off any high-frequency current.

Exercise 3-3. A 60-cycle 120-v line is connected to a coil having an inductance of 0.04 henries. What is the reactance and the current flow? Draw a diagram of vectors showing the reactance, the voltage and the current.

Exercise 3-4. How much current will pass through an inductance of 1.5 millihenries which is used as a choke coil at 600 kc if the impressed voltage is 12?

Resistance and Inductance in Series. The effect of resistance on the relation between alternating current and voltage has been studied, and it was found that the voltage is equal to the product of the current and the resistance as in the case of direct current. The current and voltage are represented, therefore, as sine waves that are at every instant in the relation

$$e = iR,$$

as shown in Fig. 3-4. The two waves are said to be *in phase*, since they pass through zero at the same time and reach a maximum at the same time.

In the case of the inductance coil, however, it was learned that the voltage causing the current is 90° ahead of the current. If a resistance is connected in series with an inductance coil, then the current flowing through the resistance and through the coil will be the same at every instant. The voltage required to maintain this current will be studied with the aid of the combined vector and sine-wave diagram of Fig. 3-16. The current which is used as a reference starts at zero and varies sinusoidally as indicated in the sine-wave diagram. This variation is indicated also in the vector diagram by the vector I drawn horizontally with a magnitude equal to I_{max}. The voltage across the resistance R from a to b is in phase with I and of a magnitude IR shown on the vector diagram. On the sine-wave diagram the voltage variation is indicated by the dashed wave in phase with the current. The voltage across the inductance from b to c is shown on the wave diagram by the wave leading the current by 90° in time phase, and is also drawn dashed. This voltage is likewise shown 90° ahead of the current on the vector diagram, and is labeled IX.

The magnitude of the reactance voltage wave is less than the magnitude of the voltage across the resistance. This means that the resistance of R in ohms is greater than the reactance of L, also in ohms. The voltage from a to c is the instantaneous sum of the voltages across R and across L, and is shown by the solid wave marked e. This wave leads the current wave by an angle θ* which is indicated on the diagram. The sum of the

* θ is the Greek letter theta and is commonly used to indicate an angle.

voltages across the resistance and the inductance is also determined in the vector diagram. The ease with which this voltage is determined is worthy of attention. It is necessary only to complete the parallelogram (rectangle in this case) indicated by IX and IR and the diagonal is the maximum value of the voltage, while the angle between this diagonal and the current is the

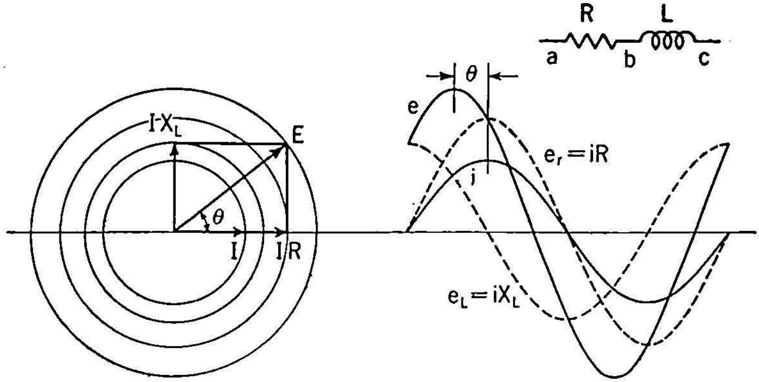


FIG. 3–16. Alternating Current and Voltage with Resistance and Inductance in Series.

angle θ by which the voltage leads the current; or said in another way, the angle by which the current lags behind the voltage. The simplicity of this method of determining the magnitude and phase relationships of currents and voltages in circuits having both resistance and inductance has led to its almost universal adoption for the solution of A.C. circuits.

Impedance and Phase Angle. The procedure described above for determining the relation between the current and voltages in A.C. circuits indicates that a circuit containing both resistance and inductive reactance has certain characteristics or constants that are independent of the current flowing. The circuit characteristic that is used to multiply the current in order to obtain the voltage is

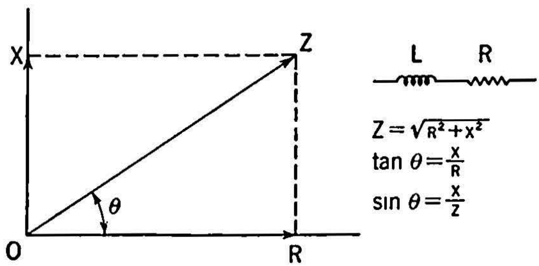


FIG. 3–17. Addition of Resistance and Reactance to Obtain Impedance.

called the *impedance* and the symbol used is Z. The unit by which impedance is measured is the ohm. In Fig. 3–16 the two voltage drops across R and L are expressed as products of current and resistance and of current and reactance. The sum of these two voltages might be expressed as a product of current and the impedance, since the impedance is a special kind of addition of the resistance and reactance. This addition must be made at right angles and such an addition is shown in Fig. 3–17. Here a line R of length equal to the resistance of the circuit is drawn horizontally to the right and a line X_L of length equal to the reactance of the circuit is drawn vertically upward. The impedance is determined by completing the rectangle and drawing the diagonal from the origin. The magnitude may be determined either by measurement or by recognizing that Z is the hypotenuse of a right angle triangle and that, therefore,

$$Z = \sqrt{R^2 + X^2}.$$

The angle θ may be determined either by measurement with a protractor or by the use of trigonometric tables from the relation that

$$\tan \theta = \frac{X}{R}.$$

The impedance is observed to have two characteristics by which it is specified. The first is the magnitude and the second is the angle by which the current lags behind the voltage. Both of these characteristics can be determined from the magnitudes of the resistance and the reactance so that it is quite common to specify an impedance by giving the magnitudes of resistance and reactance. The above method of analysis is satisfactory even though the only circuit resistance is the resistance of the winding of the inductance coil itself.

In determining the current from a given voltage it is necessary to divide the voltage by the impedance:

$$I = \frac{E}{Z}.$$

This operation gives the magnitude of the current and the angle of lag is determined as before. That is, the current always lags as far behind the voltage as the voltage leads the current.

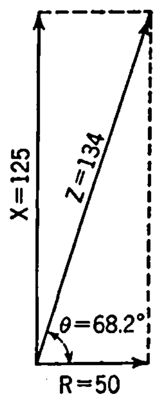


Fig 3–18. Impedance Diagram.

Example: Determine the voltage required to force a current of 20 ma through an inductance coil having a resistance of 50 ohms and an inductance of 10 millihenries. The frequency is 2,000 cycles per second. What is the angle by which the voltage leads the current?

First, determine the reactance.

$$X = 2\pi fL = 2\pi \times 2{,}000 \times \frac{10}{1{,}000} = 40\pi = 125^{+} \text{ ohms.}$$

Second, determine the impedance.

$$Z = \sqrt{R^2 + X^2} = \sqrt{50^2 + 125^2} = \sqrt{2{,}500 + 15{,}625}$$
$$= \sqrt{18{,}125} = 134^{+} \text{ ohms.}$$

$$\tan \theta = \frac{X}{R} = \frac{125}{50} = 2.5.$$

$$\theta = 68.2°.$$

Third, determine the magnitude of the voltage.

$$E = IZ = 0.020 \times 134 = 2.68 \text{ v.}$$

An A.C. voltage of 2.68 is necessary to force 20 ma through the above coil. The voltage will lead the current by a phase angle of 68.2°.

Note: Many students find that a carefully drawn diagram will give results of engineering accuracy much more quickly and with less chance for error than the mathematical calculations. The use of a diagram drawn to scale is always a desirable check on arithmetical calculations even if it is not used for the original solution.

Exercise 3–5. What 60-cycle voltage will be required to produce a current of 20 amp through an inductance coil of 20 millihenries having a resistance of 4 ohms? Determine the phase angle.

The Impedance of a Circuit of Several Elements. The sum of the voltages in a series circuit is not necessarily limited to two elements but may be of any number. In circuits of this type the voltages are added instantaneously, just as was demonstrated in Fig. 3–16. All of these voltage drops may be divided up into parts due to resistance and parts due to reactance. Since this is possible, the impedance of the entire circuit may be obtained by adding the impedances of individual parts. This computation may be done graphically, or it may be done by adding all of the resistances and all of the reactances and from these obtaining the total or equivalent impedance.

Example: Determine the current taken from a 60-cycle 220-v A.C. line when coil A, resistance B and coil C are connected across it in series. Coil A has a resistance of 0.3 ohms and an inductance of 2 millihenries. Resistance B has a magnitude of 1.2 ohms. Coil C has a resistance of 0.7 ohms and an inductance of 5 millihenries.

First, determine the reactance of the coils.

Coil A:

$$X_A = 2\pi fL = 2\pi \times 60 \times 0.002 = 0.24\pi = .754 \text{ ohms.}$$

Coil C:

$$X_C = 2\pi fL = 2\pi \times 60 \times 0.005 = 0.60\pi = 1.88 \text{ ohms.}$$

Second, determine total impedance.

$$X_{total} = X_A + X_C = 2.63 \text{ ohms.}$$

$$R_{total} = R_A + R_B + R_C = 0.3 + 1.2 + 0.7 = 2.2 \text{ ohms.}$$

$$Z_{total} = \sqrt{R^2 + X^2} = \sqrt{2.2^2 + 2.63^2} = \sqrt{4.84 + 6.92}$$

$$= \sqrt{11.76} = 3.43 \text{ ohms.}$$

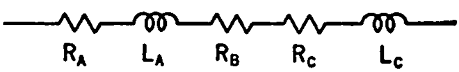
Third, determine current and phase angle.

$$I = \frac{E}{Z} = \frac{220}{3.43} = 64.2 \text{ amp.}$$

$$\tan \theta = \frac{X}{R} = \frac{2.63}{2.2} = 1.19.$$

$$\theta = 50.1°.$$

Exercise 3-6. A 10,000-ohm resistance is in series with an inductance coil having a resistance of 2,000 ohms and an inductance of 10 millihenries. This circuit is connected across a 75-v line having a frequency of 150 kc. Determine the current and the phase angle.



Exercise 3-7. Two coils and a resistance are connected across a 110-v 60-cycle power line. Coil A has a resistance of 3 ohms and an inductance of 15 millihenries. Coil B has a resistance of 7 ohms and an inductance of 5 millihenries. The resistance C has a magnitude of 2 ohms. Determine the current and equivalent impedance.

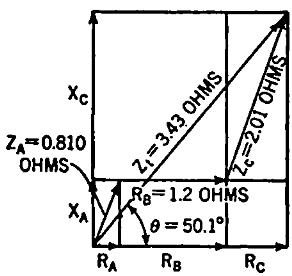


Fig. 3-19. Impedance Diagram; Example pp. 93-94.

Resistances and Inductances in Parallel. When A.C. circuits are connected in parallel the same general type of solution is used as was used for resistances in parallel. That is, the current flowing through each part of the parallel circuit is determined and the total current is the sum of the individual parts. It must be kept clearly in mind that the currents may not be in phase and that their vectors must be added at the proper difference of direction. This procedure was developed in the early part of the chapter and may be reviewed by a study of Fig. 3-3. If it is desired to obtain the equivalent impedance of the parallel circuit,* this calculation may be done by, first, assuming a pressure of 1 v across the circuit and determining the total current flow. The impedance is then found by dividing the voltage by the total current.

* A circuit constant called the *admittance* equal to the reciprocal of the impedance is sometimes used for this purpose. The characteristics of this constant are so likely to cause confusion that the constant is omitted from this analysis. Very little loss is experienced from this omission, as all usual circuit problems can be solved by the methods here presented.

Example: It is desired to know the total current flow, the phase angle of the total current, and the equivalent impedance, if coil C of the preceding illustrative example is connected in parallel with coil A and resistance B in series across a 110-v 60-cycle line.

First, determine the current in coil A and the resistance B using the values obtained in the preceding example:

$$Z_{A+B} = \sqrt{(R_A + R_B)^2 + X_A^2} = \sqrt{1.5^2 + .754^2}$$
$$= \sqrt{2.25 + .57} = \sqrt{2.82} = 1.68 \text{ ohms.}$$
$$\tan \theta_{A+B} = \frac{X}{R} = \frac{.754}{1.50} = 0.503.$$
$$\theta_{A+B} = 26.7°.$$
$$I_{A+B} = \frac{E}{Z} = \frac{110}{1.68} = 65.5 \text{ amp,}$$

which lag 26.7° behind the voltage.

Second, determine the current in coil C.

$$Z_C = \sqrt{R_C^2 + X_C^2} = \sqrt{0.7^2 + 1.88^2}$$
$$= \sqrt{0.49 + 3.54} = \sqrt{4.03} = 2.01 \text{ ohms.}$$
$$\tan \theta_C = \frac{X}{R} = \frac{1.88}{0.7} = 2.68.$$
$$\theta_C = 69.6°.$$
$$I_C = \frac{E}{Z} = \frac{110}{2.01} = 54.7 \text{ amp,}$$

which lag 69.6° behind the voltage.

Third, determine the total current by adding the individual currents.*

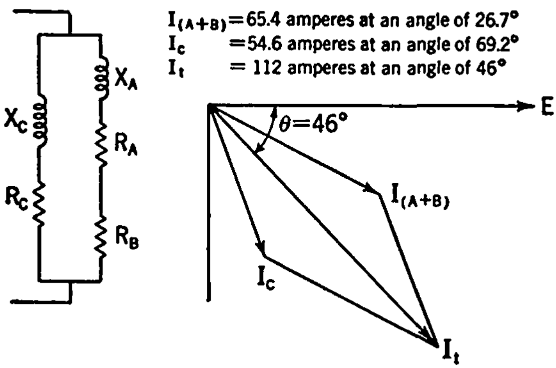


FIG. 3-20. Vector Diagram for Parallel Circuit of Example Above.

* This can be done graphically as in Fig. 3-20 or by determining the portion of the current in phase with the voltage and the portion 90° behind the voltage and adding these portions to find the portions of total current in phase with and 90° behind the voltage. The total current can then be determined from these by the use of right-triangle analysis.

Begin by determining the portion of currents in phase with the voltage.

$$I_{A+B} \cos \theta_{A+B} = 65.4 \cos 26.7° = 65.5 \times .893 = 58.5 \text{ amp.}$$
$$I_C \cos \theta_C = 54.7 \cos 69.6° = 54.7 \times .349 = 19.0 \text{ amp.}$$
$$I_{(\text{in phase})} = 58.5 + 19.0 = 77.5 \text{ amp.}$$

Next, determine the portion of the currents 90° behind the voltage.

$$I_{A+B} \sin \theta_{A+B} = 65.4 \sin 26.7° = 65.5 \times .45 = 29.4 \text{ amp.}$$
$$I_C \sin \theta_C = 54.6 \sin 69.6° = 54.7 \times 0.938 = 51.3 \text{ amp.}$$
$$I_{(90° \text{ lag})} = 29.4 + 51.3 = 80.7 \text{ amp.}$$

From the two individual currents, determine the total current flow.

$$I_{\text{total}} = \sqrt{I^2_{(\text{in phase})} + I^2_{(90° \text{ lag})}}$$
$$= \sqrt{77.5^2 + 80.7^2} = \sqrt{6,006 + 6,510}$$
$$= \sqrt{12,516} = 111.9 \text{ amp.}$$
$$\tan \theta_{\text{total}} = \frac{I_{(90° \text{ lag})}}{I_{(\text{in phase})}} = \frac{80.7}{77.5} = 1.04$$
$$\theta_{\text{total}} = 46.2°$$

Fourth, determine the equivalent impedance.

$$Z_{eq} = \frac{E}{I_{\text{total}}} = \frac{110}{111.9}$$
$$= 0.984 \text{ ohms.}$$
$$R_{eq} = Z \cos \theta = .984 \cos 46.2° = 0.984 \times 0.692$$
$$= 0.682 \text{ ohms.}$$
$$X_{eq} = Z \sin \theta = 0.984 \times 0.722$$
$$= 0.711 \text{ ohms.}$$

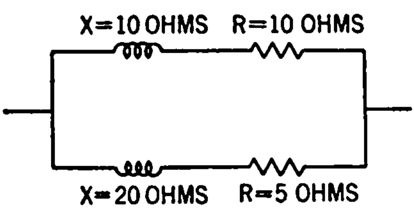


Fig. 3-21. Circuit Diagram for Exercise 3-8.

Exercise 3-8. Determine the equivalent impedance of the circuit shown in Fig. 3-21.

Exercise 3-9. A resistance of 100 ohms is connected in parallel with a reactance coil having an inductance of 25 millihenries and a resistance of 20 ohms. What total current will flow when 40 v at 1,000 cycles are impressed on the circuit?

Exercise 3-10. Three circuits are connected in parallel across a 120,000-cycle 20-v source. Circuit *A* consists of a resistance of 2,000 ohms. Circuit *B* consists of a coil having 100 ohms resistance and an inductance of 5 millihenries. Circuit *C* consists of a coil having a resistance of 50 ohms and an inductance of 1.2 millihenries, and of a resistance of 1,000 ohms. What is the total current flowing and what is the equivalent impedance?

Power and Power Factor. In the early part of the chapter it was learned that the average power for an alternating current and voltage, where the impedance consisted of resistance only, was

$$P_{\text{average}} = EI,$$

where E and I were the effective or R.M.S. values. A little later it was learned that the average power dissipated in an inductance coil of zero or negligible resistance was zero. This fact was shown in the explanation of Fig. 3–13, and it was learned that energy was alternately stored in and recovered from the magnetic field. It would be reasonable to conclude, therefore, that the only power absorbed in an inductance coil would be that which was converted into heat in the resistance of the winding. This conclusion is verified by test; and so the power in a circuit, composed of resistance and inductance, may be specified as

$$P_{\text{average}} = I^2R = (IR)I.$$

Since IR is the component of voltage which is in phase with the current, it may be determined from the total voltage by multiplying by the cosine of the angle of lag. Thus,

$$IR = E \cos \theta,$$

and, therefore,

$$P_{\text{average}} = (IR)I = IE \cos \theta.$$

The factor $\cos \theta$ is called *the power factor*, since it is the factor by which the product of the current and voltage must be multiplied in order to obtain the average power. The angle θ between the current and voltage is often called the *power-factor angle.*

The Electric Condenser. When two conducting plates are placed close to but insulated from each other, they form what is known as a *condenser.* An electric charge can be stored on these plates and will be retained as long as the plates of the condenser are insulated from each other. One of the most fundamental of all electrical concepts is the repelling action of charges of the same polarity and the attractive forces between charges of opposite polarity. If, then, the charged plate having an excess of electrons (or negative charge) is connected to the charged plate having a deficiency of electrons (or positive charge), the repelling and attracting action of the charges will cause large numbers of electrons to flow through the conductor. This flow of electrons constitutes an electric current which will be forced through the conductor against the resistive drop and will, thus, cause heat to be developed. This heat energy was stored in the condenser as potential energy owing to the repelling force of the charge. The pressure or potential due to this repelling force is proportional to the charge. The actual magnitude is dependent upon a proportionality constant that is determined by the size of the condenser plates, the distance between them, and the insulating material. This constant is called the capacitance and is indicated by the symbol C. Expressed mathematically, the relation is

$$Q = EC,$$

Q being measured in coulombs, E in volts, and C in farads.

The unit of capacitance is called the *farad* and may be defined as *that capacitance which will permit the storage of one coulomb of charge at a potential of one volt.*

Dielectric Constant. The above analysis was based entirely upon the effect of the charges within the conductors and so is valid for conductors in a high vacuum, an illustration of which is the interelectrode capacitance of vacuum tubes. Experimental studies show that the same analysis is also valid for conductors in air. When some liquid or solid insulators are placed between the plates of the condenser, however, it is found that the capacitance is considerably increased; this increase permits additional energy storage. The mechanism of this storage is indicated in Fig. 3–22. The positive charge on the upper plate attracts the negative electrons in the molecules of the dielectric. These electrons are bound so tightly to the molecule that they cannot flow as do the electrons in metals; but the force of attraction exercised by the upper plate and the repulsion exercised by the lower plate combine to produce a strain in the molecules of the dielectric similar to the strain in a spring which has been stretched. The extent to which the effect of the repelling action of the charges in the condenser plates can be neutralized by the strained condition of the dielectric is dependent upon the physical characteristics of the dielectric. An index of this ability of a dielectric to change the capacity of a condenser

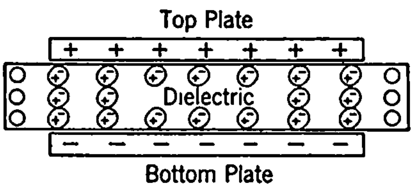


Fig. 3–22. Tension in Dielectric.

TABLE 3-2

Dielectric Constants of Insulating Materials

Material	*Dielectric constant*
Air	1
Glass	6–9
Porcelain	5–7
Steatite	5–6
Mica	6–7
Polystyrene	2.6
Bakelite	5–15

is known as the *dielectric constant,* which may be defined as *the ratio of the capacitance of a condenser with the dielectric being considered to the capacitance of the same condenser if air or a vacuum were used as the insulating medium.* A list of the dielectric constants of several common insulators is given in Table 3–2. It will be noticed that in many cases, varying methods of manufacture or varying quality cause a range of values to be given.

Capacitance of a Condenser. Most of the condensers used in radio are flat-plate condensers and the capacitance of these may be determined from their dimensions by means of the formula

$$C = 2{,}248 \frac{AK}{t} 10^{-16} \text{ f.}$$

In this equation, A is the area in square inches of the dielectric under stress, t is the thickness of the dielectric, and K is the dielectric constant determined from Table 3–2. The farad is such a large unit of capacitance that it is almost never used in radio practice. The microfarad (μf) and the micromicrofarad ($\mu\mu$f) are the units used most extensively. These units must, however, be converted to farads when substituting in equations unless the equations are converted so that the smaller units may be used directly. The value of $1\,\mu\text{f} = 10^{-6}$ f, of $1\,\mu\mu\text{f} = 10^{-12}$ f.

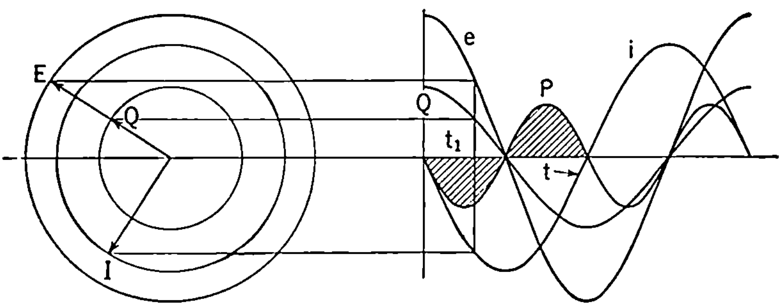


Fig. 3–23. Current, Voltage, and Power in a Capacitive Circuit.

Relation between Voltage and Current in a Condenser. Since the instantaneous charge on a given condenser is proportional to the voltage at any instant, the charge will increase and decrease as the voltage increases and decreases. In Fig. 3–23 the voltage has been plotted as a sine wave and the charge as another sine wave of different magnitude but in phase with the voltage. If the charge on the condenser is continually changing, the conductor by which the connection to the condenser is made must have a flow of electrons to and from the condenser. The current in the circuit is, therefore, the rate at which the charge on the condenser is being increased or decreased. It was learned earlier that the rate of change of a sine wave is another sine wave 90° ahead of the original wave in time phase, so the current in the condenser circuit may be shown in Fig. 3–23 as a sine wave *leading* the wave of charge and voltage by 90° and labeled i. It is seen that this current is opposite in time phase to the current in an inductance, which lags 90° behind the voltage.

The instantaneous power is also plotted in Fig. 3–23, and it is found to be a sine wave of double the voltage frequency, having zero average power as in the case of the inductance coil shown in Fig. 3–13. A careful study, however, will show that for the quarter cycle when the voltage is going from positive maximum to zero, the power is negative in the condenser circuit while it is positive in the circuit containing the inductance coil. In the following quarter cycle it is positive in the condenser circuit and negative in the inductive circuit. This is an important characteristic of these circuit elements, because it permits periodic power transfer from one element to the other and is the basis for the oscillations of tuned circuits used so extensively in radio equipment.

Capacitive Reactance. In the preceding paragraph it was determined qualitatively that the current was equal to the *rate of change* of the charge. It remains to determine the numerical relationships which involve the frequency. In the discussion on the rate of change of a sine wave it was found that the maximum value for the rate of change is $2\pi f$ times the maximum value of the original sine wave. The maximum rate of change of Q is, therefore, $2\pi f$ times Q_{max} and this is likewise the maximum value of the current. Stated mathematically, this relation is

$$I_{\text{max}} = 2\pi f\, Q_{\text{max}}.$$

It is known that the charge is equal to the product of the voltage and capacitance; thus

$$Q_{\text{max}} = CE_{\text{max}}.$$

Substituting this value in the above equation,

$$I_{\text{max}} = (2\pi fC)E_{\text{max}},$$

where $2\pi fC$ is a constant giving the relation between the current and the voltage. Since the reactance of a capacitive circuit is that value by which the current must be multiplied in order to obtain the voltage, the reactance of a condenser may be specified as $1/(2\pi fC)$. In mathematical form this statement is

$$E = IX_C = \frac{1}{2\pi fC},$$

so

$$X_C = \frac{1}{2\pi fC}.$$

Since both the capacitance and frequency are in the denominator, it follows that the impedance is decreased with increase of frequency and is decreased also with an increase in capacitance. This is the opposite of the reactance of an inductance, which increases both with an increase of frequency and of inductance. The voltage impressed across the condenser was found to lag 90° behind the current and so it is customary to draw the reactance as a line vertically downward indicating that if the current radius vector is assumed as the reference being drawn horizontally to the right, then the voltage will be vertically downward.

Resistance and Capacitance in Series. When resistance is connected in series with a condenser, a situation is obtained similar in many ways to that existing in the case of a resistance in series with an inductance. The voltage across the resistance is in phase with the current, *while the voltage across the condenser lags 90° behind the current.* The total voltage, being the sum of the two component voltages, also lags behind the current. This angle of lag, or the angle by which the current leads the voltage, has a tangent, the value of which is X_C/R.

Example: A condenser of 1 μf is connected in series with a 1,000-ohm resistance across a 500-cycle 12-v line. Determine the current, the equivalent impedance, and the power-factor angle.

First, determine the reactance of the condenser.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 500 \times 10^{-6}} = \frac{1000}{\pi} = 318.5 \text{ ohms.}$$

Second, determine the equivalent impedance.

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{1{,}000^2 + 318.5^2}$$
$$= \sqrt{1{,}000{,}000 + 101{,}500}$$
$$= \sqrt{1{,}101{,}500}$$
$$= 1{,}050 \text{ ohms.}$$

$$\tan\theta = \frac{X_C}{R} = \frac{318.5}{1{,}000} = 0.318.$$

$$\theta = 17.7\theta.$$

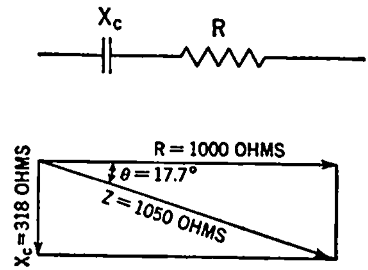


Fig. 3-24. Vector Diagram, Capacitive Circuit.

Third, determine the current.

$$I = \frac{E}{Z} = \frac{12}{1{,}050} = 0.0114 \text{ amp} = 11.4 \text{ ma.}$$

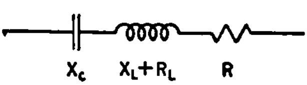
This current will lead the voltage by 17.7°.

Exercise 3-11. Determine the current drawn from a 175-v 600-kc source by a condenser of 0.001 μf in series with a 125-ohm resistance. What is the phase angle of the current? Draw the vector diagram.

Resistance, Inductance, and Capacitance in Series. When resistance, inductance, and capacitance are connected in series, the total voltage is, as before, the instantaneous or vector sum of the voltages across the individual elements of the circuit. When an analysis is made of these voltages, it is observed that the voltage across the condenser is directly opposed to the voltage across the inductance. The voltage across the combination may be less than the voltage across either element by itself. The reactance of such a series circuit is the difference between the reactance of the inductance coil and the reactance of the condenser. Since these two reactances tend to neutralize each other, it has become common practice to assign one a positive value and the other a negative value. Since the inductive reactance is drawn vertically upward with respect to the resistance as reference, it is considered as positive and the capacitive reactance is considered as negative. The power-factor angle θ on this basis is considered positive where the voltage leads the current and negative where the voltage lags behind the current. The methods of making the computations are the same as for the resistance and inductance in series and the addition of impedances may be made graphically or algebraically.

Example: A condenser of 50 μf, an inductance coil having a resistance of 5 ohms and an inductance of 0.08 henries, and a resistance of 6 ohms are connected in series across a 110-v 60-cycle circuit. Determine the equivalent impedance, the current, the power-factor angle, and the voltage across each circuit element.

First, determine the impedance Z.



$$X_{\text{condenser}} = -\frac{1}{2\pi fC} = -\frac{1}{2\pi \times 60 \times 50 \times 10^{-6}}$$
$$= -\frac{10^6}{6{,}000\pi} = -53.0 \text{ ohms.}$$

$$X_{\text{coil}} = 2\pi fL = 2\pi 60 \times .08 = 30.1 \text{ ohms}$$
$$X_{\text{total}} = -53.0 + 30.1 = -22.9 \text{ ohms of capacitive reactance.}$$
$$R_{\text{total}} = 5 + 6 = 11 \text{ ohms.}$$
$$Z_t = \sqrt{R^2 + X^2} = \sqrt{11^2 + 22.9^2}$$
$$= \sqrt{121 + 525} = \sqrt{646} = 25.4 \text{ ohms.}$$

$$\tan\theta = -\frac{X}{R} = -\frac{22.9}{11} = -2.08.$$
$$\theta = -64.4°.$$

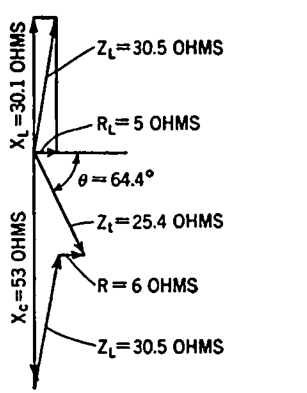


Fig. 3-25. Vector Diagram—Resistance, Inductance, and Capacitance in Series.

Second, determine the current.

$$I = \frac{E}{Z} = \frac{110}{25.4} = 4.33 \text{ amp.}$$

$$\cos\theta = \text{power factor} = 0.43.$$

Third, determine the voltages across each circuit element.

$$E_{\text{condenser}} = IX_C = 4.33 \times 53.0 = 229 \text{ v.}$$

$$Z_{\text{coil}} = \sqrt{R^2 + X^2} = \sqrt{5^2 + 30.1^2} = \sqrt{25 + 906} = \sqrt{931}$$
$$= 30.5 \text{ ohms}$$

$$E_{\text{coil}} = IZ_{\text{coil}} = 4.33 \times 30.5$$
$$= 132 \text{ v}$$

$$E_{\text{resistance}} = 4.33 \times 6$$
$$= 26 \text{ v.}$$

Note: In Fig. 3–25, showing the graphical summation of impedances, it will be noted that each impedance was determined individually. Many engineers prefer to add the resistances and reactances separately in order to obtain the final result as was done in the analysis. Many others prefer to add the impedances as shown by the heavy lines of Fig. 3–26. This method permits a more complete understanding of the voltages between different points on the circuit. It is possible, of course, to make this addition analytically as well as graphically, and in many cases it is done in this manner.

Exercise 3–12. A resistance of 20 ohms is connected in series with an inductance of 100 microhenries and a capacitance of 0.05 μf across a 10-v 100-kc line. Determine the current and equivalent impedance.

Exercise 3–13. An inductance coil having a resistance of 0.3 ohms and an inductance of 1.5 microhenries is connected in series with a 2-ohm resistor and a 10-μf condenser across a 5-v 40-kc line. Determine the difference in phase between the voltage across the coil and the voltage across the resistance. What is the difference in phase between the voltage across the coil and the impressed voltage?

Resistance, Inductance and Capacitance in Parallel. When parallel circuits are encountered which involve capacitive reactance in one circuit and inductive reactance in the other, each circuit is solved by itself and each current is obtained. The total current is then found by the vector addition of the individual currents. This addition again may be done graphically, or by determining the parts of the current in phase with the voltage and the algebraic sum of those parts that are 90° out of phase. The equivalent impedance is determined, as in other parallel circuits, by dividing the voltage by the current. The tangent of the power-factor angle is determined from the ratio of out-of-phase current to in-phase current.

Example: If the condenser and the 6-ohm resistance of the preceding illustrative example are connected in series across the 110-v line and if the coil is also connected across the line to make a parallel circuit, determine the equivalent impedance, the current, and the power-factor angle.

First, reactances may be taken from the preceding example.

$$Z_C = \sqrt{R^2 + X^2} = \sqrt{6^2 + 53^2} = \sqrt{36 + 2{,}810} = \sqrt{2{,}846} = 53.4 \text{ ohms.}$$

$$\tan\theta_C = \frac{-X}{R} = -\frac{53}{6} = -8.83.$$

$$\theta_C = -83.5°.$$

$$Z_L = \sqrt{R^2 + X^2} = \sqrt{5^2 + 30.1^2} + \sqrt{25 + 906} = \sqrt{931} = 30.5 \text{ ohms.}$$

$$\tan\theta_L = \frac{30.1}{5} = 6.02.$$

$$\theta_L = 80.5°.$$

Second, determine parts of currents in phase with voltage and 90° out of phase.

$$I_C = \frac{E}{Z_C} = \frac{110}{53.4} = 2.06 \text{ amp.}$$

$$I_L = \frac{E}{Z_L} = \frac{110}{30.5} = 3.61 \text{ amp.}$$

In-phase current:

$$\begin{aligned} I_{(\text{in phase})} &= I_C \cos\theta_C + I_L \cos\theta_L = 2.06 \cos -83.5° + 3.61 \cos 80.5° \\ &= 2.06 \times 0.112 + 3.61 \times 0.165 = 0.231 + 0.596 \\ &= 0.827 \text{ amp.} \end{aligned}$$

Out-of-phase or quadrature current:

$$\begin{aligned} I_{(\text{out of phase})} &= I_C \sin\theta_C + I_L \sin\theta_L = 2.06 \sin -83.5° + 3.61 \sin 80.5°. \\ &= 2.06(-0.993) + 3.61 \times 0.984 = -2.05 + 3.56 \\ &= 1.51 \text{ amp.} \end{aligned}$$

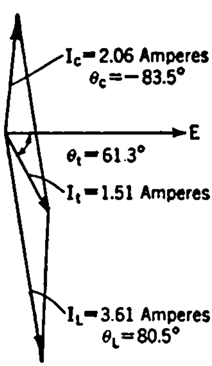


FIG. 3-26.

Third, determine total current and power-factor angle.

$$I_{\text{total}} = \sqrt{0.827^2 + 1.51^2} = \sqrt{.685 + 2.28} = \sqrt{2.96}$$
$$= 1.72 \text{ amp.}$$

$$\tan\theta_t = \frac{1.51}{0.827} = 1.83$$

$$\theta_t = 61.3°.$$

Fourth, determine the equivalent impedance.

$$Z_t = \frac{E}{I_t} = \frac{110}{1.72} = 64 \text{ ohms.}$$

Exercise 3–14. Determine the current drawn from a 275-v 180-kc source by a parallel circuit containing a condenser of 500 $\mu\mu$f in one branch and a resistance of 250 ohms in series with an inductance of 1 millihenry in the other branch. What is the maximum instantaneous power that is put into (a) the condenser, (b) the inductive circuit, and (c) into the combined circuit?

Exercise 3–15. Determine the current if a condenser of 50 $\mu\mu$f is connected in series with two parallel branches of Exercise 3–14 before being connected to the line.

Exercise 3–16. Determine the current drawn from a 750-kc 10-v line by the circuit shown in Fig. 3–27, when the values are as follows:

C_1 = 0.001 μf
L_2 = 100 microhenries
L_3 = 80 microhenries
R_1 = 100 ohms
R_2 = 100 ohms
R_3 = 50 ohms

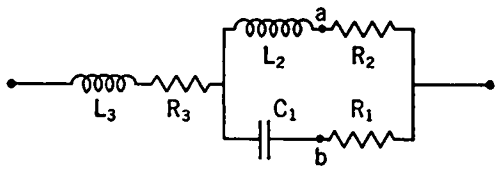


FIG. 3–27. Circuit Diagram for Exercise 3–16.

What is the voltage from a to b in the parallel circuit and what phase angle does it make with the impressed voltage?

General Concepts of Resonance. The term *resonance* comes from the characteristic of some objects which respond to or echo back sound. Usually these objects respond to sounds of certain pitches only and it is explained that they are "tuned" to those pitches. The responses to sounds are of such high frequency and small amplitude that it is difficult to observe them visually. A similar phenomenon is the pendulum clock. Here the pendulum swings back and forth, shifting kinetic energy (motion of the pendulum in the center of the swing) to stored or potential energy (raised position of the pendulum at the end of the swing). The slight impulse from an escapement wheel is all that is needed to keep this pendulum swinging back and forth indefinitely. The amplitude of the vibration or movement is much greater than could be given by a single impulse from the escapement wheel but the regular impulse from the escapement is sufficient to replace the losses of energy in the friction and air resistance of the pendulum.

A similar phenomenon may be produced in electric circuits which have both inductance and capacitance. Both of these circuit elements were found to store energy for one half cycle and to return it on the next half cycle. Since the condenser is storing energy when the inductance is returning energy, it is possible for the condenser and inductance to pass large amounts of energy from one to the other with the outside circuit supplying only the losses. A resonant circuit will operate most satisfactorily when the energy absorbed by the condenser is just equal to the energy which is being discharged by the inductance. Satisfactory operation is enhanced also when the product of the currents in and voltages across the condenser is equal to the product of the same values with reference to the inductance. In a series circuit the current is the same in both circuit elements. Since the voltage across the inductance increases with frequency, and the voltage across the condenser decreases with frequency, there must be some frequency at which the two voltages are equal and at which the energy of the condenser is equal to the energy of the inductance. This is called the *resonant frequency*. The voltage across the inductance as well as that across the condenser may be several times the voltage of the line. This condition was met in the illustrative example of series circuits having both inductance and capacitance. Here the voltage across the condenser was found to be 229, which is more than twice the impressed voltage and indicates that a condition of resonance is being approached in this circuit.

When the coil and condenser are connected in parallel, the voltage across each will be the same, so that the currents must balance for the resonant condition. If the ratio of reactance to resistance is high, as it is in most radio circuits, the resonant frequency will be approximately the same for both series and parallel connections. In fact, in some of the more complicated circuits, it is difficult to tell whether a resonant circuit is connected in series or parallel. In the elementary treatment, however, circuits will be analyzed as either series or parallel resonant circuits.

Series Resonant Circuits. If reference is made to the impedance diagram of Fig. 3–25, it will be seen that the equivalent impedance is considerably less than the impedance of either the condenser or the coil by itself. Furthermore, if a variable condenser were used and if the capacitance of the variable condenser were increased, the capacitive reactance would decrease. This adjustment might be continued until the capacitive reactance would be just equal to the inductive reactance. The circuit impedance would, then, be the resistance only, which in this case is 11 ohms. Under these conditions the circuit would be in resonance.

A series circuit is said to be in resonance when the inductive reactance in the circuit just neutralizes the capacitive reactance so that the equivalent impedance is due entirely to resistance. This condition may be obtained in practice by varying the capacitance of the condenser, as was just indicated, or by varying the inductance of the coil, or by varying the frequency. In any of the conditions above indicated, the resistance of the circuit is assumed to remain constant and the circuit impedance would, therefore, always have a constant resistance component. This would mean that regardless of how the reactances were shifted, the circuit impedance would always fall somewhere on the line *mn* in the diagram of Fig. 3–28a. If the inductive reactance were the larger, then the reactance would be located above the axis, as indicated by one of the impedances labeled Z_L. If the capacitive reactance were the larger, the impedance would be as indicated by one of the Z_C vectors. If they were equal, the impedance would become R. The manner in which the current in this circuit varies with frequency is shown in Fig. 3–28c. The current rises to a maximum at the resonant frequency, as shown by the R curve, but it does not rise very high nor is the peak very sharp. In Fig. 3–28b the same circuit is shown with considerably less resistance.

It is observed that the current will rise to much greater magnitudes as resonance is reached and also that a slight variation in reactance, due to

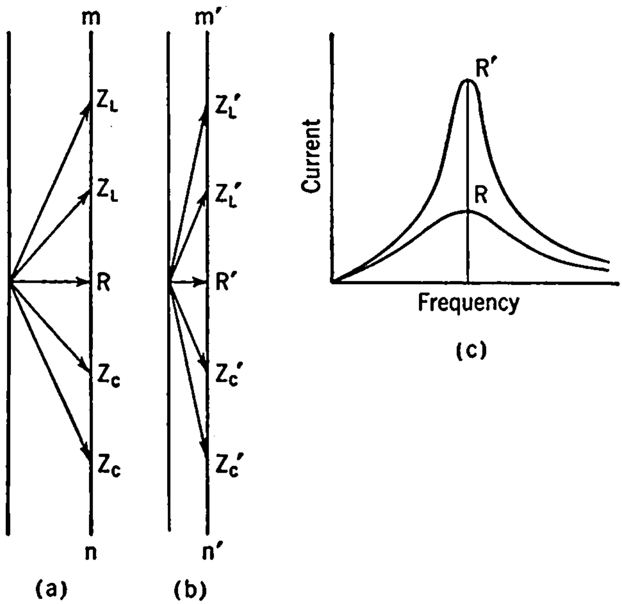


FIG. 3–28. Series Resonance—Variable Frequency.

change of frequency, will make a much greater proportional change in impedance so that curve R' is much steeper. The circuit with less resistance gives much sharper tuning than the other circuit.

DEFINITION OF Q. In this connection it is well to consider the requirements of coils and condensers for obtaining sharp resonance curves. In general, condensers have little resistance and low losses, so that they are quite often assumed to have zero equivalent series resistance. The coils, however, are made of conductors which have inherent loss in them and this quite commonly accounts for most of the loss in the circuit. From the pendulum analogy, it is realized that the less loss in an oscillatory or resonant circuit, the less energy must be supplied and the more efficient it will be. This has led to a figure of merit of reactances by which the quality of a coil may be measured. *This figure of merit is called the Q of the coil, and is defined as*

$$Q = \frac{X}{R}.$$

Thus, the higher the reactance in proportion to the effective resistance, the better is the coil or condenser.* At high frequency the losses in the dielectric of the coils and the rapid increase of resistance due to skin effect of the wire causes the resistance to increase about as fast as the reactance of the coil so the figure of merit

$$Q = \frac{2\pi fL}{R}$$

remains fairly constant over large changes in frequency. With a high Q it is possible to use the series circuit as a voltage multiplier. In many circuits, a voltage is needed to control the grid of a tube, and this voltage can be taken from the terminals of the condenser or inductance in a series resonant circuit. If the Q of the circuit is high, the voltage will be many times the impressed voltage.

Effect of source impedance. The above analysis of resonant circuits has assumed that the circuit was connected to a constant-potential line. In radio circuits this is seldom true, as the source of the voltage may itself have considerable impedance. In a series resonant circuit all of the series impedances must be considered, so that it is just as important to have low resistance in the source as in the load. If a vacuum tube of high plate resistance is used as a source, it would be expected, therefore, that a series circuit would have very little frequency selectivity.

Parallel Resonance. In the illustrative example on parallel circuits, it was found that the total current was less than the current in either branch. Furthermore, if the current drawn by the condenser were increased by increasing the capacitance, the total current would be reduced even more. If the reactive current taken by the condenser were made just equal to the reactive current taken by the inductance, then the interchange of stored energy would be balanced and the circuit would be in parallel resonance. The only current drawn from the main line would be a small in-phase current to supply the losses. As has already been stated, the condenser losses are negligible so the main item of loss is in the coil resistance. If the Q of the coil is high, this loss becomes so small that the equivalent resistance at resonance reaches very high values. This is a very interesting circuit, because in this case a reduction in the resistance of the coil makes a large increase in the equivalent circuit resistance. Although the parallel resonant circuit is definitely in a resonant condition, the

* This figure of merit Q has absolutely no connection with the symbol Q for charge. It is unfortunate that the same letter has been adopted for both, but the manner of use will indicate which meaning is intended in nearly all cases.

early application of resonance to series circuits caused the term resonance to be applied to a condition of minimum impedance. It is, therefore, very common to call the condition of maximum impedance by the name of *antiresonance.* Usually, a distinction is made between the unity power-factor condition which was here specified as parallel resonance and the condition of minimum current or maximum impedance which is called antiresonance. Where the losses in the circuits are low, as is the case in most radio circuits, this distinction has little practical significance.

EFFECT OF SOURCE IMPEDANCE. It has been learned that high source impedance destroyed the frequency selectivity of the series resonant circuit. The frequency selectivity of the parallel or antiresonant circuit, on the other hand, is increased by high source resistance. This difference is found because the equivalent impedance of the antiresonant circuit rises with the approach to resonant frequency so markedly that the terminal voltage is much higher at resonance than at frequencies on either side of resonance. Since vacuum tubes in general, and the screen-grid type in particular, have high internal resistances, the parallel resonant circuit is used extensively in radio circuits where the vacuum tube is the source of power.

Exercise 3–17. (a) Determine the size of condenser necessary to give resonance at 300 kc when the inductance is 10 microhenries. (b) Plot a resonance curve of current against frequency if the Q of the coil is 20 and the pressure is 4 v. (c) Plot a similar resonance curve for a coil with a Q of 6. (d) Plot a third resonance curve for the coil having a Q of 20, assuming that the power source has a resistance of 25 ohms.

Exercise 3–18. (a) Determine the size of condenser to give parallel resonance on 300 kc if the coil has an inductance of 1 millihenry. (b) If the Q of the coil is 20 and if this parallel circuit is connected to a vacuum tube having a plate resistance of 5,000 ohms, plot a curve of terminal volts against frequency where the generated signal is 125 v. (c) Plot a similar curve for a vacuum tube having a plate resistance of 50,000 ohms.

Impedance Matching. In most radio circuits, the power involved is small and so power efficiency is of comparatively little importance. It is, however, important to pass as much power through the circuit as is possible in order to obtain the maximum signal. It can be proved that this maximum power transfer occurs when the impedance of the load is equal in magnitude to the impedance of the power source and has a reactance equal and opposite to that of the power source. This equalization of the impedances of the load and generator is one form of *impedance matching.* In order to match impedances, the series resonant circuit should be used for power sources of low impedance and the parallel or antiresonant circuit should be used for high-impedance sources.

The use of parallel resonant circuits for impedance-matching purposes is quite common. Such a circuit is shown in Fig. 3–29. The high-impedance circuit is the standard parallel resonant circuit while the low-impedance circuit is tapped off the inductance. If this circuit is viewed from the low-impedance connection, it is difficult to determine, as mentioned previously, whether it is a series or a parallel resonant circuit, although it is unquestionably in resonance. This circuit can be used either to step up the impedance or to step it down.

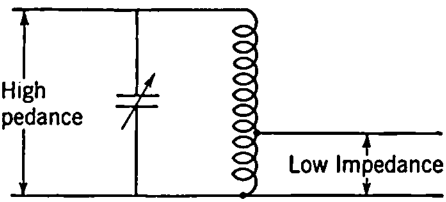


Fig. 3–29. Impedance Matching Using a Parallel Resonant Circuit.

Wave Meters. The resonant circuit has extensive use as a wave or frequency meter for approximate measurements. The meter consists of a coil and a variable condenser which is calibrated in terms of the frequency at which it resonates with the particular coil. Such a meter is placed close to a source of radio-frequency power and the condenser adjusted until the indicator, a lamp or meter, has a maximum reading. This maximum indicates that the circuit is in resonance and the frequency can be determined from the adjustment of the condenser.

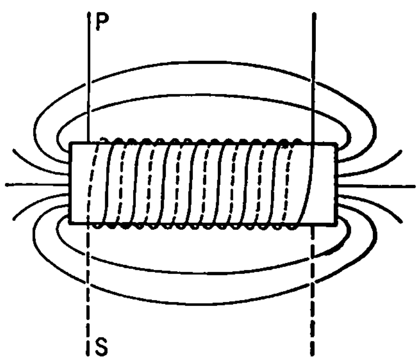


Fig. 3–30. A Mutual-inductance—Air-core Transformer.

Mutual Inductance. If an inductance coil such as has been studied in Fig. 3–10 has a second coil wound around it, as shown in Fig. 3–30, and if an alternating current flows in the first or primary coil, a voltage will be induced in the secondary coil. This secondary voltage is caused by the flux of the primary coil, which is increasing and decreasing sinusoidally with time just as the current is varying. It was found that this varying flux produced a voltage in the primary winding which by Lenz's Law tended to send a current through the coil in such a direction as to oppose the change of flux. Since the secondary coil is wound so very closely around the primary, most of the flux which threads through the primary also threads through or links the secondary. The rate of change of this flux produces a voltage in the secondary. Since the flux is directly proportional to the current in the primary, the secondary voltage is directly proportional to the rate of change of primary current.

When two coils are in such a position that a change of current in one will produce a change of flux linking the second, the two coils are said to have mutual inductance. This mutual inductance is measured in the same units as self-inductance. Thus, *when a rate of change of one ampere per second in the primary coil will produce one volt in the secondary coil, the two coils are said to have one henry of mutual inductance.*

When two coils are wound very closely together, as in the above illustration, they are said to be very closely coupled. When, however, the two coils are located so that only a small part of the flux produced by the primary coil links the secondary, the coils are said to be loosely coupled. Loose coupling is shown in Fig. 3–31, where the two coils are placed some distance apart although still on the same axis.

Effect of the Secondary. When dealing with the single coil, it was found that, neglecting resistance, the current in the coil increased until the voltage of self-inductance neutralized the impressed voltage. When a second coil is added, as shown in Fig. 3–30, it produces no effect on the primary as long as it is open circuited. If, however, resistance or some other form of load is connected across the secondary terminals, a current will flow. This current, by Lenz's Law, is in such a direction as to oppose the change of flux which is causing it. This tends to reduce the flux in the primary, which in turn tends to reduce the counter voltage. As soon as the counter voltage drops even a very small amount, an additional primary current is drawn from the line to neutralize the magnetomotive force of the secondary current and to bring the primary flux back to its original value. This increases the induced electromotive force until it is again equal to the impressed voltage. If the secondary is supplying power to a load the secondary current will be in phase with the voltage, and the primary current which neutralizes it will also be in phase with the voltage; power will thus be fed into the primary and transferred to the secondary through the medium of the common flux. Such an arrangement of closely coupled coils is called a *transformer*. If the flux is in air, as in the diagram of Fig. 3–30, it is known as an *air-core transformer*. If an iron magnetic path is provided for the flux, the transformer is known as an *iron-core transformer.*

Use and Characteristics of the Transformer. Since in most transformers, particularly those having iron cores, the coupling is very close, for the preliminary discussion it will be assumed that all of the flux threads through or links both coils. If this condition exists, the rate of change of flux for each turn is the same, regardless of whether the turn is the primary or secondary. The induced voltage across the coil will, therefore, be proportional to the number of turns. This characteristic of a transformer is used to increase or decrease the A.C. voltage. If it is desired to increase the voltage fed into the grid of a tube, a transformer may be used with many more turns on the secondary than on the primary.

Or, if a high-impedance vacuum tube is being used to supply a loudspeaker where considerable current is required at low voltage, the transformer coil having the very large number of turns will be connected to the vacuum tube and the coil having only a few turns will be connected to the loudspeaker. This is another instance of impedance matching.

A more careful study of the effects of the current in the transformer will now be made. It was learned earlier that the magnetic effects were proportional both to the current and to the number of turns. This means that the magnetic effect is proportional to the product of current and turns, usually called ampere-turns. Since the ampere-turns must be equal for neutralization, the currents in the primary and secondary will be in an inverse ratio to the number of turns. Thus, if the primary has three times as many turns as the secondary, it will require only one third as many amperes to obtain the same number of ampere turns. It should be remembered that this neutralization is not complete. A certain amount of current would flow in the primary even when the secondary is open because flux must be produced to set up the counter voltage of self-induction with which to neutralize the impressed voltage. This current is known as the *exciting current.* A very convenient rule in transformer analysis is that, *neglecting exciting current, the ampere turns on the primary are just equal and opposite to the ampere turns on the secondary.* Since the voltage is proportional to the number of turns and the current is inversely proportional to the number of turns, the product, or the volt-amperes, of primary and secondary are equal.

IRON-CORE TRANSFORMERS—POWER. In low-frequency transformers, and especially in power transformers, iron cores are used to reduce the amount of exciting current required and thus to improve the efficiency. In most cases, these transformers are designed for a certain definite maximum voltage on the winding. This is because the ratio of exciting current to flux increases very rapidly when the flux density goes beyond certain limiting values, as is indicated in the magnetization curve of Fig. 2–27. Here it is seen that the flux density rises very rapidly with increase of ampere-turns until the density has reached about 80,000 lines per square inch. Thereafter, it rises much less rapidly and if a high flux density is required to neutralize the impressed voltage, extremely large exciting currents will result. If, for instance, a 110-v transformer were connected to a 220-v power line, the flux density in the core would have to be doubled in order to produce the necessary voltage of self-induction. If the transformer were designed to operate at a maximum flux density of 65,000 lines per square inch, the required density at 220 v would be 130,000, which is entirely beyond the limits of the diagram. In circumstances of this kind the exciting current becomes so great that the I^2R losses will cause the transformer to overheat and be damaged. It is very important, therefore, that power transformers should be operated at their correct voltage and frequency rating.

IRON-CORE TRANSFORMERS—AUDIO. The preceding analysis assumed that power in unlimited quantities was available at a constant voltage, which is a true condition for most power lines. In a radio set, however, transformers are used for matching impedances, and the power source is usually a vacuum tube with high plate resistance. In this case the current in the primary is dependent upon the strength of the signal. If the strength of the signal is limited, then there is low flux density in the audio transformer and the secondary signal faithfully reproduces the primary signal. If, however, the signal strength becomes excessive, the flux density reaches the saturation point and the flux is no longer proportional to the primary current, so the secondary ceases to reproduce the signal of the primary. This effect is known as *distortion* and is one of the many limitations to accurate reproduction of signals.

AIR-CORE TRANSFORMERS. In most air-core transformers, the assumption that all of the flux links both windings is so far from true that results based on this analysis are not valid. It continues to be true, however, that the signal impressed on the primary causes a voltage and current to be set up in the secondary. The current in the secondary also causes a voltage to be produced in the primary which affects the primary current. The mathematical analysis of this type of circuit is beyond the scope of this text, but the meaning of certain terms, and a qualitative discussion of results, are given in the following pages.

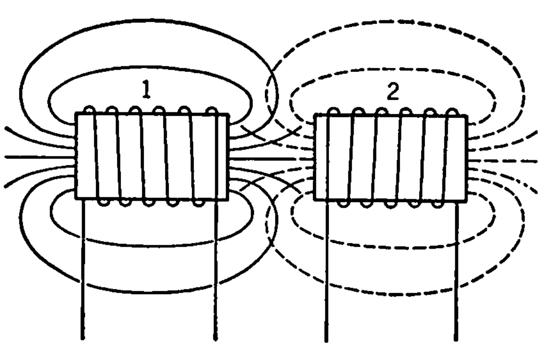


FIG. 3-31. Mutual Inductance—Loose Coupling.

Coefficient of Coupling. In Fig. 3-31 it is seen that only a small part of the flux from coil 1 threads through, or links coil 2. If the total flux of coil 1 is designated as ϕ_1,* and that part of the flux which links coil 2 as ϕ_{12}, then the ratio of ϕ_{12} to ϕ_1 is a measure of the magnetic mutuality of the two coils. This ratio is also equal to the ratio of ϕ_{21} to ϕ_2, when ϕ_{21} is the flux (caused by current in coil 2) which links coil 1 and ϕ_2 is the total flux in coil 2. This ratio is called the *coefficient of coupling of the two coils* and may be expressed mathematically as follows:

$$k = \frac{\phi_{12}}{\phi_1} = \frac{\phi_{21}}{\phi_2}.$$

* ϕ is the Greek letter phi and is the symbol usually used for flux.

It follows that

$$\phi_{12} = k\phi_1 \qquad \text{and} \qquad \phi_{21} = k\phi_2.$$

The self-inductance of coil 1 is the product of ϕ_1 per ampere and the number of turns in coil 1:

$$L_1 = \frac{\phi_1 N_1}{I_1}.$$

Similarly, the self-inductance of coil 2 is the product of ϕ_2 per ampere and the number of turns in coil 2:

$$L_2 = \frac{\phi_2 N_2}{I_2}.$$

The mutual inductance of coil 1 with respect to coil 2 is the product of ϕ_{12} per ampere of current in coil 1 and the turns in coil 2:

$$M = \frac{\phi_{12} N_2}{I_1} = k\frac{\phi_1}{I_1}N_2.$$

Similarly, the mutual inductance of coil 2 with respect to coil 1 is the product of ϕ_{21} per ampere of current in coil 2 and the turns of coil 1:

$$M = \frac{\phi_{21} N_1}{I_2} = k\frac{\phi_1}{I_2}N_2.$$

If the two definitions for mutual inductances are multiplied together,

$$M^2 = k^2\left(\frac{\phi_1}{I_1}N_2 \times \frac{\phi_2}{I_2}N_1\right).$$

By rearranging terms and by substitution,

$$M^2 = k^2\left(\frac{\phi_1}{I_1}N_1 \times \frac{\phi_2}{I_2}N_2\right),$$

$$M^2 = k^2 L_1 L_2.$$

Solving this equation for k, another relation is specified by which k may be determined:

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

If coils 1 and 2 are connected in series so that the fluxes add, the currents in both are the same, and the equivalent inductance is the sum of L_1 and L_2 plus the mutual inductance of coil 1 with respect to 2 and the mutual inductance of coil 2 with respect to 1. This becomes

$$L_t = L_1 + L_2 + 2M.$$

If the two coils are connected so that the self- and mutual-inductance fluxes are in opposition, then

$$L'_t = L_1 + L_2 - 2M.$$

If the second equation is subtracted from the first, then

$$4M = L_t - L'_t,$$

$$M = \frac{L_t - L'_t}{4}.$$

This relationship is the basis for an experimental determination of the magnitude of mutual inductance.

Selectivity and Coupling. Although it is desirable to have a very high value of coupling coefficient for power transformers, close coupling is not always desirable. In order to obtain a high degree of selectivity with respect to certain frequencies, the resonant circuits must be free to oscillate at their normal frequencies without too much interference from the currents in the mutual-inductance circuit element. Thus, if in Fig. 3–32 both the primary and secondary are tuned to resonance at a desired frequency, and if this signal of this frequency is connected to the primary, then a slight current and voltage will appear in the secondary even though it is some distance away. If the secondary is moved closer, this voltage E_2 will increase because more excitation is received from the primary coil to take care of the losses in the secondary oscillatory circuit. This increase does not continue indefinitely, however, for after reaching a certain critical value of

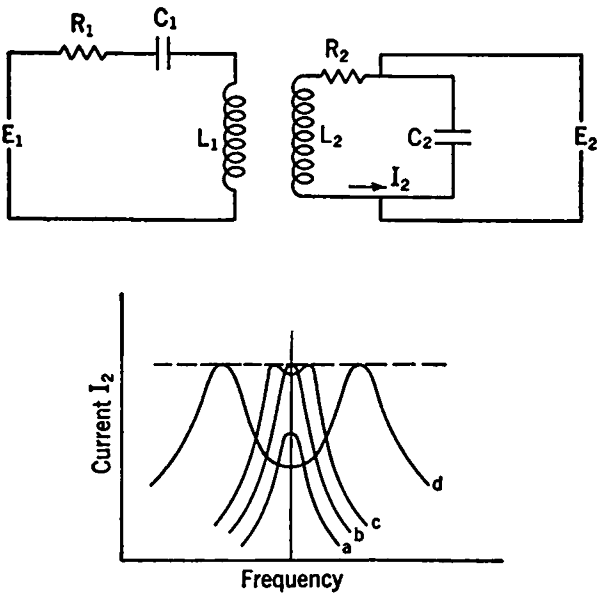
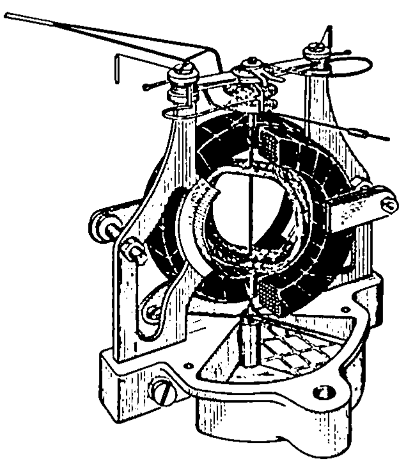


FIG. 3–32. Resonance Curves for Tuned Circuits with Different Amounts of Mutual Coupling.

mutual inductance the current and voltage in the secondary decline with increase of mutual inductance. The reactions of the secondary current on the primary circuit are such that the primary circuit is no longer in resonance at that frequency and the primary current is reduced so much that the increased coupling reduces the secondary current and voltage.

If the frequency characteristics for these several conditions of mutual-inductance coupling are obtained from test, it is found that with a very low coefficient of coupling a curve is obtained similar to Fig. 3–32-*a*. It will be noted that it is very sensitive to changes in frequency, but the coupling is *insufficient* to give the maximum signal. If the coupling is increased until the *critical* value is reached, the maximum signal strength comes through at the tuned frequency. This condition is also very sensitive to frequency changes and is desirable where only the one definite response is required. If the coupling is increased just slightly beyond critical then the frequency response is fairly constant over a narrow band, dropping steeply on either side, as shown in Fig. 3–32-*c*. This coupling is extensively used in radio circuits. If the coupling is increased still more, a frequency characteristic similar to that shown in Fig. 3–32-*d* is obtained, which has a definite double resonance response. This last type of response is an exaggerated form of the previous one and is useful only in showing the true character of the previous response.

Alternating-current Meters. Nearly all A.C. meters are current-measuring devices. To this extent they are similar to D.C. meters, but the similarity is limited, because the rapid oscillations of the current do not permit the use of the D'Arsonval type of meter. It is necessary to arrange some mechanism which will give torque in the same direction in spite of the reversal of current. Four main types of indicating meters are used for this purpose. They are the dynamometer, the iron-vane, the rectifier, and the thermocouple types of A.C. meters.



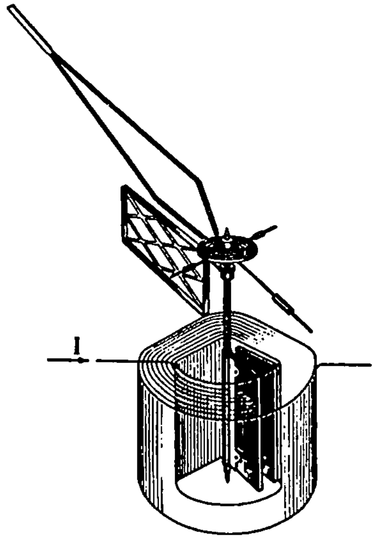
Courtesy of Weston Electrical Instrument Co.

Fig. 3–33. Dynamometer Type of Meter.

The dynamometer type. In the dynamometer type of instrument, a pair of coils carrying the alternating current to be measured produces a magnetic field. These coils are the heavy coils shown in Fig. 3–33. The same current or a portion of it is carried down to the movable coil through the springs. In this type of instrument the reversals of current in both coils come at the same instant so that the torque is always in the same direction and roughly proportional to the square of the current. This meter will operate satisfactorily on direct current, but it is not as sensitive as the usual D.C. instrument. It may be used either as an ammeter or as a voltmeter, but its greatest use is as a wattmeter.

When it is used as a wattmeter, the load current is fed through the main field coils and the moving coil is connected in series with a resistance across the line. Since power is $EI \cos \theta$ and since the field strength in the meter at the time of maximum voltage is $I \cos \theta$, the torque and hence the reading will be directly proportional to the power.

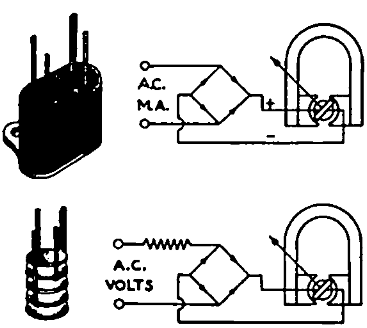
The chief disadvantage of this type of instrument is the cost of construction. Its use is limited, therefore, to wattmeters and high-precision laboratory meters.



Courtesy of Weston Electrical Instrument Co.

FIG. 3–34. Iron-vane Type of Meter.

THE IRON-VANE TYPE. The iron-vane type of instrument takes many different forms. All have a soft-iron vane attached to a shaft mounted in jeweled bearings. A spring provides the restoring torque and a pointer indicates the deflection. The torque to operate the meter depends upon the magnetic response of the soft-iron vane to the magnetic field set up by a fixed coil. A meter of this type is shown in Fig. 3–34. The alternating current is flowing in the coil and magnetizes both the stationary and movable soft-iron plates. Since the polarity of both plates is the same, they will repel each other even though that polarity is reversing rapidly. No electrical connection is necessary to the moving element. When it is desired to use such a meter as an ammeter, a few turns of large wire are used for the coil. When used for a voltmeter, many turns of fine wire are used in the coil construction.

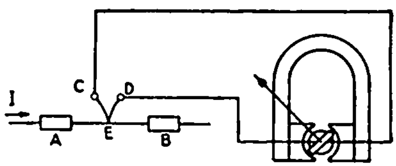


Courtesy of Weston Electrical Instrument Co.

FIG. 3–35. Rectifier Type of Meter.

RECTIFIER-TYPE METERS. It is possible to rectify alternating current and use the standard type of D.C. meter. When this is to be done it is common practice to include a small rectifier in the case of the instrument. A diagrammatic sketch of the circuit for such a meter is given in Fig. 3–35. At the left of the diagram is an illustration of the copper oxide rectifier, approximately to full scale. Meters of this type are usually limited to a few milliamperes, and so are used for low values of current and as voltmeters. Crystal rectifiers are sometimes substituted for the copper oxide type. Meters of this type may be used at extremely high radio frequencies.

THERMOCOUPLE-TYPE METERS. This type of meter uses the thermocouple principle to supply a very sensitive D.C. meter. A diagrammatic sketch is shown in Fig. 3–36. The current to be measured flows from A to B, heating up the resistance wire. The thermocouple has its hot junction at E and the cold junctions at C and D. Since this meter depends only on the heating effect, it is particularly adapted to current measurements at high frequencies. Meters operating satisfactorily up to 100 megacycles may be obtained.



Courtesy of Weston Electrical Instrument Co.

FIG. 3–36. Thermocouple Type of Meter.

Alternating-current Bridges. Although it is possible to obtain quite satisfactory measurements of resistances, of currents, and of voltages with

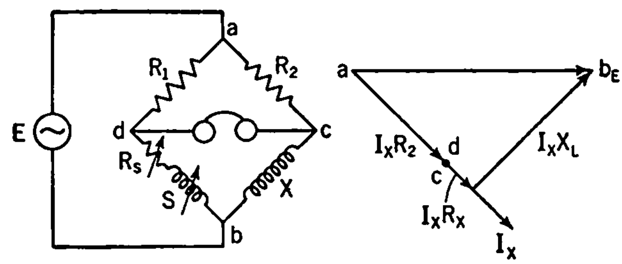


FIG. 3–37. A.C. Bridge.

the above meters, they do not lend themselves easily to the measurement of the inductance of coils or of the capacitance of condensers. The use of an A.C. type of Wheatstone bridge will permit the measurement of these quantities by direct comparison with known variable inductors and capacitors. Such a circuit is shown in Fig. 3–37. An A.C. voltage E is impressed across ab and is shown in the vector diagram at the right. The current flowing through the resistance R_2 and the unknown inductance is I_X which lags behind the voltage. The voltage drop across R_2 is shown as $I_X R_2$ and indicates that the point c has a phase and magnitude relationship as indicated. The impedance drop across the unknown inductance is shown as the summation of $I_X R_X$ and $I_X X$. If R_1 is equal to R_2, then the resistance and the inductance of the standard must be adjusted until they are equal to the corresponding values of the unknown. When this adjustment is completed, the point d is at the same voltage (both in magnitude and phase) as point c, and the signal across the detector will disappear.

Condensers can be substituted for the inductances in this bridge with no change in theory or operating procedure. After the correct adjustment is made, the value of the unknown can be determined by reading the calibrated values on the standard. Sometimes R_1 is not equal to R_2 and in this case

$$X = S\frac{R_2}{R_1}.$$

Resonance Testing Methods. Many times the easiest and most satisfactory method of determining inductance is to determine the capacity required to produce resonance at a known frequency. The resonant condition may be determined with meters or by any other convenient method, and if a calibrated condenser is used, then the inductance can be determined from the equation

$$L = \frac{1}{(2\pi f)^2 C}.$$

Note: When using equations of this type it should be remembered that the inductance is measured in henries and the capacity in farads.

A.C. Power Sources. The alternating currents which are most frequently met in radio are produced by oscillators which are combinations of tuned circuits and vacuum tubes arranged to produce these alternating currents and voltages. The study of the production of these waves is, therefore, reserved to a later chapter.

Voice and Music Waves. Voice and music waves are reproduced by the microphone as waves of electric current and voltage. These waves are extremely complicated and their calculation is beyond the tools developed in this text. The general type of procedure can, however, be indicated, and this will give a better understanding of the requirements of audio circuits and amplifiers.

Mathematicians have demonstrated that any complex wave may be reproduced by a combination of simple sine waves of different frequencies. Since the ear will respond to sounds ranging in frequency from 20 to 20,000 (in the extreme range), that is the limit of all frequencies necessary to make up the complex sound waves. Usually audio circuits do not try to reproduce accurately below 30 cycles or above 10,000 or 12,000 cycles. In solving circuits involving complex waves as above, it is usual to divide the impressed signal up into component frequencies, to apply each one by itself and determine the response, and then to put all of the individual responses together to get the final result.

Circuits Carrying Both Direct and Alternating Current. The above procedure can be used by the student when he is dealing with circuits carrying both direct current and alternating current at the same time. In this case the circuit will be solved for direct current just as if the alternating current were not there. The circuit will then be solved for alternating current just as if the direct current were not there, and the results will be added. An exception to this rule may have to be made when the D.C. flow affects the circuit constants of the A.C. circuit, as is often the case with tubes and occasionally with iron-core inductances. This general procedure is known as the *principle of superposition* and is used extensively in radio circuit analysis.

Answers to Exercises

1–1. Maximum rate of change = 1,256 amp per second; medium induced voltage $E_L = 62.8$ v.

1–2. Maximum voltage, 37.7; R.M.S. voltage, 26.63.

1–3. 15.06 ohms; 7.97 amp.

1–4. 2.12 ma.

1–5. 170.8 v leading I by 61.93°.

1–6. 4.92 ma lagging E by 38.13°.

1–7. 7.76 amp; 14.17 ohms (circuit elements connected in series).

1–8. 8.67 ohms.

1–9. 0.50 amp.

1–10. 25.7 ma; 778 ohms.

1–11. 0.596 amp leading E by 64.78°.

1–12. 0.271 amp; 36.9 ohms.

1–13. Angle between V_R and $V_L = 51.45°$; angle between $E_{applied}$ and $V_L = 52.0°$.

1–14. (a) 42.75 watts; (b) 79.4 watts; (c) 39.1 watts.

1–15. 17.9 ma.

1–16. 27.4 ma; 5.83 v lagging by 170.1°.

1–17. 0.0281 μf.

1–18. 281 $\mu\mu$f.

CHAPTER 4

Electronic Principles

Introduction. Of all the component parts used in modern systems of radio communication, the electron tube is one of the most important. Many of the basic principles of radio were well understood for years before they could be applied in practice, but their development and practical use had to await the discovery and improvement of the vacuum tube.

Vacuum tubes are made in all sizes, from the tiny acorn tube to the giant 100-kw water-cooled tubes used in large radio transmitters. They function as oscillators, amplifiers, detectors, rectifiers, modulators, as voltmeters, oscillographs, and in many other special ways. The importance of having a good working knowledge of their properties and applications cannot be overemphasized.

Thermionic Emission. As explained in Chapter 2, metallic conductors are composed of atoms and molecules, with a great number of *free electrons*, in continual random motion, which are not bound closely to particular atoms. If the temperature of a conductor is raised sufficiently, some of these free electrons will acquire enough kinetic energy to permit their escape by penetrating the surface of the conductor. This process is known as *thermionic emission*, and it is found that various metals differ widely with respect to this property. In present-day tubes, the most commonly used materials for the purpose are as follows:

TUNGSTEN. Emitters made of this metal are used principally in large transmitting tubes, where heavy emission currents are required. Tungsten emitters must be operated at 2,500°K (degrees Kelvin = degrees Centigrade + 273), and they require relatively large amounts of power for heating. The operating temperature is close to the melting point of tungsten, and consequently the heating voltage must be closely regulated. Within this limitation, however, tungsten emitters are rugged and long-lived, and capable of withstanding considerable overloads.

THORIATED TUNGSTEN. This material is tungsten containing a small percentage of thorium oxide. After suitable heat treatment, it is found that migration of thorium atoms produces an activated layer on the filament surface, with the result that electrons are emitted at a much lower temperature (1,900°K) than for pure tungsten. This type of cathode has, therefore, a greater emission efficiency, but it is much more sensitive to overload than pure tungsten, as the thorium layer is stripped off or evaporated when excessive emission currents are demanded.

OXIDE-COATED EMITTERS. Filaments and indirectly heated cathodes of this type consist of an inert base, usually a nickel alloy, covered with a layer of barium and strontium oxides. Cathodes of this type operate at a still lower temperature (1150°K), and the emission efficiency is very high. Almost all tubes of the small receiving type, as well as many of medium size, utilize cathodes of this type.

Physical Construction of Cathodes. Two types of cathodes are employed in vacuum tubes, the filamentary and the indirectly heated. The filamentary cathode consists of a wire filament, usually bent in the form of a V or a W, and is heated by the passage of a current, either direct or alternating, through it. If alternating current is used for heating, as is most frequently the case today, the different parts of the cathode are not at constant potential, on account of the IR drop of this current. Since the electron stream is determined by the potentials of the other tube elements with respect to the cathode, as explained later, this may cause fluctuations which result in the development of *hum* in the tube output. To reduce these fluctuations as much as possible, it is common practice to connect the return circuit from the other tube elements to a center tap on the transformer winding which supplies the filament-heating current. This point is always at the *average* potential of the entire filament.

The indirectly heated cathode consists of a metallic sleeve coated with emitting substance, and is heated by wires passing through channels in a ceramic insulating core placed inside the sleeve. The advantage of this construction is that the entire cathode is at a uniform potential, since it carries none of the heating current. It is therefore referred to also as a *unipotential cathode.* Another advantage is that the heater wires are so close together that the current in them produces practically no magnetic field, which otherwise would also affect the electron flow and thereby contribute to the hum in the tube output.

Diodes. Thermionic tubes are conveniently classified according to the number of active elements. The simplest is the diode, which contains two active elements, a cathode and a plate or anode. If the plate is made positive with respect to the cathode by connecting a battery or other source of potential between them, electrons emitted by the cathode will be attracted to the plate; on arriving there they deliver their electrical charges to the plate structure, then migrate as free electrons within the metal and through the battery back to the cathode, to replenish its supply. This continual flow of electrons constitutes an electric current, passing through the *plate circuit* of the tube. The magnitude of the current depends only upon the number of electrons per second arriving at the plate, and this in turn is a function of cathode temperature and plate potential.

The circuit of Fig. 4–1 may be used to investigate these relations If the heater current I_f is held constant at a value I_{f_1}, and the plate voltage E_b is increased from zero, the plate current I_b will vary in the manner shown by the lower curve in Fig. 4–2. If now the heater current is increased to a new value I_{f_2}, thereby raising the cathode temperature, the upper curve of Fig. 4–2 results. It will be noted that the two curves almost coincide at the smaller plate voltages, and that each becomes almost horizontal in the upper region, but at different values of plate current.

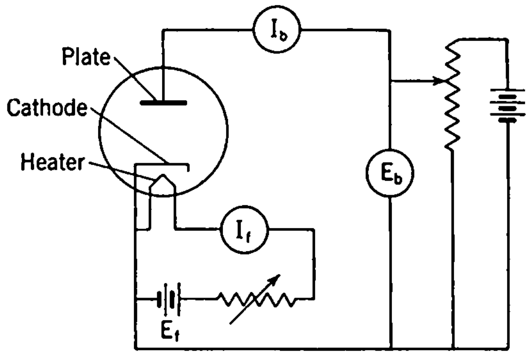


FIG. 4–1. Circuit for Obtaining Diode Characteristics.

The leveling off of plate current is known as *temperature saturation;* it is due to the fact that all the electrons which the cathode is able to emit at this temperature are reaching the plate, and consequently the current is at its maximum possible value. The only effect of further increase of plate voltage is to cause each electron to arrive at the plate with higher velocity, but as this does not increase the *number of electrons per second,* the plate current is not changed. If, however, the cathode temperature is raised, as in the second curve, more electrons are given off, and the maximum plate current is increased.

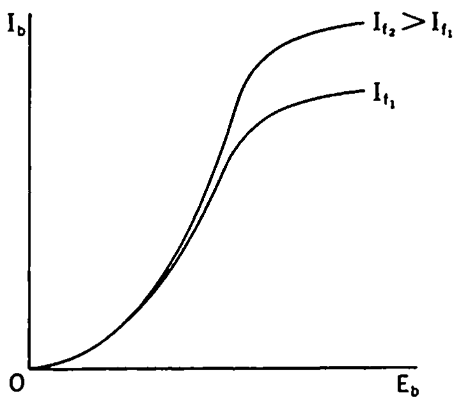


FIG. 4–2. Limitation of Diode Current by Temperature.

SPACE CHARGE. If in Fig. 4–1 the plate potential is held constant at E_{b_1} and the heater current I_f is varied, the lower curve of Fig. 4–3 is obtained. On increasing the plate voltage to a higher value E_{b_2}, it is observed that the plate current is increased, but only at the higher cathode temperatures (larger values of I_f). Evidently temperature saturation is responsible for the agreement of the two curves in the lower region. Both curves are nearly horizontal in the upper part, and this limitation of current is caused by the presence of a *space charge* near the cathode.

As electrons are given off by the cathode, they form a cloud of negative charges in the surrounding space. These negative charges lower the potential in that region, and thereby produce a field which urges the electrons back toward the cathode. If there were no plate, all the electrons emitted would eventually return to the cathode. Even with the plate at a positive potential, only those electrons emitted with sufficient velocity will have enough energy to penetrate the cloud, and their motion will constitute the plate current. If the plate voltage is increased, it will further overcome the effect of space charge, and thereby produce an increase of plate current.

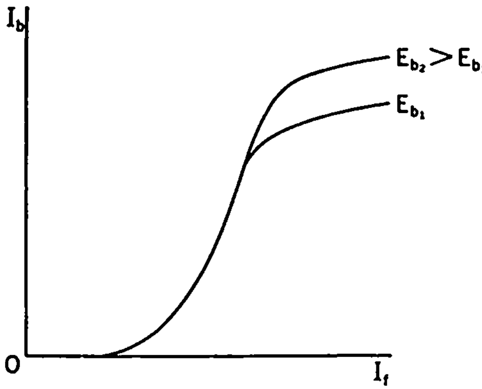


FIG. 4–3. Limitation of Diode Current by Space Charge.

This effect of space-charge limitation is extremely important, as the useful working range of nearly all vacuum tubes is precisely the region in which space charge is effective. The applications in which emission is limited by temperature saturation are very rare.

PLATE RESISTANCE. Although the relationship between plate current and plate voltage is not one of simple proportion and cannot be expressed in terms of Ohm's Law, one may nevertheless speak of the resistance of the internal plate circuit of the tube. The ratio of plate potential to plate current is known as the *D.C. plate resistance*, and is shown in Fig. 4–4 as *OA/AP*. The concept of D.C. plate resistance is useful in many cases where the current through the tube is steady or constant, but a much more important quantity is the *A.C. plate resistance*, or the ratio of *change* in plate voltage to *change* in

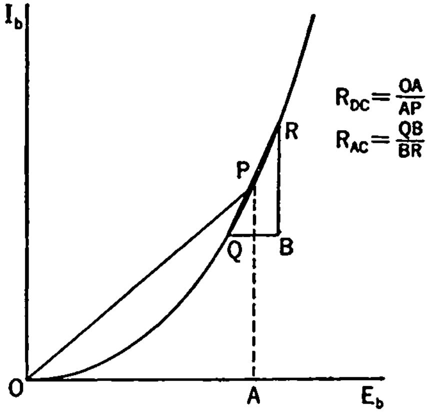


FIG. 4–4. D.C. and A.C. Plate Resistance.

plate current. In the neighborhood of point P of Fig. 4-4, the A.C. plate resistance is given by QB/BR. For many diodes, the D.C. plate resistance is approximately twice as great as the A.C. plate resistance.

It is clear from the figure that neither of these resistances is constant for all voltages, and that both tend toward lower values as plate voltage and plate current are increased.

Rectifying action. If the plate of the diode is made negative with respect to the cathode, the electrons will be driven back to the cathode and no plate current will flow. Hence, if an alternating voltage is applied to the plate, current will flow only in the positive half cycles, and it will consist of a succession of pulses, always in the same direction. The tube therefore functions as a *rectifier*, in that an alternating voltage applied to the tube produces a unidirectional current. Wide use is made of this characteristic to obtain direct voltages and currents from an A.C. source.

Effect of Gas. If the envelope of the tube is not completely evacuated, so that a small amount of gas remains enclosed, some electrons in traveling toward the plate may collide with molecules of the gas. Such collisions may result in releasing electrons from their orbits within the gas atoms, and once free these electrons likewise are attracted to the plate. They may in turn collide with other gas molecules, and liberate more electrons. This process is known as *ionization*, and it may be cumulative in nature if enough gas is present so that a large number of collisions take place in the space between cathode and plate.

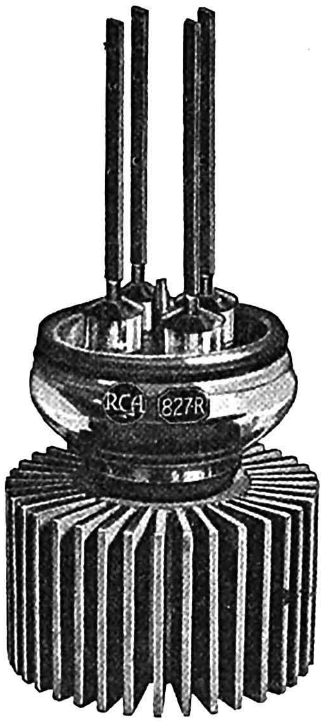
A definite amount of energy is required to dislodge an electron from an atom of any particular gas, and the voltage necessary to furnish this amount of energy is called the *ionization potential* for that gas. Typical values are listed below.

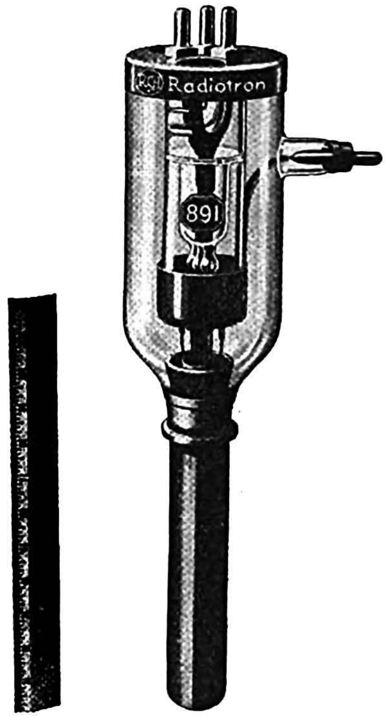
Gas	*Ionization potential (volts)*
Argon	15.7
Neon	21.47
Helium	4.0
Mercury	10.38

The atom or molecule which has lost an electron, and therefore also a definite amount of negative electricity, has acquired a net positive charge. It is attracted to the cathode, but on account of its greater mass the atom or molecule moves much more slowly than an electron traveling toward the plate. It may, however, acquire considerable kinetic energy before completing its journey, and the impact of large numbers of such positively charged particles can have a destructive action on the cathode surface, particularly if this is of the coated type. This action is referred to as *cathode bombardment*; it may be quite important in such tubes as the mercury-vapor rectifiers considered in Chapter 5.

Limitations in Operating Conditions. In attempting to operate tubes at high power, certain limitations are encountered. One of these is the peak emission current; it represents the maximum rate at which electrons can be given off without harm to the cathode structure, and without causing harmful increase of voltage drop within the tube.

Another limitation is the peak inverse voltage, or the maximum voltage that can be impressed across the tube elements without causing failure, either by electrical breakdown or by ionization of the gas within the tube.





Courtesy R.C.A. Manufacturing Co.

FIG. 4-5. Air-cooled Tube with Radiating Fins.

FIG. 4-6. Water-cooled Tube.

The third important limitation is plate dissipation. When an electron arrives at the plate, its kinetic energy is transformed to heat and serves to raise the temperature of the plate structure. If the temperature goes too high the plate may melt, or electron emission from the plate may take place. Many high-power transmitting tubes have heat-radiating fins attached to the anode for the purpose of transferring heat generated at the plate as quickly as possible to the surrounding air; others are designed to permit water circulation around the anode structure (Figs. 4-5 and 4-6).

The maximum amount of heat that can be safely dissipated by the plate of any tube is stated in terms of the corresponding electrical power, in watts, and is referred to as the *maximum plate dissipation*. With tubes containing grid structures, the heating of each grid must also be taken into account. In the case of transmitting tubes, the manufacturer furnishes information giving the maximum allowable value of each of these limiting factors, and in order to insure maximum tube life it is important to consider each of them individually.

Triodes. A triode is a tube which contains a third element, the grid, located between cathode and plate. The grid usually takes the form of a helix or spiral of fine wire, so that electrons may pass freely through it. Since the grid is nearer to the cathode, the potential of the grid has a greater effect in controlling electron flow than does the plate potential. If the grid potential is negative with respect to the cathode, as is usually true, electrons will not be attracted to the grid itself, and there will be no grid current. Variation of the grid potential will, however, have an effect on the space charge surrounding the cathode, and hence will cause variation of the plate current.

Characteristic Curves of Triodes. The effects of grid and plate potentials on flow of plate current can be investigated by means of the circuit of Fig. 4–7. In this setup, if the plate potential E_b is held constant and the grid potential varied, one of the curves of Fig. 4–8 is obtained. The other curves are found in the same way, by using different values of E_b. Such a set of curves, showing the relationship between the plate current and the grid potential for constant values of plate voltage, is known as the family of *mutual characteristics* of the tube. By use of this family, interpolating between curves where necessary, it is possible to find the plate current corresponding to any combination of grid and plate potentials.

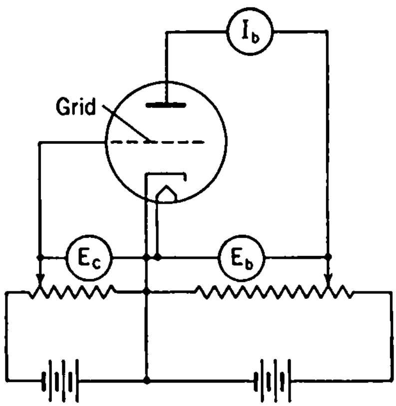


FIG. 4–7. Circuit for Obtaining Triode Characteristics.

It will be observed that each of the mutual-characteristic curves is quite similar in form to the diode characteristic of Fig. 4–2, except that the effect of temperature saturation is not apparent. The reason for the absence of saturation is that in most modern tubes the cathode is capable of furnishing many more electrons than are required for rated current. As a result, a certain amount of variation in filament or heater voltage can be tolerated without seriously impairing the tube performance, and the useful cathode life is increased. It will also be noted that the curves tend to be nearly straight lines in the upper portions, and that the curves for various plate potentials are nearly parallel. It will be shown later that these properties are important where the tube is to be used as an

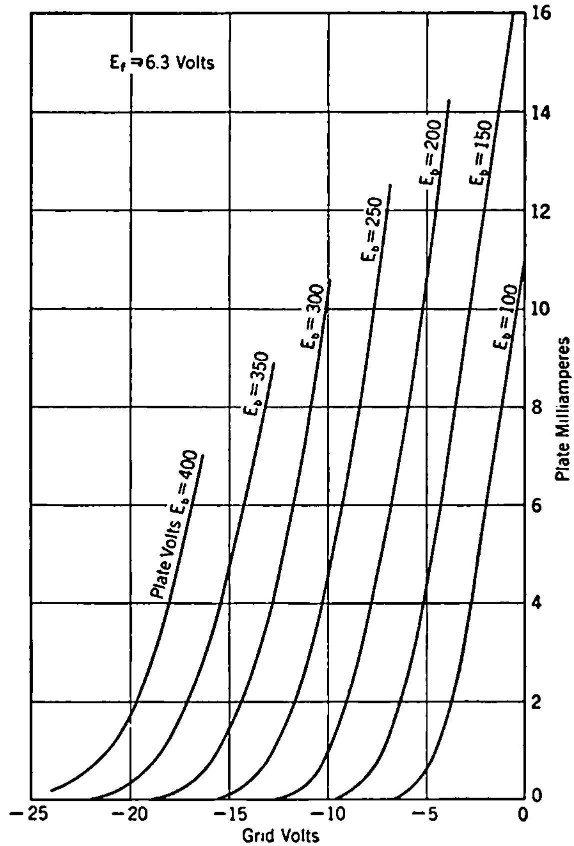


FIG. 4-8. Mutual Characteristics of 6J5 Triode.

amplifier, and where it is desirable to keep the distortion of the amplified signal as small as possible.

Another way of presenting the same information is shown in Fig. 4-9. Here the grid potential is held constant for each curve, and the plate potential is permitted to vary. This set of curves is known as the family of *plate characteristics*, and for many purposes it is more useful than the mutual characteristics. The curves of this family resemble in shape those of the mutual characteristics. The effect of increasing the negative grid potential is chiefly to shift the curves toward the right on the diagram, without causing much change in form.

Tube Parameters or Characteristics. Three important ratios, obtainable from either set of curves, are helpful in analyzing tube performance. These are the amplification factor, the mutual conductance, and the internal plate resistance. They are called the *tube parameters or characteristics.*

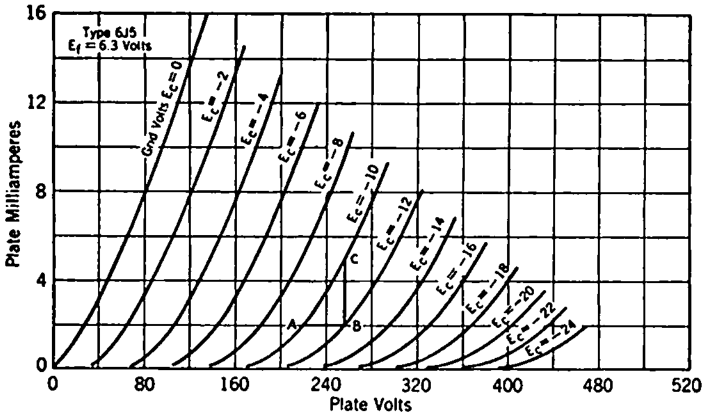


FIG. 4–9. Plate Characteristics of 6J5 Triode.

The amplification factor, symbolized by the Greek letter μ (pronounced mū), is defined as the ratio of plate-voltage change to grid-voltage change, when plate current is maintained constant. It is a measure of the relative effectiveness of the grid as compared with the plate in controlling flow of plate current. In Fig. 4–9, the plate currents at A and B are the same and by the above definition

$$\mu = \frac{AB}{\Delta E_c},$$

where ΔE_c represents the difference between the grid potentials of the two curves through A and B. In triodes, μ ranges in value from 2 to 100, with most tubes included in the range 10 to 40. For any particular triode, the amplification factor is almost constant for all operating conditions, except at very low plate currents.

Mutual conductance, g_m, is the ratio of plate-current change to grid-voltage change, when plate voltage is held constant. It is a measure of the effectiveness of the grid in controlling plate current. In Fig. 4–9,

$$g_m = \frac{BC}{\Delta E_c}.$$

Mutual conductance is stated in micromhos, and for most tubes it has a value of a few thousand. It is not nearly so constant as the amplification factor, its size depending mainly on the amount of plate current.

Internal plate resistance, r_p, is given by the ratio of plate-voltage change to plate-current change, grid voltage being held constant. It is measured in ohms, and in Fig. 4-9,

$$r_p = \frac{AB}{BC}.$$

In the case of triodes, r_p ranges in value between 2,000 and 100,000 ohms.

From the above definitions, it can be seen that the following relationship exists between the three parameters:

$$\mu = r_p g_m.$$

It is thus sufficient to specify any pair of these quantities, since the third may be computed from them.

Components of Currents and Voltages. In most tube applications the plate current, grid voltage, and plate voltage are not constant but vary with time between maximum and minimum values, as shown in Fig. 4-10,

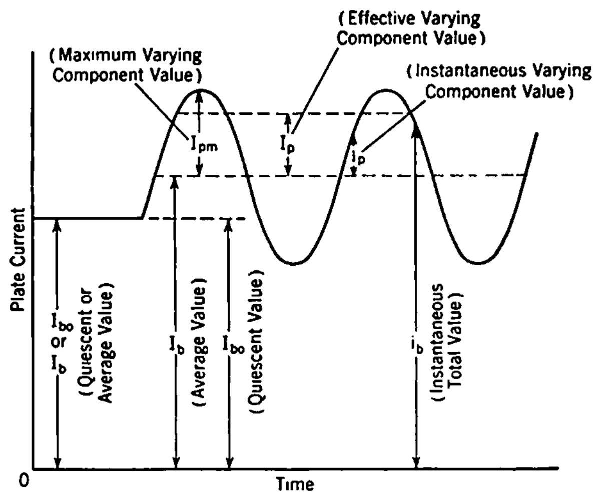


FIG. 4-10. Components of Plate Current.

which is reproduced from the I.R.E. Standards Report on Electronics. It is helpful to consider that such a current or voltage is composed of a D.C. or average component I_b, together with an A.C. or varying component of peak value I_{pm} or effective value I_p. In all but a few cases the A.C. component is the one of interest, but it should be noted that the D.C. component is necessary, since this determines the portion of the tube characteristics in which the operation takes place. It should also be noticed that the average value I_b may change when a signal is applied.

A typical instance is given in Fig. 4-11, which shows a triode tube with external resistance R in its plate circuit. E_s is an alternating voltage, the *signal*, which causes the instantaneous grid potential to fluctuate up and down about its average value E_c. The plate current i_b will vary correspondingly, and it may be regarded as having the components I_b (D.C.) and I_p (A.C.), as shown in Fig. 4-10. This current in flowing through the load resistance R produces varying amounts of IR drop, and corresponding variations in plate potential will take place. It should be noted that the average or D.C. component of plate potential E_b is less than the B-supply voltage E_{bb} by the D.C. component of drop in the load resistor R. The A.C. components of plate current and plate potential combine to give the useful output of the tube.

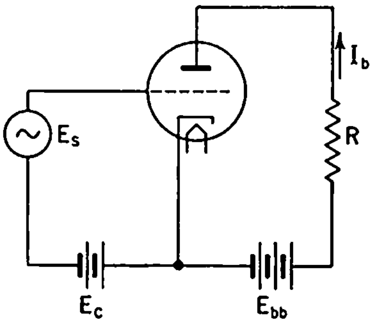


FIG. 4-11. Triode with Resistance Load.

The Load Line. A convenient graphical construction may be employed to find the output components for any operating condition. This construction is shown in Fig. 4-12, which consists of the plate characteristics of the tube in question and a superimposed *load line*. The load line is a graph of the equation

$$e_b = E_{bb} - Ri_b,$$

which is evident from the circuit of Fig. 4-11. Any point on the line represents a possible combination of plate voltage and plate current, and no other combinations are possible. It will be apparent from the equation, or by inspection of the circuit, that for $i_b = 0$, $e_b = E_{bb}$. The load line therefore intercepts the horizontal axis at this scale value. Similarly, the vertical intercept occurs at

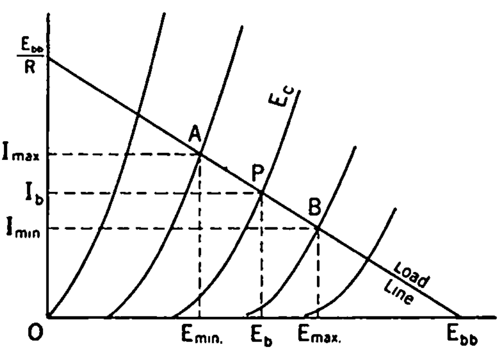


FIG. 4-12. The Load Line.

$$i_b = \frac{E_{bb}}{R}.$$

The load line is most readily drawn by use of these two points.

If the grid-bias voltage is E_c, the zero-signal or quiescent condition is represented by the point P, which is the intersection of the load line with the plate characteristic corresponding to E_c. This point will be referred to as the *operating point*, and when the signal voltage E_s is applied, the instantaneous relation between plate current and plate potential will be represented by a point oscillating about P along the load line. If the range of this oscillation is between A and B, the plate potential will fluctuate between $E_{\min}$ and $E_{\max}$, and the peak value of A.C. plate voltage will be

$$E_{pm} = \frac{E_{\max} - E_{\min}}{2}.$$

The average or D.C. component of plate potential is E_b, and for a given load line this is seen to depend on the choice of E_c.

Steady and alternating components of plate current can be read from the diagram in the same manner.

The operation of a tube with resistance load can also be represented by a line drawn on the plot of mutual characteristics, but in this case the line depends not only on the B-supply voltage and the load resistance, but also on the tube characteristics themselves. This line is known as the *dynamic mutual characteristic* of a tube and the method of constructing it is covered in Chapter 7.

The Equivalent Circuit. Another method of finding the A.C. output of a triode or other tube is by means of the *equivalent circuit*. This is shown in Fig. 4–13, and corresponds to the actual circuit of Fig. 4–11. With certain reasonable assumptions, it may be shown the equivalent circuit leads to the same values of current and voltage in the load resistor R as the A.C. components of these quantities in the actual circuit.

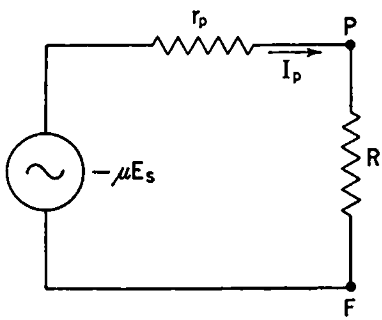


FIG. 4–13. Equivalent Circuit Corresponding to Fig. 4–11.

The assumptions required are that μ and r_p remain constant throughout the range of operation, or (stated in another way) that the curves of the tube characteristics be straight, parallel, and equally spaced. These statements may always be considered to hold for very small signals, but when a large swing is to be handled their validity should be examined in each case.

The *voltage amplification* of the circuit of Fig. 4–11 is defined as the ratio of the A.C. plate voltage to the applied signal voltage on the grid. This voltage amplification will now be computed by use of the equivalent circuit.

Since r_p and R form a simple series circuit, the current is given by

$$I_p = \frac{-\mu E_s}{r_p + R}.$$

If the direction of current flow is reversed to agree with that flowing in the actual tube, the negative sign is removed from the numerator; this is commonly done to simplify the writing. The flow of this current through the resistance R produces a voltage drop between the points P and F, which correspond with the plate and cathode of the tube, respectively. The potential of P with respect to F is

$$E_p = \frac{\mu E_s R}{r_p + R},$$

and the voltage amplification, or ratio of E_p to E_s, is

$$A = \frac{E_p}{E_s} = \frac{\mu R}{r_p + R}.$$

It should be remembered that E_p and E_s are opposite in phase. This fact can also be noted in Fig. 4–12, and it becomes an important consideration in amplifiers employing feedback, as will be shown in Chapter 7.

Tetrodes. It has been noted that the A.C. grid and plate potentials are opposite in phase. As a result, the A.C. potential difference between these two elements is considerable, being equal in magnitude to the sum of the A.C. input and output voltages. On account of the electrostatic capacitance between grid and plate, this potential difference produces a current flow from the output circuit to the grid by condenser action, and under certain load conditions such reaction or feedback causes the amplifier to become unstable. The instability may be avoided by inserting an electrostatic shield in the form of a *screen grid* between the plate and the original grid, which in this arrangement is known as the *control grid.* A tube of this type is known as a *tetrode,* or *screen-grid tube.*

To permit the flow of reasonable amounts of electron current, the screen must be operated at a positive potential. For adequate shielding of the control grid, however, the screen potential should have no A.C. component, and this must be carefully considered in any circuit application.

The fact that the screen grid acts to shield the control grid from variations of plate potential implies also that the cathode is shielded from the plate. As a consequence, electron emission is affected only very slightly by the plate potential; it is determined almost altogether by the control-grid and screen-grid potentials. A portion of the electron stream passes to the screen on account of its positive potential, and forms the screen current; the remainder, usually a much larger part, flows to the plate.

In terms of tube characteristics, since a relatively large change of plate potential causes very little change in plate current, the value of r_p, the internal plate resistance, is much larger than for a triode. For the same reason, the amplification factor μ is greater for a tetrode than for a triode. The mutual conductance g_m, however, is of about the same magnitude as for a triode, although it is somewhat reduced by reason of the fact that a portion of the emission current is diverted to the screen.

For receiving-type tetrodes, r_p is of the order of a megohm, and μ has a magnitude of several hundred.

Plate Characteristics of a Tetrode. A typical plate characteristic for a screen-grid tube is shown in Fig. 4–14. For plate voltages beyond D, plate current is nearly independent of plate voltage, as explained in the preceding section, and this region constitutes the useful working range for amplifier service. Below D, the action is complicated by another effect, namely *secondary emission.*

If an electron strikes the plate with sufficient velocity, it will dislodge

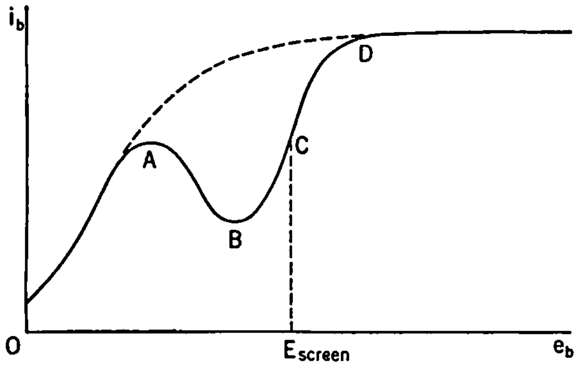


Fig. 4–14. Screen-grid Plate Characteristics.

one or more electrons from the plate structure. These are known as secondary electrons, and the process by which they are produced is called secondary emission. The secondary electrons occur in the space between plate and screen, and if the plate potential is lower than that of the screen, such electrons will be attracted to the screen and so contribute to the screen current. To the same extent they cause a decrease in plate current, and this is quite evident in the dip between A and C, Fig. 4–14. Below A, the velocity of arrival at the plate is too low for liberation of secondary electrons. Above D, secondary electrons are emitted in large numbers, but the plate potential is so much higher than the screen potential that they all return to the plate, with no net effect on plate current. Plate voltage and screen voltage are equal at C, and the decrease in plate current between C and D is explained by the initial velocity of the secondary electrons. If this velocity is large enough, it will overcome the opposing field set up by the higher plate potential, and the plate will continue to lose electrons to the screen, as shown. Beyond point D, however, the electric field between plate and screen is so strong that all the secondary electrons are returned to the plate, and the plate and screen currents remain unaffected.

The amount of secondary emission under given conditions of electron bombardment depends on the material of which the anode is constructed. Carbon and graphite, although relatively good conductors, exhibit much less of this effect than most metals, and tubes in which the anode surfaces have been coated with graphite do not have such pronounced variations of plate current in the region *ABCD* as do those with metal plate structures.

It is frequently desirable to obtain the positive voltage for the screen grid from the B-supply source, and a convenient way of doing this is shown in Fig. 4–15. The required screen voltage is usually less than the B voltage available, and the magnitude of R_s is selected so as to produce the necessary amount of drop, due to the flow of screen current. Thus, if E_{bb} is 300 v and the desired screen potential is 100 v, the required drop is 200 v. If the screen current at the working voltages is 0.5 ma, this drop will require a dropping resistor of 400,000 ohms. Without some provision for keeping the screen potential constant, however, the presence of this screen dropping resistor would cause a serious loss of output. This loss can be understood if it is remembered that increase of control-grid potential is accompanied by increase in screen current, and this increase in turn causes a fall in screen potential owing to the greater drop in R_s. This fall in turn has the effect of lowering the plate current, and thus counteracting in part the effect of the original signal on the plate current.

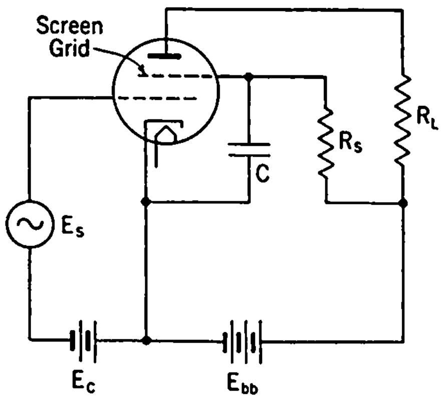


FIG. 4–15. Screen Supply for Tetrode.

It is possible to avoid this difficulty by the use of the condenser *C*, shown in Fig. 4–15. If this condenser is large enough, it will hold the potential of the screen practically constant, by by-passing the fluctuations in screen current directly to the cathode, instead of permitting them to flow through R_s. To determine the proper size of *C*, a satisfactory rule is that its reactance should not be more than one fifth the resistance in shunt with it, at the lowest frequency to be handled. In computing this reactance it is necessary to remember that there is an effective A.C. resistance within the tube between screen and cathode, exactly analogous to the A.C. plate resistance. Suppose this resistance in the present case to be 100,000 ohms.

Inspection of the circuit shows that the resistance shunting condenser C consists of the internal screen resistance in parallel with R_s (the source E_{bb} is not included, since there is no A.C. potential across it). The effective resistance across the condenser is therefore

$$\frac{100{,}000 \times 400{,}000}{100{,}000 + 400{,}000} = 80{,}000 \text{ ohms.}$$

By the above rule, the reactance of C must not exceed 16,000 ohms (one fifth of 80,000 ohms) at the lowest working frequency. If the tube is part of an audio-frequency amplifier passing frequencies down to 50 c.p.s., the required capacitance is

$$C = \frac{10^6}{2\pi \times 50 \times 16{,}000} = 0.2\mu\text{f.}$$

If a by-pass condenser of smaller size is used, the gain of the amplifier stage will be satisfactory at the higher frequencies, but will fall off at the low end.

Pentodes. One effect of the dip due to secondary emission in the plate characteristic of a screen-grid tube is to limit the available A.C. output voltage, since true amplifier action does not extend below the point D of Fig. 4–14. The range of operation would be extended materially if the secondary electrons could be sent back to the plate instead of going over to the screen, in which case the plate characteristic would have the form of the dotted curve in Fig. 4–14.

This desirable result is accomplished by the insertion of a third grid, the *suppressor*, between plate and screen grid. The suppressor is usually made with wider spacing than the other two grids, and it is connected directly to the cathode, so that its potential is zero. It has very little effect on the emission of secondary electrons from the plate, but it exerts a powerful control upon them as soon as they are produced. Even at low plate potentials the electric field between plate and suppressor is in the proper direction to send electrons toward the plate, and for this reason the secondary electrons return to the plate instead of going to other tube elements.

Tubes of the type just described are known as *pentodes*. Figure 4–16 shows the plate characteristics of the 6J7, which is a typical voltage-amplifier pentode. The suppressor-grid connection in this case is brought out to a separate pin in the base, so that is is available for special types of application, but in many pentodes the suppressor grid is connected internally to the cathode. By comparing Figs. 4–14 and 4–16 it will be quite clear that for a given B-supply voltage the pentode can furnish a much larger A.C. output voltage than the tetrode.

The values of μ, r_p, and g_m for a pentode are approximately the same as those for a tetrode of corresponding type, and therefore the circuit applications of both tubes are similar. In any circuit involving high gain it is important to isolate the input leads from the output connections, either by shielding or by physical separation. If this is not done, the advantage of internal shielding by the screen is lost, and instability occurs. For the sake of providing convenient separation between control-grid and plate wiring, many of the standard tetrode and pentode tubes have the control-grid connection brought out to a cap on the top of the tube, while all the other

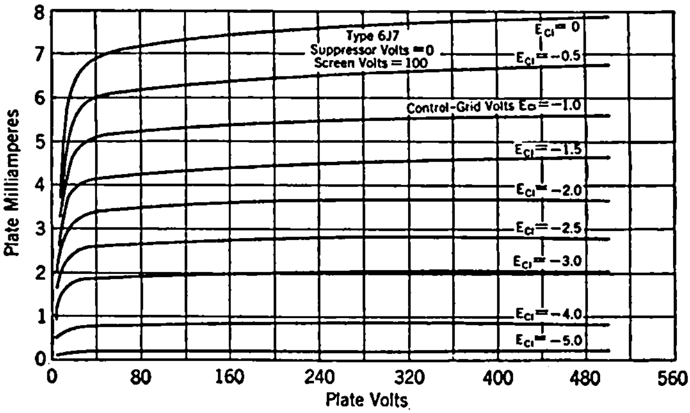


FIG. 4–16. Plate Characteristics of 6J7 Pentode.

connections are brought out at the base. The 6J7 referred to above has this construction. A more recent design, of which the 6SJ7 is an example, has all leads brought out at the base, with special internal shielding between the control-grid prong and the other elements.

The performance of a pentode amplifier may be studied in the same way as was done in the case of a triode, either by use of the load line or by the equivalent circuit. In the latter case, a simplification is possible on account of the large value of A.C. plate resistance of the pentode. In the expression for voltage amplification (p. 133),

$$A = \frac{\mu R_L}{r_p + R_L},$$

it is frequently true that r_p is so large in comparison with R_L that the latter may be neglected. By reason of the relationship between the tube constants, it is then possible to write:

$$A = \frac{\mu R_L}{r_p} = g_m R_L.$$

Variable-mu or Remote-cutoff Pentodes. In certain pentodes, such as the 6K7 and the 6SK7, the grid helix has a variable pitch, so that some of the turns are closer together than others. The closely spaced turns have a greater control over electron flow from the cathode than those with greater separation, and as a result the mutual characteristics of the tube are considerably modified. As the control grid is made more negative, cutoff of the electron stream is approached much more gradually than if the grid spacing were uniformly close. Tubes of this type are known as *variable-mu*, or *remote-cutoff*, or *super-control* pentodes. A typical example is the 6K7, and Fig. 4–17 shows a mutual-characteristic curve for this tube, in comparison with a similar curve for the 6J7. The more gradual approach to cutoff of the variable-mu type is very noticeable.

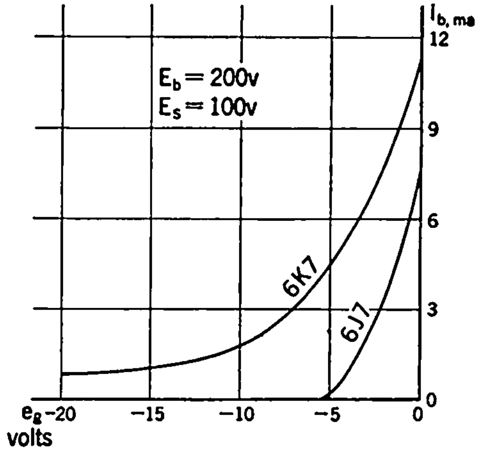


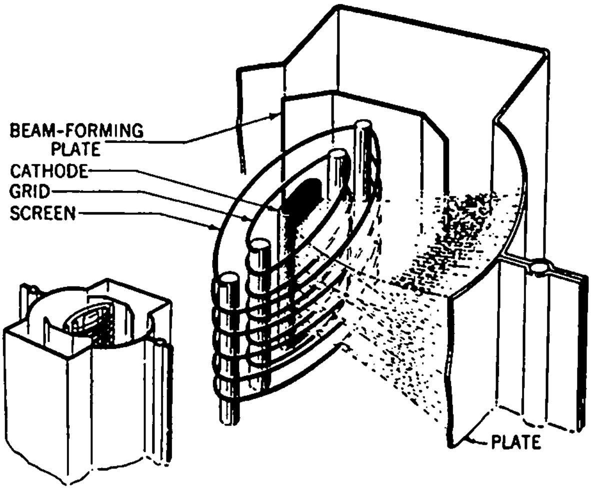
FIG. 4–17. Mutual Characteristics of 6J7 and 6K7, Showing Sharp Cutoff and Remote Cutoff.

This property can be utilized to control the gain of the tube. For a pentode, the approximate expression for voltage amplification applies:

$$A = g_m R_L.$$

The value of g_m, however, varies along the curve, being maximum at zero grid-bias voltage, and decreasing steadily to a much smaller value as the grid is made more negative. This can readily be seen from the curve if it is recalled that mutual conductance is equal numerically to the slope of the mutual characteristic. It becomes possible then to control the gain of such an amplifier by simple adjustment of the D.C. grid bias voltage. In radio receivers, which are required to respond to signals covering a wide range of intensity, the grid bias is made to depend upon the amplifier output, and as a result the sensitivity is automatically reduced when a strong signal is being received. This type of control, known as *automatic volume control* or *a.v.c.*, is incorporated in practically all modern receivers.

Beam Power Tubes. Instead of using a suppressor grid to control the secondary emission from the plate, it is possible to obtain the same effect by shaping the tube elements in such a way as to control the space charge near the plate. Figure 4–18 shows the internal structure of the 6L6, a typical beam power tube, and also the distribution of electrons within the tube. The beam-forming plate shown in this figure is connected to the cathode, and its potential is therefore zero. The field produced by this combination of elements is such as to cause a concentration of electrons to occur near the plate, as indicated in the diagram, and thereby to produce a region of minimum potential there. As long as the plate potential is higher than the potential minimum due to electron concentration, secondary elec-



Courtesy R.C.A. Manufacturing Co.

FIG. 4-18. Structure of a 6L6 Beam Power Tube.

trons will return to the plate, just as if a suppressor grid were present. The characteristic curves of this tube are shown in Fig. 4-19, and the sharp

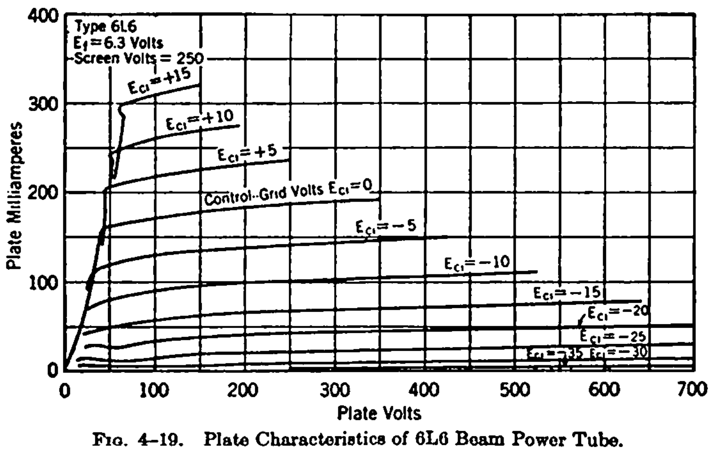


FIG. 4-19. Plate Characteristics of 6L6 Beam Power Tube.

break on each of the curves is the point at which plate voltage and potential minimum are equal.

Figure 4-18 also shows the accurate focusing of the electron beam, causing it to pass between the turns of the screen grid. When this is done, the screen current is much reduced as compared with earlier tubes, and more output power is available for a given amount of peak cathode emission.

Dual-purpose Tubes. For reasons of economy or convenience several functions that would otherwise be accomplished by two or more tubes may be handled by a single multipurpose tube. Such a tube may consist of the elements of two or more tubes all mounted within a single envelope, each unit acting independently of the others, or it may be a combination that depends for its operation on interaction of some sort between the several elements.

An example of the first class is the twin triode, such as the 6C8–G, which contains all the elements of two entirely distinct triodes, except that a single heater is used for both cathodes. It may be used in any circuit application calling for two similar triodes. Another example is the 3A8–GT, which contains a diode, a triode, and a pentode; but in this case the cathode, which is of the filamentary type, is common to all three. An added feature making for flexibility is that a filament tap is brought out, so that the tube may be operated at either 1.4 or 2.8 v for filament heating, by using the two halves in parallel or series respectively.

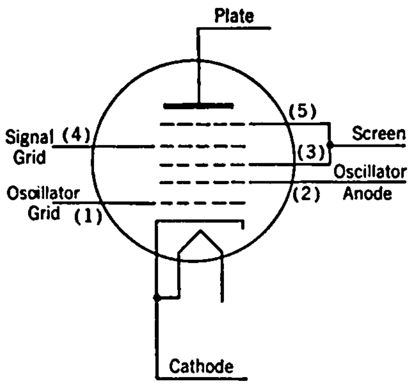
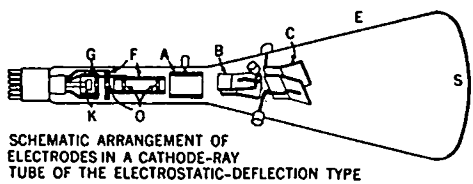


Fig. 4–20. Pentagrid Converter.

The second class of multipurpose tubes can be illustrated by the 6A8, called a pentagrid converter. As seen in Fig. 4–20, it contains five grids, which are referred to by number counting from cathode towards plate. In a typical application, No. 1 grid and No. 2 grid serve as grid and anode respectively of a triode oscillator. Grids No. 3 and No. 5 are tied together and serve as a screen, shielding No. 2 and No. 4 from each other and from the plate. Grid No. 4 receives the incoming signal, and its A.C. voltage modulates the electron stream passing from the oscillator section to the plate. The plate current is then a combination of the effects of the oscillator voltage and the signal voltage, but the only type of coupling between the two sources is by the electron flow; electrically the two circuits are isolated by the shielding action of the two screen grids. Among other tubes depending on electron coupling are the 6L7, the 6J8, and the 6K8, which should be studied in the tube manuals.

Cathode-ray Tubes. The fact that electrons after leaving the cathode can be focused into a narrow beam that can be deflected by electric or magnetic fields leads to the development of the cathode-ray tubes, used chiefly in oscillographic applications. Such a tube is shown in Fig. 4–21. In this figure, *K* represents the cathode, with heater not shown; *G* serves as the control grid, although it does not resemble in form the grids used in amplifiers; *F* is referred to as anode No. 1, or the focusing anode; *A* is anode No. 2, or the accelerating anode; *B* and *C* are two pairs of deflecting plates; *S* is a screen deposited on the inner surface of the glass envelope, and composed of a fluorescent substance that emits a luminous glow when bombarded with electrons of high velocity.



Courtesy of R.C.A. Manufacturing Co.

Fig. 4–21.

In operation, electrons emitted by the cathode are accelerated by anodes *F* and *A*, and by virtue of the apertures in the various tube elements and the form of the electric field about *F* and *A* the electrons are constricted into a narrow beam or pencil along the axis of the tube. The degree of concentration or focusing is controlled chiefly by the potential of *F*, the first anode. After passing through anode *A*, the electrons proceed at constant velocity to the screen, since there is no further accelerating field, and collide with screen *S* in a small spot at the center of the field, causing a glow to appear there.

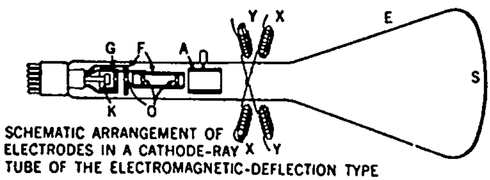
If an alternating voltage is applied between the pair of plates *B*, the electrons in the beam will be attracted to the plate which is positive at the moment, and repelled by the one which is negative. The beam is therefore deflected up and down as the voltage varies, and the luminous spot moves correspondingly on the screen. The other pair of plates, *C*, is arranged at right angles to the first, and any voltage across these plates will produce horizontal deflections, to the left or right depending on polarity.

This tube forms an extremely versatile tool for investigating electrical phenomena. It may be used for examining wave forms of currents or voltages, for comparison of frequencies, for obtaining volt-ampere characteristics, and for many other purposes. Because the electrons have so little mass, the beam will respond at much higher frequencies than any other indicating device available, and the range of applications is practically unlimited.

Figure 4–22 shows another type of cathode-ray tube, in which deflection is produced by horizontal and vertical magnetic fields set up by currents flowing in the deflecting coils *X* and *Y*. This type of deflection control is better adapted to tubes using very high accelerating voltages, and is commonly used in television receivers.

Instead of focusing the electron beam by electrostatic means as described above, it is possible to accomplish the same purpose by the use of a magnetic field directed along the tube axis. Such a field is easily produced by means of a coil surrounding the neck of the tube, approximately at the location of F in Fig. 4–22. The anode F may then be omitted, and this simplification is again advantageous with tubes using very high accelerating potentials.

A variety of fluorescent materials is available for the formation of the screen S. The color of the glow varies with the material, and this may be a matter of some importance. For oscillographic use, a screen giving a bright green color is usually employed, since this color is easily obtained and it has a good visual quality. For photographic use a deep blue glow is more useful, on account of its greater actinic power. By using a suitable mixture of phosphors, a glow approaching white in color may be obtained, and this is preferred for television screens, since the pictures present a more pleasing appearance.



Courtesy of R.C.A. Manufacturing Co.

Fig. 4–22.

When the electron beam moves away from a particular spot on the screen the glow disappears, but not instantaneously. Various fluorescent substances exhibit different rates of decay of brightness after the excitation is removed, and tubes are now available with either long-persistence or short-persistence screens. Those with long persistence are useful for the examination of extremely brief phenomena, as the glow remains long enough for visual observation. For general oscillographic use, however, a screen with short persistence is to be preferred.

Electron-ray Tubes. The ability of an electron stream to produce fluorescence is utilized in another class of tubes to indicate the presence or the magnitude of a voltage on one of the electrodes. These tubes are referred to as *electron-ray tubes*, and typical examples are the 6E5 and the 6G5. These tubes have a fluorescent screen bombarded by electrons from the cathode, and a control electrode deflects the electrons to produce a wedge-shaped shadow in the general glow covering the screen. The angular width of this shadow depends on the potential of the control electrode, the available range being from 0° to 90°. The control electrode is attached to the plate of a triode unit contained in the tube envelope, and by using the triode as a D.C. amplifier the device becomes quite sensitive as an indicator of voltages applied to the grid. This type of tube is customarily used as a tuning indicator on radio receivers, but it is also very useful as a null indicator in bridge measurements, and after calibrating the shadow angle it may even be used as a voltmeter. The 6G5 differs from the 6E5 in having a remote cutoff, and it can therefore handle a wider range of signal strength

Review Questions and Problems. *1.* What are the relative advantages and disadvantages of tungsten, thoriated tungsten, and oxide-coated filaments?

2. Why is it common practice to connect the return circuits from the other tube elements to the center tap of the filament transformer?

3. What limits the current flow in the lower regions of Fig. 4–1? In the upper regions of Fig. 4–3?

4. What three factors limit the output obtainable from a tube? What effect does each one of these have?

5. What is the fundamental structural difference between a diode and a triode? In what manner does this difference affect the operating characteristics of the two tubes?

6. From the mutual-characteristic curve for $E_b = 200$ v on Fig. 4–8, determine values of mutual conductance (g_m) for various values of plate current. Plot a curve of g_m against I_p.

7. Determine the A.C. plate resistance (r_p) for the triode plate characteristics shown in Fig. 4–8, for a grid voltage of -10.0 and for two values of plate current.

8. From the family of plate characteristics for the 6J5 triode, Fig. 4–9, determine r_p, μ, and g_m for a quiescent operating condition of $E_g = -8.0$ v and $E_b = 240$ v. Compare the value obtained for μ with the product of g_m and r_p.

9. On the plate characteristics for the 6J5 triode, construct load lines for a load resistance of 50,000 ohms and the following values of E_{bb}: 160, 240, and 320 v. Repeat the above procedure for load resistances of 40,000 ohms and 20,000 ohms (note that all the load lines for a given value of load resistance are parallel to each other).

10. Construct a load line for 50,000 ohms load resistance through the quiescent point considered in Problem 8. Using this load line, determine the voltage amplification of the tube. Using the equivalent-circuit diagram and the tube constants determined in Problem 8, determine the voltage amplification again and compare it with the value obtained above.

11. Explain completely the shape of the tetrode plate characteristics, shown in Fig. 4–14, for plate voltages less than that indicated by point D.

12. Determine the values of R_s and C, in Fig. 4–15, required to obtain 100 v at the screen from a 250-v source, if the normal screen current is 0.8 ma, the lowest operating frequency is 100 cycles, and the effective A.C. resistance in the tube between the screen and cathode is 70,000 ohms.

13. Trace the development of the multi-element high-vacuum tube from the triode through the tetrode and to the pentode. Give the reasons for and the effects of each added element.

14. From the mutual-characteristic curve for the 6J7 pentode, Fig. 4–17, determine g_m at the following values of grid voltage: 0.0, −1.0, −2.0, −3.0, −4.0, and −5.0. Repeat the above procedure for the 6K7 tube and the following values of grid voltage: 0.0, −1.0, −2.0, −3.0, −5.0, −7.0, −10.0, −15.0, and −20.0.

Plot curves of g_m against E_g for the two tubes on the same sheet of graph paper and compare them.

15. Determine the voltage amplification for the 6J7 tube operating through a quiescent point of $E_b = 240$ v and $E_g = -3.0$ v for a load resistance of 50,000 ohms. Compare the voltage amplification obtained here with that for the 6J5 tube of Problem 10.

16. In what different manners do the beam power tubes and the pentodes achieve the same results with regard to secondary emission?

17. Determine the pattern produced on the cathode-ray screen if a sine wave of voltage is supplied by the transformer secondary to the deflection plates of a cathode-ray tube, as shown in Fig. 4–23.

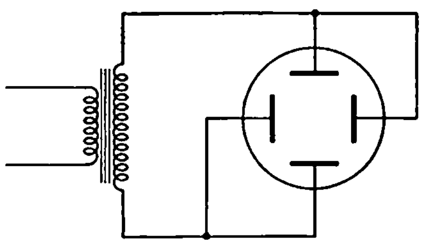


Fig. 4–23.

CHAPTER 5

Rectified Power Supplies

One of the important uses of diodes has already been referred to in Chapter 4—that of providing direct currents and voltages from an A.C. source. The tube in such service is known as a *rectifier*, and some of the properties of rectifiers and associated filter circuits will be considered in this chapter.

Half-wave Rectifier. The simplest form of rectifier circuit is that shown in Fig. 5–1, and the wave form of the current in the load resistance is shown in Fig. 5–2. The tube permits current to flow when its plate is positive with respect to its cathode, but none flows when the plate is at a negative potential. If voltage drop within the tube is neglected, the current will consist of unidirectional pulses having the form of half sine waves, and the average current throughout the cycle, or the D.C. component, will be $1/\pi$ or .319 of the peak value. Since the peak anode current is limited for any particular tube by the emitting power of the cathode, this relation determines the maximum load current that can be supplied.

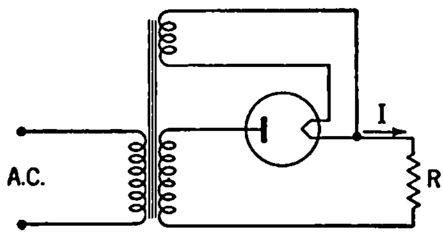


FIG. 5–1. Half-wave Rectifier.

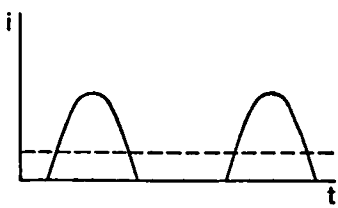


FIG. 5–2. Output of Half-wave Rectifier.

The maximum voltage across the tube occurs in the idle or nonconducting half cycle, and its value is the same as the peak or crest value of the transformer secondary voltage. This factor fixes the maximum load voltage that can be supplied by a given tube.

Full-wave Rectifier. The wave form of the rectified load current can be improved by making use of both positive and negative halves of the A.C. cycle, and one way of doing this is shown in Fig. 5–3, with the corresponding wave form of load current in Fig. 5–4. The tube has two separate anodes and one cathode. Typical tubes are the

5T4 and 5Z3. The transformer secondary is provided with a center tap. On tracing through the circuit it is seen that in each half of the cycle, one half of the secondary winding and one rectifier anode carry current, the other being idle. The D.C. component of current is $2/\pi$ or .638 of the peak value.

Besides giving a better wave form than the half-wave rectifier, the full-wave circuit has the advantage of symmetrical action in the transformer. In the half-wave circuit, the secondary current always flows in the same direction around the transformer core, and thereby produces a D.C. component of flux in the core. The A.C. flux set up by the primary current is superimposed on the D.C. flux, and to avoid oversaturation the transformer must be made larger than would be necessary if no D.C. component occurred.

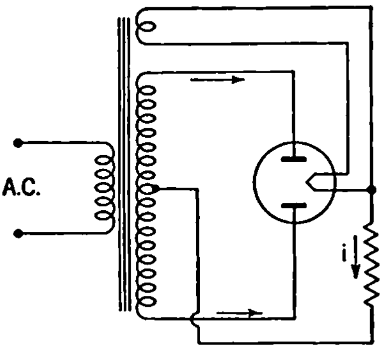


Fig. 5-3. Full-wave Rectifier.

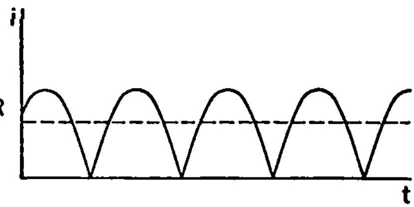


Fig. 5-4. Output of Full-wave Rectifier.

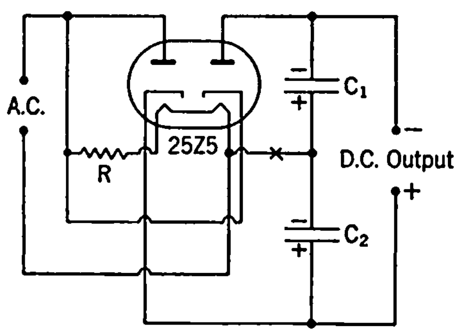


Fig. 5-5. Voltage-doubler Circuit.

Voltage-doubler Rectifier. It is possible to omit the transformer shown in Figs. 5-1 and 5-3 and still obtain a D.C. voltage high enough to serve as B supply by use of the voltage-doubler circuit of Fig. 5-5. The heater is supplied directly from the A.C. line through resistor R, which may include heaters of other tubes in the set. When the upper A.C. terminal is positive, current flows through the left anode into condenser C_2, charging it to peak line voltage. In the next half cycle current flows in the other anode, charging C_1 to the same voltage. The two condensers are in series with respect to the D.C. output terminals, and the sum of their voltages is available. When a load is connected to the output terminals, the load current is drawn from the condensers, reducing their voltage between charging intervals. The larger the condensers, the less the decrease in voltage, and therefore the better the voltage regulation. It is common practice to use condensers as large as 40 μf in this circuit.

With large condensers, the charging current attains high peak values and flows for only a brief time. To avoid excessive emission currents, a protective resistor may be inserted at the point marked ×, which is in series with each condenser as it is being charged.

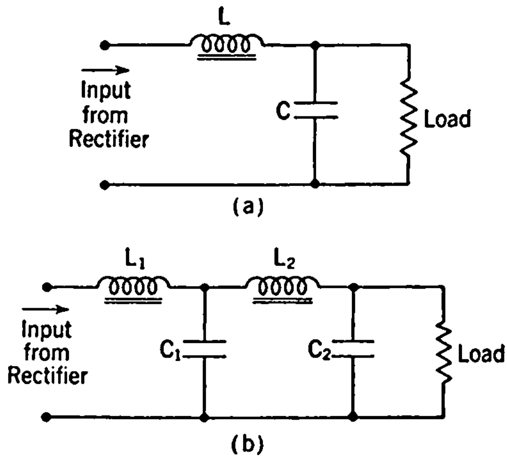


FIG. 5–6. Choke-input Filter Circuits: (a) Single-section, (b) Two-section.

Filter Circuits. The output wave forms shown in Figs. 5–2 and 5–4 are entirely satisfactory for many applications, such as the operation of relays, battery charging, and so forth, but they are not smooth and continuous enough to be useful for B-voltage supply of amplifiers and radio receivers. Service of this sort requires that the supply voltage be practically pure D.C., with very little *ripple* superimposed upon it.

Smoothing of the rectified A.C. voltage is accomplished by the use of filter circuits composed of inductance and capacitance, or resistance and capacitance. Figure 5–6 shows single-section and two-section filters of the choke-input type. The rectified current in flowing through the inductance L encounters a high reactance at ripple frequency, but very little resistance to the D.C. component, and as a result the fluctuations are greatly reduced. The condenser C in parallel with the load helps still more in this direction by absorbing most of the remaining fluctuations in current, since its reactance at ripple frequency is less than the load resistance. If the reduction in ripple is still not sufficient, another section of filter may be added, as shown in Fig. 5–6b.

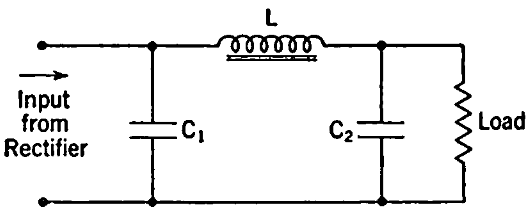


FIG. 5–7. Condenser-input Filter.

Another form of filter circuit is shown in Fig. 5–7. This is known as the condenser-input filter, since the condenser C_1 is supplied directly by the rectifier. In operation, the condenser C_1 is charged to the peak voltage available from the rectifier, and this charge is withdrawn gradually by the load current. Fluctuations in current and voltage are smoothed out by L and C_2, as in the choke-input filter. No further current is supplied by the rectifier until its voltage is again higher than that remaining on C_1. This operation is shown in Fig. 5-8.

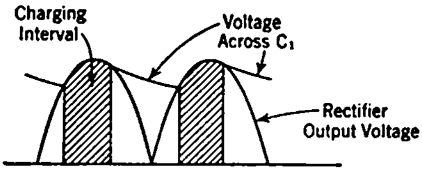


FIG. 5-8. Operation of Condenser-input Filter.

In comparing the two types of filters, it is seen that rectified current flows continuously in the choke-input filter, whereas it flows for only a brief part of each cycle in the condenser-input circuit. For the same D.C. load current, the peak anode current in the rectifier will therefore be much larger with condenser input than with choke input to the filter. Rectifiers for supplying large amounts of load current are commonly provided with filters using choke input, for this reason.

Another comparison between the two circuits concerns the voltage regulation, or the variation of output D.C. voltage with load current. By referring to Fig. 5-8, it can be seen that if the load current is increased, the voltage across C_1 will fall more rapidly, and the average or D.C. voltage will decrease. This effect is not present in the choke-input circuits of Fig. 5-6, and consequently the voltage regulation is better with choke input than with condenser input. It should be noted, however, that the input inductance loses its effect if the current through it is too small, and when this takes place the circuit behaves very much like that of Fig. 5-7. The approximate point at which this lower limit of load current is reached, in the case of a single-phase full-wave rectifier operating at 60 c.p.s., is where $L = R/1{,}200$. Here R is the D.C. load resistance in ohms, and L is the inductance of the filter choke in henries. Thus, if a rectifier is to furnish 400 ma at 2,000 v D.C., the apparent load resistance is

$$R = \frac{2{,}000}{.40} = 5{,}000 \text{ ohms},$$

and the minimum size of filter choke will be

$$L = \frac{5{,}000}{1{,}200} = 4.2 \text{ henries.}$$

If, however, the load current is likely to fluctuate between 400 and 100 ma, this choke will not be suitable, for at the lower current, the D.C. resistance has become 20,000 ohms, and the necessary inductance changes to

$$L = \frac{20{,}000}{1{,}200} = 16.7 \text{ henries.}$$

A choke of this size or larger will provide satisfactory operation over the entire range of load currents. So-called "swinging" chokes, having large inductance at low current and considerably smaller inductance at maximum current, are available for such service, at appreciable saving in weight and cost.

Still another comparison may be made between choke-input and condenser-input filters. In the case of the filter with choke input, the D.C. output voltage is approximately equal to the *average* rectified voltage of the supply. With condenser input, however, the D.C. voltage approximates the *peak* rectified voltage at light loads, and the amount of decrease with load depends on the size of the first condenser, C_1. More output voltage is therefore available for the same A.C. supply, especially if a large input condenser is used, and for this reason most of the power supplies used in radio receivers and small amplifiers are of the condenser-input type.

Filter circuits of resistance and capacitance were referred to above. A typical example of this type of filter has already been discussed in connection with the screen supply shown in Fig. 4–15, and for use with rectifiers the *R-C* filter is very practical and economical wherever the current drain is small, as in the case of cathode-ray oscillographs, photocells, vacuum-tube voltmeters, and so on.

Rectifier Tubes. Two broad classes of tubes are used in rectifier service, the high-vacuum and the gas-filled. In the high-vacuum tube, emission is controlled by space charge, and increase of current through the tube is accompanied by increase of anode potential, that is to say, voltage drop in the tube. In the gas-filled tube, ionization of the gas takes place, and consequently the voltage drop across the tube can never greatly exceed the ionization potential of the gas. Most tubes of this type use mercury vapor as the gas, and for these the tube drop is approximately 15 v, independent of the current flowing. Figure 5–9 shows current-voltage characteristics of the 5T4, a high-vacuum rectifier, and of the 83, a mercury-vapor rectifier, which have approximately the same maximum ratings.

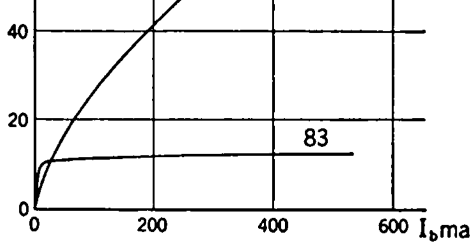


Fig. 5–9. Plate Characteristics of Rectifier Tubes.

In rectifiers employing mercury-vapor tubes, it is important that the filament or cathode be brought to full working temperature before the plate voltage is applied. If this is not done, electron emission will be limited and the voltage drop across the tube will be excessive. Greater voltage drop causes increased bombardment of the cathode by positive ions, as described in Chapter 4, and when the voltage drop exceeds 22 v this bombardment is sufficiently intense to cause disintegration of the cathode surface.

Mercury-vapor tubes should not be used with condenser-input filters unless a protective resistance is placed in series with the input condenser, to limit the peak charging current.

Regulated Power Supplies. Even though the output of a rectifier-filter be satisfactorily smooth and free from ripple, there may be some fluctuation of voltage due to variations in the A.C. supply, or in the load itself, and for some applications these fluctuations cannot be tolerated. Figure 5–10 shows one method of obtaining a stable value of D.C. voltage, depending on the use of a glow-discharge tube such as the VR150/30. This tube has the property that its anode potential remains practically constant at 150 v over a wide range of currents up to 30 ma. The size of the resistor R is selected so as to provide the necessary amount of drop due to the combination of load current and current in the glow tube. Thus, if the supply voltage is 250, and the load current is 20 ma at 150 v, one should allow for a glow-tube current of 20 ma. The resistor R is then called on to provide a drop of 100 v (from 250 to 150), while carrying a current of 40 ma. The required resistance is $100/.040 = 2{,}500$ ohms. Fluctuation of the supply voltage will cause more or less current to flow through the regulator tube, but will have practically no effect on load voltage or current.

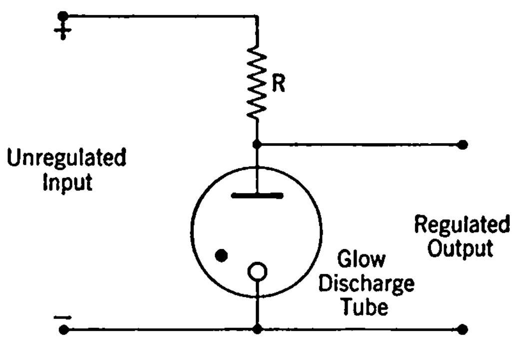


Fig. 5–10. Regulated Voltage Supply.

Triode Regulators. A still better control can be obtained by means of circuits similar to the one in Fig. 5–11. The load current flows through the 2A3 tube, and control is obtained by altering the conductance of this tube and thereby its voltage drop. This effect may be seen by tracing through the operation of the circuit. The cathode of the 6F5 tube is maintained at a constant potential by the VR75/30 glow tube. If the output voltage of the supply tends to rise, it will cause the grid potential of the 6F5 tube to rise, and therefore its plate current also. This rise produces an increased drop in R_g and decreases the grid potential of the 2A3 tube. For the given amount of load current, this decreased potential requires more voltage drop between plate and cathode of the 2A3 tube, which results in restoring the voltage at the output terminals to its original value.

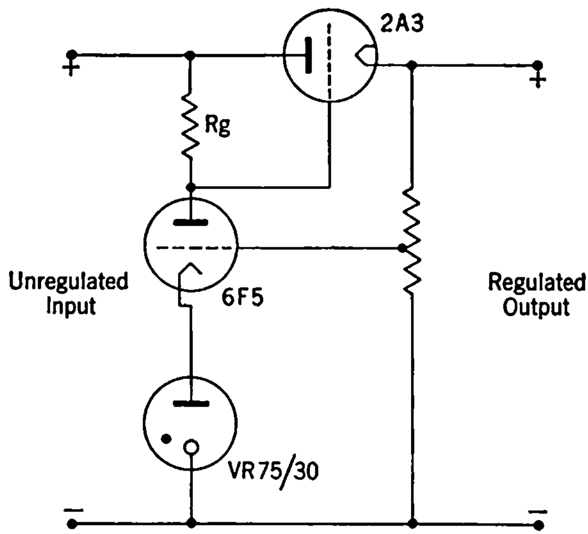


FIG. 5–11. Electronic Voltage Regulator.

Regulator circuits of this type are extremely effective over a considerable range of output currents, and no difficulty is found in holding output fluctuations below .02 v per milliampere. In addition to controlling the D.C. voltage, the regulator is capable of absorbing an appreciable amount of ripple voltage, so that the cost and weight of the filter complement can sometimes be reduced considerably in comparison with an unregulated supply.

Review Questions and Problems. *1.* What is the maximum inverse voltage applied to the tube in a half-wave rectifier circuit operating from a transformer with secondary voltage of 300 (R.M.S.)?

2. What is the maximum inverse voltage applied to the tube in a full-wave rectifier circuit operating from a transformer with a secondary voltage of 300 (R.M.S.) from the center tap to either of the outside leads?

3. Trace the operation of the voltage-doubler circuit through one complete cycle.

4. What are the relative advantages and disadvantages of choke- and condenser-input filters?

5. Determine the minimum amount of inductance that the choke in a choke-input filter should have if it is to operate with a single-phase full-wave rectifier and is to supply a minimum current of 50 ma at 400 v D.C.

6. A 5-henry choke is available for the construction of a choke-input filter to be used with a single-phase full-wave rectifier. The rectifier supplies a load which may draw a current varying from a minimum of 0 ma to a maximum of 50 ma at a D.C. voltage of 1,000. What is the maximum amount of bleeder resistance that should be used at the output of the filter?

7. What is the advantage of using a "swinging" choke in a choke-input filter?

8. What danger is incurred if the plate supply on a mercury-vapor rectifier tube is turned on before the cathode has reached its full operating temperature?

9. Explain completely the operation of a typical triode voltage-regulator circuit.

CHAPTER 6

Sound and Its Electrical Transmission

Nature of Sound. The physical effect which is interpreted by the ear as sound consists of a pressure wave in air, that is, a succession of variations in pressure, above and below the normal static pressure of approximately 15 pounds per square inch. These pressure variations travel outward from the source of the sound at a constant velocity of 1,100 feet per second, varying slightly with temperature and barometric pressure of the air. The pressure variations for ordinary sounds are only a very small fraction of the total air pressure; thus a sound which is just barely audible will be produced by an excess pressure of $1/10^9$ of the static pressure, and a sound so intense as to cause the sensation of pain in the ear requires only an excess pressure equal to $1/10^4$ of the average.

The *frequency* of the sound wave is perceived as the musical pitch of the tone, and the range of frequencies included in the field of audible sounds extends from 20 c.p.s. to 20,000 c.p.s. For most purposes it will be satisfactory to cover the range from 50 to 10,000 c.p.s., and in special cases a still more restricted band will be adequate.

The *wave form* of the sound depends upon the source which produced it, and for most voices and musical instruments it is quite complex. Just as in the case of electrical waves, the sound wave may be decomposed into a fundamental component and a succession of harmonics. In studying the properties of any electrical system for the transmission or reproduction of sound, the response due to each of the harmonic components present in the sound wave may be evaluated separately, and the total response will then be the combination of all these. This procedure is justified by the principle of superposition, as outlined at the close of Chapter 3.

Distortion. In the process of electrical transmission or reproduction of sound, the output wave form generally deviates from the original form in some degree, and this deviation is classed as distortion. There are three principal types of distortion, and these will be considered separately.

Frequency distortion. Every transmission system has a limited range of frequencies over which it can operate, and the response of the system within this range may be different at different frequencies. Two typical frequency-response curves are shown in Fig. 6–1, and it will be noted that curve A shows a much more uniform response over a range of frequencies than does curve B, besides covering a *wider* range. On the other hand, curve B shows a much larger response in the middle range of frequencies, and for this reason might represent a more desirable characteristic for certain applications. Variation of response with frequency is referred to as *frequency distortion.*

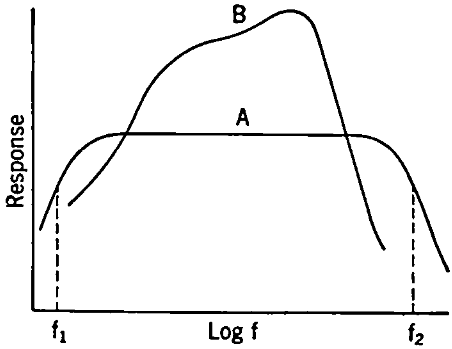


FIG. 6–1. Frequency Distortion.

It is frequently convenient to refer to the useful *band width* of a system, and for a characteristic such as A of Fig. 6–1, this is usually taken as extending between the frequencies f_1 and f_2, at which the response is .707 of the uniform response in the middle range.

It will be noted in the figure that the horizontal axis is plotted in terms of the logarithm of frequency. This logarithmic plot is used partly to avoid the crowding which would occur at low frequencies if a uniform scale were used, but also because our ability to distinguish frequencies of sound shows a logarithmic characteristic, and a truer picture of the performance of the system is obtained by this type of curve.

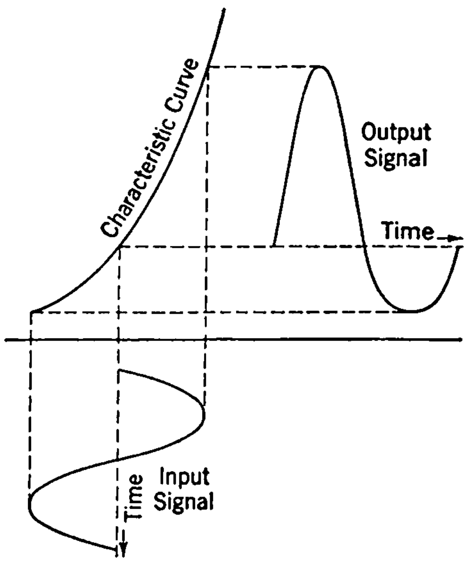


FIG. 6–2. Nonlinear Distortion.

NONLINEAR DISTORTION. A system or device is nonlinear when the relation between input and output is given by a curved characteristic, as shown in Fig. 6–2. As will be seen from the figure, the output corresponding to a pure sine wave of input will be distorted, and the result will be the formation of new frequencies not originally present. In the case illustrated, the new frequencies are the harmonics in the output wave, and they are therefore all multiples of the original frequency. However, if the input signal contains more than one frequency, as is always the situation in practice, the output will also contain frequencies equal to the sums and differences of the input frequencies, and of integral multiples of these frequencies.

DELAY OR PHASE DISTORTION. Signals in passing through a transmission system always encounter a certain amount of delay, and if the delay time is different for different frequencies, the result will be an alteration of wave form. This is because harmonics in the output wave will appear at different phase angles with respect to the fundamental than they occupied in the input wave, even though their amplitudes may be unaltered. Distortion of this type is known as *phase distortion.* It is not of much importance in sound transmission, since the relative phase of a harmonic makes practically no difference in the quality of a sound as perceived by the ear. In other applications, such as oscillography and television, however, it takes on considerable importance.

Microphones. A microphone is a device that transforms sound energy into electrical energy. In most types of microphone, the sound pressure acts upon a thin plate or diaphragm, setting it into vibration, and this mechanical motion is then utilized to produce electrical effects. The chief types are described below. In addition to these, available microphones include the condenser type and various directional types.

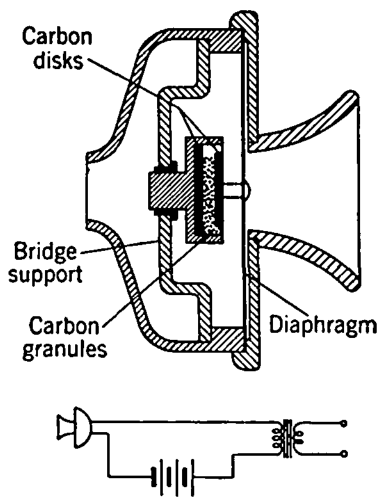


FIG. 6-3. Carbon Microphone.

THE CARBON-GRAIN MICROPHONE. One of the earliest types of microphone, and the one still most commonly used, depends for its action on the fact that the electrical resistance between carbon granules in contact with each other varies with the contact pressure. Figure 6-3 shows a simplified sectional view of a single-button carbon microphone, such as is used in telephone sets. A small brass cup contains two polished carbon disks, one fastened solidly in the cup and the other attached to the diaphragm. The space between the disks is partly filled with carbon granules, and as the diaphragm vibrates in response to the sound waves striking it, the varying pressure on the granules causes changes in the electrical resistance between the buttons. The microphone circuit is shown in the same figure, and from this it is seen that variation of microphone resistance will alter the current through the transformer primary, and so will set up induced voltages in the secondary.

By proper choice of diaphragm stiffness and mass, the moving system can be made to resonate near the middle of the speech range of frequencies. When this is done, the electrical output is large enough to operate a receiver over a considerable length of line without requiring amplification. The frequency response is then not very uniform, although it is entirely adequate for speech reproduction.

By using a very light diaphragm, tightly stretched, the frequency response is greatly improved, but at the expense of sensitivity. Carbon microphones with this type of construction, and having two *buttons*, or carbon cells, have been used extensively in broadcast practice.

THE CRYSTAL MICROPHONES. Another type of microphone, widely used in public-address systems, depends for its action on the *piezo-electric effect* possessed by certain crystals, in this case Rochelle salt. The term piezoelectric effect refers to the fact that when pressure is applied on the crystal in the proper direction, electrical potentials are produced between opposite faces of the crystal. The *sound-cell* type of microphone contains an assemblage of small crystals of this type, so connected that their piezoelectric potentials are in series. The sound falls on the crystals and vibrates them directly. The electrical output is quite small, but the frequency range and uniformity of response are excellent.

In another type of crystal microphone, a metal diaphragm is coupled mechanically to a crystal of Rochelle salt in such a way that vibration of the diaphragm causes a twisting of the crystal, and thereby the generation of a voltage at the terminals. This type has much greater output than the sound cell, but the frequency response is somewhat limited by the inertia and stiffness of the diaphragm and the associated driving members.

THE ELECTRODYNAMIC MICROPHONE. Several types of microphone depend for their action upon the induction of voltage in a conductor moving in a magnetic field. The *moving-coil microphone* contains a small coil attached to a diaphragm, so arranged that when the diaphragm vibrates the coil moves back and forth in a radial magnetic field, and thus generates the output voltage. By careful design of the moving element, and by making use of air-chamber resonance, it is possible to obtain a nearly uniform response from 40 c.p.s. to 10,000 c.p.s. An incidental advantage is that the output impedance of the microphone is low, and the microphone cable is less sensitive to hum pickup than in the case of the crystal microphones.

RIBBON MICROPHONE. In this type the moving element is a very thin and flexible aluminum ribbon, upon which the sound waves act directly. It vibrates in a transverse magnetic field, and generates an electromotive force which appears between the two ends of the ribbon. The ribbon impedance is so low that a small step-up transformer is included in the microphone mounting to raise the impedance to a level suitable for transmission over a line. Most ribbon microphones respond to air particle velocity in the sound wave, rather than to sound pressure, and they are referred to as *velocity microphones*. They can be made to have excellent frequency characteristics.

Reproducers. A reproducer is a device for converting electrical energy into sound. As in the case of microphones, this transformation usually involves an intermediate mechanical motion.

TELEPHONE RECEIVERS. The ordinary telephone receiver is the most commonly known acoustic device. A modified form, the watch-case type used in operators' headsets, is shown in section in Fig. 6–4. Two small coils are wound on soft iron pole pieces, which are attached to the poles of a permanent magnet. The pole pieces attract the steel diaphragm with a steady pull due to the permanent magnet, and with an alternating force due to the voice currents flowing in the coils. The diaphragm is set into vibration, and sets up sound waves in the air in contact with it. The permanent magnet is necessary to avoid distortion in the output, as will be apparent when it is noted that the diaphragm would be attracted *twice* in each cycle if only the A.C. attraction were present.

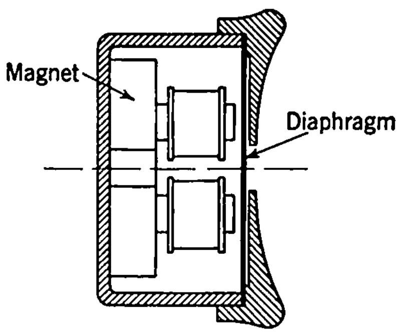


FIG. 6–4. Telephone Receiver.

The ordinary receiver used with telephone instruments is wound for about 70 ohms resistance, and it has a definite resonance peak near 1,000 c.p.s., for the sake of sensitivity. By winding with many turns of fine wire, the sensitivity to weak currents can be greatly increased, and such receivers are very useful as indicators in A.C. bridges and for radio communication systems.

LOUDSPEAKERS. The commonest type of loudspeaker is shown schematically in Fig. 6–5. The moving coil, situated in a powerful radial magnetic field, carries the operating current. The reaction of the signal current with the magnetic field causes the coil to move back and forth along its axis. In this motion it carries with it the paper cone radiator. The cone is supported at its outer edge by a flexible suspension, and, at least at the lower frequencies, it moves as a rigid piston, without appreciable bending or deformation. The result is a very effective transformation of the electrical input into sound energy radiated from the surface of the cone.

The radiation from the rear surface of the cone is opposite in phase as compared with that from the front surface, and it is the function of the *baffle* shown in the figure to prevent these two effects from canceling each other. The baffle will be effective at any frequency for which the distance from the front of the cone, around the edge of the baffle, to the rear edge of the cone, is greater than a half wave length of sound. For example, at 100 c.p.s. the wave length is $\lambda = 1{,}100/100 = 11$ ft (see Chapter 9), and the distance from front to back of cone should not be less than 5.5 feet.

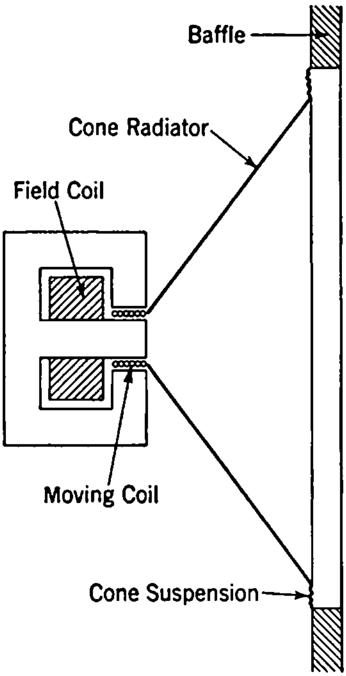


Fig. 6–5. Cone Type of Loudspeaker.

Telephone Circuits. Two-way operation is essential for satisfactory telephone service, and many of the problems of the industry arise from this fact. A simplified circuit for obtaining two-way operation is shown in Fig. 6–6. This is called a local-battery system because a separate battery is required at each end of the line, and in such a telephone system each subscriber must have a battery on his premises. The operation of the circuit is self-evident, and it is seen that speaking into either microphone will set up voice currents in both receivers. The transformers isolate the direct current required for the operation of the microphone, and also improve the efficiency by stepping up the voltage and reducing the current in the line.

Common-battery circuits. The obvious advantages of removing the batteries from the subscribers' premises to the central office led to the development of the *common-battery system*, using one large storage battery to

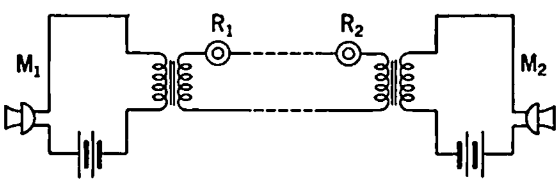


Fig. 6–6. Local-battery Telephone Circuit.

supply microphone current to all subscribers' sets. This led to a new difficulty, however, in that the *voice currents* of all circuits in use flow through the same battery, and because of its internal impedance there is a possibility that some of these currents will find their way into other circuits.

This effect is known as *cross talk*, and it is avoided by use of a transformer called the *repeating coil.*

This device, and its manner of use, are shown in Fig. 6–7. It will be recognized as essentially a one-to-one transformer with primary and secondary windings split at the battery. Other repeating coils, connecting other pairs of subscribers, may be tied in at the points *BC*, *B'C'*, and when so connected will offer extremely high impedance to flow of voice current from one channel to the other, but practically none to the flow of voice current in its own channel.

The talking circuit of the subscriber's set is different in the case of the

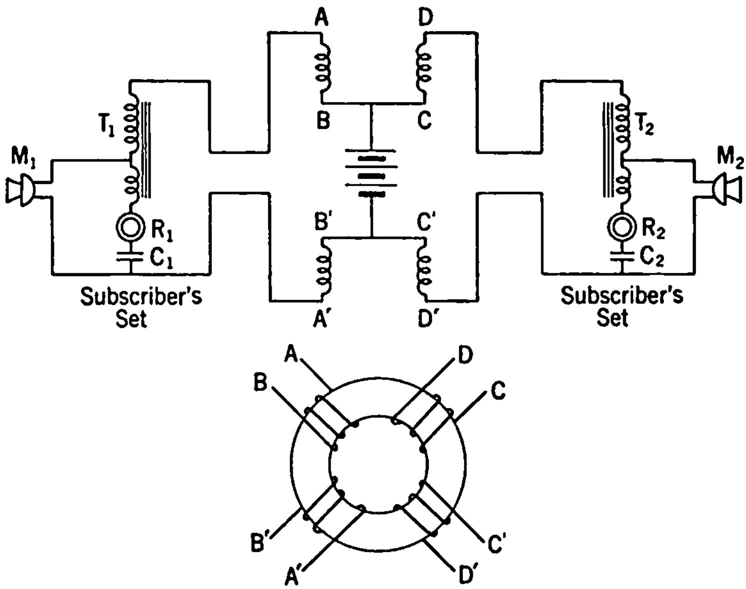


Fig. 6–7. Common-battery Telephone Circuit.

common-battery telephone. Transformers T_1 and T_2, usually called *induction coils,* and condensers C_1 and C_2, are connected in such a way that the direct current from the battery flows through the microphones M_1 and M_2 but not through the receivers R_1 and R_2. The induction coils are connected as *auto-transformers,* that is, the lower portion of the winding serves as primary and also as part of the secondary to step up the A.C. component of the microphone voltage.

Telephone systems require additional equipment for signaling the operator and subscribers and for switching connections between subscribers. Their discussion is beyond the scope of this book.

Telephone Lines. The transmission lines used in telephony are of two kinds, open-wire and cable. The open-wire lines are gradually being superseded by cable construction, either overhead or underground, for two principal reasons. In the first place, many more circuits can be accommodated—there is a standard cable containing more than 2,100 pairs of wires, and it is common practice to run 900 pairs in overhead cables on a single pole line; while a pole line carrying 50 pairs of open-wire construction would be a monstrosity. The other reason for preferring cable is that it affords much better protection against weather hazards and against electrical interference, both noise and cross talk.

THE DECIBEL. Losses in telephone lines and other equipment are stated in terms of a logarithmic unit, the decibel. It is defined as follows:

$$\text{loss in db} = 10 \log \frac{W_1}{W_2},$$

where W_1 is the input power and W_2 is the output power. For an amplifier, the output power exceeds the input power, and the loss given by the above expression becomes negative. Negative loss is referred to as *gain*, and thus

$$\text{gain in db} = 10 \log \frac{W_2}{W_1}.$$

The decibel is also used as a measure of the *amount* of power absorbed or furnished by any device, by comparing it with a standard power. This standard is referred to as *zero level*, and in telephone practice the power at zero level is 1 milliwatt. For example, if the power output of an amplifier is 4 watts, its power level expressed in decibels is

$$\text{output} = 10 \log \frac{4}{.001} = 10 \log 4{,}000 = 36 \text{ db}.$$

The decibel is a very important unit, and it is used extensively in all branches of electrical communications. The fact that it is logarithmic in nature makes it possible to obtain over-all effects resulting from a combination of lines, amplifiers, and other equipment by simply adding or subtracting their respective gains or losses. It is also true that the sensitivity of the ear is nearly a logarithmic function of sound intensity, so that a logarithmic unit is very appropriate. It turns out that a difference of sound level of 1 db is just barely perceptible to the average person, which shows further that the size of the unit is well chosen.

LOSSES IN TELEPHONE LINES. The smallest conductor used in standard open-wire telephone circuits is No. 12 B.& S. gauge, and a line of such construction will have a loss of approximately 0.06 db per mile. Cable circuits, on the other hand, make use of conductors not larger than No. 19, and a typical cable circuit with wire of this size will show a loss of about 1.0 db per mile. In this respect cable circuits are at a disadvantage as compared with open lines, especially for long distances, although recent improvements in the efficiency of telephone instruments have extended the useful range of cable circuits very materially.

LOADING. Losses in transmission circuits can be decreased by the insertion of *loading coils* at intervals in the line. These coils add series inductance to the line and provide a more favorable ratio of inductive to capacitive effect, particularly with cables. The result is that for a given power level the voltage is raised and the current lowered, just as the voltage is raised on a power-transmission line by use of transformers. Lower current results in smaller resistance losses and higher efficiency.

A limitation in the amount of loading is encountered from the fact that these loading coils, in conjunction with the capacitance of the intervening sections of the line, constitute a so-called low-pass filter which will not transmit frequencies above a certain *cut-off frequency*. This critical frequency is given very closely by the expression

$$f_c = \frac{1}{\pi\sqrt{LCd}},$$

where f_c is the upper limit of frequency transmitted, L is the effective inductance of each loading coil, C is the capacitance per mile of line, and d is the distance in miles between loading coils.

Examination of this equation shows that as L is increased to improve transmission, f_c is decreased unless the loading interval d is lowered in proportion. The loading interval has been standardized at 6,000 ft, and a material reduction of this spacing would involve a prohibitive cost. Because of greater emphasis on better speech quality, requiring transmission of the higher frequencies, recent trends have been toward a *reduction* in amount of loading on some circuits.

REPEATERS. In long-distance lines losses are so high that satisfactory operation becomes impossible without the use of amplification. Amplifiers for telephone service are known as *repeaters*, and they must of course function in both directions along the line. Repeater stations are ordinarily installed at 50-mile intervals along the line, except in the newer wide-band transmission systems, for some of which repeater spacing is as little as 5 miles. Because of the total number of stages of amplification in a long-distance transmission, the performance requirements of telephone repeaters are very severe. Any distortion is cumulative, and if present in appreciable amount would soon result in unintelligible speech. The demands of high-quality program transmission for broadcast networks are even more exacting.

One method for obtaining two-way repeater service is shown in Fig. 6–8. Two amplifiers are employed, one for each direction of transmission. Feedback and oscillation are prevented by the *hybrid coils*, which are essentially three-winding transformers of balanced construction. A signal traveling from west to east encounters the first hybrid coil, where part of the power enters the west-east amplifier by way of the center taps on the main windings. The remainder of the power is dissipated in other parts of the circuit and has no further effect. The amplified power is fed into the third winding of the other hybrid coil, and divides there into two equal portions, one passing out on the line to the east, the other into the *artificial network*. This network is constructed to have impedance properties equal to those

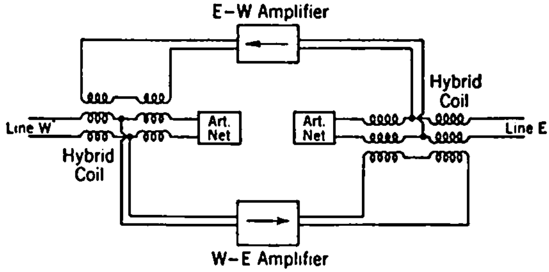


FIG. 6-8. Type 22 Telephone Repeater.

of the line over the entire frequency band, so that the combination of line and artificial network is equivalent to a balanced bridge. As a result of this balance, none of the output of the west-east amplifier reaches the input terminals of the east-west amplifier, and feedback around the loop is avoided.

It will be noted that power is diverted and lost at both hybrid coils. This loss is compensated for by raising the gain of the amplifier by an equivalent amount, which turns out to be 6 db.

Review Questions and Problems. *1.* Determine the wave lengths in air for sound waves of the following frequencies: 20, 50, 100, 1,000, 10,000, and 20,000 cycles per second.

2. Explain the difference between frequency distortion, nonlinear distortion, and delay or phase distortion.

3. Why are the curves in Fig. 6-1 plotted to a logarithmic scale?

4. Explain the operation of a carbon-grain microphone.

5. Why is it essential to have a permanent magnet in a telephone receiver?

6. Explain the operation of a typical loudspeaker of the type shown in Fig. 6-5.

7. Explain the operation of the circuit in Fig. 6-6.

8. A nonloaded telephone cable circuit using No. 19 conductors has an effective impedance (nearly pure resistance) of 450 ohms. If the power is being transmitted at a level of + 2.0 db, find the values of voltage and current present on the line.

9. Repeaters similar to Fig. 6–8 are installed at 50-mile intervals in a No. 19 cable circuit having a loss of 1.0 db per mile. The input level to the repeater is 0.0 db, and the amplifier gain is adjusted so that the input to the next repeater is the same. Find the voltage at the output of amplifier, if its load impedance is 5,000 ohms.

CHAPTER 7

Audio Amplifiers

Fundamentals of Amplifiers. In Chapter 6 it was shown how sound energy can be converted into electrical energy by means of a microphone, the electrical energy transmitted over wires, and then changed into sound by means of headphones or loudspeakers. Somewhere in this process it is generally necessary to amplify or increase the volume of the signal being transmitted. This amplification is done while the signal is in the electrical state by means of an audio amplifier or a repeater. Because energy cannot be created it is not possible to amplify the signal (that is increase its energy) without taking energy from somewhere else. In an electrical amplifier the varying signal voltage is used to control the energy output of some other source such as a battery or power supply and this greater controlled energy becomes the output of the amplifier.

An *audio amplifier* is one suitable for amplifying signal voltages whose frequencies lie within the audible range. If the amplification is uniform for all frequencies within this band the amplifier is said to be *flat* over the range of frequencies considered. In general the amplification tends to drop off for both the very high and very low frequencies and special precautions must be taken in the design of the amplifier to prevent this. A modern broadcast amplifier is usually designed to be flat from 30 to 15,000 c.p.s.

To obtain larger amplifications than can be secured with a single tube, an amplifier usually consists of several *stages*, the output of one stage being fed to the input of the next. An amplifier stage may be either a *voltage amplifier* or a *power amplifier*, and a complete audio amplifier usually consists of one or two stages of voltage amplification followed by a single power-amplifier stage.

A voltage amplifier is one designed to produce a large *voltage* amplification with very little power output. Such an amplifier could be used to furnish signal voltage to the grid of a succeeding stage, since when properly operated the grid circuit requires no appreciable power.

A power amplifier is designed primarily to supply a large amount of *power* to a loudspeaker or other power-actuated device. In this application the actual voltage step-up is of secondary importance and is usually sacrificed to improve the power-handling capacity of the stage.

Classification of Amplifiers. Audio amplifiers are also classified according to the operating conditions under which the tube works. The classifications in general use are Class A, Class AB, and Class B. A *Class A* amplifier is one in which the plate current flows continuously throughout the cycle of alternating voltage applied to the grid. The grid bias and alternating grid voltage are selected so that operation is confined to the linear or straight-line portion of the grid-voltage–plate-current characteristic curve. The shape of the output voltage wave will be similar to the input voltage wave on the grid of the tube.

A *Class AB* amplifier is one in which plate current flows for more than half but less than a complete electrical cycle.

A *Class B* amplifier is one in which the plate current flows for only one half of each cycle of the alternating grid voltage.

The subscript 1 may be used with the letter classification (for example, *Class* A_1) to indicate that the grid is not allowed to swing positive during any part of the cycle. The subscript 2 is used (for example, *Class* AB_2) to indicate that the grid does go positive for a fraction of the cycle.

These classifications are discussed more fully in the section on power amplifiers and in Chapter 11.

Resistance-capacitance-coupled amplifier. It was seen in Chapter 4 that a voltage variation applied to the grid of a tube would produce a variation in its plate current. If this varying plate current is made to flow through a resistor R, as in Fig. 7–1a, a varying voltage e_2 will be developed across the resistor, similar in all respects to the original voltage e_1, except that it may be many times larger. This larger voltage could then be applied to the grid of a second tube and still further amplification obtained. A practical circuit for connecting this second tube to the first is shown in Fig. 7–1d. In order to see why each of the components R_p, C and R_g is necessary for satisfactory performance, the operation of other simpler circuits will be considered.

Figure 7–1b represents the simplest form of coupling possible. The batteries E_f, E_{bb} and E_c are necessary to maintain the correct *operating voltages* on the tubes. It was seen in Chapter 4 that the plate and grid voltages of a tube must be selected so that it operates on a linear or straight-line portion of its characteristic (E_g-I_p) curve. This selection is necessary if the output voltage e_2 is to be a faithful reproduction of the input voltage e_1. In practice it is usually desirable to supply power for heating the filaments of the tubes from a common battery or other source of supply. If this were done in Fig. 7–1b by connecting together points a and c and points b and d, it is evident that the battery voltages E_{bb} and E_c would be shorted.

The arrangement in Fig. 7–1c gets around this difficulty by connecting the negative of both filaments to the negative of the B battery. However, this arrangement places the B battery between the grid and filament of the second tube and so puts a large positive voltage on the grid instead of the small negative voltage usually required. Therefore a large negative voltage must be put in series with the grid of the second tube by the battery E_c. This circuit is sometimes used in amplifiers built for amplifying D.C. voltages. It has the very serious disadvantage that the small D.C. voltage on the second tube is obtained as the difference of two comparatively large voltages and so a small percentage change in either of the large voltages will produce a large percentage change in the voltage on the grid. For example, if the plate voltage E_b were 90 v and the required grid voltage on the next

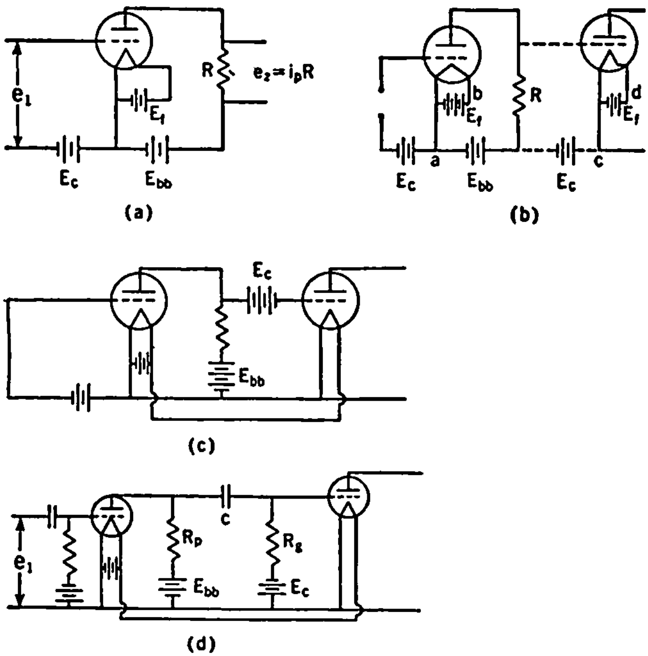
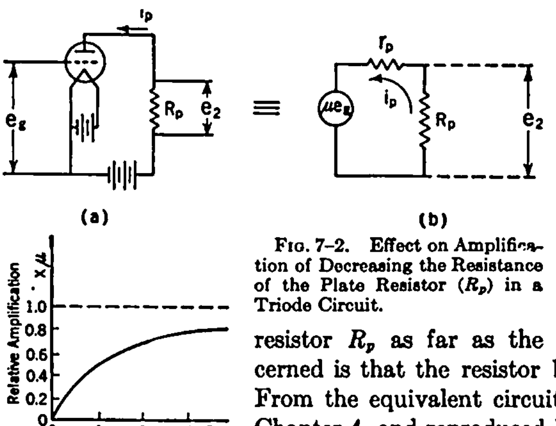


Fig. 7–1. Development of a Resistance-coupled Amplifier Circuit.

tube were $-4\frac{1}{2}$ v, the battery voltage E_c would have to be $-94\frac{1}{2}$ v. If now the voltage of E_b were to drop 5 v (that is, to 85) due to aging of the battery the voltage on the grid would increase from $-4\frac{1}{2}$ to $-9\frac{1}{2}$, while if the voltage E_c had dropped 5 v the grid voltage would have changed from $-4\frac{1}{2}$ to $+\frac{1}{2}$. The result is that an amplifier coupled directly as in Fig. 7–1c tends to be unstable and its operation may be affected too much by changes in supply voltages.

The connections of Fig. 7–1d are designed to overcome these and certain other difficulties. To prevent the voltage E_b from being applied to the grid of the next tube a condenser C, usually called a coupling condenser because it couples the two stages together, is used. However, this leaves the grid disconnected from the filament as far as direct voltages and currents are concerned. This condition of *floating grid* must not be permitted because any electrons from the filament that reach the grid have no way of leaking off. This situation allows a large negative charge to build up on the grid and the tube becomes inoperative. For this reason the grid-leak resistor R_g is used and the correct bias voltage E_c is applied in series with it. It will be noticed that the same condenser-resistor combination is used in the input of the first tube. This arrangement isolates the tube from any D.C. voltage that may be present along with the signal voltage and ensures that the correct bias voltage will be applied to the grid of the tube.

Having arrived at the circuit of Fig. 7–1d as one that will be suitable as an amplifier, the next problem is to determine what values should be used for the resistors and condensers. It is evident that each of these components has two separate functions to perform: the first is to assist in applying the correct D.C. or operating potentials to the tube and the second is to provide the best conditions for the amplification of the alternating or signal voltage. Often the two functions call for widely different values and a compromise value must be used.



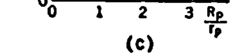


Fig. 7–2. Effect on Amplification of Decreasing the Resistance of the Plate Resistor (R_p) in a Triode Circuit.

The plate resistor. When a tube is operating as a voltage amplifier the only requirement on the plate resistor R_p as far as the signal voltage is concerned is that the resistor be as large as possible. From the equivalent circuit of a triode shown in Chapter 4, and reproduced here, it will be recalled that the tube acts like a generator having a voltage equal to μe_g in series with the plate resistance r_p of the tube. The way in which the voltage e_2 across R_p varies as the size of R_p is increased is shown in Fig. 7–2c. R_p is shown as so many times r_p, and e_2 is shown as a fraction of μe_g. When the resistance of R_p is very small nearly all of the voltage drop occurs across the resistance of the plate r_p and the voltage e_2 is small, while if R_p is made very large compared with r_p, then nearly all of the total generator voltage (μe_g) will appear across R_p, and e_2 will be nearly equal to μe_g, which is the maximum value it can have. From this it appears that R_p should be made very large, say about a hundred times as large as r_p.

But now consider the D.C. or operating voltage. The D.C. voltage E_b on the plate of the tube will be the battery voltage E_{bb} minus the D.C. voltage drop in the resistor. If R_p is made too large there is too great a voltage drop in this resistor and the resulting plate voltage will be too small. It would be possible to raise the plate voltage by increasing the battery voltage E_{bb}, but to do so might require several hundred to a thousand volts and this is not conveniently obtained. It is evident then that a smaller value of R_p would have to be used and a compromise value of 1 to 10 times r_p is generally chosen for a triode voltage amplifier, depending on whether r_p is high or low. For a power amplifier there are other considerations that dictate the value of R_p.

GRID-LEAK RESISTOR. In Fig. 7–1d not all of the signal voltage across R_p will be applied to the grid of the second tube. The reactance of the condenser C and the resistance R_g act together as a voltage divider and only that part of the voltage across R_g will be impressed on the grid. Hence, as far as the signal voltage is concerned, R_g should be as large as possible. When the circuit is properly operated there is no flow of grid current and so the resistor is not limited by the voltage-drop consideration as was the plate resistor. However, if the grid-leak resistor is made too large the effect of floating grid begins to appear, because any charge, either positive or negative, that collects on the grid cannot leak off fast enough. The actual size of the resistor required varies with the tube and usually lies in the range 0.1 to 10.0 megohms. The proper value is generally specified by the manufacturer of the tube.

THE COUPLING CONDENSER. The function of the coupling condenser is to keep the large positive voltage E_b off the grid and at the same time to offer minimum impedance to the signal voltage. This requirement indicates a large value for the capacitance, but again if C is made too large any charge that collects on the grid (and therefore charges this condenser) will require too long a time to leak off through the grid-leak resistor. It is evident that this effect depends upon both the size of the condenser and the size of the grid-leak resistor, since increasing either of them will increase the time required to discharge the condenser, or the *time constant* of the circuit as it is generally called. The time constant is given by the product of $C \times R$ where C is in farads and R is in ohms. It represents the time in seconds required to discharge the condenser to about one third (actually to 1/2.718) of the original charge. A typical value for a resistance-capacitance-coupled amplifier is .004 second, which could be obtained with a 1-megohm grid leak and a .004-μf condenser or a 0.5-megohm leak and a .008-μf condenser, or any other suitable combination.

Another practical consideration that enters into the selection of a coupling condenser is that it should have small leakage, that is, a high leakage resistance of the order of hundreds or thousands of megohms. If the leakage resistance is small, an appreciable amount of direct current will flow through it. Then the plate resistor R_p, the condenser leakage resistance, and the grid leak, all in series, will act as a voltage divider across the supply voltage E_{bb} and put a positive D.C. potential on the grid. As an example, if $R_p = 0.1$ megohm, $R_g = 1.0$ megohm and the leakage resistance is only 10 megohms, then for a plate supply voltage of 100 v there is a current of .009 ma through the circuit and this produces 9 v across the grid-leak resistor.

Frequency Response of a Resistance-coupled Amplifier. As was pointed out earlier, most applications of an audio amplifier require that it have fairly flat or uniform response over a wide range of frequencies. In other words, the amplifier must amplify all frequencies within this range approximately the same amount. The resistance-coupled amplifier is particularly suited to do this because its main impedances (the resistors) are independent of frequency over the audio range. At medium and high frequencies the reactance of the coupling condenser is so small, compared with R_g, that it

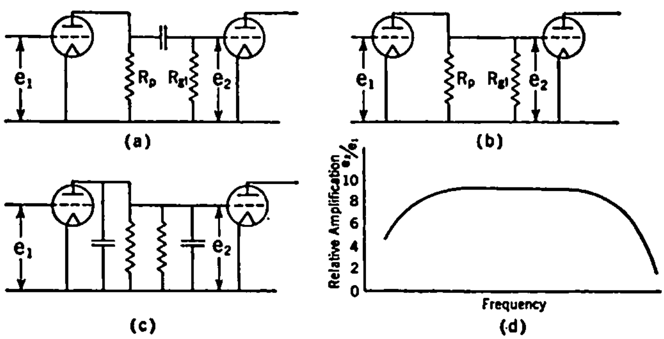


FIG. 7-3. Equivalent Circuit for a Resistance-coupled Amplifier at: (a) low, (b) medium and (c) high frequencies; and (d) the type of response obtained.

may be considered a short circuit as far as the signal frequency is concerned. At very low frequencies, however, its reactance becomes large and some of the signal voltage appears across it as well as across the grid-leak resistor where it is wanted. For this reason the amplification will fall off at low frequencies. At very high frequencies the reactances of the tube capacitances and wiring capacitances become low compared with the resistances which they shunt (see Fig. 7-3), and so the amplification drops off. The tube capacitances involved are the plate-to-cathode capacitance of the first tube, which parallels R_p, and the grid-to-cathode resistance of the second tube, which parallels R_g. This is shown in Fig. 7-3 which also shows the equivalent circuits at low and high frequencies and the frequency-response curve of such an amplifier.

Gain of an Audio Amplifier. The voltage gain of an amplifier stage is given by the ratio of the signal voltage appearing across the grid of the second tube to the signal voltage on the grid of the first tube. It is quite easily determined over most of the frequency range, for which the equivalent circuit of Fig. 7–3b will apply. It has already been shown in Chapter 4 that the voltage E_p appearing across a resistor R in the plate circuit of a triode (Fig. 4–11) is given by

$$E_p = \frac{\mu E_s R}{r_p + R}.$$

In the case of Fig. 7–3b, the resistance R consists of R_p and R_g in parallel. That is,

$$R = \frac{R_p R_g}{R_p + R_g}.$$

If the grid leak resistance R_g is very much larger than the plate resistor R_p, as is often the case, the effect of R_g in parallel with R_p will be small and R will be approximately equal to R_p. In this case the voltage amplification of the stage is the same as the voltage amplification of the tube, and is given by

$$\frac{e_2}{e_1} = \frac{\mu R_p}{r_p + R_p}.$$

It will be seen that as the plate resistor is made very large compared with r_p the plate resistance of the tube, the actual amplification approaches μ, the amplification factor of the tube (see Fig. 7–2c). When R_p is 9 times r_p, the amplification will be $\frac{9}{10}$ of μ.

At low frequencies, where the exact circuit of Fig. 7–3a must be used, it is a little more difficult to compute the gain. But if R_g is still very much larger than R_p, the gain will be given approximately by

$$\frac{R_g A}{\sqrt{R_g^2 + X_C^2}},$$

where

$$A = \frac{\mu R_p}{r_p + R_p},$$

and is the gain in the medium-frequency range. This is because R_g and X_C act as a voltage divider across R_p, and the fraction of the total voltage that appears across R_g is given by

$$\frac{R_g}{\sqrt{R_g^2 + X_C^2}}.$$

X_C is the reactance of the coupling condenser and is given by

$$X_C = \frac{1}{2\pi f C},$$

where f is the frequency (cycles per second) and C is the capacitance in farads. At that frequency which makes X_c equal to R_g, the amplification will have dropped to $1/\sqrt{2}$ times A, its value for the medium frequencies. Since power is proportional to the square of the voltage (across a fixed resistance), this means that the power will have dropped to one half which corresponds to a drop of 3 db (see chart, Chapter 1). This is the *lower half-power frequency* of this amplifier stage and is considered as being the lower limit of frequency response where reasonable fidelity is required. Similarly the *upper half-power frequency* is that frequency at the upper end of the range where gain has dropped to $1/\sqrt{2}$ times A due to the shunting effect of the tube and wiring capacities. It can be calculated by considering the effect of these capacities in parallel with the plate and grid resistors.

The Pentode Amplifier. The circuit of a resistance-coupled pentode amplifier is shown in Fig. 7–4. Except for the connections to the two new grids to apply the proper D.C. potentials to them, the circuit is similar to that of the triode. However, there are important differences in its operation, as was pointed out in Chapter 4. Because of the screening effect of the screen grid, the plate voltage has little influence on the plate current. This means that both the plate resistance r_p and the μ of the tube will be very high. The plate resistance is usually greater than a megohm, so that instead of using a plate resistor R_p having several times the value of r_p it is now necessary to use one that is very much smaller than r_p, because its size is still limited by the D.C. voltage drop across it. The gain of the stage will be still given by

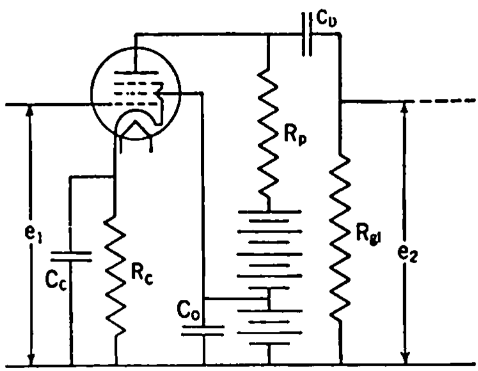


FIG. 7–4. Circuit of a Resistance-coupled Pentode Amplifier.

$$\text{Voltage gain} = \frac{e_2}{e_1} = \frac{\mu R}{r_p + R},$$

but now r_p will be much larger than R (R is even smaller than R_p), so it is possible to consider that $(r_p + R)$ is about the same as r_p and write

$$\frac{e_2}{e_1} = \frac{\mu}{r_p} R.$$

Now $\frac{\mu}{r_p}$ is equal to g_m, the mutual conductance or transconductance of the tube, so for a pentode the gain of a stage of amplification is given by

$$\text{Voltage gain} = \frac{e_2}{e_1} = g_m R,$$

which is approximately equal to $g_m R_p$ if the grid-leak resistor is several times larger than the plate resistor R_p.

From this relation it is evident that the figure of merit or worth of a pentode as a resistance-coupled amplifier is its transconductance rather than its amplification factor. With the pentode as with the triode, the D.C. voltage *on the plate* is the plate-supply voltage minus the voltage drop in the plate resistor. However, with the triode the plate current that flows is proportional to the plate voltage so that if this plate voltage should be decreased owing to use of a larger plate resistor or lower plate-supply voltage, the plate current will decrease and so reduce the voltage drop in the plate resistor. In the case of the pentode there is no such compensating effect because the plate current depends upon the *screen voltage* and is almost independent of the plate voltage. For this reason, if a larger plate resistor is used or the plate-supply voltage is reduced it is necessary to reduce the plate current by decreasing the screen voltage. If this is not done the plate voltage may be reduced to almost zero and the tube will cease to function as an amplifier.

Multistage Audio Amplifiers. It is generally desirable to have more gain than can be obtained from a single stage and for this purpose two or

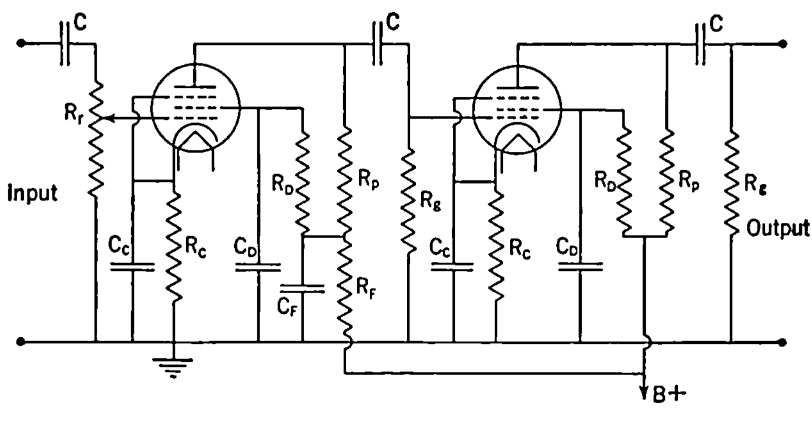


FIG. 7–5. A Two-stage Resistance-coupled Amplifier Using Pentodes.

more amplifier stages are connected in *cascade*. A typical circuit using two pentodes is shown in Fig. 7–5. The resistors R_p and R_g are the usual plate and grid-leak resistors respectively, and the condensers C are the coupling condensers. Resistors R_D are voltage-dropping resistors to reduce the screen voltages below those used on the plates, and condensers C_D are the necessary screen-resistor by-pass condensers whose function has already been discussed in Chapter 4. Resistors R_C are cathode resistors whose function is to furnish the required grid bias. Because the plate current must flow through these resistors in order to complete the path back to ground and the high voltage supply, there will be a positive voltage on the cathodes furnished by the *IR* drop across these resistors. Since the grids are operated at ground potential, the cathodes will be more positive than the grids, or the grids will be negative with respect to the cathode, by the voltage drop across the cathode resistors. The required size of resistor is then given by

$$R_c = \frac{E_c}{I_p},$$

where E_c is the required grid bias and I_p is the D.C. plate current flowing in that particular tube. The condensers C_c are cathode by-pass condensers whose function is to by-pass the signal currents around the cathode resistors. This low-impedance path is necessary, for otherwise the audio-frequency variations in the plate current would produce an audio-frequency voltage which would be introduced directly into the grid circuit.

The resistor and condenser combination R_F-C_F constitutes what is known as a filter or decoupling circuit. One of the precautions that must be taken in constructing a multistage audio amplifier is to prevent the signal voltage in the output stage from getting back to the input circuit. If this *feedback* occurs the signal which is fed back will be in such a direction that it either aids or opposes the original input signal. The first is called *regenerative* and the second *degenerative feedback.* With regenerative feedback the signal fed back from the output results in still more signal being fed into the input and the signal strength may progressively increase until the amplifier reaches an oscillating or *singing* condition. This phenomenon is particularly troublesome with three-stage amplifiers or with two-stage amplifiers having a high gain, for then the output signal is very large compared with the input signal and it requires only a very small percentage of the output signal to be fed back to produce the unwanted oscillation. This feedback may occur in several ways, but in audio amplifiers the most common cause is the coupling between stages due to the common high-voltage supply. If batteries are used they will have a low internal resistance when they are new and no feedback difficulties should be experienced. However, as the batteries age their internal resistance increases greatly. The alternating plate current of the last stage flows through this resistance and the resulting alternating voltage is applied to the plate of the first tube and hence directly onto the grid of the second tube. This condition results in feedback.

If a power supply is used instead of batteries it must be well filtered, as outlined in Chapter 5. In this case the alternating plate current must flow through the output condenser of the filter and so will produce a voltage drop across this condenser. The condenser is usually quite large, say 8 μf, so this voltage drop will be very small except at the low frequencies where the condenser reactance $1/(2\pi fC)$ becomes appreciable.* (For example, an 8-μf condenser has a reactance of only 20 ohms at 1,000 cycles, but this increases to 400 ohms at 50 cycles.) For this reason the frequency of the oscillation which occurs is generally quite low and gives rise to the sound known as *motor-boating*. This type of feedback can be eliminated by the filter circuit $C_F R_F$ (sometimes for high gain a two-section filter is required). The resistor and condenser in series act like a voltage divider across the power supply with only the voltage across the condenser being applied to the plate circuit of the first tube. If the condenser reactance is small (that is, if the condenser is large) and the resistor is large, this voltage will be only a small fraction of the original voltage appearing across the power supply.

Hum and Tube Noise. When several stages of amplification are used difficulty is often experienced with hum and other noise which is present with no signal applied to the input. This trouble occurs because in a high-gain amplifier even a very small stray voltage picked up in the first stage may be amplified to a large signal at the output. Hum may originate from stray electromagnetic or electrostatic pickup from A.C. power lines. This type of pickup can be eliminated by adequate shielding. Shielded tubes and grid leads are necessities for the first stages of a high-gain amplifier. A poorly filtered power supply is a common source of hum and for it the remedy is obvious. With A.C. operation the filament leads carry alternating current, so they should be twisted together in order to reduce their magnetic field and should be kept as far as possible from grid and plate leads. (The corner of the chassis is a good place to run filament leads.) When direct current is used for the filament the grid return may be made to one side of the filament supply (usually the negative), but when alternating current is used to heat the filament the grid return should be made to the center tap of the filament transformer. If there is no center tap, a small center-tapped resistor may be connected across the filament leads and the grid return made to the center of this. Sometimes a small potentiometer is used and a screwdriver adjustment provided for setting to the position resulting in least hum.

Tube noise may also be a source of trouble in an audio amplifier. It

* For purposes of quickly estimating the effect of a condenser it is well worth remembering that at 1,000 cycles a 1-μf condenser has a reactance of 160 ohms approximately. From this it is possible to estimate without slide rule or pencil and paper the approximate reactance of any other size of condenser at any audio frequency. For example, the reactance of a 0.1-μf condenser at 50 cycles would be

$$160 \times \frac{1}{.1} \times \frac{1{,}000}{50} = 160 \times 10 \times 20 = 32{,}000 \text{ ohms},$$

while the reactance of an 8-μf condenser at 100 cycles would be

$$160 \times \frac{1}{8} \times \frac{1{,}000}{100} = \frac{160}{8} \times 10 = 200 \text{ ohms}.$$

In the radio-frequency range it is useful to remember that 1,000 $\mu\mu$f at 1,000 kc has 160 ohms reactance.

may arise from mechanical vibration of the tube parts, in which case it is called *microphonic noise.* Replacing the offending tube with a nonmicrophonic tube will usually cure this source of trouble. Another form of tube noise appearing at very high amplifications is a hissing sound produced by the so-called *shot effect.* This effect is due to the random motion of electrons in the tubes, in particular the first tube of the amplifier, because any noise occurring there is amplified by all the other stages. A similar type of noise, which places an upper limit on the amplification that can be obtained, is that known as *thermal agitation.* The electrons in any conductor or resistor are in constant motion back and forth, with an average velocity that increases with the temperature of the material. This random motion of charges produces fluctuating voltages in the conductor or resistor and these are amplified by the amplifier. The noise due to thermal agitation in the grid resistor of the first tube is of most importance, because it receives the full amplification of the amplifier. The noise can be reduced by using a lower value of resistance but this will also decrease the signal gain and the signal-to-noise ratio will not be improved.

Over-all Gain and Frequency Response. The total gain of an amplifier consisting of several stages is the *product* of the gains of the individual stages when these gains are expressed as voltage ratios. For an amplifier having a voltage gain of 100 in each of the first two stages and 2 in the last stage, the over-all gain would be $100 \times 100 \times 2 = 20{,}000$. When the gain is expressed in decibels the total gain is the *sum* of the gains of the individual stages. From Fig. 1–13 of Chapter 1 it will be found that the gain of the above amplifier would be $40 + 40 + 6 = 86$ db. The over-all frequency response of an amplifier is also the product of the frequency response of the individual stages when this frequency response is expressed as a voltage ratio, as in Fig. 7–3d. If the gain of an amplifier stage at 10,000 cycles were only half of the gain at some reference frequency such as 1,000 cycles, then the gain at 10,000 cycles of two such stages in cascade would be only $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ of the gain at the reference frequency. If the frequency response were expressed in decibels a single stage of the above amplifier would be *down* 6 db (see Fig. 1–13) at 10,000 cycles below the gain at 1,000 cycles. Two similar stages in cascade would be *down* $6 + 6 = 12$ db at 10,000 cycles. It is apparent that if the over-all frequency response of an amplifier is to be reasonably good, the frequency response of a single stage must be very good. It is possible to *compensate* an amplifier to improve its frequency response and some methods of doing this will be discussed under *video amplifiers.*

Nonlinear Distortion in Audio Amplifiers. All of the three types of distortion discussed in Chapter 6 are found in audio amplifiers. The first of these, frequency distortion, has already been covered under resistance-coupled amplifiers, and the third, phase distortion, is not important in audio amplifiers. *Nonlinear* or *amplitude* distortion usually appears in the output or power stage of an amplifier because that is where the signal is of greatest amplitude, but it may be produced in any of the other stages if incorrect operating conditions are used. Figure 7–6 shows possible ways in which nonlinear distortion may be introduced into an amplifier.

In Fig. 7–6a are shown the correct operating conditions with the alternating grid voltage swinging over the linear or straight-line portion of the

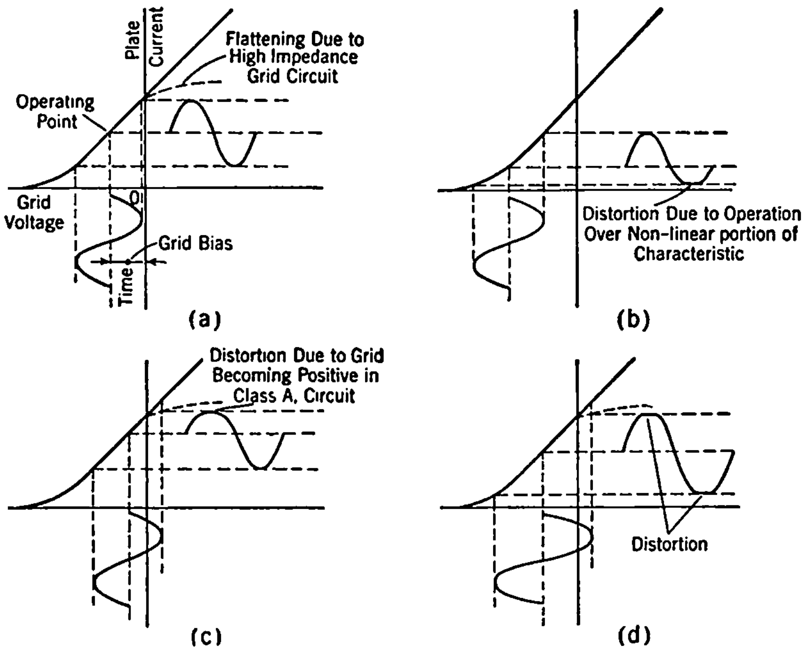


FIG. 7–6. Nonlinear Distortion Produced by Incorrect Operating Conditions. (a) Correct operating condition. (b) Grid bias too large. (c) Grid bias too small. (d) Correct grid bias but grid driving voltage too high.

characteristic curve. The result is a plate-current wave form that is a faithful reproduction of the grid voltage wave form.

In Fig. 7–6b is shown the effect of using too large a value for the D.C. grid bias. Operation has now been carried into the nonlinear portion at the lower end of the E_g-I_p curve and the resulting plate-current wave will be flattened off on the lower peaks. This flattening results in the introduction of new frequencies (particularly second harmonic) in the output that were not present in the input.

In Fig. 7–6c is shown the effect of using too small a value of grid bias. This insufficiency allows the instantaneous grid voltage to swing positive over a fraction of a cycle. If the upper part of the characteristic curve is straight (solid-line portion) no particular harm will result, but if it curves sharply and flattens off (dotted curve) after passing into the positive grid-voltage region, the upper peaks of the plate-current wave will be cut off and serious distortion will result.

This dotted curve is the one that usually applies. It is obtained when the grid circuit has a high resistance, as was the case in the resistance-coupled amplifier. As long as the grid is negative it attracts no electrons and no grid current flows. However, when the grid goes positive it attracts some of the electrons being emitted by the filament and grid current flows. If the grid circuit has zero or very small resistance, this flow of grid current has little effect on the grid voltage; but if this grid current has to flow through a high resistance (as would generally be the case), a large voltage drop occurs across the resistance and this drop reduces the grid voltage. This voltage reduction in turn reduces the plate current which would flow for positive grid voltages, and changes the characteristic curve, as shown by the dotted portion. If special low-impedance grid circuits are used it is possible to operate in the positive grid region without serious distortion and it is sometimes desirable to do this because of the larger power outputs obtainable. Such operation is indicated by use of the subscript 2 after the letter, indicating the class of operation.

Figure 7–6d shows operation with the correct grid bias but with too large a signal voltage on the grid. In this case both the upper and lower peaks are flattened off and the distortion may become quite bad.

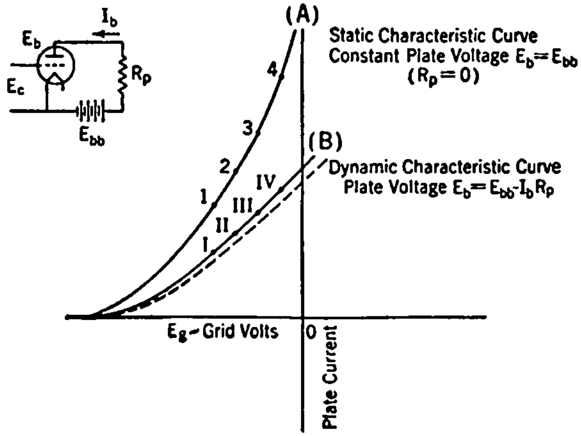


Fig. 7–7. Static and Dynamic Characteristic Curves of a Vacuum Tube.

Dynamic Characteristics of a Vacuum-tube Circuit. From the above it will be apparent that the selection of the correct operating voltages for a tube is quite important if a large signal with low distortion is desired. To determine the correct operating voltages it is necessary to consider the action of a tube *in a circuit*. Figure 7–7 (curve A) shows a mutual-characteristic curve for a tube relating grid voltage to plate current. (This curve is similar to the curves of Fig. 4–8, Chapter 4.) However, this curve cannot be used in plotting the operation of the tube in a circuit because it is for a *constant* plate voltage and makes no allowance for the fact that the actual plate voltage *decreases* as the plate current increases owing to the voltage drop in the plate resistor. The curve which takes this into account and which shows actual plate current against grid voltage when there is a resistor in the plate circuit is a lower curve, such as (B) in Fig. 7–7. This curve is the *dynamic* or operating chracteristic mentioned in Chapter 4. It depends upon the size of the load or plate resistor as well as on the characteristics of the tube. This dynamic characteristic can be obtained very easily from the plate characteristics of the tube and the load line. Each intersection of a plate-characteristic curve with the load line shows the plate current that will flow for a given grid voltage E_c and for the *actual* plate voltage on the tube. Therefore, if these corresponding values of plate current and grid voltage are plotted on the diagram of Fig. 7–7, they will lie along a line such as B. This line is the dynamic chracteristic curve for the value of load resistor used to construct the load line in Fig. 4–12. For a larger value of resistor, the load line of Fig. 4–12 would be less steep and the dynamic characteristic of Fig. 7–7 would also be less steep as indicated by the dotted curve B.

The dynamic characteristic obtained in this way can be used to plot the plate-current changes that will occur as the grid voltage is varied. This plot is shown in Fig. 7–8. If the dynamic characteristic is curved, then nonlinear distortion will occur and new frequencies (chiefly second harmonic) will be introduced into the output, as shown in Chapter 6. However, the dynamic characteristic will be straighter than the static characteristic and in general the higher the value of load resistance the straighter will be the dynamic curve. The amount of distortion, as indicated by the percentage of second harmonic present in the output wave, can be estimated roughly from the shape of the plate-current wave in Fig. 7–8. The solid line shows the actual plate current, which has a greater amplitude for the top half of the wave than for the lower half. The dotted line is a true sine wave for which the two halves are equal. Also shown dotted is the second harmonic wave, which when added to the dotted sine wave (fundamental) should give the actual wave. The percentage of second harmonic will be given approximately by $(a/b)100$.

It is also possible to determine the second-harmonic percentage directly from a set of plate-current–plate-voltage curves such as those of Fig. 4–12. This second method is dealt with in most tube manuals.

Selection of the Operating Point. Figure 7–8 shows that in order to avoid distortion the tube should be operated only over the straight-line portion of its characteristic and the lower curved portion of the dynamic curve should be avoided. Because the grid should never be allowed to swing positive, this requirement locates the operating region between the point where the curvature starts at the lower end of the curve and the point $E_c = 0$ at the upper end. The operating point P should then be midway between these limits. This same operating region can be located directly on the load line of Fig. 4–12. It extends along the load line from the curve $E_c = 0$ down to the region where the plate characteristics become curved.

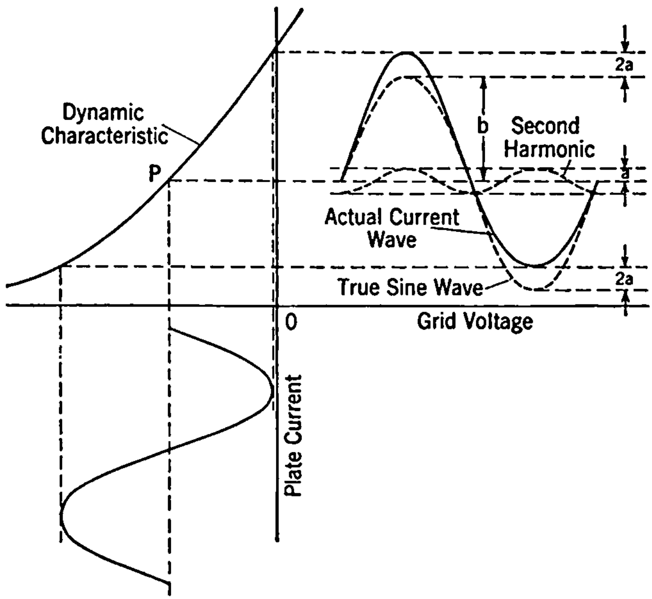


Fig. 7–8. Distortion Introduced by a Nonlinear Dynamic Characteristic.

Avoiding Nonlinear Distortion. To avoid excessive distortion in an amplifier stage certain precautions must be observed.

(1) The correct plate and grid voltages should be used so that operation takes place on the linear portion of the dynamic characteristic curve. The manufacturer usually specifies two or three sets of correct operating voltages, and where possible one of these should be used. It should be remembered that the voltage on the plate of the tube will be less than the supply voltage by the voltage drop in the plate resistor or coupling transformer. In the latter case the drop will be small, but in the former case it may be the larger part of the supply voltage.

(2) The load resistance should be of the right value. In the case of triodes this usually means about two or three times the plate resistance of the tube if maximum power output with small distortion is desired. In the case of pentodes the amount of distortion increases quite rapidly if the

wrong load resistance is used. The correct value is usually specified by the manufacturer and this recommendation should be followed closely. With pentodes it should also be remembered that if the value of plate resistor is increased it is necessary to reduce the plate current by reducing the screen voltage in order that the voltage drop across the plate resistor should not be too great.

(3) Even with correct operating potentials and load resistance, serious distortion will occur if the signal voltage on the grid is too great. In this case both peaks of the output wave will be flattened off. The remedy is to reduce the input signal voltage by means of a potentiometer or volume control in the grid circuit.

Transformer-coupled Amplifier. Another type of coupling commonly used in audio-frequency amplifiers is transformer coupling. Although its most general use is in coupling the output stage to the speaker or load, it may also be used as interstage coupling. Here it has the advantages of giving somewhat higher gain and eliminating the voltage drop that occurs in the plate resistor when resistance coupling is used. It has the disadvantages of greater cost, increased space requirement, and the possibility of a poorer frequency response. In a resistance-coupled amplifier all the voltage gain is produced by the tubes and the maximum gain possible from a stage is given by the amplification factor or μ of the tube in that stage. As was shown in an earlier section, the actual amplification is always less than μ by the factor

$$\frac{R}{R_p + R}.$$

With a transformer-coupled stage, however, the maximum possible gain is given by $\mu \times n$, where n is the step-up turns ratio of the transformer. For reasons to be given, n is generally limited to about 3, but even so with low-μ tubes this is an important factor and the early audio amplifiers were mostly transformer coupled. However, the introduction of pentodes and high-μ triodes has largely removed this advantage because it is now possible to obtain easily amplifications of the order of 100 times with a single resistance-coupled amplifier stage. For this reason transformer coupling is used mainly in the power amplifier stage or stages where certain other advantages make its use desirable.

Figure 7–9 shows a transformer-coupled amplifier, an approximate equivalent circuit, and a typical curve of amplification against frequency. The amplification is fairly uniform over the medium-frequency range from b to c but falls off at the low-frequency end (a to b) and again at the high-frequency end (d to e). The amplification represented by the flat portion b to c is approximately $\mu \times n$. For all low and medium frequencies up to the point c, the voltage e_s across the secondary of the transformer is n times the voltage e_p across the primary. For a given signal voltage e_g on the grid of the tube the voltage e_p depends upon the ratio of the reactance of the primary winding of the transformer to the plate resistance r_p of the tube. It is very small when that reactance is much less than r_p and increases as the reactance increases, approaching a

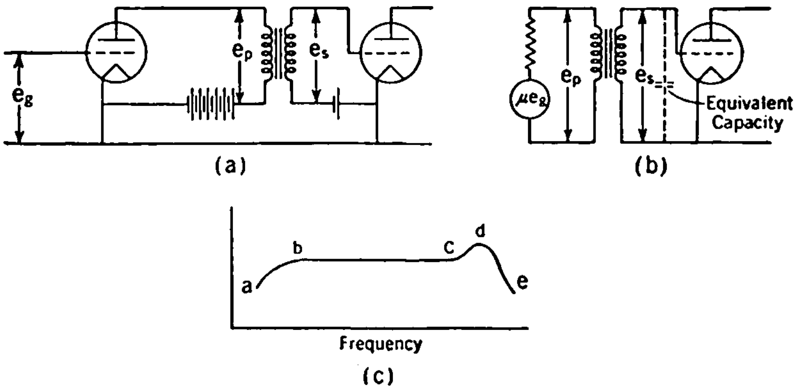


FIG. 7-9. (a) Transformer-coupled Amplifier. (b) Approximate Equivalent Circuit Showing Effect of Capacities. (c) Typical Frequency-response Curve.

constant value μe_g as the reactance of the transformer primary becomes very large compared with r_p. This behavior is similar to the way the output voltage of a resistance-coupled amplifier increases as the load resistor is increased and a curve similar to the curve of Fig. 7-2c would apply. This is shown in Fig. 7-10, where e_p is plotted against the ratio of the primary reactance X_p to the plate resistance r_p. The reactance of the primary winding depends upon the frequency and is given by $2\pi fL$ where L is the inductance of the winding, a constant, and f is the frequency. The curve of Fig. 7-10 is therefore also a picture of how e_p varies with the frequency. At a frequency which makes $2\pi fL = r_p$, e_p will be .707 times its maximum value μe_g. Above this frequency the amplification changes very little but below this frequency it falls off sharply.

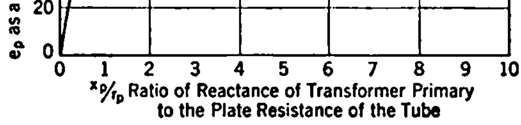


FIG. 7-10. Voltage e_p across the Primary Winding of a Transformer as a Function of the Reactance of the Winding.

It is still necessary to explain the hump at point d and the sharp drop in amplification at still higher frequencies. Both primary and secondary windings of the transformer consist of a large number of turns of wire and there is considerable capacity between windings and between the individual turns of each winding. These capacities can be represented approximately by an equivalent capacity shunted across the secondary winding (Fig. 7–9b). At low frequencies this capacity has negligible effect because its reactance will be very high, but at high frequencies where its reactance becomes smaller it acts as a shunt to by-pass the signal and the amplification is lost. This effect is shown by the portion of the curve from *d* to *e*. At some point intermediate between the high-frequency and low-frequency cases, represented by the hump *d*, there is a frequency at which resonance can occur between this equivalent capacity and the leakage reactance of the transformer. At this frequency the amplification may rise to quite large values, producing serious frequency distortion. This situation will be especially probable if the plate resistance r_p has a very low value.

Since the amplification of a transformer-coupled amplifier stage is given approximately by μn, it would appear possible to produce very high amplifications by the simple expedient of making the turns ratio extremely large. However, if this is done by winding on a large number of turns on the secondary the equivalent capacity across the secondary will be increased to such an extent that the high-frequency response of the transformer will drop off badly. On the other hand, if the turns ratio is increased by decreasing the number of primary turns, the primary reactance will be decreased and the low-frequency response will suffer. Usually a transformer ratio of 1:3 is about the greatest that can be used.

A transformer is generally designed to be used with a tube having a certain value of plate resistance. The effect of using tubes having plate resistances higher and lower than this proper value is shown in Fig. 7–11.

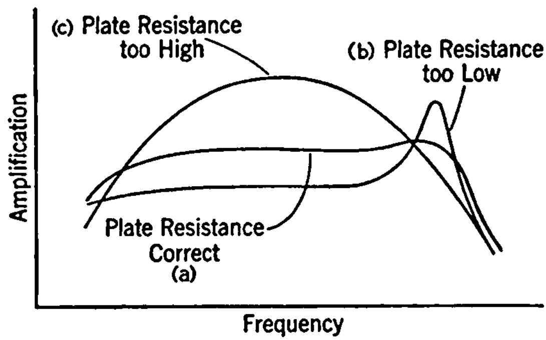


Fig. 7–11. Effects of Different Plate Resistances on the Frequency Response of a Transformer-coupled Amplifier.

Curve (*a*) is for a tube having about the correct value of plate resistance. If the plate resistance of the tube is much lower than the proper value (curve *b*) there is danger that the hump in the amplification curve will rise to a high peak. If the plate resistance of the tube is much higher than the proper value, the primary reactance of the transformer at low frequencies will not be large compared with r_p, and so the low-frequency amplification will be reduced. At the high-frequency end the shunting effect of the equivalent capacity will be greater and the high-frequency response will also be reduced. Because tubes with large plate resistances also have large amplification factors, the amplification in the middle-frequency range, still being given by μn, may be quite high and the resulting curve will be as shown in *c*. This explains why it is not feasible to use pentodes or high-μ triodes with a transformer to obtain very large amplifications. If for some reason it is desired to use a transformer with a tube having a high plate resistance, this can be done by shunting the primary winding of the transformer with a resistance equal to the plate resistance of the tube that should be used with the transformer. This shunt has the effect of reducing the amplification in the medium-frequency range and the resulting frequency response will be about the same as would be obtained with a tube having the correct plate resistance. The actual amplification that

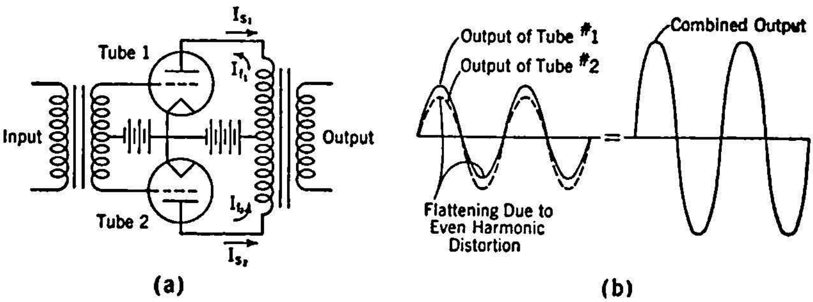


FIG. 7–12. (a) A Push-pull Amplifier. (b) Reduction of Even Harmonic Distortion by Use of Push-pull Connection.

results will be proportional to the transconductance of the tube being used and will be independent of its amplification factor.

Push-pull Circuits. To get greater ouput than can be obtained from a single tube, two tubes are often connected in a *push-pull circuit* as shown in Fig. 7–12. At the instant the voltage on the grid of tube 1 is a positive maximum, the voltage on the grid of tube 2 will be a negative maximum. At this instant the plate current of tube 1 will be increasing and the plate current of tube 2 will be decreasing. These currents will produce voltages in the secondary of the transformer that are in the *same* direction, and so twice the power output can be obtained. The circuit has the advantage that distortion from the second and all even harmonics will be reduced, because for these harmonics the currents will flow in *opposite* directions through the transformer and so cancel out. Even harmonics give to the lower half of a wave a shape different from that of the upper half, as for example the flattening off of the lower half shown in Fig. 7–6, when too large a grid bias was used. In a push-pull circuit the top half of the current wave of one tube is added to the bottom half of the other, and a symmetrical wave results. This is shown in Fig. 7–12b. Because twice the output of a single tube can be obtained with *less* distortion, it is possible to obtain more than twice the output with the *same* distortion given by a single tube. Push-pull operation also has the advantage that any A.C. hum voltage from the power supply will produce currents in opposite directions in the primary of the transformer and no hum voltage will appear across the secondary. However, this advantage applies only to hum introduced in this stage, since hum introduced in preceding stages will be amplified just as is the signal voltage. The *D.C. plate currents* to the tubes also flow in opposite directions in the transformer primary winding so that large plate currents may be used without danger of saturating the core.

Power Amplifiers. The final stage of an audio amplifier is usually a power-amplifier stage, because loudspeakers, modulation transformers, and most other devices which the radio signal must drive require power (that is, *current* as well as voltage) for their operation. Class A amplifiers are used extensively as power amplifiers and the essential difference between their operation as power amplifiers and their operation as voltage amplifiers is in the magnitude of the load resistance and the plate voltage used. For voltage amplification the load resistance is made as high as possible to obtain maximum voltage gain; but when power output is the chief consideration the load resistance is reduced to a value that gives most power for a certain permissible value of distortion. With power amplifiers the highest possible plate voltage is used because the power output possible is directly dependent on the plate voltage.

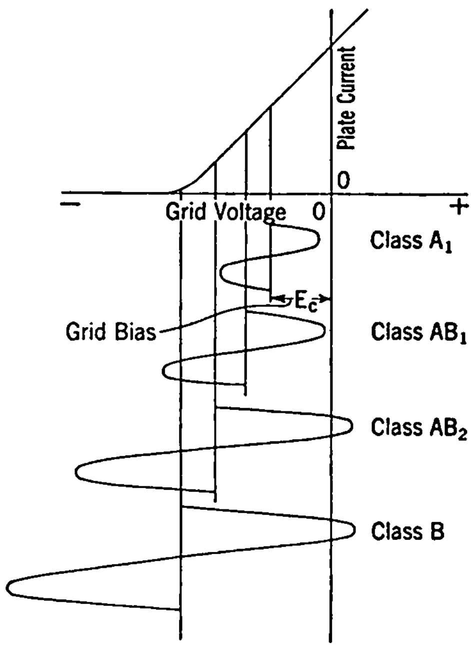


FIG. 7–13. Audio Amplifier Classifications.

To obtain larger power outputs for a given tube size and plate voltage, Class AB and Class B operation are used. The different amplifier classifications are shown in Fig. 7–13. In Class A_1 operation the grid-voltage swing is generally confined to the linear portion of the curve and the plate current flows for the entire cycle. The grid never goes positive and hence the maximum grid swing permissible is limited to that shown in Fig. 7–13.

By increasing the grid bias E_c a larger swing of grid voltage can be obtained without the grid ever going positive; such operation is known as Class AB_1. In this case the negative peak of the grid-voltage swing carries the operation off the curve and the plate current is zero for a fraction of a cycle. This would produce excessive distortion with a single tube, but by using two tubes connected in push-pull the current waves are made to supplement each other as shown in Fig. 7–12b; thus the distortion is kept within reasonable limits.

If the grid-voltage swing is increased still further with the same grid bias, a greater output can be obtained but the grid will be positive for a small portion of the cycle. Such operation is called Class AB_2. It requires a driving voltage fed from a low-impedance source, or grid-circuit nonlinear distortion of the type shown in Fig. 7–6 will result. Practically, this means that because the grid circuit draws current it requires *power* from the preceding stage. For this reason the preceding stage should be a Class A_1 amplifier feeding the grid circuit of the Class AB_2 stage through either a *one-to-one* or a *step-down* transformer.

Even greater power output can be obtained from a tube by operating it Class B, in which case the grid bias is increased almost to plate-current cutoff so that plate current flows only for the positive half of the grid-voltage swing. Because this type of operation is used chiefly to modulate the radio-frequency power of a transmitter, a more detailed discussion of it will be given in Chapter 11.

Inverse Feedback. Distortion in an audio amplifier can be reduced and the frequency response of the amplifier greatly improved by the use of

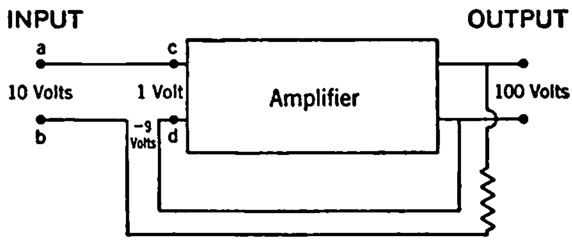


FIG. 7–14. Principle of Inverse Feedback.

inverse feedback (also called negative or degenerative feedback). This improvement in quality is obtained at the expense of the gain of the amplifier, but with modern high-μ tubes it is easy to obtain all the gain that may be desired. In an inverse-feedback circuit some of the output voltage is fed back to one of the preceding stages in a direction such that the voltage fed back opposes the original signal voltage at that point. If some distortion has occurred in the amplifier and the output contains new frequencies not originally present in the input, these new frequencies are now introduced into the input in a direction such as to tend to cancel those produced in the amplifier itself. Again, if the amplifier produces frequency distortion so that some frequencies are amplified more than others, these over-amplified frequencies produce a large signal in the input (due to the feedback). This large signal opposes the original signal and the *actual input signal* at these frequencies will be small. A numerical example will help make this situation clear.

Suppose an amplifier without feedback requires 1 v input across *cd* to produce 100 v output (Fig. 7-14). A feedback circuit is then added which feeds back 9 v to the input when the output is 100 v. If this 9 v is in the *opposite* direction to the original 1 v, it will now require 10 volts $(1 + 9 = 10)$ at the input terminals *ab* to keep 1 v at *cd* and so maintain the 100 v output. The original gain of the amplifier without feedback was

$$\frac{\text{output volts}}{\text{input volts}} = \frac{100}{1} = 100.$$

The new gain with feedback added is

$$\frac{\text{output}}{\text{input}} = \frac{100}{10} = 10.$$

The first effect of feedback has thus been to reduce the gain of the amplifier. These results can be expressed mathematically by saying that the gain of the amplifier with feedback is given by

$$\text{Gain} = \frac{A}{1 - \beta A},$$

where A is the amplification without feedback and β is the fraction of the output voltage that is fed back. In the above example $A = 100$ and $\beta = -9/100 = -.09$, so that

$$\text{Gain} = \frac{100}{1 - (100 \times -.09)} = \frac{100}{1 + 9} = 10.$$

Now consider the effect on frequency response. Suppose the amplifier alone has considerable frequency distortion and amplifies a 1,000-cycle signal 100 times but amplifies a 3,000-cycle signal 200 times. For the same input signals the ratio of 3,000-cycle to 1,000-cycle signal in the output will be $200/100 = 2$, without feedback. When feedback is applied it will require for the 1,000-cycle signal 10 v at *ab* to produce 100 v at the output. (As before, 9 v of this will be used in overcoming the feedback voltage and the other volt will appear across *cd*.) However, for the 3,000-cycle signal it requires only $9\frac{1}{2}$ v at *ab* to produce 100 v at the output, because it requires 9 v to overcome the feedback voltage and only $\frac{1}{2}$ v across *cd* to produce 100 v at the output. Thus it is found that with feedback the amplifier gain for 1,000 cycles is 100/10 = 10, but for 3,000 cycles it is 100/9.5 = 10.5 approximately. If a 10-v 1,000-cycle signal is applied at *ab* the output will be 10 × 10 = 100 v, but if a 10-v 3,000-cycle signal is applied at *ab* the output will be 10 × 10.5 = 105 v. With feedback, then, for the same input signal the ratio of the output voltages at 3,000 cycles and 1,000 cycles will be 105/100 = 1.05. Instead of having twice as much 3,000-cycle output as 1,000-cycle output (that is 100% greater) in comparison with the case without feedback, with feedback the 3,000-cycle output is only 5% greater than the 1,000-cycle output.

Of course the over-all gain has been reduced but this can be overcome by designing the original amplifier for very high gain.

Practical inverse-feedback circuits. Inverse feedback may be used over one, two, or three stages. That is, the voltage may be fed back to the

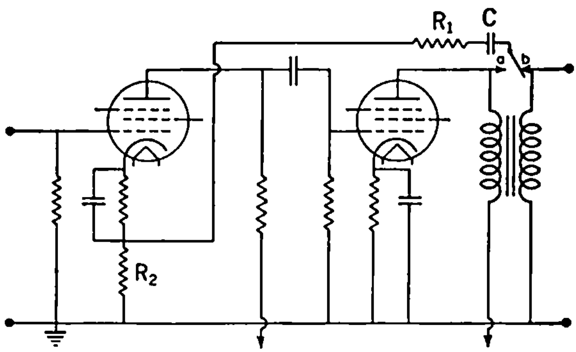


Fig. 7-15. Two-stage Amplifier with Inverse Feedback.

same stage, to a preceding stage, or to the stage before the preceding stage. One- and two-stage feedback is fairly easy to apply, but three-stage feedback is more difficult because of the possibility of phase shift or delay in the amplifier. If this phase shift amounted to nearly a half cycle at any frequency the voltage fed back would aid rather than oppose the original signal and oscillation would result. Figure 7-15 shows a typical two-stage feedback amplifier. Voltage is picked off the output stage at either of points *a* or *b* and fed back through a blocking condenser *c* and resistor R_1 to the cathode circuit of the first tube. The amount of feedback is set by adjusting the sizes of R_1 and R_2. If the voltage is picked off at *a*, any distortion introduced by the transformer is not compensated for by the feedback circuit. If the connection is made to point *b*, the feedback circuit will also compensate for frequency distortion introduced by the transformer. However, if the secondary winding is a low-impedance winding such as would be used to feed to the voice coil of a loudspeaker, there may not be sufficient voltage available for feedback at the point *b*, and it will be necessary to make the connection to point *a*.

The big advantage of inverse feedback in an amplifier is that it makes the gain and frequency response of the amplifier almost independent of changes that may occur in tube characteristics due to variations in supply voltages, aging of the tubes, and so on. The gain and frequency response depend mainly on the feedback circuit, and because this is composed only of resistances and a condenser, the characteristics will remain constant over a long period of time. For this reason inverse feedback is applied to A.C. operated vacuum-tube voltmeters, D.C. amplifiers, and other equipment that requires an amplifier of constant gain.

Video Amplifiers. For certain applications, especially in television, an amplifier is required which will pass a very wide band of frequencies, usually from about 20 cycles up to 3,000,000 cycles. Such amplifiers are called *video* amplifiers, because they are used to amplify the picture signal in television as opposed to the sound signal which requires an *audio* amplifier. The chief difference between a video amplifier and an ordinary amplifier is that it must be designed to amplify the high frequencies as well as the medium frequencies. It is not possible to build a transformer-coupled amplifier to cover such a wide range, so some form of resistance coupling is generally used. The amplification of an ordinary resistance-coupled amplifier drops off at the high frequencies because of the shunting effect of the tube capacitances. This is shown in Fig. 7–16. It is seen that the plate-to-cathode capacitance C_{pK} of tube 1, and the grid-to-cathode capacitance C_{gK} of tube 2, are in parallel with the plate and grid resistors respectively. Besides these there are the stray wiring capacitances to ground which are also in parallel with the resistors. At low frequencies these capacitances have very little effect because their reactance is so large compared with the resistances, but at very high frequencies their reactances become very small and they act like short circuits across the resistors. To reduce this effect as much as possible, tubes are used which have small capacitances combined with large values of transconductance. The amplifier is then wired in such a way as

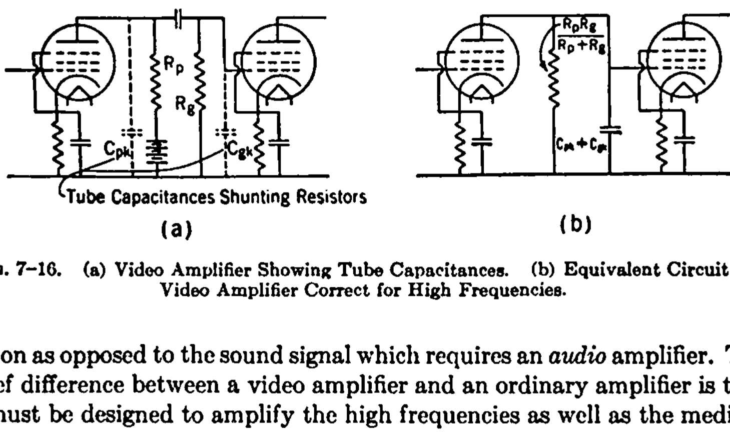


FIG. 7–16. (a) Video Amplifier Showing Tube Capacitances. (b) Equivalent Circuit of Video Amplifier Correct for High Frequencies.

to reduce wiring capacitances to a minimum. This care extends the frequency range quite considerably. It can be further extended by use of a *compensating inductance* in series with the plate resistor. This is shown in Fig. 7-17, along with the equivalent circuit accurate for high frequencies. The effect of the inductance is to form a parallel resonant circuit with the capacitance and so to produce a high impedance between the grid of the second tube and ground. By suitable selection of the inductance, the resonant frequency can be placed at the desired upper limit of the frequency range. Because of the large resistance R_p in series with L, the resonance curve for this L, C, and R circuit will be very broad, so that the impedance will be high over a wide frequency range. This condition improves the entire high-

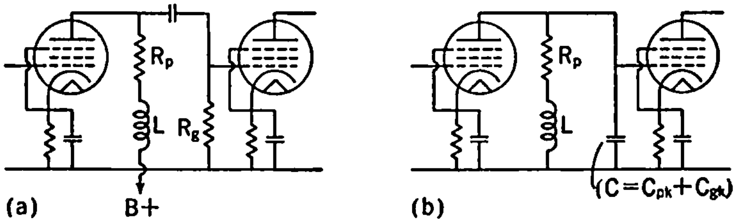


FIG. 7-17. (a) Video Amplifier Showing Compensating Inductance. (b) Equivalent Circuit Accurate for High Frequencies.

frequency response of the amplifier, and with careful selection of values it is possible to obtain nearly uniform response up to 4 or 5 mc.

Direct-current Amplifiers. Most amplifiers whose circuits are similar to those described above have poor low-frequency response and will not amplify D.C. voltages at all. Since it is sometimes necessary to amplify D.C. voltages, or A.C. voltages of very low frequency, changes must be made in the circuit of the amplifier. Many special circuits have been developed which can amplify direct voltages. Such amplifiers are known as *direct-current amplifiers.* The most common type of D.C. amplifier is the *direct-coupled amplifier.* Direct-coupled amplifiers are used in some oscilloscopes to amplify voltages before applying them to the plates of the cathode-ray tube and in vacuum-tube voltmeters to increase their sensitivity. One type of direct-coupled amplifier, known as the Loftin-White amplifier, has been used in radio sets as an A.C. amplifier. A type of direct-coupled amplifier is also used in regulated power supplies (see Chapter 5).

It was pointed out earlier in this chapter that the drop in low-frequency response in a resistance-capacitance-coupled amplifier is mainly caused by the drop in voltage across the coupling condenser between stages. The coupling condenser blocks off D.C. voltages completely so that they do not reach the grid of the tube in the next stage. Hence, if D.C. voltages are to be amplified this condenser must be removed entirely from the circuit. However, if it is removed the plate voltage of the previous stage applies a high positive bias to the grid of the tube. To overcome this positive bias, the negative bias on the grid must be greatly increased, so that a large battery is required. The circuit in Fig. 7-18 shows an amplifier of this type. The first stage is the same as used in resistance-capacitance-coupled amplifiers, except that there is no condenser in the input lead to the grid of the first tube. In the second stage, the bias battery E_{C2} has to overcome a positive voltage equal to the supply voltage of the previous stage minus

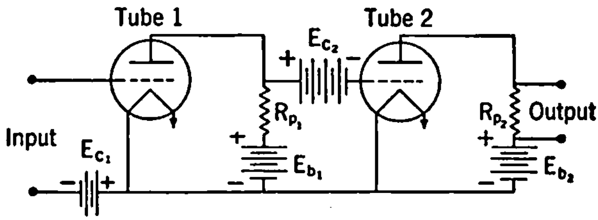


Fig. 7-18. A D.C. Amplifier Circuit.

the D.C. voltage drop in the plate resistor of this stage, and also has to supply the necessary negative bias for the tube. This circuit is an example of a direct-coupled amplifier, which could be used for amplifying either D.C. or A.C. voltages. This circuit is not often used because it requires such a large bias battery in the second stage.

In order to reduce the number of batteries required, the circuit can be changed as shown in Fig. 7-19. The plate resistor of the first tube and

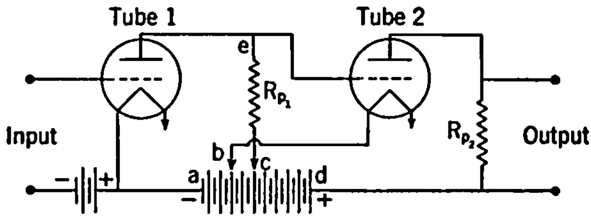


Fig. 7-19. An Alternative Method of Connection for a D.C. Amplifier.

one side of the filament, as well as the plate resistor of the second tube, are connected to taps on one battery. The plate supply for the first stage is the voltage between taps *a* and *c*. If the filament of the tube in the second stage were connected to point *c* instead of *b*, there would be a *negative* bias on this stage due to the plate current in the first tube flowing in the resistor R_{p1}. This voltage is usually too large a negative bias for the second tube, so that part of it is bucked out by moving the filament lead to tap *b*. The plate voltage on the second stage is the voltage between taps *b* and *d*. In this circuit it should be noted that the filaments of the two tubes are not at the same potential, so that either separate filament supplies, or tubes with indirectly heated cathodes, must be used.

Instead of using a tapped battery in the circuit of Fig. 7–19, a voltage divider can be used as shown in Fig. 7–20. This circuit is known as a Loftin-White direct-coupled amplifier. When A.C. voltages are to be amplified, it is necessary to by-pass the various sections of the voltage divider with low-reactance condensers. At the lowest frequency to be amplified these condensers should have reactances that are low compared with the resistances they by-pass.

If more than two stages are used in a D.C. amplifier, it becomes very difficult to make the amplifier stable. Any small changes in the voltages on the first tube are amplified to such an extent that it is hard to maintain

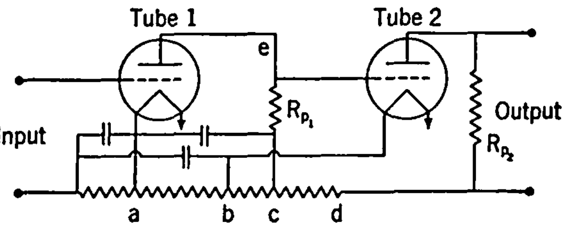


FIG. 7–20. Loftin-White Direct-coupled Amplifier.

the correct grid bias on the last tube in the amplifier. For this reason D.C. amplifiers seldom have more than two stages.

Public-address Systems. One of the important uses for an audio amplifier is the amplification of speech and music, either indoors or outdoors, so that the sound may be heard over a large area. A *public address system* (more briefly *p. a. system*) is the name given to the complete installation required for this application. A public-address system includes all necessary amplifiers, microphones, volume controls, mixing systems, loudspeakers, record turntables, and sometimes volume indicators and monitoring systems. A block diagram of the main parts of a large public-address system is shown in Fig. 7–21. Not all of these elements are included in smaller public-address systems, which may consist of only one microphone input, one phonograph input, and a simple volume-control system followed by one amplifier to feed from one to three loudspeakers. Only very large installations use volume indicators and monitoring amplifiers. In these large installations, such as in a theater, each of the parts shown in the diagram occupies a separate chassis. The usual small public-address system contains the amplifiers, the mixer, and the power supply all on one chassis.

PREAMPLIFIERS. Most microphones now used have very low voltage outputs so that considerable amplification must be used with them to produce the power required by the loudspeakers. While any properly designed amplifier with enough gain can be used, it is usual to employ a *preamplifier* ahead of the main amplifier. A preamplifier is an amplifier of moderate gain (about 50 db for example) especially designed to have very low hum and noise levels in its output. This low-noise feature is essential since any hum or noise in the first stage will be amplified by succeeding stages and will interfere with the desired audio signal. The noise level in a preamplifier can be kept low by using good-quality components in the circuit, by making careful joints and connections, and by adequate shielding. A preamplifier is particularly susceptible to hum, so that care must be taken to keep it away from power transformers, filter chokes, and filament

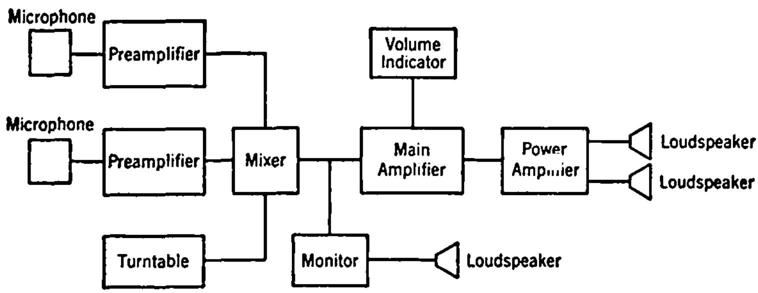


FIG. 7-21. Block Diagram of a Large Public-address System.

and power leads. A good way to reduce hum is to build the preamplifier on a separate chassis, but with proper care it is possible to include it on the chassis with the main amplifier.

Mixers, volume controls, and switching circuits are common sources of noise, so it is usual to place them in the circuit *after* the preamplifier. In this way, the microphone output is amplified enough by the preamplifier so that it overrides any noise introduced in these circuits. Noise from moving contacts in mixers and switching circuits is mainly caused by dirt, so that regular cleaning is necessary to keep down their noise level. They may be cleaned with carbon tetrachloride.

MIXERS AND VOLUME CONTROLS. When several microphones and phonograph turntables are used at one time, it is necessary to be able to control the volumes from each of them separately, and to combine or *mix* their outputs. This is accomplished by means of a *mixing circuit* or *mixer.* While mixing circuits are ordinarily used to mix the outputs of several sources to obtain the most natural balance between the various sounds, sometimes they are used to change the natural levels to obtain special effects. A mixing circuit is usually followed by a *master volume control* or *attenuator* which changes the volume of all the signals after they have been mixed.

One of the simplest mixing circuits, known as a *resistance mixer*, is shown in Fig. 7–22. In this circuit the inputs are applied to separate volume controls (variable resistors) and the variable arms of the controls connected together through the resistances R_2 and R_4. These resistances are necessary in the circuit since, if they were left out, turning one volume control to zero volume would short-circuit the output of the other. They must not be made too small, typical values being 0.25 to 0.5 megohms. The variable resistors R_1 and R_3 are usually 0.5 to 1.0 megohms. Resistance mixers of this type do not give completely independent control, since changing one control affects the volume of the other slightly. In most applications, however, the change is not noticeable, as it is usually less than 6 db.

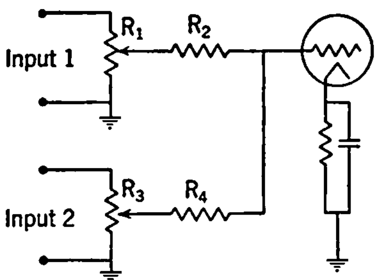


FIG. 7–22. Resistance Mixing Circuit.

Another simple type of mixing circuit is shown in Fig. 7–23, which is called an *electronic mixer*. In this circuit the two inputs are connected to the grids of a twin-triode tube and the two plates are in parallel, with a common plate resistor. An electronic mixer gives completely independent control of the two inputs. Figure 7–24 shows a slightly different type of electronic mixing circuit which is often used in preference to that of Fig. 7–23, as it has a higher gain.

Broadcasting studios and theaters employ mixing circuits which contain *constant-impedance* volume controls or attenuators. These constant-

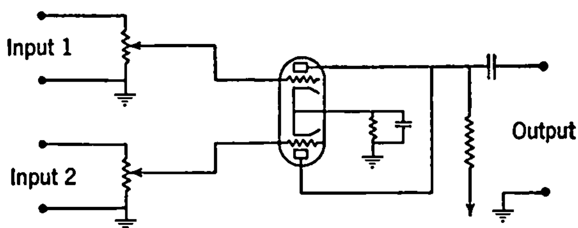


FIG. 7–23. Electronic Mixer.

impedance volume controls are of three main types, called *T*-pads, *H*-pads, and *L*-pads, which are illustrated in Fig. 7–25. Their action may be seen by considering a *T*-pad as follows. The three resistances are made variable and are connected or ganged to a single control knob. R_1 and R_2 are made equal and vary together. As the control knob is turned, resistances R_1 and R_2 decrease and resistance R_3 increases by the correct amount to maintain the impedance looking into the pad at a constant value. But as the knob is turned in this direction, the voltage drop across R_1 and R_2

decreases and that across R_3 increases, so that more and more of the input voltage appears at the output, which may be thought of as being in parallel with R_3 if R_2 is small. Thus a T-pad will change the volume in the circuit without changing the input impedance. The action of L- and

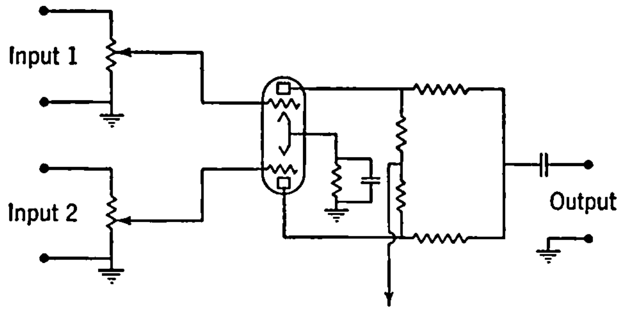


FIG. 7-24. Improved Electronic Mixing Circuit.

H-pads may be analyzed similarly. T-pads and H-pads maintain constant impedance on both the input and output since they are symmetrical. The L-pad will maintain constant impedance in only one direction, that is, on the side containing the series arm.

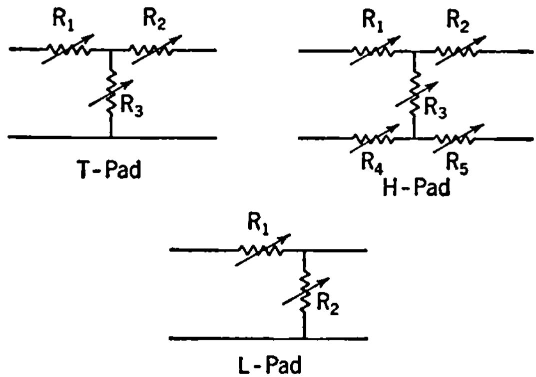


FIG. 7-25. Attenuators.

The way these pads can be connected in the circuit to form a mixer is shown in Fig. 7-26. This circuit uses two T-pads as mixers and one T-pad as a master volume control. Mixer 1 determines the volume of input 1, and mixer 2 that of input 2, while the master volume control gives control over the volume of the total signal after mixing. Since changing the setting of mixer 1 does not vary the impedance at its output terminals, there will be no change in the impedance connected to the output of mixer 2.

Consequently there is no change in the voltage supplied by mixer 2 as mixer 1 is varied. This means that the settings of the mixer controls are independent of each other. In the circuit shown, the outputs of the two mixing pads are really connected in series. Other arrangements are possible in which the pads are connected in parallel. By using more pads it is possible to mix as many circuits as desired, while always having separate control on each circuit.

Public-address amplifier circuits. The purpose of the main amplifier is to amplify the output of the mixing circuit up to a level sufficient to drive the loudspeakers to the desired volume. A wide variety of circuits is employed in these amplifiers, the circuit used depending on a number of

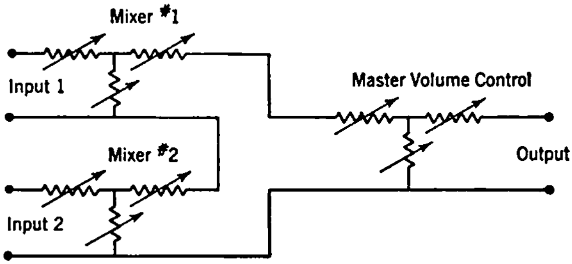


Fig. 7-26. Complete Mixing System.

factors—such as power output required, type and number of input circuits, the quality or fidelity required, portability, and cost. Public-address amplifiers are usually rated in terms of their power output, which is the amount of power delivered to the load, with no more than 5% to 10% distortion. The power-output stage is ordinarily a Class A amplifier in low-power high-fidelity amplifiers, Class AB_1 or AB_2 in medium-power amplifiers, and Class B in the large amplifiers. Class AB amplifiers are very popular for public-address work because of their excellent power output at low cost. The remainder of the amplifier is designed to provide the necessary amplification between the output of the mixing circuit and the input to the power stage. Factors in the design of such amplifiers have already been considered earlier in the chapter.

A circuit diagram of a typical public-address amplifier is shown in Fig. 7-27. This amplifier has an output of 12 to 15 watts undistorted, and is provided with one microphone input and one phonograph input. A resistance mixing circuit similar to that shown in Fig. 7-22 is used. The preamplifier for the microphone is a 6J7 connected as a pentode resistance-capacitance-type amplifier. The grid bias for this stage is obtained from a *bias cell*, a specially built battery for use only in grid circuits where no grid current flows. The cells give a voltage of from 1 to $1\frac{1}{4}$, and may be connected in series to give higher voltages. They are often used to bias the

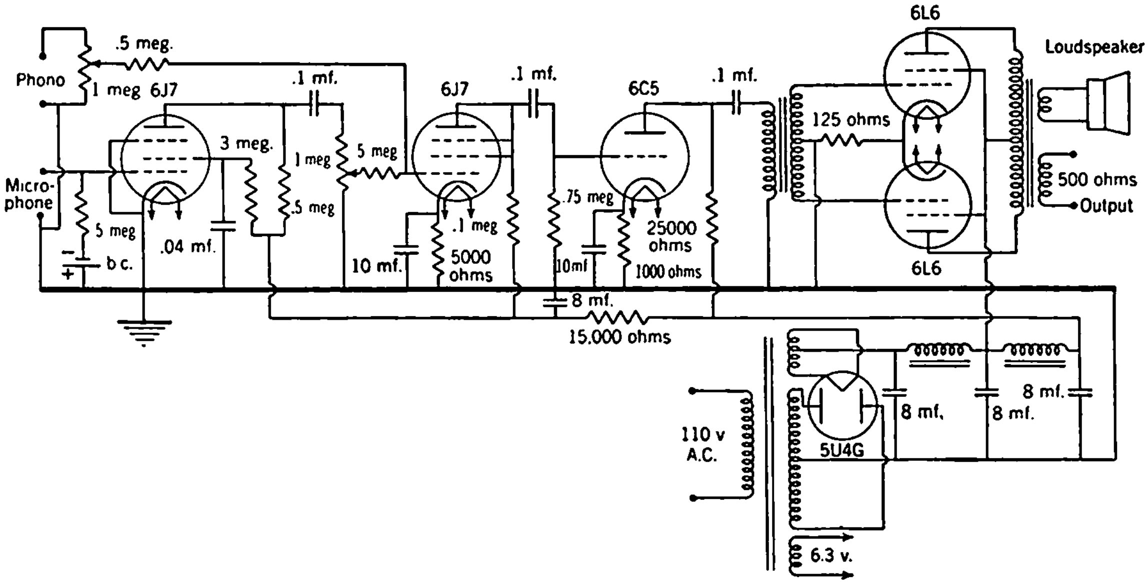


FIG. 7-27. A 15-watt Public-address System.

first tube in a preamplifier because they eliminate the cathode biasing resistor and by-pass condenser, giving decreased hum and distortion at low cost.

The output of the mixer is amplified by a type 6J7 pentode tube which is operated as a triode by connecting the screen, suppressor grid, and plate together to form the triode plate. This stage is resistance-capacitance coupled to the next stage, called the *driver*, which supplies grid excitation to the power-output stage. The plate circuit of the driver stage shown contains a shunt-fed or parallel-fed transformer. Many transformers are designed to carry the D.C. plate current of the tube to which they are connected, but often transformers are designed with relatively little iron in the core so that if the D.C. plate current should flow in the primary winding, saturation would take place, causing distortion. In the latter case, the transformer is parallel fed, by using a resistor in the plate circuit just as in a resistance-capacitance-coupled amplifier. A blocking condenser is used to keep the direct current from flowing in the primary of the transformer. This condenser must be chosen so that it will not resonate with the primary inductance of the transformer at a frequency where this resonance would be objectionable. Sometimes, however, this resonance is made to occur at a frequency which will increase the bass response of the amplifier.

The power stage in the amplifier uses a pair of 6L6 tubes connected for Class A_1. The output transformer is provided with two output windings, one to feed a loudspeaker and the other to feed a 500-ohm line. Output transformers often use tapped output windings instead of separate windings, and provision is made for supplying several loudspeakers, as well as the 500-ohm output. The 500-ohm output is used to feed power into a line such as a telephone line, or to feed power over a transmission line to a loudspeaker located a considerable distance from the amplifier.

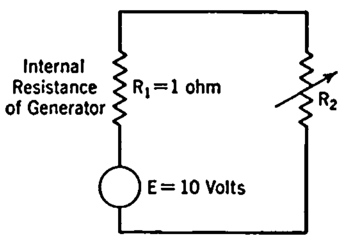


FIG. 7-28. Equivalent Circuit of a Generator Connected to a Load.

IMPEDANCE MATCHING. In connecting together the various parts of the public-address amplifier system that have been described above, there are some precautions to be observed, one of which is *impedance matching*. When two pieces of equipment are connected together, their impedances are said to be matched when the impedance of the load is adjusted to the best value to absorb power from the source. Impedances are matched for two reasons: (1) to obtain maximum power output; (2) to reduce distortion.

To see why impedance matching gives maximum power output, consider the circuit shown in Fig. 7-28, which consists of a generator (such as a battery, a power amplifier, or any device capable of supplying electrical power) and a load resistance. All generators have some internal loss which can be considered as an *internal resistance.* In the figure, R_1 represents the internal resistance of the generator, R_2 the load resistance and E the voltage of the generator. The problem is to determine what the size of R_2 must be to give maximum power output.

Suppose the internal resistance of the generator is 1 ohm and its voltage is 10 v. Then if various values of load resistance are connected to the generator, that is, if various values are assigned to R_2, the current can be calculated by Ohm's Law and the power in the resistor R_2 determined. Table 7–1 shows the results of some calculations for various values of R_2.

TABLE 7–1

R_2 (ohms)	Current (amperes)	Voltage across R_2	Power in R_2 (watts)
0.1	9.07	.907	8.20
0.5	6.67	3.33	22.2
1	5	5.00	25
4	2	8.00	16
100	.099	9.99	.998

An examination of this table shows that the maximum power that it is possible to take from the generator is 25 watts, and it occurs with a load resistance of 1 ohm, that is, a load resistance equal to the internal resistance of the generator. It can be shown that no matter what the voltage or the internal resistance of the generator, the *maximum power output* is obtained when the load resistance equals the internal resistance of the generator, that is, when their impedances are matched. However, it should be noted that the *voltage* across the load resistance increases as the load resistance is increased, and is not a maximum when the impedances are matched. Hence if the maximum *voltage* across the load is required, the impedances should not be matched, but the load impedance should be as large as possible. When the source is a vacuum tube there is a definite load resistance which will give the maximum power with a particular allowable amount of distortion. This load is specified by the manufacturer.

In a long transmission line such as a telephone line, it is found that signals transmitted over it will be distorted unless the load resistance at its output end has a particular value, called the *characteristic impedance* of the line. The characteristic impedance of the line is the impedance that must be matched for maximum output and minimum distortion. If the terminal impedance is not equal to this characteristic impedance, it is found that some of the energy sent down the line is *reflected* when it strikes the impedance mismatch at the end. If the line is long, reflected energy is largely lost, and since the amount of reflection depends somewhat on frequency, different amounts of power will reach the load resistance depending on the frequency. This variation produces distortion of the transmitted signal, which can be avoided by matching impedances. The characteristic impedance of most telephone lines is about 500 ohms.

Another case where mismatched impedances may cause distortion is in circuits involving transformers and in long microphone cables. In these cases the distortion is not usually serious except for widely mismatched impedances.

The microphone cables ordinarily used are high-capacitance cables, so that if a long length of such a cable is used to connect a high-impedance microphone to the grid of the preamplifier, the capacitance of the cable which is in parallel with the high impedance of the microphone by-passes the higher frequencies. In order to reduce the loss of high-frequency response, it is usual to transform the high impedance of the microphone to a low impedance by means of a microphone transformer or by an amplifier at the microphone with a transformer output. The capacity of the line has much less shunting effect on the low-impedance output of the transformer, so that longer lines can be used without losing high-frequency response. Microphone cables as well as transmission lines are designed to have low characteristic impedances, since low-impedance lines pick up less hum and noise than high-impedance lines.

OPERATION OF PUBLIC-ADDRESS SYSTEMS. In the operation of public-address systems it is often necessary to have the microphones and loudspeakers in the same room or auditorium. Such arrangements cause considerable difficulty due to *acoustic feedback* from the loudspeaker to the microphone. Any noise in the room picked up by the microphone is amplified and fed to the loudspeaker, which reproduces the noises louder than before. This louder noise repeats the same cycle and thus the sound builds up to the limit of the amplifier. The result is a howl at a frequency for which the over-all system has most gain. Feedback can be avoided by taking suitable precautions to limit the amount of sound from the loudspeaker that can be picked up by the microphone. Methods for doing this include the use of directional loudspeakers such as horns, curtains or drapes on the walls to reduce reflections, directional microphones, and careful control of the volume. Manipulation of the tone control on the amplifier will sometimes help to stop acoustic feedback.

Where several loudspeakers are to be used, care must be taken to see that the audience will not be able to hear the sound from two or more loudspeakers at different distances with about the same volume, otherwise echoes will be heard which make hearing difficult. When only one microphone and one or two loudspeakers are employed, a good arrangement is to put the loudspeakers above and slightly forward of the microphone. This makes the sound from the loudspeakers reach the audience at about the same time as the direct sound from the speaker and thus avoids echoes.

Since the characteristics of rooms vary so widely it is difficult to lay down specific rules for placing microphones and loudspeakers. Each particular job requires special consideration, so that either experience or trial and error must be used to determine the correct placements.

Review Questions and Problems. *1.* The energy supplied to the speaker of a radio is many times larger than the energy received at the antenna. From where does this extra energy come?

2. What is meant by a *flat response* for an audio amplifier? Is such a response desirable?

3. What determines the operating class of an amplifier? Distinguish between Class A, Class AB, and Class B operation.

4. Why aren't amplifier tubes hooked in cascade with the plate of one tube connected directly to the grid of the next tube?

5. What is meant by the *operating voltages* of a tube?

6. What is the function of the grid-leak resistor? What is the *time constant* of a coupling condenser of 1 mf together with a grid-leak resistor of 1 megohm? Is this a suitable value for use in an amplifier?

7. How does an increase in the plate resistor increase the amplification of a triode amplifier? Why is not the resistor made extremely large in practical amplifiers?

8. How would an increase in the size of the coupling condenser of a resistance-coupled amplifier affect the voltage gain of the amplifier at high frequencies and at low frequencies? How would an increase in the value of the grid-leak resistor affect the voltage gain at high frequencies and at low frequencies? Should the plate resistor be large or small compared to the grid-leak resistor of the next tube?

9. What characteristic of a pentode determines its worth in a resistance-coupled voltage amplifier? How does the value of the load resistance affect the voltage amplification?

10. Why is the condenser C_d placed in the circuit of Fig. 7–5? How is the grid bias obtained in this circuit? What components prevent feedback through the common plate supply?

11. If the gain in decibels of each stage of an amplifier is known, how can the over-all gain be obtained? What is meant by the expression "down 3 decibels at 5,000 cycles"?

12. Explain the various ways in which amplitude distortion may be produced. How does the dynamic I_p-E_g characteristic differ from the static characteristic? What is the *load line*, and how is it used in determining the dynamic characteristic?

13. Study Fig. 7–8 and devise an experimental method for testing for the presence of second-harmonic distortion. (How does the average value of the plate current change with grid excitation?)

14. Explain the salient features of the frequency response of a transformer-coupled amplifier. How does this response depend upon the plate resistance of the tube?

15. Draw a diagram of a simple resistance-coupled triode amplifier showing all of the necessary bias resistors and by-pass condensers for power-pack operation.

16. Draw a circuit diagram of a simple resistance-coupled pentode amplifier showing all the necessary bias resistors and by-pass condensers for power-pack operation.

17. Draw a circuit diagram of a push-pull amplifier. How do the output, distortion, allowable grid swing, and power-supply hum for a push-pull amplifier compare with those of a single triode stage?

18. Distinguish between power amplifiers and voltage amplifiers. Where in a radio would you use power amplification? Where voltage amplification?

19. Arrange Class A_1, AB_1, AB_2 and B amplifiers in the order of their power output and again in the order of the distortion they produce. Which classes of operation demand large input power?

20. What class of operation would give the largest undistorted power amplification in push-pull operation?

21. What is feedback, regenerative feedback, degenerative or inverse feedback? Explain how inverse feedback reduces frequency distortion.

22. What special characteristic must a video amplifier have? Why isn't transformer coupling used in video amplifiers?

23. What are the limiting factors to the high-frequency response of the circuit in Fig. 7–16? How does the circuit of Fig. 7–17 extend the frequency response over that of Fig. 7–16?

24. Does an increase in the load resistance of a triode amplifier increase or decrease the amplitude distortion? Does this apply to pentode amplifiers? How would the voltage gain be affected in these two cases?

CHAPTER 8

Vacuum-tube Instruments

A large variety of instruments depending for their operation on electronic tubes has been developed for use in laboratory investigation and measurements, not only for audio- and radio-frequency use, but also for D.C. applications. The most important of these are oscillators of various types, cathode-ray oscillographs, and vacuum-tube voltmeters.

Oscillators. As stated in Chapter 4, if a portion of the amplified output of a vacuum tube is fed back into the input circuit in the proper phase, self-sustained oscillations will be produced. These oscillations occur at the frequency at which the gain through the tube is a maximum, usually the natural frequency of any resonant circuit that may be present.

Oscillator circuits are used as *signal generators* to provide audio- and radio-frequency voltages for laboratory and shop testing; and for maximum usefulness in such service they should meet certain requirements. Usually the frequency needs to be adjustable over a wide range, but for any setting of the control dial it should be accurately known, and it should be perfectly constant and stable; there should be very little or no *frequency drift.*

Particularly in audio-frequency oscillators, it is important that the wave form of output voltage be very good, with total harmonic content not more than 1%. This high quality is necessary when distortion introduced by an amplifier is to be determined, since there is no satisfactory method of separating such distortion from that introduced into the amplifier input from the signal source.

In most cases, the magnitude of the output voltage should remain constant or nearly so over the entire frequency range of the oscillator, and this voltage should be easily and smoothly adjustable from zero to the maximum obtainable. In the true signal generators, there should also be an accurate calibration of the output voltage, so that the strength of the signal is known. This requirement can be waived in the case of oscillators used for bridge measurements, frequency comparisons, receiver alignment, and so forth.

INDUCTANCE-CAPACITANCE (*L–C*) OSCILLATORS. Figure 8–1 shows a typical oscillator circuit of the type using a resonant circuit for frequency control. In this circuit, the oscillation frequency is very nearly equal to the resonant frequency of L_g and C_1. Plate output power is fed into the grid circuit by way of the mutual inductance M between L_g and L_p. In order to produce oscillations, this coupling must be in such phase that a current flowing in the *tank circuit* composed of L_g and C_1 will produce voltages on plate and grid which are opposite in phase. Furthermore, the coupling must be large enough so that the power transferred from the plate circuit is more than enough to overcome the losses in the tank circuit and in the grid-leak resistor R_g. Roughly speaking, this requirement means that the grid voltage produced by this coupling must be slightly in excess of $1/\mu$ times the plate voltage.

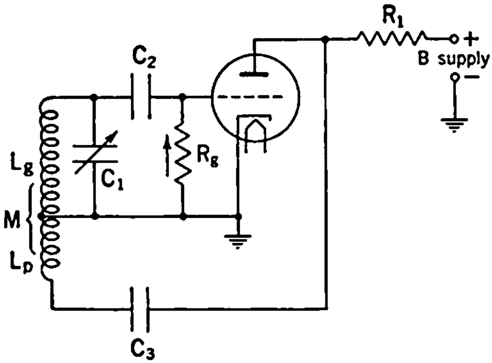


FIG. 8-1. *L-C* Oscillator.

In this circuit, grid-bias voltage is produced by flow of grid current through the grid-leak resistor R_g. This current can flow in only one direction, shown by the arrow, and in passing through R_g it produces a negative potential at the grid.

The condenser C_3 serves as a *blocking condenser*, preventing a short-circuit of the D.C. plate potential through L_p, but yet permitting the alternating plate current to flow through L_p. Resistor R_1 isolates the B supply from the A.C. plate potential, and the output voltage of the oscillator may be obtained across R_1. Resistor R_1 may be replaced by a radio- or audio-frequency choke, according to the frequency range of the oscillator, and this substitution will cause an increase in the strength of oscillations by raising the D.C. plate potential.

This circuit employs *shunt* or *parallel* plate feed, since the A.C. and D.C. components of plate current flow in separate paths. It is also possible to employ *series* plate feed, for example by connecting the B-supply voltage between L_g and L_p and omitting C_3 and R_1.

The operation of any oscillator of this type may be analyzed by considering it as a Class C amplifier, with a portion of its output utilized to supply the driving power which otherwise would come from a preceding stage or driver.

RESISTANCE-CAPACITANCE (*R-C*) OSCILLATORS. An altogether different type of oscillator which is steadily gaining prominence depends for its operation on combinations of resistance and capacitance instead of inductance and capacitance. One advantage of the *R-C* type is that the frequency produced is not so much affected by the tube constants as is the case with the *L-C* type, a condition which makes for better frequency stability. Another advantage is that a wider frequency range can be covered on each band. For example, if C_1 in Fig. 8-1 is a variable air condenser having a ratio of maximum to minimum capacitance of 10:1, the frequency range for a given coil will be nearly in the proportion $\sqrt{10}$:1, say 3:1. In the *R-C* oscillator, for the same type of condenser, the frequency range for a fixed value of resistance will be 10:1, and therefore fewer bands of frequency will be needed to cover a given total range. Still another advantage is that it is somewhat easier to obtain good wave form with the *R-C* oscillator than with the *L-C* type.

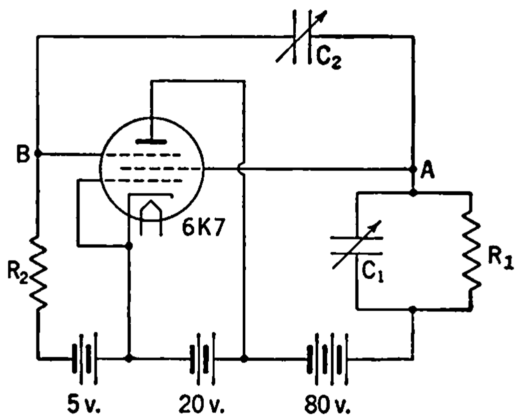


Fig. 8–2. *R-C* Oscillator.

A compact oscillator of the *R-C* type is shown schematically in Fig. 8–2. The tube employed is a standard voltage-amplifier pentode such as the 6K7, but it is used in an unconventional manner. The operation of the circuit depends upon the relation between suppressor voltage and screen current, with control-grid and plate potentials held constant. This relation is a mutual characteristic, similar to the one between grid potential and plate current discussed in Chapter 4, but with opposite slope. The characteristic is shown in Fig. 8–3, which shows that a *positive* change in suppressor potential produces a *negative* change in screen current. Another way of stating this fact is that the mutual conductance, suppressor-to-screen, is negative.

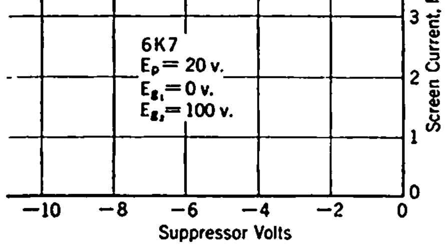


Fig. 8–3. Mutual Characteristic of 6K7 Tube. Suppressor Voltage–Screen Current.

Referring again to Fig. 8–2, assume that a small increase of suppressor potential occurs. This increase will cause a decrease in screen current, on account of the negative mutual conductance, and the decrease of current is accompanied by an increase of screen voltage because of the reduced drop in the screen load impedance. This increased potential is fed back to the suppressor through C_2, and reinforces the original change. The action is therefore cumulative and proceeds in the same direction until limited by change of tube characteristics in one form or another. When the excess charges on the condensers start leaking off, the same sequence of events takes place in reverse direction, and an oscillation is set up. If the resistances R_1 and R_2 are reduced until oscillations can just barely be maintained, the wave forms of the currents and potentials are very nearly pure sine curves, and the frequency stability is very good. The frequency of the oscillation is given by

$$f = \frac{1}{\pi\sqrt{R'_1R_2C_1C_2}},$$

where R'_1 is the parallel combination of R_1 and the A.C. screen resistance.

It is convenient to control the frequency by varying the condensers C_1 and C_2, which may be identical air condensers mounted on a single shaft. If this is done, the frequency is inversely proportional to the capacitance of either section, and the range from minimum to maximum condenser settings will be in a ratio of more than 10:1.

Cathode-ray Oscillographs. The operation of the cathode-ray tube has already been described in Chapter 4, and mention made of its most im-

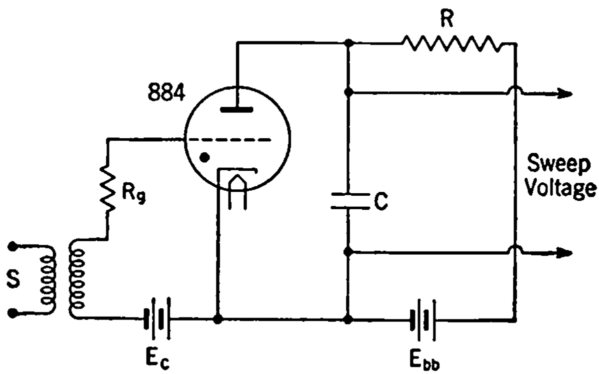


FIG. 8-4. Cathode-ray Sweep Generator.

portant application as an oscillograph. A necessary auxiliary for this instrument is some provision to produce a linear sweep across the screen of the tube, so that vertical deflecting voltages may appear as a curve plotted against time. This sweep must be recurrent at a rate which is an integral submultiple of the frequency being observed, so that successive traces coincide on the screen and give a stationary pattern. One method of obtaining a suitable voltage for this purpose is shown in Fig. 8-4.

The tube used in this circuit is a gas triode such as the 884. This tube contains a definite amount of argon gas, and conduction occurs by ionization of the gas when the plate potential exceeds a critical value. The critical plate potential depends upon the potential of the grid in the manner shown in Fig. 8–5.

When the circuits of Fig. 8–4 are first closed, the plate potential of the tube will be zero, since the condenser C is uncharged. Current flowing through R charges the condenser, raising the plate potential of the triode, until the critical potential is reached. The tube then becomes a very good conductor and effectively short-circuits C. The plate potential falls

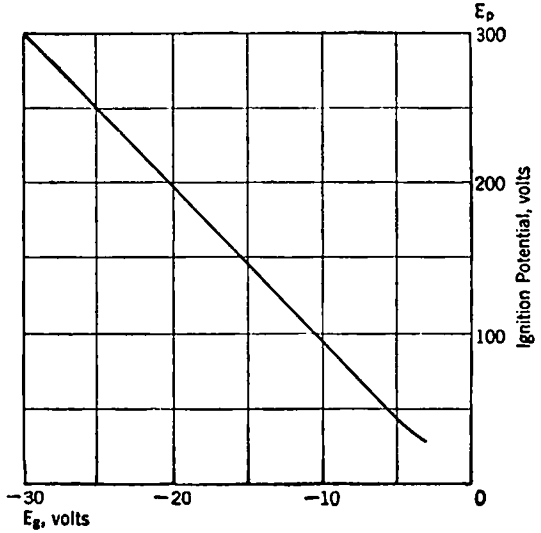


FIG. 8–5. Control Characteristic of Type 884 Gas Triode.

abruptly to a value too low to sustain the current flow through the tube. This lower value is the *extinction potential* of the gas, and is nearly equal to the ionization potential as given in Chapter 4. The cycle then is repeated, and the variation of plate potential is as shown in Fig. 8–6.

It will be seen that the curve of plate potential against time is a typical exponential charging curve, but if only the lower portion is utilized a good approximation to a straight line is obtained. If this voltage is used to produce horizontal deflection on a cathode-ray screen, the spot will travel at nearly uniform velocity during the interval *a-b*, then return almost instantaneously to its starting point, *b-c*, and repeat. The frequency with which this cycle occurs may be controlled by adjusting R, C, or E_c, as can be seen. It is customary to vary R for the fine control, and to make C adjustable in steps to give the necessary range in frequency. R_g is a protective resistance to limit flow of grid current.

The linearity of the sweep depends chiefly upon the ratio of B-supply voltage to striking voltage of the tube, being better as this ratio is higher. Variation of E_c will therefore have an effect on linearity, as well as on the amplitude of sweep voltage; hence this variation is not a desirable means of frequency control.

The sweep frequency may be *synchronized* with the signal frequency on the vertical plates, so that successive traces on the screen are exactly coincident and the pattern observed is completely stationary. This synchronizing is accomplished by applying a small voltage of the signal frequency to the terminals S, and through the transformer to the grid of the 884. The sweep frequency is adjusted to be slightly low, and then the positive swing of the A.C. grid voltage determines the instant at which striking potential is attained. The effect is indicated on the third cycle of Fig. 8–6, and the return trace is seen to occur slightly earlier than it would in the absence of synchronizing voltage.

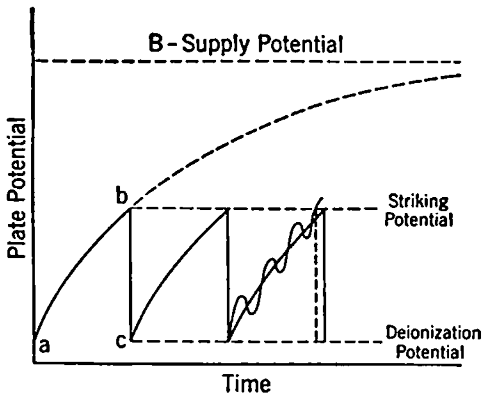


Fig. 8–6. Plate Potential of the Sweep Generator Tube.

USE OF THE CATHODE-RAY OSCILLOGRAPH. The commonest application of the cathode-ray oscillograph is the one already referred to, namely to plot a voltage against time on a linear base. Ordinarily it is necessary to amplify the voltage under observation, and the properties of the deflection amplifier must be considered. Obviously its frequency range must include that of the voltage being observed, but if the wave form as such is of importance, the amplifier must be capable of passing all the important harmonics, which ordinarily means that it must include frequencies up to ten times the signal frequency. In many cases it is important that this amplifier have no delay distortion, which is equivalent to saying that the time delay should be uniform at all frequencies, or that the phase shift should be proportional to frequency.

Another consideration is the type of circuit in which the voltage appears—whether balanced or unbalanced to ground. Some communication circuits have the two sides at equal and opposite potentials with respect to ground; others have one side of the circuit grounded, the opposite side being "hot". The same distinction occurs in input amplifiers for cathode-ray oscillographs, and if an oscillograph input amplifier having one side grounded is applied to a balanced circuit, the effect will be to ground one side of the circuit. This grounding may disturb the distribution of voltage and current in the circuit and give an altogether erroneous picture of what is taking place. Care should always be taken to preserve the balance of any circuit to ground, to avoid misleading results. Many of the more complete cathode-ray oscillographs include provision for use with either balanced or unbalanced circuits.

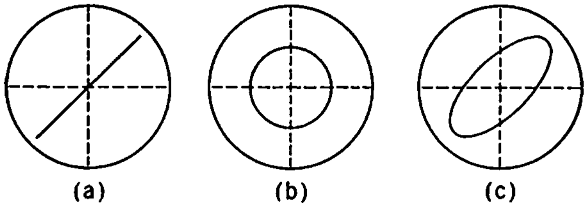


Fig. 8–7. Lissajous Figures, Equal Frequencies on Horizontal and Vertical Plates.

Lissajous figures. Another application of the oscilloscope is in the comparison of frequencies. If voltages from two separate sources are impressed on the horizontal and vertical deflection plates respectively, the pattern observed on the screen will be stationary only if one frequency has a ratio to the other represented by the ratio of two integers. Figure 8–7 shows patterns obtained when the two frequencies are equal: (a) when the two voltages are exactly in phase with each other, (b) when the voltages are 90° apart in phase, and (c) when some intermediate phase relation exists. Figure 8–8 shows the appearance of the screen pattern if the frequency applied to the vertical plates is three times as great as that on the horizontal plates. This ratio can easily be recognized by tracing out the pattern and noting that three complete vertical cycles occur in one traverse but only one horizontal cycle.

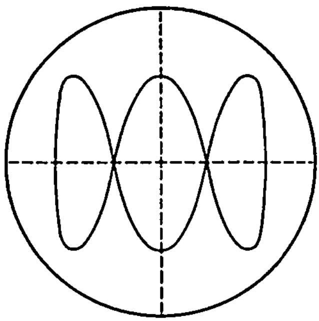


Fig. 8–8. Lissajous Figure, Frequency Ratio 3:1.

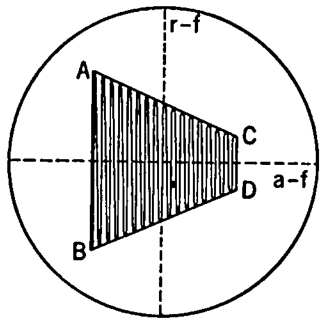


Fig. 8–9. Trapezoid Pattern Showing Modulation.

If a modulated radio-frequency voltage is applied to the vertical plates, while the modulating voltage (audio-frequency) is applied to the horizontal plates, a pattern such as is shown in Fig. 8–9 is obtained. The degree of modulation is easily determined:

$$m = \frac{AB - CD}{AB + CD}.$$

The linearity of the modulation is indicated by the straightness of the lines *AC* and *BD*, while if any phase shift occurs in the modulation process it will be shown by the appearances of elliptical curves in place of the lines *AC* and *BD*.

Vacuum-tube Voltmeters. It is essential that voltmeters for use in many communication circuits have a very high input impedance, so as not to load the circuits appreciably, and essential also that their indications be independent of frequency over a wide range. Both these requirements

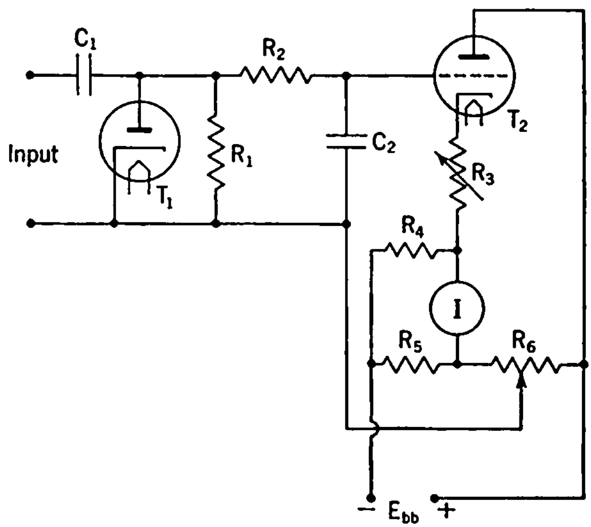


FIG. 8–10. Diode Vacuum-tube Voltmeter with D.C. Amplifier.

can be met more easily by vacuum-tube voltmeters than by any other type of instrument.

The vacuum-tube voltmeter may take many forms, but a very satisfactory type is shown in Fig. 8–10.* Essentially this consists of a diode rectifier followed by a D.C. amplifier. Alternating voltage applied at the input terminals produces a unidirectional current in the diode T_1 and in R_1. The voltage developed across R_1 by this current is pulsating, but the filter

* *General Radio Experimenter*, *9*, 12, May, 1937.

formed by R_2, C_2 passes only its D.C. component to the grid of T_2, and the resulting change of plate current is read on the meter I. The resistor R_3 provides a considerable amount of negative or inverse feedback (Chapter 8), which causes the calibration of the instrument to be nearly independent of tube characteristics. It is made adjustable to provide several ranges of full-scale voltage. The other resistors are used to provide proper bias voltage for the D.C. amplifier and to balance out the steady or zero-signal component of plate current from the meter circuit.

The input impedance to this type of voltmeter is roughly one fourth the resistance of R_1, which in the commercial instrument is 50 megohms. The amount of power taken from the circuit being measured is therefore entirely negligible in almost any application. The calibration is independent of frequency over the entire audio range, and for all radio frequencies to 50 mc per second.

For some purposes the vacuum-tube voltmeter just described does not have sufficient sensitivity, for example the measurement of radio-frequency voltages in the early stages of a radio receiver. The detecting efficiency of a diode is low for inputs less than 1 v, and increased sensitivity therefore requires preliminary amplification of the voltage to be measured. The preliminary amplifier may consist of one or more stages of audio- or radio-frequency amplification, depending on the required sensitivity and frequency range.

Review Questions and Problems. *1.* In the circuit of Fig. 8–1, the condenser C_1 has a range from 30 to 450 $\mu\mu$f. Neglecting any stray capacity, determine the inductances L_g to provide a frequency range from 50 kc to 20 mc, allowing 10% overlap in frequency from band to band.

2. In the circuit of Fig. 8–2, condensers C_1 and C_2 both have ranges from 30 to 450 $\mu\mu$f. Determine resistances R'_1 and R_2 to provide a frequency range from 40 cycles to 50 kc, allowing 10% overlap in frequency from band to band.

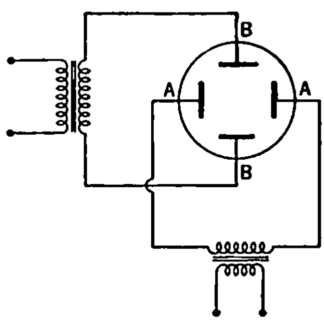


Fig. 8–11.

3. Find the mutual conductance from suppressor to screen for the 6K7 tube, making use of the curve in Fig. 8–3. Compare with the ordinary mutual conductance from control grid to plate. Can you account for the difference in magnitude?

4. Determine the pattern produced on the cathode-ray screen if a 60-cycle sine wave of voltage is supplied to plates A–A and a 120-cycle sine wave of voltage is supplied to plates B–B of Fig. 8–11. Phase relation is such that the positive peaks of the two voltage waves are simultaneous.

5. Repeat Problem 4, if the phase relation is such that the zero points of the two voltage waves are simultaneous.

6. If R_3, R_4, R_5, R_6 in Fig. 8–10 each have a value of 50,000 ohms, and $E_{bb} = 150$ v, what value of plate current in T_2 will produce zero current in the indicating meter? Does change in the value of R_3 affect this result? If the tube is a 6J5, what grid bias is required? What should be the setting of the tap on R_6 to obtain this bias? If R_3 is reduced to 25,000 ohms, what should be the tap setting on R_6 to return the meter current to zero?

CHAPTER 9

Electromagnetic Waves

Nature of Waves in Any Medium. Water waves, sound waves, radio waves—apparently widely different phenomena—have certain characteristics in common.

Each type of wave provides a means for transferring energy. It is probable that nearly all energy is transmitted by means of wave motion. When a large steamer plows through the water and sets up waves that rock a small boat half a mile away it is easy to see that the energy required to do the rocking was transmitted by means of waves. When speech is heard across a room and energy from a radio station is received a thousand miles away it is not difficult to believe that the energy was transmitted by means of waves. However, when a man pushes on one end of a steel bar and the other end pushes against an object he is moving with it, it is not so obvious that the energy has been transmitted by wave motion. But if the vibrating cone of a loudspeaker were alternately pushing and pulling at one end of the bar, the wave motion which carried the vibration to the other end of the bar would become more apparent. In this case, if the alternations were rapid enough, it might be found that when the front end of the bar was pushing forward the other end might already be pulling backward. This would be because of the time taken for the wave to travel down the bar. This same time is required when the man pushes the bar, but it is so small that its existence is not generally realized with this type of motion.

Figure 9–1 is a representation of wave motion. Figure 9–1 Part A might represent a cross-section of a water wave at a particular instant. If the wave is moving from left to right part B would be a picture of the wave an instant later. It will be seen that the crest of the wave which was at position 1 in A has moved over to position 2 in B. Part C shows the same wave at a still later instant, at which time the crest has moved to position 3. The rate at which the wave is moving from left to right is called the *velocity of the wave.* Next consider the motion of the point *a*, which might be a cork floating on the water or a particle of the water itself. At the instant represented by A the particle *a* is on the crest of the wave. An instant later the crest has moved on and the particle has dropped down, as shown in B. Still later it occupies the position shown in C. While the motion of the *wave* has been continuously forward, the motion of a *particle* such as *a* has been up and down along a vertical line such as 9. The maximum distance either side of the line *fg* traversed by the particle is called the *amplitude* of the wave. This is shown as the length *h* in part A and corresponds to the definition of the amplitude of a sine wave given in Chapter 3. The distance between successive crests of the wave is called a *wave length*

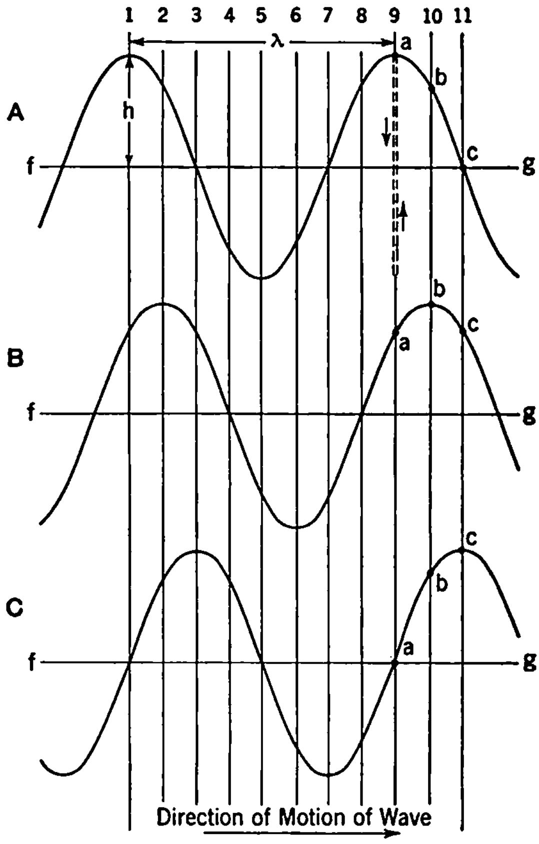


FIG. 9–1. Representation of Wave Motion.

and is represented by the Greek letter λ (lambda). Of course this is also the distance between successive troughs or any two corresponding points on successive waves.

The number of oscillations per second made by a particle such as *a* is known as the *frequency*, and is designated by the letter *f*.* If the waves

* No confusion is to be expected as between the symbol *f* for frequency and the abbreviation f for farad.

were being generated by moving a board up and down in the water the frequency would depend upon the number of times per second the board was moved up and down. That is, *the frequency depends upon the source.* On the other hand, the speed with which the waves traveled outwards would be independent of how rapidly the board was moved up and down and would depend only upon the properties of the medium—in this case upon the properties of water. If some other liquid such as oil or alcohol were used, the velocity of the waves would be different. *The velocity of propagation depends only upon the medium* and is independent of the source of the waves.

The frequency with which the particle *a* moves up and down along the line 9 is also the frequency with which the waves are going past the line, for each time *a* reaches a top peak in its journey up and down, the crest of a wave is going past.

For a given frequency f and velocity of propagation V, the wave length λ is fixed and is given by

$$\lambda = \frac{V}{f}.$$

λ is usually expressed in meters, V in meters per second, and the frequency f in number of cycles per second.

This can also be written as

$$V = \lambda f,$$

which states that the velocity with which the wave is moving is equal to the length of a wave times the number of waves per second passing a given point.

The velocity of sound in air is about 344 m per second, so that a 344-cycle oscillation would produce a wave 1 m long. The velocity of electromagnetic waves is 300,000,000 m per second and with them it requires an oscillation frequency of 300 mc to produce a wave 1 m long.

Transverse and Longitudinal Waves. The waves pictured in Fig. 9–1 are known as transverse waves because the motion of the particle is at right angles to the direction of motion of the wave. That is, the wave is

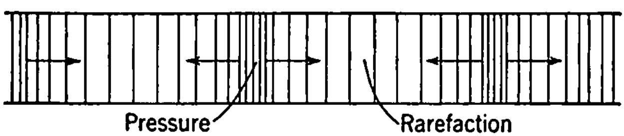


Fig. 9–2. Longitudinal Wave Motion.

moving forward, left to right, and the particle is moving up and down. With sound waves in air, on the other hand, the particle motion is back and forth in the direction in which the wave is moving. Such a wave is called a *longitudinal* wave. It is illustrated in Fig. 9–2. The density of the lines represents the pressure in that region. Pressure maxima and minima correspond to the crests and troughs of the wave of Fig. 9–1. The wave of pressure is moving from left to right and the arrows indicate the direction in which the particles are moving. The particles move in both directions from a pressure maximum in toward a pressure minimum. This particle movement results in a pressure maximum being formed where a pressure minimum existed a moment before; and in this manner the wave moves on. The individual particles, however, only oscillate back and forth in a manner similar to the up-and-down oscillation of the particle in the case of transverse waves.

Phase in Wave Motion. In Fig. 9–1, if attention is concentrated upon the motion of two particles, say *a* and *c*, it will be noticed that *c* does exactly what *a* does but at a later time. In A, the particle *a* is at the crest and the particle *c* is half way between trough and crest but on the way up. In B, *a* is on the way down and *c* is still on the way up. In C, *a* is half way down and *c* has just reached the crest. The motions could be followed through a complete cycle and it would be found that the movement of *c* was similar to that of *a* but coming after it by a constant interval of time. The particle *c* is said to *lag a* in *phase*. There is a *phase difference* between *a* and *c* and this difference may be expressed as a fraction of a cycle. In this case *c* lags *a* (or *a* leads *c*) by a quarter of a cycle. A complete cycle consists of 360°, so that a quarter of a cycle is 90°. Phase difference is usually expressed in degrees, as, for example, *c* lags *a* by 90°.

Phase difference is of importance when more than one wave is involved. If two waves of the same frequency are moving together in the same direction they will combine to give a resulting wave which will be the sum of the

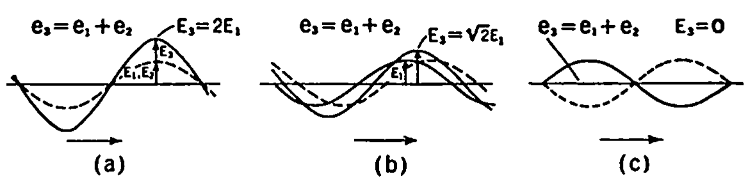


FIG. 9–3. Addition of Sine Waves: (a) in phase, (b) 90° phase difference, (c) 180° phase difference.

two waves. If the waves have the *same phase*, that is, if corresponding particles on the two waves reach the crest at the same instant, the amplitude of the resulting wave will be just double that of a single wave. This is shown in Fig. 9–3a. However, if the waves have 90° phase difference as shown in Fig. 9–3b, the amplitude of the resulting wave is only 1.414 (or $\sqrt{2}$) times that of a single wave. This figure can be proven by adding the two waves together point by point. Figure 9–3c shows the special case where the waves differ in phase by 180°. In this case, because the particle motion due to one wave is exactly equal and opposite that due to the other at all points and all times, the amplitude of the resulting wave is zero. That is, the waves have canceled each other. Complete cancellation can occur only if the two waves are of equal amplitude. This case will be of particular interest in the study of directional antenna arrays.

Although the amplitude of the resulting wave can be obtained as above by plotting the waves and adding them point by point, it can be obtained much more rapidly using a simple geometrical construction. This construction is shown in Fig. 9-4 for the three cases just discussed. The

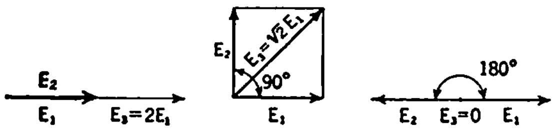


FIG. 9-4. Vector Addition of Sine Waves.

procedure is to lay off from an origin or point O two lengths proportional to the amplitudes of the two waves being added. The lengths are drawn with an angle between them equal to the phase difference. For the three cases being considered the phase differences were 0°, 90°, and 180°. With the lengths laid off the parallelogram is completed and the length of the diagonal from O will be proportional to the amplitude of the resulting wave. It will be seen that this method gives the same answers in the three cases considered as did the point-by-point addition.

Reflected Waves and Standing Waves. When a moving wave such as a water wave strikes a boundary such as a solid wall, the original wave is abruptly halted but a reflected wave is set up which travels in the direction opposite to that of the original wave. This combination is shown in Fig. 9-5, where the incident wave traveling from left to right is shown by the solid curve and the reflected wave traveling from right to left is shown by the dotted curve. This gives rise to two waves existing simultaneously at the same place, so they can be added together to show the resultant just as were the waves of Fig. 9-3. However, these waves are traveling in opposite directions so that it is not surprising to find a result different from that obtained in Fig. 9-3. Parts A to I of Fig. 9-5 show the original wave at successive instants of time. It is traveling forward, left to right, as was the wave of Fig. 9-3. The reflected wave is also shown at these same instants. It is traveling backward, right to left. It will be noted that the reflected wave is just the mirror image of the continuation of the original wave beyond the boundary. The resultant of the two waves is shown by the dashed wave in parts A to I. Figure 9-5J shows all the resultant waves of parts A to I plotted on top of one another, that is, J shows the resultant over a complete cycle. A striking feature of this is that at some points the resultant is *always zero.* This resultant occurs at one quarter of a wave length from the boundary, again at three quarters of a wave length and every odd quarter wave length from the end. Such points are called *nodes*. It will also be noticed that the resultant reaches its largest values at the boundary and at points distant from the boundary

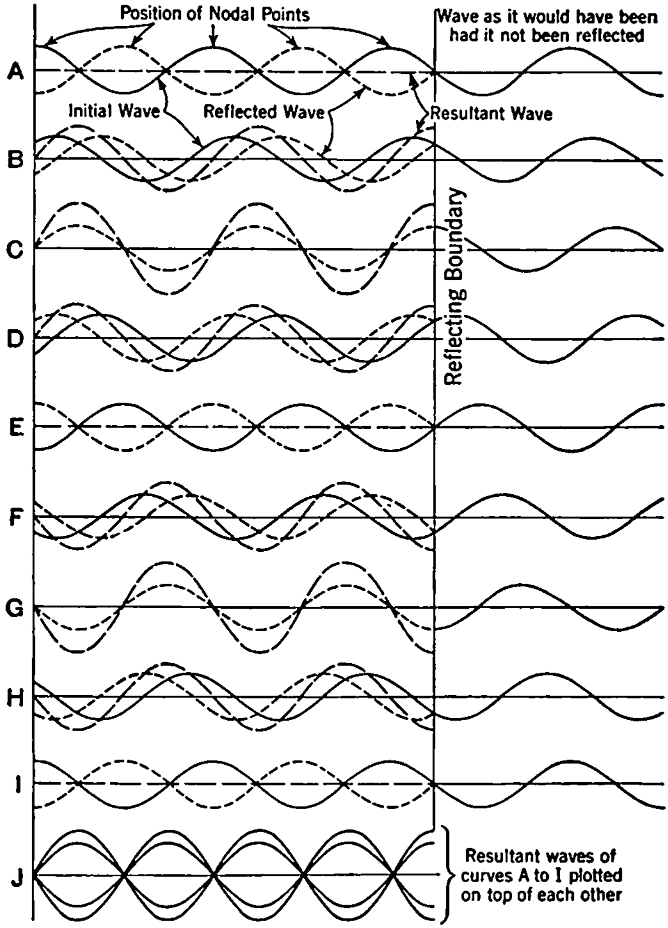


Fig. 9–5. Reflection of a Wave at a Boundary, Showing Addition of Initial and Reflected Waves to Give Standing Waves.

by multiples of one half wave length. These points of maximum amplitude are called *loops* or *antinodes*.

Figure 9–6 shows a more detailed picture of the resultant wave for the successive instants of time A through I over a complete cycle. It will be seen that although particles are everywhere in motion (except at the nodes) the wave seems to be *standing still* as there is no forward motion of the crests as was the case with the wave of Fig. 9–3. For this reason such a resultant wave is called a *stationary wave* or *standing wave* in contrast to the *traveling* or *progressive* wave of Fig. 9–3.

An effective illustration of such wave motion can be obtained with a length of rope. If the far end of the rope is left free and the rope is jerked up and down at the near end a wave of motion will be seen to travel down the rope. If the far end of the rope is then fixed solidly to a wall or other support and the near end of the rope is jerked vigorously, a wave will be seen to travel down the rope, be reflected at the boundary, and return to the sending end. The next step is to send a continuous wave motion down the rope by continuously moving the near end of the rope up and down. Interference between the incident and reflected wave will become evident, and if the frequency of the up-and-down motion is varied it will be found possible to produce standing waves on the rope. In this case the *nodes* will be at the end and half-wave-length distances from the end and the loops will come at the odd quarter-wave-length points. Whether a node or a loop appears at the reflecting boundary depends upon the *boundary conditions*. In the case of the rope fixed solidly at the far end, the boundary conditions were such that there could be no motion at this point—that is, it must be a nodal point.

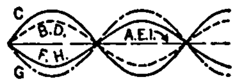


FIG. 9–6. Resultant Waves of Fig. 9–5 Over a Complete Cycle.

An understanding of the ideas of wave motion, that is, traveling waves and standing waves, is of great assistance in the study of antennas and their feeder systems.

Electromagnetic Waves on Wires. When a pair of parallel wires is used to connect a battery or generator to a load there will be a voltage V between the wires and a current I flowing through them as shown in Fig. 9–7a. Because of the voltage between the wires there will be an *electric field* about them which will everywhere have the direction shown by the *lines of electric force*. The density of these lines in any region is proportional to the *strength of the electric field* E (volts per centimeter) in that region (Fig. 9–7b). The electric field is strong near and between the wires and becomes weaker the further away from the wires one goes. It will be noticed that the lines terminate on the wires at right angles to their surfaces.

Because of the current I flowing through the wires there will be a magnetic field H surrounding them, as shown. The density of these lines is proportional to the *magnetic field strength* and their direction indicates the direction of the magnetic field. The lines of magnetic field strength H are everywhere at right angles to the lines of the electric field E.

The arrows on both the electric and magnetic field lines indicate the directions of the fields which correspond to the directions of voltage and current shown in Fig. 9–7a. If the generator terminals were reversed so that the top line became negative and current flowed from right to left in the top line, the directions of both electric and magnetic fields would be the reverse of those indicated by the arrows. If the D.C. generator were replaced by an A.C. generator so that voltage and current on the wires were alternating, the electric and magnetic fields would also be alternating;

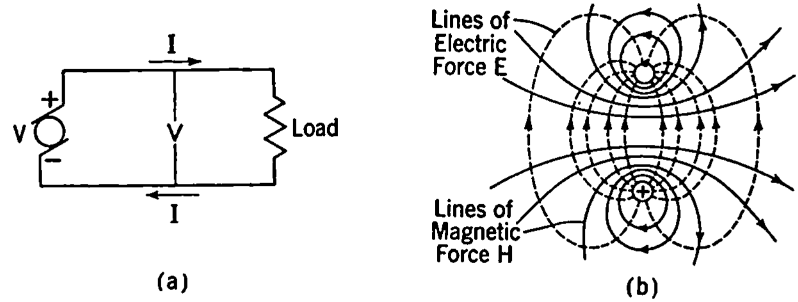


FIG. 9-7. Electric and Magnetic Fields about a Pair of Parallel Wires.

that is, their positive and negative directions would change with the reversal of voltage and current. Moreover, at the instant that the voltage and current were zero the electric and magnetic fields would also be zero, so that with A.C. operation the electric and magnetic fields are continually being built up and then collapsed back again to zero.

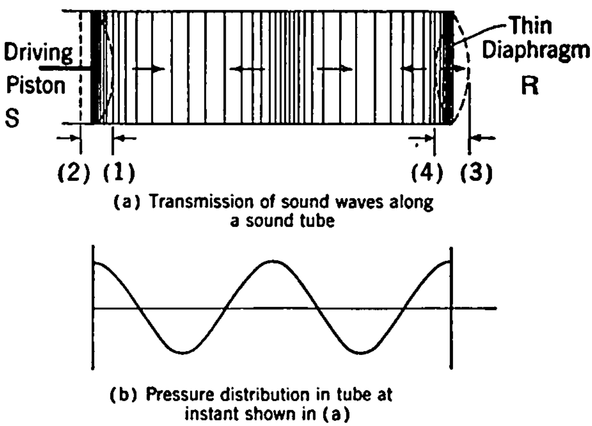


FIG. 9-8. Transmission of Sound Waves along a Tube, Showing the Instantaneous Pressure Distribution.

Sound Waves and Electromagnetic Waves. It will be interesting and instructive to compare the transmission of sound waves in a speaking tube with electromagnetic waves along a pair of parallel wires. In Fig. 9-8 the sound tube is shown with an oscillating piston or diaphragm at the sending end *S* and a flexible diaphragm at the receiving end *R*. When the piston is moved back and forth quite slowly, the pressure is the same all along the tube and the diaphragm R moves in and out with the driving piston S. R and S are then *in phase.* Actually there is a small time interval between the maximum forward position of S, position 1, and the maximum forward position of R, position 3. This time interval is so small compared with the time for a complete oscillation (a period) that it can be neglected in this case. However, as the piston is speeded up to a higher frequency, the time for an oscillation becomes small and the time interval required for the pressure produced at S to reach the diaphragm at R becomes important. When the piston is vibrating fast enough it will be possible to have it back at position 2 by the time R has reached position 3. In this case R would be lagging S by one half cycle or 180°. R and S are then said to be 180° out of phase. As the frequency is increased still further, the driving piston may have moved from 1 to 2 and back to 1 again by the time the pressure has reached R to move it to position 3. In this case R and S would move in and out together, but R would lag S by a complete cycle or 360°. For the case shown in Fig. 9–8 the piston has made two complete movements in and out before the disturbance has reached R, and R lags S by two complete cycles or 720°. Figure 9–8b shows the pressure which would exist along the tube at the instant that S and R are in their most forward positions. It is evident that *for this frequency* the tube is just two wave lengths long.

It is important to note the differences that exist between the slow and rapid operation of the driving piston. When the frequency of oscillation was low so that the corresponding wave length $\lambda = V/f$ was very large compared with the length of the tube, the pressure in the tube was everywhere the same and the direction of motion of the air particles was the same in all parts of the tube. However, when the frequency was increased so that the wave length was reduced to the same order of magnitude as the length of the tube, these conditions changed. The pressure was different in different parts of the tube and the particle velocity was forward in some places and backward in others. Quite similar effects will be observed with electric waves.

Figure 9–9 illustrates the transmission of electric energy along a pair of parallel wires, or a transmission line as it is generally called in communication work. As in the case of sound in a tube, as long as the alternations of the generator are slow so that $\lambda = V/f$ is large compared with the length of the line, the voltage will be practically the same all along the line (resistance drop considered negligible) and the current will be in the same direction along the line; that is, left to right on the top wire and right to left on the bottom, or vice versa. This situation is just the familiar 60-cycle case, because the wave length at a frequency of 60 cycles per second is

$$\lambda = \frac{300,000,000}{60} = 5,000,000 \text{ m} = 3,100 \text{ miles},$$

and this is long compared with any transmission line that might be used.

However, when the frequency is increased to very high values such as are used in radio work, the corresponding wave length becomes small and even short transmission lines may be several wave lengths long. Figure 9–9 shows the case for which the frequency has been increased until the wave length is just one half the length of the line or the line is two wave lengths long. The instantaneous direction of current is shown in Fig. 9–9a and the instantaneous voltage distribution is shown in Fig. 9–9b. The similarity to particle velocity and pressure in the sound tube will be evident. No longer is the voltage between the wires constant along the line, for as a voltage maximum leaves the generator and starts on its journey down the

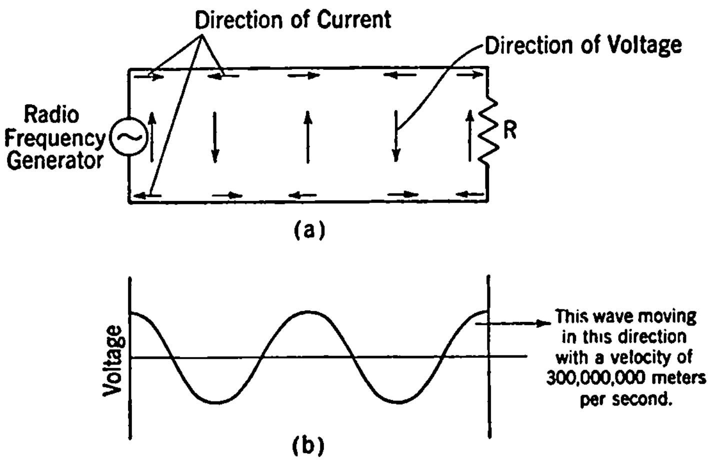


FIG. 9–9. Voltage and Current on a Transmission Line at the Instant of Maximum Generator Voltage.

line the generator voltage changes and goes through two complete alternations (in the case shown) before the first voltage maximum reaches the load at the end of the line. Similarly, the direction of current from the generator changes four times (two complete cycles) while a particular current crest is traveling down the line.

The idea of current leaving one terminal of the generator, going around the circuit and returning to the other terminal now becomes rather awkward, and it is better to think of it as positive current leaving one terminal and negative current leaving the other, and these two current mates traveling down the line together. Of course, when the alternator voltage and current reverse, the terminals from which the positive and negative currents leave will reverse and the currents along the line will be as shown in Fig. 9–9.

So far the transmission of energy from the generator to the load has been considered in terms of voltages and currents along the line. As was seen earlier, corresponding to these voltages and currents are electric and magnetic fields that surround the wires and travel down the line with their respective voltage and current mates. Actually, then, the energy is conveyed from generator to load *by these fields* through the space surrounding the wires, and the wires themselves merely *guide* the energy to its destination. This guidance corresponds to the acoustic case, where the sound tube serves merely as a guide, the sound energy being conveyed by the motion of the air in the tube. For the transmission of electromagnetic waves along wires it is immaterial whether one considers voltages and currents or electric and magnetic fields, but for the transmission of electromagnetic waves in space *where there are no wires*, consideration of the electric and magnetic fields is necessary.

Standing Waves. If the end of the sound tube of Fig. 9–8 is closed by a solid plate instead of the flexible diaphragm, the pressure wave is reflected

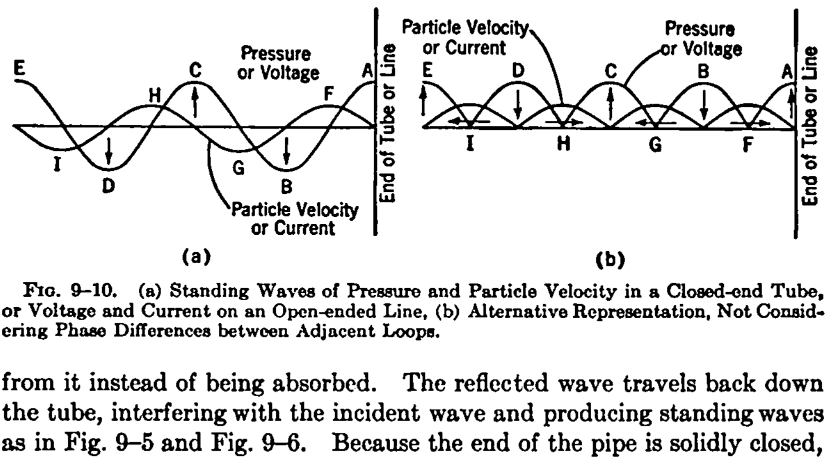


FIG. 9–10. (a) Standing Waves of Pressure and Particle Velocity in a Closed-end Tube, or Voltage and Current on an Open-ended Line, (b) Alternative Representation, Not Considering Phase Differences between Adjacent Loops.

from it instead of being absorbed. The reflected wave travels back down the tube, interfering with the incident wave and producing standing waves as in Fig. 9–5 and Fig. 9–6. Because the end of the pipe is solidly closed, the pressure can build up to a maximum and there is a pressure loop at the end and a pressure node one quarter of a wave length from the end as in Fig. 9–6. The particle velocity, however, must be zero at the end because the solid plate is immovable, and the layer of air next the plate must also have zero velocity. The distribution of particle velocity is therefore as shown in Fig. 9–10, with a node at the end and at half wave length from the end. Fig. 9–10b is an alternative representation of Fig. 9–10a. The only difference between two adjacent crests such as B and C is that they have 180° phase difference, the one being negative when the other is positive and vice versa. Because a pressure meter (in the case of sound) and a voltmeter (in the case of electricity) cannot measure *phase*, the pressure or voltage indicated by such instruments would be as shown in Fig. 9–10b.

If the transmission line of Fig. 9–9 is open at the end instead of being terminated in a resistance, standing waves of voltage and current are set up as in the case of the sound tube. With an open-ended line the current must be zero at the end but the voltage can go to a maximum, so that the voltage and current distribution are as indicated in Fig. 9–10. If, on the other hand, the line is shorted instead of being left open, reflection also occurs but in this case the voltage must be zero at the end (because of the short circuit) and the current can go to a maximum. For this case the voltage and current distributions shown in Fig. 9–10 are interchanged.

Waves in Three Dimensions. If either the sound tube of Fig. 9–8 or the transmission line of Fig. 9–9 is left open at the end, a certain amount of

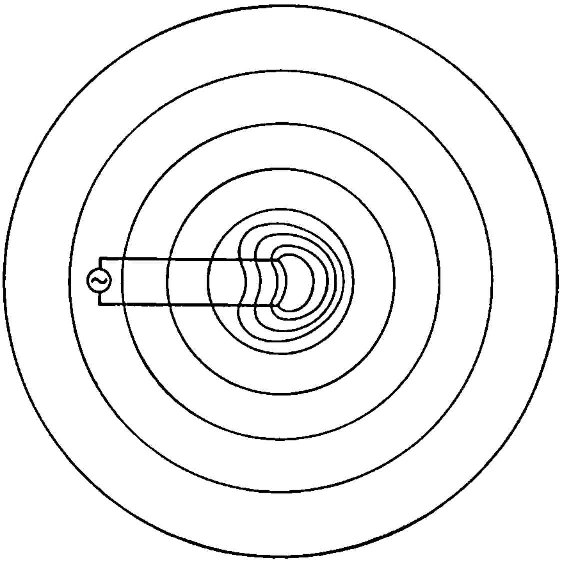


Fig. 9–11. Radiation from the Open End of a Transmission Line or Sound Tube.

the energy escapes or is *radiated* from the end. This energy spreads out in all directions in space in the manner shown in Fig. 9–11. Because the same amount of energy is spreading out through surfaces of ever-increasing size, the energy that flows *through a given area* decreases as the distance from the open end increases. This statement means that the pressure or electric-field strength, as the case may be, decreases with increasing distance. It is found that both pressure and electric-field strength are inversely proportional to r, the distance from the source. For the electromagnetic wave,

$$E = \frac{K}{r},$$

where K is a constant that depends upon the amount of energy radiated per second.

Because a radio wave released into space becomes weaker as the distance from the transmitter increases, it is important that as much energy as possible be radiated from the source. In the case of the sound tube, radiation from the end can be increased by opening the end out into a horn. The effect of this enlargement is to set a larger volume of air into motion and so increase the amount of energy radiated. Of course this extra radiated energy is furnished by the driving piston; the effect of the horn is to increase the back pressure on the piston and so make it do more work.

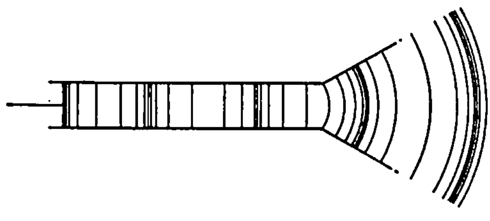


FIG. 9-12. Use of a Horn to Increase Radiation from the End of a Sound Tube.

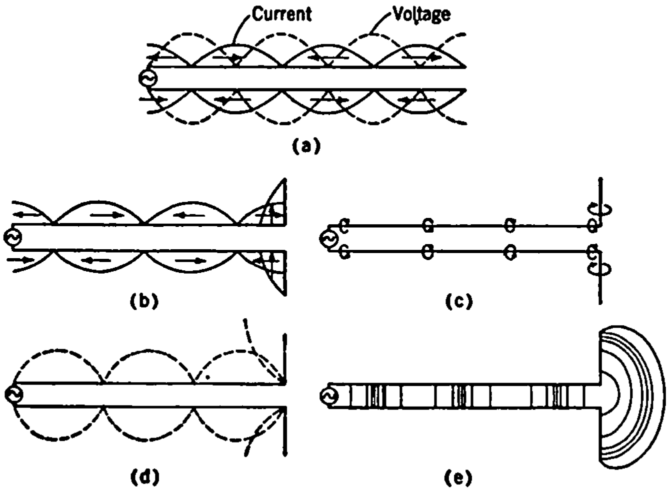


FIG. 9-13. End of a Transmission Line Opened Out to Increase Radiation: (a) Line before opening out; (b) opened-out line showing current distribution; (c) corresponding magnetic field; (d) voltage distribution; (e) corresponding electric field.

In this respect the horn is like a transformer or impedance-matching device because it matches the low "impedance" of the air to the relatively high impedance of the driving mechanism.

The end of the transmission line can be "opened out" in a somewhat similar manner, as shown in Fig. 9-13, when the ends of the transmission line have been turned back to form a *radiator* or *antenna*. The current distribution and a cross section of the corresponding magnetic field are shown in Fig. 9–13b and Fig. 9–13c; the voltage distribution and cross section of the corresponding electric field are shown in Fig. 9–13d and Fig. 9–13e. As in the case of the horn, the opening out of the line into an *antenna* increases the radiated energy and this energy is taken from the generator. Here again, the antenna acts as an impedance-matching device to couple or match the generator to "space."

Dimensions of an Antenna. For efficient radiation an antenna should have dimensions at least of the *order* of a quarter of a wave length. The reason for this can best be seen in Fig. 9–13b. Upon examining the currents in the turned-back or radiating portions of the line it will be found that they are in the *same* direction; likewise, the corresponding magnetic and electric fields in space will be in the same direction and will reinforce each other to produce relatively strong fields. This reinforcement in turn means a large amount of energy radiated. The longer the antenna, up to one wave length, the more of this current there will be in the radiating portion and the stronger will be the fields. When each of the turned-back portions is longer than one half wave length so that the total antenna length is greater than a wave length, there will be included on the antenna portions of other current loops in which current is flowing in the reverse direction (Fig. 9–14). These reverse currents change the directions in which maximum energy is radiated so that the antenna becomes a *directional* antenna (see Chapter 16). As will be seen later, there are certain advantages to making an antenna just one half wave length long, and this is a very commonly used length.

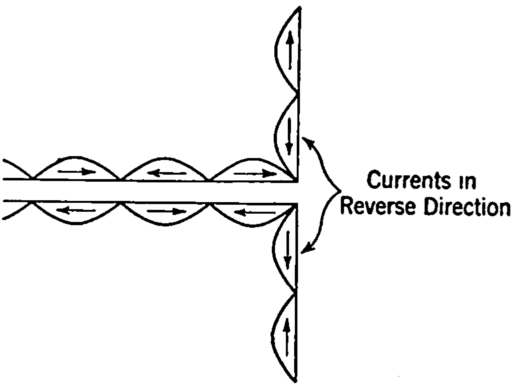


FIG. 9–14. Antenna Longer than One Wave Length, Showing Canceling Current Loops.

Radiation Resistance. When a transmission line is left open or is shorted at the end, the electromagnetic wave is reflected and sent back up the line so that there is no net flow of energy along the line (except a small amount to supply resistance losses). However, as the end of the line is opened out into a radiator some of the energy is radiated into space instead of being reflected back, and energy is taken from the generator. As far as the generator is concerned, then, the antenna is like a resistor absorbing power at the end of the line. The particular value of resistance which would absorb the same amount of power as the antenna is called the *radiation resistance* of the antenna. The power absorbed by a resistor is I^2R; the power absorbed, and therefore radiated, by an antenna is I^2R_a, where R_a is the radiation resistance of the antenna and I is the current flowing in at the feed point.

Mechanism of Radiation. So far, nothing has been said as to how the electromagnetic waves manage to leave the wires and travel on in free space where there are no wires and therefore no currents and no charges. The mechanism of radiation is complicated and any simple picture can but suggest the manner in which it occurs. Figure 9–15a indicates the electric-field distribution about an antenna at the instant that the voltage between the two halves of the antenna is a maximum. As the voltage goes to zero the charges upon which the lines of electric force end flow towards the center (the antenna current) and the lines contract or collapse back to zero. However, it takes a certain length of time for them to travel outward and back in again (they move with the speed of light); if, therefore, the voltage alternations on the antenna are very rapid, the voltage may have reached zero and be building up in the opposite direction before some of the outermost lines have collapsed. This condition is illustrated in Fig. 9–15b for a particular line. This line then becomes detached and is pushed out into space by the new set of lines expanding outward with the increasing voltage on the second half cycle. This process continues and results in loops of electric force or electric strain moving out into space with the velocity of light. Along with these lines of electric force, and at right angles to them, is a magnetic field or lines of magnetic force also moving outward with the speed of light. These lines of magnetic force form circles

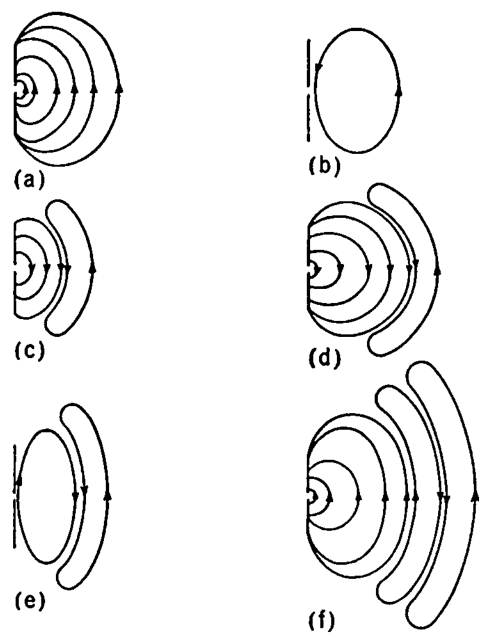


FIG. 9–15. Electric-field Distribution about an Antenna at Successive Instants over a Complete Electrical Cycle.

about the antenna, the circles increasing in diameter as the field moves outward.

When electromagnetic waves are guided along wires, there exists both a moving electric field corresponding to the moving charges (electrons) and a moving magnetic field which is generally considered as being produced by the current or moving charges. In turn, the moving or collapsing magnetic field generates a voltage which corresponds to the electric field. It was Clerk Maxwell who pointed out that the intermediate step of charges and currents is not necessary and that a *changing electric field is equivalent to a current*, which he called displacement current, and so can produce a magnetic field directly. In a similar way a moving or changing magnetic field generates an electric field.

In this manner the propagation of electromagnetic fields in regions where there are no conductors is explained.

Direction of the Electric and Magnetic Fields. From the above it will be evident that the direction of the electric field at any point remote from an antenna is at right angles to a line from the antenna to the point and lies in the plane through the antenna and the point. This arrangement is shown in Fig. 9–16, where the plane through the antenna and the point is the plane of the paper. The direction of the magnetic field is perpendicular to this plane and therefore perpendicular to the electric field. The strengths of both the electric and magnetic fields decrease with distance, being inversely proportional to r, the distance from the antenna. Also the field is stronger at points along lines perpendicular to the antenna (such as along the line OA) than in other directions making a smaller angle with the antenna (such as along OP). This relation is expressed by saying that the field strength is proportional to sin θ where θ is the angle between the antenna and the direction of radiation. When θ is zero, that is, when the point considered is directly above the antenna, the field strength is zero. The above relation is true for very short antennas, considerably less than a half wave long. Longer antennas become *directional* and may radiate much more at certain angles than at others.

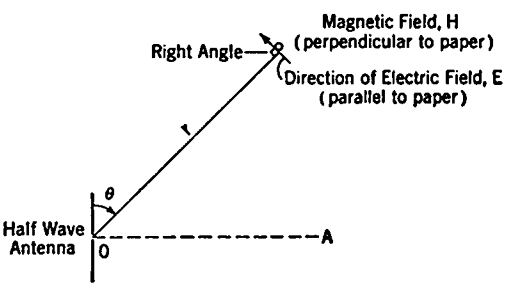


FIG. 9–16. Direction of the Electric Field about an Antenna.

The Receiving Antenna. If the generator or transmitter at the end of the transmission line in Fig. 9–13 were replaced with a receiving set, the antenna would be a *receiving* antenna (Fig. 9–17). As the electromagnetic waves radiated from a distant transmitting antenna speed by, the electric field induces a voltage in the receiving antenna. This voltage alternates in direction at the same frequency as the voltage in the transmitting antenna. It is also possible to consider that it is the lines of magnetic force moving past the antenna which generate a voltage in it. Either point of view is acceptable and both lead to the same answer. However, in calculating the voltage actually induced in the antenna, the voltage should be considered as the result of the electric field *or* the magnetic field, but not both, since these are but two different ways of viewing the same phenomenon.

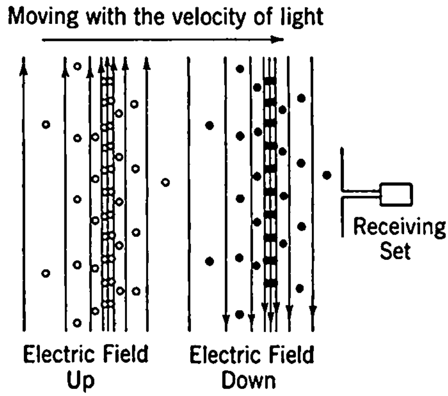


FIG. 9–17. Reception of an Electromagnetic Wave by a Receiving Antenna.

Review Questions and Problems. *1.* In connection with wave motion, define the following: frequency, wave length, velocity of the wave, phase difference, nodes and loops, transverse and longitudinal waves.

2. Calculate the wave length (a) of a 1,000-cycle sound wave; (b) of a 1,000-cycle radio wave. *Answer:* (a) 34.4 cm, (b) 300 km.

3. What are the differences between a standing wave and a traveling wave? Explain fully.

4. Show how the voltage (as measured by a voltmeter) varies along a line having zero resistance carrying (a) a traveling wave; (b) a standing wave.

5. Find the resultant of two waves, each of amplitude A, which differ in phase by 60°, by (a) plotting and adding instantaneous values, (b) vector addition. Repeat for phase differences of 90°, 120°, 180°, 270°.

6. Figure 9–5 shows a wave reflected at a boundary *without change of phase*. This could be a voltage wave reflected from the open end of a line or a current wave reflected from the short-circuited end of a line. Redraw Fig. 9–5 to show a wave reflected with 180° phase reversal. This would represent a voltage wave reflected from a shorted end or a current wave reflected from the open end of a line.

7. What is meant by the radiation resistance of an antenna?

8. Would one expect to receive much signal on a horizontal antenna when the transmitting antenna is vertical? Explain in terms of the directions of the electric and magnetic fields about a vertical transmitting antenna.

CHAPTER 10

Transmission of Signals by Radio

Radio Communication Systems. A communication system is established for the purpose of rapidly transmitting information from one point to another. The two principal means for "instantaneous" communication are by wire and radio. Radio is the only practical means for communicating with such moving conveyances as boats, airplanes, and motor vehicles.

When a body of water, such as the Atlantic Ocean, is located between stationary points of communication, radio often proves to be the best system. Radio is also the most logical system to use where the transmissions are to be received at a great many points, as in broadcasting to the general public. The reason for this is that the *medium* of transmission is present everywhere and such a physical connection as a wire circuit between the two points of communication is unnecessary.

Radiotelegraphy. The information to be transmitted may take the form of a message sent in dots and dashes. This method is called *radio telegraphy*. The transmission is instantaneous.

Radiotelephony. The human voice, music, and other sounds may be transmitted by radio. This method is called *radiotelephony*. Transmission by radiotelephony is instantaneous.

Radio Facsimile. It is also possible to transmit the contents of pages containing written matter and pictures. The page of material is placed in the transmitting device and a page of blank paper is placed in the receiving device. By means of electrical impulses transmitted by radio, a reproduction or *facsimile* of the original page is obtained at the receiving point. Transmission of information by this means is called *facsimile*.

Television. The fourth principal means of transmitting information by radio is called *television*. This is the process of transmitting continuous instantaneous visual pictures of events occurring at a distance. Motion pictures may also be transmitted. In addition to viewing the action, the observer also may hear what is going on at the pickup point through transmissions over a radiotelephony channel that accompanies the visual transmissions.

This chapter will be concerned principally with the fundamentals of radiotelegraph and radiotelephone systems.

Audio-frequency Electromagnetic Wave Radiation. It was shown in the previous chapter that electric waves could be started out in all directions by causing current to flow back and forth in a wire called an antenna. If means are provided at a distant point for detecting these electromagnetic waves the system can be used for communication.

The system which would first occur to an experimenter would be to connect the output of a microphone and an audio-frequency amplifier directly to an antenna and radiate waves at the audio frequency corresponding to the original sound and music. However, there are reasons why audio-frequency electromagnetic waves cannot be used directly for communication without wire connections. It was shown in the previous chapter that the antenna should have a length which is an appreciable part of the wave length. The audio-frequency band extends approximately from 30 to 20,000 cycles, dependent on the listeners' ears. The highest audio frequency ordinarily considered in communication work is 10,000 cycles; the wave length corresponding to this frequency is 30,000 m or 187 miles. The lowest usable audio frequency is about 50 cycles; this corresponds to a wave length of 6,000,000 m or 3,720 miles. It has not been practical to build antennas much over 1,000 ft high. From the discussion in Chapter 9 on the dimensional requirements for an efficient antenna, it is seen that any practical height will be a small fraction of the height required for efficient radiation at these frequencies. Therefore the so-called audio frequencies cannot be transmitted with an efficiency high enough to be useful for distant communication.

At audio frequencies, even were it possible to erect an efficient antenna for the transmission of audio-frequency electromagnetic waves, there would still be other limitations. It is seen that a communication system employing audio-frequency electromagnetic radiation directly is impractical.

Modulation. It was shown in Chapter 9 that electromagnetic waves could be radiated efficiently from antennas of moderate physical size if a sufficiently high frequency is used. The question arises—can such a radiation be used for communication?

It is obvious that a high-frequency radiation which is transmitted continuously can contain no information. However, if the transmission is turned on and off by a key to form the dots and dashes of the telegraph code, information can be sent if a proper receiver or detecting system is provided. The operation of the key in turning the signal on and off is called *modulation.*

If the amplitude of the transmitted power can be varied continuously in accordance with the complicated variations characteristic of speech or music, then it is possible to transmit the more complicated information involved in such signals. The impressing of a variation in amplitude, whether it is the on-or-off variation of the telegraph signal or the continuous and complicated variation of sound, is called *amplitude modulation.* The wave upon which the modulation takes place is called the *carrier.*

Common Use of the Transmission Medium. If only one transmitter and one receiver were in existence the problem of radio would be very simple. However, the whole world must use the same medium and thousands of radio transmitters are operating simultaneously. A given receiver must be able to select the particular transmission to which it wishes to respond. This it can do in part by using resonant or selective circuits which make it sensitive to the frequency or frequencies involved in the desired signal and insensitive to waves at other frequencies.

The useful radio spectrum at present is approximately from 10,000 cycles up to 10,000 mc (1,000,000 cycles = 1,000 kilocycles = 1 megacycle, abbreviated 1 mc).

Each radio transmitting station requires a certain segment of the radio spectrum. Assume that a radio signal is being transmitted on 1,000 kc. In order to transmit intelligence, the radio signal must be modulated. The effect of modulation causes the signal to have components on either side of its original frequency, as will be explained later in the chapter. The segment required in the radio spectrum for the transmission of a message on a given carrier frequency is called a *channel.* The center frequency of a channel is the *carrier frequency.*

The transmitters of the world operate on various channels in the radio spectrum. There are not enough channels so that each transmitter may have exclusive use of one. For certain services, a given channel can be shared, within a short distance of the order of 100 miles; for other services there can be only one transmitter on the channel in the entire world. The radio channels of 1,230 kc, 1,240 kc, 1,340 kc, 1,400 kc, 1,450 kc and 1,490 kc have been set aside on the North American continent for the use of low-powered transmitters designed to render broadcast service to a single city and its immediate vicinity. Low-powered stations are placed on the same channel within distances as low as 70 to 100 miles and still each station performs a useful service to its own community. On the other hand, international code transmitters and international broadcast stations, used for the purpose of transmitting from country to country and from continent to continent, require the exclusive use of a channel.

As will be explained in Chapter 15, the propagation characteristics of radio channels at different frequencies vary greatly. Fortunately, the entire radio spectrum is useful for one type of service or another. It would have been unfortunate indeed if it had been found that only a small segment of the presently known radio spectrum had the optimum characteristic for all services.

There is a shortage of radio channels and they must be used very carefully and wisely. There are many uses for radio transmission besides those that are in common use today. However, there are not enough channels to go around for all of these uses. It is therefore necessary to allocate the use of the frequencies in the order of importance of the service.

Allocation of Channels. By international agreement the radio spectrum has been divided up for specific uses. The part of the radio spectrum from 550 to 1,600 kc has been set aside for domestic radio broadcasting in each country. The channels most suited for international transmission have been carefully divided up among the nations of the world so that each transmitter may have an exclusive channel.

Each country in the world regulates the use of radio channels in its own country. In the United States this function is performed by the Federal Communications Commission. The F.C.C. determines who may use the international facilities available to the United States. It also allocates channels for the use of all other civil radio services in the country. The channels are carefully allocated so that each transmitting station can render a useful service.

Besides international agreements on the use of channels and national regulation of their domestic use by each country, it is sometimes necessary for adjacent countries to have additional agreements for the use of certain facilities. An example of this is the North American Regional Broadcasting Agreement. The parties to this agreement are Canada, Mexico, Cuba, Haiti, the Dominican Republic, Newfoundland, and the United States of America. These countries are so close together that if they did not agree on certain divisions of the domestic broadcasting channels and certain uniform methods for using them there would be a great deal of interference between the broadcast stations in the various countries of North America.

As a result of international agreement and proper regulation in each country, thousands of radio transmitters throughout the world can use the comparatively few channels of the radio spectrum in the same medium of transmission.

Radio Communication System. The three principal parts of a communication system are the *transmitting station, the medium,* and *the receiving station.* A block diagram of a communication system is shown in Fig. 10-1. The *transmitting station* is composed of a transmitter that generates and modulates the radio-frequency power and an antenna that produces the electromagnetic radiation.

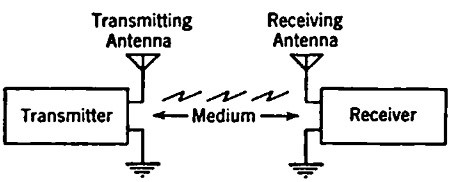


FIG. 10-1. Radio Communication System.

The medium conducts the electromagnetic radiation out into space. A *receiving station* consists essentially of an antenna and a receiver. The receiving antenna is in the electromagnetic field produced by the transmitting station and therefore a current is made to flow through it. The receiver performs the function of converting the current in the receiving antenna into the intelligence that is contained in the transmission.

Radiotelephone Communication System. The transmitter and the receiver of a radiotelephone communication system is shown in Fig. 10–2

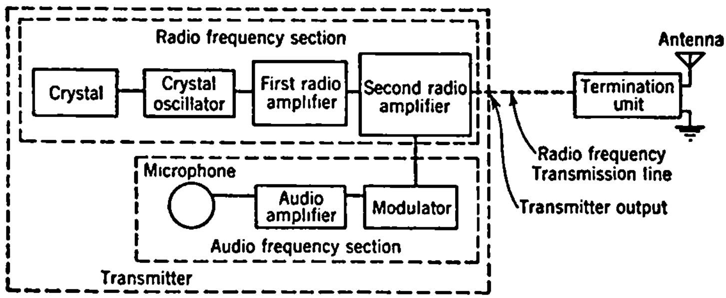


FIG. 10–2. Radiotelephone Transmitting Station.

and Fig. 10–3. The various components of the transmitter and receiver are shown in block form in these diagrams.

Transmitter. A radio transmitter consists of two principal sections. These are the radio-frequency section and the audio-frequency section. The radio-frequency section generates energy at the carrier-wave frequency and amplifies it to the output power of the transmitter. The audio-frequency section is concerned with modulation of the radio signal. The radio-frequency section of the transmitter is shown in a dotted box. This consists of a crystal, crystal oscillator, and two successive stages of radio-frequency amplification, labeled *first radio amplifier* and *second radio amplifier*. This section of the transmitter produces continuous-wave radio-frequency power. Modulation by the audio frequency takes place in the modulated amplifier. In order to obtain the audio-frequency power necessary for proper modulation from a microphone it is necessary to insert audio-frequency amplification between the microphone and the modulated amplifier. Microphones produce a very weak audio-frequency energy and this therefore must be amplified. The energy is further amplified by the *modulator*. Even though it is an audio-frequency amplifier, the modulator is so named because it supplies the necessary power for modulation in the modulated amplifier.

The output of the modulated amplifier is a modulated radio-frequency wave. The modulated radio-frequency energy is used to excite the antenna to produce an electromagnetic radiation. The modulated radio-frequency energy is carried to the site of the transmitting antenna by a *radio-frequency transmission line*. This is often several hundred feet long. The *termination unit* couples the transmission line to the antenna.

The medium conducts the radiation from the transmitting antenna out into space.

Medium. Electromagnetic radiation is conducted out through space at the speed of light. The behavior of electromagnetic waves, or *radio waves* as they are commonly called, is well known. However, little is actually known about the *medium* that carries them from the transmitting antenna to the receiver. The *medium*, for radio-wave transmission, is better known by the behavior of radio waves in space. This behavior is discussed in detail in Chapters 9 and 15.

Receiver. The electromagnetic field causes a radio-frequency current to flow in the receiving antennas. Referring to Fig. 10–3, it is seen that the first important component of a receiver is the *selective circuit*. If no selective circuit were used the receiver would respond to a number of transmissions at once, causing interference with the desired signal. The selective circuit is tuned to allow only the desired signal to pass into the receiver. The radio-frequency signal is *demodulated* or *detected* in the *detector* so as to

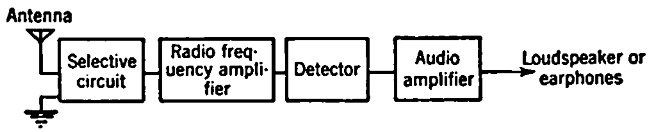


Fig. 10–3. Radio Receiver.

obtain, in the form of audio-frequency currents, the intelligence transmitted by the radio-frequency signal. In order to increase the sensitivity of the receiver, thereby extending the range of the system, a *radio-frequency amplifier* is inserted between the selective circuit and the detector as shown in Fig. 10–3. The radio-frequency amplifier consists of one or more stages of radio-frequency amplification.

In Fig. 10–3 the selective circuit is shown as the first component of the receiver. In practice there is usually a tuned circuit between the antenna and the first vacuum tube and between the successive vacuum tubes up to the detector. These tuning circuits all constitute the selective circuit, shown as a block component in Fig. 10–3. In some cases, a pair of earphones is used on the output of the detector. If it is desired to listen to a loudspeaker rather than earphones, then the weak audio frequency from the detector is amplified to the proper level to operate the loudspeaker. This audio-frequency amplification is shown in Fig. 10–3. In most receivers, andio-frequency amplification is provided even for earphone reception.

There are two principal types of receiver, the radio-frequency receiver (RF receiver) and the superheterodyne. The one shown in Fig. 10–3 is a RF receiver and employs straight detection.

Superheterodyne Receiver. The superheterodyne type of receiver is more sensitive and tunes sharper than does the radio-frequency receiver. The block diagram of a typical simple superheterodyne receiver is shown in Fig. 10–4a. The incoming signal is impressed upon the *first detector.* Assume that the incoming signal is at 1,000 kc. The *local oscillator* is adjusted so that it produces a signal at a frequency such as 1,455 kc. These two signals mix in the first detector, producing a new signal at 455 kc, called the *intermediate frequency.* The intermediate-frequency signal is amplified in an *intermediate-frequency amplifier* and it is then detected in the *second detector* to obtain the audio-frequency currents to actuate the earphones.

In order to make the superheterodyne receiver more sensitive, a tuned radio-frequency amplifier is added ahead of the first detector as shown in

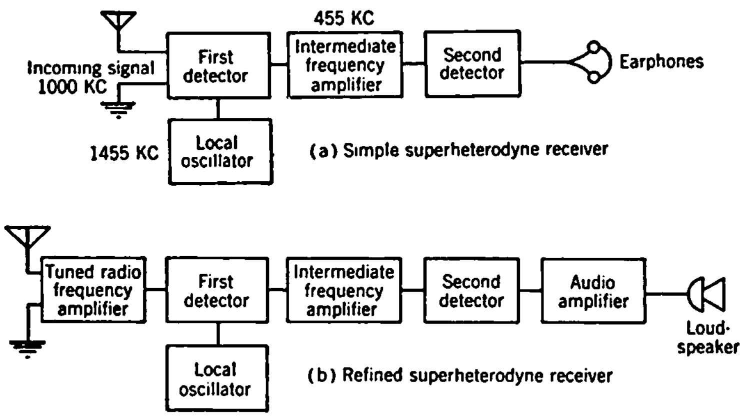


FIG. 10–4. Superheterodyne Receivers.

Fig. 10–4b. Another important reason for adding the radio-frequency amplifier is discussed in Chapter 13. In order to increase the audio-frequency output of the receiver an audio amplifier is added after the second detector.

The superheterodyne type of receiver is used almost to the exclusion of the radio-frequency receiver.

Radiotelegraphy. How a radio-frequency signal can be modulated has already been explained. The carrier may be modulated in such a way as to produce combinations of dots and dashes forming a code. This form of communication is called *radiotelegraphy.* There are two types of radiotelegraph. These are *continuous-wave* (CW) and *interrupted continuous-wave* (ICW).

Continuous-wave (CW) Radiotelegraphy. The carrier wave itself may be interrupted in such a fashion as to form dots and dashes which constitute code transmissions. Figure 10–5 shows graphically what happens to the wave. A telegraph key may be used to start and stop the carrier wave indirectly. Referring to Fig. 10–5, as an example, the telegraph key can be depressed for a fraction of a second, left open for a fraction of a second, and then closed for three times as long as the first time, thus transmitting two impulses or radio-frequency groups that may be interpreted by a receiving system as being a dot and dash.

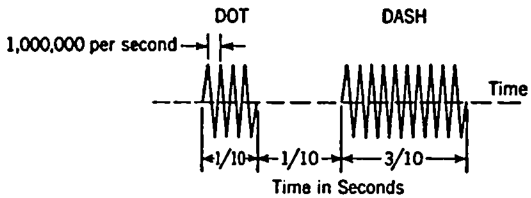


Fig. 10–5. Continuous-wave (CW) Telegraph. (The waves are sine waves. They are shown as straight lines for convenience only.)

Assume that the carrier frequency in use is 1,000 kc. The impulses received will be at 1,000 kc. This frequency is much above the audible range. A continuous-wave (CW) receiver has incorporated in it a radio-frequency oscillator. Adjusting the oscillator to 1,001 kc produces two additional waves in the receiver, one the sum of 1,000 and 1,001 kc and the other the difference of the two waves, that is, 1,001 − 1,000 = 1 kc. The 1-kc or 1,000-cycle component may be used to actuate a pair of earphones or a loudspeaker to produce an audible tone. It is seen that the transmission of the dots and dashes on 1,000 kc produces impulses that will actuate earphones or a loudspeaker to produce audible tones of 1,000 cycles.

Interrupted Continuous-wave (ICW) Radiotelegraphy. Besides CW, the other principal method used for radiotelegraphy is *interrupted continuous-wave* (ICW). In this method the carrier is modulated at a fixed frequency by one of several methods. The desired audio frequency, to actuate the earphones or loudspeaker, is produced in the receiver without the use of a local oscillator such as is necessary with the continuous-wave type.

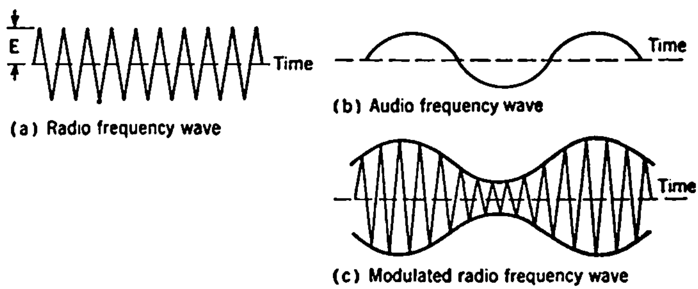


Fig. 10–6. Modulation of a Radio-frequency Wave.

Amplitude-modulated Wave. A radio-frequency wave is shown in Fig. 10–6a. The audio-frequency wave shown in Fig. 10–6b is low in frequency compared to the radio-frequency wave shown in Fig. 10–6a. An audio-frequency wave (such as that shown in Fig. 10–6b) can be combined with a radio-frequency wave (such as that shown in Fig. 10–6a) in such a way that the audio-frequency wave varies the magnitude of the radio-frequency wave exactly in accordance with its own wave form. The result of this operation is an *amplitude-modulated wave* and is shown graphically in Fig. 10–6c. The heavy lines in Fig. 10–6c form what is called the *modulation envelope*, which is a replica of the modulating audio-frequency wave. If a more complicated signal were used, the modulation envelope would still follow it in detail.

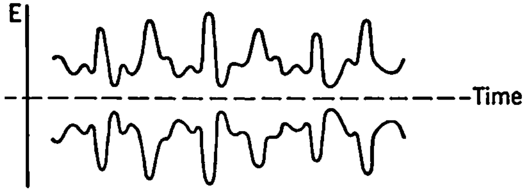


FIG. 10–7. Envelope of Modulated Radio-frequency Signal over Short Period of Time.

In radiotelephony the *envelope* of the carrier wave is very complex as compared with the envelope of the wave when a single frequency modulates the transmitter as shown in Fig. 10–6. Figure 10–7 shows graphically what the *envelope* of a radiotelephone modulated wave might look like over a period of short duration.

Side Frequencies. Analysis of the electrical circuits is always on the basis of simple sine waves. It is apparent that the modulated wave shown in Fig. 10–6 is not such a simple sine wave. It may be shown both by mathematics and experiment that this wave really contains three frequencies. These are

f_0 (called the carrier wave),
$f_0 - f_1$ (called the lower side frequency),
$f_0 + f_1$ (called the upper side frequency),

where f_1 is the signal or modulation frequency. As an example, if f_0 is 1,000 kc and f_1 is 1,000 cycles, then the resultant waves will have frequencies of 999 kc, the lower side frequency; 1,000 kc, the carrier wave, and 1,001 kc, the upper side frequency.

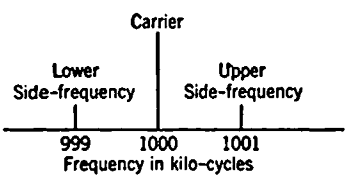


FIG. 10–8. Side Frequencies in Modulation.

Figure 10–8 shows a so-called spectrum analysis of such a wave.

SIDE BANDS. If the modulating signal contains several frequency components, each component will produce an upper and lower side frequency. The group of lower side frequencies is called the *lower side band*, and the group of upper side frequencies the *upper side band*. The width of *each* band is equal to the highest frequency component of the signal to be transmitted.

Therefore the total width of the segment occupied in the spectrum by an amplitude-modulated wave is twice the highest frequency in the signal.

Receiver Selectivity. In order to receive satisfactorily a modulated wave, it is necessary that the receiver be able to accept the band of frequencies that is transmitted. It is also necessary that the receiver reject signals on other frequencies. The dotted lines in Fig. 10–9 show an ideal response characteristic for a receiver. The carrier wave of the received signal is 1,400 kc and the highest modulation frequency is 5,000 cycles. The response is uniform from 1,395 kc to 1,405 kc. There is no response outside this band. Unfortunately, for practical reasons, such response curves cannot be obtained.

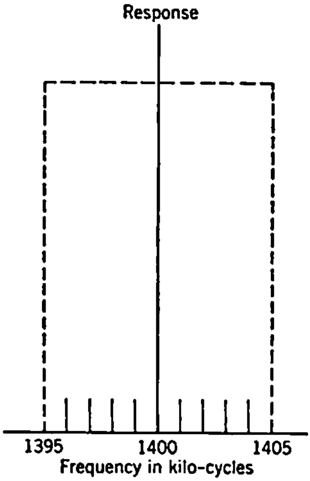


Fig. 10–9. Ideal Selectivity Response Curve of Receiver.

Figure 10–10a shows the response curve of a typical tuned radio-frequency receiver. Figure 10–10b shows the response curve of a typical superheterodyne receiver. Assume that the transmitter carrier is on 1,400 kc and that the highest modulation frequency is 5,000 cycles. The desired frequency response of the receiver would then be as shown in Fig. 10–10c. This is 5 kc on each side of the carrier frequency of 1,400 kc. As may be seen from the response curves, neither the curve of Fig. 10–10a nor that of Fig. 10–10b is the ideal *box*. The superheterodyne response curve, as shown in Fig. 10–10b, however, comes the closest. It will therefore receive the desired signal and still not respond, to any great extent, to signals on other frequencies. On the other hand, the response curve of the radio-frequency receiver, as shown in Fig. 10–10a, shows an appreciable response even at 1550 kc and at 1250 kc. Such a radio-frequency receiver would be subject to a great deal of interference from other stations between 1,250 and 1,550 kc. Relatively, the superheterodyne receiver would only be interfered with up to about 1,420 kc and down to about 1,380 kc.

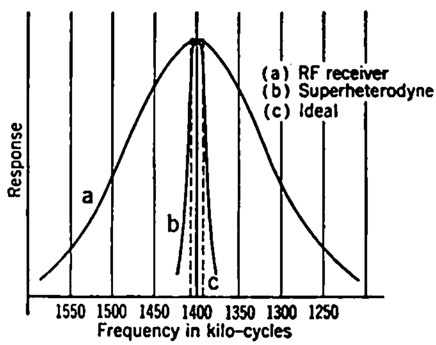


Fig. 10–10. Selectivity Response Curves of Receivers.

Review Questions and Problems. *1.* Under what conditions may radio communication be used to advantage?

2. What are the four principal types of radio communication?

3. Why is it that audio-frequency electromagnetic wave radiation cannot be used for communication?

4. What is modulation?

5. Name the three principal parts of a communication system and briefly tell what each does.

6. (a) Describe an amplitude-modulated radio transmitter. (b) Tell the function of each part.

7. What is known about the medium of transmission for radio communication?

8. What function does a receiver perform?

9. (a) Name the principal parts of a receiver. (b) Describe what each does.

10. Name the two general types of receivers and describe how they function.

11. What is the difference between continuous-wave (CW) and interrupted continuous-wave (ICW) radiotelegraphy?

12. What are side frequencies?

13. What are side bands?

14. Why is selectivity important in receivers?

CHAPTER 11

Radio-frequency Amplifiers and Detectors

Radio-frequency Amplifiers. Radio-frequency amplifiers are used in both transmitters and receivers for the purpose of amplifying a radio-frequency voltage or power.

Fundamentally, vacuum-tube amplifiers operate in the same way regardless of the frequency of the voltage to be amplified. A signal voltage impressed upon the grid of a vacuum tube will control a relatively large output in the plate circuit. The grid of the tube must be properly biased and the tube must work into the proper load impedance. The circuit diagram of a conventional resistance-coupled audio-frequency amplifier is shown in Fig. 11–1a. The bias for vacuum tube T is obtained through the resistor R. The signal input voltage is impressed across R as well as across the grid and filament of the tube. The value of R must therefore be large enough so that it does not unduly load the input circuit and reduce the signal input voltage. Resistor R_1 is the proper output load impedance for the vacuum tube T to work into. Assume that R and R_1 are pure resistances. Over a wide frequency range the vacuum tube T functions as an amplifier.

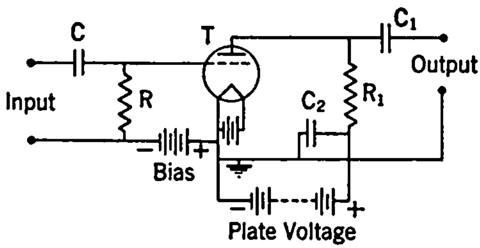
An amplifier such as that shown in Fig. 11–1a can function as a Class A radio-frequency amplifier. It is obvious that such a circuit does not respond to the desired signal to the exclusion of others. A receiver using such an amplifier will be unsatisfactory because it does not reject undesired signals. Because of shunting capacities, moreover, the gain would also be limited.

In Chapter 3 it was shown that a parallel circuit, consisting of an inductance and a capacity, has a relatively high impedance at the antiresonant frequency. It was also shown that the parallel-circuit impedance rapidly decreased on either side of the antiresonant frequency. A parallel circuit may be substituted for the load impedance R_1 of Fig. 11–1a. Such a substitution is shown in the diagram of Fig. 11–1b. The air-core inductance L and the capacity C_3 form a parallel antiresonant circuit. By proper selection of L and C_3, the impedance across points a and b may be

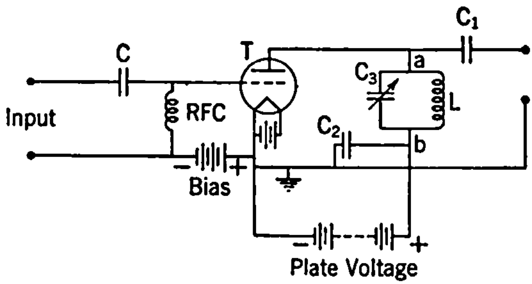
* In the radio trade, *radio-frequency* amplifiers, chokes, and so forth are commonly called simply *radio* amplifiers, chokes, and so forth. When there is no likelihood of confusion, this practice will be followed in this book. So also for *audio-frequency* and *audio*.

made large for any given radio frequency. The amplifier will therefore function properly at the resonant frequency. Because the impedance of the parallel circuit rapidly decreases on each side of resonance, the amplifier output will be reduced at other frequencies. Selectivity has therefore been obtained by substituting a parallel resonant circuit for the load resistance R_1 of Fig. 11–1a.

The use of a parallel resonant circuit as a load impedance for a vacuum tube is important in transmitters. The resistor R_1 of Fig. 11–1a wastes plate-supply power. In itself this is unimportant in radio-frequency amplifiers used in receivers, because of the low power. The relatively high powers involved in radio-frequency amplifiers used in transmitters require that efficiency be carefully considered. The parallel-resonant-circuit load impedance avoids the loss of power. Another important consideration is that the plate voltage must be raised to compensate for the drop in the resistor R_1. In Fig. 11–1b the plate current flows through the inductance L. The resistance of the inductance to direct current is negligibly small compared with that of R_1 in Fig. 11–1a.



(a) Resistance coupled amplifier



(b) Resonant circuit coupled amplifier

Fig. 11–1. Resistance- and Resonant-circuit-coupled Amplifiers.

The selectivity of the parallel resonant circuit is somewhat important in transmitters but not to the extent that it is in receivers.

The inductance or radio-frequency choke marked RFC in the circuit diagram of Fig. 11–1b is an air-core inductance presenting a high impedance at radio frequencies. Therefore it will not short the radio-frequency input signal. If a resistor, such as R in Fig. 11–1a, is used to supply bias to a vacuum tube used as a radio-frequency amplifier in a transmitter, the power loss may become appreciable. Radio-frequency chokes, such as RFC in Fig. 11–1b, are therefore extensively used in transmitters for feeding bias voltage to a radio-frequency amplifier vacuum tube.

The principal difference between a Class A resistance-coupled audio amplifier circuit and a radio amplifier circuit is that parallel resonant circuits are used in the latter for coupling. In addition, bias voltages are usually fed through radio-frequency chokes. However, resistances are sometimes used in low-powered radio amplifiers.

Resonant-circuit Coupling for Radio Amplifiers. In audio-frequency amplifiers, the coupling devices are usually transformers, resistances, or impedances. Parallel-resonant-circuit coupling is usually used in radio amplifiers.

In receivers, the resonant coupling circuits act to accept the desired signal and reject the undesired signals. A resonant coupling circuit is, in effect, an impedance-matching circuit. It is easy to adjust the impedance of the circuit to any desired value. This adjustment is particularly important in transmitters, where large amounts of power are used and where efficiency is therefore important. The impedance of a resonant coupling circuit can readily be adjusted for the optimum load impedance of the vacuum-tube amplifier, thereby attaining maximum efficiency.

Figure 11–2a shows a parallel resonant circuit consisting of a capacity C in parallel with inductance L which is in series with resistance R.

Antiresonant frequency. The impedance Z looking into terminals X–X is a pure resistance at antiresonance. The circuit shown in Fig. 11–2a has the approximate constants for resonance at 1,000 kc. At 1,000 kc, when R is zero, Z is infinite; when R is 10 ohms, Z is 10,000 ohms; and when R is 100 ohms, Z is 1,060 ohms. R may be selected so that the impedance Z will be whatever value is the proper load impedance for the amplifier. This selection is impedance matching. As an example, the 10-ohm load has been matched to 10,000 ohms. The use of parallel resonant circuits for impedance matching is more fully discussed in Chapter 3.

A parallel resonant circuit may be adjusted so as to present the proper load impedance to a vacuum tube operating at a specific carrier frequency, such as 1,000 kc. In radiotelephony, however, the amplifier must also properly amplify the side bands.

Referring to Fig. 11–2a, assume that the circuit is resonant under three conditions, namely R equals 0, 10 ohms, and 100 ohms. It is also assumed that there is no appreciable resistance in the condenser C, which is true in a well-constructed condenser up to the ultrahigh frequencies; and that L is a pure inductance and has no resistance, which is not true in practice because of the resistance of the wire. The impedance Z over a band of frequencies under the three conditions is shown in Fig. 11–2b.

A tuned circuit such as is shown in Fig. 11–2a appears as a pure resistance at its terminals only at the antiresonant frequency. At other frequencies the impedance Z is partly reactive.

Curve A shows that when R equals zero the impedance Z rises to infinity at the resonant frequency of 1,000 kc. Curve B shows Z when the resistance is 10 ohms. If the highest modulation audio frequency is 10,000 cycles, the band of frequencies to be transmitted is from 990 to 1,010 kc. Curve B shows that the impedance varies somewhat over this range. A radio amplifier can work satisfactorily with this difference if simple precautions are taken. Curve C of Fig. 11–2b is a plot of Z when R is equal to 100 ohms. Over the band of frequencies from 990 to 1,010 kc the impedance is substantially equal at all frequencies. However, it is to be noted that the impedance Z is much lower than that usually required for the plate load of a vacuum tube. If an attempt is made to make the impedance characteristic too flat, the impedance at antiresonance is then too low to operate as a coupling device for a vacuum tube.

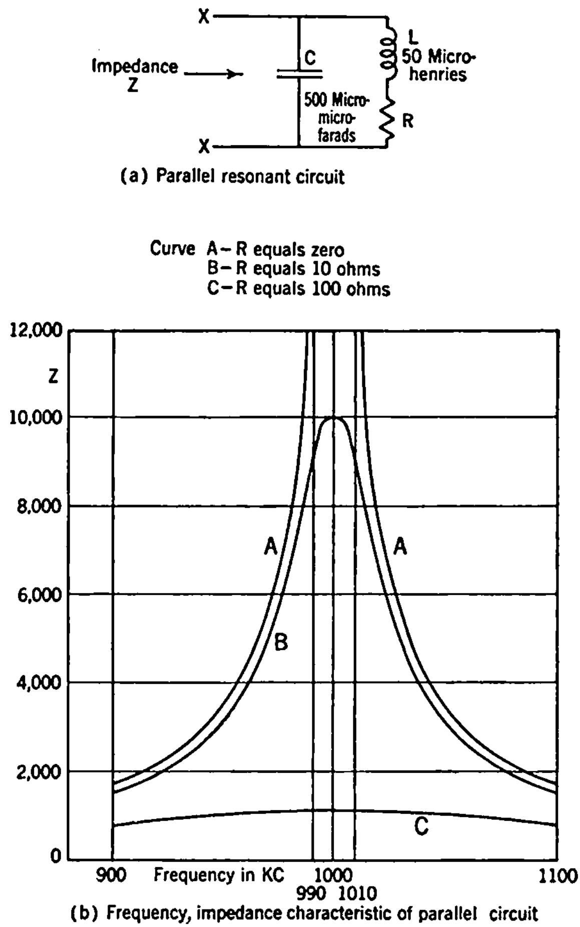


Fig. 11-2. Impedance of Parallel Circuit.

Harmonic suppression. Vacuum tubes generate harmonics that produce undesirable radiations when coupled to an antenna if precautions are not taken to suppress them. The second harmonic of 1,000 kc is 2,000 kc, the third harmonic is 3,000 kc, and so forth. If the harmonics were allowed to radiate, interference with other stations on the harmonic frequencies may result. The curves of Fig. 11–2b extend only up to 1,100 kc. However, it can be seen that if they were extended to 2,000 kc the value would be very low. The harmonics are reduced because the amplifier would be working into a very low impedance at the harmonics.

Resonant-circuit Coupling for Audio Amplifiers. Faithful reproduction through an electrical system is dependent on the ratio, usually expressed in per cent, of the band width to the lowest band frequency and not on the band width itself.

If a 1,000-kc carrier is modulated by 10,000 cycles, then the band width is 20,000 cycles. The band extends from 990 to 1010 kc. The band width,

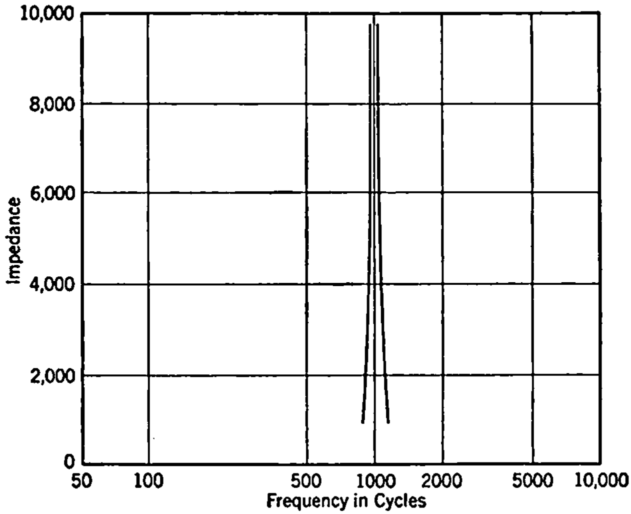


FIG. 11–3. Parallel Resonant Circuit at Audio Frequencies. Electrical Constants Equivalent to (B) of Fig. 11–2.

20 kc, is 2.02% of the lowest frequency, 990 kc, of the band. This ratio corresponds to an audio-frequency band of 990 to 1,010 cycles. The electrical equivalent of Fig. 11–2a could be set up for 1,000 cycles. In this case C and L would have much higher values than for 1,000 kc. It would be expected that the electrical behavior would be the same over the same *fraction* (per cent) of band width. The impedance Z of such a circuit is plotted in Fig. 11–3. It is seen from Fig. 11–3 that the resonant circuit presents a high impedance only in the neighborhood of 1,000 cycles and that over the rest of the audio band, from 50 to 10,000 cycles, the impedance is very low. This impedance curve causes a great difference in amplification at frequencies other than 1,000 cycles. Resonant circuits therefore cannot be used as coupling devices in audio amplifiers, whereas they may be used and are desirable in equipment designed to handle an equal band of

frequencies after it has been transferred to a higher place in the radio-frequency spectrum.

Radio-frequency Amplifier Circuit. The fundamentals of all classes of radio amplifiers are the same. The schematic circuit diagram of a simple radio-frequency amplifier is shown in Fig. 11–4. The vacuum tube T works into the resonant circuit composed of condenser C and inductance L. In some circuit applications, the wire forming the inductance L has enough resistance so that the resonant-circuit impedance is the right value for proper operation of the vacuum tube. This often is the case where voltage amplification is required, as in radio receivers.

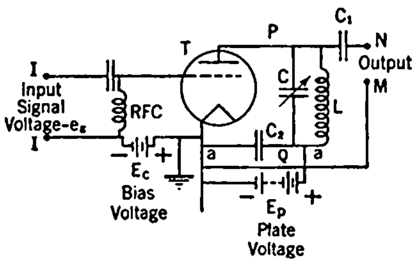


FIG. 11–4. Simple Triode Radio Amplifier.

The signal input voltage e_s to be amplified is impressed on the grid through terminals I–I. The bias voltage E_c reaches the grid through the radio-frequency choke RFC. RFC prevents the radio-frequency input signal voltage from being shorted through the bias supply. The radio-frequency by-pass condenser C_2 is of such a value that points *a-a* are at the same radio-frequency potential without shorting the plate voltage. The resonant output-coupling circuit is composed of inductance L and condenser C. So far as radio frequency is concerned, the coupling circuit is connected directly from the filament to the plate of the tube. Blocking condenser C_1 is of such a capacity that output terminal N is effectively connected to one end of the resonant circuit, so far as radio frequency is concerned. C_1 prevents the plate voltage from appearing across the output terminals. Output terminal M is effectively connected to the other end of the resonant circuit through capacity C_2. Whatever radio-frequency voltage develops across PQ also appears across the output terminals NM.

There are actually many different ways for coupling the signal input voltage into an amplifier and for coupling the output of the amplifier to the load. There are also many methods for connecting the bias voltage to the grid, and there are many methods for supplying plate voltage to the tube. The resonant output circuit or *tank circuit,* as this is often called, takes many different forms.

Radio Amplifier Input Circuits. A number of different input circuits may be used for impressing the input signal voltage on the grid of a radio amplifier. Four different methods are shown in Fig. 11–5. L and C in Fig. 11–5a represent the output circuit of a radio amplifier. The radio-frequency input voltage is impressed upon the grid of the tube through the radio-frequency blocking condenser C_1. Radio-frequency choke RFC prevents the bias supply from shorting the radio-frequency voltage between the grid and filament. The blocking condenser C_1 prevents the plate voltage of the preceding amplifier from being impressed on the grid.

C and L_1 represent the input tuning circuit to the radio amplifier in Fig. 11–5b. The resistor R is the cathode bias resistor. Condenser C_1 is a radio-frequency by-pass around the cathode resistor, therefore effectively placing the input voltage directly across the grid and cathode. In Fig. 11–5c the condenser C and the inductance L represent the output circuit of a preceding amplifier. L induces a radio-frequency voltage into L_1. Condenser C_1 by-passes the radio-frequency voltage directly to the cathode, thereby impressing the voltage from L_1 across the grid and cathode. The circuit shown in Fig. 11–5d may be used where it is desirable to use a separate bias supply and still be able to ground one end of the tuning condenser. As may be seen from the diagram, condenser C is grounded at one end. The condenser C_1 is large enough so that it is a short circuit for radio frequency. It prevents the bias voltage from being shorted. The bias voltage is fed to the grid of the tube through inductance L_1. Radio amplifier input circuits may take many different forms. The four shown in Fig. 11–5 are typical.

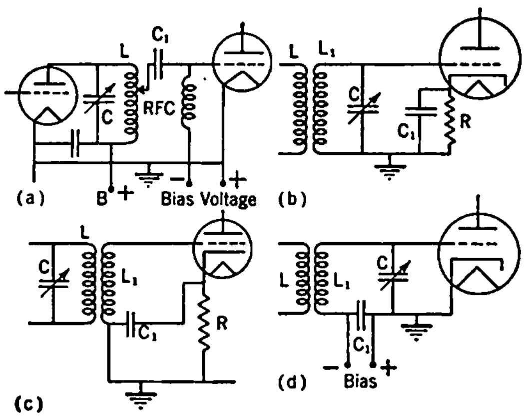


FIG. 11–5. Types of Radio-amplifier Input Circuits.

Single-ended Radio Amplifier Output Circuits. There are many different types of output circuits that may be used with radio amplifiers. The general fundamentals of most of these are illustrated in Fig. 11–6. In Fig. 11–6a, L and C comprise the tuned output circuit. The plate supply is fed to the plate of the tube through inductance L. C_2 is a radio-frequency by-pass condenser effectively connecting one end of the output circuit to the filament. Condenser C_1 allows the passage of the output radio frequency but stops the plate voltage from being impressed upon the output terminals. The plate supply is said to be *series fed* because the plate voltage is fed in series through the tuning inductance L to the plate.

Figure 11–6b illustrates *parallel feed* of the plate voltage. The plate voltage is fed to the plate through the radio-frequency choke RFC. As may be seen from the diagram, the plate voltage and the output circuit are, in effect, in parallel. The radio-frequency choke RFC prevents the radio energy from passing into the plate supply and the condenser C_1 prevents the D.C. plate voltage from passing into the tuning circuit. Inductances L and L_2 form a radio-frequency transformer. The output of the amplifier is inductively fed from L to L_2.

A somewhat different type of output circuit is shown in Fig. 11–6c. In this circuit tuning is accomplished by the variable capacity C. Inductances L_1 and L_2 form a radio-frequency transformer. Therefore, the tuning circuit is reflected into the inductance L_1 and is effectively in series with the plate of the tube.

An output circuit used extensively in transmitters is shown in Fig. 11–6d.

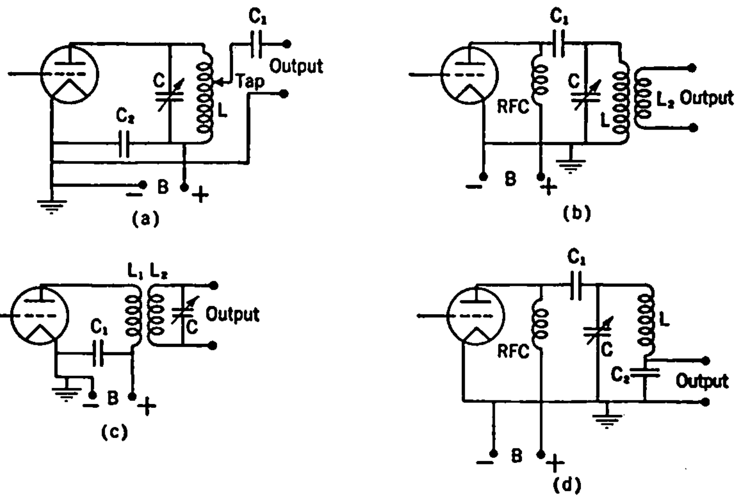


Fig. 11–6. Types of Single-ended Radio-amplifier Output Circuits.

Inductance L and condensers C and C_2 form the output tuning circuit. In this circuit the output of the amplifier is obtained from across condenser C_2. This type of circuit helps to reduce harmonics.

Output circuits such as those shown in Figs. 11–6a, b, d are used extensively in transmitters. The output circuit shown in Fig. 11–6c is extensively used in receivers. Variations of Fig. 11–6a are also used in receivers.

Grid-bias Voltage Supplies. Grid-bias voltage may be supplied to an amplifier by any one of a number of different methods. Several of these are shown in Fig. 11–7. In each of the five figures shown, L and L_1 represent the input and output circuits respectively of the amplifier.

Figure 11–7a shows a battery in use as a grid-bias supply. This type of bias supply is extensively used in portable battery-operated receivers. It is seldom used in amplifiers deriving their power from a 60-cycle A.C. source because the bias voltage can readily be supplied by other means. So-called *cathode biasing* is shown in Fig. 11–7b. The resistor *R* is in series with the cathode of the tube. The plate current drawn by the tube returns to the cathode through the resistor *R* and therefore there is a D.C. voltage drop across the resistance. The polarity of this voltage is positive at the cathode end and negative at the other end. By connecting the grid return lead as shown in Fig. 11–7b, the grid is biased negative in relation to the cathode. The voltage across resistor *R* is equal to the plate current multiplied by the resistance. By using the proper value of resistance, the proper bias voltage is obtained for the amplifier.

Figure 11–7c shows cathode biasing except that the grid return is made

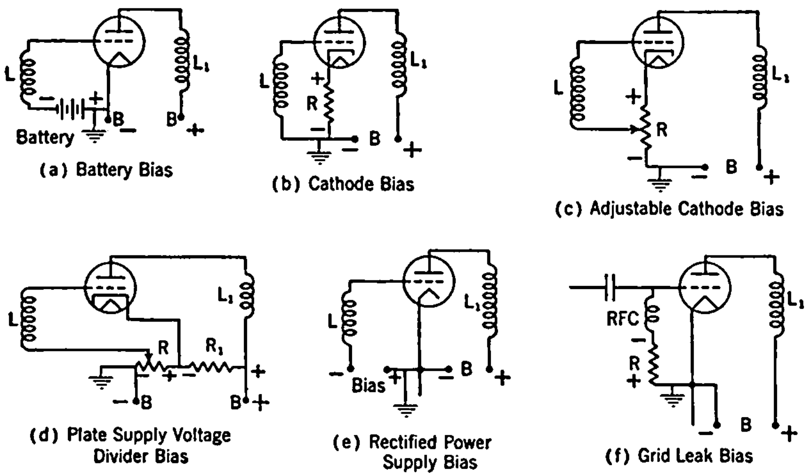


Fig. 11-7. Grid-bias Voltage Supplies.

to the arm of a potentiometer so that the bias voltage to the tube may be varied at will. This type of biasing is not generally used but is convenient in experimental and test equipment.

Figure 11–7d shows biasing by means of a voltage divider across the plate supply. The voltage divider is composed of resistance R_1 in series with a potentiometer *R*. The flow of current through the voltage divider is such that negative bias is supplied to the grid of the tube.

Biasing such as is shown in Fig. 11–7b is extensively used in receiver radio amplifiers and in low-powered audio amplifiers. It is also used in transmitters to some extent. The circuit of Fig. 11–7c, as explained, is a variation of that in Fig. 11–7b and has a limited use in experimental and test equipment. The use of biasing as shown in Fig. 11–7d is principally limited to low-powered amplifiers such as those in receivers. However, it is not in such general use as is the biasing shown in Fig. 11–7b.

Figure 11–7e shows a method of biasing that is extensively used in transmitters. In this method a separate rectifier is used to supply the bias voltage.

Grid-leak biasing is shown in Fig. 11–7f. This type of biasing is used extensively in transmitter radio amplifiers, particularly in code transmitters. When the radio-frequency grid input voltage exceeds the bias voltage, an amplifier draws grid current. As shown in Fig. 11–7f, the grid current returns to the filament through the radio-frequency choke RFC and the resistor *R*. The flow of current through *R* produces a voltage drop across *R* and this drop biases the grid of the tube. Grid-leak bias has the disadvantage that if the signal input voltage to the grid is removed, the bias on the tube is removed because there is no flow of current through *R*. Without bias the tube will draw an abnormally high plate current and may be damaged. To avoid this possibility, sometimes a fixed bias supply is inserted between the resistor *R* and ground. The bias supply contributes part of the bias voltage and the voltage drop across *R* contributes the rest. The fixed bias is high enough to limit the plate current to a safe value in the event that the signal input voltage to the grid of the tube is removed.

Voltage and Power Amplification. The terms *voltage amplification and power amplification* have the same meaning in radio amplifiers that they have in audio amplifiers, as explained in Chapter 7.

Radio Amplifier Classifications. Radio-frequency vacuum-tube amplifiers divide into three classes dependent on the method of operation. These are Class A, Class B, and Class C.

Class A amplifier.* *A Class A amplifier is an amplifier in which the grid bias and alternating grid voltages are such that plate current in a specific tube flows at all times.*

Class A radio amplifiers are usually used for voltage amplification. Their efficiency and power output are low.

Class B amplifier.* *A Class B amplifier is an amplifier in which the grid bias is approximately equal to the cutoff value, so that the plate current is approximately zero when no exciting grid voltage is applied, and so that plate current in a specific tube flows for approximately one half of each cycle when an alternating grid voltage is applied.*

Class B radio amplifiers are usually used as power amplifiers to amplify modulated radio frequency. Class B amplifiers operate at a much higher efficiency than do Class A.

Class C amplifier.* *A Class C amplifier is an amplifier in which the grid bias is appreciably greater than the cutoff value, so that the plate current in each tube is zero when no alternating grid voltage is applied, and so that plate current in a specific tube flows for appreciably less than one half of each cycle when an alternating grid voltage is applied.*

* The definition in italics is from the Standards on Electronics by the Institute of Radio Engineers.

Class C radio amplifiers are usually used for power amplification of unmodulated radio frequency. They will not amplify modulated radio signals. A Class C radio amplifier may, however, be modulated in its plate output circuit. Class C radio amplifiers operate at the highest efficiency.

Radio amplifiers operate much the same as do audio amplifiers. One principal difference is that resonant circuits may be, and usually are, used for input and output coupling circuits.

Class B radio amplifiers can function either single ended or push-pull, while Class B audio amplifiers must be push-pull.

Class C amplification may be used for some radio amplification applications but not for audio amplification. Class C produces the highest plate efficiency and the highest power output. This efficiency is important where large amounts of radio-frequency power are involved, as in radio transmitters.

Class B audio amplifier. Class B audio-frequency vacuum-tube amplifiers operate in a manner similar to Class B radio amplifiers. They are frequently associated with radio-frequency equipment as *modulators*. For these two reasons they are discussed later in this chapter rather than in Chapter 7.

Class A Radio-frequency Amplifier. The principal characteristics of Class A radio amplifiers are high voltage amplification, faithful reproduction of the input wave, low plate efficiency and low power output. Triode, screen-grid, and pentode tubes may be used. Class A radio amplification is usually used where high voltage amplification and faithful reproduction of the wave is required. Class A radio amplifiers are used extensively in receivers where high voltage gain per stage is important. The plate-supply power used by radio amplifiers in receivers is small and therefore power efficiency is unimportant.

In transmitters the plate power is large and therefore output efficiency becomes important. Efficient amplifiers of the Class B and Class C type have been developed to the point where they have almost entirely eliminated the use of Class A radio amplifiers in transmitters.

The radio amplifier shown in Fig. 11–4 could operate as a Class A radio amplifier if the input and output impedances were of the right value and if the bias voltage, the plate voltage, and the signal input to the grid were proper.

Figure 11–8 shows the relation between grid voltage and plate current for a properly operated Class A radio-frequency amplifier. Grid voltage is plotted along the abscissa and plate current along the ordinate. Curve A of Fig. 11–8 is the characteristic curve of a vacuum tube with a load and is a plot of the plate current I_p as the grid voltage E_c is varied. A Class A radio amplifier must operate on the straight portion of the characteristic curve between points *c* and *d*, as shown in Fig. 11–8, for faithful reproduction. Because c to d is a straight portion of the characteristic curve, it can be seen that the plate-circuit signal output current of the vacuum tube will be a faithful reproduction of whatever signal input voltage is impressed on the grid circuit of the vacuum tube.

The bias voltage E_c is of such value that the operating point o is half way between c and d. With no signal input to the grid, the plate current will be o-n. Class A amplifiers usually are used to amplify modulated

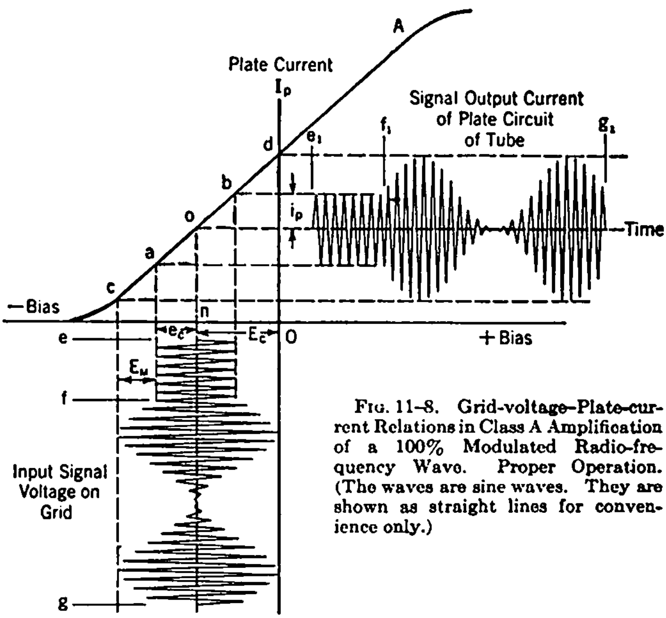


FIG. 11–8. Grid-voltage–Plate-current Relations in Class A Amplification of a 100% Modulated Radio-frequency Wave. Proper Operation. (The waves are sine waves. They are shown as straight lines for convenience only.)

waves. In Fig. 11–8 an unmodulated radio-frequency input voltage is shown along the time axis between e to f and has an amplitude of e_c. This voltage causes a radio-frequency current i_p to flow in the plate circuit during a corresponding time interval of e_1 to f_1. The time interval f to g indicates 100% modulation of the carrier wave by an audio frequency. A modulated wave impressed on the grid circuit causes a current to flow in the plate circuit of the tube during the corresponding time interval f_1 to g_1. The modulated radio signal voltage input to the grid is faithfully reproduced in the output of the tube because the operation has taken place over the linear portion of the characteristic curve between c and d.

It was shown in Chapter 7 that the efficiency of Class A audio amplifiers is low. For the same reasons Class A radio amplifiers operate at a low efficiency. If a Class A amplifier were used to amplify an unmodulated carrier, then the entire portion of the characteristic curve from c to d could be used for the carrier. Both e_c and i_p would be twice the value shown in Fig. 11–8. The efficiency would be approximately the same as for Class A audio amplification. In radiotelephony the peak of the modulation voltage E_m must not exceed the portion of the characteristic curve between c and d. In order to provide for 100% modulation it is necessary to reduce the carrier wave to operation over the portion of the characteristic curve between a and b. This reduction requires that the radio-frequency signal input to the grid be half of what it would be if an unmodulated wave were to be amplified; conversely the output current i_p is half. This condition means that the carrier output power is low. The plate current o-n is the same regardless of the input signal voltage, and therefore the efficiency of the amplifier is low.

Improper operation. Improper operation of a Class A radio ampiifier results if operation on the characteristic curve takes place below point c or above point d in Fig. 11–8.

If operation extends below point c, part of the operation takes place on

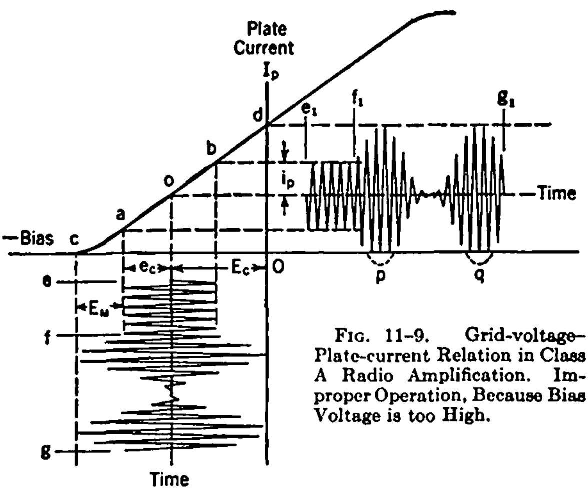


Fig. 11–9. Grid-voltage–Plate-current Relation in Class A Radio Amplification. Improper Operation, Because Bias Voltage is too High.

a curved portion of the characteristic curve and therefore faithful reproduction is not obtained. Figure 11–9 illustrates this type of improper operation. The bias voltage E_c has been increased over that shown in Fig. 11–8 so that the operating point o is lower on the characteristic curve.

In Fig. 11–9, 100% modulated radio-frequency input voltage to the grid is shown during a time interval f to g. The characteristic curve is straight from a to d and therefore there is faithful reproduction of the input wave over this portion of operation. The portion of the characteristic curve from a to c is not a straight line and therefore the reproduction is not linear, as demonstrated by the lower half of the envelope of the output current. The dotted lines *p* and *q* indicate what the shape of the envelope would be if the reproduction had been true.

When the voltage on the grid of a vacuum tube becomes positive a current flows through the grid. This current flows through the resistance of the input circuit, causing a voltage drop that changes the bias point of the amplifier. It is therefore necessary, if grid current is drawn, to make the resistance of the input circuit low. When it is made low, a comparatively high input power is required to produce the voltage necessary on the grid to obtain the desired output. Class A amplifiers are used to obtain high voltage amplification, and therefore high-impedance input and output circuits are used. Owing to the high-impedance input circuit, grid current cannot be allowed to flow in a Class A amplifier. Improper operation of a Class A radio amplifier, where grid current is drawn over part of the operation, is shown in Fig. 11-10. The bias voltage E_c is exceeded by the peak grid signal voltage input e_m during part of the time, and therefore the grid draws current when it is positive during the portion of time when e_m exceeds E_c. Operation over the characteristic curve between *c* and *d* is linear, and therefore the input voltage is faithfully reproduced over that portion of the curve. Over the portion of the curve from *d* to *e* the grid current flows, causing the reproduction to be nonlinear, as indicated by the envelope of the modulated wave. The dotted lines *p* and *q* indicate what the envelope would look like if the reproduction had been faithful.

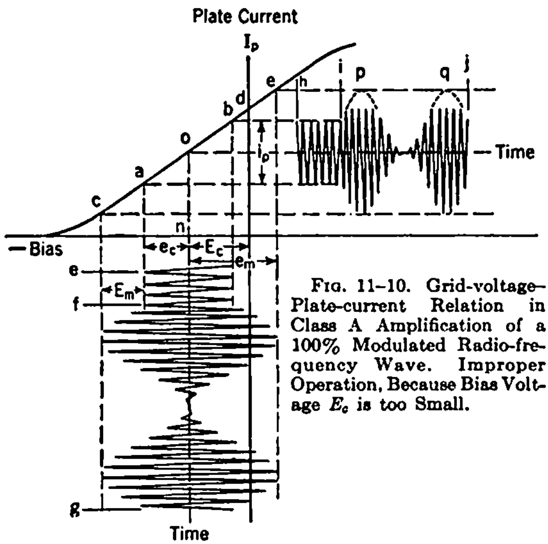


FIG. 11-10. Grid-voltage–Plate-current Relation in Class A Amplification of a 100% Modulated Radio-frequency Wave. Improper Operation, Because Bias Voltage E_c is too Small.

The D.C. plate ammeter indicates the average plate current *o-n* as shown in Fig. 11-8. If the operation is improper, as shown in Fig. 11-9 or Fig. 11-10, then the average plate current will change during modulation and the D.C. plate ammeter will change indication. If 100% modulation is not exceeded and the input modulated wave is sinusoidal, then proper operation of a Class A radio amplifier, as shown in Fig. 11–8, is indicated by no change in the D.C. plate ammeter during modulation.

PLATE DISSIPATION LIMIT. A given tube will dissipate only a certain amount of power in heat without being damaged. Class A radio amplifier tubes in receivers can have a plate dissipation which is large compared to the small amount of plate input power required, so there is usually little or no difficulty from dissipating heat. In transmitters, however, where large amounts of power may be involved, the matter of plate dissipation is of particular concern. The bias voltage must be adjusted so that the safe plate dissipation of the tube is not exceeded. Referring to Fig. 11–8, the bias voltage E_c on some tubes may need to be higher than that shown so that the tube operates at some point between *o* and *c*. The 100% modulated power output of the amplifier is limited then because operation on the characteristic curve is limited to twice *o* to *c*.

Neutralization. The grid-to-plate capacity of a triode vacuum tube is very small, and therefore at audio frequencies the impedance of this capacity is extremely high. At radio frequencies, however, the impedance is much lower. Therefore, at radio frequencies, an appreciable amount of energy from the plate circuit may *feed back* to the grid circuit through this capacity. The energy feedback from the plate circuit to the grid circuit may cause the tube to start oscillating. This oscillation interferes with the normal operation of the triode as a radio amplifier. It is possible to counteract the energy feedback from the plate to the grid. This process is called *neutralization.* In Fig. 11–11a C_{pg} represents the interelement plate-to-grid capacity. Energy from the output of the tube is fed from plate to grid through C_{pg}. This feedback usually causes oscillation. The feedback through C_{pg} may be neutralized; how this may be done is shown schematically in Fig. 11–11a. The tuning condenser for the output circuit is really two condensers in series, allowing the ground return to be made on the center of the two condensers. The voltage from N to M is across the plate and filament. This feeds voltage back to the grid through the capacity C_{pg}. The voltage from M to O is opposite in polarity or 180° out of phase with the voltage across MN. The voltage across MO causes a current to flow through the neutralizing condenser C_n to the grid. If capacity C_n is made equal to the plate-to-grid capacity C_{pg}, then the same voltage is fed to the grid through C_n as is fed through C_{pg}. Inasmuch as the resultant voltages on the grid are equal and of opposite phase, they cancel out at the grid of the tube, thus preventing oscillation.

Neutralization may be accomplished in a great many manners. Figures 11–11b and 11–11c show two practical forms of plate neutralization. C_n in both cases is the neutralizing condenser. In practice the neutralizing condenser is made variable and of a capacity in excess of the plate-to-grid capacity. The reason for this higher capacity is that stray capacities between wiring and pieces of equipment in the circuit will require a neutralizing capacity higher than the plate-to-grid capacity. The type of neutralization shown in Figs.11–11b and 11–11c is called *plate neutralization* because energy is fed from the plate circuit back to the grid.

Grid neutralization may also be employed. This type of neutralization is illustrated in Fig. 11–11d. The operation is identical with plate neutralization except that voltage is fed from the grid circuit to the plate. C_n is the neutralizing condenser.

Figure 11–11e illustrates a type of neutralization which is being used more and more in transmitters. C_{pg} represents the plate-to-grid capacity of the tube. Condenser C and inductance L represent the neutralizing circuit. Inductance L, of the proper value, is connected from plate to grid. In practice, a condenser C is inserted in series with inductance L so that the plate voltage is not impressed on the grid of the tube. The current fed

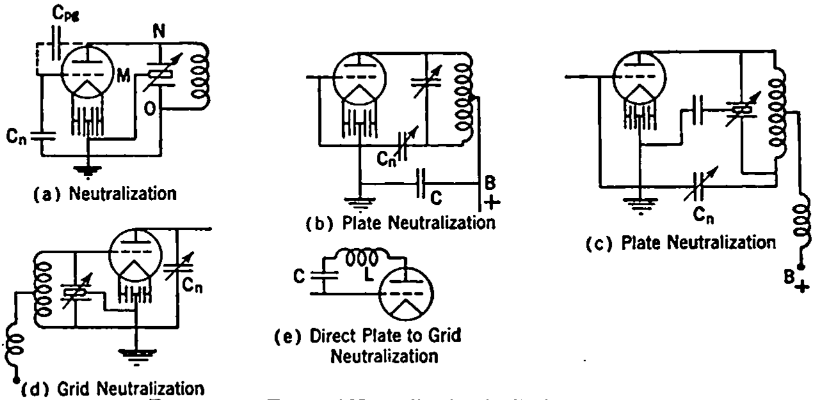


FIG. 11–11. Types of Neutralization for Radio Amplifiers.

from the plate to the grid through inductance L is out of phase with the current fed through C_{pg}. If L is of the proper value then the resultant voltages on the grid will be equal and out of phase. The voltage fed back through C_{pg} is neutralized on the grid by the voltage fed through inductance L.

Another type of neutralization is illustrated in Fig. 11–16. In push-pull the neutralization is usually what is called *cross neutralization*. The two condensers in Fig. 11–16 labeled C_n are the neutralization condensers. As may be seen from the diagram, they cross feed from the plate of one tube to the grid of the other.

Class A Pentode Radio Amplifier. Pentode vacuum tubes have a number of characteristics that make them ideal for use as voltage amplifiers. The principal advantages are large voltage or power amplification per stage, and freedom from the need for neutralization. The elements in a pentode are the filament or cathode, control grid, screen grid, suppressor grid, and plate. The suppressor grid is located in the tube between the plate and the other elements. When the suppressor grid is connected to the cathode of the tube, it behaves as an electrostatic shield between the plate and the control grid.

Figure 11–12 shows the diagram of a pentode radio-frequency amplifier. In the vacuum tube T, G_3 is the suppressor grid. There is an interelement capacity between the plate and the suppressor grid. As may be seen from the diagram, because of the radio-frequency by-pass condenser C_3, this capacity is virtually across the output inductance L_3 and thus forms part of the output circuit. There also is an interelement capacity between the suppressor grid G_3 and the control grid G_1. This capacity is virtually across the grid input circuit and therefore becomes part of the input tuning circuit. The two capacities are independent of each other owing to the fact that the suppressor grid is connected to the cathode of the tube. The capacity from the plate to the control grid G_1 is, for all practical purposes, not present. As a practical matter, there is a slight capacity present between the plate and the control grid through the meshes of the suppressor grid and the screen grid. The grid-to-plate capacity of a typical pentode tube is about .008 $\mu\mu$f. If the suppressor grid and the screen grid were removed from the tube, the capacity would be something on the order of several micromicrofarads. The suppressor grid reduces the plate-to-grid capacity to a value so small that it cannot feed enough energy back from the plate to the grid circuit to cause oscillation. The suppressor grid, accordingly, eliminates the need for neutralization.

Capacitive or inductive coupling between the plate and grid circuits of a radio amplifier, if of sufficient value, will cause oscillation. Referring to Fig. 11–12, if the output circuit coils L_3 and L_4 are placed in inductive relation to the input circuit comprised by L_1 and L_2, there will be a feedback causing oscillation. Oscillation could also be caused by running the leads from the grid and plate too close to each other. Other circuit components of the input and output circuits, if placed in proximity to each other, may cause enough feedback to make the vacuum tube oscillate.

Shielding. Shielding is resorted to in order to prevent oscillation from taking place as a result of these causes. The pentode vacuum tube itself is usually placed in a metal shield. In receivers this shield may be a small metal cylinder that fits rather snugly over the tube. Many types of tubes are made with a metal envelope instead of glass. These tubes are self-shielded and usually require no further shielding. The input and output coupling circuits also should be shielded. Usually the tuning inductances are fully shielded and limited shielding is sometimes employed around the variable tuning condensers. Small by-pass condensers are sometimes placed in metal shield housings. This shielding is not necessary in all parts of the circuit. It is also common practice in some applications to shield the lead from the input tuning circuit to the grid of the tube.

It is also important that the circuits of one amplifier be shielded from the circuits of other stages of amplification. When using pentode tubes shielding is automatically accomplished because the individual components are quite well shielded from each other, and therefore from components of other stages of the amplifier.

In triode amplifiers it is not so important as in other types to keep the various elements of an amplifier stage isolated from each other. However, it is good practice to follow the same general principles. In order that the stages of amplification shall be isolated from each other, each stage is sometimes shielded. Shielding the tuning inductances is usually sufficient in receivers and low-powered transmitters. In high-powered transmitters, in addition to shielding the individual tuning circuit inductances, each amplifier stage is often shielded completely.

Typical Pentode Radio Amplifier. Figure 11–12 is a circuit diagram of a typical pentode radio-frequency amplifier. The fundamentals of the

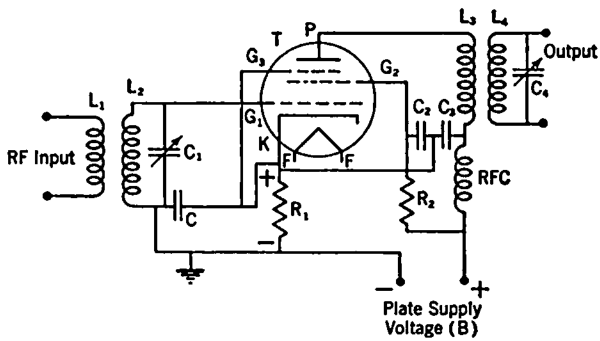


Fig. 11–12. Diagram of Pentode Radio-frequency Amplifier

circuit are the same whether the amplifier is used in a receiver or in a transmitter. An indirectly heated cathode is shown in Fig. 11–12. This type of cathode is common practice where pentodes are used in receivers. For power use in transmitters a directly heated filament is usually preferred to the cathode. The input resonant circuit to the vacuum tube T is composed of L_2 and C_1. The signal input voltage impressed on L_1 is induced into L_2. L_1 and L_2 comprise a radio-frequency transformer. L_4 and C_4 inductively couple into the plate circuit through the transformer formed by L_4 and L_3. This sequence constitutes the output resonant circuit of the amplifier. The plate of the vacuum tube receives its current through RFC and L_3. RFC is a radio-frequency choke and along with by-pass condenser C_3 stops the radio-frequency energy from passing through to the source of the plate voltage. The need for this blocking will be discussed in detail in a later paragraph of this chapter. The screen grid G_2 receives its voltage through resistor R_2. Screen grids usually operate at a voltage a little bit lower than the plate-supply voltage and they also draw a small amount of current. The screen-grid current passing through resistor R_2 causes a voltage drop that effectively reduces the voltage on the screen grid G_2. Condenser C_2 is a radio-frequency by-pass from the screen grid to the cathode circuit. R_2 and C_2 form a *filter* to keep the radio-frequency from passing through to the plate supply. Resistor R_1, connected between the ground and the cathode, is called a *cathode resistor*. The current drawn from the plate supply by the plate and the screen grid G_2 passes through resistor R_1 to the cathode, causing a voltage drop in R_1. The voltage across R_1 has the polarity shown in the diagram of Fig. 11–12. The control grid G_1 is connected to the negative end of R_1 through inductance L_2. Owing to the polarity of the voltage drop in resistor R_1 the control grid is negative in respect to the cathode. By proper selection of resistor value, the bias on the control grid may be set at the proper operating value. Condenser C is a by-pass condenser and serves two purposes. It is virtually a short circuit for radio frequency and therefore the input resonant circuit is effectively connected across the cathode and control grid. Condenser C also effectively by-passes the radio frequency around the resistor R_1.

Figure 11–12 is schematic of a practical circuit hookup of a radio amplifier. Many circuit variations are possible. There is no one circuit which is more fundamental than any one of a number of others. The class of the amplifier, the service that it is to perform (for example, whether it is used in radiotelegraphy or radiotelephony), and the power that the radio amplifier is to handle govern, to a marked degree, the specific circuit that is used and the values of the component parts.

Screen-grid Voltage Supplies. In some applications the voltage to the screen grid of a tetrode or pentode is the same as the plate voltage, and in

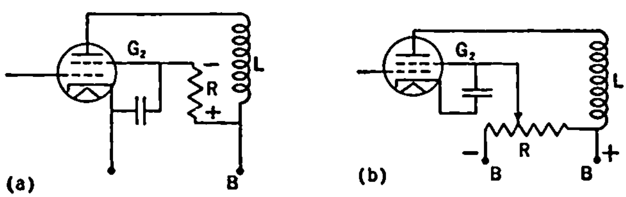


Fig. 11–13. Screen-grid Voltage Supplies.

such applications the screen grid is connected to the positive terminal of the B or plate supply. Usually, however, the screen grid operates at a voltage lower than the plate voltage. There are two principal means for obtaining the lower voltage from the plate supply. The two methods are shown in Fig. 11–13. L indicates the output circuit of the vacuum tube. A resistor R is shown in series with the screen grid G_2 in Fig. 11–13a. The screen grid draws current and therefore there is a voltage drop across the resistor

R. The net voltage on the screen grid is the plate-supply voltage minus the voltage across the resistor. The second principal method of obtaining the lower screen grid voltage is shown in Fig. 11–13b. A potentiometer or voltage divider is connected across the plate supply. A tap at the proper voltage is taken off the voltage divider to supply voltage to the screen grid G_2.

Tetrode Radio Amplifiers. Tetrode or screen-grid tubes contain a filament or cathode, control grid, screen grid, and plate. Screen-grid tubes may be used in radio-frequency amplifiers. The pentode tube has displaced the screen-grid tube in most receiving-set applications. The screen grid in a tetrode tube functions to reduce the plate-to-control-grid capacity. In addition to the screen grid, the pentode tube also has a suppressor grid. The plate-to-grid capacity of a pentode is less than it is in a tetrode tube. Therefore an amplifier using a pentode tube is less liable to oscillate than is one using a tetrode tube, and the pentode tube is usually preferred over the screen-grid tube in radio amplifiers.

The explanation of the operation of a pentode radio amplifier is applicable to an amplifier using the screen grid tube.

Filtering D.C. Supply Circuits to Radio Amplifiers. In a previous paragraph mention was made of *filtering circuits* in the plate and screen-grid leads from the plate-supply sources. The necessity for isolating various circuit components from each other has also been discussed. The importance of isolating cascaded amplifiers from each other was pointed out. A series of amplifier stages is usually operated from the same plate and bias supplies. If two cascaded stages of amplification are properly shielded it is still possible for feedback to occur through the plate or grid power supplies if precautions are not taken. A common way to prevent feedback through the power supplies is to use filters. Referring to Fig. 11–12, the radio-frequency choke RFC and the condenser C_3 form a filter which prevents the radio-frequency energy in the plate-circuit output from feeding into the plate supply, where it in turn could feed into the plate circuit of the preceding tube and from there to the grid circuit of the vacuum tube T. In some cases, this feedback could be enough to cause oscillation. The radio-frequency choke RFC is a high inductance presenting a high impedance to the radio frequency. The condenser C_3 presents a very low impedance to the radio frequency and therefore by-passes the radio frequency energy around RFC and the plate supply. In this manner the radio-frequency energy in the output of the vacuum tube is prevented from feeding into the plate power supply. In some cases a resistance may be substituted for the radio-frequency choke RFC. Resistance R_2 and condenser C_2 form a filter for the screen grid G_2 in the circuit of Fig. 11–12.

Figure 11–14 shows four different types of power-supply filters. These are labeled types 1 to 4. The type numbers given are arbitrary. Type 1 is the same as the filter composed of choke RFC and C_3 in Fig. 11–12. Type 3 is the same as the filter composed of R_2 and C_2 in Fig. 11–12. Type 2 and type 4 are the same as types 1 and 3 except that an additional condenser C_a has been added. The condenser C_a adds to the filtering action. In most cases there is a large capacity across the output of the D.C. supply circuit and therefore the condenser C_a is unnecessary.

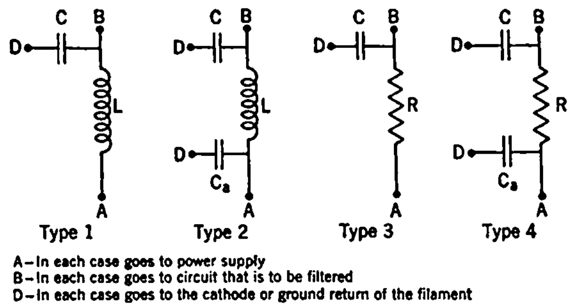


Fig. 11–14. D.C. Power-supply Filters.

In transmitters, there is another reason why radio-frequency supply circuit filters are important. Mercury-vapor rectifying tubes are in general use for changing 60-cycle alternating current into direct current for use as plate and bias power supplies. Radio frequency affects the operation of mercury-vapor rectifying tubes and therefore it is important that radio frequency be isolated from them.

Transmitter Output Tube Complements. The output power of radio transmitters varies from a fraction of a watt up to several hundred kilowatts. The amount of power required is determined by the service of the communication system. It would be desirable to have a single vacuum tube in the output amplifier of a radio transmitter. For economic reasons, the power output capacity of the output tubes should not be much higher than is required for the operating power of the transmitter. Thus, if the transmitted carrier power is 1,000 watts, it is uneconomical to operate with tubes that can produce an output power of 5,000 watts. The larger the tubes are, the more expensive they are. The large tubes also require more power to light the filament.

It is impractical for the tube manufacturers to produce a single tube with the proper power output for each and every use. Neither is it practical to build tubes larger than a certain size; therefore, if the use demands a power higher than the capability of the largest tube, it becomes necessary to use two or more tubes in the output amplifier.

By using combinations of two or more tubes in transmitter amplifiers, it has been found feasible to standardize in some degree on the power capability of vacuum tubes. A multitude of different power outputs is obtained by proper combination of a limited number of vacuum tubes of various power ratings. When two or more tubes are used in an amplifier, they may be operated either in *parallel* or in *push-pull*. Tubes so operated must be similar.

Parallel operation of vacuum tubes. Any number of vacuum tubes may have their elements connected in parallel. The power output capability of the combination is equal to the sum of the power output capabilities of the individual tubes. Figure 11–15 (a) shows three vacuum tubes with the filaments connected in parallel, the grids connected in parallel, and the plates connected in parallel.

When tubes are connected in parallel there is a tendency for *parasitic oscillations* to be set up at a very high frequency. Parasitic oscillations are extra oscillations occurring at a frequency different from the main or desired oscillations. They take power away from the intended output and sometimes cause undesirable heating of the tube elements and erratic operation. This tendency may be present even in a single tube. How-

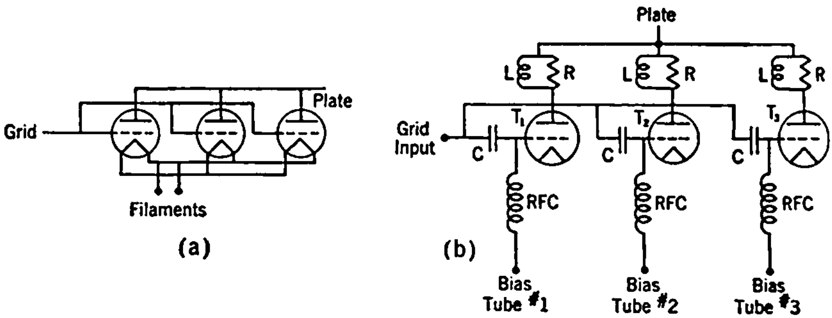


Fig. 11-15. Parallel Operation of Vacuum Tubes.

ever, it is not nearly as serious as it is when two or more tubes are connected in parallel. Parasitic oscillation interferes with the normal operation of the tube. The circulating currents from parasitic oscillations can readily become so high that the parasitic current flowing through the grid return in the glass seal of the tube will heat the conductor to the extent that it will damage the tube. Parasitic oscillations may be prevented by inserting a simple circuit in series with the plate of each tube. Figure 11–15b shows the scheme of parasitic chokes associated with three vacuum tubes operated in parallel. The inductances L are small, usually something on the order of about 15 turns on a diameter on the order of an inch or two. The inductance presents a very high impedance to the parasitic frequency because the parasitic frequency is usually very high. The inductance, however, is small enough so that it has no effect at the operating frequency. The resistance R is usually of the order of 50 ohms and is placed in parallel with the inductance L. This circuit serves to damp out the parasitic oscillations. The values for L and R given above would be suitable for use in the medium frequencies, such as the broadcast band of 550 to 1,600 kc. At higher frequencies the inductance is smaller.

Vacuum tubes of the same type do not possess absolutely identical characteristics. It is sometimes necessary to allow for this fact in the circuit of an amplifier using vacuum tubes in parallel. Usually this allowance takes the form of a separate bias-voltage adjustment for each tube. A method for supplying the proper bias to each individual tube is shown in Fig. 11-15b. The signal voltage input is supplied to each tube through condenser *C*. Separate bias voltage may therefore be fed to each grid through the radio-frequency chokes RFC.

Push-pull operation. The second method of using multiple tubes in a radio amplifier is to operate them in push-pull. This method of operation requires that the number of tubes in use be even. Push-pull operation

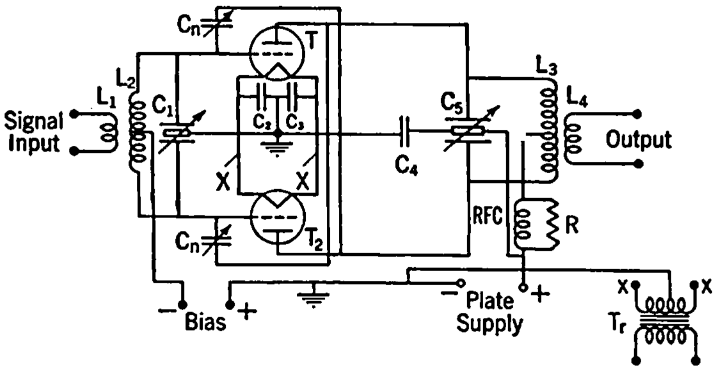


Fig. 11-16. Push-pull Triode Radio-frequency Amplifier.

of tubes has several advantages over parallel operation. Cross neutralization of a push-pull amplifier does not complicate the circuits as much as does neutralization of a single-ended amplifier.

Two tubes in push-pull operation will give a slightly higher power output than will the same two tubes operating in parallel. Another advantage push-pull has is that theoretically the even harmonics produced by the vacuum tubes are canceled out. In practice this cancellation amounts to a marked reduction in the amplitude of the even harmonics provided circuits are carefully balanced. A circuit diagram of a push-pull triode radio-frequency amplifier is shown in Fig. 11-16. L_2 and C_1 comprise the resonant input circuit. The signal input voltage in L_1 is transformed to L_2, where it is impressed upon the grids of the vacuum tubes. The tuning condenser C_1 is a double-section condenser. Each section is a condenser by itself and the two sections are connected in series so that the midtap may be grounded to form a balanced input circuit to the grid of each tube. Balance is necessary in order that the voltage impressed on the grid of each tube shall be the same. The output circuit is comprised of inductances L_4 and L_3 and condenser C_5 and is similar to the input circuit. Condenser C_4 is connected from the midtap of the tuning condenser to ground, thus giving a balanced output circuit so that each tube works into the same impedance. The plate supply to the vacuum tubes is obtained through the radio-frequency choke RFC and the inductance L_3. In this type of plate-supply circuit there is a tendency for parasitic oscillations to form. A resistor R is placed across the radio-frequency choke RFC to damp out these parasitic oscillations. The two condensers labeled C_n are the neutralizing condensers and are connected for *cross neutralization.*

The radio-frequency amplifier shown is typical in transmitters. The transformer T_r is used to light the filaments of the tubes. The plate current returns to the filaments through the midtap of the filament trans-

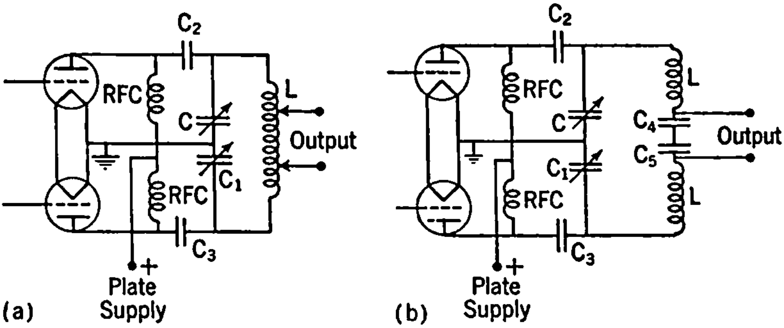


FIG. 11-17. Types of Push-pull Radio-amplifier Output Circuits.

former. It is not feasible to return the plate current through a filament midtap resistance because the resistance would need to have a very high current-carrying capacity and would necessarily be unduly large. It is therefore common practice in transmitters to return the plate current through the midtap of the filament transformer, whereas in receivers it is more common to make the return through midtap resistors. In some low-power transmitter amplifiers the midtap-resistor return is used.

OUTPUT CIRCUITS, PUSH-PULL RADIO AMPLIFIER. One type of push-pull radio amplifier output circuit is shown in Fig. 11-16. Two other push-pull radio amplifier output circuits are shown in Fig. 11-17. The plate supply was series fed in the diagram shown in Fig. 11-16. In both diagrams shown in Fig. 11-17 the plate supply is parallel. The plate voltage is fed to the plates of the tubes through the radio-frequency chokes RFC. The blocking condensers C_2 and C_3 prevent the plate-supply voltage from being impressed upon the tuning inductance. In Fig. 11-17a the output of the amplifier is obtained from taps on the tuning inductance L. In the circuit shown in Fig. 11-17b, the output is obtained across the condensers C_4 and

C_5. The circuit shown in Fig. 11–17b is preferred in amplifiers used in the final stage of a transmitter because the condensers C_4 and C_5 aid in reducing harmonics. The condensers C_4 and C_5 have a lower impedance at the harmonics and therefore partially shunt the harmonics around the output terminals. The hookup shown in Fig. 11–17a is more applicable to interstage radio amplifiers.

Class B Radio-frequency Amplifier. Class B radio-frequency amplifiers are mainly used in transmitters for the purpose of amplifying a modulated signal. The main characteristics of Class B radio amplifiers are faithful reproduction together with higher plate efficiency and greater power output than Class A radio amplifiers. Class B amplifiers are biased at the cutoff point so that, with no signal input to the grid, there is no flow of plate current.

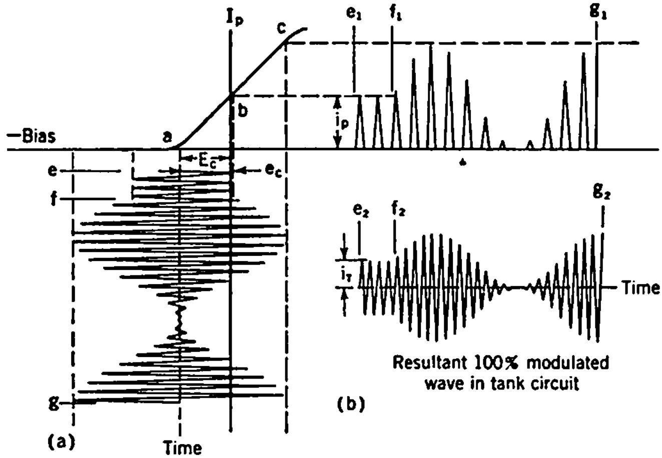


FIG. 11–18. Grid-voltage–Plate-current Relation in Class B Amplification of a Modulated Radio-frequency Wave. Proper Operation.

Referring to Fig. 11–18a, for proper operation a Class B radio amplifier is biased to point *a*. During the time interval *e* to *f*, unmodulated radio-frequency signal voltage is impressed on the grid of the tube. Owing to the operation at the cutoff point, the negative half of each cycle of the signal voltage impressed on the grid produces no current in the plate circuit. The result is that half-cycle pulses of plate current are made to flow as shown in time interval e_1 to f_1. These half-cycle pulses are impressed on the resonant output tuning circuit. A resonant circuit has the property that if part of a cycle at a given frequency is impressed upon it, the pulse excites the resonant circuit and produces a full sine wave. This fact makes it possible to operate radio amplifiers at the cutoff point single-ended. Figure 11–18b represents the instantaneous tank current circulating in the output resonant circuit. Time interval e_2 to f_2 corresponds to time interval e_1 to f_1 and represents the full-sine-wave current flowing in the tank circuit produced by the half-sine-wave pulses received from the plate circuit of the vacuum tube.

Class B radio amplifiers may be operated over a portion of the characteristic curve extending up to where the top slopes off. The operating portion of the characteristic curve as shown in Fig. 11–18a is from a up to where the curvature starts at point c. To amplify faithfully a 100% modulated radio signal, the signal input voltage without modulation must be half of ac or up to point b. During the time interval f to g the carrier is 100% modulated by an audio frequency. As may be seen from Fig. 11–18a, the part of the signal input voltage to the left of the operating axis has no effect on the instantaneous plate current i_p. During modulation the pulses of radio-frequency current in the plate circuit have an envelope (represented

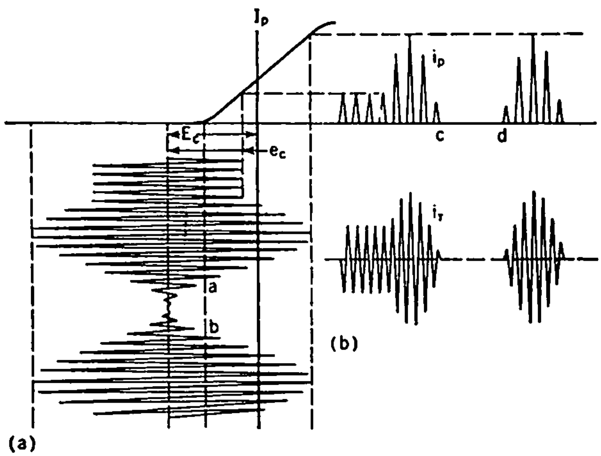


FIG. 11–19. Grid-voltage–Plate-current Relation in Class B Radio Frequency Amplification. Improper Operation. Bias Too High.

in the time interval f_1 to g_1) that corresponds with the modulation frequency. The half-cycle pulses of plate current excite the resonant output circuit, producing full sine waves of radio frequency to produce a modulation envelope as shown in Fig. 11–18b between f_2 and g_2.

EFFICIENCY. The efficiency of Class B radio amplifiers for the carrier only without modulation is much higher than that of Class A radio amplifiers. Class B radio amplifiers operate over a greater portion of the characteristic curve than do Class A radio amplifiers and therefore a given tube will give a much greater output. As may be seen from Fig. 11–18a, the signal input voltage at 100% modulation exceeds the bias voltage and therefore the grid goes positive. For this reason the impedance of the grid input circuit must be low. Therefore a higher input power is required for a Class B amplifier than is required for a Class A radio amplifier where the input impedance is high. A carrier efficiency as high as 35% is obtainable from Class B radio amplifiers.

Figure 11–19 shows the effect produced when the bias of a Class B amplifier is too high. During modulation the signal input voltage between points *a* and *b* is such that there is complete cutoff of the plate current and therefore the envelope of the radio-frequency pulses in the plate circuit is not a sine wave. Part of the full envelope has been cut off between points *c* and *d*. Figure 11–19b shows the current in the tuned circuit; the diagram shows that the envelope of the modulated wave is not a sine wave.

The single-ended amplifier circuits shown so far and the push-pull radio amplifier shown in Fig. 11–16 may be operated as Class B radio amplifiers if the constants of the circuits are properly selected and the voltages are properly adjusted.

Class B Audio Amplifiers. Class B audio amplifiers may operate at efficiencies as high as 50% to 70% for maximum output. They are there-

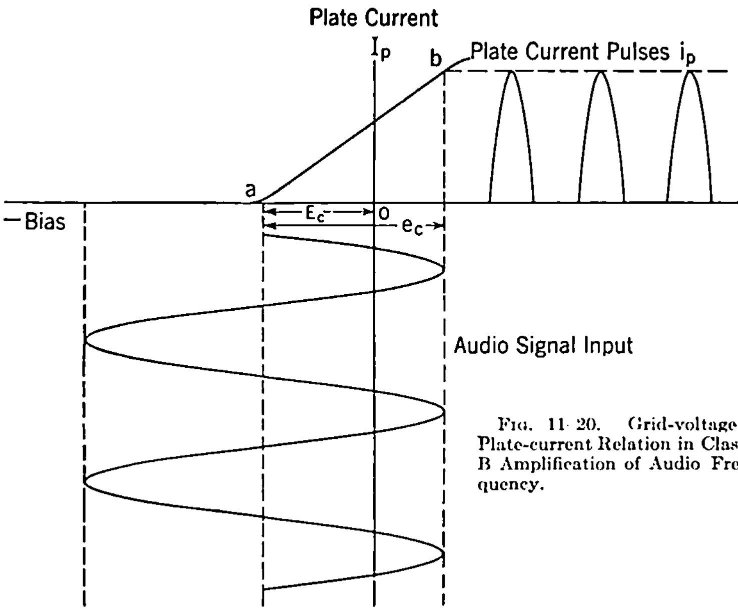


Fig. 11–20. Grid-voltage-Plate-current Relation in Class B Amplification of Audio Frequency.

fore used whenever a large amount of audio power is required. Many radiotelephone transmitters are modulated in the output amplifier. The amount of audio power required for 100% modulation is half of the plate power input to the amplifier. As an illustration, a 10,000-watt transmitter operating at an efficiency of 80% has a plate input power of 12,500 watts.

The audio power required for 100% modulation is 6,250 watts. Audio powers of this magnitude are difficult and expensive to obtain from Class A audio amplifiers. High-efficiency Class B audio amplifiers are therefore used in transmitters to produce the audio power required for modulation. An amplifier so used is termed a *modulator*.

Figure 11–20 shows the grid-voltage–plate-current relation of a vacuum tube used in Class B audio amplification. The bias voltage E_c fixes the operation point at the cutoff point of the characteristic curve. The voltage e_c is an audio signal voltage impressed on the grid of the tube. This voltage causes half-cycle pulses of plate current to flow, as shown by i_p. As

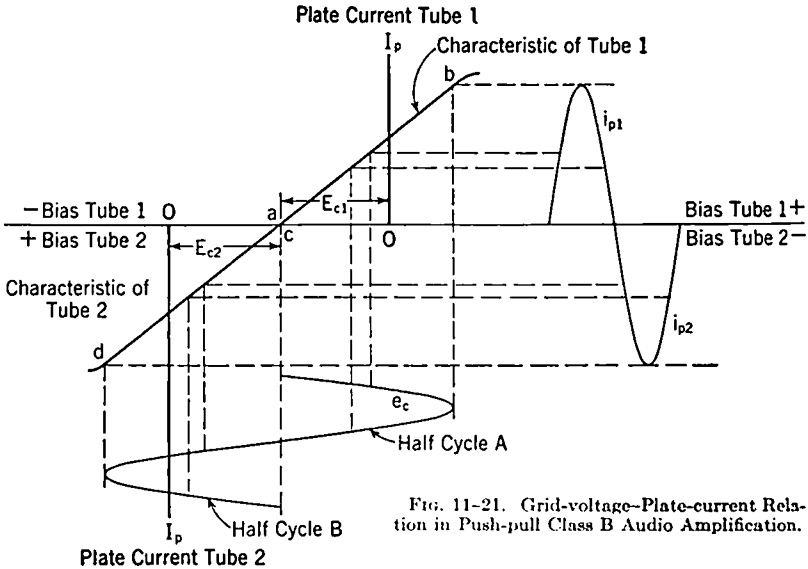


FIG. 11–21. Grid-voltage–Plate-current Relation in Push-pull Class B Audio Amplification.

explained earlier in the chapter, resonant circuits cannot be used for audio amplifier coupling. Class B audio amplifiers use a transformer for the output coupling. Half-cycle sine-wave pulses impressed on a transformer produce a distorted output wave and not a full sine wave as does a resonant circuit. It is necessary, therefore, to operate Class B audio amplifiers in push-pull so that, in effect, the first tube will produce a half cycle while the second tube is idle. The second tube then produces the other half of the cycle while the first tube is idle. The two tubes in combination produce the full-sine-wave output.

Figure 11–21 represents graphically the operation of two tubes in a Class B audio amplifier. The characteristic curve of tube 1 is shown in the upper portion of the diagram. The characteristic curve of tube 2 is shown turned around in such a way that the two characteristic curves form a straight line. The voltage e_c represents the audio-frequency signal input voltage to the grids of the tubes. Both tubes are biased to the cutoff point so that there is little or no plate current flowing when there is no signal input voltage. During the half cycle of e_c labeled A the grid of tube 2 is made more negative and therefore the plate current of tube 2 remains at zero. In tube 1, during the half cycle A, the total grid voltage (the sum of E_c, the bias voltage, and e_c, the signal input voltage on the grid) becomes more positive and operation takes place over the portion of the characteristic curve of tube 1 from *a* to *b*. A half-cycle pulse of plate current i_{p1} is produced. During the half cycle B of grid input, the grid of tube 1 becomes more negative and therefore the plate-current flow from tube 1 remains at zero. The total voltage on the grid of tube 2, however, becomes more positive and operates over the portion of the characteristic curve of tube 2 from *c* to *d*. This causes a half-cycle pulse of plate current

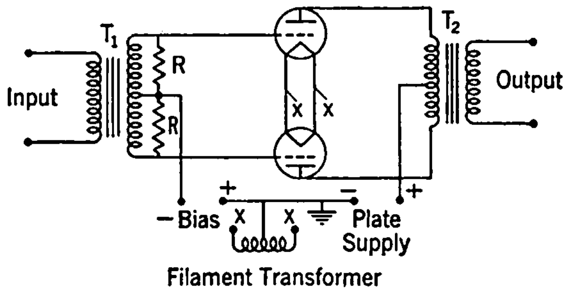


FIG. 11–22. Circuit Diagram of a Push-pull Class B Audio Amplifier.

I_{p2} to flow. I_{p1} supplied by tube 1 and I_{p2} supplied by tube 2 form a full-wave reproduction of the input signal voltage to the grids of the tubes.

CLASS B AUDIO AMPLIFIER CIRCUIT DIAGRAM. A simple circuit diagram of a push-pull Class B audio amplifier is shown in Fig. 11–22. The signal input voltage is impressed on the grids of the tubes through the input audio-frequency transformer T_1. The output power is obtained from the plates of the tubes through the audio-frequency transformer T_2.

Figures 11–20 and 11–21 show that in operation the grids of the vacuum tubes in Class B audio amplifiers go positive and therefore draw grid current. It is therefore necessary to use a low-impedance input to the grids. The low-impedance input requires a relatively high input power to produce the necessary grid input voltage across the resistors *R* and *R*. For proper operation it is necessary that the bias voltage shall not change when the input signal is applied. Without an input signal no current is drawn from the bias supply. However, as soon as the input signal becomes high enough so that the total voltage on the grid is positive, current is drawn by the grid of the tube. In order that the bias voltage may be held as constant as possible, it is necessary that the regulation of the bias-voltage supply be good.

Pulses of current are drawn from the plate supply by the plates of the vacuum tubes. The amount of current drawn from the plate supply varies from zero to the total plate current at the crest of the output audio-frequency wave. This range of current is comparatively large. As in the grid circuit, it is necessary to maintain the plate-voltage supply at a nearly constant voltage. It is therefore necessary to have a plate supply with good regulation for supplying power to a Class B audio amplifier. To operate properly, Class B audio-frequency amplifiers require very careful design in order to assure that all the rigid requirements for proper operation be met.

POWER AMPLIFICATION. Class B audio amplifiers using triode tubes have a relatively low power amplification. Pentode tubes produce a higher power amplification. However, they have other disadvantages. This problem will be discussed in a later paragraph. The low power-amplification gain of a class B audio amplifier using triode tubes is more than offset by the high power output and high output efficiency for applications where a large amount of audio power is required.

Class C Radio Amplifiers. The chief characteristics of Class C radio-frequency amplifiers are that they have the highest output power and output efficiency of all classes of radio amplifiers, but they have power amplification less than Class B radio amplifiers and can amplify only unmodulated waves. Class C radio amplifiers, because of their high efficiency, are used in the output amplifiers of radiotelegraph transmitters and in the modulated stages of radiotelephone transmitters when using plate or grid modulation.

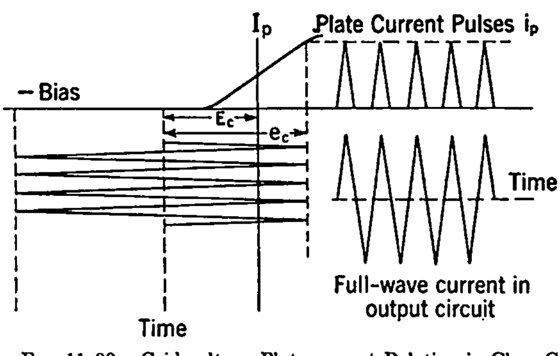


FIG. 11-23. Grid-voltage–Plate-current Relation in Class C Radio Amplification.

Figure 11–23 shows the grid-voltage–plate-current relation in Class C amplification. A high bias voltage is used so that the tube is operating beyond the cutoff point. Usually the bias voltage is about twice the voltage required for cutoff. Referring to Fig. 11–23, e_c represents the radio-frequency signal voltage impressed on the grid of the tube. Pulses of plate current i_p are produced in the plate circuit of the tube. In discussing

Class B radio amplification, it was shown that these pulses were a full half cycle. In Class C amplification, it can be seen by an inspection of Fig. 11–23 that the plate pulses i_p are less than a full half cycle. These pulses, however, being impressed on the resonant output tuning circuit of the amplifier, produce a full-sine-wave reproduction of the radio-frequency signal voltage impressed on the grid of the tube. This type of pulse excitation of the resonant circuit produces a higher plate-circuit output efficiency than does the full-half-wave signal-pulse excitation in the Class B radio amplifier. Class C radio amplifiers operate at efficiencies as high as 80%. If an attempt is made to modulate the input signal voltage e_c the plate circuit output modulation envelope will be greatly distorted because the entire length of the characteristic curve has been used to produce the unmodulated signal. However, the output signal can be properly modulated. This output modulation is discussed in Chapter 12. The radio-frequency amplifier circuits shown previously in this chapter may be used for Class C radio amplification by proper selection of component parts and voltages.

Pentode Radio Amplifiers, Class B and Class C. Because of their high voltage amplification, pentode tubes are ideally suited for Class A amplifiers when the maximum voltage amplification is desired. Pentode tubes, capable of a power output of several hundred watts, are available for use in transmitters. In addition to having a large voltage amplification factor, pentodes also possess a high power amplification. Pentode tubes, operated as power amplifiers, possess the undesirable property of high distortion. In services where voice transmission only is required, the high distortion is not objectionable and therefore they are coming into wide use for these services. Pentode tubes are coming into extensive use in transmitters in stages previous to the modulated amplifier. Research work is constantly improving pentode tubes and new ways are being devised to use them so that their usefulness steadily increases.

Pentode Audio Amplifier, Class B. Because of their high power amplification, pentodes are being used more and more in Class B audio amplifiers for services where distortion is not particularly important. They are extensively used for output audio amplifiers in inexpensive receiving sets. They are also extensively used in the output amplifiers of public-address systems. They are also becoming more widely used in voice radiotelephone transmitters.

Pentodes have extremely high plate resistance. Fortunately, their characteristics are such that a low-impedance load may still be used when the tube is employed as a power amplifier.

Detection or Demodulation. The process of modulating a radio-frequency carrier to transmit intelligence is described in Chapter 10. The envelope of the modulated wave varies at an audio-frequency rate. An audio-frequency wave derived from the modulation envelope may be used to actuate earphones or a loudspeaker, producing an audible sound to the listener. The process for obtaining the signal from the modulated wave is called *detection* or *demodulation.*

There are many different types of detectors. However, fundamentally, they all operate in the same way. The operation of a detector or demodulator is accomplished in three steps. These are *rectification, filtering,* and *separation of the audio-frequency component.*

RECTIFICATION. A radio-frequency signal is shown in Fig. 11–24b. During the time interval *f* to *g* the carrier wave is unmodulated. During the time interval *g* to *h* the carrier wave is amplitude-modulated by an audio frequency. As described in Chapter 4, if an A.C. wave is impressed upon a rectifier, the positive halves of each cycle will pass through and the negative halves will be cut off. Figure 11–24a shows a radio-frequency input, a rectifier, and an output in series. If the radio-frequency signal shown in Fig. 11–24b is impressed on the input, then the output will be as shown in

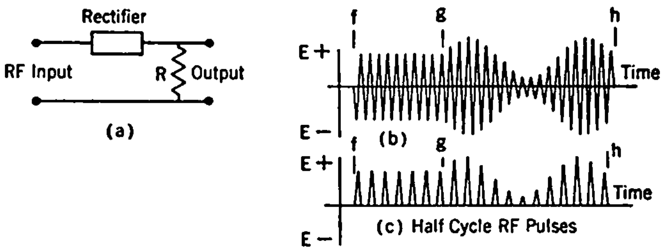


FIG. 11–24. Rectification of Radio-frequency Signal.

Fig. 11–24c. The signal in the output will be half-cycle pulses of radio frequency. The envelope of the half-cycle pulses will be identical with the envelope of the input signal shown in Fig. 11–24b. Rectification of the radio-frequency signal is the first step in detection.

FILTERING OF THE RECTIFIED SIGNAL. The second step in detection is to filter the rectified radio-frequency wave so as to obtain current that varies at an audio frequency in accordance with the radio-frequency signal input. Condenser *C* and resistor *R*, as shown in Fig. 11–25a, constitute such a filter. The rectified radio-frequency input signal to the filter is shown in Fig. 11–25b. This signal is the same as the signal shown in Fig. 11–24c. As was shown in Fig. 11–24, if the output of the rectifier is a resistance, then the voltage across the output resistance *R* is a series of half-wave pulses at a radio-frequency rate. The addition of a condenser *C*, as shown in Fig. 11–25a, smooths out or filters the pulses, thereby producing a varying voltage across *R*. The current will vary in accordance with the envelope of the input signal. At the time when the first part of the first half-wave signal is impressed upon the filter, the voltage across *C* and *R* increases. This increase occurs between *a* and *b* of Fig. 11–25b. At the point *b* there is a voltage *e* across the resistor and condenser. As the voltage of the half wave decreases between *b* and *c*, the voltage across *R* would normally decrease with it. However, during *a* to *b* the condenser *C* has charged up. The condenser, during the time interval *c* to *d*, discharges through *R*, maintaining a constant voltage across *R*. The result is a direct current flowing through *R*, thereby producing a D.C. voltage across *R*. The positive half-wave radio-frequency pulses have produced a D.C. voltage across *R*. Referring to Fig. 11–25b, during the time interval *f* to *g* the input signal is unmodulated and therefore a constant D.C. voltage is produced across *R*, as shown by e_1 of Fig. 11–25c. During the time interval *g* to *h* of Fig. 11–25b, the input signal is modulated. This causes the output voltage across *R* to vary in accordance with the modulation. The current variation from *g* to *h*, as shown in Fig. 11–25c, is the same as that of the

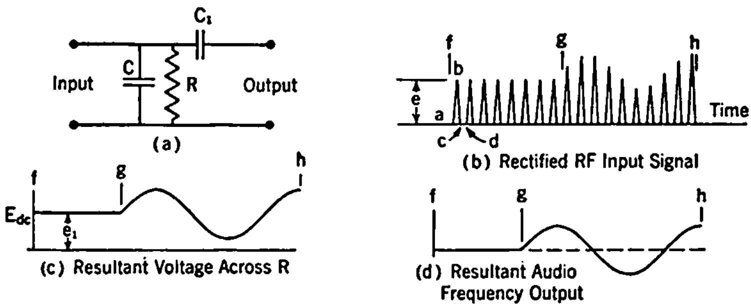


FIG. 11–25. Filtering of Radio-frequency Wave to Obtain Audio Frequency.

envelope of the modulated wave as shown between *g* and *h* of Fig. 11–25b. The voltage across *R* is a varying voltage during the time interval *g* to *h*.

SEPARATION OF AUDIO-FREQUENCY COMPONENT. Audio frequency is required for the operation of earphones or a loudspeaker in order that sound waves may be produced that can be heard by the human ear. It is therefore desirable to obtain an audio frequency that is in accordance with the envelope of the varying voltage shown in Fig. 11–25c. Condenser C_1 shown in Fig. 11–25a is an audio-frequency by-pass condenser. C_1 allows alternating current to flow through it but it does not allow direct current to pass. The result at the output terminals will be an audio frequency, as shown by Fig. 11–25d. The audio frequency is in accordance with the modulation envelope of the original radio-frequency signal wave, as shown in Fig. 11–24b.

In the three steps of detection, namely rectification, filtering, and separation of the audio-frequency component, an audio frequency has been obtained that conforms with the envelope of the modulated radio-frequency signal input.

Types of Rectifiers. Any device that conducts electricity in only one direction is a rectifier. There are three principal classes of rectifiers used in detectors. These are *crystals, diode vacuum tubes,* and *triode vacuum tubes.*

CRYSTAL DETECTORS. Crystal detectors were extensively used until the advent of the vacuum tube. They have been used little since the early 1920's. Recent developments have produced detector requirements that are admirably filled by the characteristics of crystals. They are, therefore, again becoming important. The diagram of a simple crystal detector is shown in Fig. 11–26. The input circuit, consisting of inductance L_2 and condenser C, is tuned to the carrier frequency. The incoming modulated signal is rectified by the crystal X. Condenser C_2 and resistor R comprise the filtering circuit. Condenser C_3 separates the audio-frequency component from the varying voltage across R. Figure 11–26 shows the signal in the various stages of detection. The individual steps were discussed in detail when considering Fig. 11–24 and 11–25.

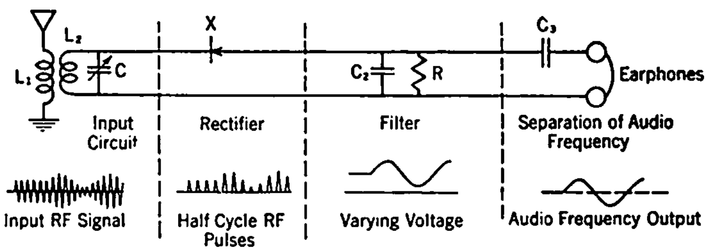


FIG. 11–26. Simple Crystal Detector.

For all practical purposes, the D.C. resistance of the earphones could be substituted for resistor R and the condenser C_3 eliminated. In a simple receiver such as this, the earphones can handle the D.C. component of the output of the filter. With this modification the circuit diagram shown in Fig. 11–26 would be typical of some of the receivers used previous to the advent of the vacuum tube.

Crystal detectors have several advantages as well as disadvantages, in comparison with vacuum-tube rectifiers. A detector circuit using a crystal rectifier is simple because there are only two leads from a very small instrument. In comparison with crystals, vacuum tubes have the disadvantage that the filaments must be heated and the filaments burn out. The vacuum-tube rectifier is also more bulky than is a crystal detector. Crystals have a sensitivity that compares quite favorably with diode detectors. At extremely high carrier frequencies the interelement capacity of a vacuum-tube detector and the spacing of the elements is important. A crystal rectifier, on the other hand, has an extremely low capacity between terminals and the detecting element is a point. Therefore crystal rectifiers have a decided advantage in the extremely high frequencies.

A piece of crystal is usually imbedded in some soft metal with a low melting point, such as Wood's metal, with one face of the crystal exposed. This crystal face forms one terminal of the rectifier. A sharp-pointed piece of wire forms the other terminal. The crystal does not have good rectifying properties all over its face, and therefore it is necessary to move the point of the wire around on the face of the crystal until a good rectifying point is found. The contact wire is usually easily dislodged from the sensitive point and therefore the rectification is unstable and the detection erratic. In order to avoid this difficulty, a good spot is found on the crystal and the wire is securely held in place by some sort of sealing compound. This sealed position makes a stable and reliable rectifier and therefore greatly extends the usefulness of the crystal.

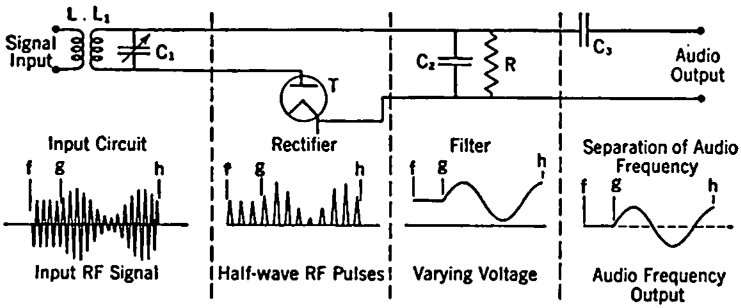


FIG. 11-27. Diode Detector.

DIODE DETECTORS. The principal characteristics of diode detectors are faithful reproduction and low sensitivity. In a diode detector, the rectifier is a diode or a two-element vacuum tube containing a filament or cathode and a plate. This type of tube is a unilateral conductor of electricity and therefore will rectify an impressed alternating current. A diode detector circuit is shown in Fig. 11-27. The radio-frequency signal input voltage is induced into the resonant circuit by inductance L. The tuned circuit, consisting of inductance L_1 and condenser C_1, is tuned to resonance. The radio-frequency input signal wave form is shown directly below the input circuit. During the time interval f to g, the carrier wave is unmodulated. During the time interval g to h, the carrier is modulated by an audio frequency. Rectification of the input signal occurs in the diode vacuum tube T. Directly below T is the wave form of the half-wave radio-frequency pulses formed as a result of the rectification. Condenser C_2 and resistance R comprise the filter. The wave form of the varying voltage across R is shown directly below the filter. During the time interval from f to g, when the carrier is unmodulated, the voltage across R is a constant D.C. voltage. During the time interval g to h, when the input carrier wave is modulated, the voltage across R is varying at an audio rate. The varying voltage is the same as the envelope of the positive half of the radio-frequency input signal voltage.

Condenser C_3 is an audio-frequency blocking condenser separating the audio frequency from the D.C. component of the voltage across R. The wave form of the audio frequency is shown directly below C_3. During the time interval f to g, when there is no modulation, there is also no audio output. During the time interval g to h, when the carrier is modulated, an audio frequency appears at the output terminals.

Diode detectors must work into a high load resistance and therefore resistor R will have a value on the order of $\frac{1}{2}$ to 1 megohm. If the condenser C_2 is made too large, it will not only filter out the half-wave radio-frequency pulses but it will also filter out the audio modulation of the voltage across R. The higher the modulation frequency, the smaller the value of C_2 may be. C_2 may have a value on the order of 150 to 200 $\mu\mu$f. The values given for C_2 and R would be about right for the medium frequencies in the neighborhood of the standard broadcast band of 550 to 1,600 kc. At higher frequencies, C_2 could be made smaller.

Figure 11–28 shows the rectification characteristics of a diode vacuum tube. The static relation between input voltage and output current is shown by the line *diode plate-current characteristic.* The voltage e_c represents the radio-frequency signal input voltage to the diode. During the time interval f to g the carrier is unmodulated. During the time interval g to h the carrier is modulated by an audio frequency. It can be seen from the diagram that the negative half-wave cycles of radio frequency are cut off while the positive ones are allowed to pass through the tube. This separation produces the half-wave radio-frequency pulses labeled i_p. The dotted curve marked d is the average of the radio-frequency half-wave pulses. The envelope of d is in accordance with the envelope of the modulated input signal. In other words, the curve d is the same as the varying voltage shown in Fig. 11–27.

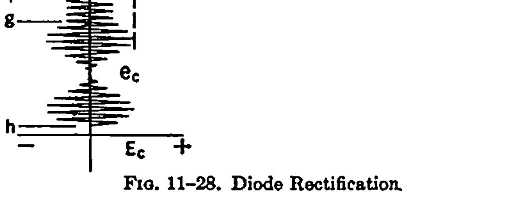


Fig. 11–28. Diode Rectification.

Figure 11–28 represents the first two steps of a complete detecting system. The current i_p represents the result of rectification and d represents the result of filtering.

The third step, *separation of audio frequency*, is really not part of the detecting process. However, for purposes of clarity it has been included in the explanation of detector action so far.

Triode Detectors. Triode vacuum tubes may be used in detector circuits. The methods of detection divide into two general classes. These are *plate-circuit detection* and *grid-circuit detection.* There are two types of plate-circuit detectors, the *linear plate detector* and the *square-law plate detector.* Grid-circuit detection and square-law plate detection possess undesirable characteristics and are little used. Linear plate detectors and diode detectors, because of their fidelity of reproduction, are extensively used. In linear plate detection, the triode vacuum tube also acts as an amplifier and therefore produces some gain in signal.

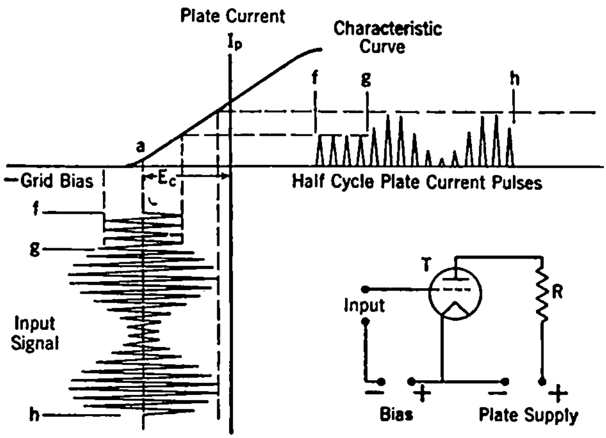


Fig. 11–29. Triode Rectification.

Linear plate detection. Figure 11–29 shows a triode vacuum tube connected as an amplifier with a pure resistance output *R*. The grid of the tube is biased to the cutoff point by the negative bias voltage. A radio-frequency signal impressed on the grid of the tube will produce half-cycle pulses of radio frequency through resistance *R*. The envelope of the half-cycle pulses in *R* will be the same as the envelope of the positive half of the signal input voltage envelope. This process in effect is rectification, and if the half-cycle pulses are put through a filtering circuit the full process of detection has been performed. Because the vacuum tube is connected as an amplifier, there will also be a certain amplification between the input and output.

A linear plate-circuit detector diagram is shown in Fig. 11–30. The triode vacuum tube T is biased to the cutoff point. The signal input voltage is impressed on the grid of T through the tuning circuit composed of inductance L_2 and condenser C. The input circuit is tuned to resonance. The signal input wave form is shown directly below the input circuit. The vacuum tube T performs the rectification. The half-cycle radio-frequency pulses are shown below the vacuum tube. The output load circuit of the vacuum tube is composed of the resistance R shunted by the condenser C_1. R and C_1 form the filter for the detector action. Directly below C_1 and R is shown the varying voltage. Separation of the audio frequency is attained by the condenser C_2. The resultant audio frequency is shown directly below.

Referring again to Fig. 11–29, the linear portion of the characteristic curve must be used or else there is distortion, as was explained in discussing Class B radio amplifiers. The curvature in the characteristic at point a

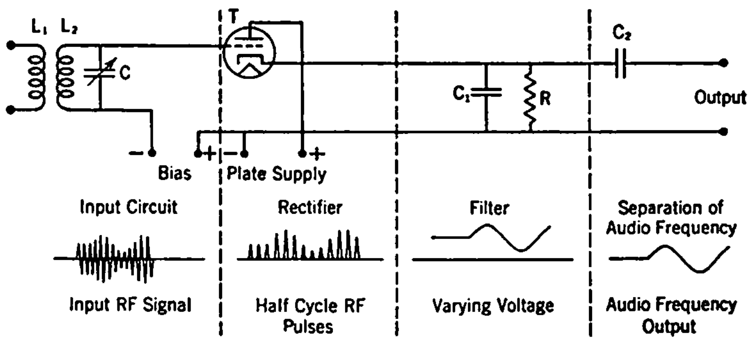


FIG. 11–30. Linear Plate Detector.

introduces a certain amount of distortion, which can be made relatively small as the maximum extent of the characteristic curve is utilized.

SQUARE-LAW PLATE DETECTION. Square-law plate detection has been used in the past because of a lack of good radio-frequency amplifiers in spite of its great distortion. With modern tubes the signal input voltages may easily be built up by radio amplifiers to a value high enough to operate properly a diode detector or a linear plate detector. Figure 11–31 shows the lower end of the characteristic curve of a triode vacuum tube used in square-law detection. This portion is in the vicinity of a in Fig. 11–29. A small signal input voltage is impressed on the grid of the tube. This causes a plate current i_p to flow in the plate circuit. As may be seen from Fig. 11–31, the curvature of the characteristic causes a distortion of the output current i_p. The dotted curve d indicates the average current of the radio-frequency waves. This current varies approximately in accordance with the modulation envelope of the input signal. Owing to the curvature in the characteristic, distortion is introduced and therefore the envelope of d is not a true reproduction of the envelope of the input signal. The output signal pulses i_p may be put through a filter, producing a varying current as shown by d. The audio-frequency component may then be separated by means of a blocking condenser. A circuit diagram would be the same as Fig. 11–30.

GRID-CIRCUIT DETECTION. In the triode detectors described so far, the rectification has taken place in the plate circuit. By the use of the proper circuit, rectification may take place in the grid circuit. An exami-

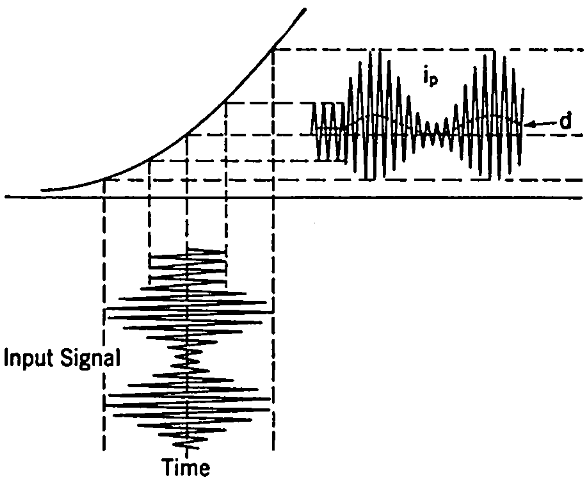


FIG. 11–31. Square-law Plate Detection.

nation of the diagram in Fig. 11–32 will show that the input circuit composed of inductance L_2 and condenser C_1, the filter circuit composed of resistor R and condenser C, and the grid and filament of the tube, together comprise a diode detector circuit such as that shown in Fig. 11–27. There are only two differences. One is that a grid is used in Fig. 11–32, whereas

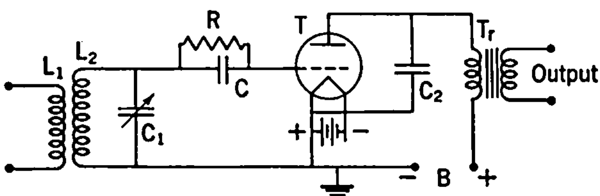


FIG. 11–32. Grid-circuit Detection.

a plate is indicated in Fig. 11–27. The other difference is that an output circuit is not shown across the filter circuit.

In effect, the detector action of rectification takes place in this circuit causing a varying current to flow through R. The D.C. component of the current through R is such that the voltage across R is negative at the grid end and the proper bias voltage is therefore automatically supplied to the vacuum tube T so that it may function as an amplifier. The audio-frequency component of the varying voltage across R is impressed across the grid and filament of the vacuum tube through inductance L_2. The radio-frequency inductance L_2 is small enough so that at audio frequencies it has negligible impedance. The audio-frequency component across the grid and filament is amplified by the vacuum tube, producing an audio frequency that is transformed from the plate circuit to the output terminals through the audio-frequency transformer T_r. Part of the radio-frequency signal voltage appearing across the tuned input circuit is impressed upon the grid of the tube through resistor R and condenser C. This is also amplified by the vacuum tube. The by-pass condenser C_2 is provided to shunt the radio frequency in the output circuit to ground.

Grid-circuit detection is characterized by high sensitivity, high output audio-frequency signal for a given radio-frequency signal voltage input, good fidelity for small input voltages, and increasing distortion as the input voltage is increased.

Grid-circuit detectors have been replaced, to a great extent, by diode and linear plate detectors—for the same reason that square-law plate detectors have been replaced, namely because the signal input voltage may easily be built up in radio amplifiers to a value high enough to operate properly a diode or linear plate detector.

Heterodyne Detection of Continuous Waves. As may be seen from the explanation of the foregoing detector actions, an unmodulated carrier wave that has been detected produces no audio frequency at the output terminals of the detector. There is therefore no audio-frequency current present to actuate a loudspeaker or a pair of earphones. An incoming continuous-wave signal f_s at 1,000 kc is shown in Fig. 11–33a. The signal wave from a local oscillator in the receiver is shown in Fig. 11–33b. The local oscillator frequency f_h is 1,001 kc. The frequency f_h is coupled into the input circuit to the heterodyne detector. The two frequencies, f_s the incoming signal, and f_h the local oscillator signal, combine to form a modulated wave such as that shown in Fig. 11–33c. The modulation envelope varies at a rate that conforms with the difference in frequency between f_s and f_h, that is, 1,000 cycles. Figure 11–33c shows a resultant radio-frequency

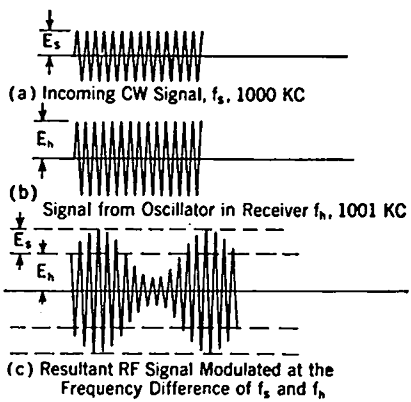


Fig. 11–33. Heterodyne Detection.

signal modulated at the frequency difference between f_s and f_h. The resultant radio-frequency signal may be passed through a detector to give a 1,000-cycle audio frequency that may be used to actuate a loudspeaker or earphones. The audio-frequency output of the detector may be put at any frequency desired by adjusting the local oscillator to a frequency that is different from the incoming continuous-wave signal by an amount equal to the audio frequency desired. Assume that a continuous-wave transmitter is being keyed as a code transmitter. Referring to Fig. 11-34, during the time interval from f to g the key is open at the transmitter and no signal is being transmitted. During this interval the local oscillator in the receiver is producing a signal in the input of the receiver f_h. During this time interval there is no audio frequency in the output of the detector because a continuous-wave signal by itself does not produce an audio frequency. During the time interval from g to h the key at the transmitter has been closed and a continuous-wave signal is emitted. This wave f_s combines with f_h in the receiver input circuit to produce a signal modulated at an audio frequency that is the difference between f_s and f_h. By detector action, an audio frequency is produced in the output terminals of the detector during the time interval g to h. At h the key is again opened and therefore only the signal f_h is on the input of the detector. There is therefore no audio-frequency signal in the output of the detector. By properly keying the transmitter in code groups of dots and dashes it is possible to transmit code signals.

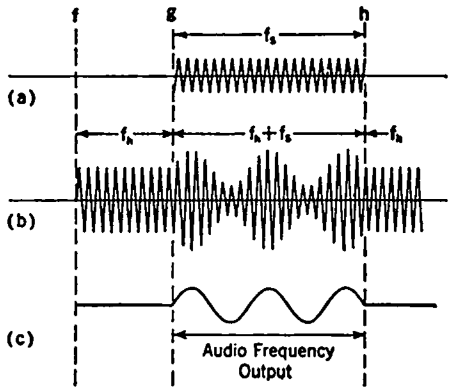


FIG. 11-34. Detection of Keyed Transmitted Continuous Wave.

Detection of Interrupted Continuous-wave Signals. Interrupted continuous-wave signals may be put through any one of the various types of detectors, thereby producing an audio output signal that is proportional to the envelope of the interrupted continuous-wave signal. Straight detection and not heterodyne detection is required for interrupted continuous-wave signals because the incoming wave is modulated. It should be remembered that a continuous-wave carrier amplitude modulated by speech or music audio frequencies is detected by straight detection and not heterodyne detection, because the signal input to the detector is a modulated wave. Only in the case of continuous-wave code transmission is a heterodyne detector required.

Review Questions and Problems. *1.* What is the principal difference between audio-frequency and radio-frequency amplifiers?

2. Why do tuned circuits reduce harmonics?

3. Why is it that resonant-circuit coupling cannot be used for audio amplifiers?

4. Describe three methods of supplying grid-bias voltage to a radio amplifier.

5. Name the three classes of radio amplifiers and describe briefly their characteristics.

6. Why is the operation of a Class B audio amplifier similar to that of a Class B radio amplifier?

7. (a) How is a single-ended radio amplifier neutralized? (b) How is a push-pull radio amplifier neutralized?

8. What are the advantages of pentode radio amplifiers?

9. In cascaded radio amplifiers, why is it desirable to filter the supply leads from the grid-bias supply and the plate-voltage supply?

10. Why is shielding employed in radio amplifiers?

11. Why is it not necessary to neutralize tetrode and pentode radio amplifiers?

12. (a) What trouble is likely to be encountered in operating tubes in parallel? (b) How is this trouble overcome?

13. (a) What type of radio amplifier is used in the plate modulated stage? (b) Why?

14. What is detection or demodulation?

15. Name the three steps of detection and explain.

16. Name three types of detectors and explain their characteristics.

17. How are continuous-wave telegraph signals received?

CHAPTER 12

Amplitude-modulation Radio Transmitters

The part of a radio communication system called the radio transmitter produces radio-frequency energy for the purpose of exciting an antenna, causing electromagnetic radiation. Fundamentally, an amplitude-modulated radio transmitter consists of means for generating radio-frequency energy and means for controlling or modulating the radio-frequency energy by means of audio frequencies.

A transmitter divides into two principal sections. These are the *radio-frequency section* and the *audio-frequency section.* A block diagram of an amplitude-modulation radio transmitter is shown in Fig. 12–1. The audio- and radio-frequency sections are individually blocked out by dotted lines.

Radio-frequency Section. The frequency of a transmitter must be controlled very carefully so that transmission will occur only on the assigned channel. If the transmitting frequency is allowed to vary, the result may be interference with services on other channels. Vacuum-tube oscillators, such as those described in Chapter 8, may be used in transmitters. Crystal oscillators, however, are almost universally used for maintaining transmitters on their assigned frequencies. Crystal oscillators will maintain frequency within very narrow limits. Crystal-oscillator output power, however, is small. It is therefore necessary to amplify the small radio-frequency energy produced by the crystal oscillator up to the output power of the transmitter. The number of radio amplifier stages necessary is dependent mainly on the transmitter output power. It is customary, however, to have at least one stage of radio amplification between the crystal oscillator and the modulated amplifier. Such an amplifier is called a *buffer amplifier.*

During the process of modulation the power output of the transmitter is varied. The variation in modulated amplifier output power may react on the previous radio amplifier or oscillator. In order to maintain frequency properly, the crystal oscillator is usually isolated from the reaction that takes place in the modulated amplifier by a buffer amplifier whose function will be explained later. The particular transmitter shown in Fig. 12–1 has two radio amplifiers between the crystal oscillator and the modulated amplifier. In some low-powered transmitters no buffer amplifiers are used. This practice is followed where high-precision frequency control is not necessary and where fidelity of transmission is not important.

The intermediate power amplifier, preceding the modulated amplifier, must supply enough power for the input to the Class C modulated amplifier. The last radio amplifier shown in Fig. 12–1 is labeled *modulated power amplifier.* Modulation takes place in this amplifier and it is also the output amplifier that supplies the radio-frequency power to the output terminals of the transmitter. The intermediate amplifiers between the crystal

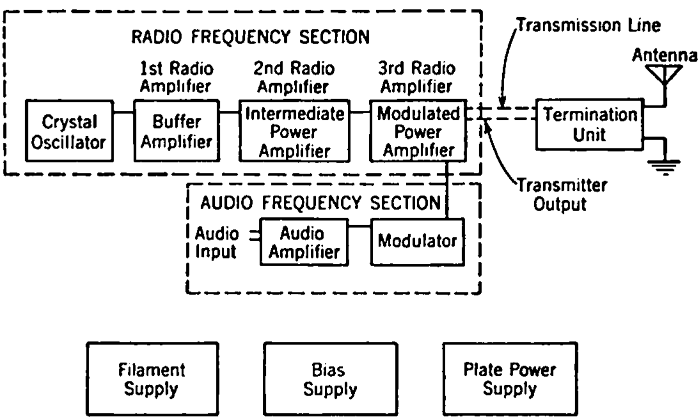


FIG. 12–1. Block Diagram of Crystal-controlled Amplitude-modulated Transmitter (high-level modulation).

oscillator and the modulated power amplifier may be either Class A, Class B, or Class C. The amplifier preceding the modulated amplifier is usually a Class C amplifier.

Audio-frequency Section. Two components are shown in the audio-frequency section box of Fig. 12–1. These are the *audio amplifier* and the *modulator.* The modulator supplies the audio-frequency power to amplitude-modulate the radio-frequency power amplifier. An audio-frequency amplifier is usually provided in a transmitter ahead of the modulator so that the audio-frequency input power to the transmitter need not be higher than approximately the power required to feed a program over a telephone line. Three other important components of a transmitter are the filament, bias, and plate power supplies. These three are shown as blocks in Fig. 12–1.

Transmitters are classified according to whether they are ***high-level modulated*** or *low-level modulated.* The transmitter shown in Fig. 12–1 is a high-level modulated transmitter because modulation takes place in the output amplifier at the highest level. Some transmitters are modulated in one of the radio amplifiers preceding to the output amplifier. These are called low-level modulated transmitters. The amplifiers after modulation must be linear, and therefore they are either Class B radio amplifiers or one of several types of so-called "high-efficiency" linear amplifiers.

The block diagram of a low-level amplitude-modulated transmitter is shown in Fig. 12–2. Modulation takes place in the third radio-frequency amplifier. The modulated radio-frequency output of the third amplifier is raised to the transmitter output power by the fourth linear radio ampli-

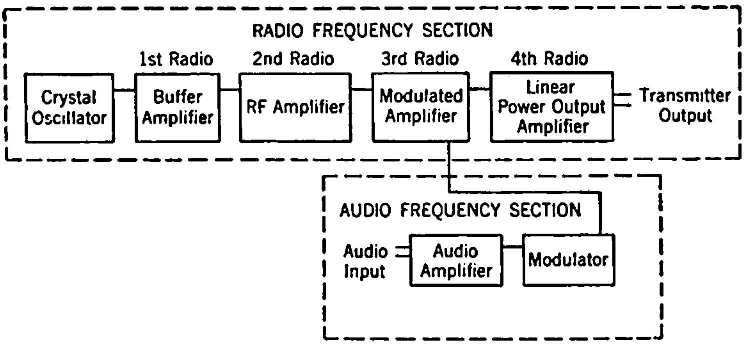


FIG. 12–2. Block Diagram of Crystal-controlled Amplitude-modulated Transmitter (low-level modulation).

fier. In some transmitters there may be two or more linear amplifiers after the modulated amplifier, but modulation is usually accomplished either in or next to the output radio-frequency amplifier.

Over-all Transmitter Efficiency. The over-all power efficiency of a transmitter is important. As distinguished from the power output efficiency of an amplifier, this efficiency is the radio-frequency carrier output power of the transmitter divided by the total power input to the transmitter, multiplied by 100 to express it in per cent. A 50-kw transmitter with an over-all power efficiency of 20% draws 250 kw from the power mains. If the over-all efficiency is 50% the total power drawn from the mains is 100 kw.

A low-level modulated 50 kw radiotelephone transmitter, with a Class B radio amplifier output, requires a power input of about 250 kw. As was explained previously, the radio amplifiers after the modulated stage must be linear. A linear Class B radio amplifier usually operates at an efficiency of about 33%. In order to produce a 50-kw carrier at the output of the transmitter, the power input to the plate of the last amplifier would accordingly be 150 kw. This difference alone accounts for 60% of the total power input to the transmitter. Transmitters such as this are still in use. However, they are rapidly being supplanted by transmitters with much higher efficiency. The two types of transmitters with high over-all efficiency are the *low-level high-efficiency linear amplifier transmitter* and *high-level modulated transmitter.*

LOW-LEVEL HIGH-EFFICIENCY LINEAR AMPLIFIER TRANSMITTERS. There are several types of high-efficiency linear amplifiers. One in common use is the Doherty amplifier. The Doherty amplifier takes the place of the Class B radio amplifier in the output stage of a low-level modulated transmitter. The efficiency of a Doherty amplifier under unmodulated carrier conditions is 60%–65%. The plate input power to the output amplifier of the transmitter is 78 kw as compared with the 150 kw required by a linear Class B radio amplifier. As a result of this, and of increases in efficiency in other components of the transmitter, a low-level modulated Doherty amplifier transmitter will produce an output carrier power of 50 kw with a total over-all transmitter efficiency of about 35%.

HIGH-LEVEL MODULATED TRANSMITTER. In a high-level modulated transmitter, modulation takes place in the output amplifier. This amplifier is therefore a Class C radio amplifier and operates at a high efficiency. Such an amplifier may operate at an efficiency of 80%. For a transmitter carrier power of 50 kw, the input power to the plate of the last amplifier would be 62.5 kw. The total over-all efficiency of a typical 50-kw high-level modulated transmitter is slightly better than the high-efficiency linear amplifier transmitter. As may be seen from these figures, the high-level modulated transmitter has the greatest efficiency. This efficiency has been made possible by the development of the high-power-output Class B audio-frequency amplifier that can be used as a modulator. To modulate the high-level modulation transmitter just described, 31.2 kw of audio-frequency power is required. Large amounts of audio-frequency power have been made available for modulators by the development of the highly efficient Class B audio amplifier.

The transmitter proper ends at the terminals marked *transmitter output* in Fig. 12–1. Transmitting antennas are usually placed at a distance of several hundred feet from the transmitter. The radio-frequency energy output of the transmitter is conducted from the transmitter to the antenna by a *radio-frequency transmission line.* The energy is connected from the transmission line to the antenna by means of a *termination unit.* Radio-frequency transmission lines and termination units are discussed in Chapter 16.

Crystal Oscillators. It was shown in Chapter 6 that a conversion from sound to electrical energy is accomplished by a microphone. It also was shown that a conversion from electrical to sound energy is accomplished by an earphone or loudspeaker. In each case there is an intermediate transfer to mechanical energy.

Certain crystalline substances also possess the property of converting energy from electrical to mechanical form and vice versa. When such a crystal is mechanically stressed, an electric field appears between its faces. Conversely when a voltage is applied to electrodes on two parallel faces of the crystal a mechanical displacement occurs. This behavior is called the piezo-electric effect. Among substances having this property are quartz crystal.

A piezo-crystal that vibrates naturally at a certain frequency is the electrical equivalent of an *LC* circuit tuned to that frequency. Such a quartz crystal may be substituted for the input circuit to a vacuum tube, thus producing an output oscillation at a rate determined mainly by its thickness. A piezo-electric quartz crystal, about the thickness of a half dollar, vibrates mechanically at a rate of about 1,000,000 times per second. The thickness of the crystal will vary with the temperature. Because frequency varies with thickness, the temperature of the crystal is maintained constant by a heat control, usually within a range of about 1° F or less.

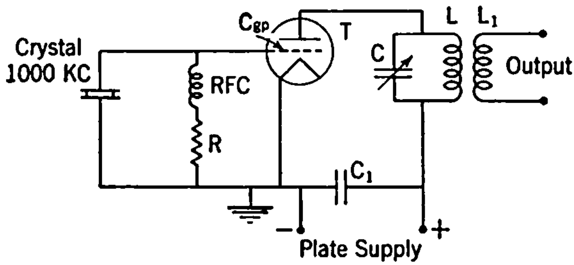


FIG. 12-3. Simple Crystal-control Circuit—Triode Tube.

The precision of heat control necessary is determined by the type of crystal and the type of service the transmitter is to perform.

A simple crystal control circuit is shown in Fig. 12-3. In Fig. 12-3, assume that the crystal has a natural vibration of 1,000 kc and that the output circuit of the vacuum tube T, composed of condenser *C* and the inductance *L*, is tuned to 1,000 kc. The grid and the plate circuits are both effectively tuned to the same frequency and therefore the feedback through the plate-to-grid capacity C_{pg} causes oscillation. Inasmuch as the crystal vibrates at a natural frequency of 1,000 kc, the output radio-frequency power will also be at 1,000 kc. Crystal oscillator circuits may take many different forms. However, Fig. 12-3 is typical of a crystal triode oscillator. In the operation of a crystal oscillator a grid current will flow through the radio-frequency choke RFC and the grid-leak resistor *R* of Fig. 12-3. There is also a radio-frequency voltage across the crystal and a radio-frequency current is caused to flow through the crystal. If the current is too high the crystal is liable to crack, and therefore the power output of a crystal oscillator tube must be limited. The power output, using triode tubes, is usually not more than 5 watts.

Tetrode and pentode oscillators place a lower strain on the crystal for a given power output than do triode tubes. It is also possible to get a little higher output from these tubes. Figure 12–4 shows a crystal control circuit using a tetrode or screen-grid tube. The crystal is connected between the control grid G_1 and the ground. Across the crystal are placed a radio-frequency choke RFC and the resistor R in series. This circuit supplies part of the bias voltage for the tube. The cathode resistor R_1 also supplies part of the bias voltage. G_2 is the screen grid. The output circuit of the vacuum tube is composed of condenser C_4 and inductance L. The plate-to-grid capacity of a tetrode tube is very small and in some cases is not large enough to cause oscillation. Then it is necessary to aid feedback by adding a small capacity between the plate and the control grid such as C_5 in Fig. 12–4.

PENTODE-TUBE CRYSTAL OSCILLATOR CIRCUIT. Figure 12–4 would show the diagram of a crystal control circuit using a pentode tube if a suppressor

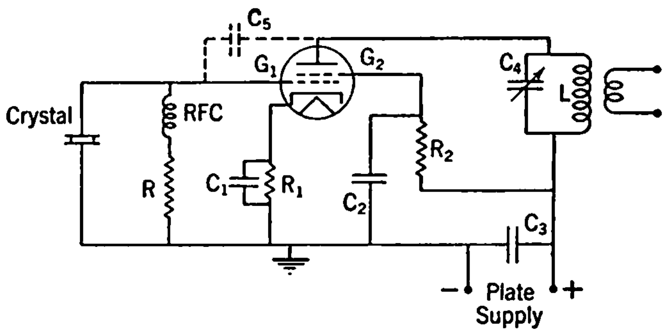


FIG. 12–4. Crystal-controlled Circuit—Tetrode Tube.

grid were added between the screen grid G_2 and the plate and if it were connected to the cathode. Tetrode and pentode crystal control circuits function much the same except that the extra capacity C_5, as shown in Fig. 12–4, is more often required with pentode than it is with tetrode tubes. There are a great many different crystal oscillator circuits. A circuit is shown in Fig. 12–13 in which the crystal is connected between the control grid and the screen grid of a pentode.

The thickness of the crystal is the main condition that determines the frequency of a crystal oscillator. Other conditions affect frequency to a smaller degree, such as the capacity and inductances associated with the crystal oscillator tube. The filament, plate, and bias voltages also may slightly affect the frequency of the oscillation. Where high precision of frequency control is required, the tuning circuit, such as C and L in Fig. 12–3, is also maintained at a constant temperature.

Buffer Amplifiers. In order to maintain the frequency constant, the crystal oscillator must operate under constant conditions. In order that the crystal oscillator may function independently to produce a constant frequency it is customary to insert at least one radio amplifier between the crystal oscillator and the modulated amplifier. Amplifiers used for this purpose are called *buffer amplifiers.*

In order to prevent reaction and undesirable feedbacks to the crystal oscillator, buffer amplifiers are often carefully shielded, as are the crystal oscillator circuits. If a triode is used for the buffer, it is often desirable to have it neutralized as well as possible to prevent feedback from its plate circuit to its grid circuit and thence to the crystal-oscillator tube. Because of their low feedback qualities, as described in Chapter 11, pentodes are ideal tubes for use as buffer amplifiers. They also have the added advantage of greater amplification.

Modulated Amplifiers. There are many different methods for accomplishing amplitude modulation. The methods divide into two general classifications, *grid modulation* and *plate modulation.* An amplifier using plate modulation is much more efficient than one using regular grid modulation and therefore this type is used to a greater extent.

PLATE-MODULATED AMPLIFIER. The schematic of a plate-modulated radio amplifier is shown in Fig. 12–5. Radio amplifier tube T is operated as a Class C radio amplifier. The output efficiency of a Class C radio amplifier under unmodulated carrier conditions may be as high as 80%. The plate voltage E is supplied to the vacuum tube T through the audio-frequency choke CH and the radio-frequency choke RFC_1. Modulation takes place by varying the plate supply voltage. At 100% modulation the variation is between zero and $2E$. The vacuum tubes T_2 and T_3 are connected as a Class B audio-frequency amplifier. They are the modulators. The power output of the modulator is impressed across points a and b through the audio-frequency transformer T_{r2} and condenser C_4. Condenser C_4 prevents the D.C. supply voltage from being impressed on the secondary of the transformer T_{r2}. If the audio-frequency choke CH were not used, the audio-frequency output from the modulator would be shorted through the plate supply. The audio-frequency voltage from the modulator is effectively placed in parallel with the D.C. voltage from the plate supply. During the positive half cycle of audio frequency it adds to the D.C. plate voltage and during the negative half cycle of audio frequency it subtracts. The plate voltage to the amplifier tube T is then varied. This variation varies the output power of the modulated amplifier, thus producing a modulated wave.

The variation of power output with variation in plate voltage is shown in Fig. 12–6. Figure 12–6a shows the radio-frequency output of a radio-frequency amplifier. During the time interval from f to g the carrier power only is being supplied, and therefore the cycles of radio-frequency voltage are all the same. During this time interval there is no input to the modulator and therefore no audio-frequency voltage is produced in the output of the modulator. The audio-frequency voltage output of the modulator is shown in Fig. 12–6b. The plate-voltage supply to the modulated amplifier is shown in Fig. 12–6c. The D.C. plate-voltage supply e_b is constant during the time interval f to g. During the time interval g to h the audio-frequency voltage e_a, as produced by the modulator, is shown in Fig. 12–6b. The audio-frequency voltage adds and subtracts from the D.C. plate voltage and a varying voltage results. This resultant is shown

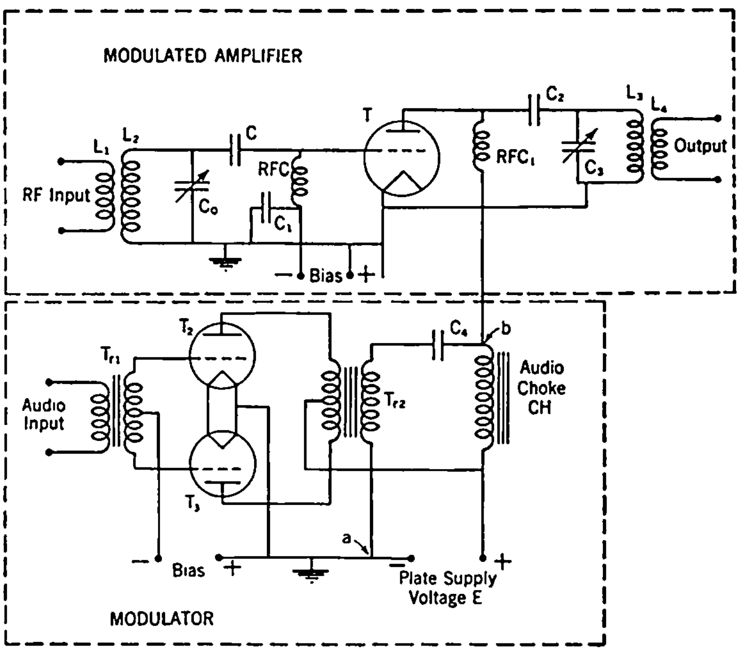


Fig. 12–5. Plate Modulation.

in Fig. 12–6c for the time interval between g and h. The varying plate voltage causes the output power of the modulated amplifier to vary. Figure 12–6a during time interval g to h shows the variation in the voltage across the radio-frequency output circuit L_3, C_3 of Fig. 12–5. Figure 12–6a shows that during the time interval g to h the radio-frequency output voltage is varying or modulated at the rate of the audio-frequency voltage output of the modulator. Referring to Fig. 12–6, at 100% modulation e_a, the audio-frequency voltage, is equal to e_b, the D.C. plate voltage. The varying plate voltage, as shown in Fig. 12–6c, varies between zero and twice e_b. The radio-frequency output voltage shown in Fig. 12–6a also varies between zero and twice e_{RF}, the voltage under no modulation. Inasmuch as power varies as the square of the voltage, the *peak* output power of the amplifier at 100% modulation will be four times the carrier power. It is therefore necessary that the tube or tubes used in a modulated amplifier be able to put out a peak power four times that of the unmodulated carrier.

GRID-MODULATED AMPLIFIER. In grid modulation the function is performed in the grid circuit of the tube. A grid-modulated amplifier operates

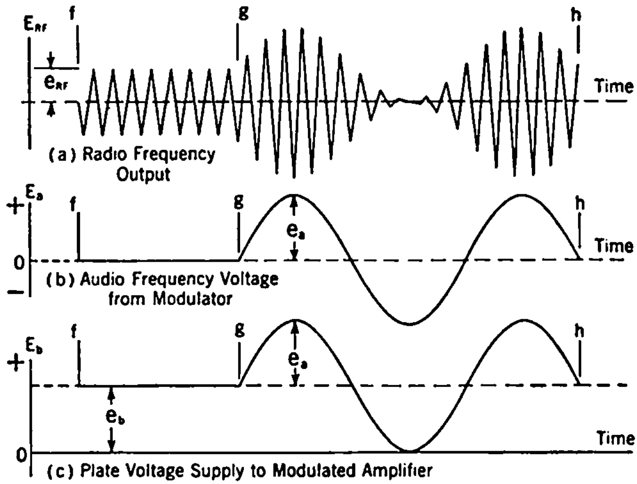


FIG. 12–6. Plate Modulation: Variation of Power Output with Variation in Plate Voltage.

at an efficiency that is below 35% and usually about 25%. It is usually only used in low-power stages. The fundamental circuit of a grid-modulated amplifier is shown in Fig. 12–7. A grid-modulated amplifier is biased as a Class C amplifier. The efficiency is low, however, because only a limited portion of the characteristic curve can be used for the carrier condition. The circuit diagram of the grid-modulated amplifier shown in Fig. 12–7 is conventional except that the secondary of an audio-frequency transformer T_r has been inserted in series with the grid-bias supply lead. The audio-frequency voltage E_a appearing across the secondary of the transformer T_r is effectively placed in series with the D.C. voltage of the bias supply. Therefore the audio frequency varies the total bias voltage on the grid of the tube.

The grid-voltage–plate-current relation in a grid-modulated amplifier is shown in Fig. 12–8. The grid-bias voltage E_c is at least twice that required for cutoff and therefore the amplifier is biased as a Class C amplifier. As explained before, however, the efficiency is low. During the unmodulated carrier condition, as shown during the time interval f to g, the radio-frequency grid voltage input e_c causes part-cycle pulses of plate current i_p. During time interval g to h, the bias voltage is varied by an audio-frequency voltage impressed in series with the bias-voltage supply.

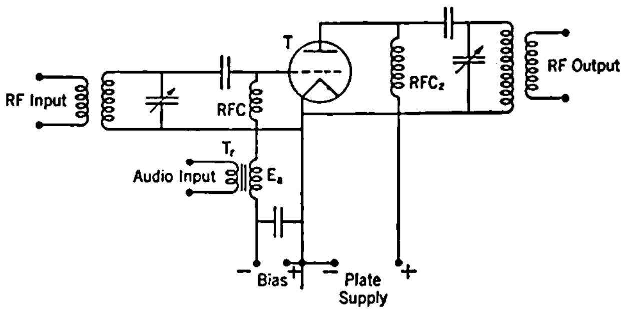


Fig. 12–7. Grid-modulation Circuit.

This variation causes the radio-frequency voltage input to the grid of the tube to vary in accordance with the audio-frequency voltage as shown over time interval g to h. This variation in turn causes part-cycle pulses of plate current to flow, which have an envelope in accordance with the

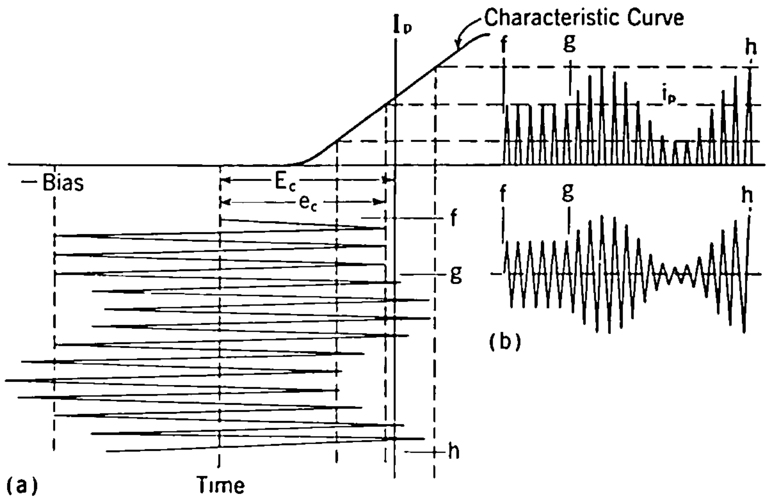


Fig. 12–8. Grid-voltage–Plate-current Relation in Grid Modulation.

audio frequency. As explained in Chapter 11, the plate current pulses i_p produce full sine waves of radio frequency in the output circuit of the amplifier. This effect is shown in Fig. 12–8b.

Modulation of screen-grid tubes. Screen-grid tubes of the pentode type may be used as plate-modulated radio-frequency Class C amplifiers. The circuit diagram of a modulated screen-grid tube is shown in Fig. 12–9. In order properly to plate-modulate a screen-grid tube, it is necessary to modulate the screen-grid voltage also. The screen grid G_2 is connected to the plate-supply circuit through its dropping resistor R_2. The plate-supply voltage is obtained through the secondary of the modulator output audio transformer T and therefore the screen grid and the plate voltage are modulated simultaneously. In this type of modulation

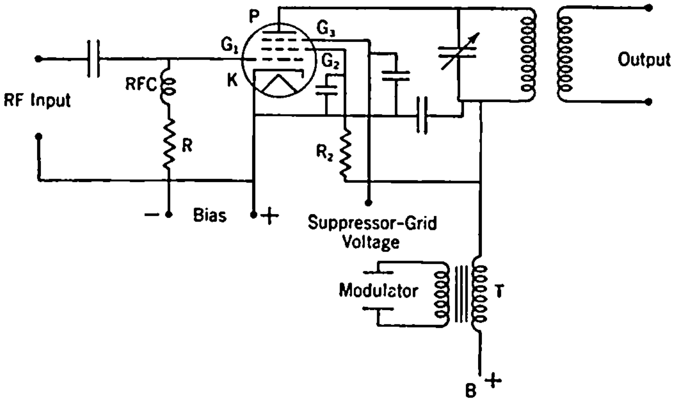


FIG. 12–9. Modulation of Screen-grid Tube.

the audio-frequency power required for 100% modulation is one half the input power to the plate and the screen grid.

This type of modulation may be used where the highest fidelity is not required.

SUPPRESSOR-GRID MODULATION. A screen-grid tube of the pentode type may be modulated in the suppressor-grid circuit. The circuit diagram of such a modulated amplifier is shown in Fig. 12–10. The amplifier hookup is a conventional pentode amplifier hookup except that the secondary of the modulator output transformer T_r is inserted in series with the lead to the suppressor grid G_3. The suppressor-grid-modulated amplifier operates under conditions similar to those for grid-bias modulation, and therefore the plate efficiency of suppressor-grid modulation is low.

Modulators. It has previously been explained that amplifiers after the modulated stage must be linear. They must, therefore, either be Class B radio amplifiers operating at a low efficiency or linear radio amplifiers such as the Doherty amplifier operating at a relatively high efficiency. The use of the Class B radio amplifier is rapidly being restricted because of its low efficiency. The Doherty type of amplifier employs a more complicated radio-frequency circuit and requires a plate voltage approximately 50% higher than a high-level modulated amplifier requires. The high level, however, requires a large audio-frequency power for modulation. The plate-modulated amplifier is in more general use than are the Doherty types of amplifiers.

As a consequence of the trend towards high-level plate-modulated amplifiers, the Class B audio-frequency amplifier, which can produce a high audio-frequency power output, has assumed major importance.

In plate modulation the modulator, at 100% modulation, must supply an audio-frequency power that is 50% of the plate input power to the modulated amplifier. The D.C. plate-supply input power to a 5,000-watt

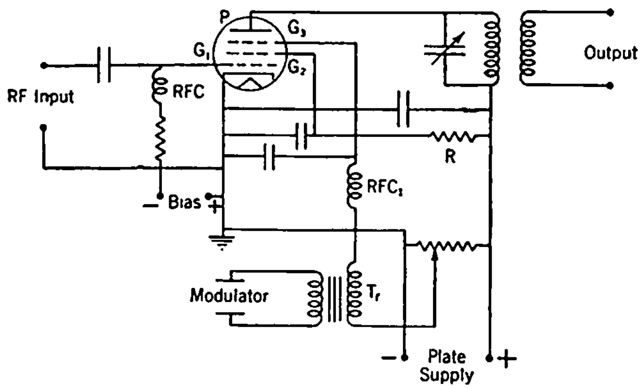


FIG. 12-10. Suppressor-grid Modulation.

modulated amplifier operating at 80% efficiency is 6,250 watts. For 100% modulation the modulator must supply an audio power of 3,125 watts.

Regulation of Plate Power Supply for Class B Radio Amplifiers. Figure 12-11a shows a vacuum tube T operating as a linear Class B radio amplifier. L is a filter choke in the plate-supply rectifier filter. Condenser C is a filter condenser of the plate-supply rectifier filter. Figure 12-11b shows the current drawn by the plate of the tube. During time interval f to g the carrier is unmodulated and therefore the plate current has an amplitude of i_{b1}. During the time interval g to h the carrier wave is modulated with an audio frequency. The plate current is a varying current during this time interval. During the positive half cycles of modulation the plate current is i_{b2} and during the negative peaks of modulation the plate current is i_{b3}. During modulation there is a large variation in plate current. At the same time the plate voltage must remain constant. The filter condenser C is constantly being charged by the voltage output of the rectifier. During the positive half cycles of modulation the plate of the vacuum tube draws additional current. During this period the condenser C discharges, furnishing the additional current for the positive peaks of modulation. The filter inductance L prevents this current from coming from the rectifier because its inductance offers a high impedance to the passage of a current varying at an audio frequency. The condenser C, on the other hand, has a very low impedance and therefore

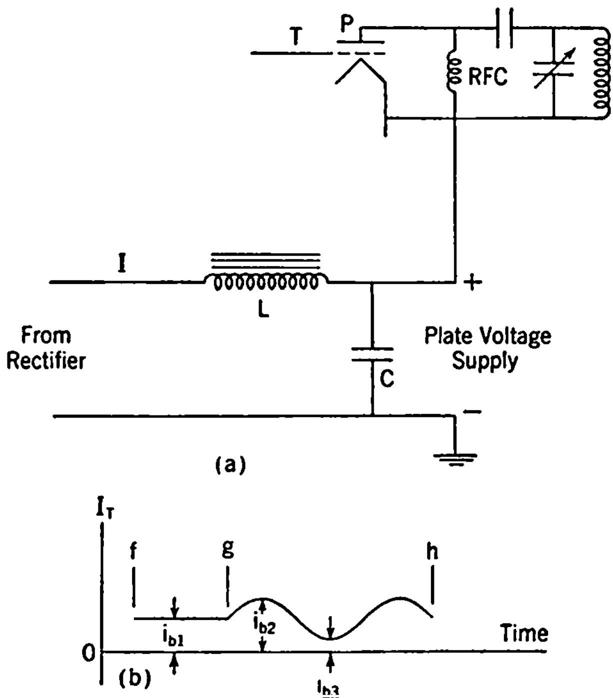


FIG. 12-11. Class B Radio Amplifier Plate-supply Regulation.

it readily gives up its charge to supply current for the positive peaks of modulation. During the negative peaks of modulation the grid of the vacuum tube is made more negative and therefore the flow of plate current is reduced. During this time interval the condenser C regains the charge it lost during the positive half cycle of modulation. The condenser C and the inductance L act as a filter so that the audio-frequency variations are not imposed upon the rectifier. The output current I of the rectifier, therefore, remains substantially constant. The inductance L and the condenser C do double duty because they are also serving to iron out the A.C. pulses from the rectifier tubes.

The term *regulation* was explained in Chapter 5. This is a measure of the variation of the voltage output from the rectifier filter between no load and full load. As current is drawn from the filter of a rectifier, the output voltage is reduced owing to the D.C. resistance of the filter chokes and also owing to the D.C. resistance of the power transformer. In Class B linear amplifiers, the current drawn from the rectifier is substantially constant and therefore the plate-supply regulation need not be very exact. However, for reasons of efficiency, the filter chokes and the plate power-supply transformer are built with rather low resistance to avoid losses due to heating in the windings.

Regulation of Plate Power Supply for Class B Audio Amplifiers. Figure 12–12 shows the characteristic curves of two tubes used as a Class B au-

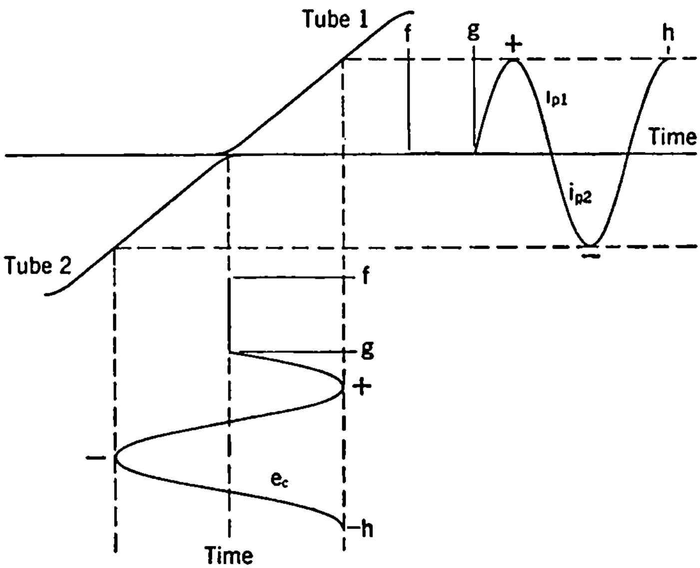


FIG. 12–12. Regulation of Plate Supply for Class B Audio Amplifier.

dio amplifier. As explained in Chapter 11, the tubes are biased at approximately the cutoff point. With no input to the Class B audio amplifier, the plate current drawn by the amplifier is approximately zero. In Fig. 12–12, during the time interval f to g, there is no audio-frequency input to the grids and therefore during the corresponding period the plate current i_p is small. During the time interval g to h an audio-frequency input has been impressed upon the grids of the tubes. During the positive half cycle of audio-frequency input, tube 1 draws plate current I_{p1}. During the other half cycle of audio-frequency input, tube 2 draws plate current i_{p2}. The plate current drawn by the two tubes when amplifying audio frequency varies over a large range. The plate-supply voltage to the amplifiers must remain constant for linear reproduction. Therefore it is necessary to employ a plate supply with good regulation. This requirement means that the filter choke and the power transformer must be made with relatively low D.C. resistance. Also, a large condenser must be used in the output of the filter. Also, the audio-frequency output transformer of the amplifier must have a relatively low D.C. resistance. The plate current passes through the primary of the output audio transformer and therefore the D.C. drop must be kept low.

Regulation of Plate Power Supply for Modulated Class C Radio Amplifiers. In modulating a Class C amplifier, the plate voltage is varied and therefore the plate current varies. This variation is added to the one caused by the Class B modulator and therefore plate-supply regulation must be good.

Regulation of Bias Supplies. Amplifiers that operate without drawing grid current need not have bias supplies with very good regulation because there is no change in the current drawn from the bias supply. Such applications are linear Class A radio amplifiers and linear Class A audio amplifiers where the signal input voltage does not exceed the bias voltage. When an amplifier draws a steady or substantially steady grid current, there is no need for good regulation in the bias supply. The radio-frequency amplifier stages ahead of the modulated amplifier in a radiotelephone transmitter and some of the amplifiers in an unmodulated radiotelegraph transmitter are examples. The factors that affect regulation in plate supplies also are applicable to bias supplies.

Schematic Circuit of a Transmitter. The schematic circuit diagram of a 1,000-watt high-level modulation radiotelephone transmitter is shown in Fig. 12–13. The diagram is complete except that it does not show the primary power circuits nor the control circuits such as overload relays. These circuits have been omitted for clarity. Primary power circuits and overload relays will be discussed separately later in the chapter.

The circuit diagram of Fig. 12–13 has been divided into sections by dotted lines. The divisions are: *radio-frequency section*, *audio-frequency section*, *second audio bias-supply rectifier*, *low-power plate-supply rectifier*, *modulated bias rectifier*, and *high-voltage plate-supply rectifier*. The radio-frequency section shows the circuits associated with the production of the radio-frequency wave and its amplification up to the carrier output power. The audio-frequency section includes the audio-frequency amplifiers used for the purpose of producing the audio-frequency power for modulation. The four rectifiers produce the bias supplies and the plate-voltage supplies for the various amplifiers. The various sections of the transmitter will be discussed individually.

Radio-frequency Section. The radio-frequency section is composed of a crystal oscillator and three successive stages of radio amplification. The successive stages of radio amplification are labeled the buffer amplifier, the intermediate power amplifier, and the power amplifier.

OSCILLATOR. The first part of the radio-frequency section of a transmitter is the oscillator. The frequency of the oscillator shown in Fig. 12–13 is controlled by the crystal X. In order to maintain the frequency constant the crystal is placed in a *crystal heat chamber*. The crystal heat chamber is insulated so that a constant temperature may be maintained inside. The resistor R shown in Fig. 12–13 is a heating element employed to keep the temperature of the chamber well above normal room temperatures that may be encountered. The selected temperature is usually on the order of 140° F. A thermostat S_1 is connected in series with the resistor R. As the chamber temperature reduces, the thermostat contacts close, thus allowing current to pass through the heater R. As soon as the temperature in the chamber has increased to the proper point the thermostat contacts again open, stopping the flow of current through R. The thermostat S_1 therefore functions to maintain a constant temperature inside the crystal heat chamber. The condenser C is placed across the contacts of the thermostat to avoid sparking and arcs. Sparking and arcs injure the thermostat contacts and must therefore be avoided.

In Fig. 12–13, the crystal oscillator tube VT_1 is a pentode. The crystal X is connected between the control grid and the screen grid of the tube. The screen-grid voltage is supplied through resistor R_3 from the voltage divider R_5. The source of the voltage across R_5 will be described later. The control-grid bias is obtained from two sources. Part of the bias is obtained from the cathode resistor R_2 and part is obtained from the grid leak R_1. The frequency of a crystal is affected, to a certain degree, by a capacity across it. The variable condenser C_1 is connected across the crystal for the purpose of making minor frequency adjustments. It is, therefore, not necessary to grind the crystal to exactly the proper thickness, inasmuch as C_1 provides a certain amount of adjustment. The output circuit of the oscillator is composed of condenser C_8 and inductance L_1. The plate voltage is obtained through the radio-frequency choke RFC_1. The entire oscillator is contained in a shield so that the operation of the oscillator will be as independent as possible. It is also to be noted that the power-supply circuits are all carefully filtered. C_4, C_7 and R_3 form the filter for the screen grid. C_5 and R_4 form the filter for the suppressor grid. C_6 and RFC_1 form a filter for the plate supply. T_1 represents the secondary of the transformer that heats the filament of the vacuum tube VT_1. For diagram simplicity, the primary power supply to the various filament transformers is not shown. These are T_1, T_2, T_3, T_4, T_6, T_9, T_{11}, T_{12}, T_{13}, T_{15}, T_{16}, T_{18}, and T_{19}. The ammeter A_1 in series with the cathode resistor R_2 measures the plate current, the screen-grid current, and the suppressor-grid current of VT_1. The radio-frequency voltage output of the oscillator is fed to the grid of the buffer amplifier through the condenser C_9.

BUFFER AMPLIFIER. The buffer amplifier functions to prevent reaction on the crystal oscillator from the succeeding radio amplifiers. The vacuum tube VT_2 is a pentode. The self-neutralizing qualities of the pentode tube prevent anything that happens in the amplifiers subsequent to the buffer amplifier from reacting back on the oscillator through the tube itself. The resistor R_7 functions to damp out any high-frequency parasitic oscillations. The control-grid bias for the buffer amplifier is obtained from the voltage drop across the grid-leak resistor R_6. An ammeter A_2 is connected between ground and the midtap of the secondary of the filament transformer. This ammeter measures the current drawn by the plate, suppressor grid, and screen grid of VT_2. The output circuit composed of C_{13} and L_2 is the same type as the one employed in the output of the oscillator. The output of the buffer amplifier is impressed upon the grid of the intermediate power amplifier through condenser C_{14}. It is to be noted throughout the transmitter diagram that *grounds* are frequently shown. The output circuit of the buffer amplifier is connected between plate and ground through the radio-frequency by-pass condenser C_{12}. This voltage is then effectively impressed across the grid and filament of VT_3.

INTERMEDIATE POWER AMPLIFIER. Vacuum tube VT_3 functions as the second radio amplifier. R_8 is a parasitic-damping resistor. The bias for VT_3 is supplied by the cathode resistor R_{10} and the grid leak R_9. When the radio-frequency power becomes appreciable it is necessary to keep it out of the grid leak. The radio-frequency choke RFC_3 performs this function. The resistance R_9 is comparatively small and therefore would place a heavy load on the output circuit of the previous amplifier. This load would lower the radio-frequency voltage on the grid of VT_3. There is a considerable amount of *distributed capacity* in many types of resistors. This capacity acts the same as if a condenser were placed across the resistor, and forms a low-impedance by-pass for the radio frequency. This additional factor would reduce the voltage from grid to plate if the radio-frequency choke RFC_3 were not used. The distributed capacity also would cause radio-frequency currents to circulate in the resistor. Such currents sometimes cause the grid leak to burn out. It is therefore usual practice to employ a radio-frequency choke in series with the grid leak when the radio-frequency power becomes appreciable.

The plate voltage for VT_3 is parallel fed through RFC_4. The radio-frequency output of the intermediate amplifier is fed to the output circuit composed of C_{18} and L_3 through the radio-frequency blocking condenser C_{17}. When the radio-frequency power in a circuit becomes appreciable it is usually customary to provide a meter for measuring the current at some place in the circuit. Radio-frequency ammeter A_3 in series with C_{18} measures the current in the output circuit of the intermediate power amplifier. The output of the intermediate power amplifier is fed through the radio-frequency transformer composed of L_3 and L_4 to the grids of the power amplifier. The midtap of L_4 is grounded. The grids in the power amplifier are then fed equal radio-frequency input voltages.

POWER AMPLIFIER. The output power amplifier of the transmitter employs two tubes VT_4 and VT_5 in a push-pull circuit. The power amplifier is plate modulated and therefore the transmitter is *high-level modulated.* The operation of the amplifier is Class C. The control-bias voltage is obtained from a combination of grid-leak bias and cathode bias. R_{11} and R_{12} are the grid-leak resistors and R_{13} and R_{14} are the cathode resistors. The ammeters A_4 and A_5 measure the direct current flowing through the grid leaks; A_6 and A_7 in series with each cathode resistor measure the plate current drawn by each tube. Ammeter A_8 measures the total plate current drawn by the amplifier.

When the power of an amplifier gets to be on the order of several hundred watts or more, a combination of inductance and resistor in a choke circuit is used for suppressing parasitic oscillations. Resistor R_{15} and inductance L_5 provide the parasitic-suppression circuit. If a plain resistance is used, large enough to stop the parasitic oscillations, there will be excessive heating in the resistor if the power of the amplifier is too high. This heating occurs because the grid current flows through the resistor. In small amplifiers this heating is not objectionable. In larger amplifiers, however, a small inductance such as L_5 or L_6 is placed across the resistor. The small inductance has a high impedance at the parasitic frequencies. They are therefore damped out by the resistance such as R_{15} or R_{16} across the inductance. The D.C. resistance of the choke, however, is very low and therefore the grid current flows principally through the inductance and not through the resistor. Excessive heating in the resistor by the grid current is thus avoided. The output circuit of the power amplifier is balanced to ground by the use of a double-section condenser composed of C_{28} and C_{29}. Plate voltage is supplied through the radio-frequency choke RFC_7 and through the output inductance L_7. The amplifier is cross neutralized by the condensers C_{24} and C_{25}. The condensers C_{26} and C_{27} are filament by-pass condensers. As may be seen from Fig. 12–13, one side of each condenser is grounded. Due to these condensers, the filament of vacuum tube VT_5 is effectively at ground potential so far as radio frequency is concerned. It is to be noted that similar filament by-pass condensers are provided on the other amplifiers. In low-powered radio-frequency tubes it is not necessary to use the midtap filament by-pass condensers. Referring to C_3 of Fig. 12–13, it may be seen that a single filament by-pass condenser has been used on the low-power oscillator circuit.

The output of the amplifier is transferred to the *impedance-matching network* through the radio-frequency transformer formed by L_7 and L_8. The proper adjustment of inductances L_9 and L_{10} and condenser C_{31} makes it possible for the transmitter output power to be fed into a wide range of load resistances.

The power amplifier of the transmitter is, of necessity, Class C because it is plate modulated. Class C amplifiers are less critical in operation than are other classes. Their amplification per stage is a little lower than other classes. However, the plate-circuit efficiency is higher. Therefore the buffer amplifier and the intermediate power amplifier are also operated as Class C amplifiers.

Audio-frequency Amplifier. There are three stages of amplification in the audio-frequency section of the transmitter. Enough audio amplification is provided so that the audio input to the transmitter need not be higher than that required to feed a telephone line. The first two stages of amplification employ pentode tubes and the third or modulator stage employs triode tubes. All three stages of audio amplification are in push-pull. The audio-frequency input voltage is impressed upon the grids of the first audio amplifier through the transformer T_{10}. By the proper selection of the load resistors R_{22} and R_{23}, the frequency-response characteristic of the transformer T_{10} may be made uniform over the audio-frequency band. R_{26} and R_{27} form the cathode resistor. The grid return is made between R_{26} and R_{27} and therefore the control grid bias voltage is the voltage appearing across R_{26} only. The resistors R_{24}, R_{25}, R_{36}, R_{37}, R_{38}, and R_{39} and the connecting wires indicated by Z comprise the *inverse feedback* circuit. This circuit improves the fidelity of reproduction of the amplifier and will be described in more detail later. The screen-grid voltage for the first audio tubes VT_{11} and VT_{12} is obtained from the voltage dividers R_{28} and R_{29}. It is to be noted that the voltage divider is across the plate supply to the first audio amplifier. The interstage coupling between the first and second amplifiers is resistance coupling. R_{30} and R_{31} are the plate resistors. C_{43} and C_{44} are the blocking condensers. R_{32} and R_{33} are the grid resistors for the second audio amplifier tubes VT_{13} and VT_{14}. Separate bias adjustments are provided for each of the tubes in the second audio amplifier. The second audio amplifier must supply enough input power for the Class B final audio amplifier. It is therefore necessary to adjust the second audio amplifier for optimum performance by separate adjustment of the bias voltage on each of the tubes. The ammeters A_{10} and A_{11} measure the plate current drawn by each tube.

T_{14} is the interstage transformer between the second and third audio amplifiers. It has a double secondary so that each section may feed the grid of one of the tubes in the modulator. R_{34} and R_{35} are the load resistors on the secondaries of the transformer T_{14}. Because optimum operating conditions are required in a Class B audio amplifier, separate bias adjustments are provided for the grid of each tube. The plate current to each tube is indicated by the ammeters A_{12} and A_{13}. The output transformer of the modulator is T_{17}; the plate voltage is fed to the plates through a midtap on its primary. If the radio-frequency Class C power amplifier is operating at an efficiency of 80% for a 1,000-watt carrier, then the power drawn by the plates of the power amplifier is 1,250 watts. The audio-frequency power output of the modulator required for 100% modulation is 625 watts.

The plate voltage to the radio-frequency power amplifier is obtained from the high-voltage plate supply rectifier through the audio-frequency inductance L_{17} and the radio-frequency choke RFC_7. The plate-supply voltage is prevented from shorting to ground through the secondary of transformer T_{17} by the condenser C_{45}. The condenser C_{45} is large and therefore has a very small reactance at audio frequencies. The resistor R_{40} is also comparatively small. The audio-frequency output voltage in the secondary of T_{17} is therefore effectively impressed between ground and point Y of the diagram in Fig. 12–13. The audio-frequency choke L_{17}, called the *modulation choke*, prevents the audio frequency from shorting through the high-voltage plate-supply rectifier. The audio-frequency output of the modulator therefore varies the plate supply voltage and hence causes modulation in the plate circuit of the radio-frequency power amplifier.

Resistors R_{36} and R_{37} are in series between the plate and ground of VT_{15}. Likewise, R_{38} and R_{39} are in series between the plate and ground of VT_{16}. The audio-frequency output of the plates of the modulator tubes is across these two pairs of resistors. The two pairs of resistors act as voltage dividers so that a small portion of the audio-frequency output voltage can be fed to the grids of the first audio amplifier through the two circuits indicated by Z. The audio-frequency voltage across R_{37} is impressed across the grid resistor R_{24} of the first audio amplifier. Likewise, the audio-frequency voltage across resistor R_{38} is impressed across the resistor R_{25}. The feedback voltages across R_{25} and R_{24} are out of phase with the input voltages to the grids across the resistors R_{22} and R_{23}. This feedback cancels out distortions that are introduced by the amplifiers and it also cancels out a certain amount of internal noise in the amplifiers. This is inverse feedback. as described in Chapter 7.

It is to be noted that there is a D.C. component across the resistors R_{37} and R_{38} from the plate-supply voltage to the tubes of the modulator. This D.C. voltage is also impressed across the resistors R_{24} and R_{25}. This voltage is positive in respect to the grids of the first audio amplifier. It is therefore necessary to make the bias voltage across R_{26} equal to the sum of the normal bias voltage required plus the voltage drop across R_{24} or R_{25}. The feedback D.C. voltage is, however, very small because R_{37} and R_{38} are a very small portion of the plate-voltage dividers of the modulator tubes.

The audio-frequency current in the secondary of the transformer T_{17} passes through condenser C_{45} and resistor R_{40}. There is, therefore, an audio-frequency voltage across the resistor R_{40}. The audio frequency appearing across R_{40} is used to monitor the output of the transmitter. Any distortion that may appear in the radio-frequency amplifier is reflected in the modulator circuit through the transformer secondary to T_{17} and is detected in the monitor output. Likewise, distortion that is due to the radio-frequency power amplifier is reflected to a certain degree into the feedback circuit through the transformer T_{17} and therefore it is partially nullified by the feedback circuit.

Second Audio Bias-supply Rectifier. This rectifier supplies the bias voltage for the second audio amplifier. The bias voltage to each vacuum tube in the second audio amplifier is separately adjustable by means of the potentiometers R_{17} and R_{18}. These are across the output of the rectifier filter circuit. Filter reactors L_{11}, L_{12} and condensers C_{32}, C_{33} constitute the filter for the rectifier. The rectifier is full-wave, employing a single tube with two plates. The power supplied by this rectifier is rather small and therefore the small type of full-wave rectifier tube may be used. In such a small rectifier, it is feasible to employ only one transformer such as T_5 for supplying both the filament voltage and the plate voltage.

Low-power Plate-supply Rectifier. This rectifier supplies the suppressor-grid, screen-grid and plate voltages for all amplifiers except the radio-frequency power amplifier and the modulator. The various voltages required for the different grids and plates are obtained from the voltage divider R_{19}. The full-wave rectifier employs two half-wave rectifier tubes VT_7 and VT_8. T_6 is the filament transformer and T_7 the high-voltage transformer. The filter is composed of C_{34}, C_{35}, L_{13}, and L_{14}. Voltmeter V_2 measures the output voltage of the low-power plate-supply rectifier. C_{36} is a by-pass condenser for the screen grids of the second audio amplifier tubes.

Modulator Bias Rectifier. This is a full-wave rectifier supplying the bias voltages for the modulator tubes. Two half-wave rectifier tubes VT_9 and VT_{10} are employed. The filament transformer is T_9 and the plate voltage transformer T_8. The main filter is composed of C_{37}, L_{15}, and L_{16}. Separate bias adjustment for each tube in the modulator is provided by the voltage dividers R_{20} and R_{21}. Additional filtering is provided by the condensers C_{38} and C_{39}.

High-voltage Plate-supply Rectifier. This rectifier supplies plate voltage for the radio-frequency power amplifier and the modulator. It is a bridge rectifier employing the four half-wave rectifier tubes VT_{17}, VT_{18}, VT_{19}, and VT_{20}. T_{18} and T_{19} are the filament transformers. T_{20} is the high-voltage plate transformer. The filter is composed of L_{18}, L_{19}, C_{40}, and C_{41}. D.C. voltmeter V_1 measures the plate voltage on the modulated power amplifier.

The voltage divider R_{19}, as shown in Fig. 12–13, is pictured as a single resistor with five taps on it to supply the variable voltages required. In practice, however, this voltage divider is usually made up of several separate resistors. Figure 12–14 shows a plate-supply voltage divider made up of seven separate resistors. Five of the resistors, R_1, R_2, R_3, R_5, and R_7, are potentiometers and may be used to tap off voltage of a desired value. By the proper arrangement of resistors and potentiometers the taps may be located at the proper point on a voltage divider so that the potentiometer for any given voltage supply will provide the necessary operating adjustment.

In practice, similar voltage dividers are sometimes used in such places as R_5, R_{17}, R_{18}, R_{19}, R_{20}, and R_{21} of Fig. 12–13.

For clarity in understanding the circuit diagram of a radiotelephone transmitter, certain pieces of equipment and certain circuits are not shown

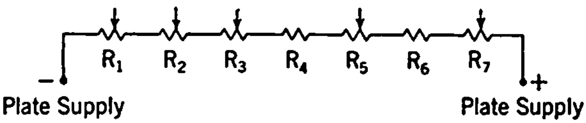


FIG. 12–14. Plate-supply Voltage Divider.

in Fig. 12–13. These are such items as overload relays, switches, time-delay relays, interlock switches, primary power circuits to the primaries of filament and plate transformers, and so on.

Figure 12–15 shows the filament ground-lead circuit of a vacuum tube through an inductance L. The plate current of the tube flows through L. The inductance L is a coil in an overload relay. When the current through L becomes excessive, the relay operates to open the contacts A. The contacts A may be inserted in the primary power-supply circuit to the high-voltage plate transformer, thus shutting off the plate supply to the vacuum tube T of Fig. 12–15. This circuit provides a means for protecting a vacuum tube against overload that may damage the tube.

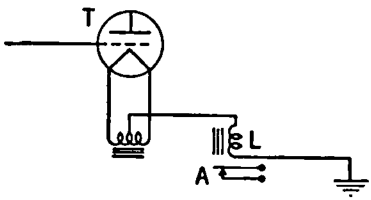


FIG. 12–15. Overload Relay.

The schematic circuit diagram of the primary power-supply circuit of a simple transmitter is shown in Fig. 12–16. S_1 is the main switch that controls power to the entire transmitter. The primary fuses for the entire transmitter are indicated by F_1.

The main primary power-supply switch S_1 is usually left closed so that the crystal heat circuit may have power at all times. It is usually undesirable to let the crystal heat chamber cool off if or when the transmitter is shut down. The next switch in the sequence of the transmitter primary-supply circuit is S_2. This controls power to the rest of the transmitter. The filament transformer and the bias-voltage rectifier transformer are connected on the output of S_2. S_3 controls the power to the primaries of the plate-supply rectifier transformers. With S_3 open and S_2 closed the filaments of the tubes in the transmitter can be warmed up to operating temperature before the plate voltage is applied. S_3, upon being closed, supplies power to the primaries of the plate-supply rectifier transformers. The contacts of all overload relays, such as that shown at A in Fig. 12–15, are put in series with S_3 so that if any one of the relays has an overload, the power to the plate-supply rectifier transformers will be interrupted. For safety reasons, transmitters are enclosed in housings so that the operating personnel cannot come in contact with the dangerous high voltage. Access doors to the equipment are provided in case work must be done on any of the equipment. In order to prevent someone from accidentally opening one of the access doors and coming in contact with the high voltage, an *interlock switch* is provided on each door. The interlock switch contacts are in series with S_3. Therefore, if an access door is opened, the contacts of the interlock switch open, thereby shutting off the high-voltage plate supply.

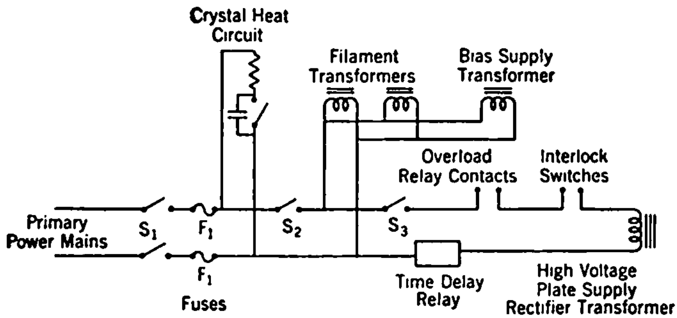
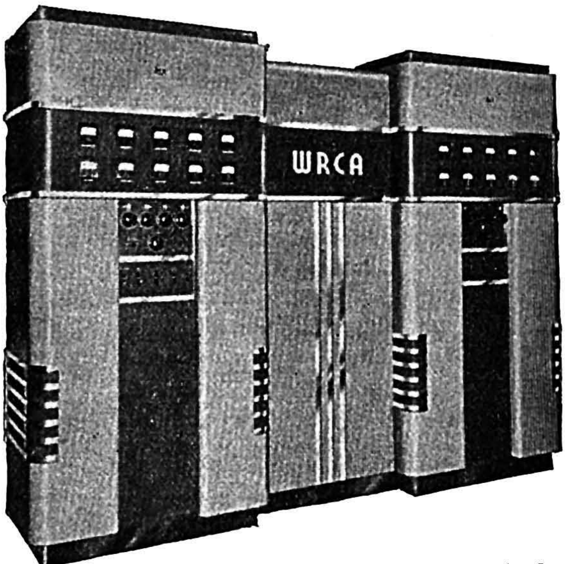


FIG. 12–16. Transmitter Control Circuit.

Most transmitters are provided with automatic start, so that only one switch such as S_2 need be closed in order to start the transmitter. In such a procedure S_3 would be normally closed. Upon closing S_2 the filaments and bias supplies would receive power but not the high-voltage plate-supply rectifier transformer because the contacts in a *time-delay relay* would be open. These would be in series with S_3. The closing of S_2 would energize a coil in the time-delay relay. After a predetermined time the contacts would close, thus closing the power circuit to the high-voltage rectifiers.

In small transmitters, S_2, S_3, the overload-relay contacts, the interlock contacts, and the time-delay relay contacts would themselves be in series with the power circuits. In larger transmitters, where the current drawn in the primary power-supply circuit is high, the switches and contacts would be used to operate power-control relays called contactors. Contactors are similar to the overload relay described in connection with Fig. 12–15 except that they sometimes have more than one set of contacts, depending on how many circuits must be opened and closed. The coil to operate the contactor requires a comparatively small amount of power. The contacts may be built to handle a large amount of power.

The circuit diagram shown in Fig. 12–13 is somewhat similar to the circuit diagram of the R.C.A. 1,000-watt broadcast transmitter type 1–K. The regular circuit diagram of the type 1–K is too complicated for instruction purposes. A picture of the R.C.A. type 1–K transmitter is shown in Fig. 12–17. The rear of the transmitter is shown in Fig. 12–18. Various components of the transmitter are identified and a reference given to the



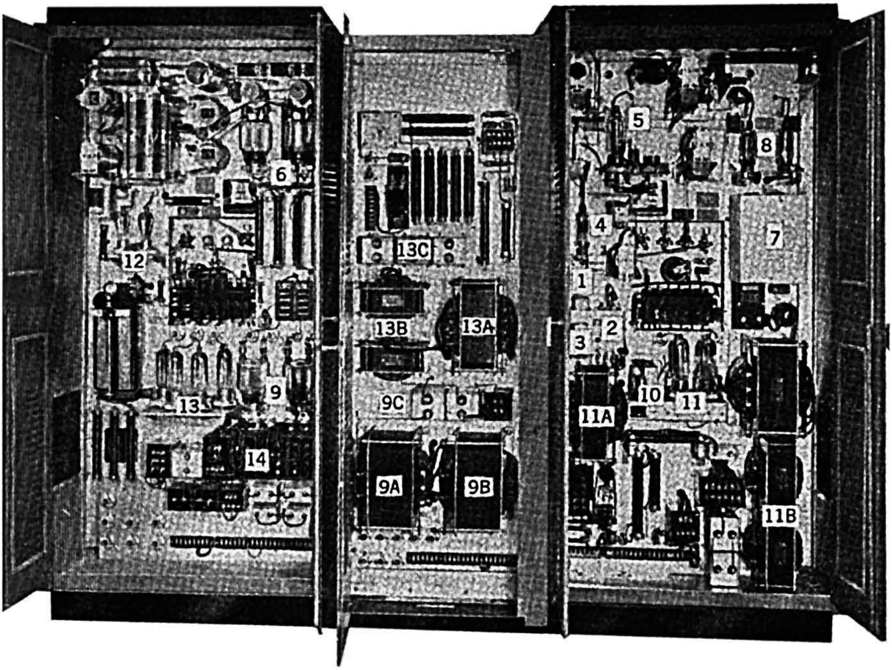
Courtesy of R.C.A. Manufacturing Co., Inc.

FIG. 12–17. Front View of 1,000-watt Broadcast Radiotelephone Transmitter, R.C.A. Type 1-K.

similar piece of equipment as it would be used in the transmitter circuit diagram shown in Fig. 12–13.

The complete schematic diagram of the type 1–K transmitter is shown in Fig. 12–20. The description of this diagram will not be given. However, a study of the circuit diagram in Fig. 12–20 compared with the circuit diagram in Fig. 12–13 will yield information pertinent to the over-all circuit diagram of a transmitter. It should be kept in mind that the circuit diagram of Fig. 12–13 is similar to that of Fig. 12–20 but that it is not a simplification of the circuit diagram of the 1–K transmitter. Many of the circuits have been omitted, particularly the control circuits and primary power circuits, and some of the other circuits have been modified for clarity in explaining their operation.

An inside view of the R.C.A. type 5–E transmitter is shown in Fig. 12–19. The left-hand panel is the output amplifier. Two tubes are shown and they are both used for 10-kw output. For 5-kw operation only one of the tubes

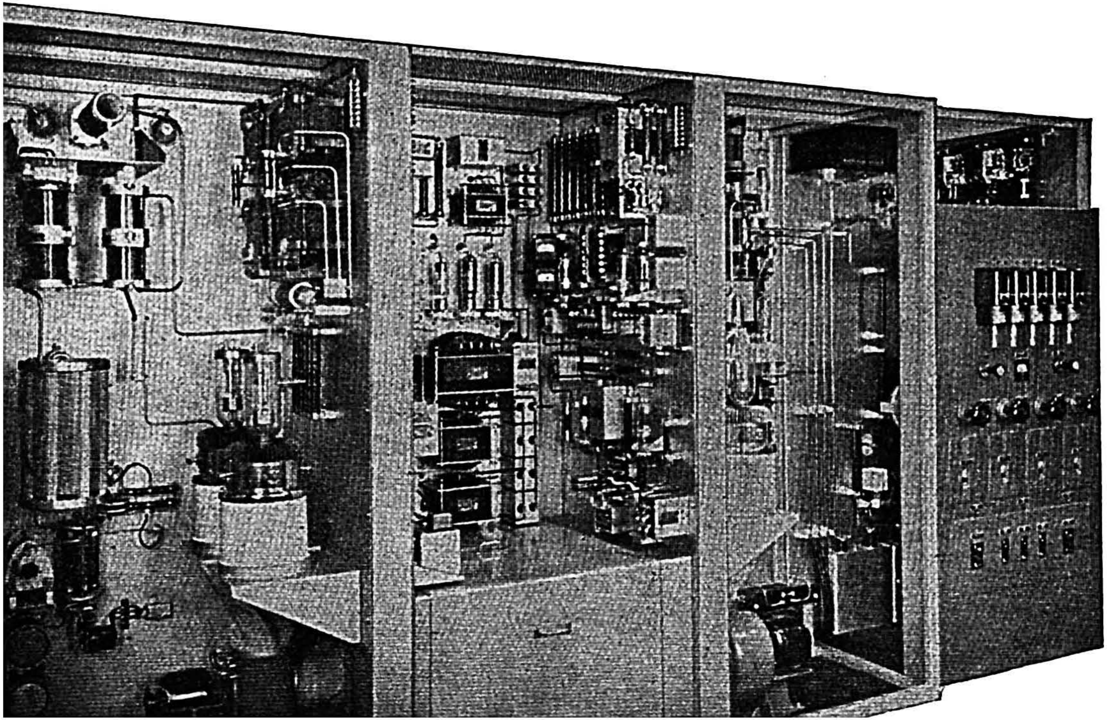


Courtesy of R.C.A. Manufacturing Co., Inc.

Fig. 12–18. Rear View of 1,000-watt Radiotelephone R.C.A. Type 1-K.

KEY TO COMPONENTS

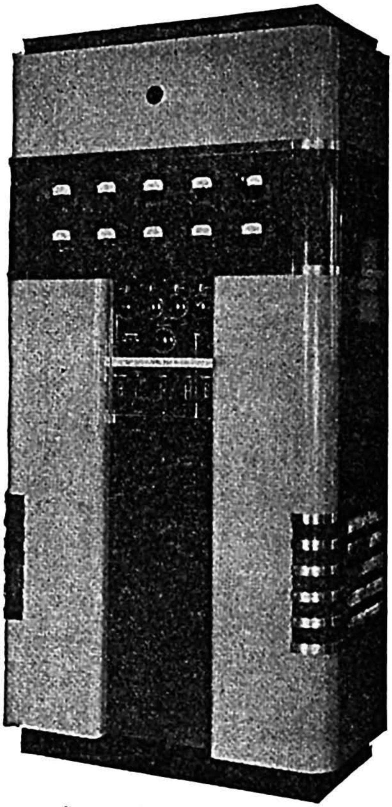
umber and Description		*Corresponding Equipment in Fig. 12–13*
l.	Crystal holder and heat chamber	Crystal heat chamber
2.	Crystal oscillator and shield	Oscillator
3.	Spare crystal holder	
4.	Buffer amplifier tube	VT_2
5.	Intermediate power amplifier tube	VT_3
6.	Radio-frequency power amplifier tubes	VT_4 and VT_5
7.	First audio amplifier shield	
8.	Second audio amplifier tubes	VT_{13} and VT_{14}
9.	Modulator tubes	VT_{15} and VT_{16}
9A.	Modulator output transformer	T_{17}
9B.	Modulation reactor	L_{17}
9C.	Audio by-pass condenser	C_{45}
0.	Second audio bias rectifier tube	VT_6
1.	Intermediate plate-supply rectifier tubes	VT_7 and VT_8
1A.	Intermediate plate-supply high voltage transformer	T_7
1B.	Intermediate plate-supply filter reactors	L_{13} and L_{14}
2.	Modulator bias rectifier tubes	VT_9 and VT_{10}
3.	High-voltage plate-supply rectifier tubes	VT_{17}, VT_{18}, VT_{19}, and VT_{20}
3A.	High-voltage plate-supply transformer	T_{20}
3B.	High-voltage plate-supply filter reactors	L_{18} and L_{19}
3C.	High-voltage plate-supply filter condensers	C_{40} and C_{41}
4.	Filament transformers	Such as T_4, T_{15}, T_{16}, T_{18} and T_{19}



Courtesy of R.C.A. Manufacturing Co., Inc.

Fig. 12–19. Inside View of 5-kw Broadcast Radiotelephone Transmitter, R.C.A. Type 5-E.

is used. It is to be noted that the tubes are inserted in radiator fins. Directly below each tube is an air blower that forces air up around the tube. The heat dissipation of tubes of this size is high and the heat must therefore be removed by forced air circulation. The second rack from the left contains some of the rectifiers and the low-powered audio and radio amplifiers. The transmitter is high-level modulated and therefore a high audio power is required for modulation. One of the two modulator tubes may be seen to the left in the third rack from the left. Two of the high-voltage filter condensers may be seen in the lower right-hand corner of the third rack from the left. The right-hand panel contains the control circuits, the voltage adjustments, and the overload relays.



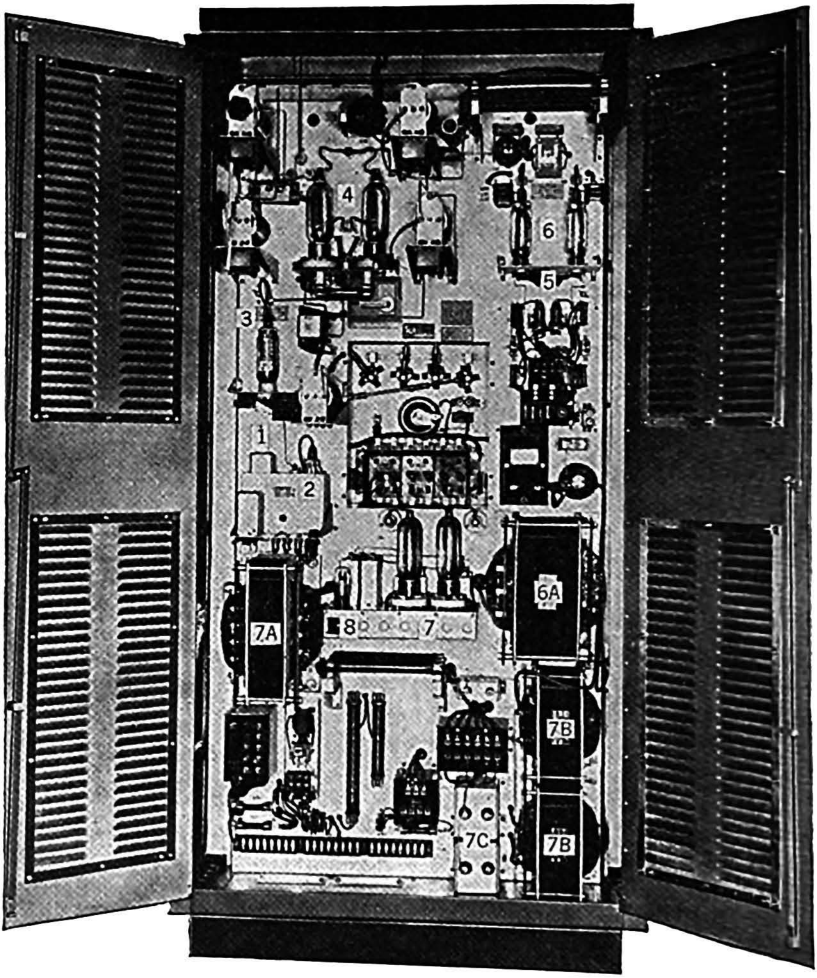
Courtesy of R.C.A. Manufacturing Co., Inc.

FIG. 12–21. Front View of 250-watt Broadcast Radiotelephone Transmitter, R.C.A. Type 250K.

250-watt Transmitter. A front-view picture of a 250-watt broadcast transmitter is shown in Fig. 12–21. This is a high-level plate-modulated transmitter employing a crystal oscillator and two stages of radio amplification. There are also two stages of audio-frequency amplification. A rear view of the transmitter is shown in Fig. 12–22, in which various components are identified.

Inverse Feedback in Transmitters. Inverse feedback for audio amplifiers was described in Chapter 7. One method of applying inverse feedback to a transmitter was explained in connection with Fig. 12–13.

Inverse feedback may be applied to a transmitter in a different way. Referring to Fig. 12–23, the inductance L_2 picks up radio-frequency energy from the output circuit inductance L of the transmitter. This energy is fed through a detector to obtain the audio frequency to be used in the inverse feedback. This audio frequency is fed across resistor R_2 in the grid circuit of the first audio-frequency amplifier of the transmitter. It is out of phase with the audio frequency input to the first audio amplifier and, therefore, the benefits of inverse feedback are derived.



Courtesy of R.C.A. Manufacturing Co., Inc.

FIG. 12–22. Rear View of 250-watt Broadcast Radiotelephone Transmitter, R.C.A. Type 250K: (1) crystal holder and heat chamber; (2) crystal oscillator; (3) buffer amplifier tube; (4) radio-frequency power amplifier tubes; (5) first audio amplifier; (6) modulator tubes; (6A) modulator output audio transformer; (7) plate voltage-supply rectifier tubes; (7A) plate-supply rectifier high-voltage transformer; (7B) plate-supply filter reactor; (7C) plate-supply filter condensers; (8) bias rectifier tube.

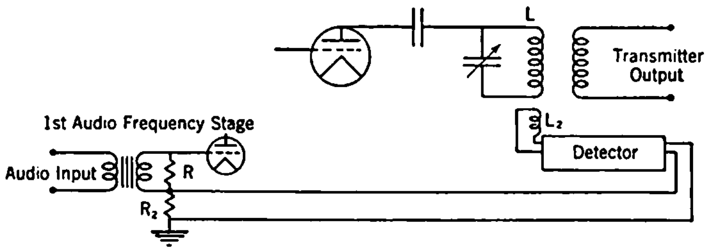


FIG. 12-23. Inverse Feedback Applied to a Modulated Transmitter.

Distortion introduced by the transmitter is reduced, the frequency response of the transmitter is improved, and the noises introduced by the transmitter are reduced.

Review Questions and Problems. *1.* What function does a radio transmitter perform in a communication system?

2. (a) What are the two principal sections of an amplitude-modulation radio transmitter? (b) Describe in general the components of each section and how they function.

3. Why are buffer amplifiers used?

4. What are the two types of modulated transmitters and why are they so named?

5. What two precautions are taken to insure that the frequency of a crystal oscillator will remain within proper limits?

6. What type of audio-frequency amplifier is usually used for modulators? Explain why.

7. (a) What is power-supply regulation? (b) Why is good regulation necessary in Class B audio amplifiers and in Class C modulated amplifiers?

8. How may inverse feedback be applied to a transmitter?

CHAPTER 13

Amplitude-modulation Radio Receivers

A radio receiver performs the function of producing sound waves that are in accordance with the modulation envelope of a radio-frequency signal. The main receiving-set factors are sensitivity, selectivity, and fidelity. *Sensitivity* is a measure of a receiver's ability to receive weak signals. *Selectivity* is concerned with the receiver's ability to reject undesired signals. *Fidelity* is a measure of a receiver's ability to reproduce faithfully audio-frequency currents that are in accordance with the modulation envelope of the received signal.

Sensitivity. The sensitivity of a receiver is determined by the amount of radio amplification preceding the detector. The sensitivity of receivers used for various types of communication varies a great deal. A broadcast receiver for local reception need not be very sensitive and may be built with as few as three or four tubes. On the other hand, a receiver employed commercially for transoceanic communication must have a high sensitivity. It is not uncommon for such a receiver to employ fifteen or more vacuum tubes.

There is a practical limit to the amount of sensitivity that can be employed in a receiver. Static is a limitation, and there is no point in building a receiver that will respond to signals lower than the static level. Man-made noise is also a limitation; however, steps may be taken to reduce this. Man-made interference may be minimized by filters on the electrical apparatus causing the interference. Commercial receiving installations, however, are often located in the country away from power lines and other electrical apparatus.

Another important source of noise is the random motion of electrons in the first stage of the amplifier, both in the grid circuit and in the tube itself. Since these random motions produce small varying voltages which are amplified through all the stages of the receiver, they set an upper limit on the total amplification which can be used.

A narrower band width is required for code than for radiotelephone reception. The narrower the band width, the less static and man-made noise the receiver accepts. Therefore, owing to the narrower band width, code receivers may employ a greater sensitivity. With a given transmitting power the range of the station is greater for radiotelegraphy than it is for radiotelephony.

In commercial communication installations, high-gain antennas are usually used. High-gain receiving antennas are directional and a greater signal input voltage is therefore impressed upon the receiver input by a transmitter in the proper direction from the antenna. Such antennas effectively increase the sensitivity of a receiving system.

Selectivity. The greater the selectivity of a receiver the more it will discriminate against undesired signals on channels adjacent to the one that the desired station is on. If the sensitivity of a receiver is increased by the addition of radio amplifiers the selectivity is automatically improved. The selectivity of the individual tuned coupling circuits between radio amplifiers functions to improve the over-all selectivity. This latter is greater than that produced by an individual tuned coupling circuit.

If the selectivity is made too great, then the receiver will not respond uniformly to the band of frequencies contained in an amplitude-modulated wave. The successive tuning stages may be slightly detuned to produce a uniform response over all of the band. This process will be explained later.

Fidelity. A great many factors in a receiving set have a bearing on the fidelity of reproduction. In the paragraph on selectivity, mention was made of manipulating the successive tuning circuits so as to obtain a uniform response over a band of frequencies. If a uniform response is not obtained, then the higher frequencies will be attenuated and the quality of the received music or voice will sound boomy.

The response curve of a tuned circuit is shown in Fig. 13–1a. It is assumed that the carrier frequency is 1,000 kc and that the highest modulation frequency is 10,000 cycles. The desired band width is from 990 to 1,010 kc. The response is not uniform over this band and therefore the higher modulation frequencies are discriminated against. In a receiver with multiple tuned circuits the successive circuits may be tuned slightly off the carrier frequency. Figure 13–1b shows a plot of the response of two tuned circuits A and B, slightly detuned from the carrier frequency of 1,000 kc. The over-all response curve of the two circuits is shown in Fig. 13–1c. The response is comparatively uniform over the desired band from 990 to 1,010 kc.

Radio-frequency transformers are usually used for interstage coupling in receivers. Examples are shown in Fig. 13–9. Assume that both the primary and secondary of a radio-frequency interstage coupling are tuned to resonance. The coupling between the primary and secondary of the radio-frequency transformer may be adjusted so as to produce a response such as that shown in Fig. 13–1c. In this method of obtaining a flat response the circuits are not staggered but are actually tuned to resonance and the effect is obtained by varying the coupling between primary and secondary.

As the band width of the receiver is expanded, the receiver responds to more static and man-made noises. It therefore requires a higher signal strength to give interference-free reception.

The detector in a receiver is very important. It is often the source of poor quality. Pains must therefore be taken to see that the detector is linear.

The audio system of a receiver is composed of an audio amplifier and a pair of earphones or a loudspeaker. The audio amplifier must be able

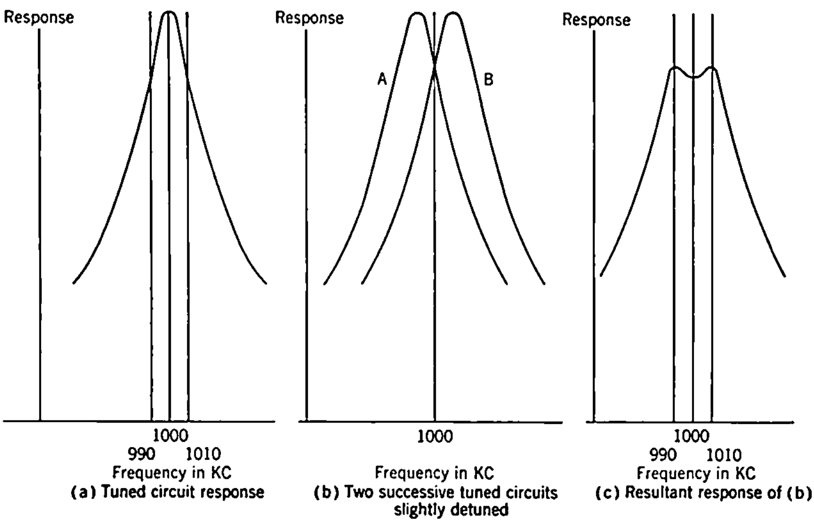


Fig. 13-1. Multiple-circuit Selectivity.

to reproduce faithfully all of the audio-frequency components of the modulation envelope of the received signal. The earphones or loudspeaker must likewise be able to reproduce these frequencies faithfully. In a high-fidelity loudspeaker, two or more units are sometimes used. In a two-unit loudspeaker one unit reproduces the low frequencies and the other reproduces the high frequencies.

Simple Receiver. The simplest type of receiver employs an antenna, an input tuning circuit, a detector, and a pair of earphones. A simple receiver such as this is unsatisfactory for two reasons. The first reason is that it is not sensitive enough. The second reason is that a single input tuning circuit does not effectively reject undesired signals.

A simple crystal receiver is shown in Fig. 13-2. The signal from the antenna is transformed to the tuning circuit by the radio-frequency transformer composed of inductances L and L_2. L_2 and C comprise the tuned circuit. X is the crystal rectifier. Condenser C_1 and the resistance of the earphones form the output filter of the detector. Such a receiver has very low sensitivity because the highest output power available is the detected signal itself. The selectivity of the receiver is poor because there is only one tuned circuit.

In order to increase the sensitivity of a receiver, radio-frequency amplification is employed between the antenna and the detector. There is a tuning

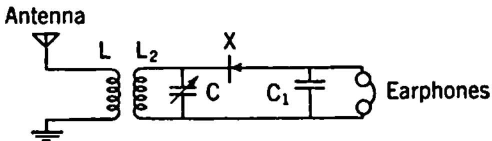


FIG. 13–2. Simple Crystal Receiver.

circuit on the input to the first radio amplifier, a tuning circuit between each succeeding radio amplifier, and a tuned input circuit to the detector. The additional tuning circuits add to the selectivity of the receiver.

Tuned Radio-frequency Receiver. The block diagram of a simple tuned radio-frequency receiver is shown in Fig. 13–3. The elements of the receiver are a radio-frequency amplifier, a detector, an audio-frequency amplifier, and a loudspeaker. The radio-frequency amplifier amplifies the signal input voltage from the antenna and impresses it upon the input to the detector. The audio-frequency output of the detector is amplified in the audio-frequency amplifier to a volume great enough to operate a loudspeaker or a pair of earphones.

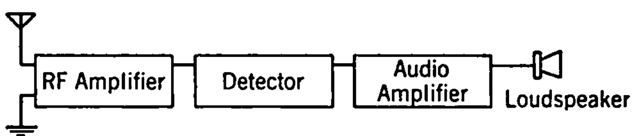


FIG. 13–3. Block Diagram of Simple Tuned Radio-frequency Receiver.

SELECTIVITY. There is a tuned input circuit to the radio amplifier and a tuned input circuit to the detector. The selectivity curves of the two circuits function to produce a greater over-all selectivity. Tuned radio-frequency receivers usually have two or more radio amplifiers. Each time a radio amplifier is added to the receiver another tuned circuit is added, thereby increasing the selectivity of the receiver.

A given selectivity is more easily attained with a superheterodyne circuit. This greater selectivity is one of the reasons why superheterodyne receivers are in greater general use than are tuned radio-frequency receivers.

Multistage Radio-frequency Receiver. The block diagram of a multistage tuned radio-frequency receiver is shown in Fig. 13–4. Three radio amplifiers are shown ahead of the detector. The additional radio amplification increases the sensitivity and the selectivity of the receiver.

Threshold Sensitivity of a Receiver. Mention has been made of the increased selectivity resulting from the additional tuned circuits employed with radio amplifiers. A single tuning network may be built with selec-

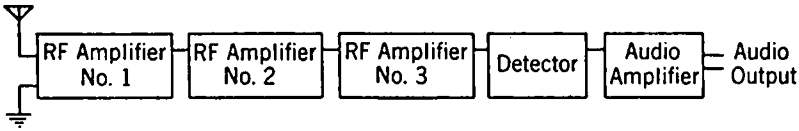


FIG. 13-4. Block Diagram of Multistage Tuned Radio-frequency Receiver.

tivity equivalent to that possessed by several radio amplifiers. This network could be inserted between the antenna and the detector.

A receiving set is shown in Fig. 13-5. The selective circuit is composed of multiple inductances and condensers to simulate the selectivity of several interstage tuning circuits. The selectivity characteristic is the same as that obtained by the combined selectivities of the coupling circuits in a

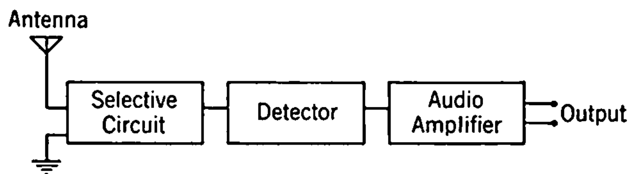


FIG. 13-5. Receiver with "Lumped" Selectivity.

multistage radio amplifier. In order to bring the output of the receiver up to the desired level, it is assumed that the audio-amplifier contains sufficient amplification.

In the light of this it would seem that the output of the detector could be amplified at audio frequency to whatever level was desired. There are noises generated internally in a vacuum tube which are amplified through audio-frequency amplifiers. The radio-frequency signal input voltage to the detector must be high enough so that there is a proper ratio between the incoming signal and the internal noises in the detector. Therefore the signal input voltage to the detector must not be below a certain *threshold* or the reception will be marred by internal noises in the detector. In order to operate a receiver at signal voltage inputs far below the detector threshold, it is necessary to add radio amplification ahead of the detector. Linear detectors such as the diode and the linear plate detector require comparatively large signal inputs. Radio amplification ahead of the detector is used for this reason also. There is a discussion of this point in Chapter 11.

Superheterodyne Receiver. Superheterodyne reception is accomplished by converting the frequency of the incoming signal to a second frequency called the *intermediate frequency.* The radio amplification takes place at the intermediate frequency. The signal at the intermediate frequency is then detected to obtain the audio frequencies. Assume that two radio-frequency signals of frequency f_1 and f_2 are brought together in the input to the vacuum-tube detector. Two new signals are formed. One is the sum of f_1 and f_2 and the other is the difference between f_1 and f_2. Assume that f_1 is 1,000 kc and f_2 is 1,455 kc. The resultant new signals will have frequencies of 2,455 kc and 455 kc. If f_1 were amplitude-modulated, then its side bands would be converted over to the new signal of 455 kc. The new modulated signal at 455 kc may be passed through radio amplifiers and then detected to obtain audio frequencies in accordance with the envelope of the original modulated signal on 1,000 kc. This is the general principle of superheterodyne reception.

The block diagram of a simple superheterodyne receiver is shown in Fig. 13–6. It is assumed that the incoming signal has a frequency of

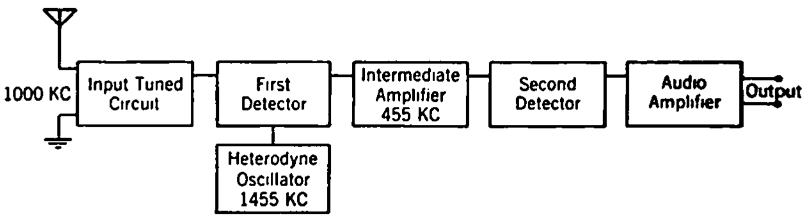


FIG. 13–6. Block Diagram of a Simple Superheterodyne.

1,000 kc. The input circuit is tuned to this frequency. A local *heterodyne oscillator* is adjusted to the frequency of 1,455 kc and is coupled into the first detector along with the incoming signal. Through heterodyne action, a signal with a frequency of 455 kc is formed. This signal is amplified in the *intermediate-frequency amplifier.* The output of the intermediate amplifier is impressed upon the grid of the second detector, where detection takes place; thereby audio frequencies are obtained that are in accordance with the modulation envelope of the incoming 1,000-kc signal. The audio frequencies are amplified, in the audio amplifier, to a level high enough to actuate a loudspeaker or a pair of earphones. In order that signals with different carrier frequencies may be received, the tuning of the input circuit and the heterodyne oscillator are variable and simultaneous on one dial. The frequency of the oscillator is always 455 kc higher than the frequency of the incoming signal. Therefore, whatever the frequency of the incoming signal is, the intermediate frequency is 455 kc. Other intermediate frequencies are also used, but 455 kc is common in broadcast receivers. Since the intermediate frequency is fixed in any given receiver, it is possible to employ fixed tuning in the intermediate-frequency amplifier. It is easier to obtain a desired selectivity and gain at a fixed frequency than it is when the amplifier must tune over a range of frequencies. The superheterodyne therefore has the advantage over a tuned radio-frequency receiver in that the radio-frequency amplification may be obtained at a fixed frequency.

The heterodyne action in itself increases the selectivity, another advantage that the superheterodyne circuit has over the tuned radio-frequency circuit. Assume that a desired station is on 1,000 kc and that an undesired station is on 1,010 kc. The difference in frequency between the two stations is 1% of the carrier frequency of the desired station. If the desired station is properly tuned in, its intermediate frequency may be, say, 455 kc. The undesired signal will be at an intermediate frequency of 465 kc. The stations are still 10 kc apart. The difference in frequency is now 2.2% of the frequency of the desired signal. The superheterodyne action has increased the percentage difference between the undesired and the desired signal. As is pointed out in Chapter 11, the frequency response of tuned circuits is determined by the relative change, expressed per cent. The selectivity of a circuit, in other words, is increased as the numerical ratio between the undesired and desired signals increases. The superheterodyne action, therefore, contributes materially to the selectivity of the receiver.

The superheterodyne circuit has one disadvantage, namely that the heterodyne oscillator is in reality a low-powered transmitter. Referring to Fig. 13–6, the 1,455-kc signal generated by the heterodyne oscillator is connected to the input of the first detector and thence through the input tuning circuit to the antenna. Radiation from the receiver takes place on 1,455 kc. If another receiver in the vicinity is tuned to receive a signal on 1,455 kc there will be interference. In order to prevent this, *preselection radio amplification* is employed. One or more radio-frequency amplifiers are placed between the antenna and the first detector. One stage of preselection radio amplification is not always entirely satisfactory because there is still a slight feed through the plate-to-grid capacity of the radio amplifier.

In order to stop radiation effectively, the radio-frequency section of the receiver should be carefully shielded.

Automatic Volume Control. The received intensity of a signal from a transmitting station at a distance may vary over wide limits. Variations due to sky-wave transmission are explained in Chapter 15. The signal input voltage to an automobile receiving set will vary over wide limits dependent on the location of the vehicle. In tuning from one station to another or in tuning over a range of frequencies to listen to a number of different transmitters the received signals will have intensities that are different to a great extent. It is desirable in either of these common situations that the signal out of the loudspeaker remain substantially the same even though the input signals at various frequencies vary greatly in intensity. Receiving sets incorporate *automatic volume control* circuits that make it possible to obtain substantially the same audio-frequency output from a receiver with large differences in radio-frequency signal input to the receiver. The voltage amplification of radio amplifiers varies with the voltages on the screen grid and the suppressor grid and with the bias voltage on the control grid. The automatic volume-control circuit functions to vary one or more of these voltages so as automatically to change the amplification of the radio amplifiers. This circuit includes the radio-frequency as well as the intermediate-frequency amplifier. The voltage to accomplish this is obtained from the detector. It is easier to obtain the control voltage from a diode detector than it is from other types of detectors and this is an additional reason why diode detectors are extensively employed in receiving sets.

The circuit diagram of a diode detector is shown in Fig. 13–7. Resistor R and condenser C_1 comprise the output circuit of the detector. As was explained in Chapter 11, when a radio-frequency signal input voltage is impressed upon the diode detector, a D.C. voltage is produced across the output resistor R. As the radio-frequency signal voltage is increased, the D.C. voltage across R is increased. The D.C. voltage across R is fed back to the bias leads of the radio amplifiers and is there employed to vary the amplification. As the signal input voltage to the diode detector increases, the D.C. voltage across R increases, thereby reducing the amplification of the radio amplifiers. By proper adjustment, the audio-frequency output of the diode detector can be maintained at approximately the same intensity despite large changes in radio-frequency signal input voltages to the receiver.

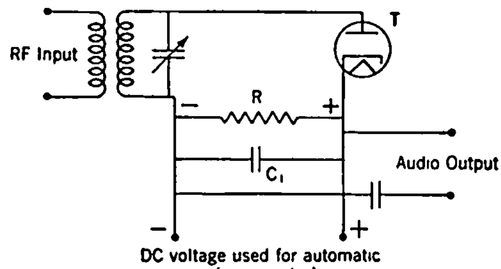


FIG. 13-7. Automatic Volume Control.

Squelch Circuit. When a receiver is tuned to a station, the amplification of the receiver is reduced by the automatic volume control to a point where the proper audio-frequency output is obtained from the receiver. If the receiving set is then tuned in search of another station, the sensitivity of the receiver greatly increases between stations. The noise level from static and man-made sources greatly increases and may produce an annoying disagreeable noise from the loudspeaker. The noise received when an incoming signal is not present in a receiver may be reduced by employing a *squelch circuit.* Such a circuit is shown in Fig. 13–8. Vacuum tube T_1 is employed in a diode detector circuit. L and C constitute the tuned radio-frequency input circuit. The output circuit of the diode detector is composed of C_1 and R. C_3 is a conventional audio-frequency coupling condenser. The audio-frequency output of the diode detector is impressed on the grid of the vacuum tube T_3 through C_3. The cathode K of the diode rectifier, at audio frequencies, is effectively connected to the cathode of T_3 through the by-pass condensers C_4 and C_5. T_3 functions as an audio-frequency amplifier. T_3 obtains its plate supply from the section of the voltage divider marked *c*. Its bias voltage is obtained from the section of the voltage divider marked *b* and from the voltage drop in the resistor R_2. R_2 supplies plate voltage to the vacuum tube T_2. The vacuum tube T_2 and its associated circuits produce the *squelch* action. The plate current drawn by T_2 determines the bias voltage on the audio-amplifier

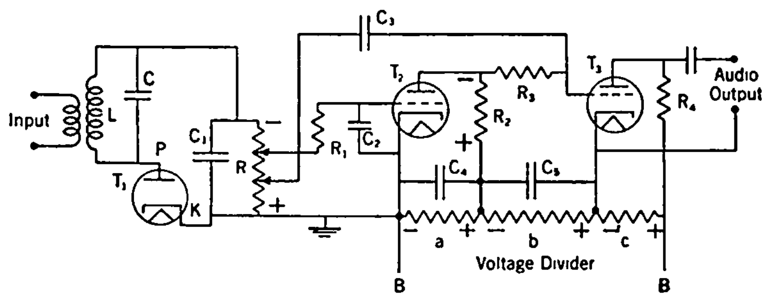


FIG. 13-8. Squelch Circuit.

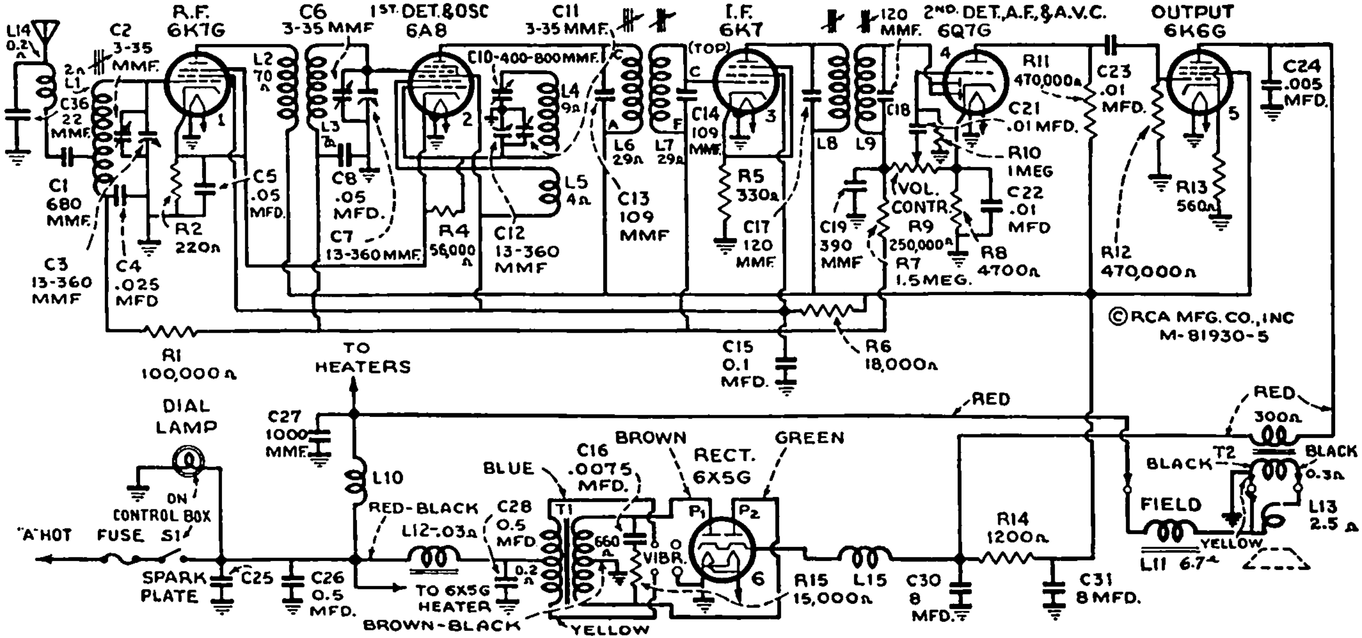
T_3. When there is a radio-frequency signal input voltage to the diode detector, there is a D.C. voltage across the resistor R. This D.C. voltage furnishes bias to T_2. R_1 and C_2 constitute an audio-frequency filter. This filter irons out the audio frequency in the varying voltage across R and thereby furnishes a steady D.C. bias voltage for T_2. The circuit constants are so adjusted that when there is a radio-frequency input to the diode detector, the bias voltage on T_2 will be beyond the cutoff point, thus stopping the flow of plate current to T_2. Therefore there is no voltage drop across R_2. The bias voltage on the audio amplifier T_3 will only be the voltage across section *b* of the voltage divider. This voltage is adjusted so that T_3 will function as an audio-frequency amplifier.

When there is no radio-frequency signal input voltage to the diode detector there is no D.C. voltage drop across resistor R, and therefore there is no bias on the *squelch tube* T_2. T_2 therefore draws a large plate current through resistor R_2 and thereby causes a large D.C. voltage to appear across R_2. This additional voltage biases the audio amplifier T_3 beyond cutoff, thus preventing T_3 from amplifying any audio frequency that may be impressed on its grid. Therefore T_3 can only function as an amplifier when there is a radio-frequency input voltage to the diode rectifier. In this way, during tuning from station to station, the output of the receiver is cut off.

Squelch circuits are also used in mobile communication systems, such as police receivers. Normally the transmitter is inactive, and if no squelch were supplied the receiver would be very noisy. With the squelch the receiver is silent until the carrier frequency is received from the transmitter; then the receiver becomes normally operative.

Superheterodyne Receiver Circuit. A complete schematic diagram of the circuit of a six-tube superheterodyne automobile receiver is shown in Fig. 13–9. The receiver has a preselection radio-frequency amplifier, a combination first detector and oscillator, one stage of intermediate-frequency amplification, a combination second detector and audio-frequency amplifier, an automatic volume control, and one stage of audio-frequency amplification. The primary power supply for the receiver is obtained from a 6-v automobile battery. All vacuum-tube filaments are heated by the 6-v battery supply. Inductance L_{10} and condenser C_{27} form a filter to prevent noises originating in the electrical system of the automobile from entering the receiver through the filaments. The instrument marked VIBR is a vibrator that causes pulses of the 6-v direct current to pass through the primary of the transformer T_1. T_1 is a step-up transformer used as the plate-supply transformer for the 6X5G rectifier tube. The output voltage of the rectifier is filtered by the inductance L_{15}, the resistor R_{14}, and the condensers C_{30} and C_{31}. The loudspeaker used is electrodynamic and current must therefore be supplied to the field coil. This current is supplied from the battery. The field coil of the speaker is indicated by L_{11}. The audio-frequency output transformer is T_2 and the plate voltage to the audio-frequency amplifier is supplied through the primary of this transformer.

The circuit of the preselection radio-frequency amplifier will be recognized as one of those shown in Chapter 11. A single 6A8 vacuum tube functions as the first detector and oscillator. The oscillator circuit, composed of C_{10}, C_{11}, C_{12}, L_4, and L_5 is connected to two grids of the first detector tube. This circuit produces the local oscillation which in turn produces the intermediate frequency of 260 kc. The signal input voltage from the radio-frequency amplifier is impressed upon one of the grids of the 6A8 detector tube. The control grid of the tube is biased in such a manner that the tube also functions as a detector, thereby causing the intermediate frequency to be formed. The 6K7 tube is the intermediate-frequency amplifier. The tuning between the first detector-oscillator and the intermediate-frequency amplifier, and between the intermediate-frequency amplifier and the second detector, is fixed at the intermediate frequency of 260 kc. The 6Q7G second detector tube contains a diode rectifier and a triode tube in the same envelope. The diode elements perform the rectification for the second detector. The control grid of the



Courtesy of R.C.A. Manufacturing Co., Inc.

FIG. 13–9. Schematic Circuit Diagram of Superheterodyne Receiver, R.C.A. Model 8M1.

triode is connected to the arm of the potentiometer R_9. R_9 is the manual volume control of the receiver. R_9 is in the diode detector circuit and therefore audio frequency is supplied to the grid of the triode. The triode element then functions as an audio-frequency amplifier. Automatic-volume-control voltage is fed from the diode through the resistor R_9 to the grids of the previous tubes. Variation in signal input to the diode detector produces a change in the D.C. voltage across the diode circuit resistors R_9 and R_8. This variable D.C. voltage is fed to the grids of the previous tubes, changing their bias and thereby changing their voltage amplification. Automatic volume control is thus obtained, as explained in detail earlier in this chapter. The audio-frequency output of the triode in the 6Q7G vacuum tube is resistance-coupled to the 6K6G output audio amplifier tube. The output of the 6K6G tube is fed to the loudspeaker through the matching transformer T_2.

Review Questions and Problems. *1.* What function does a radio receiver perform?

2. What are the three main receiving-set factors? Explain what each is and their importance.

3. What are the principal components of a radio-frequency receiver and what function does each perform?

4. What are the principal components of a superheterodyne receiver and what function does each perform?

5. Why is a superheterodyne receiver more selective than a radio-frequency receiver?

6. What are the limiting factors in radio-receiver sensitivity?

7. (a) Why is an automatic volume control used in a receiver? (b) Explain how it functions.

8. Why is a squelch circuit used in receivers?

9. Explain the superheterodyne action.

CHAPTER 14

Frequency Modulation

General Principles. A station whose radio wave does not change in any respect cannot be used to send signals from one point to another, because all information is conveyed by changes in the radio wave. The process of changing the amplitude or strength of the wave has already been discussed. This chapter deals with a second method of changing the radio wave, by varying its *frequency* in a manner which corresponds to the information to be transmitted. Such a variation is called *frequency modulation* (F.M.).

The frequency is the number of alternations (cycles) per second that the wave makes, and is the same as the figure often used to identify a radio station—such and such a number of kilocycles, for instance. If some means were used to vary or change this frequency a slight amount on either side of its average or assigned value—faster or slower—these changes might be discovered or detected by a suitable device at the receiving end. Such changes can be made by suitable transmitters, and receivers can be constructed which will respond to a frequency-modulated wave.

An example of frequency modulation for code transmission may be used as an illustration of how this frequency variation might be effected. Code signals consist of dots and dashes in groups to make up letters and words. One way to send these dots and dashes is to press a switch or key which will start the transmitter or "put it on the air" for the time necessary to make a dot or a dash and leave it off the air for the rest of the time. This is amplitude modulation, because the amplitude or strength of the signal is changed from nothing (with the key open) to its maximum (with the key closed) as often as necessary to spell out the letters and words of the message. An alternative system is to keep the transmitter on the air all of the time but use an extra circuit which may be switched into operation when the key is open and cut out again when the key is closed. This extra circuit is so arranged as to cause the transmitter to operate at a different frequency from the one used when there are no signals. The receiver is tuned to the second frequency and the signal will appear when the extra circuit is connected. Signals may be transmitted which seem (at the receiving end) to be like those of the amplitude-modulation system. The transmitter is operating all of the time a message is being sent, the only change being a shift of frequency in step with the keying used to make the Morse characters. The operation depends upon changes of frequency to convey the information, although this system usually is not called a frequency-modulation system and has been practically obsolete for some time.

Such a system will not work to transmit voice or music because it is an "all or none" arrangement. However, if a scheme is used which will increase the frequency slightly when a positive voltage is applied and decrease the frequency when a negative voltage is applied (or vice versa) this difficulty will be overcome. Moreover, the variations have to be perfectly smooth, with the frequency change following instantly the slightest change of voltage in either direction, and they must be linear—that is, if a change of 1 v changes the frequency by 1 kc, then 2 v must change the frequency by 2 kc, 3 v make a change of 3 kc, and so on.

Now speech or musical sounds have a complex wave form which by its character tells the listener whom or what he is hearing. If a voltage produced by sound striking a microphone and subsequently amplified is applied to the system just suggested, the frequency of the transmitter will increase whenever the complex audio voltage is positive and the frequency will decrease whenever the complex audio voltage is negative. It is assumed that the system will operate fast enough so that any audio frequency within the useful range will be followed without lagging. This process will alternately increase and decrease the frequency of the wave from the station. Increasing the frequency means that there will be more cycles in a given period of time and decreasing it means that fewer cycles will appear; this variation is shown in the diagram of Fig. 14–1. The process of frequency modulation treats the radio wave something like an accordion, squeezing it in and out according to the needs of the speech or music being transmitted. The corresponding picture for an amplitude-modulated wave is shown in Fig. 14–2 for comparison.

The height or amplitude of the amplitude-modulated (A.M.) wave changes according to the loudness of the sound in the microphone, the greatest permissible variation being shown in the diagram. This is called 100% modulation, which occurs when the wave is pressed down to zero during one half cycle of the audio frequency and reaches double the unmodulated amplitude during the other half cycle. It may be shown (with a measuring instrument or by mathematics) that the power in the wave increases during modulation—as much as 50% in case of 100% modulation. The amplitude of the frequency-modulation wave, on the other hand, remains the same all of the time and the total power supplied by the transmitter also remains the same, instead of increasing with increased modulation as it does in the amplitude-modulation wave.

The modulation percentage in the amplitude-modulation system is limited to 100%, because any further increase would shut the station off the air momentarily, and cause other technical troubles. In the frequency-modulation system there is a different limitation to the amount of modulation which may be applied. An increase in modulation simply means that the station occupies more of the assigned radio frequencies. If this process goes too far, the station being considered will interfere with other stations by trespassing upon their frequency range, and so the frequency swing is limited by law (for frequency-modulation broadcasting service to a channel 200 kc wide, 100 kc on each side of the carrier frequency which is in the center of the channel). In practice the present stations do not try to swing the entire 200 kc but consider a total of 150 kc, or 75 kc on each side of the carrier, to be enough. The modulation percentage in frequency-

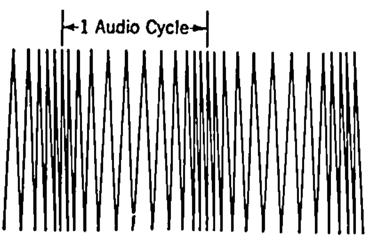


Fig. 14–1. Frequency-modulated Wave.

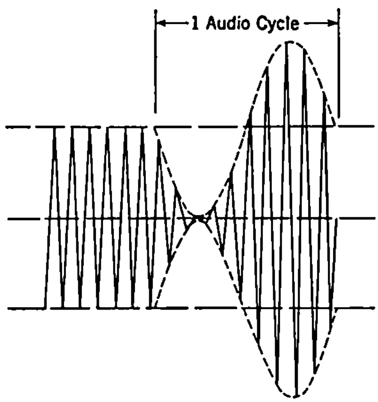


Fig. 14–2. Amplitude-modulated Wave.

modulation broadcasting is defined as the percentage of this 75 kc that the frequency actually swings through at any given time.

This frequency swing or *frequency deviation* depends upon the loudness of the sound in the microphone. It is convenient here to define another term, the *modulation index*, which is the frequency deviation divided by the audio frequency. For example, if the maximum frequency deviation from the unmodulated frequency is 75 kc and the audio frequency transmitted is 15,000 cycles (15 kc), then the modulation index is 75/15 = 5. For the highest audio frequency this is called the *deviation ratio*. If a single audio frequency of 1,500 cycles (1.5 kc) were broadcast at full modulation, the modulation index would be 75/1.5 = 50, ten times as great as before. For an audio frequency of 150 cycles (0.15 kc) the modulation index would be 500, and so on. The modulation index is large for low audio frequencies and small for high audio frequencies of the same intensity.

The modulation index also varies with the loudness of the sound in the microphone. In the first example above, when the maximum loudness of sound was acting, the frequency deviation was 75 kc. If the sound decreased to one half of this intensity the deviation would be half as much as before and the modulation index would drop to half its former value (37.5/15 = 2.5). If there were no sound in the microphone the frequency deviation would be zero, and no sound would be heard from the receiver. To sum up: for any given audio frequency the modulation index will vary directly with the sound level (or audio-frequency voltage); for any given sound level the modulation index will vary inversely as the audio frequency; and both of these effects may occur at the same time.

A suitable network may be connected in the audio-frequency channel to make the voltage of the high audio frequencies somewhat greater than their normal values; this is called *pre-emphasis* and tends to equalize the modulation index for the various audio frequencies as well as to aid in the reduction of noise.

Nearly any wave which is not a pure sine wave may be analyzed into component parts which are true sine waves. As a simple example, the blunt-topped wave shown in Fig. 14–3a is made up of a pure sine wave and another pure sine wave of three times the frequency of the first (the component waves are dotted). This wave may also, be pictured by a frequency spectrum as in Fig. 14–3b. In this spectrum the height of the line indicates the peak amplitude of each component and the horizontal

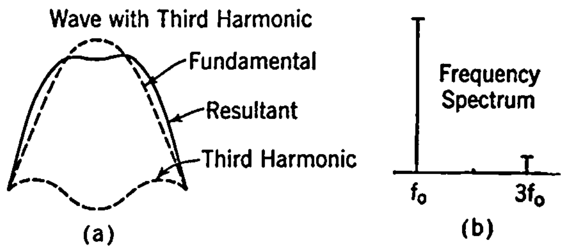


Fig. 14–3. Distorted Wave and Components.

distance their separation in terms of frequency. The amplitude of the third harmonic shown is one fifth of the height of the fundamental and the distance between them is $2f_0$, where f_0 represents the frequency of the fundamental or lowest-frequency wave.

In a frequency-modulation system the wave is alternately squeezed and stretched (as mentioned above) so that the shape of each individual cycle is no longer that of a pure sine wave but is slimmer or fatter than it would be if unchanged. Figure 14–4a shows one of the squeezed cycles and Fig. 14–4b shows the result of analyzing a frequency-modulated wave into pure sine waves and plotting the amplitudes of these sine waves on a frequency-spectrum diagram. The center line, marked f_c, indicates the carrier frequency, and the height of the line indicates the amplitude of the carrier under certain conditions. The other vertical lines indicate frequencies which are called *side bands*; there may be a great many of these. With a large modulation index the number of important side bands is greater than with a small modulation index. The examples above show that when the modulation index is large the audio frequency is low and vice versa. The side bands are spaced out from the carrier frequency at distances which are proportional to the audio frequency; that is, each of the side bands set up by a 100-cycle note will be separated from its neighbors by 100 cycles, the side bands of a 1,000-cycle note will each be separated by 1,000 cycles, and so forth. At the time when there are a large number of important side bands, then, these side bands will lie closer to the carrier frequency than when the modulation index is small (with a small number of side bands). In general, the side bands decrease in amplitude as their frequency becomes farther from that of the carrier. The result is that the station may stay within a certain assigned frequency

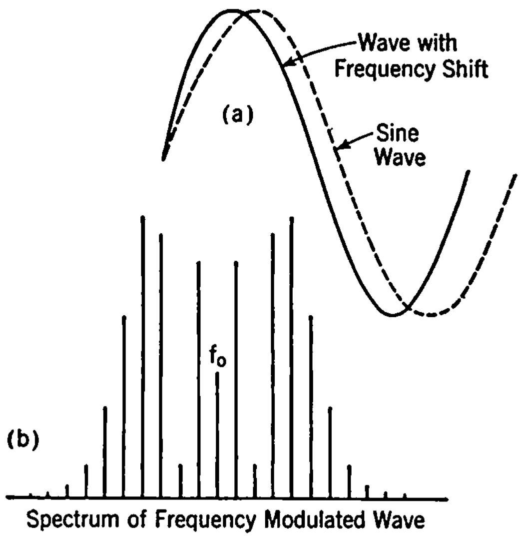


FIG. 14-4. Distorted Wave and Frequency-modulation Spectrum.

band and the side bands do not cause interference outside of the station channel.

Wide-band Frequency Modulation, an Outgrowth of the Search for Static Eliminators. Noise in the output of a radio set may be broadly defined as any sound appearing which was not present at the microphone. In case of message service, often called "communication service," this definition includes any disturbance which interferes with understanding the message coming over the air. Noise comes from many sources, such as atmospherics, lightning, magnetic storms, ignition systems, diathermy machines, interfering stations, and so on—even from the random motion of electrons in circuit components within the set itself. These disturbances are like radio signals in character, appear at all radio frequencies, and cannot be tuned out.

To overcome the effects of noise, some means must be used to enable the receiver to separate the desired signal from the noise. Some of the methods intended to effect this are the use of high power in the transmitter, increase of modulation, selective circuits at the receiver, and directional antenna systems. Each of these methods has certain drawbacks. Beside the above methods, which are effective under certain conditions, many inventions which do not work have been proposed for reducing noise.

In an amplitude-modulation system the interfering signal or noise will add directly or cumulatively to the desired signal and the result in the output of the receiver may be serious interference if the undesired signal is as small as 1% of the desired signal. With the frequency-modulation system, on the other hand, the addition of the noise and the desired signal is quite different and a condition results which is much more favorable to the reduction of noise. This fundamental difference has been proven both by mathematical analysis and by experiment. In frequency modulation, when two signals are added and the modulation index of the desired signal is fairly large while the undesired signal (either noise or radio transmission) is less than half the amplitude of the desired signal, the modulation of the sum is determined almost completely by the larger or desired signal. In fact, no appreciable disturbing effect will be noticed in the output of the receiver under these circumstances. This is a great improvement compared with the amplitude-modulation system, and is the situation observed in actual installations. This same effect leads to "blanketing" of a weak station by a stronger one, which is sometimes an advantage and occasionally a disadvantage.

Other Factors in Frequency-modulation Systems. The only portion of the radio-frequency spectrum where wide-band channels (200 kc) are available to accommodate frequency-modulation stations is among the ultrahigh frequencies. The band assigned to commercial frequency-modulation stations is from 43 to 50 mc.

The fact that the range of signals at these frequencies is about one to three times the distance from the transmitting antenna to the horizon (or less than 100 miles in most cases) means that the same carrier frequency may be used by stations separated by intervals of about 200 miles without causing serious interference. In communication or emergency work the same frequency may be used by stations much closer together or by a whole group of stations; in some such cases the blanketing effect of a strong carrier might be undesirable, if it concealed the fact that the weaker station was on the air.

If two amplitude-modulation stations operate with carrier frequencies which are very nearly the same and signals from both stations appear in a radio receiver, a squeal or heterodyne is set up. This squeal often prevents the reception of a desired signal and may be caused by an interfering station a thousand miles or more away, whose signal strength is very small compared with that of the desired station. Heterodyne effects between frequency-modulation stations on the same or very nearly the same frequency are eliminated by the limited distance covered by the signal and by the blanketing effect of the stronger of the two signals.

In broadcasting entertainment programs, the usual narrow channel has always been a barrier against transmitting the higher audio frequencies which give brilliance and life to music and speech. Frequency modulation, with a top audio frequency in commercial transmitters of 15,000 cycles, removes this limitation to high fidelity.

In amplitude-modulation systems the power output under conditions of 100% modulation must average 50% more than the power required with no modulation, and reaches peaks of power which are four times the unmodulated value. This radio-frequency power must be supplied by changes in the efficiency of the linear radio-frequency amplifiers or by the audio-frequency high-level modulator stage. With large stations this situation may lead to great engineering difficulties. In frequency-modulation systems, on the contrary, the modulator may be a receiving-type tube and the output stage, operating efficiently as a Class C amplifier, delivers the same amount of power to the antenna at all times. Changes in the modulation simply change the distribution of the power between the carrier and the various side bands.

Because of the noise-reducing properties, high fidelity, and limited interfering range of frequency modulation, it has often been proposed and sometimes used in radio relay circuits to supplement or substitute for wire transmission lines or program circuits. Frequency-modulation transmissions may have several services multiplexed, or sent simultaneously on the same carrier—such as a combination of sound and facsimile, or two sound channels to give binaural reproduction, or combinations of television picture and sound signals. In the latter case the picture may use amplitude modulation and the sound frequency modulation.

Frequency-modulation Systems for Communication Services. The primary purpose of communication or message services is to convey information; tone quality and similar factors are of minor importance. For example, a restricted audio-frequency range of, say, 250 to 3,000 cycles is found to be adequate. With this compressed audio-frequency range the frequency deviation may be considerably reduced and stations assigned to narrower channels, which would permit the simultaneous operation of more stations in a given limited area. The deviation ratio may still be large to retain the noise-reducing properties of frequency-modulation;

thus a top frequency of 3,000 cycles and a deviation ratio of 5 means a frequency swing of only 15 kc on either side of the carrier frequency or a channel width of 30 kilocycles. As a matter of fact, experiments have shown that near the limit of the useful range of frequency-modulation signals for emergency service it is better to use a small deviation ratio, which will still further reduce the channel width.

The useful sensitivity of any receiver is limited finally by the noise level. Because of their low noise level, frequency-modulation receivers are well suited for communication service. On the other hand, it is imperative that frequency-modulation receivers have very high gain, because the

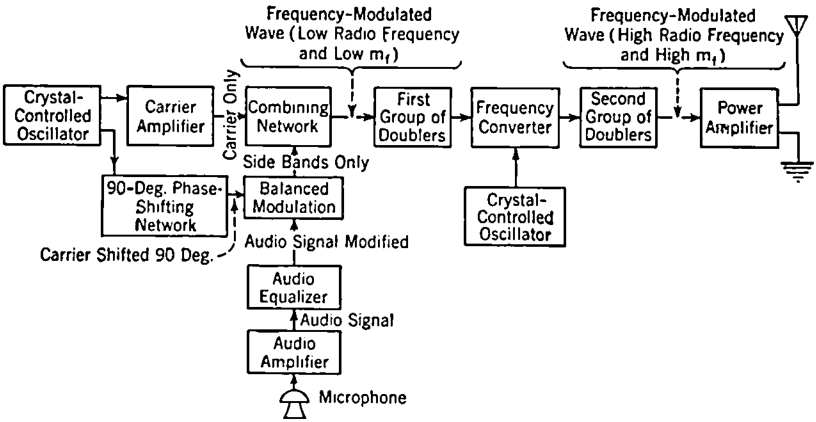


Fig. 14–5. Block Diagram of Armstrong Frequency-modulation Transmitter System.

limiter circuit (to be described later) needs a certain minimum voltage to operate properly. When no carrier is being received the frequency-modulation receiver is open to any random noise or static from the air and its high gain and broad amplifiers make it more troublesome in this respect than a comparable amplitude-modulation receiver. For communication work, where the carrier is switched on only when needed for messages, the continuous noise of the receiver becomes very annoying. Some sort of effective squelch circuit should be included to prevent any sound from coming out of the audio channel when no carrier is applied to the receiver.

Frequency-modulated Transmitters. The operation of many of the circuits in a frequency-modulated transmitter is identical with the corresponding circuit in an amplitude-modulated transmitter. There are, however, certain circuits which are unique to the frequency-modulation system and these will be described in some detail.

Two methods of obtaining frequency-modulated signals are in present use: (1) the Armstrong system, and (2) the reactance-tube or Crosby system. Other methods are possible but have not been widely used.

In the Armstrong system of frequency modulation, the stages shown in the block diagram, Fig. 14–5, are used.

This transmitter has an oscillator whose frequency is held at a constant value by means of a quartz crystal at about 200 kc. This constant-frequency wave is sent through two channels; one channel consists of a radio-frequency amplifier which builds up the amplitude of the wave, and the other channel includes a phase-shifting network and a balanced modulator. The balanced modulator is similar in most respects to the standard modulated stage used in amplitude-modulation transmitters, whose output can be analyzed into a carrier and a pair of side bands for each audio-frequency component in the signal from the microphone. However, in this case the carrier is not wanted and is removed by the action of this modulator. The phase-shifting network is needed because it has been found that if the side bands set up by this balanced modulator are added to the original carrier after a 90° phase shift frequency modulation may be produced. The deviation ratio is small and must be multiplied to reach a useful amount.

The output of the combining network is fed into a series of frequency-doubling stages. These consist of Class C radio-frequency amplifiers whose plate circuit is tuned to double the frequency of the grid circuit; the radio frequency applied to the grid is thus multiplied by two and the deviation ratio is also multiplied by two.

In frequency-modulation transmitters for broadcasting service it is found that a continuation of this multiplying process does not give a sufficiently large deviation before the carrier frequency becomes too high to be used. A frequency converter is then used to lower the carrier frequency without affecting the deviation ratio. This converter, which is the same sort used in a superheterodyne receiver, is supplied with a fixed frequency from another crystal-controlled oscillator.

The output of the converter is passed through more doubler stages to increase the deviation and carrier frequency still further until the correct carrier frequency is obtained. All of these operations may be carried out using receiver-type tubes. The output, with high carrier frequency and high deviation, may then be applied to as many Class C power-amplifier stages as are necessary to build up the power output of the transmitter to any amount desired.

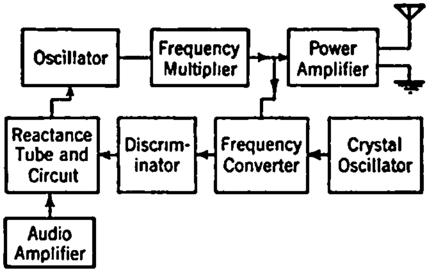


Fig. 14–6. Reactance-tube or Crosby System Frequency-modulation Transmitter.

In this system the modulation index is directly proportional to the signal voltage. The audio frequency supplied to the modulator must be passed through a network which makes the audio amplitude and hence the modulation index inversely proportional to the audio frequency.

The reactance-tube system is shown in block diagram Fig. 14-6.

In this system the oscillator is not crystal-controlled but self-excited. Another tube, the reactance tube, is connected in parallel with the tank circuit of the oscillator. The reactance tube is supplied with the usual plate voltage and cathode-bias voltage as shown in Fig. 14-7. In addition,

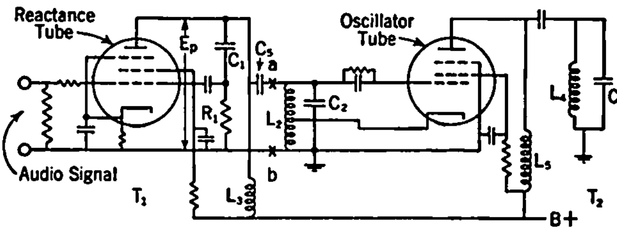


FIG. 14-7. Circuit Diagram of Oscillator and Reactance Tube.

its plate is supplied with alternating voltage from the tank circuit of the oscillator through condenser C_5. By means of a suitable circuit, as shown, this tube is caused to act like a reactance (coil or condenser) and its reactance is varied in accordance with the audio frequency. The frequency of the oscillator is also changed because of the changing reactance connected across its tank circuit, and a frequency-modulated signal appears in its output.

This carrier with frequency modulation is passed through a number of frequency doublers or multiplier stages to increase the carrier frequency and deviation ratio. Finally, power amplifiers step up the power and feed the signal into the antenna.

Some means must be provided to keep the transmitter upon its assigned carrier frequency. A self-excited oscillator, such as that used here, would not hold closely enough to a particular frequency to be satisfactory in this service. Stability is obtained by comparing the output of the transmitter with a standard, crystal-controlled oscillator and feeding back a suitable correcting voltage from a frequency converter and discriminator (to be explained later) to keep the self-excited oscillator on the correct frequency.

Frequency-modulation Receivers. The typical frequency-modulation receiver is a superheterodyne, conventional in many respects but with certain special features. Among the latter are (a) a *limiter* to remove any amplitude modulation, including noise, appearing along with the desired frequency-modulation signal; (b) a *discriminator* circuit to change the frequency variations into audio-frequency voltages.

In entertainment receivers, the audio-frequency channel and especially the loudspeaker should be of the highest quality to take advantage of the high fidelity possible in frequency-modulation systems. This high quality may require the use of special loudspeakers for the high audio frequencies in addition to the usual speaker for low and medium frequencies.

The *pre-emphasis* at the transmitter to increase modulation at the higher audio frequencies must be corrected with a *de-emphasis* circuit in the re-

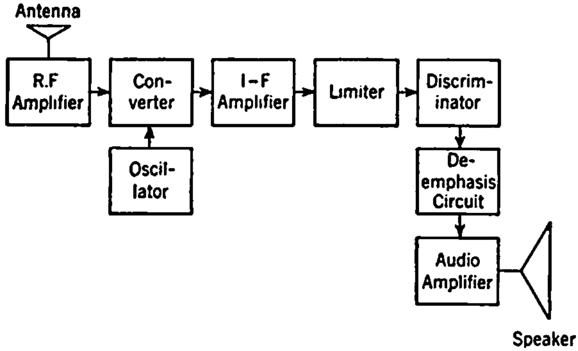


FIG. 14–8. Block Diagram of Typical Frequency-modulation Receiver.

ceiver to restore the correct balance of tone. This circuit may consist of a resistor and condenser in series whose time constant is about 100 microseconds (such as a 100,000-ohm resistor and a 0.001 μf condenser).

A block diagram of a typical frequency-modulation receiver is shown in Fig. 14–8.

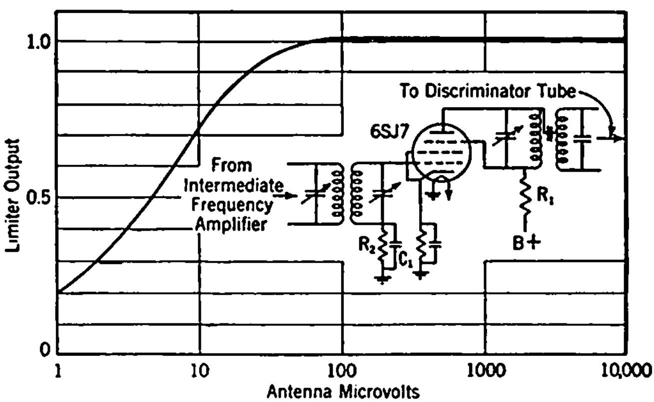


FIG. 14–9. Circuit of Limiter Stage and Curve Showing Limiter Action.

The circuit for a typical limiter stage and its action are shown in Fig. 14–9. The limiter tube operates with a low plate voltage (40 to 80) so that it will reach its maximum output with a relatively low signal voltage. The circuit operates like a Class C amplifier, and even though large signals are applied to the input of the set, the output of the limiter remains practically constant. Hence amplitude modulation arising from any cause is wiped off the signal; the limiter does not interfere with the frequency modulation because its action does not affect the instantaneous frequency of the wave.

The circuit diagram of a typical frequency-modulation detector (or discriminator) is shown in Fig. 14–10. When a signal is applied which is

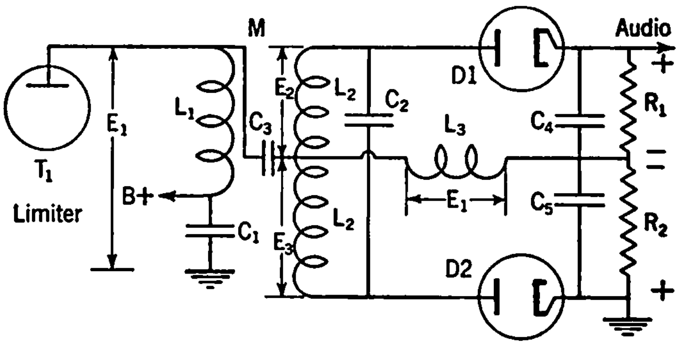


Fig. 14–10. Frequency-modulation Detector or Discriminator.

exactly the same frequency as that to which the secondary of the intermediate-frequency transformer is tuned the voltage across each diode is the same and no audio-frequency voltage appears. If the applied frequency is varying about this same frequency, an audio-frequency voltage appears at the audio terminals and is amplified with the audio-frequency amplifier of the set.

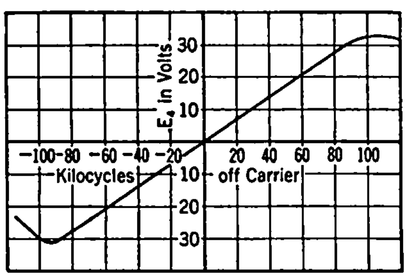


Fig. 14–11. Frequency Variation Plotted against Audio-frequency Output in a Discriminator.

A typical curve showing the relation of the applied frequency to the resulting audio voltage is that of Fig. 14–11.

Another method of converting frequency-modulation signals is to apply the incoming signal to a circuit tuned to a slightly different frequency from that of the signal, as shown in Fig. 14–12. A higher frequency will cause an increase in the current in the tuned circuit and a lower frequency will cause a corresponding decrease in current. This current change converts the variation in frequency into a variation in the amplitude of a current, which may then be passed on to a standard amplitude-modulation type of detector. This effect explains the fact that frequency-modulation signals may be brought in with amplitude-modulation receivers under certain conditions—although the quality is usually very poor owing to nonlinear action of the circuits.

The circuit diagram of a typical frequency-modulation receiver (which also includes an amplitude-modulation receiver) is shown in Fig. 14–13.

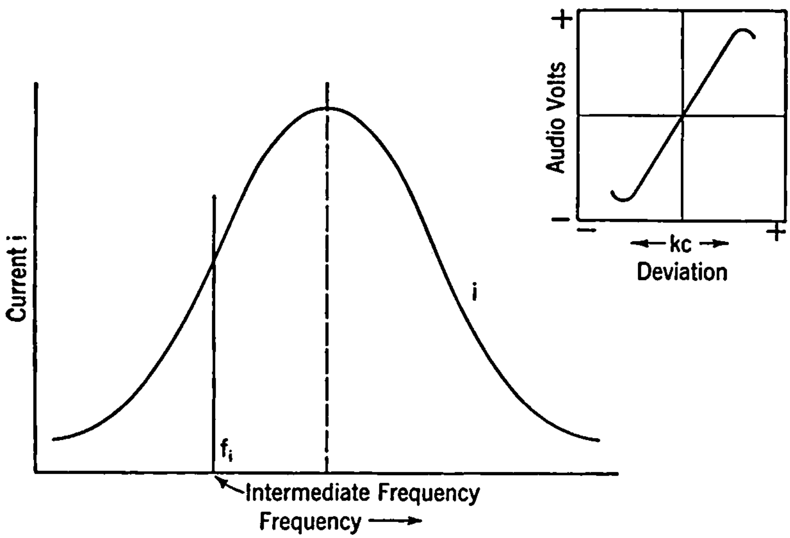


Fig. 14–12. Off-resonance Tuning to Convert Frequency-modulation Signals into Amplitude-modulation Signals.

Review Questions and Problems. *1.* Point out the main differences between a frequency-modulation and an amplitude-modulation radio system.

2. What are some of the reasons for the noise-reducing properties of frequency-modulation systems?

3. Why is a limiter necessary in a frequency-modulation receiver?

4. What is the frequency-band width for frequency-modulation transmission at present?

5. Compare the variations of output power when modulation level is changed in a frequency-modulation system and then in an amplitude-modulation system.

6. What is the ratio of band width to mid-frequency in the intermediate-frequency amplifiers of a frequency-modulation system as compared with an amplitude-modulation system?

7. Why is it not practical to have frequency-modulation stations in the regular broadcast band?

8. In an entertainment receiver using frequency modulation, why is it desirable to have the best possible audio amplifier and loudspeaker?

9. Why is it practical to use a narrow band of audio frequencies in communication service?

10. What advantages would the frequency-modulation system offer in mobile service?

CHAPTER 15

Radio Wave Propagation

General Nature of Propagation. When a radio wave leaves an antenna it spreads out in all directions as indicated in Fig. 15–1. Part of the radiated energy travels along near the ground and is *guided* around the surface of the earth much as electromagnetic waves are guided by wires. This portion of the radiation is called the *ground wave* or *surface wave.* The remainder of the energy is called the *sky wave* or *space wave.* It is radiated upward into space and would be lost completely were it not for reflecting

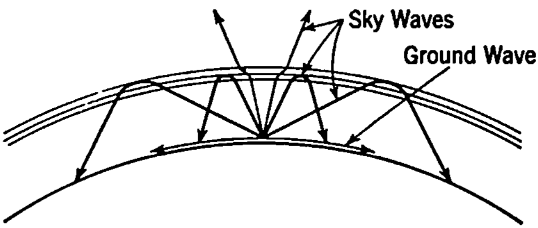


FIG. 15–1. Ground-wave and Sky-wave Radiations.

layers of ions and electrons that exist some 30 to 250 miles above the earth's surface. These ionized layers can reflect or refract a portion of the incident radiation back towards the earth and so produce a signal at distant points.

Polarization. If the signal radiated by a vertical antenna is received a short distance away it will be found that the received signal is a maximum when the receiving antenna is vertical and zero when the receiving antenna is horizontal. This is because the electric field is in the vertical direction at this point. The radiation from the antenna is said to be *vertically polarized.* In a direction at some vertical angle θ to the sending antenna (see Fig. 9–16) the receiving antenna must lie in the direction shown by E for maximum signal pickup. It is then not vertical but it will still be in the *vertical plane* through the antenna and the point P, and there will be zero voltage induced in an antenna perpendicular to this plane. The radiation from the antenna is still said to be vertically polarized, or more correctly, *plane polarized in the vertical plane.* Similarly the radiation from a horizontal antenna is said to be *horizontally polarized* and there will be no direct pickup of such a signal on a vertical antenna.

The Ground Wave. The part of the radiated energy that travels along near the ground, that is, the ground wave, induces voltages and currents in the ground that subtract energy from the wave. If the ground were a perfect conductor, these ground currents could flow without any losses and the wave would not be affected. However, the ground does have resistance, or a *finite conductivity*, so that energy is required to make these ground currents flow and this energy is absorbed from the wave. The result is that the ground wave is thus *attenuated* or decreased in strength even more than by the distance factor $1/r$ which is due to the spreading out of the wave through larger surfaces as it recedes from the antenna. The amount by which the wave is attenuated due to an imperfectly conducting ground is important in determining how far the wave will travel before the signal becomes too weak to be of any use.

The attenuation due to the ground depends both upon the conductivity (or resistance) of the ground and upon the frequency being used. A high-

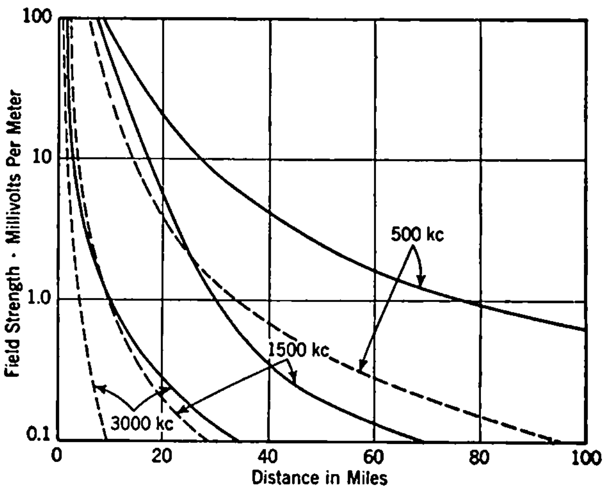


FIG. 15–2. Effect of Frequency and Ground Conductivity upon the Strength of the Ground-wave Signal. Solid lines: fairly good conductivity (100×10^{-15} e.m.u.). Dashed lines: poor conductivity (20×10^{-15} e.m.u.).

frequency wave is attenuated much more than a low-frequency wave over the same ground.

Figure 15–2 illustrates the manner in which the ground-wave signal depends upon frequency and upon the conductivity of the ground. The curves show the field strength in millivolts per meter against distance from the transmitter. The expression millivolts per meter (or microvolts per meter) refers to the voltage that would be induced in a wire 1 m long when placed parallel to the direction of the field at that point. A field strength of 0.1 mv per meter is about the lowest signal strength that will give satisfactory reception, although this depends upon the amount of noise present in the receiving location. A noisy location will require as high as 10 or 100 mv per meter for worthwhile reception, while in a very quiet locality, at a time when static is weak, a signal of 10 μv per meter may be entirely adequate.

The solid curves in Fig. 15–2 are for a fairly good ground conductivity while the dashed curves are for a poor ground. Ground conductivity is usually expressed in electromagnetic units (e.m.u.); a conductivity of 100×10^{-15} e.m.u. represents a fairly good ground while a conductivity of 20×10^{-15} is considered a very poor ground. The ground conductivity seems to depend to a large extent upon the nature of the terrain. Flat prairie country usually shows a high value of conductivity while mountainous or rugged, broken country has a low conductivity.

Because the ground wave is attenuated so much at the higher frequencies its chief usefulness lies in the long-wave and broadcast bands. Daytime reception of broadcast stations is entirely by means of the ground wave.

The ground wave is always vertically polarized because any horizontal component of electric force would be shorted out by the ground. For this reason vertical antennas must be used for ground-wave transmission.

The Sky Wave. The energy radiated upwards by an antenna, that is, the sky wave, would be wasted energy as far as radio communication is concerned if it continued on its path and did not return to earth. Fortunately, under certain circumstances it is reflected from the *ionosphere*, or *Kennelly-Heaviside layer* as it used to be called. The reflected wave may return to earth at distances from the antenna much greater than can be reached by the ground wave, and this reflected wave makes extreme long-distance communication possible.

The Ionosphere. The ionosphere consists of several ionized layers, that is, layers which are electrically conducting. These layers exist at high altitudes, in the upper parts of the earth's atmosphere. Radio waves that strike these conducting layers have their paths changed while passing through the layers. Often the waves penetrate all the layers and are lost, but more often the waves are bent in their paths so much that they return to earth at distant points. The heights of the layers and the *degree of ionization* (that is, the number of ions and electrons in a given volume) determine how far radio waves will go, and what frequencies give the best transmission. The ionized layers are found usually at heights between 50 km (30 miles) and 400 km (250 miles) above the surface of the earth.

In Chapter 4 it was shown that in a gas under very low pressure it is possible to knock one or more electrons out of a molecule of the gas, leaving a positive charge on the molecule. The positively charged molecule is no longer a true molecule, but is an *ion*, and can be attracted or repelled by electric forces. Electrons can be knocked out of a molecule not only by fast moving particles like electrons, but also by certain types of radiation such as ultraviolet rays and cosmic rays. In the high atmosphere, where the pressure is low, conditions are excellent for ionization to take place. The sun constantly gives off ultraviolet rays, and when these reach the upper atmosphere they cause a large proportion of the air particles to become ionized. Cosmic rays are believed to cause some ionization also.

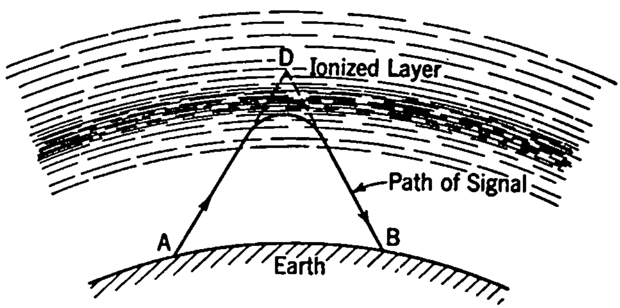


FIG. 15–3. Refraction of the Sky Wave by an Ionized Layer.

Since the atoms, ions, and electrons in a gas are in constant motion, frequent collisions take place between them. When a positive ion collides with an electron, it may keep the electron to neutralize its charge so that it once more becomes a molecule. This process of *recombination* goes on all the time, so that a molecule, once it has been ionized, does not remain ionized indefinitely. The time that it takes for recombination to occur will depend on several factors, but particularly on the average distance between the particles in the gas. If there are only a few particles present, as high in the upper atmosphere, collisions will not occur very frequently, so that the air particles remain ionized for long periods. In the lower parts of the earth's atmosphere, collisions take place so often that the air molecules do not remain ionized for very long. Another reason why there is relatively little ionization in the lower atmosphere is that the ultraviolet rays from the sun are largely absorbed by the upper parts of the atmosphere. As a result there is very little ionization below about 30 miles. Above 250 miles there are so few air particles present to be ionized that the density of ionization is again very low. However, at intermediate heights there is considerable ionization. The region between 30 and 250 miles above the earth is thus the region where the ionosphere exists, and this region therefore has the most influence on the propagation of the sky waves.

Sky waves that return to earth from the ionosphere are found to come from different heights above the earth, depending on the frequency and on the time of reflection. This phenomenon shows that the ionosphere is not one layer, but several layers. The reason why there are several layers in the ionosphere is because the different gases in the earth's atmosphere ionize at different pressures, that is, at different heights above the surface of the earth, and because there are other ionizing agents (cosmic rays, for example), which penetrate to different depths. The number of layers, their heights above the earth, and the amount they bend the sky wave all vary from day to day, from month to month, and from year to year. There are two principal layers, called the *E* layer and the *F* layer. The *E* layer is usually found at a height of 110 km (68 miles), but may vary from 90 to 140 km (55 to 85 miles). The other principal layer is one layer only at night, but splits into two parts during the daytime. The designation *F* layer is used to refer to the night layer. The designations F_1 and F_2 are given to the two parts of the layer which exist in the daytime, the F_1 layer being the lower one. The *F*, F_1, and F_2 layers are always above the *E* layer. Another layer, the *D* layer, exists only in the daytime at very low heights, but its effects are not as important as those of the other layers, and so will not be considered further. The heights at which the various layers exist is shown in Table I.

TABLE 15–1

Name of layer	*Height of layer, miles*
E	55 to 85
F (night only)	110 to 250
F_1 (daytime only)	85 to 155
F_2 (summer day)	155 to 220
F_2 (winter day)	90 to 185
D	30 to 55

Effect of the Ionosphere on the Sky Wave. The way in which radio waves are bent by the ionized layers in the ionosphere may be seen with the aid of Fig. 15–3. In this figure the ionosphere is shown as one layer, the dashes representing ions and the number of dashes in any region indicating the density of ionization in that region. Suppose a wave is sent upwards along the path shown, from the transmitter at point *A*. The path is a straight line until it reaches a region where there are ions. The path of the signal now becomes bent, and bends more and more as it gets into regions of higher and higher ion density. The wave is always bent *away* from regions of high density to regions of low density. If the signal is bent sufficiently in the layer, it finally emerges and returns to earth as shown.

The actual path of the wave in the ionized layer is a curve, as shown, and it is said to be caused by *refraction* of the wave. This is similar to the refraction of light that takes place in a prism. Since it is usually simpler and more convenient to think of the wave as being *reflected*, rather than refracted, the path can be assumed to be the straight lines *AD* and *DB* as indicated in the figure. This assumption is made in measurements of the height of a layer. To measure the height of an ionized layer, a wave is sent out from a transmitter at A and the time taken by the sky wave traveling over the path AD–DB to reach the receiver is compared with the time taken by the ground wave along the direct path AB. From this information, and knowing the distance AB, it is possible to calculate the height of D above the earth. This height is called the *virtual height* of the ionosphere, since it is not the true height. In order to measure the true height of the layer it would be necessary to know the shape of the curved path. In measuring virtual heights the points A and B are usually placed very close together so that the wave is sent nearly vertically upward.

If the frequency of the transmitted wave in Fig. 15-3 is increased sufficiently it will be found that a point is reached beyond which the wave is no longer reflected back to earth. The path taken by the ray in this case is shown in Fig. 15-4. The path is straight until it reaches the ionized region, when it again is bent, but not as much as before. This is because

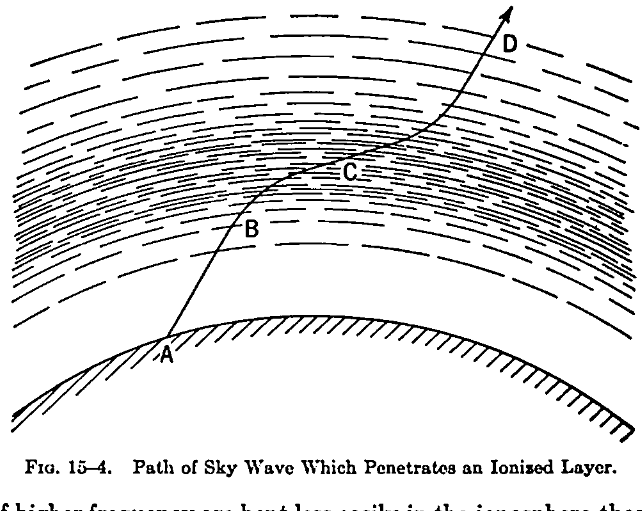


FIG. 15-4. Path of Sky Wave Which Penetrates an Ionized Layer.

waves of higher frequency are bent less easily in the ionosphere than waves of low frequency. The path is bent *away* from the region of high ion density to that of low density, as shown from points B to C. At C the wave passes through the region of highest ion density and is again bent away from this region to a region of low density. As a result, the path is now curved in the opposite direction and the wave is not bent back towards the earth. The wave emerges from the layer and is lost unless it should be reflected from another higher layer of greater ion density. Whether a wave such as that just considered is reflected from a layer or whether it penetrates the layer depends on the frequency, on the density of the layer, and on the angle at which the wave first strikes the layer. In the example of Fig. 15–4, if the density of ionization in the layer should increase for some reason, it may happen that the wave path will be bent enough to make the ray return to earth, so that higher frequencies would have to be used to make the ray penetrate the layer.

Now suppose that a wave of the same frequency as the signal in Fig. 15–4 is sent out from the transmitter along a path such as AFB (Fig. 15–5) which makes a lower angle with the ground. When the wave strikes the ionized

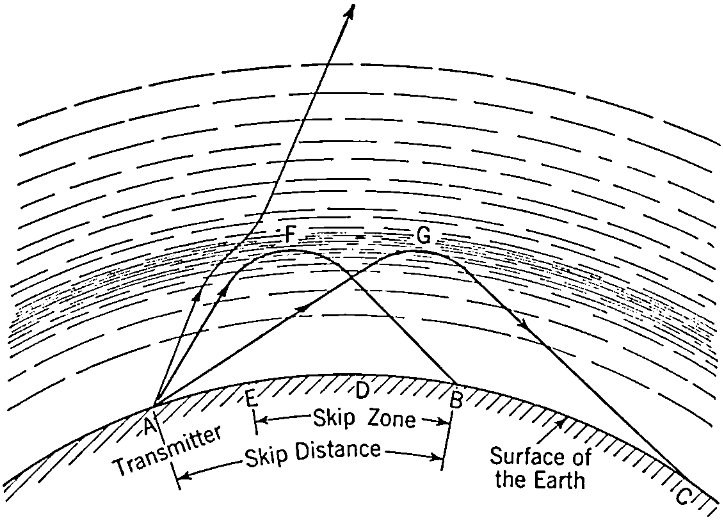


FIG. 15–5. Effects on Transmission Due to the Angle at Which the Sky Wave Is Transmitted.

layer, its path begins to bend as before. The amount which the wave is bent is now greater than before, since the wave spends a longer time in the ionized region owing to the low angle at which it is traveling. Thus it may happen that if the angle is low enough, the wave can become bent enough to return the wave to earth again, even though waves of this same frequency would penetrate the layer at steeper angles. Waves sent out at all angles less than that of the wave *AFB* shown in Fig. 15–5, on the same frequency, will be bent back to earth. Thus there will be an area, say from *B* to *C*, in which it is possible to receive the sky wave from the transmitter at *A*. If a receiver is located at a point nearer to the transmitter, as at *D*, no sky wave will be received, and unless the ground wave is strong enough, as at *E*, it will be impossible to receive the signals. There will be therefore an area from *E* to *D* in which it is impossible to receive signals for the transmitter, even though points farther away are able to receive. This area of no signal is known as the *skip area*, or *skip zone*, and the distance *AB* from the transmitter to the point where the sky wave first can be received is known as the *skip distance*.

When a wave is returned to the earth as at *B* or *C* in Fig. 15–5 it can be reflected from the earth, since the earth also is partially conducting. This reflected wave from the earth will then strike the ionosphere and be reflected once more, returning to earth at a great distance from the transmitter. This type of path is known as *two-hop* transmission as compared with the *one-hop* transmission in Fig. 15–3 and 15–5.

Critical Frequencies. It was mentioned in connection with Fig. 15–3 that it is possible to measure the virtual heights of the layers in the ionosphere by sending signals vertically upward and measuring the time taken by the signals to return. Since the speed that radio waves travel is known (186,000 miles per second), it is easy to calculate the distance the wave has traveled. Because the time taken is extremely short, a few thousandths of a second, the signal that is sent up must be a very short pulse in order that the ground wave signal and the reflected signal may be separated. An oscilloscope is used to observe the two signals for measuring the time difference. If such measurements of virtual height of a layer are made at successively increasing frequencies it will be found (as mentioned earlier) that a frequency will be reached for which the waves are no longer reflected back to earth from this layer. The highest frequency for which waves sent vertically upward are returned by a layer is called the *critical frequency*

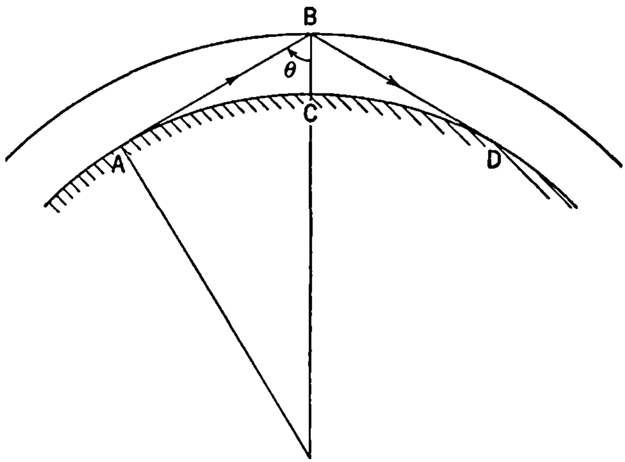


FIG. 15–6. Transmission by Means of Ionosphere Reflection, Showing Angle θ.

for the layer at the time and place the measurements are made. The critical frequency for a particular layer is *not* the highest frequency that can be used for communication using that layer. As shown in Fig. 15–5, it is only necessary to decrease the angle that the path of the wave makes with the earth in order to have a wave of higher frequency reflected by the layer. For communication between two fixed points the angle the path of the wave makes with the layer depends upon the height of the layer and the distance between the points (see Fig. 15–6). For a given layer height BC there will be a particular angle θ corresponding to each distance AD between transmitting and receiving points. It has been found that the maximum frequency that can be used for sky-wave communication between two such points is given by

$$f_{max.} = \frac{f_c}{\cos\theta}$$

where f_c is the critical frequency at the point of reflection and θ is the angle between the ray and the vertical. This maximum frequency that can be used for transmission between two points is called the *maximum usable frequency.*

As the distance between transmitting and receiving points is increased a limit occurs where, owing to the curvature of the earth, the path of the wave is tangent to the surface of the earth at these points. The angle θ corresponding to this limiting distance is about 74° for the F layer. For this case the maximum usable frequency will be

$$\frac{1}{\cos 74°} = 3.6f_c$$

and this will be the *maximum* frequency that will be reflected back to earth.

If measurements of the virtual height of the ionosphere are made at frequencies higher than the critical frequency for the lowest layer, it is found that the waves are still returned to earth, but from a greater height. This difference shows that the waves are penetrating the lower layer and are being reflected from a higher layer. As the frequency is increased more, critical frequencies for the higher layers will be found; finally a frequency will be reached for which the waves sent vertically upward are no longer returned to earth, showing that the waves are penetrating all the layers. Exceptions to this behavior occur occasionally when *sporadic E* reflections are present. Owing to reflections at a boundary caused by sharp changes in ionization density, strong reflections from a layer at the height of the E layer sometimes occur at frequencies considerably in excess of the normal critical frequencies for the E layer.

Absorption in the Ionosphere. In addition to the virtual heights and critical frequencies for each of the layers, the attenuation or absorption of energy from the waves by the ionosphere is an important factor in limiting radio transmission over large distances. When a radio wave passes through an ionized region, it causes the electrons to vibrate. The vibrating electrons collide with neighboring molecules and ions and give up some or all of their energy. This energy is used up in heating the air and is thus wasted. The amount of energy that will be taken from a radio wave and wasted in this way will be greater, the greater the distance the wave travels in the ionized region and the greater the density of the ions and air molecules in the layer. Since ultraviolet rays from the sun cause ionization to be present at lower levels in the daytime than at night, and since there are more air particles at the low altitudes, the absorption of energy will be much greater during the daytime than during nighttime, the absorption occurring mostly in the D and E layers. On account of a sort of resonance condition for the electrons moving in the earth's magnetic field at 1,400 kc, the maximum absorption occurs at this frequency. The further the frequency of a wave is from this resonance frequency, the less the attenuation. For long-distance transmission, frequencies near the maximum usable frequency for the distance are most desirable.

Regular Variations in the Ionosphere. The characteristics of the ionosphere go through regular variations which affect the propagation of radio waves. These variations can be predicted with fair accuracy. They are

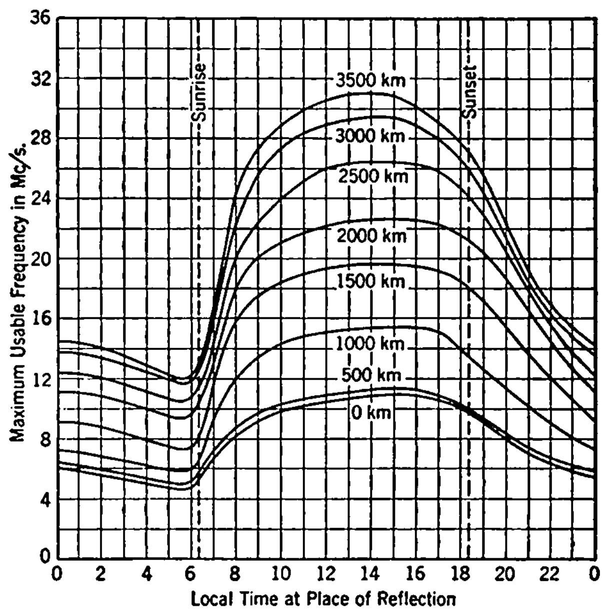


Fig. 15-7. Graph of Maximum Usable Frequency.

of three principal types, called diurnal variations, seasonal variations, and sunspot-cycle (11-year) variations. These changes in the ionosphere are due largely to changes in the radiation from the sun, so that they are mostly changes in ion density in the layers rather than in the virtual heights of the layers. With the exception of the F_2 layer, the heights of the layers show only moderate fluctuations. Therefore the variations in the ionosphere are mostly exhibited as changes in absorption and in critical frequencies.

Changes in the *E* layer are particularly regular from day to day and from season to season, and depend almost entirely on the position of the sun in the sky. When the sun is directly overhead, the ionization density is highest, as shown by a high critical frequency for this layer. The critical frequency is thus higher during the day than it is at night, and higher in summer than in winter. Variations in the critical frequency for the *E* layer occur from year to year owing to the 11-year sunspot cycle. The density of ionization in the ionosphere seems to vary with sunspot activity, and is greatest during the most active sunspot periods. Thus the critical frequency and the absorption for the *E* layer are lowest during a minimum in sunspot activity.

The regular changes in the ionosphere can be predicted fairly accurately. The diurnal and seasonal variations follow regular patterns, so that when allowance is made for the changes brought about by the 11-year sunspot cycle, predictions are readily made. The Bureau of Standards publishes

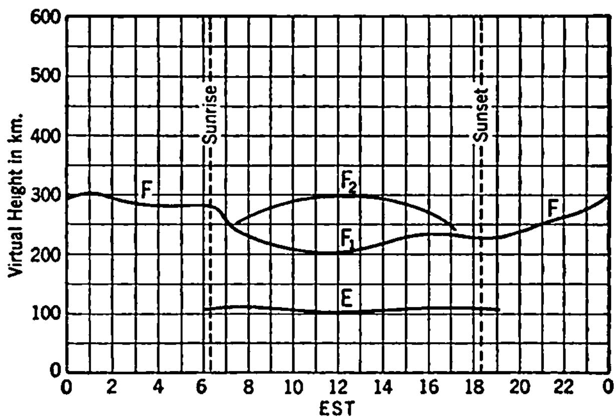


Fig. 15–8. Graph of Virtual Heights.

each month ionosphere data including predicted maximum usable frequencies for the ensuing month. Figure 15–7 shows a sample graph similar to those published. The graphs show the predicted maximum usable frequency which may be used for transmission over any given distance with one-hop transmission. By making use of these graphs, communication services can lay out ahead of time a schedule of the best frequencies (from among those they have available) to use for communicating over particular distances at various times of the day. These predictions of maximum usable frequency are based upon measurements of virtual heights and critical frequencies made regularly by the Bureau of Standards in Washington. These measurements are also published in graph form (see Fig. 15–8) along with predictions of maximum usable frequency.

Fading. Fading of radio waves is the name given to undesirable changes in the intensity or loudness of the waves at the receiving point, and is caused by variations in the height and density of ionization in the layers of the ionosphere. To see why fading occurs, consider Fig. 15–9. A transmitter at point *A* is sending out signals at various angles above the surface of the earth. One signal follows the path *ABC* and is the ground wave from the transmitter. Another wave follows a path such as *AEG*, being reflected from the *E* layer (which is shown as a line for simplicity) and received at *G*, but not at *C*. A wave sent out at a higher angle follows

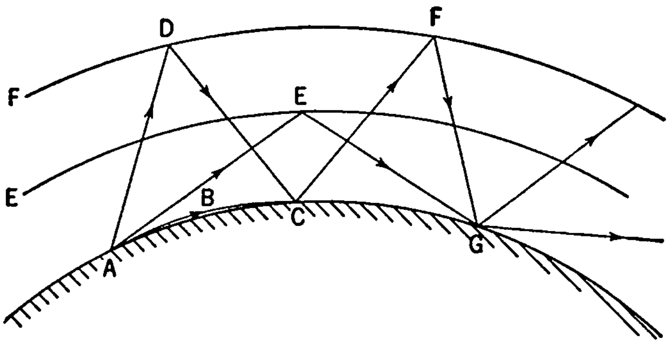


FIG. 15–9. Fading Due to Interference between Waves Which Have Traveled Different Distances.

the path *ADC*, since it penetrates the *E* layer but is reflected by the *F* layer and is received at points *C* and *G*. At point *C* the received signal is the result of two waves, one a ground wave and the other a sky wave, which have reached that point by traveling different paths. Depending on the differences in the lengths of the paths followed by these two waves, they will add together in phase, giving a loud signal, or they will add out of phase giving a very weak signal. The difference in path will depend greatly on the height of the *F* layer in this case, so that small changes in its height may change the two signals from the in-phase condition to the out-of-phase condition. This change of phase causes large variations in the received signal strength; that is, the signal *fades*.

In the above illustration it was shown that fading at point *C* is the result of interference between the ground wave and a sky wave. Fading can also occur as a result of interference between two sky waves, as for example at the point *G*. Here the received signal is the result of two sky waves which have traveled different paths so that changes in either or both layers can cause fading. One signal arrives at *G* by one hop and the other by two hops.

All modulated signals consist of a band of frequencies, not just one frequency, the width of the band depending on the type of signal. Because of the difference between the component frequencies, only part of the signal may fade at any given time, so that, for example, one part of a side band fades independently of another part, giving a peculiar form of distortion in the audio signal, which is known as *selective fading.*

Fading in a received signal may take place very slowly or it may take place quite rapidly, since the ionosphere varies from a number of causes, some slow and some rapid. It was pointed out above that conditions in the ionosphere go through regular variations from day to day, season to season and year to year. Variations in signals due to these variations are not usually thought of as fading because they take place so slowly. However, in addition to these regular changes in the ionosphere, other more or less irregular changes take place and cause severe fading.

One of the most startling of the irregular variations in the ionosphere is that known as a *radio fadeout.* A radio fadeout is the result of a sudden burst of ionizing radiation from the sun which causes the ionization in the *D* layer to increase suddenly, which in turn greatly increases the absorption of sky waves of all frequencies. The effect on radio transmission is the sudden fadeout of all signals on frequencies above about 1.5 mc. The drop in signal strength occurs suddenly and lasts from about ten minutes to an hour or more.

Another and important irregular change in the ionosphere is known as an ionosphere storm. During such a storm the ionosphere becomes quite unstable in its effects on radio waves, causing signals on frequencies above about 1.5 mc to drop in level and fade badly. A type of fading known as *flutter fading* takes place, especially at night. The effect of these storms is usually to weaken the sky wave on the broadcast band at night, but sometimes it is increased in strength. Ionosphere storms may last from one or two days on the high frequencies to several weeks on the low frequencies. Radio communication is very erratic during these storms.

Reduction of Fading. There are several ways in which fading can be reduced so that usable signals can be received. The most common means employed is *automatic volume control* (a.v.c.) in the receiver, described in Chapter 13. In this system, the strength of the carrier of the signal being received is used to control the volume of the receiver so that the output is held reasonably constant in spite of variations in the signal strength. Automatic volume control is not a complete solution to the problem of fading, since the signals often drop so much that they are below the noise and no amount of amplification in the receiver will make the signal usable. Automatic volume control cannot help selective fading since components of the same signal fade out at different times.

One of the best means for reducing fading is known as *diversity reception.* It has been found that fading does not usually occur at two different places at exactly the same time. Hence, if two or more receiving antennas are used, and spaced several wavelengths apart, it is unlikely that the signals will fade out completely on all of them at the same time, so that if the signals from all of them are combined there will nearly always be some signal present. Complicated receivers must be used with diversity receiving systems since it is necessary to combine the outputs of the antennas in the correct phase relation to give a large signal. Another type of diversity system makes use of the fact that signals received on vertically and horizontally polarized antennas usually do not fade at the same time.

Telegraph transmitters sometimes employ *frequency diversity* systems, which are based on the fact that signals spaced even as little as 500 to 1,000 cycles apart fade independently (selective fading). This is done by using 500- or 1,000-cycle modulation on the carrier, and keying this modulated carrier. The modulation on the carrier produces a side band on each side of the carrier frequency; each side band may be considered to be another carrier, which conveys the same signal when keyed. Since fading is usually selective as regards frequency, such a telegraph signal is less affected by fading than a single unmodulated keyed carrier.

Static and Man-made Noise. The output of a receiver which is tuned to a relatively weak signal usually contains some noise in addition to the desired signal. This noise may come from any one of several different sources, natural and man-made. Noise in a receiver is usually the limiting factor in determining the lowest signal strength that can be used for communication or broadcasting purposes.

STATIC. Noise picked up by the antenna is referred to as *static* or *atmospherics* when it is due to natural causes. Static is caused by natural electrical disturbances, principally thunderstorms, and its energy is found distributed throughout most of the frequency range used in radio. The energy in static decreases as the frequency is increased, so that most static is found at low frequencies. Static is relatively unimportant at ultrahigh frequencies. Since static is really a radio signal produced by nature, it is propagated the same way radio waves are, and is reflected by the ionosphere under suitable conditions. Static impulses may therefore travel great distances under the right conditions, and cause interference in a receiver at a great distance from its origin. A large part of the static heard at any particular place comes from a considerable distance, and the rest is from local thunderstorms and the like.

Since the sky wave on the broadcast band and lower frequencies is greatly attenuated in the daytime, very little daytime static on these bands comes from great distances. Most of the daytime static on these frequencies is due to local thunderstorms. At night, however, the sky wave propagates with less attenuation, so that the static noise level is usually greater than in the daytime. In the short-wave region, the noise level due to static is much less than it is at lower frequencies; the higher the frequency, the less the static. Short-wave static can travel great distances in the daytime with very little attenuation, so the static heard at any particular place may have come from some distant point.

In the frequency range from about 9 to 21 mc it has been found that even when there is no ordinary static or man-made interference, some noise is still picked up by the antenna. It is believed that this noise comes from some region in space, probably in the Milky Way. This form of noise is often the limiting noise in this frequency range at a good receiving location.

Above about 30 mc no energy is reflected from the ionosphere and so distant static has no effect. Normally there is no static interference in this range except during local thunderstorms.

Man-made noise. Man-made noise is generated by most electrical applicances and electrically-operated devices. Ignition systems, diathermy machines, power-line discharges, sparking brushes on motors and generators all can cause interference with radio reception. In fact, almost any device that produces an electrical spark can interfere with reception. Such noises, once produced by the device, are carried by the power lines connected to the device and are either carried directly into the receiver by the power lines, or radiated in the neighborhood of the antenna and picked up along with the desired signal.

Man-made noise is of two general types, *hiss* types and *impulse* types. Impulse types of noise consist of separate and distinct pulses of very high amplitude and are produced by separated electrical sparks such as occur in ignition systems, A.C. power leaks, switch and key clicks, and so forth. In hiss types of noise the pulses occur so closely together that they overlap and sound like a continuous noise. Hiss noise is produced, for example, by commutator sparking in D.C. motors and A.C. series motors. Static resembles the hiss type of noise in that the separate pulses overlap.

Noise-reducing systems. There are several ways in which the noise in a receiver may be reduced in order to improve reception, the system used depending on the type of noise it is desired to reduce. The best noise-reducing system is, of course, the elimination of the noise at its source if possible. This method is often practicable as far as electrical appliances are concerned. An appropriate electrical filter placed in the line at the source of the noise will often eliminate or greatly reduce the interference. A simple line filter consists of a 0.1-μf condenser placed across the A.C. line right at the appliance, or in some cases, of two such condensers in series with their midpoint connected to ground. When these simple filters do not effect a cure, choke coils must also be inserted in series with the A.C. power line. The sizes of condensers and coils for most effective suppression of noise can best be determined by experiment. A power-line filter at the receiver will sometimes reduce noise.

In most cases it is not possible to eliminate the noise at its source, so other means of reducing the noise must be found, the system depending on the type and source of the noise. For static and hiss types of noise, increased selectivity at the receiver will usually reduce the interference. This reduction is possible because the noise energy is spread out over a considerable frequency band, so that the wider the band the receiver accepts, the more noise it picks up. Crystal filters in the receiver are very helpful in reducing hiss types of noise. If the noise is coming from a definite direction a highly directive receiving array will reduce the interference with reception.

Owing to their special characteristics, impulse types of noise require different treatment. Impulse noises are pulses of extremely short duration separated by much longer time intervals. The energy in each pulse, and hence the amount of interference it produces, will depend both on the duration and the amplitude of the pulse. If the noise is sufficient to interfere with a signal, it must therefore have a pulse amplitude very much greater than the signal.

The simplest noise-reducing circuit in a receiver is the *audio limiter*, used only for reception of telegraph signals. Such circuits operate on the principle that a high-amplitude noise pulse causes transient disturbances in the loudspeaker or earphones that last much longer than the pulse itself, so that by limiting the amplitude of the pulse, the transients are reduced. Circuits used for audio limiting include triodes operated with extremely low plate voltage (of the order of 8 to 10 v), pentodes with low screen voltage (30 to 40), and biased diodes shunted across the output so that any noise or signal exceeding the bias is by-passed by the diodes. Audio-limiting noise-reducing systems are not very effective with low level signals or with hiss types of noise, and cannot be used for radiotelephone reception on account of the distortion they produce by the limiting process.

Superheterodyne receivers often have noise-reducing systems placed at the second detector, which operate on the principle that since the noise pulses in impulse types of noise last for such a short time, no appreciable loss of intelligibility of the signal occurs if the receiver is made inoperative during the pulse. Two types of circuits are used, both employing diodes for their action. In one circuit a low-impedance path for the noise pulse is provided so that the noise is short-circuited to ground. In the other, a diode is connected so that when the noise pulse occurs, the circuit of the audio amplifier is opened and nothing is heard.

A third type of noise-reducing system employs a fast-acting automatic-volume-control system in the intermediate-frequency amplifier to reduce the sensitivity of the receiver during a noise pulse. This type of noise silencer is very effective against impulse types of noise and is of some help with hiss noise. Intermediate-frequency noise silencers are usually used with receivers with crystal filters, so that all types of noise can be reduced.

Mention has been made of crystal filters, which are used to increase the selectivity of receivers. Quartz crystals can be cut so that they act like high-Q resonant circuits and therefore can be used in place of the usual coils and condensers in the intermediate-frequency amplifier of a superheterodyne receiver to increase the selectivity. Special circuits have been devised which allow the selectivity to be varied. Some circuits also allow interfering stations on nearby frequencies to be tuned out.

Ultrahigh-frequency Propagation. For ultrahigh frequencies, that is, frequencies higher than about 30 mc, the sky wave is no longer reflected back to earth from the ionosphere. Moreover, at these frequencies ground-wave propagation is not possible even with vertically polarized waves, for at very high frequencies both vertically and horizontally polarized waves traveling along the surface of the earth are shorted out by it. (At low frequencies this shorting out occurs only for horizontally polarized waves.) The result is that transmission and reception at these frequencies depends upon straight-line propagation from the transmitter to the receiver. For best results, both transmitting and receiving antennas should be located as high as possible above the surface of the earth. The height of the antennas determines how far apart they may be located and still receive the signal, because of the curvature of the earth and because, except for diffraction and refraction effects (to be discussed later), the waves travel in straight lines from transmitter to receiver. Figure 15–10 shows direct-ray trans-

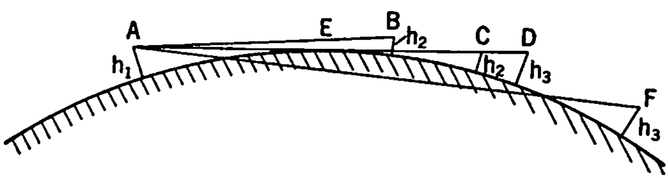


FIG. 15-10. Direct-ray Propagation at the Ultrahigh Frequencies.

mission from an antenna at A having a height h_1 above the earth to a receiving antenna of height h_2 located at B. As the receiving antenna is moved further away from A, a point will be reached where the line of sight from A to the receiving antenna will just graze the surface of the earth. This situation is shown by the location C; the distance from A to C represents the maximum over which this direct line-of-sight transmission can occur for antennas of height h_1 and h_2. However, if the height of either of the antennas is increased, this distance will be extended. In the figure the receiving-antenna height is shown increased to h_3 and the maximum distance of propagation has been increased from AC to AD. When this same height h_3 is used for a receiving antenna located at F no direct-path signal will be received.

DIRECT AND REFLECTED RAYS. When the distance between the transmitter and receiver is in line of sight, the received signal is the result of two waves, one the direct wave shown as A-C in Fig. 15–11, and the other a wave reflected by the surface of the earth and shown as ABC in the figure. These two waves add together at the receiving point C and will *reinforce* or *cancel* each other depending upon whether they arrive *in phase* or *out of phase*. This addition of waves in phase and out of phase was shown in Fig. 9–3. The reflected wave will be *reversed in phase* upon reflection. This is because the incident wave induces currents in the ground which set up a new wave (the reflected wave) which has the direction of its electric field reversed from what it was in the original wave. That is, the reflected wave is 180° out of phase with the initial wave. This phase reversal *always* occurs when horizontally polarized waves are reflected from the

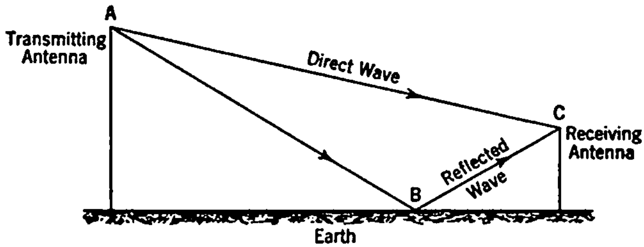


FIG. 15–11. Reception of Direct and Reflected Waves at Ultrahigh Frequencies.

ground but occurs for vertically polarized waves only at the very high frequencies.

Whether the two waves from the transmitting antenna at A arrive at C in phase or out of phase depends upon the relative path lengths of the two waves. If the path lengths are the same (as they very nearly are when the antennas are close of the ground) the waves arrive *out of phase* because of the phase reversal suffered by the reflected ray. However, if the path of the reflected wave is one half wave length longer than that of the direct ray, it takes one half of a cycle longer to travel from A to C and so arrives in phase. It is evident that it would also arrive in phase if the path difference were three halves, five halves or any odd number of half wave lengths. The signals will arrive out of phase when the path difference is any even number of half wave lengths. Whether cancellation or reinforcement occurs at C, then, depends upon the heights of the antennas and their distance apart. These should be picked, if possible, so that the path difference between the two rays is an odd half wave length and reinforcement will occur.

DIFFRACTION AND REFRACTION. It has been found that reception is possible somewhat beyond the line-of-sight distance. This reception is due to two factors not yet considered, namely, *diffraction* and *refraction*. It is possible for waves, whether sound waves, light waves or radio waves, to be diffracted or bent around obstacles in their path. The amount of bending or diffraction depends upon the size of the obstacle as compared with the wave length of the wave. If the obstacle is very large in terms of *wave lengths*, the bending will be small. This case embraces light waves, where the wave length is so small that nearly all objects are very large compared with it and little diffraction occurs. In the case of sound, however, where the wave length may be several feet, bending occurs quite readily around most objects. At low frequencies the length of radio waves is sufficiently large compared with ordinary obstacles that the waves bend around them. At very high radio frequencies the wave length becomes much shorter and only a small amount of bending or diffraction occurs. However, it is sufficient to enable signals to be received at these frequencies several miles beyond the line of sight.

Refraction of radio waves occurs as the result of changes in the density of the air with height and changes in the temperature, pressure, and amount of water vapor in the air. This refraction tends to bend the waves back to earth and has an effect in the lower atmosphere somewhat similar to the effect of the ionosphere in the upper atmosphere. The increase in distance of transmission obtained as the result of refraction can be allowed for by considering that the earth is flatter than it really is, that is, by assuming its radius to be increased by about 20% to 35%. Signals received by reason of refraction are not as stable as those for direct-ray transmission, because slight changes in the condition of the atmosphere change the amount of refraction and so produce fading.

Summary of Radio Wave Propagation. LOW FREQUENCIES (50 TO 550 KC). At frequencies below the broadcast band the ground wave is attenuated very little and it may be used for communication up to distances of 1,000 miles or more. The signals are very stable and show no diurnal or seasonal variations. At greater distances the sky wave becomes of more importance than the ground wave. The sky wave is fairly reliable but there are slight fluctuations due to variations in the ionosphere. The absorption of the sky wave is less at night than in the daytime, but at the very low frequencies the absorption is low even in the daytime. Sky-wave transmission at these frequencies is good for distances from 500 miles up to 8,000 miles. At the upper end of this band, that is, near the broadcast band, the daytime absorption of the sky wave becomes large and the sky wave is useful only at night.

BROADCAST FREQUENCIES (550 TO 1,600 KC). In the broadcast band the range of the ground wave varies from 50 miles at the higher frequencies to about 200 miles at the lower frequencies, depending also upon the power of the transmitter. Sky-wave reception is not possible in the daytime owing to high absorption of the sky wave by the *D* and *E* layers, but at night the sky wave gives reception at distances from 100 to 3,000 miles. At night there is generally an area where the sky wave and ground wave are of nearly equal magnitudes and in this region severe fading will occur. At greater distances the sky wave alone exists but there will still be some fading caused by variations in the ionosphere. The absorption of the sky wave in the broadcast band increases with frequency up to 1,400 kc. At this frequency the absorption is a maximum. The absorption decreases with decreasing frequency below 1,400 kc and decreases with increasing frequency above 1400 kc.

High frequencies (1,600 kc to 30 mc). The attenuation of the ground wave at frequencies above about 1,600 kc is so great as to render the ground wave of little use for communication except at very short distances of the order of 15 miles. The sky wave must be employed, and since it makes use of the ionosphere for its propagation, communication by means of it, although not perfectly reliable, is possible over distances as great as 12,000 miles. The sky wave absorption in this range decreases with increasing frequency, so that the higher the frequency the more efficient the transmission. It is therefore desirable to use a frequency as near as possible to the maximum usable frequency (defined above). However, if a frequency too near the maximum usable frequency is employed, the irregular changes in the condition of the ionosphere make communication uncertain. Therefore there is a range of frequencies, from about 50% to 85% of the maximum usable frequency for the given distance and time, that may be used satisfactorily. The particular layers utilized in transmitting over a given distance will depend on the distance, the time of day, and so on. For very short-distance communication (say a few hundred miles) using the sky wave, the frequency must be below the critical frequency for the layer used since the sky wave strikes the ionosphere at almost vertical incidence.

Ultrahigh frequencies (30 mc to 300 mc) and microwaves (above 300 mc). There is no ground-wave propagation on frequencies above about 30 mc, and there is usually no reflection from the ionosphere, so that communication is possible only if the transmitting and receiving antennas are raised sufficiently above the surface of the earth to allow use of a direct wave. The reflected wave from the surface of the earth combines with the direct wave to give a strong or weak signal depending on the distance from the transmitting antenna. These signals will be quite stable however, with very little static interference. For distances beyond the line of sight, diffraction and refraction must be relied on to produce signals. The refracted signal is variable because it depends on conditions in the lower atmosphere, so that fading occurs. At times when the sporadic E layer occurs, reflections of the sky wave can occur, giving transmission over long distances at frequencies up to 60 mc. Since the sporadic E layer is patchy, transmission is possible only between certain localities and for short times.

In the microwave region the direct wave must be used and communication is not possible much below the line of sight. Since there is practically no static or fading in this region, reception is very satisfactory. Highly directive antenna arrays can be built in a small space to concentrate the energy in a narrow beam, thus increasing the signal strength.

Review Questions and Problems. *1.* Why is there a low density of ionization in the atmosphere below about 30 miles and above 250 miles from the earth's surface?

2. What factors determine whether a radio wave is reflected by or penetrates an ionized layer?

3. What is meant by virtual height? Critical frequency? Maximum usable frequency for a given layer?

4. Why is energy absorbed from a wave passing through an ionized layer? How does the amount of attenuation vary between day and night conditions?

5. What causes fading of signals from broadcasting stations at night? Why is fading not present in the daytime on the broadcast band?

6. What causes selective fading? Flutter fading? Radio fade-out?

7. How does the energy in static vary with frequency?

8. What causes fading on ultrahigh frequencies?

9. What factors must be considered by a communication company in deciding which frequency to use for communication between two fixed stations?

CHAPTER 16

Radio Antennas

Functions of an Antenna System. An antenna system usually serves a twofold purpose. Its chief function is to radiate efficiently the energy furnished by the transmitter. Its second function may be to direct this energy into directions where it is wanted and to prevent radiation in other directions where the signal is not wanted. A simple antenna can do the first job quite well. To do the second may require a very complicated structure.

Practical antennas fall into one of two classes, *elevated antennas* or *grounded antennas*. An elevated antenna is operated some distance above the ground. It may be either horizontal or vertical. A grounded antenna operates with one end grounded through the output of the transmitter or the coupling coil at the end of the feed line. Elevated antennas are used at the higher frequencies, above about 2 mc, while grounded antennas are generally used at frequencies below this. At the lower frequencies, a wave length becomes very long and the necessary size of antenna for efficient radiation becomes quite large. Because of the difficulties of elevating large structures above the ground, grounded antennas are used at these frequencies. Grounded antennas are also used at high frequencies in certain particular applications such as airplane antennas where the airplane itself becomes the ground.

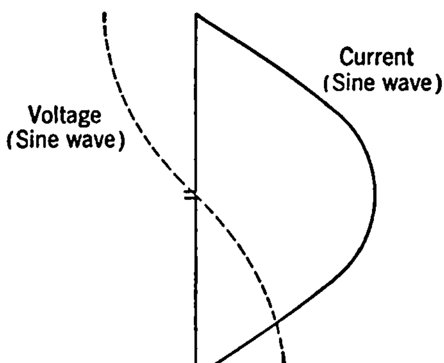


FIG. 16-1. Voltage and Current Distribution on an Elevated Half-wave Antenna.

The Elevated Half-wave Antenna. When an elevated antenna is used it is generally made a half wave length long. The antenna is then of a *resonant length*, that is, it is tuned to resonance and its input impedance is a pure resistance equal to its radiation resistance. The advantage of operating it this way will be shown later, in the section on transmission lines. The current and voltage distributions are similar to those obtained at the end of an open-circuited transmission line. These are illustrated in Fig. 16–1. The current is zero at the ends (necessarily) and the voltage is a maximum at these points. This voltage is the voltage between the two halves of the antenna, one end being positive and the other negative at any given instant. The current distribution is very nearly *sinusoidal,* that is, the current loop shown has the shape of a half sine wave. This distribution occurs because it is obtained by the addition of two traveling sine waves (the outgoing wave and the reflected wave), as illustrated in Fig. 9–5. The voltage wave is also sinusoidal with a node at the center of the antenna, which corresponds to a point one quarter wave length back from the open end of the transmission line of Fig. 9–5. The voltage loop is generally shown crossing over at the nodal point to indicate the 180° phase difference between the top and bottom halves.

Because the voltage between the two halves of the antenna is zero (or very nearly so) at the center, it is possible to short them together without affecting the operation. Of course it is still necessary to feed power to the antenna, but this can be done by tapping the feed line directly to the antenna or by coupling into the antenna by means of a radio-frequency transformer.

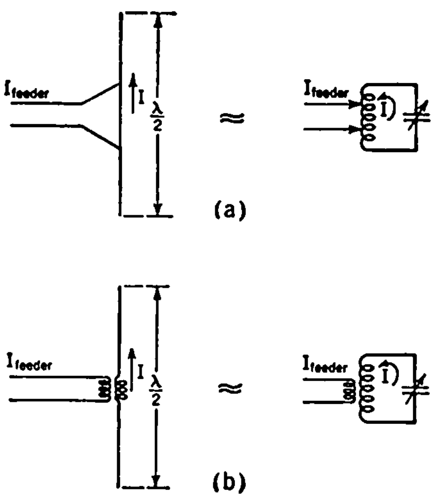


Fig. 16–2. The Half-wave Antenna Compared with a Parallel Resonant Circuit.

The half-wave antenna operated in this manner is similar in many respects to a parallel resonant circuit (Fig. 16–2) in which a large circulating current can be built up with but a small current in the feed line. As with the resonant circuit, if the antenna is detuned slightly the circulating current, or in this case the standing-wave current on the antenna, will drop sharply. The antenna will be detuned if it is shorter or longer than is required for resonance at the frequency being used. If it is too short, that is, if it resonates at a higher frequency, it can be tuned by adding inductance at the center exactly as in the case of the parallel circuit. This added inductance is known as loading. If the antenna is too long it can be tuned by inserting capacitance at the center, which is equivalent to adding series capacitance in the resonant circuit.

Radiation Characteristics of a Half-wave Antenna. As was mentioned in Chapter 9, an antenna radiates a stronger field at right angles to its axis than in other directions. The *radiation characteristic* of a half-wave antenna far away from the ground is shown in Fig. 16-3. The radiation characteristic is a method of representing graphically the relative field strength in different directions. The distance from the center of the antenna to the curve, along any particular line, represents the relative field strength in that direction. In Fig. 16-3 the line OA is twice the length of OB, so that field strength E in the direction of OA is twice as great as the field strength in the direction of OB. The field is symmetrical about the axis of the antenna and the characteristic shown is merely a cross section of the solid figure, which would represent the three-dimensional characteristic. This solid figure could be obtained by rotating the characteristic of Fig. 16-3 about the axis of the antenna. It is a doughnut shape with practically no hole and with the antenna sticking vertically through the middle.

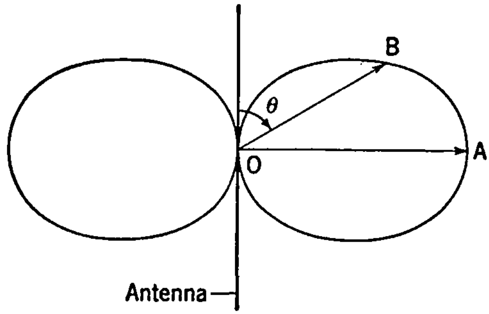


Fig. 16-3. Radiation Characteristic of a Half-wave Antenna.

The characteristic has been shown for a vertical antenna, and for such it is the vertical radiation characteristic; that is, it indicates the relative field strength at various vertical angles. The *horizontal characteristic*, which shows the relative field strength at various horizontal angles, is just a circle of radius OA for this antenna. This is because the antenna radiates uniformly in all directions perpendicular to its axis.

If this same antenna were operated in the horizontal position, the same radiation characteristic would apply, except that it would be turned over with the antenna. The characteristic of Fig. 16-3 would then be the horizontal characteristic. The vertical characteristic looking along the length of the antenna would be a circle.

The Grounded Antenna. At the lower frequencies a half wave length becomes quite long and an elevated half-wave antenna would be a large and costly structure. This is particularly true because at these frequencies the ground wave is used for transmission and so the antenna would have to be vertical.

In the middle of the broadcast band, at a frequency of say 1,000 kc, a half wave length is 150 m or about 500 ft. The difficulties of constructing a vertical antenna of this length and elevating it above the ground are evident. Fortunately, under these circumstances it is possible to use an antenna only a quarter wave long (or even shorter if inductance loading is used) and operate it with one end grounded. In this case the ground takes the place of the lower half of the antenna. This is shown in Fig. 16–4, where the lower half of the antenna has been replaced by the *image* of the upper half in the ground. It is possible to do this because if the ground were a perfect conductor the distribution of the electric field about the antenna would be as shown by the solid lines of Fig. 16–4b. This distribu-

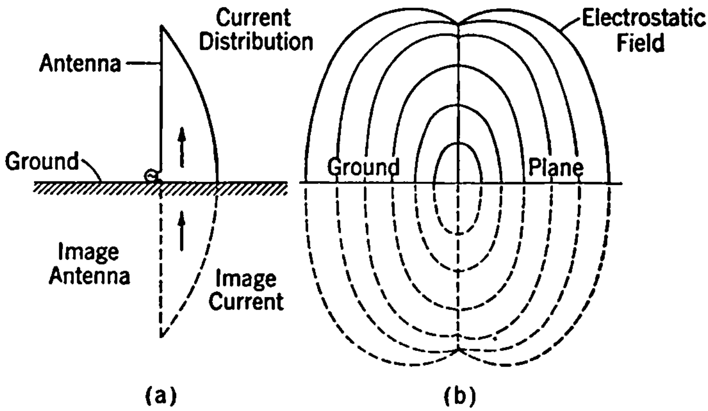


Fig. 16–4. A Quarter-wave Vertical Antenna at the Surface of the Earth. (a) Antenna and its image, showing current distribution. (b) Electrostatic field about the antenna, showing how the image antenna can be used to account for the effect of the ground.

tion above the ground is exactly similar to that obtained about one half of a half wave antenna *in free space* (far away from the ground). Therefore, in considering the operation of a grounded antenna it is merely necessary to replace the ground by the image of the antenna and then determine the fields of this *complete* antenna as if it were in free space, that is, remote from the ground.

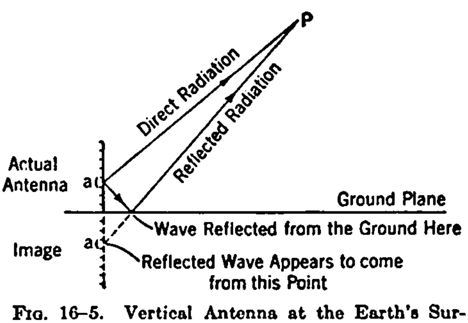


Fig. 16–5. Vertical Antenna at the Earth's Surface Showing How the Image Can Be Used to Account for the Reflected Wave.

This behavior can also be understood by considering the current in the antenna. The strength of the electric field at a point in space (not too close to the antenna) *due to any portion of the antenna* is proportional to the current flowing in that portion. The total electric field at the point is the *sum* of all the electric fields produced at that point by the various portions. (This addition must be a *vector addition* and the full meaning of this term will be discussed later under Directional Antenna Arrays.) In Fig. 16–5, to determine the relative electric field strength at the point P, the contributions due to the currents in each of the small lengths a must be summed up. However, besides the energy which reaches P by direct radiation from each of these points there is the *reflected* energy which is radiated towards the ground and then reflected up to the point P. As far as the electric field strength at the point P is concerned, it is exactly as if this reflected energy were coming from another antenna which is the *mirror image* of the actual antenna. (A *mirror image* is the type of image that is seen in a mirror, with right-hand points on the right, left-hand points on the left, close points close and distant points far away.)

Antennas of Other Heights. As in the case of the half wave antenna, a quarter wave grounded antenna is of a *resonant length* and acts like a circuit tuned to the frequency being used. Essentially this condition means that looking in at the feed point to the antenna the impedance presented is a pure resistance. It also means that with this length of antenna the antenna current for a given applied voltage will be large. If the antenna is lengthened or shortened slightly from this length the current will drop sharply. However, if it is necessary to use an antenna longer or shorter than this resonant length, such an antenna may be *resonated* or tuned by adding capacity or inductance respectively in series with it. This tuning is often done with broadcast antennas. In particular, because of cost considerations, it is often desirable to operate with a shorter antenna, say one sixth or one eighth of wave length. In this case it is necessary to load the antenna with series inductance to tune it to the frequency being used.

Losses and Efficiency. It has already been seen that because an antenna radiates energy it acts like a resistance at the end of the line feeding it. This resistance, because it is a result of the energy or power radiated, is called radiation resistance. However, there are also some other resistances in the circuit. The antenna will be constructed of wires or rods or steel girders, depending upon its size, and these will have resistance. If the antenna is loaded with an inductance coil this coil will also have a resistance, which may amount to several ohms. These resistances are generally called *ohmic* or loss resistances to differentiate them from the radiation resistance of the antenna. The antenna current must flow through these ohmic resistances and so there is a power loss. In addition there is a dielectric loss in the ground due to the penetration of the electric field into it. The relative amounts of power lost and power radiated will depend upon the ratio of the loss resistances to the radiation resistance. For this reason it is desirable to have an antenna with a high radiation resistance, since the radiation resistance gives a measure of power radiated or the useful power. The radiation resistance of a half-wave antenna is about 73 ohms. If it is properly constructed its loss resistance may be only 1 or 2 ohms or even less. This means that the losses will be small compared with the useful output power and the efficiency will be high. For a loss resistance of 2 ohms, the efficiency in this case would be given by

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} \times 100 = \frac{\text{output}}{\text{output} + \text{losses}} \times 100$$

$$= \frac{73}{73 + 2} \times 100 = 97\%.$$

For an elevated antenna the only other losses likely to occur are losses in supporting insulators and *absorption losses* due to currents induced in other conducting structures near the antenna. These can be kept small

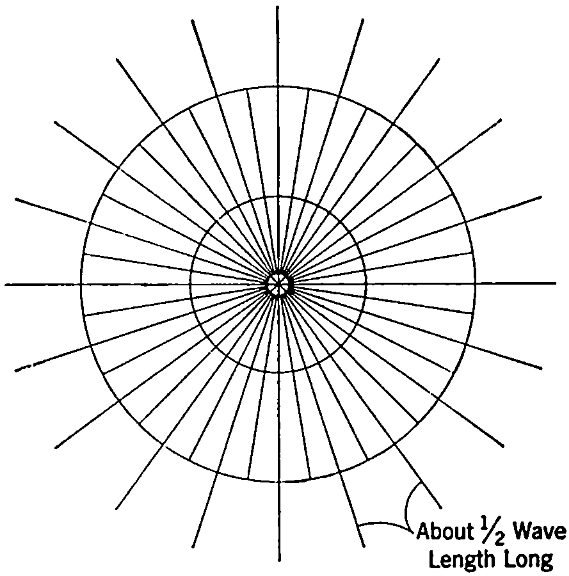


Fig. 16–6. Radial Type of Ground System Used for Grounded Antennas.

by placing the insulators at low-voltage points and by keeping the antenna away from nearby conductors.

In the case of grounded antennas the losses *may* be considerably greater than this. The ground that is taking the place of one half of the antenna, and which was assumed to be a perfect conductor for this purpose, actually has resistance that may be quite high. Referring to Fig. 16–4, it will be seen that the current flowing in and out of the bottom of the antenna must also flow out of and into the ground connection. Once in the ground it spreads radially in all directions from the antenna, keeping near the surface or penetrating to considerable depths depending upon whether the wave length is short or long. These ground currents flowing through the resistance of the ground cause losses which must be supplied from the input power. This loss decreases the efficiency of the antenna system.

As the antenna is made shorter (in terms of a wave length) the losses become quite large, for three reasons. (1) As the antenna is made shorter it requires more current in it to produce the same amount of radiated power. The losses increase as the square of the current ($P = I^2R$), and so mount rapidly. (2) The antenna must be loaded with an inductance coil to tune it to resonance and the large current flowing through the resistance of this coil may absorb an appreciable portion of the power. (3) The electric field at the base is high, which increases the dielectric losses.

Ground Systems. To reduce the losses in the ground which occur with a grounded antenna, a *ground system* is used quite genefally, especially with broadcast station antennas. The ground system is often constructed as indicated in Fig. 16–6. Wires of about a half wave length are stretched radially outward from the ground connection. The usual angular separation is 3° so that there are 120 of them. To save wire, alternate radials are often made shorter; this does not affect the operation much. If the wires are soldered or preferably welded together at the joints the resulting ground system will have very small losses. The dielectric loss at the base can be reduced by placing a screen under the antenna which is also connected to the ground system.

Radio-frequency Transmission Lines. RESONANT LINES. It is generally not possible to locate the transmitter right at the input to the antenna and some means is required for conveying the radio-frequency energy from the output of the transmitter to the antenna. Some form of radio-frequency transmission line is therefore used. One type of transmission line, known as a *resonant line*, has already been mentioned (Chapter 9). Two examples of such a line feeding a half-wave antenna are shown in Fig. 16–7. The standing waves of voltage and current existing on the antenna continue back along the line to the generator. In order not to complicate the diagrams, only the current distributions are shown. The voltage distributions are similar except that the voltage nodes occur at the current loops and the voltage loops occur at the current nodes. Figure 16–7a shows the antenna being fed by a transmission line that is approximately a half wave length long. The line feeds the antenna at a point of current maximum and the current distribution on the line is as shown. In this case (with a half-wave line), the transmitter feeds the line at a point of current maximum. This point is also a point of voltage minimum, and since impedance is given by the quotient voltage divided by current, it is also called a *point of minimum impedance* or a *low-impedance point.*

In Fig. 16–7b the same antenna is fed by a line which is three quarters of a wave length long, and in this case the transmitter feeds the line at a current minimum and voltage maximum. This point is, of course, a *high-impedance point*, and a different method of coupling the transmitter to the line is required, as will be described later. If the line were a quarter wave long, or any odd number of quarter wave lengths long, the input to the line would be at a high-impedance point. This condition exists as long as the line feeds the antenna at a current loop or low-impedance point. Also, for this case when the line is any integral number of half wave lengths long, the input to the line is at a low-impedance point. When the lines

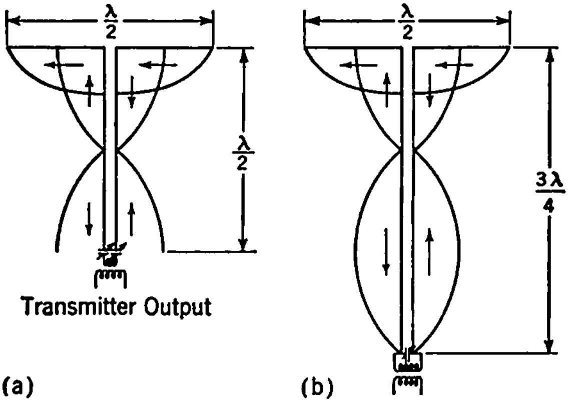


Fig. 16-7. Resonant Line Feed to Half-wave Antenna. (a) Line has low input impedance when it is an integral number of half wave lengths Long. (b) Line has high input impedance when it is an odd number of quarter wave lengths long.

are some length other than an integral multiple of a quarter or half wave length, it is necessary to tune them by adding inductance or capacity, usually at the input point of the lines. Tuning is discussed in the next section.

Characteristic Impedance and Nonresonant Lines. The lines of Fig. 16-7 are called resonant lines because, like the antenna itself, they have standing waves of current and voltage and so must be tuned to resonance. With resonant lines the input impedance, or the impedance which the transmitter has to feed, varies with the length of line. It will be recalled from Chapter 9 that standing waves of voltage or current exist on a line when there is a *reflected wave* as well as the initial wave. A wave is completely reflected at the end of a line if the line is open-circuited or short-circuited. This is so because no energy can be dissipated or used up at the end of the line under either of these conditions and so it must be reflected back. However, if the line is terminated in a *resistance* at least some of the energy will be absorbed by the resistance. If the resistance is very large (nearly an open circuit) only a small amount of energy will be absorbed and most of it will be reflected. If the resistance is very small (nearly a short circuit), again only a fraction of the energy in the initial wave will be absorbed. There is one particular value of resistance which will absorb *all* the energy in the initial wave. This particular value of resistance is called the *characteristic resistance* or *characteristic impedance* of the line. Its value depends upon the characteristics of the line, in particular upon the inductance and capacity per unit length of the line. The symbol representing it is Z_0 and its value for a line is given by

$$Z_0 = \sqrt{\frac{L}{C}},$$

where L and C are the inductance and capacity per unit length of the line. For an open-wire line of the type usually used to feed half-wave antennas, the value of the characteristic impedance is about 500 ohms. That is, if such a line were terminated by a 500-ohm resistance, all the energy would be absorbed and there would be no reflected wave and therefore no standing wave on the line. Now, the radiation resistance of a half-wave antenna (fed at the center) is only 73 ohms, and since this is small compared with 500 ohms, most of the energy will be reflected back instead of being absorbed by the antenna system and radiated. Hence there are standing waves on the transmission line. They are not perfect standing waves (that is; the voltage and current nodes do not go right down to zero), because *some* energy is absorbed and therefore the reflected wave is not quite as large as the initial wave. However, if it were possible to use some sort of a transformer to transform the 73-ohm resistance of the antenna into 500 ohms, that is, to *match* the antenna to the line, then all the energy traveling down the line would be absorbed, there would be no reflected wave, and the line would carry a *traveling wave* of current (that is, there would be a wave traveling in one direction only). The line would then be called a *nonresonant line* because there would be no standing waves on it and it would not be necessary to tune it to a resonant length. In this case its input impedance would be independent of its length.

Losses on Lines. The losses on a nonresonant line on which there are no standing waves will be less than on a resonant line which has standing waves of voltage and current. The reason is that when standing waves are present the current and voltage on the line may at certain points be many times that required when a nonresonant line is used to feed the antenna. The larger value of current increases the losses, which are porportional to the square of the current. The voltage becomes very large at the voltage antinodes, and may be sufficient to cause arc-overs if much power is being delivered. The reason for these large currents and voltages is easy to understand in terms of initial and reflected waves. If reflection occurs at the end of the line, as it does on resonant lines, the only energy radiated by the antenna is the *difference* between the initial and reflected waves. If the reflected wave is nearly equal to the initial wave, it will require very large values for both of them to make their difference appreciable. As an example, suppose $\frac{1}{5}$ of the initial current wave is absorbed (that is, radiated by the antenna) and $\frac{4}{5}$ of it is reflected back down the line. The initial and reflected waves will add in phase at certain points to produce a standing wave equal to $\frac{9}{5}$ (or $\frac{5}{5} + \frac{4}{5}$) of the amplitude of the initial wave. Since the current effective in producing radiation is only $\frac{1}{5}$ of the initial current wave, the standing-wave current (at a current loop) will be 9 times the current required to transmit the same power if the line were nonresonant. Since the losses are proportional to the *square* of the current, the losses at the current loops will be 81 times what they would be on a line without standing waves. This explains the *hot spots* that occur on a resonant line transmitting appreciable amounts of power.

Quarter-wave Matching Sections. One form of network commonly used to match a line to the antenna is a *quarter-wave matching section.* This makes use of the standing waves which exist on a resonant line feeding the antenna (see Fig. 16–7). In this case the *resonant line* is made just a quarter wave long and has the current distribution shown in Fig. 16–8. The end nearest the antenna has a current loop and is a low-impedance point, while the open end has a current node and is a high-impedance point. Somewhere in between these low- and high-impedance points there is a point where the impedance is just 500 ohms. This point is shown as point *a* in Fig. 16–8. If a 500-ohm line is attached at this point, it will then be terminated in its characteristic impedance and all the energy of the initial wave traveling down it will be absorbed. There will then be no reflected wave and therefore no standing waves on this line and it will operate as a nonresonant or untuned line. This is because the quarter-wave section has *matched* the line to the antenna.

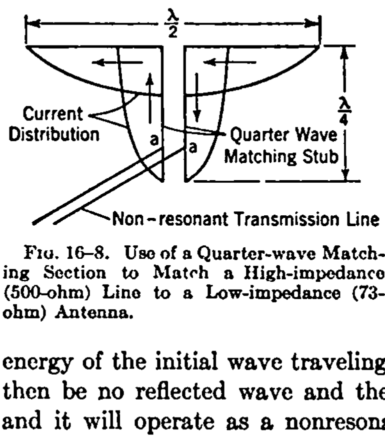


FIG. 16–8. Use of a Quarter-wave Matching Section to Match a High-impedance (500-ohm) Line to a Low-impedance (73-ohm) Antenna.

Parallel-wire and Concentric Lines. The two types of transmission lines most commonly used for conveying power to antennas are the *parallel-wire* line already mentioned and the *concentric* or *coaxial* line. Most amateur installations use parallel-wire lines because of their relative ease of installation and low cost. Commercial stations favor the concentric line, which consists of a small copper tube or a copper wire within a larger copper tube. The current flows on the outside of the inner tube and the inside of the outer tube. The electric and magnetic fields are shown in Fig. 16–9. Such a line is completely shielded from possible interference. Since the electric and magnetic fields are entirely on the inside there is no radiation from the line. The size of the tubes used depends upon the amount of power to be transmitted. Sizes of the outer tube range from $\frac{3}{8}$-inch diameter for a power of a few watts up to 3 or 4 inches for a 50-kw station. The loss depends upon the size of tubes and upon the ratio of the diameters of the outer and inner tubes. The optimum value of this ratio for least

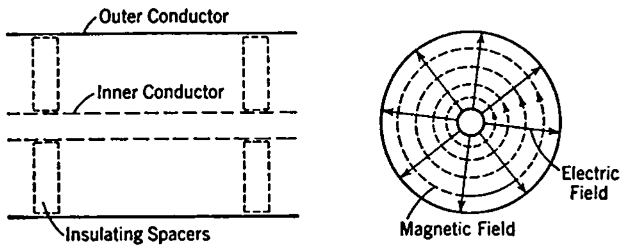


FIG. 16–9. Concentric Transmission Line Showing Electric and Magnetic Fields.

loss is 3.6, but considerable variation from this value is possible without increasing the losses much.

The characteristic impedance of a concentric line is lower than that of a parallel-wire line, being around 75 ohms. It can be calculated from dimensions of the conductors by use of the formula

$$Z_0 = 138 \log_{10} \frac{b}{a},$$

where b is the inside diameter of the outer conductor and a is the outside diameter of the inner conductor. For the optimum ratio of $b/a = 3.6$ this gives

$$Z_0 = 138 \log_{10} 3.6 = 77 \text{ ohms}.$$

The formula for the characteristic impedance of parallel-wire lines is

$$Z_0 = 276 \log_{10} \frac{b}{a},$$

where in this case b is the spacing between wires, center to center, and a is the *radius* of the wire. For No. 10 wires at 3-inch spacing this gives about 490 ohms, while for No. 10 wires at 6-inch spacing the characteristic impedance is about 570 ohms.

Another difference between these two types of lines is that the parallel-wire line is a *balanced line* while the concentric line is *unbalanced to ground.* With a balanced line the two conductors have the same capacity and inductance per unit length and also the same physical relation to the ground. With the concentric or unbalanced line the outer conductor is at ground potential and the voltage on the inner conductor is then a voltage with respect to ground. Balanced lines are required for properly feeding elevated half-wave antennas and other similar types that are balanced to ground. Concentric lines, on the other hand, are particularly suited to feeding grounded antennas because then one connection of the line *must* be grounded if it is to be directly connected to the antenna. It is possible to feed grounded antennas with balanced lines and elevated antennas with concentric lines, but suitable coupling networks must be used.

Coupling Networks. Coupling networks are used to couple or connect the transmitter to the transmission line and the transmission line to the antenna. (If there is no transmission line the coupling network is used to couple the transmitter output directly to the antenna.) Coupling networks serve to isolate the transmission line and antenna from the large D.C. potentials at the output of the transmitter. They are used to tune to resonance the circuits they connect, and for this purpose are generally provided with one or more variable elements, such as variable condensers or variable inductances. Finally, they provide a means for varying the coupling between the circuits and therefore can be used for impedance matching so that the maximum possible amount of power is transferred from the transmitter to the antenna.

There are many different types of coupling networks, some of which are shown in Figures 16–10 to 16–15. However, the two functions of tuning and impedance matching can be explained with reference to the simple coupling network shown in Fig. 16–10.

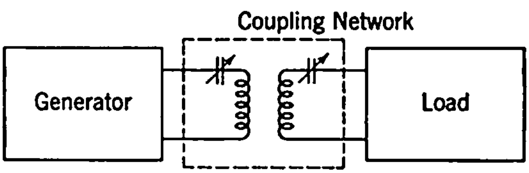


Fig. 16–10. Typical Series-tuned Coupling Network for Connecting a Low-impedance Generator to a Low-impedance Load.

Tuning. In order to obtain large currents in an antenna or any other circuit having reactance it is necessary to *tune the circuit to resonance*. In Chapter 3 it was seen that this can be accomplished by adding either inductance or capacitance to the circuit until the inductive reactance is just equal to the capacitive reactance If the circuits to be coupled together are not already tuned to resonance the coupling network can be made to tune them.

In the network of Fig. 16–10, the variable condensers are used to tune the generator and load circuits. Of course it might happen that either or both of these circuits would require added inductance instead of capacity for correct tuning. However, in this case, since the coupling coils used are now part of these circuits, it is only necessary to make the reactance of these coils large enough so that the circuits (without the condensers) are inductive, and so can be tuned by means of the condensers. Variable condensers are preferred to variable inductances in all low-power applications because they are easier to construct and adjust, but with high-power transmitters the opposite is sometimes true. In that case the condensers would be fixed and the inductances made variable.

If only one of the circuits being coupled together requires to be tuned (the other already being resonant), then a variable element will be required only in the untuned side. However, if both circuits are detuned and the coupling between them is *loose*, a variable element will be required in both input and output sides of the coupling unit. Coupling is said to be *loose* when it is insufficient, as defined in Chapter 3. In this case the tuning of one circuit does not have much effect on the tuning of the other and so the circuits can and must be separately tuned to the resonant frequency. However with *close* coupling, that is, where the coupling is sufficient or greater than critical, the tuning of one circuit affects the tuning of the other so that only one variable element is required for tuning both circuits.

It is necessary nevertheless to adjust the amount of coupling in order to transfer the maximum amount of power from one circuit to the other, and this adjustment affects the tuning so that the correct over-all adjustment is more difficult to obtain. With a tuning element in both circuits the tuning procedure is simpler. In this case the coupling between circuits is made loose and each circuit is then tuned to resonance independently of the other. The coupling is then increased until maximum current flows in the second circuit and the correct adjustment has been obtained. Because of the ease of adjustment that results, two tuning elements are often used even where only one is necessary.

IMPEDANCE MATCHING. Besides tuning the circuits they couple together, coupling networks are used to control the amount of power transferred from one circuit to the other. This control is known as *impedance matching.* In Fig. 16–10 the amount of power transferred is controlled by varying the coupling between the two coils. Often, though not always, it is desired to transfer the maximum possible amount of power from one circuit to the other. In this case the coupling is increased until the resistance coupled into the generator circuit by the coupling is just equal to the resistance of the generator circuit alone. It has been previously shown (Chapter 7) that this is the condition for maximum power transfer.

Coupling between two circuits can be varied in ways other than varying the mutual position of two coils. If the coils are fixed or wound on the same form the coupling may be varied up to the maximum obtainable by providing the coils with taps and tapping on at different points.

The amount of mutual inductance that can be obtained conveniently with air-core coils, such as are used in antenna coupling networks, is quite small. If the resistances of the generator and load are small so that large currents are required for a given amount of power, the large current flowing through the primary will produce a large amount of magnetic flux and sufficient voltage will be induced in the secondary. A *series* circuit, such as that shown in Fig. 16–10, in which the coupling coil and condenser are in series with the generator or load, is satisfactory in this case. However, when high resistance circuits are to be connected together, only small currents will flow, and it would require a very large number of turns to produce sufficient magnetic flux to induce the required voltage in the secondary. In this case the parallel connection of Fig. 16–11 is used. The coil and condenser are in parallel with the generator or load. Here a relatively small current (but correspondingly large voltage) from the high-impedance generator will produce a large circulating current in the primary parallel resonant circuit. This large current produces sufficient magnetic flux to induce the required voltage in the secondary. If both generator and load circuits are high-impedance circuits (that is, high-voltage–low-current circuits) the parallel connection is used on both sides of the coupling network. If one is a high-impedance circuit and the other is a low-impedance circuit, the parallel connection will be used on the high side and the series connection on the low side.

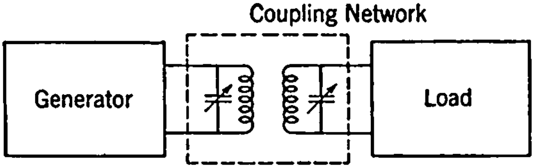


FIG. 16–11. A Parallel-tuned Network for Coupling a High-impedance Generator to a High-impedance Load.

TYPICAL COUPLING NETWORKS. Some typical coupling networks used to couple the transmitter to the transmission line and the transmission line to the antenna are shown in Figs. 16–12 to 16–15.

Figure 16–12 shows methods of coupling from the final stage of the transmitter to a resonant line. The plate circuit of a tube is a high-impedance circuit and parallel tuning is used on that side. On the secondary side series tuning is used, as in Fig. 16–7a, when the resonant line is being fed at a low-impedance point. If the line is being fed at a high-impedance point (Fig. 16–7b), the parallel circuit of Fig. 16–12b is used.

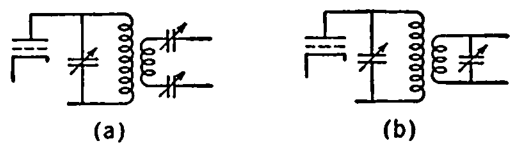


FIG. 16–12. Coupling a Transmitter to a Resonant Line. (a) Series feed for coupling at a low-impedance point. (b) Parallel feed for coupling at a high-impedance point.

When a nonresonant parallel-wire transmission line is being used, the input impedance will be a pure resistance (usually about 500 ohms) and its value will be independent of length. The coupling circuits of Fig. 16–13 are suitable for this. In Fig. 16–13a the condenser is omitted on the line side and the small amount of inductive reactance introduced by the second-

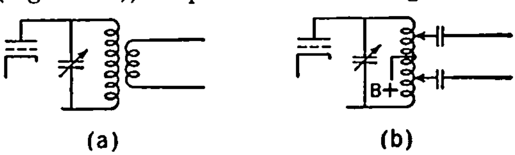


FIG. 16–13. Coupling a Transmitter to a Nonresonant Line. (a) Inductive coupling. (b) Direct coupling.

ary coupling coil is tuned out on the primary side. Close coupling is necessary to accomplish this tuning. In Fig. 16–13b the same result is obtained by direct coupling to the primary coil. The amount of the coupling is varied by moving the taps on the coil. Loose coupling requires only a few turns between taps; close coupling is obtained with a larger number of turns. The fixed condensers are necessary to isolate the line from the high D.C. voltages present in the plate circuit of a transmitting tube.

Figure 16–14 shows a method of coupling a transmitter to a concentric or unbalanced transmission line. Such a line has a low impedance and so series feed is used. A *Faraday shield* is often used in this case. A Faraday shield is a shield which prevents electrostatic coupling between the coils while allowing magnetic coupling.

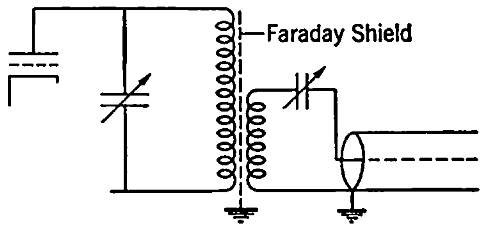


Fig. 16–14. Coupling a Transmitter to an Unbalanced Line.

Coupling the transmission line to the antenna. Several methods of coupling a transmission line to an antenna are shown in Fig. 16–15. Figure 16–15a shows a series-feed connection to a vertical grounded antenna. If the antenna is shorter than a quarter wave length it will have a capacitive reactance and this can be tuned out by proper adjustment of the loading coil. There is no provision for impedance matching. The radiation resistance of the antenna is about 36 ohms for a quarter wave length and drops rapidly as the antenna height is decreased, being only about 13 ohms at one sixth of a wave length and 7 ohms at one eighth of a wave length. Because of this drop, considerable mismatch may occur and the feed line will be operating as a resonant line. Figure 16–15b shows a parallel-feed connection to couple a high-impedance (500-ohm) line to a grounded antenna. The coil in series with the antenna and ground is used to tune the antenna circuit to resonance, and the impedance match is obtained by adjusting the coupling between the coils. In Fig. 16–15c is shown a shunt-feed arrangement sometimes used with broadcast antennas. The feed line is run up at an angle and connected directly to the antenna tower at a point where the impedance is the same as the line impedance. The impedance to ground along the antenna is zero at the base and increases to a maximum at the top. (This is similar to an elevated half-wave antenna in which the voltage and therefore the impedance between halves is low at the center and increases to a maximum at the ends.)

Connections to elevated antennas are shown in Fig. 16–15d to 16–15g. The center-feed connection (Fig. 16–15d) makes the feed line a resonant line (with standing waves) because of the impedance mismatch between the 500-ohm line and 73-ohm antenna. However, tuning can be accomplished at the transmitter end of the line and the only effect of the mismatch will be to increase somewhat the losses due to the standing waves. If the line is to be quite long the direct feed connection of Fig. 16–15e is to be preferred, since the impedance of the antenna and line can be matched and standing waves on the line eliminated or at least reduced. The matching

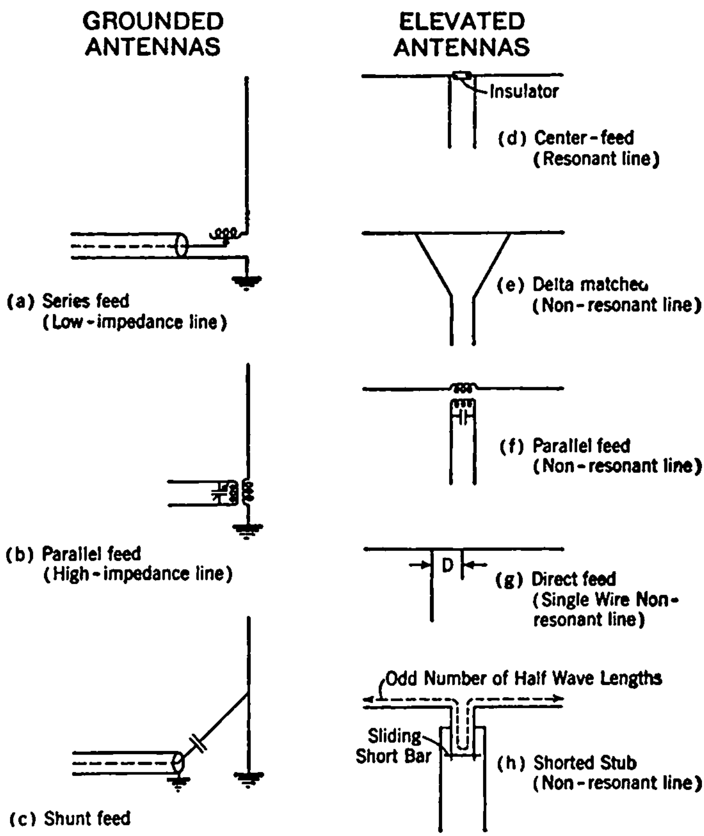


Fig. 16–15. Methods of Coupling the Transmission Line to the Antenna.

comes about in the same way as it does for the shunt feed arrangement of Fig. 16–15c. The impedance between two points on opposite sides of the center of the antenna increases as the distance between the points is increased. By tapping the transmission line to the antenna at points where the impedance is equal to that of the transmission line an impedance match is obtained. However, because of the loop that then exists at the end of the line some inductive reactance is present along with the resistance, so that a slight detuning results and there will be some standing wave on the line. Figure 16–15f illustrates a parallel type of feed similar to that in Fig. 16–15b. It is possible to obtain both an impedance match and proper tuning, but the arrangement is not well suited to outdoor work.

Figure 16–15g shows a single-wire feed line direct-connected to the antenna. By tapping the line to the antenna at a suitable distance off center an impedance match can be obtained and standing waves on the feed line eliminated.

Figure 16–15h shows a coupling system using a portion of a line as a matching and tuning network. The total length of the antenna from one end down around the shorting bar on the stub and out to the other end should be an odd number of half wave lengths. This length will make the antenna circuit resonant. Tuning can be accomplished by adjusting the shorting bar. A current loop will appear at the shorting bar and current nodes at the ends of the antenna. A match to the open-wire transmission line can be obtained by tapping the line to the stub at an appropriate distance from the shorting bar.

Directional Antenna Systems. Besides acting as efficient radiators or receivers of electromagnetic energy, antennas may be used to select the directions along which the energy shall be transmitted or received. This selection is usually accomplished by using two or more single antennas properly spaced and suitably fed. Such a system is known as a *directional array.* For transmitting antennas directional arrays are used to direct the energy into certain directions where it is desired and to prevent its radiation in other directions where it would be wasted or where it would create interference. For receiving antennas a directive array can be used to discriminate against undesired signals coming from directions other than that in which reception is desired.

The gain of a directional array is the ratio of the signal strength in the desired direction using the directional antenna system to the signal strength obtained using a nondirectional antenna. The gain is usually expressed in decibels, a 6-db gain corresponding to an increase of two-to-one in signal strength. Since gains of 10 to 15 db are possible in both transmitting and receiving antenna systems, the reduction in transmitted power necessary for a given received signal strength in point-to-point communication is very marked. According to the reciprocity theorem, the radiation pattern of an antenna system used for transmitting is the same as its pattern when used for receiving. (Exceptions to this occur in the case of the V and rhomic antennas owing to the polarization of the transmitted wave.)

Directivity of a Single Half-wave Antenna. A single half-wave antenna radiates uniformly in all directions perpendicular to its axis. However, the radiation in directions making other angles with the axis depends upon this angle. The radiation reduces to zero as the angle goes to zero, that is, along the line of the antenna. The graphical representation showing the relative radiation or field strength in different directions is known as the *radiation characteristic* or *radiation pattern* of the antenna system. It is also called the *directional characteristic* or *directional pattern* of the antenna and this expression is usually shortened to *pattern.*

The *horizontal pattern* is the directional pattern seen when looking down on the antenna from above. It shows the relative field strengths at all *horizontal angles.* The *vertical pattern* shows the relative radiation at all *vertical angles.* The horizontal and vertical patterns are shown in Fig. 16–16 for vertical and horizontal antennas. Figure 16–16a shows the horizontal pattern of a vertical antenna, and since the radiation is uniform in all horizontal directions this pattern is simply a circle. Figure 16–16b shows the vertical pattern of a vertical antenna. Figure 16–16c shows the horizontal pattern of a horizontal antenna far away from the earth and is, of course, exactly the same as Fig. 16–16b. The vertical pattern of a horizontal antenna looking along the line of the antenna is shown in Fig. 16–16d. This is the same as Fig. 16–16a. The vertical pattern of the same horizontal antenna looking along a perpendicular to the axis of the antenna is shown in Fig. 16–16e. This pattern is similar to Fig. 16–16b and Fig. 16–16c.

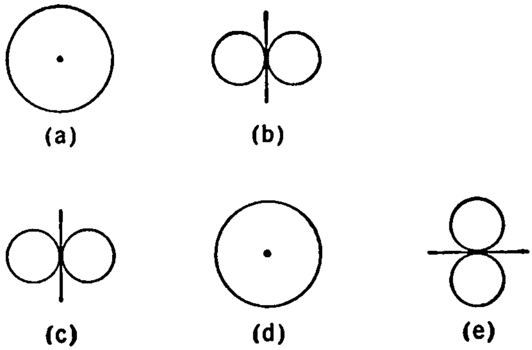


FIG. 16–16. Radiation Patterns of a Single Half-wave Antenna. (a) Horizontal pattern of a vertical antenna. (b) Vertical pattern of a vertical antenna. (c) Horizontal pattern of a horizontal antenna. (d) Vertical pattern of a horizontal antenna (looking along the line of the antenna). (e) Vertical pattern of a horizontal antenna (looking along a line perpendicular to the axis of the antenna).

Vertical Antennas Spaced One Half Wave Length. One of the simplest directional arrays consists of two vertical antennas spaced one half wave length apart and fed in-phase with equal currents. The expression in-phase means that the currents in the two antennas reach their maxima (in the same direction) at the same instant. Figure 16–17a shows the horizontal pattern which would be obtained for this array. It is the so-called "figure eight" pattern with the line of zero radiation parallel to the line of the antennas (the line drawn between the antennas). The pattern is simply explained by reference to Fig. 16–17b. For the direction OA the distances from the two antennas to the receiving point are equal so that radiations leaving the two antennas at the same instant will arrive simultaneously at the point A. The currents in the antennas are assumed to be in-phase, so the radiations will leave the antennas in-phase and will arrive at point A in-phase. The addition of sine waves having a phase difference between them was covered in Chapter 3. The addition is accomplished by drawing vectors with the appropriate phase angle between them. In this case the phase angle between the waves arriving at A is zero (they arrive in-phase) so that the resultant field strength is just twice that from a single antenna (Fig. 16–17c).

For radiation in the direction of OB, conditions are different. Antenna No. 2 is one half wave length further from the receiving point than antenna No. 1, so that it requires a longer time for a wave from No. 2 to reach B

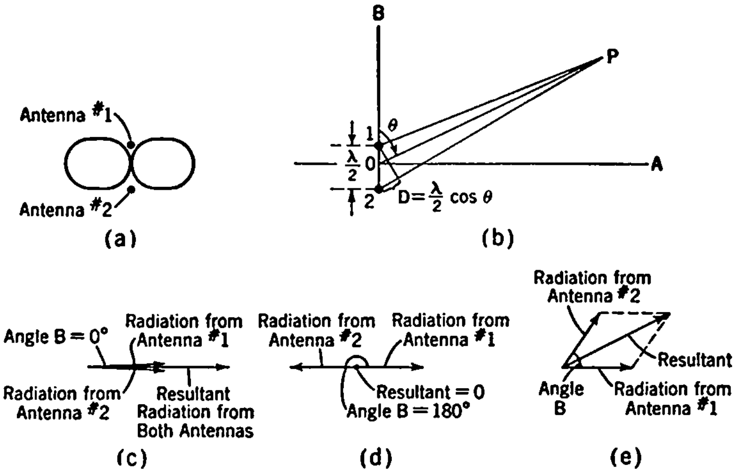


Fig. 16–17. Horizontal Radiation Pattern for Two Vertical Antennas Spaced One Half Wave Length and Fed In-phase. (a) Radiation pattern. (b), (c), (d), and (e) Method of obtaining pattern.

than it does for a wave from No. 1. A wave leaving antenna No. 2 at a moment of a current maximum will arrive at antenna No. 1 just one half cycle later when the current in No. 1 is a maximum *in the opposite direction.* The two waves which leave No. 1 and travel on together toward point B will then be 180° out of phase and they will cancel each other. Their resultant, as shown by the vector addition in Fig. 16–17d, is zero.

For waves traveling toward some other point P along a line which makes an angle θ with the line of the antennas, the difference in distance between antenna No. 2 and P and antenna No. 1 and P is something less than a half wave length. This difference in distance is shown as D in Fig. 16–17b, and by geometry it will be seen to be approximately equal to $\frac{\lambda}{2} \cos \theta$ where λ is a wave length. In this case the phase of the current in antenna

No. 1 will change by an angle B which is less than 180° while the radiation from No. 2 is traveling the distance D. The angle B will be given by

$$B = 180 \times \frac{D}{\lambda/2} = 360 \times \frac{D}{\lambda} \text{ degrees.}$$

The waves which travel on together towards point P will differ in phase by B degrees, and so their resultant at P will be as shown in Fig. 16–17e.

In this manner the radiation at any angle θ can be computed, and if this is done for all angles a directional pattern as shown in Fig. 16–17a will be obtained. It is a maximum along OA, zero along OB, and varies from the maximum to zero as the angle θ changes from 90° to 0°.

As an example, the radiation along an angle $\theta = 60°$ will be computed. The distance D for this angle is

$$D = \frac{\lambda}{2} \cos 60° = \frac{\lambda}{4}.$$

The angle B will be

$$B = 360 \times \frac{D}{\lambda} = 90°.$$

For $B = 90°$ the two vectors are at right angles and the resultant vector has a length which is $\sqrt{2}$ times that of a single vector. Therefore the radiation in a direction making an angle $\theta = 60°$ is $\sqrt{2}/2$ times the maximum radiation along OA.

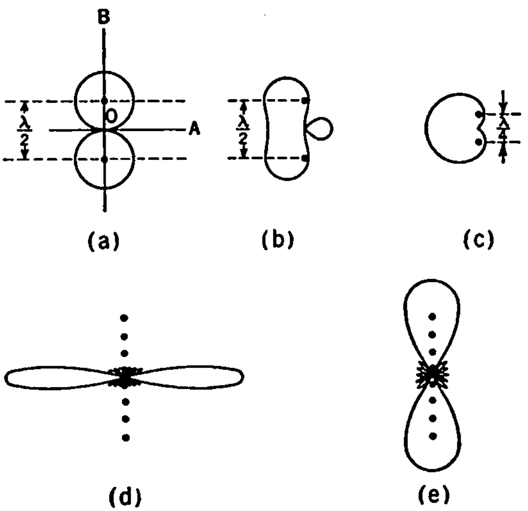


FIG. 16–18. Radiation Patterns for Commonly Used Antenna Arrays. (a) Two antennas spaced one half wave length and with 180° phase difference. (b) Same with 90° phase difference. (c) Two antennas spaced one quarter wave length and fed with 90° phase difference. (d) Broadside array (currents in-phase). (e) End-fire array (currents 180° out-of-phase).

Antennas Fed with Currents Having 180° Phase Difference. When the same two vertical antennas as above are fed with currents which have 180° phase difference the resulting radiation pattern is another "figure eight," this time with the zero radiation direction perpendicular to the line of the antennas. This pattern is shown in Fig. 16–18a. The explanation of this pattern is similar to that of the antennas fed in-phase, except that now radiations which leave the antenna at the same instant are 180° out of phase so that they cancel each other along the line *OA*. However, along the line *OB*, the current in antenna No. 1 changes by 180° while the wave from No. 2 travels the distance between the antennas so that the two waves leaving antenna No. 1 and traveling along together towards *B* are now

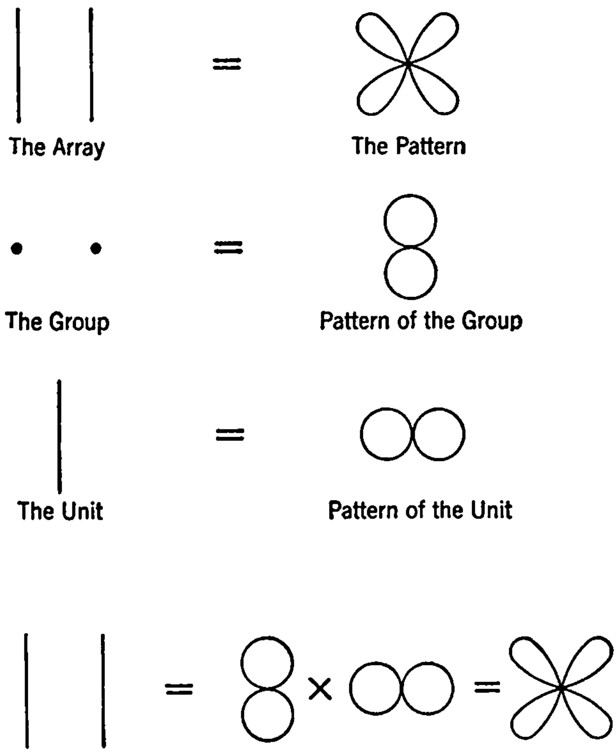


Fig. 16–19. Horizontal Pattern of Two Horizontal Antennas Spaced One Half Wave, and Fed In-phase—Obtained by Multiplication of Patterns.

in-phase and so produce a large resulting signal in this direction. The radiation in other directions can be computed in a manner similar to that of the in-phase case of Fig. 16–17.

Other Phases. When the two antennas are fed with equal currents having other phase differences the directions of zero radiation will be different and different patterns will result. The pattern for a phase difference of 90° is shown in Fig. 16–18b.

Other Spacings. A variety of different patterns can be obtained by changing the spacing between the antennas. An example of particular interest is the *cardioid* pattern obtained with a spacing of one quarter wave length and a phase difference of 90°. This pattern, shown in Fig. 16–18c, is sometimes known as a unidirectional pattern because most of the energy is transmitted in one direction.

Line of Antennas. When more gain or sharper directivity than can be obtained with two antennas is desired, a line of antennas is often used. A line of antennas spaced one-half wave length and fed in-phase is known as a broadside array because the energy is radiated broadside to the line of the array. This is evidently so because the radiation from all the antennas would add in phase along the perpendicular to the line. A typical pattern is shown in Fig. 16–18d. If the same array were used with alternate antennas fed 180° out of phase, the resulting pattern would be as shown in Fig. 16–18c and the array would be known as an end-fire array.

Horizontal Patterns of Horizontal Antennas. The patterns of Figs. 16–17 and 16–18 are horizontal patterns of vertical antennas *or* vertical patterns of horizontal antennas. (The effect of the ground on these patterns will be considered later.) The horizontal pattern of two horizontal antennas spaced one half wave length and fed in-phase can be obtained as in Fig. 16–17, but there will be one difference. The individual vertical antennas radiate equally in all the directions considered, that is, their individual patterns are circles. The horizontal antennas will each have a (horizontal) pattern as shown in Fig. 16–16c. This latter pattern is called the pattern of the unit, while the pattern of Fig. 16–17, which is due to a combination of antennas (each radiating uniformly), is known as the pattern of the group. The directional characteristic of an array of (similar) antennas, whether they radiate uniformly or not, can be obtained by multiplying the pattern of the group by the pattern of the unit. This computation is illustrated in Fig. 16–19.

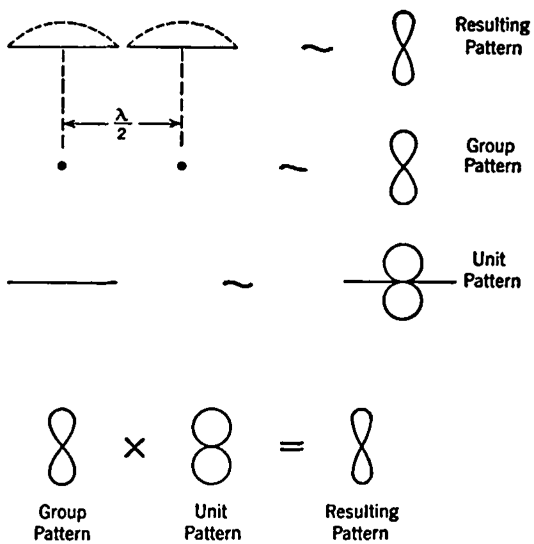


FIG. 16–20. Pattern of a Colinear Array (Half Waves In-phase)—Obtained by Multiplication of Group and Unit Patterns.

where the horizontal pattern of two horizontal antennas spaced one half wave length and fed in-phase is determined by multiplying the pattern of the group (Fig. 16–17a) by the pattern of the unit (Fig. 16–16c). This principle of multiplication of patterns can also be used quite effectively in determining the effect of the ground on the vertical patterns of antenna arrays.

Colinear Array. A colinear array consists of half wave elements in a line parallel to the axis of the element. The elements are fed in-phase. Since the centers of the antennas are one half wave apart the group pattern would be that of two (nondirectional) units spaced one half wave length and fed in-phase, that is, similar to Fig. 16–17a. The pattern of the unit is, of course, Fig. 16–16c. Multiplied together as in Fig. 16–20, they give the resulting pattern of a colinear array.

Effect of the Ground on Vertical Radiation Patterns. The ground is not a perfect conductor and therefore not a perfect reflector of radio waves. However, at low and medium-high frequencies it can be considered as such for the purpose of determining its effect on the radiation characteristics of antennas. Figure 16–21 shows a vertical antenna elevated one half wave length above the ground. As far as the effect at some point *P* is concerned, the wave reflected from the ground appears to come from an *image antenna* located one half wave length beneath the surface. This image antenna is in-phase with the actual antenna and the two antennas constitute a directional array. Figure 16–21b shows a horizontal antenna elevated one quarter wave length (say) above the earth. Again the image appears an equal distance below the surface, but this time its current is 180° out of phase with the current in the actual antenna. This opposite phase occurs because in this case the wave is horizontally polarized and the wave reflected from the ground is *reversed in phase* upon reflection.

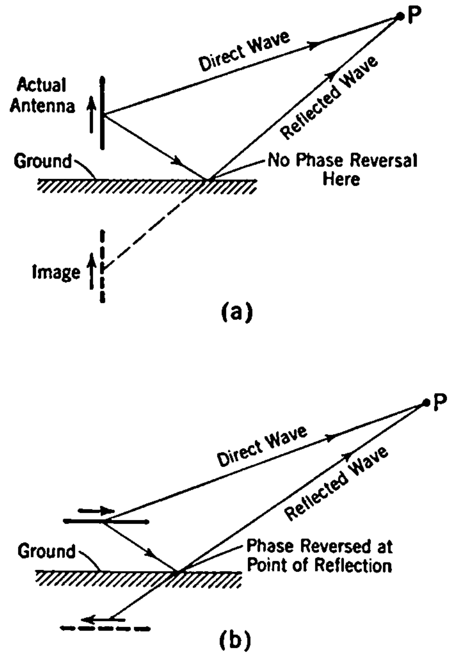


FIG. 16–21. Image Antenna Due to a (Perfectly) Reflecting Ground. (a) Vertical antennas (reflection without phase reversal). (b) Horizontal antennas (phase reversal on reflection).

The effect of the presence of the ground may be determined by considering the antenna and its image as a directional array and using the principle of multiplication of patterns to obtain the vertical radiation characteristic. This procedure is illustrated in Fig. 16–22, where the vertical radiation pattern of a horizontal antenna located a quarter wave length above the ground is determined, as shown. Of course, only the upper half of the resulting pattern will apply. The patterns for vertical or horizontal antennas at any other heights above the ground can be determined in a similar manner. The patterns for a horizontal antenna located at heights λ/4, λ/2, 3λ/4, and λ above the ground are shown in Fig. 16–23.

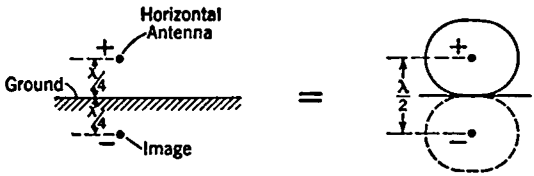


Fig. 16–22. Vertical Radiation Pattern of a Horizontal Antenna One Quarter Wave Length above the Ground, Determined by Considering the Antenna and Its Image.

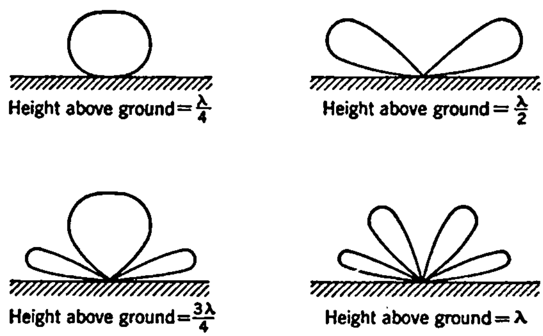


Fig. 16–23. Vertical Radiation Patterns of a Horizontal Antenna at Various Heights above a (Perfect) Ground.

Loop Antennas. An antenna that has useful directional characteristics is the loop antenna shown in Fig. 16–24. The dimensions of this antenna are usually small compared with a wave length, so that the currents are in phase all around the loop (as shown). The two vertical sides of such a loop are then equivalent to two vertical antennas having a spacing that is a fraction of a wave length and having currents in *opposite phase* (one current is up when the other is down). The resulting horizontal pattern is a "figure eight" with the zero line perpendicular to the plane of the loop. The top and bottom parts of the loop are equivalent to two horizontal antennas, also 180° out of phase so that their pattern is also a "figure eight" (this time vertical) having its zero line perpendicular to the plane of the loop. This sharp null or zero-signal line makes the loop very useful in the direction-finding applications discussed in the section on radio beacons, later in this chapter.

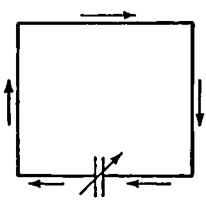


Fig. 16–24. Loop Antenna.

Radio Beacons. By making use of the directional properties of antennas, systems of *radio beacons* for aircraft have been developed and are now used extensively for marking the courses for aircraft to fly. The type of beacon used in the United States operates in the following manner: when a plane is flying on its proper course as marked out by the beacon, a continuous tone signal from the beacon is heard in a receiver, but if the plane gets off course two tone signals will be heard, the code letter N (– .) and the code letter A (. –), one being louder than the other, depending on whether the plane is off course to one side or the other of the beam. By turning the plane to make the two code signals blend together into a continuous tone, the pilot gets back on the beam from the beacon.

The principle on which the beacons operate may be seen with the aid of Fig. 16–25. Four antennas are used, with the radio-frequency power being

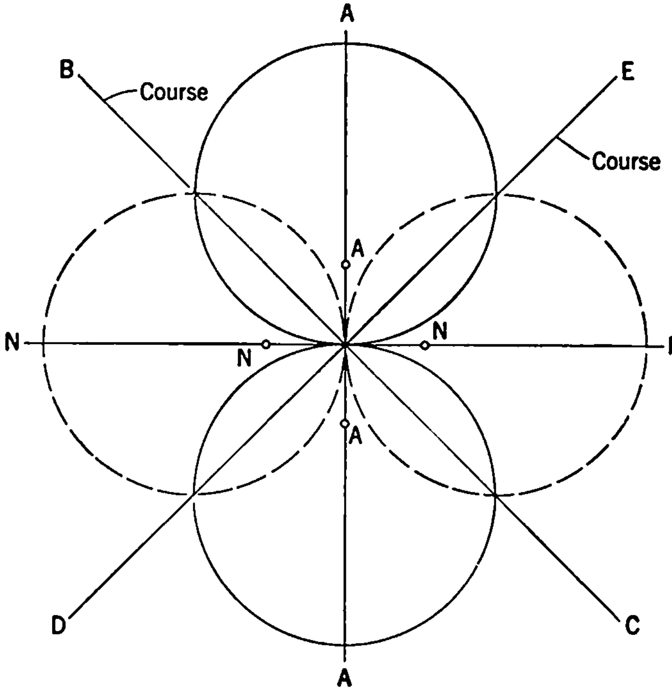


Fig. 16–25. A Radio Beacon Showing Directional Patterns of the Pairs of Antennas, and Directions of Equal Signals.

fed to them in pairs so that they really consist of two directional antenna systems with their patterns crossed. The antennas are arranged on the corners of a square, the diagonally opposite antennas forming pairs. A switching arrangement is used to feed the power to one pair of antennas, the N antennas, at one moment, and then to the A antennas at the next moment, but never to all the antennas at once. When power is being fed to the A antennas, a directional pattern shown by the solid line in the figure is produced. This is just the "figure eight" pattern of Fig. 16–18a. When power is fed to the N antennas, the "figure eight" pattern shown by the dashed line is obtained, with its zero-signal line at right angles to the zero-signal line of the A antennas. In general, a plane will receive stronger signals from one pair than from the other, depending on the direction of the plane from the beacon. However, along the directions of the diagonal lines *BC* and *DE* the signals will be equal. The switching system used to feed the power alternately to one pair and then to the other pair is such that the signal sent out from A antennas consists of a short signal (dot), a short pause, a long signal (dash), a long pause (space), and then this sequence repeats. This signal is a series of A's in the International code. During the pauses, the energy is fed to the N antennas, so that the signal from the N antennas is a series of N's in the International code. When the plane is off the course line *BC* (or *DE*), either the A signal or the N signal will be louder than the other, depending upon which side of the course the plane is on. When the plane gets back on course the A and N signals are of equal loudness, and since the N signal just fills in the pauses in the A signal a continuous tone will be heard. The continuous tone is the on-course signal.

The on-course signals are actually heard over a small zone, instead of the mere line shown, because the ear is unable to detect extremely small differences of volume. The on-course zone usually is a sector, a degree or two wide, so that it becomes narrower as the plane approaches the beacon. The courses defined by these zones will lie along lines at right angles to each other, that is along the lines *BC* and *DE*. In order to change these courses so that they pass over air routes which are not located 90° apart, the directional patterns of the antennas must be changed. The ways of doing this include changing the positions of the antennas so that they lie on the corners of a rectangle instead of a square, the feeding of different amounts of power to the pairs of antennas, and the addition of more antennas to make more complicated arrays. Radio beacons commonly use vertical antennas or towers, but sometimes crossed-loop antennas are employed. Since a single loop antenna has the same "figure eight" pattern shown for a pair of vertical antennas, a loop may be substituted for each pair, forming the crossed-loop beacon.

Directly above the beacon there will be very little signal from either pair of antennas, so that a *cone of silence* exists, which tells the pilot he is over the beacon. As the plane approaches the cone of silence, the signal builds up rapidly to a loud volume and then disappears. Sometimes, however, false cones of silence exist, caused by rapid flight over changing conformations of surface of the earth, as for example while flying over a deep canyon. False cones of silence are easy to detect because of the absence of the loud build-up before the signal disappears.

Radio beacons are subject to what is known as *night error*. Radio beacons usually operate on the low frequencies, so that, just as in the broadcast band, a sky wave is present at night. Reflections from the ionosphere can cause an apparent change in the directional pattern of the beacon antennas, so that the direction of the on-course signal is changed slightly. The explanation of the night effect is given below in connection with direction finders. The night effect is very small with vertical type antennas, but may be appreciable with crossed-loop antennas.

Direction Finders and Radio Compasses. Radio beacons keep an airplane flying on a predetermined course, but tell the pilot very little about his actual position. When the plane's actual position must be known, and for plotting courses when not flying on a beam, radio direction finders and radio compasses are used. Direction finders are also used by vessels for navigating, particularly near harbors and dangerous locations. Radio direction finders and radio compasses make use of directional antennas at the receiver, instead of at the transmitter as in beacons. By using a directional receiving antenna which can be rotated, the direction from which a signal is coming can be determined quite accurately.

Most direction finders use loop antennas, as these are much smaller than other types of directional antennas. When used for receiving, the pattern of a loop antenna is a "figure eight," as it is in transmitting. When the plane of the loop is at right angles to the direction from which vertically polarized waves are coming, the same voltage will be induced in each of the vertical sides of the loop; but since these voltages send currents around the loop in opposite directions, they will cancel and no signal will be heard. The same is true for waves that come from the opposite direction. Now, suppose the loop is rotated 90° so that one side of the loop is nearer to the transmitting station than the other. The voltages induced in each side of the loop are now not quite in phase opposition because of the time the radio wave takes to go from one side to the other. Therefore the two voltages will not quite cancel each other and a signal will be heard. Signals will be heard for waves that come from the opposite direction also. In order to increase the signal output from a loop antenna, it is usual to use a condenser at the terminals to tune the loop to resonance.

If a loop antenna is mounted so that it can be rotated about a vertical axis, and tuned to a station, the loudness of the signal received will depend on the direction of the loop with respect to the station. If the plane of the loop is at right angles to the line to the station, no signals will be heard, but if it is turned 90° from this position, maximum signals will be heard. Thus, either the position of zero signal (the *null* position) or the position of loudest signal could be used for direction finding. The null is usually used since it gives a sharper indication.

When a loop is turned to either the null position or the maximum-signal position, there is still an uncertainty as to whether the signal is coming from the front or the back directions of the loop. This uncertainty exists because there are two positions of the loop for zero signal. Often it is known in which general direction the station lies, especially in marine navigation, so that the uncertainty is removed. However, in cases where this is not known, other means must be used. To determine the *sense* of the reading of the loop (that is, which of the two possible directions is the correct one), *sense antennas* are sometimes used. A sense antenna is a small vertical antenna that picks up a signal which is then fed into the loop in such a way as to unbalance the loop. The unbalance of the loop makes one of the directions of maximum signal louder than the other. To use the sense antenna, a bearing using the null position is first taken with the sense antenna disconnected. Then, to determine the sense, a switch is closed and the loop is turned until the direction of loudest signal is found. A pointer on the loop then shows the true direction of the station.

Loop antennas are subject to certain errors, most of which can be avoided by proper construction and calibration. If a loop antenna is not balanced with respect to the ground, unequal stray currents will flow in the sides of the loop, changing the directions in which zero signal occurs. These two directions of zero signal are no longer exactly 180° apart, so that incorrect bearings will be obtained, depending on which null direction is used. Errors from this cause are eliminated by using a loop that is shielded electrostatically and carefully balanced with respect to ground. Sometimes small compensating condensers are used to eliminate any residual unbalance. The presence of wires, large metal objects, and conductors in the neighborhood of a loop can cause errors in the bearing indications.

There will be some horizontally polarized waves present in signals that have been reflected from the ionosphere, as the plane of polarization may be rotated on this path. These waves are coming downward instead of traveling horizontally as do the ground waves. The horizontal component of the wave induces voltages in the horizontal arms of the loop, thus affecting the null position and giving incorrect bearings at night. This *night error* can be largely eliminated by using an *Adcock antenna*. An Adcock antenna is simply a pair of very short vertical antennas, spaced a small distance apart and crossed over at their centers to form an H. This antenna is shown in Fig. 16-26. The action of the antenna is just the same as that of a loop as far as vertically polarized waves are concerned. Horizontally polarized waves, such as those present in downcoming sky waves from the ionosphere, induce voltages in the horizontal parts of the antenna in such a way that they cancel and do not produce any signal. Small Adcock antennas give very low signals since they are equivalent to a loop antenna with only one turn.

Fig. 16-26. Adcock Antenna.

Adcock antennas give accurate bearing indications under conditions that render loop antennas completely useless.

It is possible to attach a meter or other indicator to a loop antenna in such a way as to give indications on a dial telling the pilot whether the plane is headed directly towards the transmitting station or whether it is flying to the right or left of this direction. An arrangement of this type is called a *radio compass.* One type of compass uses a fixed loop at right angles to the line of flight. Another vertical antenna is used with it to give a maximum signal from the front of the plane. When the plane is flying directly towards the station, the needle on the meter rests in the center of the dial. If the plane deviates from this course the needle moves from the center position, showing which way the plane has deviated from its course. This arrangement is often called a *homing device,* since it guides the plane to its home base. Homing devices are very useful in guiding planes back to an aircraft carrier.

Another form of radio compass operates on the same principle as that described above, but with the difference that the loop can be rotated. This rotation allows the pilot to take bearings on stations off his course without changing the direction of flight. An important advantage of this type of compass is that it allows corrections to be made for drift, which cannot be done with fixed-loop homing devices.

Review Questions and Problems. *1.* (a) Sketch the voltage and current distributions on an elevated half-wave antenna and on a quarter-wave grounded antenna. (b) Sketch the vertical and horizontal radiation characteristics of these antennas (in the vertical position).

2. Why should an antenna be *tuned?* How can this be accomplished? What is meant by *loading* an antenna?

3. Why can the effect of the ground be simulated by an *image* antenna?

4. What are the power losses which can occur in an antenna system? Why is a short antenna likely to be less efficient than one which is a quarter or half wave long?

5. Explain fully why long resonant lines are less efficient than non-resonant lines of the same length?

6. Define *characteristic resistance* of a transmission line.

7. Explain the use of quarter-wave matching sections.

8. Number 12 wire has a diameter of .081 inches. Compute the characteristic impedance of line having two such wires spaced (a) 6 inches apart; (b) 3 inches apart. *Answer:* (a) 600 ohms, (b) 520 ohms.

9. The plate circuit of a vacuum tube is to be connected to a 500-ohm transmission line. Show a coupling network suitable for this connection.

Why would a parallel resonant circuit rather than a series resonant circuit be used on the tube side of the network?

10. Explain how airplanes can be kept *on course* by means of radio beacons.

11. How do direction finders operate? What is a radio compass? A homing device?

Index