AF MANUAL 52-8

RADAR

circuit analysis

DEPARTMENT OF THE AIR FORCE

AF MANUAL 52-8

RADAR

circuit analysis

DEPARTMENT OF THE AIR FORCE

AIR FORCE MANUAL
NO. 52-8

DEPARTMENT OF THE AIR FORCE
WASHINGTON, 30 JUNE 1951

FOREWORD

1. PURPOSE. This Manual provides the graduate radar technician with reference information on fundamentals of electronics and provides the student radar technician with a basic coverage of the subject.

2. CONTENTS. This Manual develops the subject of radar circuit analysis from applied mathematics through fundamentals of radar circuits typical to Air Force equipment.

3. RECOMMENDATIONS. Recommendations or suggestions for the improvement of this Manual are encouraged. Comments may be forwarded to the Director of Training, Headquarters USAF, Washington 25, D. C.

BY ORDER OF THE SECRETARY OF THE AIR FORCE:

HOYT S. VANDENBERG
Chief of Staff
United States Air Force

OFFICIAL:
K. E. THIEBAUD
Colonel, USAF
Air Adjutant General

DISTRIBUTION

ZONE OF INTERIOR

Headquarters USAF	50
Major air commands	2
Subordinate air commands	2
Services — MATS & AMC	5
Air divisions	2
Air Force bases	2
Wings	2
Groups	1
Squadrons	1
AACS detachments	1
Radar air traffic control centers	2
Personnel holding SSN 867 and SSN 955	1

OVERSEAS

Major air commands	2
Subordinate air commands	2
Air Force bases	2
Wings	2
Groups	1
Squadrons	1
AACS detachments	1
Radar air traffic control centers	2
Personnel holding SSN 867 and SSN 955	1

This training manual is published by the Commander, Air Training Command, in accordance with provisions of AFR 5-39

REPRINTED FEB 1955
REPRINTED OCT 1959

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. — Price $3.25

Introduction

Radar, one of the great inventions emerging from World War II, has many uses in aviation. Aircraft are equipped with various types of radar equipment for navigational purposes. Radar equipment is employed at many airports for landing aircraft during periods of restricted visibility. Meteorologists employ radar equipment for collecting certain weather data for aviation. Air-traffic controllers use radar equipment in some phases of air-traffic control operations. These are only a few of the many present day applications of radar to aviation. A greater use of radar equipment in aviation is planned for the future.

Since radar equipment is so widely used, you may think that the job of the radar technician is a difficult one. However, this is not necessarily true, for all types of radar equipment are based on the same fundamental principles and basic circuits. If you learn these principles and circuits, you should not have a great amount of difficulty when working with any type of radar equipment. In many cases radar circuits are just radio circuits which have been modified and rearranged. Though some of these circuits may require considerable study, you will find that for the most part they are easy enough to understand, once you understand the basic principles and basic circuits of radar.

It is the purpose of this manual to give you the basic principles and basic circuits of radar. A prerequisite to your understanding these basic concepts is a knowledge of mathematics, electricity, and radio. Several chapters review these subjects rapidly. If you are not too well grounded in mathematics, electricity, and radio, you may have to consult other information sources for greater detail. The other chapters analyze the circuits which serve as the basic components of radar equipment. For the most part, the circuits discussed are the circuits common to most radar sets and the circuits which the radar technician understands. Only in a few cases are they taken from specific sets. In short, this manual gives you the basic aspects of radar and its circuits. As a radar technician, you should find its contents valuable for study and reference both during and after training.

Table of Contents

CHAPTER 1

Mathematics for Radar

No doubt you can learn to understand the operation of many electrical and radar circuits by a study of current flow and by mechanical analogy, but to make the circuits actually work, you need to apply mathematics. Algebra, a common mathematical subject, is applied frequently in the analysis and study of electronic and radar circuits. In high school you probably studied this subject, but might have forgotten many of the principles and thus need a review. Logarithms, trigonometry, vector algebra, and the slide rule are not commonly taught in high school, but also are very useful tools for analyzing radar circuits.

This chapter deals with the mathematics you need in understanding the electronic and radar circuits in this manual. Algebra is reviewed rapidly. Logarithms, trigonometry, vector algebra, and the slide rule, new subjects to many, are treated less rapidly and with emphasis on the type of problems you will encounter in the manual. Each mathematical principle stated is followed by examples and solutions to give you a better understanding of its application.

ALGEBRA

An *algebraic expression* is any combination of signs, numerals, and letters used for numbers, written according to the rules of algebra. For example, $a + b$, l, lw, and $-\frac{x}{2y}$ are algebraic expressions. Numbers represented by letters are called *literal* numbers. Thus a, b, l, w, x, and y are literal numbers. Signs may indicate whether the numbers are positive or negative, or they may indicate operations; for example, addition, subtraction, multiplication, division, a root, or a power.

Multiplication of two algebraic numbers need not be indicated by the symbols for multiplication — the times sign (×) or the dot (·). Just putting them together indicates multiplication. Thus, the product of a and b can be written ab. In the expression ab, a and b are called *factors* of the product. Likewise 6, a, and b are factors of $6ab$.

Each factor of a product is the *coefficient* of the other factors. In the algebraic expression $6ab$, 6 is the coefficient of ab, $6a$ of b, and $6b$ of a. In this expression, 6 is called the numerical coefficient and b the literal coefficient of a. The word coefficient is usually restricted to mean only the numerical or arithmetic coefficient.

The size or magnitude of a number is called its *absolute value.* Absolute value refers only to the magnitude with no consideration to the sign preceding the number. For example, the absolute value of +7 is 7, and the absolute value of −9 is 9. Both +3 and −3 have the same absolute value, 3.

An *exponent* is a number or letter which indicates the power to which a quantity is to be raised, or the number of times the quantity (called the base) is multiplied by itself. For example, 4^3 is read *4 to the third power*, and means 4 x 4 x 4. The expression e^x is read *e to the x power*, and means that e is multiplied by itself x times.

Any arithmetic or literal number, or the product or the quotient of these numbers, is called a *term*. For example 4, x, y, $25b$, ab, $\frac{x}{y}$, $\frac{3a^2b}{c}$ are terms.

Terms that have the same literal parts are called *similar* or *like* terms. Thus a, $7a$, and $12a$ are like terms. Terms such as x^2, y, lh, which do not have the same literal parts, are called *unlike* terms.

You can add or subtract two or more numbers if they are like terms. For example, you can add $4x$ to $5x$ to produce the sum $9x$. Likewise, you can add $5x^2$ to $10x^2$, since the literal parts x^2 are alike. (Literal numbers raised to the same power are like terms.) The sum of numbers such as $4x$ and $5m^2$ must be indicated as $4x + 5m^2$ since x and m^2 are not like terms.

Some Rules of Algebra

ADDITION. To add two numbers with the same sign, add their absolute values and prefix the common sign.

Thus, $+6$ *added to* $+3$ *equals* $+9$
-6 *added to* -3 *equals* -9
$+6a^2$ *added to* $+3a^2$ *equals* $+9a^2$
$-6m^2n$ *added to* $-3m^2n$ *equals* $-9m^2n$

To add two numbers with opposite signs, take the difference between their absolute values and prefix the sign of the number with the larger absolute value.

Thus, $+6$ *added to* -3 *equals* $+3$
-6 *added to* $+3$ *equals* -3
$+6a^2$ *added to* $-3a^2$ *equals* $+3a^2$
$-3m^2n$ *added to* $+6m^2n$ *equals* $+3m^2n$

SUBTRACTION. The process of subtraction is opposite to the process of addition. To subtract a quantity from another, change the sign of the quantity to be subtracted, then add the quantities, following the rules of addition.

For example

$$(-6)-(-3)=(-6)+(+3)=-3$$
$$(-3)-(-6)=(-3)+(+6)=+3$$
$$(+6)-(-3)=(+6)+(+3)=+9$$
$$(+6)-(+3)=(+6)+(-3)=+3$$

POLYNOMIALS. An algebraic expression containing two or more terms joined together by a plus or minus sign is called a *polynomial*. Thus, the expressions $a + b - c$ and $ab - ac + \frac{a}{c}$ are polynomials. A polynomial with just two terms such as $5x + y$ is a *binomial*. A polynomial with only three terms is a *trinomial*. The expression $5x^2 + ab + c$ is a trinomial. A single term, such as a, x, $10x^2y$, is a monomial.

Polynomials are added and subtracted in the same way as single terms (monomials).

For example

Addition	*Subtraction*
$-4a^2 + 3az - 3y$	$3m^2 - 6n^2$
$3a^3 \quad - 2az + 5y$	$-3m^2 - 8n^2$
$3a^3 - 4a^2 + az + 2y$	$6m^2 + 2n^2$

Note in these examples that the plus sign is omitted before $3a^2$, $3m^2$ and $6m^2$. Algebraic practice permits the omission of the sign before the first term of an expression if it is positive. Likewise, the numerical coefficient 1 is omitted before a literal number, such as before az in the addition problem.

SIGNS OF GROUPING. Certain symbols, called *signs of grouping*, tie or group together several quantities affected by the same operation. The most commonly used symbols of grouping are parentheses ( ), brackets { }, braces [ ], and the vinculum_____. All these symbols indicate that the quantities affected by them are to be treated as single quantities. For example, the expression $(4a^2 - 3ab) - (3a^2 + 2ab)$ means that the quantity $3a^2 + 2ab$ is to be subtracted from $4a^2 - 3ab$. To perform this operation, change the signs of the expression $3a^2 + 2ab$, and add algebraically,

$$\begin{array}{r} 4a^2 - 3ab \\ -3a^2 - 2ab \\ \hline a^2 - 5ab \end{array}$$

In removing a symbol of grouping preceded by a minus sign, change the signs of all terms included by the symbol. When one symbol of grouping occurs within another, remove the innermost symbol first. You may remove both symbols at the same time if you are quite careful, but this practice can easily lead to errors of sign. Therefore, it is best to remove only one symbol at a time. The following example illustrates the procedure for removing a symbol of grouping occurring within another:

$$3a^2 - [2am - (2a^2 + 5am) + a^2]$$
$$3a^2 - [2am - 2a^2 - 5am + a^2]$$
$$3a^2 - 2am + 2a^2 + 5am - a^2$$
$$4a^2 + 3am$$

MULTIPLICATION. In the multiplication of algebraic terms, there are three things to consider — the sign, the exponent, and the coefficient.

The Sign. The product of two terms with like signs is positive. The product of two terms with unlike signs is negative.

The Exponent. The exponent of any letter in the product is the sum of the exponents of the factors with the same base.

Thus, $x^2 \times x^3 = x^{(2+3)}$, or x^5
Likewise, $y^3 \times y^7 = y^{10}$

The Coefficient. The arithmetic coefficient of the product is the product of the absolute values of the coefficients of the terms being multiplied.

Thus, $6x^2 \times 2x^3 = 12x^5$
Likewise, $3y^3 \times 3y^7 = 9y^{10}$

Examples

$$(3a^2) \times (5a^3) = 15a^5$$
$$(2ax) \times (-3a^2xy) = -6a^3x^2y$$
$$(m^2n) \times (-7mn^2) = -7m^3n^3$$
$$(4xy) \times (-xy^3) = -4x^2y^4$$
$$(-a) \times (-5a^3) = 5a^4$$

A literal quantity with no written exponents means the quantity is to the first power — that is, it has an exponent of one.

Multiplication of polynomials is similar to the multiplication of numbers consisting of several digits in arithmetic. Multiply each term of the multiplier by every term of the multiplicand and add the partial products. Thus, in multiplying $2e^2 - 3e - 5$ by $-4e + 7$, proceed as follows:

$$\begin{array}{r} 2e^2 - 3e - 5 \\ -4e + 7 \\ \hline -8e^3 + 12e^2 + 20e \qquad \\ +14e^2 - 21e - 35 \\ \hline -8e^3 + 26e^2 - e \quad - 35 \end{array}$$

DIVISION. In the division of one term by another, the sign, the coefficient, and the exponent must be considered in obtaining the quotient.

The Sign. The quotient of two positive or two negative quantities is positive; the quotient of a positive and a negative quantity is negative.

The Exponent. The quotient of powers having like bases has the same base as the given powers, and an exponent which is the difference of the exponents of the given powers.

Thus, $\frac{x^4}{x^2} = x^2$

Likewise, $\frac{x^{10}}{x^3} = x^7$

The Coefficient. To obtain the coefficient of the quotient, divide the absolute value of the coefficient of the dividend by the absolute value of the coefficient of the divisor.

Thus, $\frac{6x^4}{3x^2} = 2x^2$

Likewise, $\frac{8x^{10}}{2x^3} = 4x^7$

In division there is the possibility of zero and negative exponents, as the following examples indicate.

$$\frac{a^5}{a^5} = a^{(5-5)} \text{ or } a^\circ, \text{ or } \frac{a^4}{a^5} = a^{-1}, \ \frac{a^3}{a^5} = a^{-2}$$

Any quantity with a zero exponent is equal to *one.*

Thus, $\frac{a^5}{a^5} = a^\circ = 1$ or $\frac{x^3}{x^3} = x^\circ = 1$

Any quantity with a negative exponent is equal to the reciprocal of that quantity with the corresponding positive exponent.

Thus, $a^{-2} = \frac{1}{a^2}$ or $\frac{1}{x^{-3}} = x^3$

Any factor can be moved from the numerator to the denominator of a fraction or vice versa without changing the value of the fraction, if the sign of the exponent is changed.

Thus, $\frac{3x^2y}{a^3b} = \frac{3x^2ya^{-3}}{b}$

Division of one polynomial by another is similar to long division in arithmetic. One difference is that the dividend, divisor, and the remainder (if there is one) must be arranged in order of ascending or descending powers of some letter.

Example

Divide $30c^4 + 3 - 82c^2 - 5c + 11c^3$ *by* $3c^2 - 4 + 2c$

Solution

$$\begin{array}{r|l} & 10c^2 - 3c - 12 \\ \hline 3c^2 + 2c - 4 & 30c^4 + 11c^3 - 82c^2 - 5c + 3 \\ & 30c^4 + 20c^3 - 40c^2 \\ \hline & -9c^3 - 42c^2 - 5c \\ & -9c^3 - 6c^2 + 12c \\ \hline & -36c^2 - 17c + 3 \\ & -36c^2 - 24c + 48 \\ \hline & 7c - 45 \text{ remainder} \end{array}$$

Solution of Equations

An equation is a statement that two quantities are equal. To solve an equation means to find the value or values of the literal numbers for which the equation holds true. In solving an equation, you will have to use *axioms,* statements accepted as true without proof. Here are a few of the more commonly used axioms:

1. If the same number is added to or subtracted from both sides of an equation, the result is still an equation.
2. If both sides of an equation are multiplied or divided by the same quantity (not zero), the result is still an equation.

3. If like roots or powers are taken of both sides of an equation, the result is still an equation.

Other processes which are based upon the axioms and are often applied in the solution of equations are *transposition*, *changing signs*, and *cancellation*. Transposition is the transferring of a term from one side of an equation to the other and changing its sign. This is merely another way of saying that the term was added to or subtracted from both sides (axiom 1). Changing signs of terms on both sides of an equation is merely another way of multiplying or dividing both sides by -1 (axiom 2). Cancellation is the means of collecting terms or of transposing and collecting terms.

Examples

1. *Given* $x-5=3$, *find* x.

Solution

Add 5 to both sides of the equation (axiom 1),

$$x-5+5=3+5$$

Then collect terms.

$$x=8$$

2. *Given* $5x-4=21$, *find* x.

Solution

Add 4 to both sides of the equation (axiom 1),

$$5x-4+4=21+4$$

Then collect terms on both sides,

$$5x=25$$

Divide both sides by 5 (axiom 2),

$$x=5$$

3. *Given* $\frac{1}{3}x+5=8$, *find* x.

Solution

Subtract 5 from both sides of the equation (axiom 1),

$$\frac{1}{3}x+5-5=8-5$$

Collect terms,

$$\frac{1}{3}x=3$$

Then multiply both sides of the equation by 3 (axiom 2),

$$x=9$$

4. *Given* $\frac{4}{5}x+5=25-1\frac{1}{5}x$, *find* x.

Solution

Subtract 5 from both sides of the equation (axiom 1),

$$\frac{4}{5}x=20-1\frac{1}{5}x$$

Then add $1\frac{1}{5}x$ *to both sides (axiom 1),*

$$\frac{4}{5}x+1\frac{1}{5}x=20-1\frac{1}{5}x+1\frac{1}{5}x$$

Collect terms,

$$2x=20$$

Divide both sides by 2 (axiom 2),

$$x=10$$

5. *Given* $16-5(x+3)=4(2x+1)-9\frac{1}{2}$, *find* x.

Solution

Remove the parentheses,

$$16-5x-15=8x+4-9\frac{1}{2}$$

Collect terms,

$$1-5x=8x-5\frac{1}{2}$$

Subtract 1 from both sides (axiom 1),

$$-5x=8x-6\frac{1}{2}$$

Subtract 8x from both sides (axiom 1),

$$-13x=-6\frac{1}{2}$$

Then divide both sides by -13 *(axiom 2)*

$$x=\frac{1}{2}$$

6. *Given* $\sqrt{x}+2=4$, *find* x.

Solution

Subtract 2 from both sides of the equation,

$$\sqrt{x}=2$$

Then square both sides (axiom 3),

$$x=2^2 \text{ or } 4$$

7. *Given* $z^2=r^2+x^2$, *find* r.

Solution

Subtract x^2 *from both sides of the equation,*

$$z^2-x^2=r^2$$

Take the square root of both sides of the equation (axiom 3),

$$\sqrt{z^2-x^2}=r$$

Monomial Square and Square Roots

When a monomial is multiplied by itself, the product is the *square* of the coefficient multiplied by each literal quantity with its exponent doubled.

The square root of a monomial is the square root of the coefficient times each literal number with its exponent halved. There are two roots—a negative root and a positive root. This is true because the sign of the product of two negative or two positive quantities is positive.

Examples

$$(3a^2)^2 = 9a^4$$

$$(-5mn^3)^2 = 25m^2n^6$$

$$\sqrt{64a^4} = \pm 8a^2$$ (*The symbol* $\pm$ *means that it could be either* $+8a^2$ *or* $-8a^2$)

$$\sqrt{50a^3} = \pm 7.07a^{1.5}$$

Special Products and Factoring

Factoring is the process of finding two or more quantities, each called a factor, whose product is equal to a given quantity. The most simple form of polynomial factoring is finding the common monomial factor. The expression, $5x + 5y$, is a polynomial in which 5 is a factor common to both monomials. Factor this expression by dividing $5x + 5y$ by 5. Write the result as $5(x+y)$.

Examples

$$ax + ab - az = a(x + b - z)$$

$$4mx - 8my + 2mz = 2m(2x - 4y + z)$$

$$5a^3b - 10a^2b^2 - 15ab^3 = 5ab(a^2 - 2ab - 3b^2)$$

PRODUCT OF TWO BINOMIALS WITH COMMON TERM. A frequently encountered product has the form $(x + a)\ (x + b)$, which, when multiplied, gives $x^2 + x\ (a + b) + ab$. This relationship can be stated as a rule: The product of two binomials having a common term is the square of the common term, plus the product of the common term times the algebraic sum of the unlike terms, plus the algebraic product of the unlike terms.

Examples

$$(a+3)\ (a+4) = a^2 + a(3+4) + (3 \times 4)$$
$$= a^2 + 7a + 12$$

$$(a-5)\ (a-3) = a^2 - 8a + 15$$

$$(a-7)\ (a+3) = a^2 - 4a - 21$$

$$(x+7)\ (x-9) = x^2 - 2x - 63$$

A trinomial such as $x^2 + x\ (a + b) + ab$ is often expressed in the general form $x^2 + mx + n$ in which m represents the sum of the unlike terms and n, their product. *Factoring*, or finding the binomial factors, in trinomials of this form involves finding two quantities whose sum equals m and whose product equals n. After you determine these quantities, write the binomial factors by adding these quantities to the square root of the first term of the trinomial. Thus, for example, in $b^2 - 9b + 14$, the numbers -7 and -2 must be part of the binomial factors, for their sum is -9 and their product is $+14$. This means that the binomial factors are $(b - 7)$ and $(b - 2)$.

Examples

$$e^2 - 9e + 20 = (e-5)\ (e-4)$$

$$r^2 + 11r - 42 = (r+14)\ (r-3)$$

$$p^2 - p - 42 = (p-7)\ (p+6)$$

$$t^2 + 11t + 24 = (t+3)\ (t+8)$$

SQUARE OF A BINOMIAL. When the product is of the form $(x + a)\ (x + a)$ or $(x + a)^2$, the middle term of the product is $2ax$, and the final term is a^2. Thus, $(x + a)^2 = x^2 + 2ax + a^2$.

Expressed in rule form, this relation is: The square of a binomial is the sum of the squares of the two terms added to twice their algebraic product.

Examples

$$(r+t)^2 = r^2 + 2rt + t^2$$

$$(k-2j)^2 = k^2 - 4jk + 4j^2$$

$$(3e-d)^2 = 9e^2 - 6ed + d^2$$

Factoring polynomials of the type $x^2 + 2ax + a^2$, called perfect trinomial squares, depends largely on your ability to recognize them as such. A trinomial is a perfect square if two of its terms are perfect squares and the third term is twice the product of their square roots.

Examples

$$g^2 + 2gk + k^2 = (g+k)^2$$

$$p^2 + 2pq + q^2 = (p+q)^2$$

$$r^2 - 10r + 25 = (r-5)^2$$

OTHER BINOMIAL PRODUCTS. When multiplying binomials of the type $mx + a$ and $nx + b$, whose product is $mnx^2 + x\ (na + mb) + ab$, little time is saved by following a rule. However, knowing the rule for such multiplication may make it easier for you to factor the product of binomials of this type.

When multiplying any two binomials,

1. To find the first term of the product, multiply the first terms of the binomials by each other.

2. To find the second term of the product, obtain the product of the outer terms of the binomials and the product of the inner terms of the binomials and add them algebraically.

3. To find the third term of the product, find the algebraic product of the second terms of the binomials.

Thus, in finding the product of $(2e - 3)\ (3e + 4)$, you would proceed as follows:

First term: $2e \times 3e$, or $6e^2$.

Second term: $(2e \times 4) + (-3 \times 3e) = 8e - 9e$ or $-e$.

Third term: -3×4, or -12.

The total product is therefore: $6e^2 - e - 12$.

Factoring trinomials of the form $mnx^2 + (mb + na)x + ab$ usually involves a trial-and-error method. As you saw in the rule just given, the factors must be such that the product of the first terms of the binomials equals the first term of the trinomial, the product of the second terms of the binomials equals the third term of the trinomials, and the algebraic sum of the cross products is the second term of the trinomial. Therefore, the usual procedure is to try two binomials which satisfy the first and second conditions, then check to see if they also satisfy the third condition.

Example

Factor $6a^2 - 17a + 12$.

Trial Factors		*Product*
$(3a - 2)\ (2a - 6)$	=	$6a^2 - 22a + 12$ *Wrong* *middle term incorrect*
$(6a - 1)\ (a - 12)$	=	$6a^2 - 73a + 12$ *Wrong—same thing*
$(6a - 3)\ (a - 4)$	=	$6a^2 - 27a + 12$ *middle term still incorrect*
$(3a - 4)\ (2a - 3)$	=	$6a^2 - 17a + 12$ *Right*

With practice, you can eliminate some elements in guessing the right combination. For example, you can tell by the signs of the trinomial whether the signs of the binomials will be both plus, both minus, or one plus and one minus. Do not use a binomial factor that contains a monomial factor. Remove monomial factors from the trinomial before extracting the binomial factors. This means that in the example just given it was not necessary to try the first combination, $(3a - 2)\ (2a - 6)$, since 2 is a factor of $(2a - 6)$ but not of the trinomial. Likewise, the third combination $(6a - 3)\ (a - 4)$ was unnecessary since 3 is a factor of $(6a - 3)$, but not of the trinomial.

Fractions and Fractional Equations

A fraction is an indicated division in which the *numerator* (number above the line) is the dividend and the *denominator* (number below the line) is the divisor. The ratio of two quantities (the value of the fraction) is unchanged if they are both multiplied or divided by the same number (not zero). Dividing both numerator and denominator by the same number is called reduction to lower terms.

Examples

1. $\frac{14}{21} = \frac{(2)\ (7)}{(3)\ (7)} = \frac{2}{3}$

2. $\frac{6ax}{9a} = \frac{(2x)\ (3a)}{(3a)\ (3a)} = \frac{2x}{3a}$

3. $\frac{a^2 - b^2}{a^2 - 2ab + b^2} = \frac{(a + b)\ (a - b)}{(a - b)\ (a - b)} = \frac{a + b}{a - b}$

Note that a quantity can be used as a divisor only if it is a *factor* of the *complete* numerator and *complete* denominator.

4. $\frac{e^2 - 14e - 51}{e^2 - 2e - 15} = \frac{(e - 17)\ (e + 3)}{(e - 5)\ (e + 3)} = \frac{e - 17}{e - 5}$

To change a given fraction to an equivalent fraction with a new denominator, multiply both numerator and denominator by a number which will make the new denominator the desired value. For example, to change $\frac{1}{5}$ to 25ths, multiply both the denominator and numerator by 5. The equivalent fraction is $\frac{5}{25}$.

Examples

1. *Given:* $\frac{2a}{5b} = \frac{?}{10ab}$, *find the equivalent numerator.*

Dividing 10ab by 5b gives 2a, so both terms of the fraction must be multiplied by 2a to change to the desired denominator.

Thus, $\frac{2a}{5b} \times \frac{2a}{2a} = \frac{4a^2}{10ab}$ *and* $\frac{2a}{5b} = \frac{4a^2}{10ab}$

2. *Given:* $\frac{2m + 1}{3m - 5} = \frac{?}{6m^2 - 7m - 5}$

The factors of $6m^2 - 7m - 5$ *are* $(3m - 5)\ (2m + 1)$

Hence $\frac{(6m^2 - 7m - 5)}{(3m - 5)} = (2m + 1)$

Multiplying,

$$\frac{2m + 1}{3m - 5} \times \frac{2m + 1}{2m + 1} = \frac{(2m + 1)\ (2m + 1)}{(3m - 5)\ (2m + 1)} = \frac{4m^2 + 4m + 1}{6m^2 - 7m - 5}$$

Fractions may be combined (added or subtracted) only if they have the same common denominator. When you need to combine fractions with different denominators, first change them to equivalent fractions with the same denominator (called *common denominator*). The least common denominator (LCD) is best since the use of any other necessitates further reduction of the answer to lower terms. When dealing with polynomial denominators, factor each denominator as far as possible. The least common denominator is then the product of all the factors of the denominators, taking each factor the greatest number of times it occurs in any one denominator.

Examples

1. *Combine:* $\frac{x}{4}+\frac{3x}{5}-\frac{2x}{6}$

$$L.C.D.=60$$

$$\frac{(15)x+(12)3x-10(2x)}{60}$$

$$\frac{15x+36x-20x}{60}=\frac{31x}{60}$$

2. *Combine:* $\frac{3}{4xy}-\frac{2}{3x^2}$

$$L.C.D.=12x^2y$$

$$\frac{(3x)3-(4y)2}{12x^2y}=\frac{9x-8y}{12x^2y}$$

3. *Combine:* $\frac{1}{R^2-1}+\frac{3}{R^2-2R+1}$

$$=\frac{1}{(R+1)\ (R-1)}+\frac{3}{(R-1)\ (R-1)}$$

$$=\frac{1(R-1)+3(R+1)}{(R-1)\ (R-1)\ (R+1)}$$

$$=\frac{R-1+3R+3}{(R-1)\ (R-1)\ (R+1)}$$

$$=\frac{4R+2}{(R-1)^2\ (R+1)}$$

4. *Combine:* $3R+5-\frac{5}{2R+1}$

$$=\frac{3R+5}{1}-\frac{5}{2R+1}$$

$$=\frac{(3R+5)\ (2R+1)-5}{2R+1}$$

$$=\frac{6R^2+13R+5-5}{2R+1}=\frac{6R^2+13R}{2R+1}$$

The expression combined in example 4 is known as a *mixed expression* and the answer, as an *improper fraction*. Handle them as mixed numbers and improper fractions are handled in arithmetic.

To multiply two fractions together, find the product of the numerators and of the denominators. Frequently it is possible to reduce the result to lower terms by dividing out *factors* common to both numerator and denominator. This is sometimes called *cancellation of common factors*. To divide one fraction by another, invert the second fraction and multiply. Factoring of polynomial numerators and denominators is advisable in order to eliminate common factors from the result.

Examples

1. *Perform the indicated operations:* $\frac{4a}{3b}\times\frac{6ab}{2a^2}\div\frac{7ac}{9c^2}$

$$=\frac{4\cancel{a}}{3\cancel{b}}\times\frac{\overset{2}{\cancel{6}}\cancel{a}\cancel{b}}{\cancel{2}\underset{a}{\cancel{a^2}}}\times\frac{9\overset{c}{\cancel{c^2}}}{7a\cancel{c}}=\frac{36c}{7a}$$

Note the use of cancellation in this example. Wherever fractions are being multiplied, you can cancel out like terms. Note in the example that an a above the division line is cancelled and that the a^2 below is replaced by an a. Remember,

$$\frac{a}{a^2}=\frac{a}{aa}=\frac{a}{a}\times\frac{1}{a}=\frac{1}{a}$$

Therefore, when you see an a above and a^2 below, you can do the steps just outlined mentally, leaving, in this case, an a below the line. But since there is an a in the first fraction, this left-over a can be cancelled into it. Further cancellation removes practically two-thirds of the fraction.

2. *Multiply:* $\frac{m^2+5m+6}{m^2-4}\times\frac{m^2-6m+8}{m^2-9}$

$$=\frac{\cancel{(m+2)}\ \cancel{(m+3)}}{\cancel{(m+2)}\ \cancel{(m-2)}}\times\frac{\cancel{(m-2)}\ (m-4)}{\cancel{(m+3)}\ (m-3)}$$

$$=\frac{m-4}{m-3}$$

3. *Perform the indicated operations:*

$$\frac{5t^2+8t+3}{2t^2-6t}\times\frac{6t^2-18t}{5t^2-2t-3}\div\frac{t^2+5t+4}{t^2+3t-4}$$

$$=\frac{5t^2+8t+3}{2t^2-6t}\times\frac{6t^2-18t}{5t^2-2t-3}\times\frac{t^2+3t-4}{t^2+5t+4}$$

$$=\frac{\cancel{(5t+3)}\ \cancel{(t+1)}}{2t\cancel{(t-3)}}\times\frac{6t\cancel{(t-3)}}{\cancel{(5t+3)}\ \cancel{(t-1)}}\times\frac{\cancel{(t-1)}\ \cancel{(t+4)}}{\cancel{(t+1)}\ \cancel{(t+4)}}$$

$$=\frac{6t}{2t}=3$$

Complex fractions are fractions whose numerator or denominator, or both, contain fractions. They can be simplified by treating them as division problems.

Example

Simplify:

$$\frac{R_1\left(\frac{R_2R_3}{R_2+R_3}\right)}{R_1+\frac{R_2R_3}{R_2+R_3}}$$

(The numbers 1, 2, 3 below and to the right of the R's are subscripts. They indicate that the R's are different quantities and are to be treated as if they were different letters.)

$$=R_1\left(\frac{R_2R_3}{R_2+R_3}\right)\div\left(R_1+\frac{R_2R_3}{R_2+R^3}\right)$$

$$=\frac{R_1R_2R_3}{R_2+R_3}\div\frac{R_1R_2+R_1R_3+R_2R_3}{R_2+R_3}$$

$$=\frac{R_1R_2R_3}{\cancel{R_2+R_3}}\times\frac{\cancel{R_2+R_3}}{R_1R_2+R_1R_3+R_2R_3}$$

$$=\frac{R_1R_2R_3}{R_1R_2+R_1R_3+R_2R_3}$$

The solution of fractional equations is not so difficult as you might think, for you can multiply both sides of an equation by the same number (axiom 2). Multiplying all the fractions by the common denominator, you obtain an equation involving whole numbers, which you can then solve as previously outlined.

Examples

1. *Solve for a in the equation* $\frac{6a}{7}-\frac{1}{3}=\frac{5a}{14}+\frac{13}{6}$.

   Given: $\frac{6a}{7}-\frac{1}{3}=\frac{5a}{14}+\frac{13}{6}$

   42 is the common denominator

   Multiply both sides by 42,
   $$36a-14=15a+91$$
   Add 14 to both sides,
   $$36a=15a+105$$
   Subtract 15a from both sides,
   $$21a=105$$
   Divide both sides by 21,
   $$a=5$$

2. *Solve for x in the equation,* $\frac{8}{x}+\frac{1}{2}=2-\frac{7}{x}$

   Given: $\frac{8}{x}+\frac{1}{2}=2-\frac{7}{x}$

   Multiply both sides by 2x,
   $$16+x=4x-14$$
   Subtract 16 from both sides,
   $$x=4x-30$$
   Subtract 4x from both sides,
   $$-3x=-30$$
   Divide both sides by −3,
   $$x=10$$

3. *Solve for m in the equation* $\frac{2m+1}{4m+1}=\frac{2m+5}{4m+7}$.

   Given: $\frac{2m+1}{4m+1}=\frac{2m+5}{4m+7}$

   Multiply both sides by (4m+1) (4m+7),
   $$(2m+1)\ (4m+7)=(2m+5)\ (4m+1)$$
   $$8m^2+18m+7=8m^2+22m+5$$
   Subtract $8m^2$ from both sides,
   $$18m+7=22m+5$$
   Subtract 7 from both sides,
   $$18m=22m-2$$
   Subtract 22m from both sides,
   $$-4m=-2$$
   Divide both sides by −4,
   $$m=0.5$$

4. *Solve for r_t in the equation,* $\frac{1}{r_t}=\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}$

   Given: $\frac{1}{r_t}=\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}$

   Multiply both sides by (r_t) (r_1) (r_2) (r_3),
   $$r_1r_2r_3=r_2r_3r_t+r_1r_3r_t+r_1r_2r_t$$
   Factor the right hand member,
   $$r_1r_2r_3=r_t(r_2r_3+r_1r_3+r_1r_2)$$
   Divide both sides by $(r_2r_3+r_1r_3+r_1r_2)$,
   $$\frac{r_1r_2r_3}{r_2r_3+r_1r_3+r_1r_2}=r_t$$

Simultaneous Linear Equations

As long as equations contain only one unknown, the simple rules for solution apply, but when you encounter a problem with two unknowns, you need to use *simultaneous linear equations.*

Simultaneous linear equations are two or more equations which contain only first powers of the unknown quantities and no products of unknowns. They are true for certain values of the unknowns. Among the various methods for finding the values of the unknowns are solution by graphing, elimination by addition or subtraction, elimination by substitution, and the use of determinants. Of these, the most common method is elimination by addition or subtraction. Therefore, it is well to understand this method.

If the letter you wish to eliminate has the same coefficient in the two equations, merely add (or subtract) the two equations. If the coefficients are different, you can make them the same by multiplying both members of each equation by some number, as you can see in the sample problem.

Example

Solve for e and i in the following pair of equations,
$$2e+10i=25 \quad (1)$$
$$5e-\ 8i=46 \quad (2)$$
Multiply equation (1) by 4,
$$8e+40i=100 \quad (3)$$
Multiply equation (2) by 5,
$$25e-40i=230 \quad (4)$$
Add equation (4) to equation (3),
$$33e=330 \quad (5)$$
Divide both sides by 33,
$$e=10 \quad (6)$$
Substitute 10 for e in (1),
$$20+10i=25 \quad (7)$$
Subtract 20 from both sides,
$$10i=5 \quad (8)$$
Divide both sides by 10,
$$i=0.5 \quad (9)$$

In this example, the coefficients of i were made equal by multiplying equation (1) by 4, and equation (2) by 5. You could have eliminated the e's by multiplying equation (1) by 5, and equation (2) by 2. While equations are not always in the form of (1) and (2), you can put them in such a form by simplifying and transposing.

If you have three equations with three unknown quantities, follow the same procedure of elimination. Eliminate one unknown by combining one pair of equations, then eliminate the same unknown from another pair of equations. This results in two equations with two unknowns, which you can solve as explained before.

Example

$$9i_1 - 4i_2 - 7i_3 = 4 \quad (1)$$
$$5i_1 + 3i_2 - 10i_3 = 3 \quad (2)$$
$$2i_1 - 5i_2 + 9i_3 = 2 \quad (3)$$

First, eliminate i_3 between (1) and (2)

Multiply equation (1) by 10,

$$90i_1 - 40i_2 - 70i_3 = 40 \quad (4)$$

Multiply equation (2) by 7,

$$35i_1 + 21i_2 - 70i_3 = 21 \quad (5)$$

Subtract equation (5) from equation (4),

$$55i_1 - 61i_2 = 19 \quad (6)$$

Next, eliminate i_3 from (1) and (3),

Multiply equation (1) by 9,

$$81i_1 - 36i_2 - 63i_3 = 36 \quad (7)$$

Multiply equation (3) by 7,

$$14i_1 - 35i_2 + 63i_3 = 14 \quad (8)$$

Add equation (7) and equation (8),

$$95i_1 - 71i_2 = 50 \quad (9)$$

Now you have two equations with two unknowns—equations (6) and (9).

$$55i_1 - 61i_2 = 19 \quad (6)$$
$$95i_1 - 71i_2 = 50 \quad (9)$$

Multiply equation (6) by 19,

$$1045i_1 - 1159i_2 = 361 \quad (10)$$

Multiply equation (9) by 11,

$$1045i_1 - 781i_2 = 550 \quad (11)$$

Subtract equation (11) from equation (10),

$$-378i_2 = -189 \quad (12)$$

Divide equation (12) by −378,

$$i_2 = 0.5$$

Substitute 0.5 for i_2 in equation (6),

$$55i_1 - 30.5 = 19$$

Add 30.5,

$$55i_1 = 49.5$$

Divide by 55,

$$i_1 = 0.9$$

Substitute 0.5 for i_2 and 0.9 for i_1 in equation (1),

$$8.1 - 2.0 - 7i_3 = 4$$

Collect terms,

$$6.1 - 7i_3 = 4$$

Subtract 6.1,

$$-7i_3 = -2.1$$

Divide by −7,

$$i_3 = 0.3$$

Check by substituting in equations (1), (2), and (3),

$$8.1 - 2.0 - 2.1 = 4$$
$$4 = 4$$
$$4.5 + 1.5 - 3.0 = 3$$
$$3 = 3$$
$$1.8 - 2.5 + 2.7 = 2$$
$$2 = 2$$

Exponents, Radicals, and Complex Numbers

You have already noted that a number raised to the zero power is equal to one, and that a negative exponent is the same as the reciprocal of the quantity to the same positive exponent. Now consider the significance of a fractional exponent. Squaring the quantity $x^{1/2}$ gives $(x^{1/2})\ (x^{1/2})$. By adding the exponents $x^{(1/2+1/2)}$ as in any other multiplication with exponents, you find that the result is x^1, or x. By this reasoning, you can see that $x^{1/2}$ is the square root of x. In like manner, $x^{1/3}$ is the cube root of x, $x^{1/4}$ is $\sqrt[4]{x}$. The expression $x^{2/3}$ is read x to the two-thirds power. This means that x is raised to the power of the numerator of the fractional exponent, and reduced by the root of the denominator. Thus, $x^{2/3}$ is the cube root of x squared. Mathematically, this becomes $\sqrt[3]{x^2}$.

Sometimes it is necessary to simplify an expression involving *radicals*, (square roots, cube roots, etc.) without changing its value. Sometimes you can divide the quantity under the radical (called *radicand*) into two factors, and take the indicated root of one of the factors.

Examples

1. $\sqrt{80} = \sqrt{16 \cdot 5} = 4\sqrt{5}$
2. $\sqrt{54} = \sqrt{9 \cdot 6} = 3\sqrt{6}$
3. $\sqrt[3]{16m} = \sqrt[3]{8 \cdot 2m} = 2\sqrt[3]{2m}$

When the quantity under the radical is a fraction, multiply both numerator and denominator by a quantity which will make it possible to take the indicated root of the denominator.

Examples

1. $\sqrt{\frac{2}{3}} = \sqrt{\frac{2}{3} \cdot \frac{3}{3}} = \sqrt{\frac{6}{9}} = \sqrt{\frac{1}{9} \cdot 6} = \frac{1}{3}\sqrt{6}$
2. $\sqrt[3]{\frac{40}{x^2}} = \sqrt[3]{\frac{40}{x^2} \cdot \frac{x}{x}} = \sqrt[3]{\frac{40x}{x^3}} = \sqrt[3]{\frac{8}{x^3} \cdot 5x} = \frac{2}{x}\sqrt[3]{5x}$

You may combine (add or subtract) radicals if they differ only in the coefficient, that is, if they have the same radicand and root.

Examples

1. $4\sqrt{x}+3\sqrt{x}-2\sqrt{x}=5\sqrt{x}$

2. $\sqrt{4x}+5\sqrt{x}=2\sqrt{x}+5\sqrt{x}=7\sqrt{x}$

3. $\sqrt[3]{54}+\sqrt[3]{6\frac{3}{4}}=\sqrt[3]{27\cdot 2}+\sqrt[3]{\frac{27}{4}}=3\sqrt[3]{2}+\sqrt[3]{\frac{27}{4}\cdot\frac{2}{2}}$

$=3\sqrt[3]{2}+\sqrt[3]{\frac{54}{8}}=\sqrt[3]{2}+\sqrt[3]{\frac{27}{8}\cdot 2}$

$=3\sqrt[3]{2}+\frac{3}{2}\sqrt[3]{2}=4\frac{1}{2}\sqrt[3]{2}$

You can multiply two radicals together provided they have the same index (indicated root). You can do this by multiplying the coefficients together for the coefficient of the product and multiplying the radicands together for the radicand of the product.

Examples

1. $(2\sqrt{3})(3\sqrt{2})=6\sqrt{6}$

2. $(7\sqrt{x})(2\sqrt{y})=14\sqrt{xy}$

3. $(2\sqrt{x})(3\sqrt{x})=6\sqrt{x^2}=6x$

4. $a\sqrt{b}\,(2\sqrt{3}+4)=2a\sqrt{3b}+4a\sqrt{b}$

5. $(5\sqrt{3}+2\sqrt{6})\,(5\sqrt{3}+4\sqrt{6})$

Multiply:

$$\begin{array}{r} 5\sqrt{3}+2\sqrt{6} \\ 5\sqrt{3}+4\sqrt{6} \\ \hline 25\sqrt{9}+10\sqrt{18} \qquad\qquad \\ 20\sqrt{18}+8\sqrt{36} \\ \hline 25\sqrt{9}+30\sqrt{18}+8\sqrt{36} \end{array}$$

$=25\cdot 3+30\sqrt{9\cdot 2}+8\cdot 6$

$=75+30\cdot 3\sqrt{2}+48$

$=123+90\sqrt{2}$

To divide expressions containing radicals, use the opposite process—that is, divide coefficient by coefficient and radicand by radicand. This gives an expression with a radical in the denominator so frequently that a special process, called *rationalizing the denominator*, becomes necessary. This process changes the denominator to a whole number, a fraction, or a mixed number.

You can rationalize the denominator by multiplying both the numerator and the denominator by a quantity that will eliminate radicals from the denominator. In the case of monomials this is quite simple, and is the same process as discussed previously under *simplification of fractional radicands.*

Examples

$$\frac{\sqrt{5}}{\sqrt{6}}=\sqrt{\frac{5}{6}}=\sqrt{\frac{5}{6}\cdot\frac{6}{6}}=\sqrt{\frac{30}{36}}=\sqrt{\frac{1}{36}\cdot 30}=\frac{1}{6}\sqrt{30}$$

When the denominator is a binomial, rationalize by multiplying by the binomial with the sign between terms changed. This changed binomial is called the *conjugate* of the denominator. Since nearly all applications of this process in radio and electrical circuits deal with binomials, it is not necessary to discuss nor give examples involving more terms.

Examples

1. $\frac{5}{3-\sqrt{2}}=\frac{5}{3-\sqrt{2}}\cdot\frac{3+\sqrt{2}}{3+\sqrt{2}}=\frac{15+5\sqrt{2}}{9-\sqrt{4}}=\frac{15+5\sqrt{2}}{9-2}$

$=\frac{15+5\sqrt{2}}{7}$

Here the denominator was $3-\sqrt{2}$ and the conjugate was $3+\sqrt{2}$

2. $\frac{x}{\sqrt{x}+a}=\frac{x}{\sqrt{x}+a}\cdot\frac{\sqrt{x}-a}{\sqrt{x}-a}=\frac{x\sqrt{x}-ax}{\sqrt{x^2}-a^2}=\frac{x\sqrt{x}-ax}{x-a^2}$

3. $\frac{4-\sqrt{m}}{2+\sqrt{m}}=\frac{4-\sqrt{m}}{2+\sqrt{m}}\cdot\frac{2-\sqrt{m}}{2-\sqrt{m}}=\frac{8-6\sqrt{m}+\sqrt{m^2}}{4-\sqrt{m^2}}$

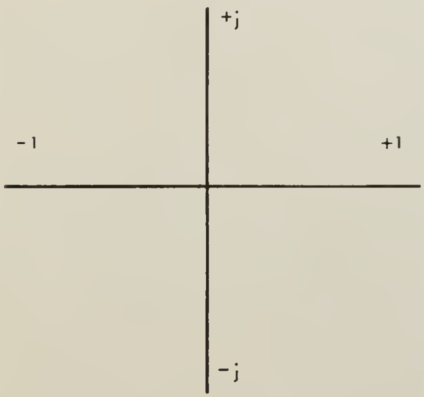
$\frac{8+m-6\sqrt{m}}{4-m}$

4. $\frac{\sqrt{r}-\sqrt{s}}{\sqrt{r}+2\sqrt{s}}=\frac{\sqrt{r}-\sqrt{s}}{\sqrt{r}+2\sqrt{s}}\cdot\frac{\sqrt{r}-2\sqrt{s}}{\sqrt{r}-2\sqrt{s}}$

$\frac{\sqrt{r^2}-3\sqrt{rs}+2\sqrt{s^2}}{\sqrt{r^2}-4\sqrt{s^2}}=\frac{r+2s-3\sqrt{rs}}{r-4s}$

Note that in each of the foregoing examples the final result has no radicals in the denominator.

Thus far you have dealt with roots of positive quantities. Consider now negative quantities appearing under radicals, such as $\sqrt{-x}$ or $\sqrt{-3}$. Obviously, there is no quantity which, when squared, will give $-x$ or -3. The indicated square roots of negative quantities are designated *imaginary numbers*. In algebra, treat them as follows: $\sqrt{-x}=\sqrt{x}\cdot\sqrt{-1}=i\sqrt{x}$; $\sqrt{-3}=\sqrt{3}\cdot\sqrt{-1}=i\sqrt{3}$. In each case, i stands for the imaginary quantity, the square root of -1. In electrical work the symbol j is used instead of i because the letter i is used to represent current. The quantity $\sqrt{-1}$ is usually referred to as the *j operator* and is used frequently in the solution of AC current problems.



Positions of a Unit Vector

Real and imaginary quantities can be graphically represented by four positions of a unit vector, as shown above. Positive real numbers are plotted to the right of the origin, and negative real numbers to the left along the horizontal axis, which is known as the *axis of reals*. Imaginaries which have *positive* signs are plotted *above* the origin and negative imaginary quantities below the origin along the vertical axis which is known as the axis of imaginaries.

Real and Imaginary Numbers

A *complex number* is the sum or difference of a real quantity and an imaginary quantity. For example, $5-jb$ and $3+j2$ are complex numbers. The term complex number is really inappropriate because the system is not complex at all. *Rectangular notation* is a better designation. It is not difficult to apply the fundamental processes to this special notation, for two or more complex numbers can be combined by combining the real portions and imaginary portions separately.

Examples

Addition.

(1) $$\begin{array}{r} 5+j3 \\ 2-j2 \\ \hline 7+j \end{array}$$ (2) $$\begin{array}{r} 6-j4 \\ -5-j3 \\ \hline 1-j7 \end{array}$$

Subtraction.

(1) $$\begin{array}{r} 5+j3 \\ 2-j2 \\ \hline 3+j5 \end{array}$$ (2) $$\begin{array}{r} 6-j4 \\ -5-j3 \\ \hline 11-j \end{array}$$

To multiply complex numbers, use the same procedure as with binomials, except that, where j^2 occurs in the final result, you replace it by its equivalent, -1.

Examples

1. *Multiply* $5+j3$ *by* $2-j2$

$$\begin{array}{r} 5+j3 \\ 2-j2 \\ \hline 10+j6 \\ -j10-j^2 6 \\ \hline 10-j4-j^2 6 \end{array}$$

$$=10-j4-(-1)(6)=10-j4+6=16-j4$$

Divide one rectangular quantity (complex number) by another by rationalizing the denominator, then dividing the real number into the numerator. Remember that you rationalize the denominator by multiplying both numerator and denominator by the conjugate of the denominator.

Examples

1. *Divide* $(5+j3)$ *by* $(2-j2)$

$$\frac{5+j3}{2-j2}\cdot\frac{2+j2}{2+j2}=\frac{10+j16+j^2 6}{4-j^2 4}=\frac{10+j16-6}{4+4}$$

$$=\frac{4+j16}{8}=\frac{1+j4}{2} \text{ or } 0.5+j2$$

2. *Divide* $(6-j4)$ *by* $(-5-j3)$

$$\frac{6-j4}{-5-j3}\cdot\frac{-5+j3}{-5+j3}=\frac{-30+j38-j^2 12}{25-j^2 9}$$

$$=\frac{-30+j38+12}{25+9}=\frac{-18+j38}{34}=\frac{-9+j19}{17}$$

Quadratic Equations

The degree of an equation in which the unknown has only positive integral exponents and does not appear in the denominator of a fraction is the same as its term of highest degree. The degree of a term in a letter means its exponent in that term. To illustrate, the degree of the term $4x^2y^3$ in x is the second. The equation $ax^2 + bx + c = 0$ is a second degree or quadratic equation in x. The following discusses the solution of equations of this form.

There are several methods you can use to solve a quadratic equation, such as graphing, completing the square, factoring, and using a formula derived from the general form $ax^2+bx+c=0$ by the completion of the square method. The formula is $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ where a, b, and c are, respectively the coefficient of x^2, the coefficient of x, and the term which does not contain x. Before applying the formula, always put the equation in the standard form $ax^2+bx+c=0$ by transposition. The following examples illustrate the processes for solving quadratic equations by the formula.

Examples

1. *Solve for x in the equation* $3x^2+8x-10=0$.

   By the formula, $x = \frac{-8 \pm \sqrt{64+120}}{6}$

   $$= \frac{-8 \pm \sqrt{184}}{6} = \frac{-8 \pm 13.56}{6}$$

   $$= \frac{-21.56}{6} \text{ or } \frac{5.56}{6}$$

   $$= -3.59 \text{ or } 0.93$$

2. *Solve for a in the equation* $2a = 3a^2 - 8$.

   Subtract $3a^2$ *from both sides*

   $$-3a^2 + 2a = -8$$

   Add 8 to both sides

   $$-3a^2 + 2a + 8 = 0$$

   By formula

   $$a = \frac{-2 \pm \sqrt{4+96}}{-6} = \frac{-2 \pm \sqrt{100}}{-6}$$

   $$= \frac{-2 \pm 10}{-6}$$

   $$= \frac{-12}{-6} \text{ or } \frac{8}{-6} = 2 \text{ or } -1.33$$

3. *Solve for R in the equation* $\frac{R^2}{R+5} = \frac{R}{3} + \frac{5}{6}$

   $$6R^2 = 2R(R+5) + 5(R+5)$$

   $$6R^2 = 2R^2 + 10R + 5R + 25$$

   $$6R^2 = 2R^2 + 15R + 25$$

   $$4R^2 - 15R - 25 = 0$$

   $$R = \frac{15 \pm \sqrt{225+400}}{8}$$

   $$= \frac{15 \pm \sqrt{625}}{8} = \frac{15 \pm 25}{8}$$

   $$\frac{40}{8} \text{ or } \frac{-10}{8} = 5 \text{ or } -1.25$$

The quantity b^2-4ac which appears under the radical in the formula is called the *discriminant.* It indicates the type of roots. If b^2-4ac is positive, there are two real and unequal roots; if b^2-4ac is negative, the roots are imaginary and unequal; if b^2-4ac equals zero, the roots are real and equal.

LOGARITHMS

The *logarithm* of a quantity is the exponent, or the power, to which a given number, called the base, must be raised to equal that quantity. To illustrate, in the quantity $3^2=9$, the exponent, *2*, is called the logarithm of *9* to the base *3*. This relation is usually written $\log_3 9 = 2$. Any positive number greater than 1 might serve as a base. Two numbers have been selected, resulting in two systems of logarithms. One base, 2.718, usually indicated by the Greek letter epsilon (ϵ), is used in the *natural* logarithm system. The other base is 10; it is used in the *common* system of logarithms. In the common system, the base *10* is usually omitted in the logarithmetic expression. Thus $\log_{10} 1000 = 3$ is usually written $\log 1000 = 3$. In the natural system the base (ϵ) may be written in.

In the common system, logarithms that are exact powers of 10 are integers. Thus,

log 100	$=2$,	since 10^2	$=100$
log 1000	$=3$,	since 10^3	$=1000$
log 10000	$=4$,	since 10^4	$=10000$
log 0.1	$=-1$,	since 0.1	$=10^{-1}$
log 0.01	$=-2$,	since 0.01	$=10^{-2}$
log 1	$=0$,	since 0	$=10^0$

For numbers not exact powers of ten, the logarithm consists of two parts, an integral part (whole number) and a decimal part. The integral part is called the *characteristic* and the decimal part the *mantissa*. Thus, for example, $\log 595 = 2.7745$ (in words, the logarithm of *595 is 2.7745*), the characteristic is *2*, and the mantissa is *.7745*. The characteristic is found by inspection and the mantissa from logarithmetic tables.

CHARACTERISTIC. You can determine the characteristic by the following rules:

1. The characteristic of the logarithm of a number greater than 1 is positive and is one less than the number of digits to the left of the decimal point. For example, in the case of the log *595*, the characteristic is *2*, and for the log of *59.5*, the characteristic is *1*.

2. The characteristic of the logarithm of a number less than 1 is negative, and is equal to one more than the number of zeros immediately to the right of the decimal point. For example, for the log of *.0595*, the characteristic is *−2*, and for log *.00595*, the characteristic is *−3*.

When the characteristic is negative, do not put the minus sign in front of the logarithm, since it applies only to the characteristic and not to the entire logarithm. Instead, add *10* to the negative characteristic and indicate the subtraction of *10* at the end of the logarithm.

Thus the characteristic *−2* is written, *8.(mantissa)−10*, and the characteristic *−3* is written, *7.(mantissa)−10*.

Another method of indicating that the characteristic is negative is to place the minus sign above the characteristic. For example:

$\bar{2}$. (mantissa), and

$\bar{3}$. (mantissa)

MANTISSA. Find the mantissa from tables of logarithms. Numbers which have the same figures in the same order and differ only in the position of the decimal point have the same mantissa in their logarithms. For example, the mantissa of *595* is *.7745*; the mantissa of *59.5* is also *.7745*.

TO FIND THE LOGARITHM OF A NUMBER—1. Determine by inspection the characteristic of the number.

2. Find the mantissa from the tables. The mantissa of the number is independent of the position of the decimal point, so you can disregard the decimal point in the number when finding the mantissa. The mantissas in the table are the decimal part of the logarithm and therefore should be preceded by the decimal point.

In four-place logarithm tables, the first column in the table contains the first two digits of the numbers whose mantissas are given in the table, and the top row contains the third digit. Thus, to find the mantissa of *595*, find *59* in the left-hand column and *5* at the top. In the column under *5*, and opposite *59*, is *7745*, the mantissa. The logarithm of *595* is then *2.7745*.

3. To find the logarithm of a quantity with more than three digits, use the process called interpolation. Suppose you want to find the logarithm of *5956*. The tables do not give the mantissa for *5956*. However, they give the mantissas for *5950* and *5960*. (The mantissa for *5950* is the same as that for *595*. Likewise, the mantissa for *5960* is the same as that for *596*). Since *5956* lies between *5950* and *5960*, its mantissa must lie between the mantissa for these two numbers.

By arranging the mantissas in the following tabular form,

Mantissa of *5960* = *.7752*

Mantissa for *5956* = *?*

Mantissa for *5950* = *.7745*

you can see that *5956* is *6/10* of the way between *5950* and *5960*. Obviously, the mantissa of *5956* must be *6/10* of the way between *.7745* and *.7752*. Since the difference between the two is *.0007* and since *6/10* of *.0007* is *.00042*, add *.00042* to *.7745* (the mantissa of *5950*). The result, *.77492*, is the mantissa of *5956*. Therefore, the logarithm of *5956* = *3.77492*.

ANTILOGARITHMS. The number corresponding to a given logarithm is called the antilogarithm of that number. It is written *antilog* or log^{-1}. To find the antilog reverse the process for finding logarithms.

Examples

1. To find the antilog (log^{-1}) of *1.8987*, first look in the logarithm table and locate the mantissa *.8987*. It is in line with the number *79* and under the column headed *2*. Thus the number corresponding with the mantissa *.8987* has the digits *792*.

To determine the location of the decimal point, reverse the rule for finding a characteristic. If the characteristic were zero, the decimal point would be placed after the first digit (*7.92*). Since the characteristic is *1*, count two places to the right and place the decimal point after the *9*. Thus the antilog of *1.8987* is *79.2*.

2. Find the antilog (log^{-1}) of *2.4325*.

The tables do not show the mantissa *.4325*; therefore you must interpolate. The mantissa *.4325* lies between the two mantissas *.4330* and *.4314*. These mantissas correspond to *271* and *270* respectively. The difference between

.4330 and *.4314* is *.0016*, and the difference between *.4325* (the given mantissa) and the mantissa for *270* is *.0009*. Then find the number corresponding to *.4325* by adding $\frac{.0009}{.0016}$ of 1, or *.563* to *270*, thus giving the sum *270563*. Since the given characteristic is *2*, the antilog of *2.4325* is *270.563*.

Computations with Logarithms

To multiply two quantities, add their logarithms and find the antilog of the result.

Example

Find the product of 6952 and 437.

Solution

$$\text{Log } (6952 \times 437) = \log 6952 + \log 437.$$
$$\text{Log } 6952 = 3.8421$$
$$\text{Log } 437 = 2.6405$$
$$\text{Adding, } \log 6952 + \log 437 = 6.4826$$
Find the antilog 6.4826.

The antilog 6.4826 = 3,038,000.

Actual multiplication of 6952 by 437 would give the result 3,037,842. This indicates an error of 176 in over 3,000,000 or .006 of one percent. The error is due to the fact that the logarithm tables go to four places only. Greater accuracy would result in using five place tables. In general, though, accuracy obtained with four place tables is sufficient.

To divide two quantities, subtract the logarithm of the divisor from the logarithm of the dividend and find the antilog of the result.

Example

Find the quotient of 6952 divided by 437.

Solution

$$\text{Log } (6952 \div 437) = \log 6952 - \log 437$$
$$\text{Log } 6952 = 3.8421$$
$$\text{Log } 437 = 2.6405$$
$$\text{Subtracting, } \log 6952 - \log 437 = 1.2016$$
Find antilog 1.2016
$$\text{Antilog } 1.2016 = 15.908$$

To raise a quantity to any power, multiply the logarithm of the quantity by the exponent, or the power, and find the antilog of the result.

Examples

1. *Find the value of* $(5.2)^6$

Solution

$$\text{Log } (5.2)^6 = 6 \times \log (5.2)$$
$$\text{Log } (5.2) = 0.716$$
$$\text{Log } (5.2)^6 = 6 \times (0.716) = 4.296$$
$$\text{Antilog } 4.296 = 19768$$
$$\text{Therefore, } (5.2)^6 = 19768$$

2. *Find the value of* $(3.7)^{1/5}$. *(Exponent is fractional)*

Solution

$$\text{Log } (3.7^{1/5} = \frac{1}{5} \log (3.7)$$
$$\text{Log } (3.7) = 0.5682$$
$$\text{Then } \log (3.7)^{1/5} = \frac{1}{5} (0.5682) = 0.1136$$
$$\text{Antilog } 0.1136 = 1.2991$$
$$\text{Therefore, } (3.7)^{1/5} = 1.2991$$

3. *Find the value of* $(45.6)^{-3}$ *(Exponent is negative)*

Solution

$$\text{Log } (45.6)^{-3} = -3 \log (345.6)$$
$$\text{Log } 45.6 = 1.659$$
$$\text{Then, } \log (45.6)^{-3} = -3 (1.659) = -4.977$$
Since logarithm tables list only positive values of montissa, change -4.977 *to* $\overline{5}.023$ *(or* $5.023 - 10$*) by subtracting* -4.977 *from 10.*
$$\text{The antilog of } \overline{5}.023 = 0.000010544.$$
$$\text{Therefore, } (45.6)^{-3} = .000010544$$

To find the root of a quantity obtain the logarithm of the quantity, divide it by the index of the root, and find the antilog of the result.

Examples

1. *Find the value of* $\sqrt[3]{1.572}$

Solution

$$\text{Log } \sqrt[3]{1.572} = \frac{1}{3} \log 1.572$$
$$\text{Log } 1.572 = 0.19646$$
$$\text{Then } \log \sqrt[3]{1.572} = \frac{1}{3}(0.19646) = 0.06549$$
$$\text{Antilog } 0.06549 = 1.1611$$
$$\text{Therefore, } \sqrt[3]{1.572} = 1.1611$$

NATURAL LOGARITHMS

The natural system of logarithms uses a base ϵ which is approximately *2.718*. Natural logarithm tables give the complete logarithm rather than the mantissas only.

To find the natural log of a quantity which does not appear in the table and cannot be obtained by interpolation or if there is no natural logarithm table, use the common logarithm table and use the following relationships:

$$\log_\epsilon N = 2.3026 \log_{10} N$$
$$\log_{10} N = 0.4343 \log_e N$$

In multiplying, dividing, raising to powers, or finding roots with the natural system, use the same rules in the common logarithm system.

Example

Find the value of $\frac{54.3 \times 27.5}{13.65}$

Solution

$$\log_\epsilon \frac{54.3 \times 27.5}{13.65}$$

$$= \log_\epsilon 54.3 + \log_e 27.5 - \log_e 13.65$$

$$\text{Thus, } \log_\epsilon \frac{54.3 \times 27.5}{13.65}$$

$$= 3.9945 + 3.3190 - 2.6131$$
$$= 4.7004$$

$$\text{Antilog}^e\ 4.7004 = 109.9$$

Natural logarithms are frequently used in analyzing the charge and discharge action of condensers. Analysis of condenser action discussed in detail in chapter 5 makes use of the expression ϵ^{-x} in which ϵ is the base in the natural logarithm system.

TRIGONOMETRIC FUNCTIONS

Several special relationships, called *trigonometric functions*, hold true in a right triangle. Electrical problems when reduced to a right triangle can be easily and quickly solved by use of tables based upon these functions.

In the diagram below, θ is the angle *ZOR*, *OR* is the projection of *OZ* on the horizontal axis, *OX* is the projection of *OZ* on the vertical axis. The letters *r*, *x*, and *z* represent the lengths of *OR*, *OX* and *OZ* respectively. There are, in all, six different ratios between the sides *r*, *x* and *z*. They are called trigonometric ratios or *trigonometric functions*.

Notice that three of the functions, *sin θ*, *cos θ*, and *sec θ*, are the reciprocals of *cot θ*, *sec θ*, and *csc θ*, respectively. Generally you will use only the first three, sin θ, cos θ, and tan θ, in your work. You can save much time by memorizing them.

If you suppose that in the diagram, *OZ* has a unit length of 1 and is rotated in a counterclockwise direction beginning with angle θ at 0° value and continuing until it is 90°, then the functions will vary within the following limits:

sin θ increases from 0 to 1.0
cos θ decreases from 1.0 to 0
tan θ increases from 0 to ∞
cot θ decreases from ∞ to 0
sec θ increases from 1.0 to ∞
csc θ decreases from ∞ to 1.0

When θ varies between the values of 90° and 180° (second quadrant), the projection *r* is negative and the functions which involve *r* become negative. Thus, cos θ and tan θ are negative in this quadrant. In the third quadrant, both *r* and *x* are negative. Therefore, sin θ and cos θ which involve only one or the other, are negative, while tan θ, which involves both *r* and *x*, is positive. In the fourth quadrant, *r* is positive, but *x* is still negative. For this reason, sin θ and tan θ are negative in the fourth quadrant, while cos θ is positive.

VARIATIONS IN VALUE OF FUNCTIONS

Quadrant		Sin θ From	Sin θ To	Cos θ From	Cos θ To	Tan θ From	Tan θ To
I	(0°-90°)	0	1.0	1.0	0	0	∞
II	(90°-180°)	1.0	0	0	−1.0	−∞	0
III	(180°-270°)	0	−1.0	−1.0	0	0	∞
IV	(270°-360°)	−1.0	0	0	1.0	−∞	0

Trigonometric tables give functions up to 90° only. Therefore you will have special rules for angles in the other quadrants.

Sine of θ (written sin θ) $= \frac{x}{z}$

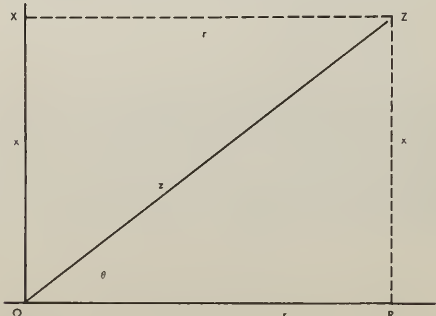
Cosine of θ (written cos θ) $= \frac{r}{z}$

Tangent of θ (written tan θ) $= \frac{x}{r}$

Cotangent of θ (written cot θ) $= \frac{r}{x}$

Secant of θ (written sec θ) $= \frac{z}{r}$

Cosecant of θ (written csc θ) $= \frac{z}{x}$



Trigonometric Functions

The following rules apply for functions of angles greater than 90°.

In quadrant II: $\theta = 90° +$ some angle which is designated α (alpha).

$\sin (90° + \alpha) = \sin (90° - \alpha)$
$\cos (90° + \alpha) = -\cos (90° - \alpha)$
$\tan (90° + \alpha) = -\tan (90° - \alpha)$

In quadrant III: $\theta = 180° + \alpha$

$\sin (180° + \alpha) = -\sin \alpha$
$\cos (180° + \alpha) = -\cos \alpha$
$\tan (180° + \alpha) = \tan \alpha$

In quadrant IV: $\theta = 270° + \alpha$

$\sin (270° + \alpha) = -\sin (90° - \alpha)$
$\cos (270° + \alpha) = \cos (90° - \alpha)$
$\tan (270° + \alpha) = -\tan (90° - \alpha)$

Examples

1. Find sin 29°.

Solution

Find the angle 29° in the table. Opposite this angle and under the heading sine is 0.4848. Therefore sine 29° = 0.4848

2. Find cos 129°.

Solution

Cos 129° = cos (90° + 39°) According to the rules for angles larger than 90°, cos (90° + 39°) = −cos (90° −39°) or −cos 51°. The tables show that cos 51° is 0.6293. Therefore cos 129° = −0.6293.

Interpolation is necessary only occasionally since the tables give functions of angles to tenths of a degree.

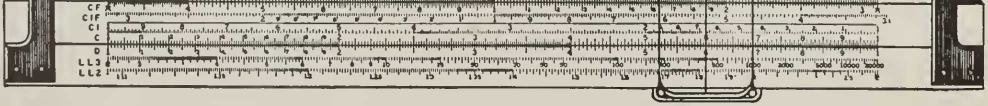
The angle corresponding to a given function is called the *inverse function* and is written *arc sin* θ or $\sin^{-1} \theta$. There are two angles corresponding to any given function. For example, arc sin 0.6428 is 40° or 140°. To avoid confusion, certain values of arc sin, arc cos, etc., are designated the principal values and are indicated by capitalizing the letter A of arc. The principal values of arc sin θ and arc tan θ are those in the first and fourth quadrants, and the principal values of arc cos θ are those in the first and second quadrants. Therefore Arc sin 0.6428 = 40°.

In finding the function of an angle, or vice versa, note that the functions repeat themselves every 360°. Thus sin 400° = sin (400° − 360°) = sin 40° = 0.6428. You can say that a negative angle is equivalent to a positive angle which is 360° plus the negative angle. To illustrate, −50° = 360° + (−50°) = 310°. Hence sin (−50°) = sin 310° = −0.7660. Angles in the fourth quadrant are frequently expressed as negative angles.

USE OF THE SLIDE RULE

Paper and pencil methods of making computations are at the best slow and often introduce unintentional errors. Mathematical computations can be done very quickly and quite accurately by the use of a slide rule. This instrument consists of two rulers, one sliding back and forth on the other. It can be used in solving many types of problems. It is only a tool, however, and without you, it can do nothing.

The log-log decitrig slide rule shown indicates the final answer of a computation in the form of three or four digits. For placing the decimal point there have been developed many rules. However, the surest method is to know the approximate answer, and from it locate the decimal point in the exact answer provided by the slide rule. For example, in multiplying 513 by 13, the sequence of digits is too confusing for mental calculation. According to the slide rule, the digits in the answer are 667. To locate the decimal, round off the numbers 513 and 13 to 500 and 10, respectively, and mentally calculate the product. From this product, 5000, you can determine that the decimal point should be located so that the answer given by slide rule reads 6670. By adopting such



Log-Log Decitrig Slide Rule

a rule for locating the decimal, you can avoid answers that are unreasonable.

The slide rule consists of several logarithmic scales so arranged that you can make computations by setting a movable portion, the *slide*, in a certain position relative to the fixed part, the *body* of the slide rule and reading the answer. The types of computations that are possible with a slide rule depend on the type of slide rule, but usually they include multiplication, division, squaring, cubing, extracting square and cube roots, finding logarithms and finding trigonometric functions. These processes are discussed briefly here, but before taking up the processes, it is well to know the meaning of several terms. The slide and body have already been identified. The glass runner is called the *indicator* and the mark upon it is the *hairline*. The mark associated with the number one on the body or slide is the *index*. Two positions on different scales are opposite each other when the hairline covers both positions at the same time.

Multiplication

Use the C and D scales to find the product of two numbers. To perform this operation, set the index of the C scale opposite the multiplicand on the D scale, and read the product on the D scale, opposite the multiplier on the C scale. This gives a number of three or four digits. Place the decimal point in accordance with a rough estimate. Note that the C scale has two indices, one at either end. If you find that when you set the left-hand index opposite the multiplicand, the multiplier on the C scale is beyond the index of the D scale, then you use the right-hand index of the C scale and proceed as previously indicated.

Division

In dividing two numbers, set the slide so that the divisor on the C scale is opposite the dividend on the D scale. The number on the D scale opposite the index of the C scale is then the quotient.

Squaring and Extracting Square Root

To square a number, set the indicator so that the hairline covers the number to be squared on the D scale, then read the result as the opposite number on the A scale. (C and B scales may be used in like manner.)

To extract the square root of a number, set the hairline covers the number to be squared on read the root opposite it on the D scale. Since the A scale is a double scale, the same number appears on it at two places. To avoid confusion resulting from this, use the *left-hand* portion of A for numbers between *1* and *10*, and the *right-hand* portion for numbers between *10* and *100*. In general, use the *left-hand* portion if the number has an *odd* number of digits and the *right-hand* portion if the number has an *even* number of digits to the left of the decimal point.

Logarithms

The L scale gives the mantissa of the common logarithm for the opposite number on the D scale. Determine the characteristic in the usual manner from the number of digits. Find the antilog by locating the number on the D scale opposite the mantissa on the L scale and putting the decimal point in the place indicated by the characteristic.

Reciprocals

The CI scale is made up of the reciprocals of the opposite numbers on the C scale. The D and DI scales give the same information.

Trigonometric Functions

To find the sine of an angle, set the hairline to the black number (representing the size of the angle in degrees) on the S scale (use the ST scale if the angle is less than any angle shown on the S scale) and read the sine from the C scale. When using the ST scale, the left index of C is 0.01 and the right index is 0.1; when using the S scale, the left index is 0.1 and the right index is 1.0.

You can find the cosine from the same scales. Since $\sin\theta = \cos(90° - \theta)$, it is possible to find the cosine of an angle by finding the sine of 90° minus that angle. The *red* figures on the S scale are the values of $90° - \theta$, (called complement). Thus to find $\cos\theta$, set the hairline to the red number on S scale representing θ and read the opposite number on the C scale.

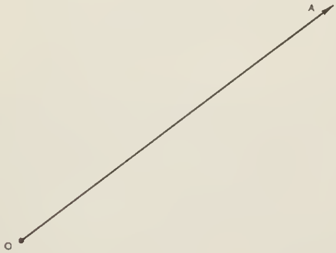
Use the following method in finding the tangent of an angle. For small angles, set the hairline to the angle on the ST scale and read the tangent on the C scale. The values on the C scale range from 0.01 at the left index to 0.1 at the right. For angles from 5.74° to 45°, set the hairline to the black figure on the T scale, and read the tangent as the opposite number on the C scale. These values range from 0.1 to 1.0. For angles greater than 45°, use the red figures on the T scale and read the result from the CI scale. The values range from 1 to 10.

Many other computations are possible with the slide rule. However, those given cover the bulk of calculations you will have to perform.

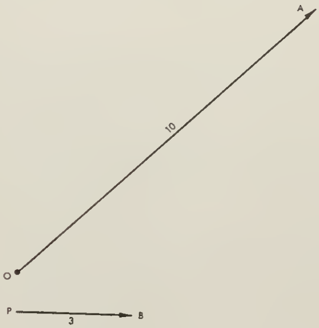
VECTOR QUANTITIES AND VECTOR ALGEBRA

Some quantities have magnitude only while others have both magnitude and direction. Quantities with magnitude only are known as *scalar* quantities; those with both magnitude and direction, as *vector* quantities. Forces, motion, and acceleration are examples of vector quantities. Scalar quantities may be added, subtracted, or multiplied directly. But since a second dimension enters into vector quantities, a number of special methods must be employed in dealing with vector quantities.

A vector quantity may be represented by a line segment. The length of the line indicates the magnitude and its angular position and an arrow point indicates the direction, as shown to the left. Usually this vector is called the *vector OA*. However, you may find it written $\overline{OA}$ or $\underline{OA}$. Other common notations, to indicate vector quantities, are $\overline{M}$, $\dot{M}$ and $\underset{\cdot}{M}$.

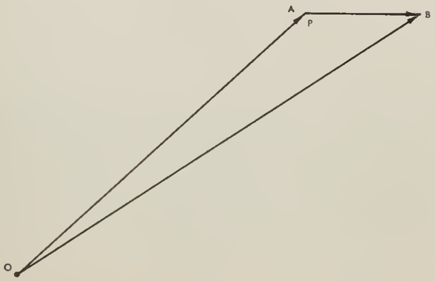


Representing a Vector Quantity



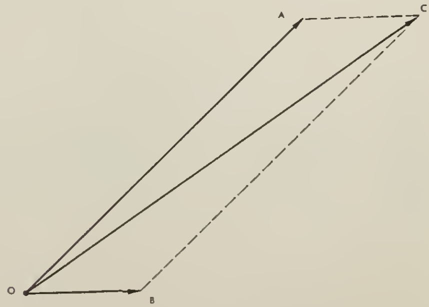
Representing Two Forces

Vector quantities may be combined, as you see in the illustrations. Suppose the two vectors OA and PB represent two forces applied to the same object. OA might be the force on an airplane due to the action of its engines and PB the force due to the wind. The vector sum of OA and PB is obtained by placing PB, without changing its magnitude or direction, so that P coincides with A. Then OB is the vector sum and represents the magnitude and direction of the resultant force.



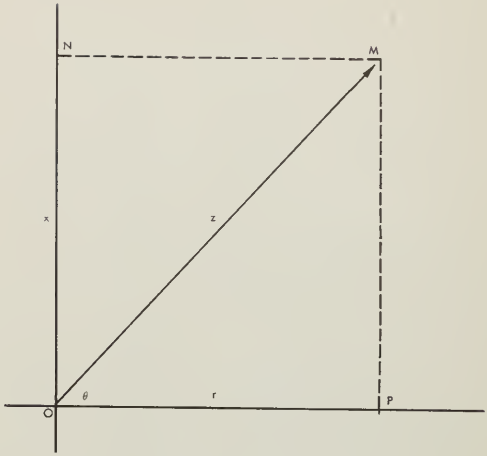
Addition of Vectors

You can obtain the same result by an alternate method. If you place the two vectors at the origin, and form a parallelogram with OA and OB as adjacent sides, the diagonal will be the resultant force. This is referred to as the *parallelogram law of forces* and may be applied to the addition of any set of vector quantities.



Parallelogram Method of Adding Vectors

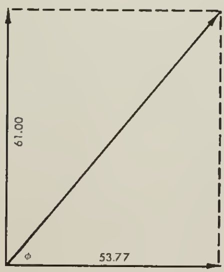
You can describe a vector in *polar* form or in *rectangular* form. The polar form shows the magnitude of the vector and the angle which it makes with the horizontal axis. For example, $10\underline{/30^\circ}$ describes a vector 10 units in length and at an angle of 30 degrees. In the rectangular form, the vector is resolved into two components which are the projections of the vector on the horizontal and vertical axis and have as their origin the initial point of the vector as to the right. In this work, you can designate the horizontal component as the *real* component and the vertical component as the *imaginary* or *j* component. From the relations in the definitions of sine and cosine, ($\sin\theta = x/z$ and $\cos\theta = r/z$) you can see that $r = z\cos\theta$ and $x = z\sin\theta$. Thus, $z\underline{/\theta} = z\cos\theta + jz\sin\theta$. Hence the vector $10\underline{/30^\circ} = 8.66 + j5$.



Components of a Vector

Addition and Subtraction

You have already observed how complex quantities (such as rectangular forms of vectors) can be added and subtracted. Since the addition of vectors by graphical means, as described before, is not satisfactory unless you use accurate instruments to construct the vectors, the usual method of combining two vectors is to convert them to rectangular form and to add them algebraically.



Examples

1. *Add* $35\underline{/40^\circ}$ *and* $47\underline{/55^\circ}$.

$$35\underline{/40^\circ} = 35\cos 40^\circ + j35\sin 40^\circ = 35(.7660) + j35(.6428) = 26.81 + j22.50$$

$$47\underline{/55^\circ} = 47\cos 55^\circ + j47\sin 55^\circ = 47(.5736) + j47(.8192) = 26.96 + j38.50.$$

Adding, $26.81 + j22.50$
$26.96 + j38.50$

The sum is $53.77 + j61.00$

The graphical addition of the vectors in this example is shown in the diagram directly above.

To convert this result to polar form, remember that $\frac{x}{r} = \tan\theta$. Therefore $\tan\theta$ in this case is $\frac{61.00}{53.77} = 1.134$, and $\theta = \text{arc}\tan 1.1343 = 48.6^\circ$. Since $\cos\theta = \frac{r}{z}$, $z = \frac{r}{\cos\theta} = \frac{53.77}{\cos 48.6^\circ} = \frac{53.77}{.6613} = 81.3$. In polar form the answer is $81.5\underline{/48.6^\circ}$.

2. *Subtract* $45.6\underline{/21.5^\circ}$ *from* $51.4\underline{/-10.5^\circ}$

$$45.6\underline{/21.5^\circ} = 45.6\cos 21.5^\circ + j45.6\sin 21.5^\circ = 45.6(.9304) + j45.6(.3665) = 42.4 + j16.7$$

$$51.4\underline{/-10.5^\circ} = 51.4\cos(-10.5^\circ) + j51.4\sin(-10.5^\circ) = 51.4(.9833) + j51.4(-.1822) = 50.5 - j9.36$$

Subtracting, $50.5 - j9.36$
$42.4 + j16.7$

The remainder is $8.1 - j26.06$

$$\theta \text{ arc}\tan\frac{-26.06}{8.1} = \text{arc}\tan -3.217 = -72.7^\circ$$

$$z = \frac{8.1}{\cos(-72.7^\circ)} = \frac{8.1}{.2957} = 27.1$$

In polar form the answer is $27.1\underline{/-72.7^\circ}$

Multiplication

To multiply two vectors in polar form, multiply the magnitudes together and add the angles.

Example

Multiply $55\underline{/40^\circ}$ *by* $47\underline{/55^\circ}$.
$(55) \bullet (47) = 2585$
$40^\circ + 55^\circ = 95^\circ$

Hence $(55\underline{/40^\circ})(47\underline{/55^\circ}) = 2585\underline{/95^\circ}$

You can check this result by converting from polar to rectangular form, multiplying and then converting back to polar form.

Example

$$55\underline{/40^\circ} = 42.13 + j35.35$$
$$47\underline{/55^\circ} = 26.96 + j38.5$$

$$\begin{array}{ll} & 42.13 + j35.35 \\ & 26.96 + j38.5 \\ \hline \textit{Multiplying} & 1135.825 + j953.036 \\ & \qquad j1622.005 + j^2 1360.975 \\ \hline \textit{The product is} & 1135.825 + j2575.041 - 1360.975 \\ & = -225.15 + j2575.04 \end{array}$$

$$\theta = \text{arc tan } \frac{2575.04}{-225.15} = \text{arc tan } -11.43 = 95^\circ$$

$$z = \frac{-225.15}{\cos 95^\circ} = \frac{-225.15}{-.0872} = 2582$$

The answer is $2582\underline{/95^\circ}$, *practically the same as obtained before.*

Division

To divide one vector by another, divide the magnitudes and subtract angles.

Example

Divide $(55\underline{/40^\circ})$ *by* $(47\underline{/55^\circ})$

$55 \div 47 = 1.17$

$40^\circ - 55^\circ = -15^\circ$

Hence $$\frac{55\underline{/40^\circ}}{47\underline{/55^\circ}} = 1.17\underline{/-15^\circ}$$

Since a power of a quantity is a repeated multiplication process, note that $(15\underline{/20^\circ})^2 = 225\underline{/40^\circ}$, $(15\underline{/20^\circ})^3 = 3375\underline{/60^\circ}$, and so on. Raise the magnitude of the vector in polar form to the desired power and multiply the angle by the exponent.

Conversely, to extract a root of a vector in polar form, extract the required root of the magnitude and divide the angle by the index of the root taken.

Examples

$$\sqrt{10\underline{/50^\circ}} = 3.16\underline{/25^\circ}$$
$$\sqrt[3]{10\underline{/50^\circ}} = 2.15\underline{/16.7^\circ}$$

When either extracting a root or raising to a power, convert rectangular quantities to polar form before finding the root or power. Addition and subtraction are easier in the rectangular form. Multiplication, division, raising to a power, and extracting a root are easier in the polar form.

CHAPTER 2

DC and AC Circuits

ELECTRICAL FUNDAMENTALS

Radar, like all other applications of electricity from the simplest appliance to the most complicated piece of electrical machinery, is based primarily on a number of fundamental principles of electricity. Any attempt to study radar without first learning these fundamentals could at the best result only in mere drudgery or most likely in complete failure. The purpose of this chapter is to give you the more important fundamentals of electricity; those which you need in order to study radar profitably. This chapter discusses the basic concepts of electricity and magnetism, describes the behavior of direct and alternating current in a large variety of electrical circuits, and presents a large number of formulas and equations essential to your understanding of fundamental electricity.

ELECTRON THEORY

According to science, the smallest particle into which any substance can be divided and still retain its characteristics is the *molecule*. The substance might be anything, such as the air you breathe, the water you drink, or the food you eat. There are thousands of kinds of substances in the world. The molecule, in turn, is built of atoms; the molecule of water, for example, is made up of two hydrogen atoms and one oxygen atom. A substance that contains only atoms which have the same chemical properties is called an *element*. Oxygen is an element since it contains only oxygen atoms. Likewise, nitrogen is an element because it is entirely composed of nitrogen atoms. There are some ninety odd elements known to science.

Finally, the atom itself consists of protons and electrons. The proton is a *positive* charge of electricity, and the electron is a *negative* charge of electricity. The protons together with another type of particle—neutrons—which are really combinations of protons and electrons form a closely packed nucleus in the center of the atom. The electrons spin around this nucleus in the same manner as the planets move in orbits around the sun. It is impossible to subdivide electrons and protons further.

One of the fundamental laws of electricity is that *like* charges *repel* each other, and *unlike* charges *attract* each other. This law explains the bond that exists in the atom between the positively charged nucleus and the electrons revolving about the nucleus in planetary elliptical orbits. The radii of the orbits of the various electrons within an atom differ. Some electrons are close to the nucleus; others are farther away. Electrons in the inner orbits near the nucleus are bound tightly in the atom. Electrons in the outer orbits are rather loosely bound and when influenced by outside forces, may be torn away from the nucleus. This leaves the nucleus with a deficiency of electrons, a condition in which it is called an *ion*, or a *positively* charged atom. The electrons that are loosely bound to the nucleus are called *free electrons* and exist in greater numbers in the atoms of some elements than in others. Metals contain more free electrons than do such substances as rubber and glass.

When a metal bar or wire is connected between a point with an excessive number of electrons and another point with a normal amount or a deficiency of electrons, there will be a movement of the free electrons. This movement is explained by the fact that an electrical pressure exists whenever there are more electrons at one point than at the other. The greater the difference in the number of electrons, the greater the electrical pressure and the greater the amount of electron flow. Likewise, the less the difference in the number of electrons, the less the pressure, and the less the amount of electron flow.

In addition to the electrical pressure exerted, the amount of electron flow depends upon the physical characteristics of the metal connector. A substance such as a metal, which contains many free electrons, is called a *conductor* and offers little opposition to the flow of electrons. Glass, which has few free electrons, is called an *insulator* and offers great opposition to electron flow. Actually, there is no such thing as a perfect conductor or a perfect insulator. A substance is a good conductor or a poor conductor, a good insulator or a poor insulator.

ELECTROSTATICS

The fact that an object such as a comb when rubbed with a cloth will attract light objects, such as pieces of paper, has been known for a long time. The early Greeks were familiar with this phenomenon, and unknowingly discovered the type of electricity which today is known as *static* electricity and which for some time was referred to as electricity at rest. The Greeks knew that when they rubbed a piece of amber, which they called *electron*, with a piece of cloth, it would attract other objects such as bits of cloth or pith, in a manner very much like a magnet. From the Greek word for amber are derived the English words, *electron* and *electricity*.

In general, much of the knowledge of electricity has been obtained from experiments in static electricity. Among these, for example, is that of Guericke, who, in 1663, discovered that a globe of sulphur which was mounted on an axis became electrified when it was rotated while being pressed with the dry palm of the hand; and that of Benjamin Franklin whose experiment with a kite led to the development of the lightning rod. Although radar deals with electricity in motion (dynamic electricity), an understanding of the effects of static electricity is basic to your study of radar.

Originally, static electricity was considered electricity at rest, but with the acceptance of the electron theory which states that electrons are in continual motion in the atoms, static electricity is usually thought of as electricity associated with insulators and dielectrics. A dielectric is the insulating material between a pair of oppositely charged bodies. Insulating materials are excellent dielectrics because they contain very few free electrons. Although there is no or little movement of free electrons in a dielectric, a dielectric is capable of transmitting a force from one body to another. The electric condenser, which you will study later, is an example of two oppositely charged bodies separated by a dielectric which under certain conditions transmits a force. A simple condenser consists of two metal plates or conductors separated by air for the dielectric. When it is connected to a source of voltage, such as a battery, one plate will be charged negatively (excess of electrons) and the other plate positively (lack of electrons). No electrons move through the air (dielectric), but if the condenser is charged too high (a condenser is charged when there are more electrons on one plate than the other), the dielectric will break down and a spark will jump between the plates, discharging the condenser.

The clouds and earth form an excellent example of a charged condenser in which the breakdown of the dielectric, which is air, manifests itself as lightning. The generally accepted theory of lightning is that the moisture in the clouds and the movement of the clouds themselves cause electrons to be stacked up or wiped off the clouds. When the number of electrons on a cloud becomes excessively high or excessively low, the dielectric will break down and a spark, which is called lightning, will jump from one cloud to another, or between the earth and the cloud. A lightning rod on the ground during a lightning discharge acts as a partial conductor and conducts the electrons to the earth or permits electrons to go up and balance the deficiency of electrons in the cloud.

Charged Bodies

A *charged* body is one that has more or less than the normal number of electrons. It may be either positively or negatively charged. A positively charged body is one in which some of the electrons have been removed from the atoms, and there is a deficiency of electrons, or fewer electrons than protons. A negatively charged body is one in which there are more than the normal number of electrons in each atom—that is, there are more electrons than protons. A body in which there is an equal number of electrons and protons in each atom is an *uncharged* body.

Removing electrons from a body involves physically attaching them to another body, and then moving the other body some distance away. The second body will have an excess of electrons, and thus will be negatively charged. The first body will have a deficiency in electrons and thus will be positively charged. This can be illustrated by rubbing glass with silk. Some of the electrons are rubbed off the glass onto the silk. This leaves the glass with a positive charge (lack of electrons) and the silk with a negative charge (surplus of electrons). So long as the two bodies, the silk and the glass, are not brought into contact, they will retain the charges. However, when they are allowed to touch, the surplus of electrons on the silk will move onto the glass and neutralize the charge on the two bodies.

Force Between Charged Bodies

Experimentally, it has been proved that charged bodies act upon each other with a force of attraction when they are oppositely charged, and act upon each other with a force of repulsion when similarly charged. This is because electrons and protons *attract* each other; electrons *repel* other electrons, and protons repel other protons. The forces of attraction or repulsion change with the magnitude of the charges and also with the distances between them. This is dealt with in Coulomb's Law, which states that charged bodies attract or repel each other with a force that is directly proportional to the product of the charges and inversely proportional to the square of the distance between them. Mathematically, Coulomb's Law can be expressed as follows:

$$F = \frac{Q \times Q'}{D^2}$$

where Q and Q' represent the charges and D the distance separating them.

The charge on one electron or proton might be used as the unit of electrical charge, but it would not be practical because of its small magnitude. Generally, the practical unit of charge used is the coulomb, which is equal to the charge of 6.28×10^{18} electrons.

Field of Force

The region of space around and between charged bodies where their influence is felt is called their dielectric field of force. The dielectric field requires no physical or mechanical connecting link but can be applied through air or through a vacuum. Electrostatic field and electric field are other names given to this region of force.

Fields of force permeate the space surrounding certain objects and, in general, diminish in proportion to the square of the distance from their source or origin. The force of gravity is a field of force that permeates the space surrounding the earth, and acts through free space, causing all unsupported objects in its region to fall to the earth. Newton discovered the law of gravitation, which states that every object attracts every other object with a force that is directly proportional to the product of the masses and inversely proportional to the square of the distance between them. (Note the similarity between the law of gravitation and the law of attraction of charged bodies.) The gravitational fields hold the universe together, for with no gravitational field, the planets, including the earth, would fly off at a tangent and travel through space instead of revolving around the sun. The moon would cease to revolve about the earth, and, due to the earth's rotation, objects on the earth's surface would fly out into space like mud from a bicycle wheel. Electrons likewise revolve at a tremendous velocity around the positive protons of the atom, but they, like the planets revolving around the sun, do not fly off at a tangent. Therefore, there too must be a field of force between them and the protons.

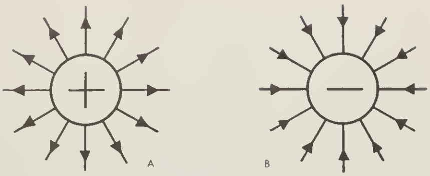
Relatively speaking, there are great distances between the electrons and the protons, even in apparently solid matter. If a copper one-cent piece were enlarged to the size of the earth's orbit around the sun (approximately 186,000,000 miles in diameter), the electrons in it would be the size of baseballs and they would be, on the average, three miles apart.

Since the field of force between the electrons and protons in the atom is the same as the dielectric field associated with charged bodies, any discussion of dielectric or electric fields is understood to mean the external fields about the charged bodies, unless specific reference is made to the field within the atom.

Lines of Force

In diagrams, imaginary lines are used to represent the direction and intensity of the field of force. The intensity (field strength) is indicated by the density (number of lines per unit area), and the direction of the field is indicated by arrowheads on lines drawn in the direction a small test charge moves or tends to move when acted upon by the field of force.

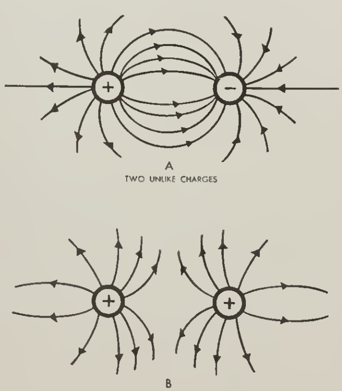
Either a small positive or negative test charge can be used to test the direction or force, because the force of a dielectric field will act on either. Arbitrarily, however, it has been agreed to use



Direction of Fields about Isolated Charges

a small positive charge when determining the direction of a dielectric field. The test shows that the direction of the field about an isolated positive charge is *away* from the charge, for a positive test charge is *repelled*; and that the direction about an isolated negative charge is *toward* the charge, for a positive test charge is *attracted*. From this, you can see that the direction of the field between the positive and the negative charge is from positive to negative.

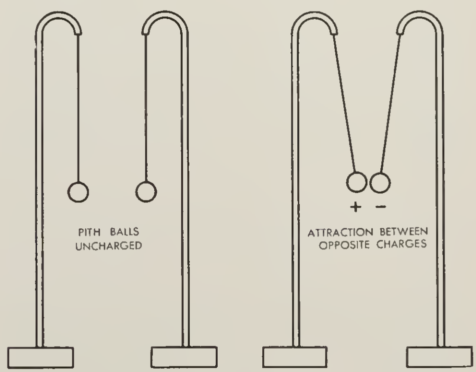
Below in the illustration showing the dielectric fields about like and unlike charges, notice that the lines of force apparently repel each other. At A, although the two charges are attracted, the lines of force between them are not parallel. They bulge out at the center as if they were repelling each other, and are in the same direction; that is, from left to right on the page. At B, the lines of force located in the region between the charges apparently repel each other, as you can see by their bent appearance. They also are in the same direction. Thus, instead of saying that *like charges repel*, the law can be modified to say that *dielectric lines of force in the same direction repel each other*. This rule, you will find, is very convenient and useful in dealing with certain electrical phenomena.



Dielectric Fields about Like and Unlike Charges

Experimenting with Charged Bodies

When a rubber rod is briskly rubbed with a piece of woolen cloth, a number of electrons from the cloth adhere to the rubber rod. If the two objects are separated immediately, there will be an excess of electrons on the rubber rod. In other words, it is negatively charged. If two pith balls such as are shown below are oppositely charged by touching one of them with the rubber rod and the other with the cloth, they will have an attraction for each other. This shows that a force is present and indicates that a dielectric field has been established. In establishing this field, it was necessary to do work against the force of attraction in separating the charged bodies. If the bodies were allowed to come together as a result of the force of attractions between them, this energy would be regained. Hence, the conclusion is that energy can be stored in a dielectric field.



Experimenting with Dielectric Fields

If you move the negatively charged rubber rod an appreciable distance away from the cloth, a dielectric field will still exist in the space around it. You can prove that the field still exists by picking up bits of paper with the rod or by charging both of the pith balls with it. In this condition the pith balls show that a force of repulsion exists between them, a condition that indicates that the dielectric field still exists about the rod.

If an external force is used to bring the two charged pith balls closer, work is done and the force of repulsion is increased. The energy consumed in increasing the field is recovered whenever the external force is removed. It will be used up in returning the pith balls to their original positions. Here again, energy is necessary to establish or increase a field; force is necessary to maintain it, and recoverable energy is stored in the field.

If one negatively charged pith ball is isolated and the negatively charged rubber rod is brought toward it from any direction, a force of repulsion is present. If the pith ball is positively charged, there will be a force of attraction no matter from which direction you bring the negatively charged rod toward the pith ball. The conclusion, then, is that a dielectric field entirely surrounds a charged body.

Distribution of Charges

Michael Faraday, the famous English scientist, and others proved that charges on various shaped objects distribute themselves according to a fixed pattern. For example, if a hollow sphere is charged either positively or negatively, there is no electric field on the inside, but there are uniformly distributed charges on the outside. On pointed objects, charges tend to accumulate on the pointed parts. Thus, on a charged tear-drop shaped object, the intensity of the electric field is greater in the region of the sharp point. Use of this fact is made in the design of spark gaps. In spark plugs, the shape of the electrodes determines the voltage at which a spark jumps across the spark gap. The sharper the points, the lower the voltage at which the spark will jump for a given separation of the electrodes.

Electroscope

While the detection of a charge can often be made by the ability of a charged body to attract other light bodies, this method is not sensitive enough to be of practical use. A simple and very sensitive instrument for detecting the presence of a charge of electricity is the *electroscope.* It consists of two very thin strips of metal called leaves, suspended from one end of a metal rod whose other end terminates in a ball. The leaves are usually enclosed in glass to prevent their being affected by air current. The electroscope serves to detect the kind of charge (positive or negative) and, to some extent, the amount of charge on a body.

The thinness and type of material used for the leaves chiefly determine the sensitivity of the instrument.

ELECTRICAL QUANTITIES

Charge

The unit of electrical charge or the unit quantity of excess electrons is the coulomb. The coulomb is the charge on a single charged body containing 6.28 x 10^{18} free electrons. For practical purposes this unit is quite large, but it is important in that it serves to define other units.

Current

Current is the movement of electrons through a conductor. The total amount of charge transferred by moving electrons is measured in coulombs or fractions of coulombs. The rate at which they flow is measured in amperes. An ampere is equal to one coulomb passing a given point per second. An ampere, for example, is approximately equal to the current flowing in a 100-watt lamp.

Resistance

The opposition which a conductor (or insulator) offers to the flow of electrons is called *resistance.* For any given conductor, the resistance depends upon the cross-sectional area, the length, and the relative resistance of the material.

The following are the relative resistances of several different materials for the same length and cross section, with silver as the standard.

Silver	*1.00*	*Aluminum*	*1.8*
Copper	*1.08*	*Lead*	*13.5*
Gold	*1.4*	*Platinum*	*7.0*

The numbers in this table are only approximate, for the resistance of each metal depends, not only on its purity, but also on its condition; that is, whether the metal is hard drawn or annealed. Resistance is also affected by temperature. Generally, the higher the temperature the greater the resistance of the wire. Each kind of metal has its own temperature coefficient. Temperature coefficient is the change in resistance per degree change in temperature. For very accurate work, the temperature at which a resistor has a certain value must be specified. To calculate the resistance of wire at any temperature use the formula,

$$R_t = R_0\ (1 + {}_aT)$$

where R_t is the resistance in ohms, R_o the resistance at 0° C, and a the temperature coefficient.

The following are the temperature coefficients for some of the common conductors:

Aluminum	*0.0042*	*Manganese*	*0.00001*
Copper (hard-drawn)	*0.00408*	*Silver*	*0.0040*
Tungsten	*0.0051*	*Platinum*	*0.0051*

Ordinarily, gases are considered to have a high resistance, but under certain conditions as when a gas is ionized, they may be fairly good conductors. A gas is ionized when a relatively small number of its electrons are changed into ions. An ion is an electrified particle formed when a molecule loses or gains an electron. Gases may be ionized while under the influence of an electrostatic field or by bombardment of energy rays, such as light or X-ray.

The unit of resistance to electron flow is the ohm. The ohm is specified by international agreement as the resistance which, at a temperature of 0° C, a column of mercury of uniform cross section with a length of 106.3 cm, and a mass of 14.45 grams offers to the flow of electrons. The cross section of the column is approximately a square millimeter. A copper wire 1000 feet long and 0.1 inch in diameter has a resistance of one ohm. Likewise, an iron rod one kilometer long and one centimeter square has a resistance of one ohm.

Electromotive Force

The external force or electrical pressure which tends to produce a flow of electrons is known as electromotive force (abbreviated as emf), potential difference, or voltage. All these terms practically mean the same thing and are used interchangeably.

The unit of electromotive force is the volt. The volt is the amount of electrical pressure required to maintain a current of one ampere through a resistance of one ohm. A dry cell produces 1.5 volts. Each cell of a storage battery supplies 2.0 volts.

Another term relative to electromotive force is *voltage drop*. Voltage drop implies that the voltage (emf) has been reduced by elements that oppose current flow and that there is a difference between the original voltage and the new voltage. This difference is called the *drop* or the voltage drop.

Ohm's Law

A very useful relationship exists between the volt, the ampere, and the ohm. This relationship is stated in various ways and each way is sometimes used to define one of the three units. The relation between the volt, the ampere, and the ohm, is known as Ohm's law, in honor of the discoverer. It states that in any circuit or part of a circuit, the current in amperes is equal to the electromotive force in volts divided by the resistance in ohms. Mathematically, this is stated, $I = E/R$, where I represents the current in amperes, E represents the voltages and R the resistance in ohms. Other ways of stating this formula are $E = IR$ and $R = E/I$.

Power

Another factor to be considered in electricity is power. Power is the rate of doing work. The watt is the unit of electrical power and is equal to work done at the rate of one joule per second. (A joule is the practical metric unit of work.) One joule equals 10^{17} ergs (the basic unit of work). Power in an electrical circuit is equal to the product of voltage and current and is expressed by the formula,

$$P = EI$$

where P equals the power in watts, E the electromotive force in volts, and I the current in amperes.

Since, according to Ohm's law $E = IR$, this equation can also be written $P = I^2R$. Likewise, since $I = E/R$, the formula is also written $P = E^2/R$. You will use all three forms frequently.

PRODUCING ELECTROMOTIVE FORCE BY CHEMICAL ACTION

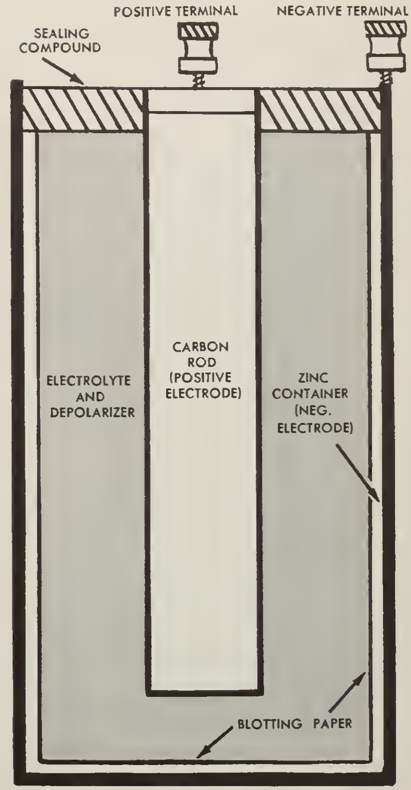
There are two common methods of producing an electromotive force or voltage. One method is mechanical and involves rotating electrical conductors in a magnetic field in such a manner that an electrical voltage is generated. Such an arrangement is called a generator. (Generators are discussed later in this chapter.) The other method is chemical and involves transforming chemical energy into electrical energy by utilizing the chemical-electrical phenomena displayed by dissimilar substances when immersed in a chemical solution.

Primary Cell

If two dissimilar metals (or a metal and carbon) are immersed in a solution that produces greater chemical action on one metal than on the other, a difference of potential will exist between the two, and if a conductor is connected between them, a current will flow. This arrangement is called a *primary cell;* the two metals are known as *electrodes*, and the solution is called the *electrolyte*. The difference of potential results from the fact that material from one or both of the electrodes goes into solution in the electrolyte, and in the process, ions form in the vicinity of the electrodes. Due to the electric field associated with the charged ions, the electrodes acquire charges.

The amount of difference in potential between the electrodes depends principally on the metals used. The type of electrolyte and the size of the cell have little or no effect on the potential difference produced.

There are two types of primary cells, the wet cell and the dry cell. In a wet cell, the electrolyte is a liquid. A cell with a liquid electrolyte must remain in an upright position and is not readily transportable. The dry cell, much more commonly used than the wet cell, is not actually dry, but contains an electrolyte mixed with other materials to form a paste. The top of the cell is sealed by a substance such as sealing wax, or by a metal cap. This arrangement prevents the paste from drying out and electrolyte from spilling, and thus makes it possible to use the cell in any position. The container is zinc and acts as the negative electrode. Next to the zinc container is a layer of blotting paper which is saturated with the electrolyte. The positive electrode is a carbon rod. The space between the positive electrode and the blotting paper is filled with a mixture of carbon, manganese dioxide, and the electrolyte itself. The voltage of a dry cell is 1.5 volts.



Construction of a Dry Cell

Internal Resistance

In any conductor there is a certain amount of resistance. Similarly there is some resistance in a cell. This resistance is largely due to the resistance in the electrodes and in the electrolyte. It depends upon the kind of materials used, and the size and spacing of the electrodes. Large cells with close spacing of electrodes have less internal resistance than smaller cells made of the same materials.

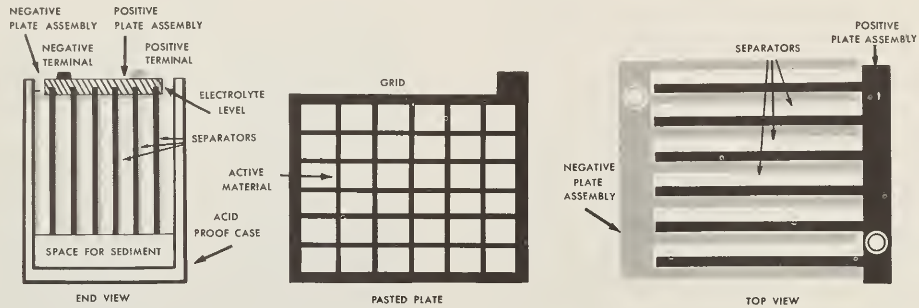
Battery

A collection of cells connected together is called a *battery*. When connected in series, the individual cell voltages add to produce a greater voltage than a single cell. When connected in parallel, the maximum usable current is increased without changing the voltage. Portable radio sets use batteries containing small cells connected in series to furnish the high voltages required for the operation of the tubes. A battery designed for this purpose is called a *B* battery.

Primary Cell Faults

If the primary cell is to give satisfactory service, you must guard against two ill effects of chemical action. These are local action and polarization. Although primarily problems of the manufacturer, you should understand the effects of these undesirable actions in a cell.

LOCAL ACTION. Local action is identical to the chemical action occurring in the cell as a whole. Instead of this, however, it takes place only between the zinc electrode and impurities in the zinc. Chemical action between these elements forms small scale cells, in which a voltage causing a minute current flow is set up. This causes the zinc to be consumed which, in turn, causes an increase in the internal resistance of the cell itself and lowers the voltage output. Local action can take place even when the cell is not in use. For this reason, some cells have a date stamped on them, indicating when they must be put into service. The remedy for local action is to clean the zinc electrode thoroughly and coat its surface with mercury. The alloy of mercury and zinc thus formed, called amalgam, covers the impurities, and thereby retards local



Secondary Cell Construction

action until the time when sufficient material is consumed in the cell, leaving the impurities exposed.

POLARIZATION. Polarization takes place only while the cell is in use. It manifests itself in the formation of hydrogen bubbles on the positive electrode. These bubbles are nonconducting, and tend to increase the internal resistance of the cell, reducing the current flow. Current may cease to flow entirely if the entire electrode becomes coated. If the bubbles are removed, or prevented from forming, polarization is reduced and current can flow until the negative electrode does not contain sufficient material to furnish the energy. A chemical which prevents the formation of hydrogen bubbles is called a *depolarizer*. An example of a depolarizer is the chemical, manganese dioxide.

Secondary Cell

A primary cell delivers electrical energy directly from the reaction of chemicals mixed together. However, a *secondary* or *storage* cell must have energy stored in it by a process called *charging* before it will deliver energy. Charging the cell consists in forcing a current to flow through it in a direction opposite that in which it flows when the cell is delivering energy (discharging) to the circuit. During charge, electrical energy is converted to chemical energy and stored in the cell. During discharge the chemical energy is reconverted to electrical energy.

A simple secondary cell contains two lead electrodes immersed in a dilute solution of sulphuric acid. If a current is forced through the cell, the surface of the electrode which is connected to the positive terminal of the charging voltage will be changed to lead peroxide and the surface of the other electrode will be changed to spongy lead. After a period of charge the cell will have a voltage of slightly over two volts. It will then furnish current to an external circuit for a period of time depending upon the amount of current drawn. This discharge process is the result of chemical action which forms lead sulphate on both electrodes.

When the lead peroxide and spongy lead are converted to lead sulphate, the cell is discharged, and will not furnish more current until charged again.

Storage Battery

The automobile storage battery is a typical battery of secondary cells, consisting of three such cells connected electrically and assembled in a single container. While the principle of operation is the same as that of the simple secondary cell, the construction is somewhat more complicated. For the cells to furnish large amounts of current, the surface area of the electrodes must be great and very closely spaced. The electrodes are rectangular, flat plates consisting of latticeworks, or grids, of lead alloy. The holes in the grids are filled with a paste of lead oxide. The electrolyte is a weak solution of sulphuric acid. To increase the area of each electrode, several grids or plates are connected together. The cell is made up of alternate positive plates and negative plates which are kept apart by spacers. The separators may be of any one of several insulating materials, but quite often are of treated wood. The separators must be porous so that the electrolyte passes through them freely and the flow of electrons is not impeded. Such a cell always has an odd number of plates, the two outside plates being negative. This is because the positive plate expands when the cell is charged and must have negative plates on each side otherwise the side next to the negative plate would expand and the off side would not. As a result, the plate would buckle. The expansion and contraction of plates cause some of the lead oxide to become loose and drop off from time to time. There is provision made for the sediment to drop to the bottom of the cell. This arrangement prevents the sediment from settling in a layer thick enough to touch the bottom ends of the plates. If it should touch the plates it would cause a short circuit between the positive and negative plate and discharge the cell.

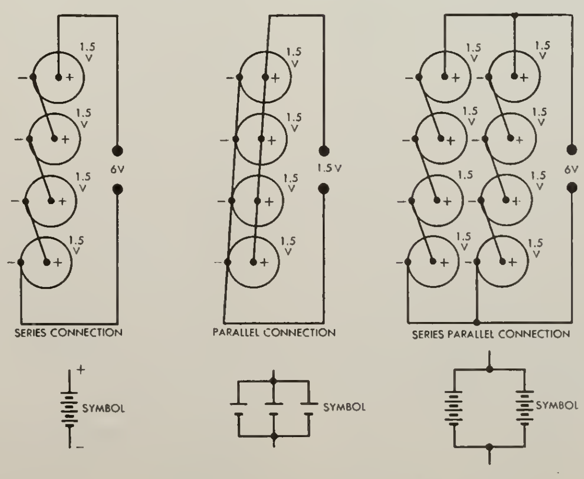
The state of charge or discharge of a battery can be determined by a hydrometer, an instrument for measuring the specific gravity of a liquid. During the charging process, water in the electrolyte is used up and the sulphuric acid solution becomes more concentrated, that is, its specific gravity is increased. A fully charged cell has a specific gravity of 1.285 to 1.300 while a discharged cell has a specific gravity of 1.150 to 1.175. Intermediate values indicate the relative proportion of full charge in the battery.

Rating of Storage Cells

A storage cell is rated in ampere-hour capacity. Ampere-hour capacity indicates the number of amperes that the cell is able to furnish continuously for a stated period of time. For example, a battery rated at 100 amp-hours at an 8-hour discharge rate can deliver 12.5 amperes continuously for 8 hours. Generally, the normal discharge rate is based upon an 8-hour discharge period. If a battery is discharged at a rate higher than normal, its capacity is decreased. Thus, a battery rated at 100 amperes hours would not furnish 50 amperes continuously for two hours. Conversely, if discharged at a lower than normal rate, the capacity would be increased. The ampere hour rating of a battery depends principally upon the size and number of plates. Batteries commonly are described as 11 plate, 13 plate, or 15 plate batteries. The number of plates gives a rough indication of the correct capacity of the battery.

Series and Parallel Connection of Cells

Many electrical devices require higher voltage or higher current than a single cell is able to furnish. Therefore, it is often necessary to connect several cells together in the form of a storage battery. The manner in which cells, primary or secondary, are connected depends upon whether you want to increase the voltage or the current. If *greater voltage* is needed, connect the cells *in series*. To connect cells in series, connect the negative terminal of each cell to the positive



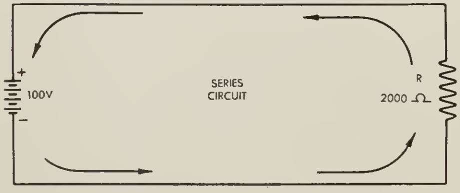
Methods of Connecting Primary Cells

terminal of the succeeding cell. The voltage of the battery is then equal to the total of the voltages of the separate cells. The same current, however, flows through the cells in succession. Therefore, the current that the battery can supply is equal to the current of a single cell. Thus, a battery composed of cells in series provides a higher voltage, but not a greater current capacity. To obtain a greater *current* than one cell is able to supply, connect the cells *in parallel.* In this case, the total current is the sum of the individual currents since the current of one cell does not flow through the other cells. To connect cells in parallel, connect all positive terminals together and all negative terminals together. The positive terminals of the cells will form the positive terminal of the battery, and the negative terminals of the cells, the negative terminal. Each cell must have the same voltage, otherwise a cell with higher voltage will force current through the lower voltage cell and carry the greater part of the load. The output voltage of the battery made up of cells connected in parallel is therefore the same as that of a single cell.

Another method of arranging cells is to connect them series-parallel. In this method the cells are connected partially in series and partially in parallel. This connection provides both a greater voltage and greater current output.

CURRENT, VOLTAGE AND POWER RELATIONS IN SERIES-CIRCUITS

The simplest form of electrical circuits is the series circuit. In the series circuit, current flow is the same at all points. To find the current flow in a series circuit, use the Ohm's law formula, $I = E/R$.



In the series circuit just above, the applied voltage is 100 volts and the resistance of the resistor R is 2000 ohms. To find the current flow in this circuit substitute the given quantities in the formula, $I = E/R$ and solve as follows:

$$I = \frac{100}{2000}$$

The current, $I = .05$ *ampere*

When this current flows through R, power is expended. To find the power expended, substitute in the power formula, $P = EI$ and solve as follows:

$$P = 100 \times 0.05$$

The power expended, $P \doteq 5$ *watts*

Resistors

A resistor is a circuit element designed to offer a certain amount of opposition to current flow. Usually, a resistor is cylindrical in shape and is constructed in one of two forms. Some resistors are sections of carbon rod or some combination of carbon and other substances molded into cylindrical form. Others are wire wound. They make possible more precise manufacture of the desired resistance and are capable of carrying heavier current than carbon resistors without damage.

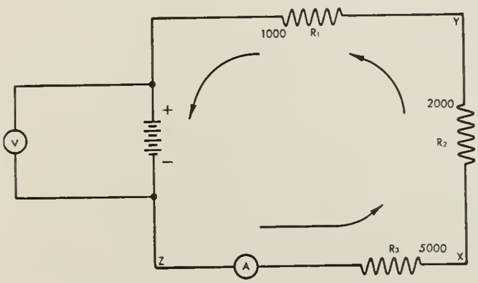
Another factor associated with resistors is power rating. Power rating indicates the maximum current that can flow through the resistor without danger of burning it out because of overheating. In the illustration of the series circuit, resistor R requires a power or wattage rating of at least 5 watts, the amount of power expended in the circuit.

Colored bands in combination, or some other arrangement of colors, are used to indicate the resistance value of resistors. (See a resistor color code for explanation of procedure for determining value of resistors.) Each resistor is marked with a certain value along with a percent of tolerance since it is difficult to manufacture a resistor to exact value. Carbon type resistors usually have a tolerance of 5%, 10%, or 20%. Wire-wound resistors have much lower tolerance. The tolerance in this type is 1% or 2%.

A 2000-ohm resistor with a 10% tolerance, for example, has an actual value of 2000 ohms ± 10%, or a range of 1800 to 2200 ohms. Although this may seem like a wide range of values, the majority of radio circuits will function satisfactorily if the circuit components are within 10% or even 20% of the desired value. Where precision is required, resistors of lower tolerance are used. Circuit problems in this manual assume the value of resistance to be exact, and do not require your considering tolerance unless specifically indicated.

Variable Resistors

Some circuits require a resistor which can be varied between limits. Such a resistor is called a *variable resistor.* Schematically, it is represented by the symbols or . Variable resistors are either tubular or circular in form and are constructed with one permanently connected contact and one movable contact. In the circular form the movable contact is connected to a shaft which, when rotated, moves the contact along the arc of the resistor. Tubular form resistors are wire wound and are usually designed to carry an appreciable amount of current. The circular type may or may not be wire wound. Variable resistors designed to carry large amounts of current are known as *rheostats*.



Series Circuit

The resistors in the circuit just above are connected in series. The current in the circuit flows through each resistor in succession, as well as through the ammeter represented by Ⓐ, and is equal in value at all points. (The construction and operation of an ammeter, the instrument for measuring current flow, is described in detail in Chapter 3.) The ammeter may be connected at any point in the circuit since current flow is the same through the circuit. The encircled V represents a voltmeter, an instrument for measuring potential difference. (See Chapter 3 for details in construction and operation of a voltmeter.) Always connect a voltmeter between the two points where you desire to read the voltage. This connection places the voltmeter in *parallel* with the part of the circuit developing or expending the voltage. Voltmeters can be used to measure the emf of a battery or the voltage drop across a resistor or other circuit elements.

In the series circuit shown above, with the values indicated, calculate the current flow and the power rating of each resistor. Since all the current flows through each resistor, the total opposition (resistance) is the sum of the individual oppositions (resistors). Mathematically, this is expressed by the equation,

$$R_T = R_1 + R_2 + R_3$$

where R_T represents the total resistance.

By substituting,

$$R_T = 1000 + 2000 + 5000$$

The total resistance $R_T = 8000$ ohms.

To find the total current, use the Ohm's law formula, $I = E/R$.

Thus, $$I = \frac{100}{8000}$$

The total current, $I = 0.0125$ *amperes.*

(Because the ampere is rather a large unit, current is often measured in milliamperes (ma), thousandths of an ampere, or in microamperes (μa), millionths of an ampere. Thus, the current, 0.0125 ampere in the problem, can be expressed as 12.5 ma or 12,500 μa. The milliampere is widely used in radar.)

To find the power ratings of each resistor, substitute in the power formula, $P = R \times I^2$ and solve as follows:

$$P_1 = R_1 \times I^2 = 1000 \times (0.0125)^2 = 0.15625 \text{ watts}$$
$$P_2 = R_2 \times I^2 = 2000 \times (0.0125)^2 = 0.3125 \text{ watts}$$
$$P_3 = R_3 \times I^2 = 5000 \times (0.0125)^2 = 0.78125 \text{ watts}$$

Manufacturers do not generally manufacture resistors with the wattages determined in the problem. In practice, you would use a resistor with the next higher rating. For R_1, it would be ¼ watt, for R_2, ½ watt, and for R_3, it would be one watt.

The arrows in the two preceding circuit illustrations indicate the direction of electron flow. On the positive terminal of the battery, there is a deficiency of electrons, while on the negative terminal there is a surplus. The emf of the battery tends to force electrons to flow through the circuit in the direction indicated from the negative terminal to the positive terminal; and, thus, neutralize the charges of both electrodes. Since continued chemical action within the cells of the battery maintains the emf, current flow is continuous.

By applying the formula $E = IR$, you can calculate the voltage drop across each resistor in the problem. Thus,

$$E_1 = R_1 \times I = 1000 \times 0.0125 = 12.5 \text{ volts}$$
$$E_2 = R_2 \times I = 2000 \times 0.0125 = 25 \text{ volts}$$
$$E_3 = R_3 \times I = 5000 \times 0.0125 = 62.5 \text{ volts}$$

Note that the sum of the voltages across the individual resistors is 100 volts, a sum equal to the applied voltage. In any series circuit, the sum of the voltages across the individual resistors is always equal to the applied voltage.

The total power P_t expended by the battery is equal to the total of the power expended by the individual resistors. Therefore,

$$P_t = E_t \times I_t$$
$$P_t = 100 \times 0.0125$$
$$P_t = 1.25 \text{ watts}$$

This is equal to the sum of 0.15625, 0.3125, and 0.78125 watts, which are the wattages of the individual resistors.

Voltage Dividers

The arrangement of resistors shown on the preceding page is called a voltage divider, a device very useful for supplying voltages of various values. If for some purpose, you want a voltage of 25 volts, you can use the voltage across R_2. If you need 30 volts instead of 25, you would have to select other values of resistance in the circuit, so that one resistor would have 3/10 of the total resistence. (The total voltage across the circuit is 100 volts)

To make the voltage across a resistor variable between fixed limits, a kind of voltage divider called a *potentiometer* must be used. A potentiometer is like a variable resistor except that there are fixed contacts at both ends. The movable contact is located between the fixed contacts. Schematically, the potentiometer is represented by the symbol

To illustrate the use of a potentiometer, assume that R_2 in the circuit is a 2000 ohm potentiometer. In this condition the voltage at X (with respect to Z) is equal to 62.5 volts. At Y the voltage is equal to 87.5 volts. By correctly setting the variable tap, it is possible to obtain any voltage between these two limits.

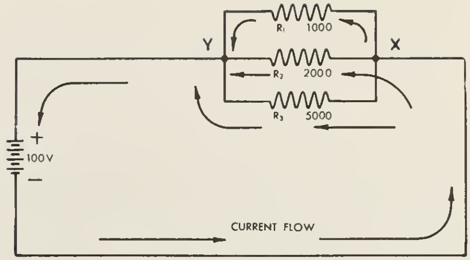
Summary

In summary, the current, voltage and power relations in a series circuit are the following:

1. $R_t = R_1 + R_2 + R_3 + \ldots\ldots\ldots$
2. $I_t = I_1 = I_2 = I_3 = \ldots\ldots\ldots$
3. $E_t = E_1 + E_2 + E_3 + \ldots\ldots\ldots$
4. $P_t = P_1 + P_2 + P_3 + \ldots\ldots\ldots$

Current, Voltage and Power in a Parallel Circuit

In the parallel circuit shown on this page, the resistors R_1, R_2, and R_3 have their terminals joined together at points X and Y. Current divides at point X with part of it going through R_1, part through R_2, and the remainder through R_3. All of the individual currents re-combine at point Y and flow to the positive terminal of the battery. The voltage across each of the three resistors is equal to the potential difference be-



Parallel Circuit

tween points X and Y and is equal to the applied voltage of the battery. Therefore, the voltage across each resistor is 100v ($E_1 = E_2 = E_3 = 100$v).

To find the total current supplied by the battery to the combination of parallel resistors, first find the current through each individual resistor, and then find the sum of the individual currents. You can readily find the individual currents by Ohm's law ($I = E/R$).

1. Current through R_1

$$I_1 = \frac{E_1}{R_1}$$
$$I_1 = \frac{100}{1000}$$
$$I_1 = 0.1 \text{ amperes}$$

2. Current through R_2

$$I_2 = \frac{E_2}{R_2}$$
$$I_2 = \frac{100}{2000}$$
$$I_2 = 0.05 \text{ amperes}$$

3. Current through R_3

$$I_3 = \frac{E_3}{R_3}$$
$$I_3 = \frac{100}{5000}$$
$$I_3 = 0.02 \text{ amperes}$$

4. The total current

$$I_t = I_1 + I_2 + I_3$$
$$I_t = 0.1 + 0.05 + 0.02$$
$$I_t = 0.17 \text{ amperes}$$

The total power expended by the battery is equal to the sum of the power expended in the individual resistors. This can be found by the power formula, $P = EI$.

1. Power in R_1

$$P_1 = E_1 \times I_1$$
$$P_1 = 100 \times 0.1$$
$$P_1 = 10 \text{ watts}$$

2. Power in R_2

$$P_2 = E_2 \times I_2$$
$$P_2 = 100 \times 0.05$$
$$P_2 = 5 \text{ watts}$$

3. Power in R_3

$$P_3 = E_3 \times I_3$$
$$P_3 = 100 \times 0.02$$
$$P_3 = 2 \text{ watts}$$

4. The total power

$$P_t = P_1 + P_2 + P_3$$
$$P_t = 10 + 5 + 2$$
$$P_t = 17 \text{ watts}$$

Since you already know the total current, you can find the total power by substituting the total current I_t in the power formula. Therefore,

$$P_t = E_t \times I_t$$
$$P_t = 100 \times 0.17$$
$$P_t = 17 \text{ watts}$$

Resistance in Parallel Circuits

When the total current is known, the total resistance in a parallel circuit may be found by the Ohm's law formula, $R = E/I$.

Thus, the total resistance in the circuit shown on the preceding page is,

$$R_t = \frac{E_t}{I_t}$$
$$R_t = \frac{100}{0.17}$$
$$R_t = 588 \text{ ohms}$$

Note that the total resistance 588 is less than any one of the individual resistors R_1, R_2, and R_3. The resistance of two or more branches in parallel is always less than that of any of the component branches.

A certain relationship exists between the total resistances R_1, R_2, and R_3, and the total current I_t in the parallel circuit. This relationship is $R_t = \dfrac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$

This can be proved as follows:

1. $I_t = I_1 + I_2 + I_3$

2. By substituting the following values of I,

$$I_t = \frac{E_t}{R_t},\ I_1 = \frac{E_1}{R_1},\ I_2 = \frac{E_2}{R_2}$$

and $I_3 = \dfrac{E_3}{R_3}$ in equation (1), it becomes,

$$\frac{E_t}{R_t} = \frac{E_1}{R_1} + \frac{E_2}{R_2} + \frac{E_3}{R_3}$$

3. Since $E_t = E_1 = E_2 = E_3$,

$$\frac{E_t}{R_t} = \frac{E_t}{R_1} = \frac{E_t}{R_2} = \frac{E_t}{R_3}$$

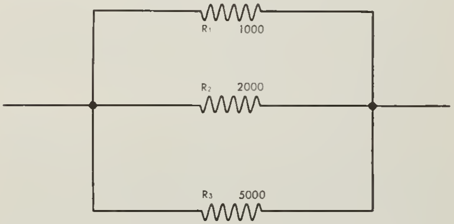
4. Dividing by E_t, the equation becomes,

$$\frac{I}{R_t} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

5. And then dividing both sides into 1 gives,

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

This expression states that the combined resistance of a parallel circuit is equal to the reciprocal of the sum of the reciprocals of the individual resistances.



Reducing Parallel Resistances to Equivalent Resistance

Thus, in the circuit directly above which shows three resistors, R_1, R_2, and R_3, with values of 1000 ohms, 2000 ohms, and 5000 ohms, respectively, you can solve for the total resistance by the formula

$$R_t = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$
$$R_t = \frac{1}{\frac{1}{1000} + \frac{1}{2000} + \frac{1}{5000}}$$
$$R_t = \frac{1}{\frac{10+5+2}{10{,}000}} \text{ or } \frac{10{,}000}{17}$$
$$R_t = 588 \text{ ohms}$$

When only two parallel resistors are involved, the equivalent resistance is given by the equation,

$$R_t = \frac{R_1 R_2}{R_1 + R_2}$$

Stated as a rule, this reads, "The total resistance of two parallel branches is the product of the two resistors divided by their sum." When there are three or more branches, you can use this formula by combining resistors into an equivalent circuit containing only two resistors.

Example

Find R_t when $R_1 = 1000$ ohms, $R_2 = 2000$ ohms, and $R_3 = 5000$ ohms

Solution:

1. Combine R_1 and R_2

$$R_e \text{ (equivalent)} = \frac{1000 \times 2000}{1000 + 2000}$$

$$R_e = 667 \text{ ohms}$$

2. Then find the total resistance of the two resistors R_e and R_3

$$R_t = \frac{667 \times 5000}{667 + 5000} = \frac{3335000}{5667} = 588 \text{ ohms}$$

When two equal resistors are connected in parallel, the combined resistance of the two resistors, equals one-half of either resistor. Similarly, the total resistance of three equal resistors in parallel is one-third the resistance of one resistor.

Conductance

The reciprocal of resistance is conductance. Conductance is the property of a conductor that makes it possible for current to flow. It is represented by the letter G. The unit of conductance is the *mho* (ohm spelled backward). For example, the conductance of a conductor with a resistance of 1000 ohms equals .001 mho (1000 micromhos) since $G = \frac{1}{R} = \frac{1}{1000}$.

In a parallel circuit with several branches, the total conductance is the sum of the conductances of the several branches. In formula, it can be written,

$$G_t = G_1 + G_2 + G_3 + \ldots.$$

Division of Current in Parallel Branches

The total current in a parallel circuit divides between the branches. In a circuit with two branches R_1 and R_2, the current flowing in R_1 is expressed,

$$I_1 = I_t \times \frac{R_2}{R_1 + R_2}.$$

The current flowing in R_2 is expressed,

$$I_2 = I_t \times \frac{R_1}{R_1 + R_2}.$$

Summary

The following are the general relationships of current, voltage, and power in a parallel circuit:

1. In a parallel circuit the same voltage appears across each element ($E_t = E_1 = E_2 = E_3 = \ldots$).

2. The total current in a parallel circuit is equal to the sum of the currents flowing in the individual branches ($L_t = I_1 + I_2 + I_3 + \ldots$).

3. The total resistance is equal to the reciprocal of the sum of the reciprocals of the individual resistances,

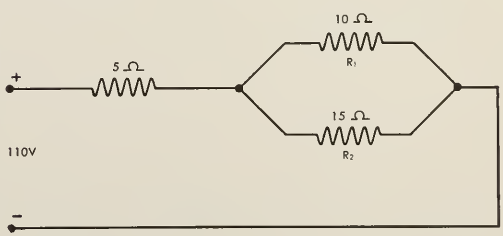
$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots}$$

4. The total conductance of the resistances is equal to the sum of the conductances of the individual resistors, ($G_t = G_1 + G_2 + G_3 + \ldots$).

5. The total power is the sum of the power expended in each individual resistor, ($P_t = P_1 + P_2 + P_3 + \ldots$).

SERIES-PARALLEL CIRCUITS

Frequently, you will find circuits having combinations of resistors in series and in parallel. To solve problems of this type, apply the principles you learned in circuits having only resistors in series and those having only resistors in parallel.



Example 1.

In the circuit directly above, find the total current, the current through the 10-ohm resistor and the voltage drop across the 5-ohm resistor.

Solution:

1. *First find the equivalent resistance (R_e) of the 10-ohm and the 15-ohm resistors in parallel.*

$$R_e = \frac{10 \times 15}{10 + 15}$$

$$R_e = \frac{150}{25}$$

$$R_e = 6 \text{ ohms}$$

2. *R_e is now in series with the 5-ohm resistor. Therefore, $R_t = 6 + 5 = 11$ ohms.*

3. *The total current*

$$I_t = \frac{E_t}{R}$$

$$I_t = \frac{110}{11} = 10 \text{ amperes}$$

4. The portion of the total current flowing through the 10-ohm resistor can be found by the formula,

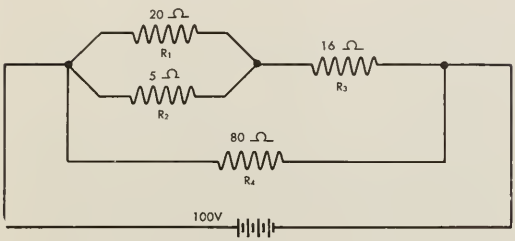
$$I = I_t \times \frac{R_2}{R_1 + R_2}$$

$$\text{Therefore, } I = 10 \times \frac{15}{10 + 15}$$

$$I = 10 \times \frac{15}{25} \text{ or 6 amperes.}$$

5. The voltage drop across the 5-ohm resistor can be found by the formula, $V = I_t \times R$.

Therefore, $V = 10 \times 5 = 50$ *volts.*



Example 2.

Given the circuit shown above with the indicated values. Find the total current and the current I_2 through R_2.

Solution:

1. Resolving R_1 *and* R_2 *into an equivalent resistance,*

$$R_e = \frac{20 \times 5}{20 + 5} = \frac{100}{25}$$

$$R_e = 4 \text{ ohms}$$

2. Combining the 4 ohms with the 16 ohms in series, the total resistances of R_1, R_2 *and* $R_3 = 20$ *ohms.*

3. Since R_4 *is in parallel with the combination having a resistance of 20 ohms,*

$$R_t = \frac{20 \times 80}{20 + 80}$$

$$R_t = \frac{1600}{100} = 16 \text{ ohms.}$$

4. The total current

$$I_t = \frac{E_t}{R_t}$$

$$I_t = \frac{100}{16}$$

$$I_t = 6.25 \text{ amperes.}$$

5. Current I_3 *through* R_3 *equals* E_t *divided by 20 ohms, the total resistance of* R_1, R_2 *and* R_3 *in combination. Therefore* $I_3 = \frac{100}{20}$ *or 5 amperes*

The 5 ampere current divides between R_1 *and* R_2*, thus,*

$$I_2 = 5 \times \frac{R_1}{R_1 + R_2}$$

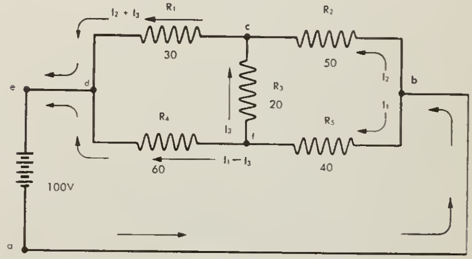
$$I_2 = 5 \times \frac{20}{25} = 4 \text{ amperes.}$$

KIRCHOFF'S LAWS

More involved combinations of resistances and voltages are called networks. The solution of network problems is best handled by using Kirchoff's laws which, in turn, are based on Ohm's law. There are two Kirchoff laws. They are the following:

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 loop of the circuit is equal to the sum of the voltage drops across the resistances in that loop.



Examples

Problem 1. In the circuit illustrated above, find the current through each resistor, the voltage drop across each resistor and resistance of the network by Kirchoff's laws.

Solution:

Let I_1 *be the current through* R_5, I_2 *be the current through* R_2*, and* I_3 *be the current through* R_3. *(Assume that electron flow is from f to c. This assumption, if incorrect, will not affect the solution except that* I_3 *will be found to be negative if this is a wrong assumption in the direction of current flow.)*

Then, by Kirchoff's first law, which states that the current leaving a point must equal the current flowing into that point, current I_1 *flowing to f through* R_5 *and* I_3 *flowing away through* R_3, $(I_1 - I_3)$ *equals the current through* R_4. *Likewise* $I_2 + I_3$ *equals the current through* R_1. *Since there are three unknown quantities, there will be three equations, which will give you the equations for the voltages in three loops. Here they are*

In loop abfde, $40\ I_1 + 60\ (I_1 - I_3) = 100$ *v.*
or $100\ I_1 - 60\ I_3 = 100$ (1)

In loop abfcde $40\ I_1 + 20\ I_3 + 30\ (I_2 + I_3) = 100$ *v.*
or $40\ I_1 + 30\ I_2 + 50\ I_3 = 100$ (2)

In loop abcde, $50\ I_2 + 30\ (I_2 + I_3) = 100$ *v.*
or $80\ I_2 + 30\ I_3 = 100$ (3)

Multiplying (1) by 2
$200\ I_1 - 120\ I_3 = 200$ (4)

Multiplying (2) by 5

$200\ I_1 + 250\ I_3 + 150\ I_2 = 500$ (5)

Subtracting (5) from (4)

$-150\ I_2 - 370\ I_3 = -300$ (6)

Multiplying (6) by 8

$-1200\ I_2 - 2960\ I_3 = -2400$ (7)

Multiplying (3) by 15

$1200\ I_2 + 450\ I_3 = 1500$ (8)

Adding (8) to (7)

$-2510\ I_3 = -900$

Dividing by 2510

$I_3 = 0.359$ *amperes.*

Substituting 0.359 for I_3 *in (1),*

$100\ I_1 - 60\ (.359) = 100$

Removing parenthesis

$100\ I_1 - 21.5 = 100$

Adding 21.5

$100\ I_1 = 121.5$

Dividing by 100

$I_1 = 1.215$ *a.*

Substituting 0.359 for I_3 *in (3)*

$80\ I_2 + 30\ (0.359) = 100$

Removing parenthesis

$80\ I_2 + 10.77 = 100$

Subtracting 10.77

$80\ I_2 = 89.23$

Dividing by 80

$I_2 = 1.115$ *a.*

then,

$I_1 - I_3 = 1.215 - 0.359 = 0.856$ *a*

and

$I_2 + I_3 = 1.115 + 0.359 = 1.474$ *a.*

The voltage drops across R_1, R_2, R_3, R_4 *and* R_5 *respectively are* E_1, E_2, E_3, E_4 *and* E_5. *By Ohm's law,* $E = IR$, *thus,*

$E_1 = 30\ (1.474) = 44.22$ *v.*

$E_2 = 50\ (1.115) = 55.75$ *v.*

$E_3 = 20\ (0.359) = 7.18$ *v.*

$E_4 = 60\ (0.856) = 51.36$ *v.*

$E_5 = 40\ (1.215) = 48.60$ *v.*

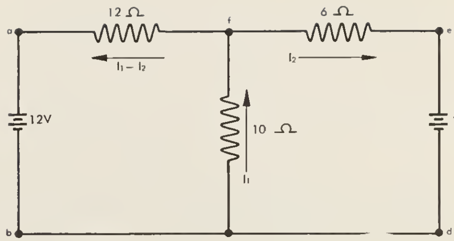
Find the total current I_t *at b or d and the total resistance*

At b, $I_t = I_1 + I_2 = 1.215 + 1.115 = 2.33$ *amps,*
and the total resistance

$R_t = E_t\ /\ I_t = 100/2.33 = 42.9$ *ohms*

NOTE

In forming the voltage equations for the loops, place plus in front of voltage drops when you are going in the direction of electron flow and minus when you are going in the other direction.



Problem 2.

In the circuit shown above, find the current and voltage drop across each resistor.

Solution:

Let I_1 be the current through the 10-ohm and I_2 the current through the 6-ohm resistor. By Kirchoff's first law, the current from point f to point a is $I_1 - I_2$. Thus, the voltage equations from the second law are the following:

Around loop dcfe

$10\ I_1 + 6\ I_2 = 10$ *v.* (1)

Around loop bcfa

$10\ I_1 + 12\ (I_1 - I_2) = 12$ *v.* (2)

or

$22\ I_1 - 12\ I_2 = 12$ *v.* (3)

Multiplying (1) by 2

$20\ I_1 + 12\ I_2 = 20$ *v.* (4)

Adding (3) to (4)

$42\ I_1 = 32$

Dividing by 42

$I_1 = 0.762$ *a.*

Substituting

0.762 for I_1 *in (1).*

$7.62 + 6\ I_2 = 10$

Subtracting 7.62

$6\ I_2 = 2.38$

Dividing by 6

$I_2 = 0.396$ *a.*

Therefore,

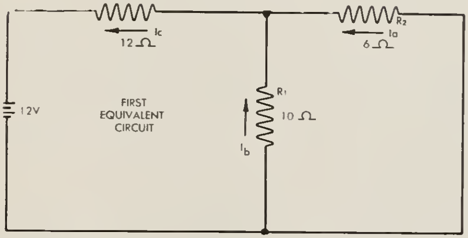
$I_1 - I_2 = 0.366$ *amperes*

The various voltage drops are the following:

10-ohm resistor $= 10 \times 0.762$ *or 7.62 volts*
6-ohm resistor $= 6 \times 0.396$ *or 2.38 volts*
12-ohm resistor $= 12 \times 0.366$ *or 4.39 volts*

Superposition Method

A second method for solving networks containing more than one source of voltage is the superposition method. This method is based on the principle that the current in any part of a circuit is the sum of the currents produced by each voltage acting separately and with the other voltages shorted out.



Example

Solve the circuit illustrated on page 2-16, with the values indicated, by superposition. First, draw the equivalent circuit as shown above with the 10-volt battery shorted out, and find the total current in the equivalent circuit.

1. To find total current I_c, calculate the equivalent resistance of the 6-ohm and 10-ohm resistors

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

$$Re = \frac{6 \times 10}{6+10} = \frac{60}{16} = 3.75 \text{ ohms}$$

Total resistance = 12 + 3.75 or 15.75 ohms

$$I_c = E/R$$

$$I_c = \frac{12}{15.75} = 0.763 \text{ ampere}$$

2. Find the current in each branch of the equivalent circuit.

$$I_a = I_t \times \frac{R_1}{R_1 + R_2}$$

$$I_a = 0.763 \times \frac{10}{10+6}$$

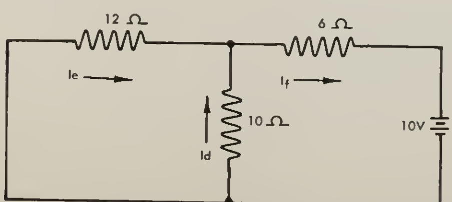
$$I_a = 0.477 \text{ ampere}$$

$$I_b = I_t \times \frac{R_2}{R_1 + R_2}$$

$$I_b = 0.763 \times \frac{6}{10+6}$$

$$I_b = 0.286 \text{ ampere}$$

Now, draw the equivalent circuit as shown below with the 12-volt battery shorted out and find the total current and the current in the branches.



SECOND EQUIVALENT CIRCUIT

1. Find the total current.

The equivalent resistance of the 12-ohm and 10-ohm resistors in parallel is $\frac{10 \times 12}{10+12}$ *or 5.46 ohms.*

The total resistance in the circuit is 6 + 5.46 or 11.46 ohms.

The total current, $I = \frac{10}{11.46}$ *or 0.873 amps.*

2. Find the current in branches

$$I_d = 0.873 \times \frac{12}{10+12}$$

$$I_d = 0.873 \times \frac{6}{11} \text{ or } 0.476 \text{ amperes}$$

$$I_e = 0.873 \times \frac{10}{10+12}$$

$$I_e = 0.873 \times \frac{5}{11} \text{ or } 0.397 \text{ amperes.}$$

Next combine the results of the two sets of computations.

1. The current through the 12-ohm resistor is $I_c - I_e$*. Therefore, this current equals 0.366 ampere (0.763 − 0.397). (These currents are subtracted since they flow in reverse directions (*I_c *to the right;* I_e *to the left). The difference is their net result.*

2. The current through the 10-ohm resistor equals $I_b + I_d$ *or 0.286 + 0.476 = 0.762 ampere.*

3. The current through the 6-ohm resistor equals $I_f - I_a = 0.873 - 0.477$ *or 0.396 ampere.*

On comparing these results with those obtained in the Kirchoff law problem, note that the results obtained here are identical.

MAGNETISM

A magnet is an object which has the property of attracting iron and steel and which, if permitted to turn freely, will rotate to a definite direction. Magnets attract other materials such as nickel and cobalt but not with as much force as iron and steel. Materials which can be attracted by magnets are called *magnetic substances*. There are two principal kinds of magnets—*natural* magnets and *artificial* magnets. Natural magnets are found in nature already magnetized while artificial magnets must be made by magnetizing iron or steel. Artificial magnets can be either permanent or temporary.

Natural Magnets

A black mineral ore called lodestone or magnetite found in plentiful supplies in Asia Minor exhibits magnetic properties. Lodestone is called a natural magnet because it exists in nature already magnetized. Historically, lodestone has played an interesting role, being used in the middle ages to magnetize compass needles. Today, however, lodestone has very little value as a magnet chiefly because of its unstable physical structure and low magnetic strength.

Artificial Magnets

Artificial magnets are made of iron and steel. They are magnetized by induction from some exterior object, by stroking with some other artificial magnet or by being placed in the field of an electromagnet. (Electromagnets are described later.) A bar of hard steel will hold magnetism for a long period of time, and for this reason is called a permanent magnet. Soft iron can be magnetized easily but loses its magnetism quite rapidly. The property of a substance which causes it to remain magnetized is called *retentivity* and the magnetism which remains is called *residual magnetism*. Steel has high retentivity while that of soft iron is low.

Magnetic Poles

When a bar magnet is dipped into iron filings, a large number of filings will cling to the magnet near its ends, but few will attach themselves to the magnet near its center. This action indicates that the magnetism is concentrated at the two ends. These ends are called the poles of the magnet. The magnetic strength of the two poles of any magnet is equal.

A magnet, free to rotate, will always turn to a north-south direction, aligning itself with the earth's magnetic field. The pole of the magnet which always turns toward the north is called the *north-seeking pole*, or simply the north pole (N); and the pole at the opposite end, the *south-seeking* pole, or south pole (S).

Force Between Poles

Experiments show that if you bring the S-pole of one bar magnet near the N-pole of another, there will be an attraction between the two poles, and if you bring two N-poles or two S-poles together there will be a force of repulsion between them. This action is summed up in the basic law of magnetism which states that *unlike* poles *attract* each other and that *like* poles *repel* each other.

The force of a reaction or repulsion between two poles varies directly as the product of the strength of the poles and inversely as the square of the distance between them. This relation is expressed mathematically in the following form:

$$F = \frac{M_1 \times M_2}{\mu R^2}$$

where M_1 and M_2 are the strengths of the two poles, and R is the distance between them. The Greek Letter μ is a constant whose value depends upon the medium in which the poles are located. If the space between the poles is a vacuum (air is approximately the same), μ is unity. If the space between the poles contains a magnetic material, μ is some number greater than one. (This is discussed further under permeability.)

The unit of pole strength (referred to as a unit pole) is derived from the equation for the force between two poles. If M_1 and M_2 are equal and, at one cm apart, exert a force of one dyne, they are unit poles. A unit magnetic pole, then, has such strength that it will exert a force of one dyne upon an equal pole in a vacuum (or air) when placed one cm away from it.

Magnetic Field

The region around a magnet in which its effect can be detected is known as its magnetic field. When a magnetic pole is moved about the magnetic field, there will be exerted on the pole, forces that vary both in direction and intensity. At any place about a magnet, the field has a certain direction and intensity. The direction of a magnetic field at any one place is the direction in which a unit north magnetic pole would have a force applied to it. The intensity or strength (number of lines of force per unit area) at any point is equal to the force which the field exerts on a unit north-pole placed at that point. This relation is represented by the equation,

$$H = \frac{F}{M}$$

where H is strength of the field in oersteds, M the strength of the pole brought into the field, and F the force which the field exerts on this pole.

Example

Problem. Find the strength of a magnetic field at a point where it exerts a force of 50 dynes, acting on a pole to the right. Assume the strength of the pole to be 10 units.

Solution:

$$H = \frac{F}{M}$$

$$H = \frac{50}{10}$$

$$H = 5$$

Therefore, the strength of the field at the point in question is 5 oersteds to the right.

In the region around a pole, the intensity of the field is not uniform, but diminishes rapidly as the distance from the pole increases. Experiments have shown that the field of intensity around a pole varies directly as the pole strength, and inversely as the square of the distance from the pole. This relationship takes the following form:

$$H = \frac{M}{\mu r^2}$$

or in a vacuum

$$H = \frac{M}{r^2}$$

where M represents the strength of the pole, r represents the distance of the place from the pole being determined, and H the force in oersteds.

Experiments show that the direction of the magnetic field about a magnet is radially outward from the N-pole and radially inward toward the S-pole. To find the intensity and direction at any point about the magnet, apply the equation $H = \frac{M}{\mu r^2}$ to both poles, and add the results vectorially. The resultant vector direction is the direction of the field.

Lines of Force

Placing a sheet of glass or some other nonmagnetic substance over a magnet and sprinkling iron filings on it provides a means of observing the configuration of the magnetic field, since tapping the glass causes the filings to align themselves into chains or lines, producing a facsimile of the magnetic field itself. At any particular point in the space around the magnet, there is a state of stress which exerts a force on any pole brought into the vicinity of the magnet. The direction that such a force takes indicates the direction of the magnetic field at that point, and lines connecting the direction of the field at a series of points, form lines called *lines of force*. It is along these lines of force that iron filings align themselves.

According to the illustration on this page which shows the lines of force about a bar magnet, lines of force travel from the N-pole to the S-pole, have only one direction at a given point and form closed loops. The path of each loop is from the N-pole to the S-pole in space and thence through the magnet to the N-pole.

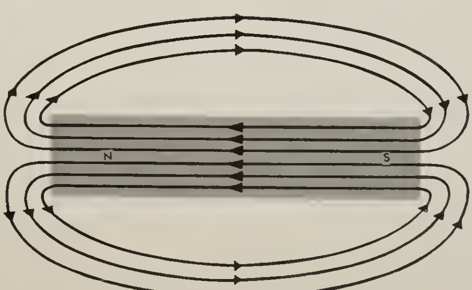
A convenient way of regarding lines of force is to think of them as rubber bands which in trying to shorten themselves establish a stress field of mutual repulsion. When two magnets are brought together the fields of stress interact causing repulsion or attraction, depending upon the polarity of the poles.

Earth's Magnetic Field

That magnetic lines of force surround the earth much in the same way that lines of force surround a bar magnet, and that the earth has two magnetic poles, one near the geographic north pole and the other near the geographic south pole, are facts responsible for man's most important navigational instrument—the magnetic compass. Dating back to the middle ages, the magnetic compass still is today the most used navigational instrument. Basically, it is an artificial magnet mounted in such a way that it aligns itself with the earth's magnetic lines of force with one end of the magnet always pointing toward magnetic north.

Molecular Theory of Magnetism

In experiments consisting of breaking a magnet in two and then breaking each of these parts in two, and so on as far as physically possible, it is found that the parts are magnets themselves. From this discovery it is presumed that if the



Magnetic Field about a Bar Magnet

breaking process were continued until the parts were the size of molecules, each part would be a magnet with N and S poles. This assumption that a magnet is made up of molecular sized magnets, is called the *molecular theory* of magnetism. According to this theory, the molecular magnets in a piece of unmagnetized magnetic material are not arranged but are in the form of small stable groups pointing in various directions and displaying no noticeable magnetic characteristics. When you place the unmagnetized material in a magnetic field, the molecular magnets tend to align themselves with the field to a degree depending upon the strength of the magnetic field and the kind of material. For example, the molecular magnets in a piece of steel require a stronger field for alignment than do those in a piece of soft iron. However, upon removal of the magnetic field, the molecules in the piece of steel remain aligned longer than do the molecules in the piece of soft iron.

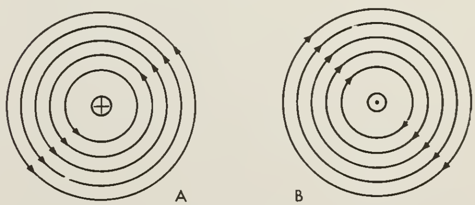
Soft iron conducts magnetic lines of force more readily than steel because soft iron has higher permeability than steel. Steel holds magnetism longer than soft iron because steel has greater retentivity than soft iron. Materials with high permeability can be magnetized easily, but do not hold their magnetism long after the magnetizing agent has been removed. In other words, their retentivity is poor. On the other hand, materials which are hard to magnetize, once magnetized, retain magnetism for an appreciable period after removal of the magnetizing agent. These materials have high retentivity.

Another fact which the molecular theory explains is that you can magnetize a piece of steel by holding it parallel with the earth's magnetic field and striking it several times with a hammer. The blows struck with the hammer shift the molecular magnets within the piece of steel into a state of alignment, which according to the molecular theory is the condition of magnetism. Similarly, this theory accounts for the fact that you can demagnetize a permanent magnet by heat. Heating the magnet accelerates the motion of the molecules in the magnet causing them to be thrown into a state of disalignment, the condition of demagnetization.

ELECTROMAGNETISM

Over a century ago, Oersted, a Danish physicist, discovered the fact that a current-carrying conductor is surrounded by a magnetic field. He also discovered that at any point the direction of this field is tangent to a circle about the conductor, that its strength diminishes as the distance from the concentric circles increases, and that the lines of force form about the conductor.

The lines of force about a current-carrying conductor travel in either a clockwise or a counterclockwise direction depending upon the direction of electron flow. At A in the illustration just below, the electrons are moving into the page. In this case, the direction of the field about the conductor is counterclockwise. At B the electrons are moving out of the page. In this case, the direction of the field is clockwise.



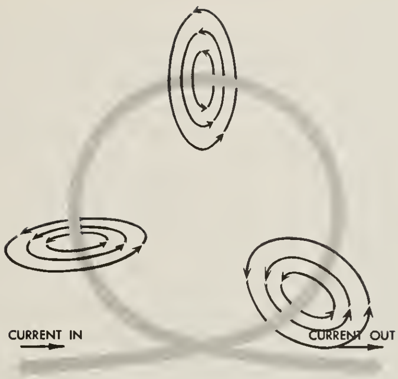
Magnetic Fields about Conducting Wires

To determine the direction of the lines of force about any conductor use the left-hand rule which states that if you grasp a conductor with your left hand in such a manner that your thumb points in the direction of electron flow, your fingers will indicate the direction of the lines of force.

To find the strength of the field at any point about a long, straight, current-carrying conductor, use the equation, $H = \frac{2I}{10r}$ where H is the strength of the field in oersteds, I, the current in amperes, and, r, the distance from the conductor in centimeters.

Magnetic Field about a Loop

If you bend the straight conductor into a single turn loop as is shown on the next page, the lines of force will concentrate within the loop. This concentration is due to the fact that all lines of force enter the loop from one side and leave at the other. To find the intensity of the field at the center of the loop use the equation, $H = \frac{2\pi I}{10r}$ where I is the current in amperes, r the radius of the loop in centimeters, and H the field intensity in oersteds.



Effect of Loop on Fields

Field in a Solenoid

If you wind several turns of wire close together into the form of a coil, the magnetic fields about each turn will all have the same direction. When the length of the coil is small compared to its radius, you can find the strength of the field at the center by the equation,

$$H = \frac{2\pi NI}{10r}$$

where N equals the number of turns, r the radius in centimeters and I the current in amperes. When the length of the coil is longer than the radius, the coil is called a *solenoid.*

When current flows through it, the coil (or solenoid) is surrounded by a magnetic field like that shown directly below. One end of the coil is called the north magnetic pole, and the other end the south magnetic pole. To determine the polarity of a coil use the left hand rule as follows. Grasp the coil with the left hand in a manner that the fingers point in the direction of electron flow and note the direction the thumb points. This direction is the north pole. To obtain the intensity of the field at the axis and near the middle of a coil, use the equation,

$$H = \frac{4\pi NI}{10\ l}$$

where I is the current in amperes, N the number of turns, l the length of the coil in centimeters, and H the intensity in oersteds.

If you substitute $n = \frac{N}{l}$ (number of turns per unit length) the equation becomes $H = 4\pi nI$. The significance of this substitution is that by the second equation it is apparent that the strength of the field about a solenoid depends only upon the magnitude of current and the number of turns per unit length.

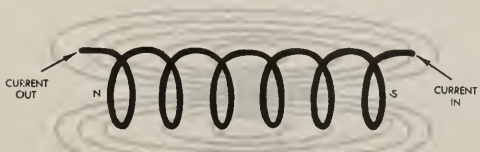
Inserting a soft iron core into a solenoid greatly increases the number of magnetic lines of force. This increase in magnetic lines is not from an increase in the intensity of the field, which you have just learned depends only on current and turns per unit length, but from additional lines produced by the magnetization of the iron core.

Magnetic Flux

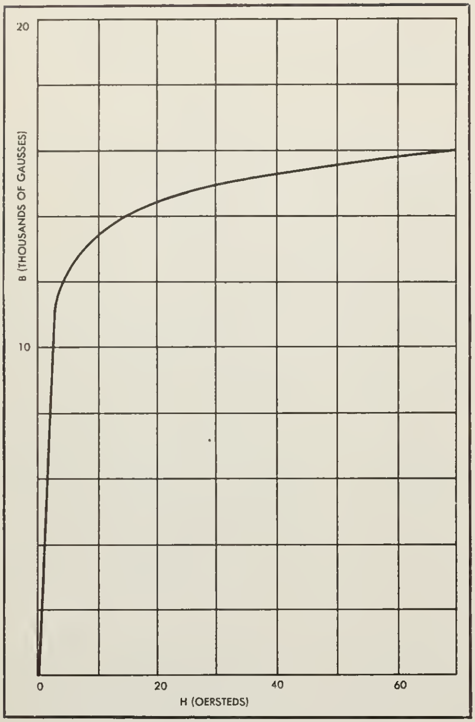
The total number of lines of force occupying a region of magnetic activity is called the *magnetic flux* and is represented by the Greek letter ϕ. The unit of magnetic flux is one line of magnetic flux and is called the *maxwell.* The number of lines passing perpendicularly through a square centimeter is called *flux density* and is represented by the letter B. The unit of flux density is the *gauss* and is equal to one maxwell per square centimeter.

Permeability

The amount by which the flux density (B) exceeds the field intensity (H) in a coil or solenoid depends upon the type of core. The ratio between these two quantities, which is written $\frac{B}{H}$, is called the *permeability* of the core. For example, soft iron has a permeability about 2000. Earlier in the discussion of magnetism, permeability was defined as the ease with which a substance conducts magnetic lines of force. Permeability is represented by the Greek letter mu (μ). In any ferromagnetic substance permeability is not a constant quantity but a quantity which depends largely upon the intensity of the field. For example, magnetization curves, the curves which indicate the relation of flux density B to field



Magnetic Field Surrounding a Solenoid



Magnetization Curve of Steel

intensity H, show that in the case of silicon steel, the flux density increases practically in a direct proportion to H, at low field intensities, but that at large field intensities, the steel becomes saturated with flux and a large change of H is required to increase B further.

You can find permeability from the values of H and B in the diagram by solving the equation,

$$\mu = \frac{B}{H}$$

Example

Assume $H=70$, and $B=1600$. Find μ.

$$\mu = \frac{B}{H}$$

$$\mu = \frac{1600}{70} = 22.86$$

Likewise, find μ when $B=10000$ and $H=3$.

$$\mu = \frac{10000}{3}$$

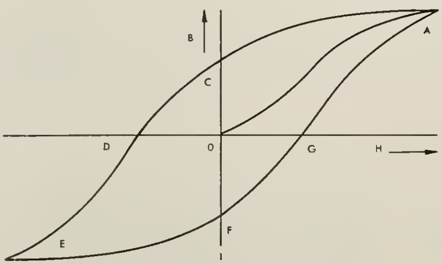
$$\mu = 3333$$

(Note the difference in permeability at the different points on the curve.)

Hysteresis

When a piece of iron is magnetized, considerable energy is expended in lining up the molecular magnets in the iron in a definite direction. Therefore when they are aligned first in one way and then in the other many times per second, as for example, in an alternating current electromagnet, considerable energy is wasted in the form of heat. Such a waste of energy is called hysteresis loss and in electromagnets causes the magnetization of the core not to reverse polarity at the same time that the magnetization current does.

The best way to understand the effects of hysteresis is to study hysteresis loops. A hysteresis loop is a curve which shows the variation of magnetic intensity to flux density, or magnetic induction. For example, in the hysteresis loop shown below, the flux density increases along OA as the magnetizing force is increased, but does not decrease to zero along AO when the magnetizing force is decreased, but follows a new curve, AC. When the magnetizing force again reaches zero, it has a value equal to OC. The magnetism left in the magnet (OC) when the magnetizing force is zero is called *residual* magnetism. When the current in the electromagnet reverses, the magnetizing force likewise reverses and the flux density decreases along CD. However, it requires applying the magnetizing force OD to reduce the magnetism in the material to zero. OD is called the *coercive force.* As the magnetizing force continues to increase, the material becomes magnetized in the opposite polarity, and the flux density increases along DE. Then, as the current decreases, the flux density decreases along EF.



Hysteresis Loop

OF, which equals CO, is the amount of residual magnetism in this direction. As the magnetizing force increases in the original direction, the flux density first drops to zero at G, and then increases to point A. Thus, you see that each time the magnetizing current in the electromagnet reaches zero the magnetism is not at zero but must be brought to zero by a coercive force. If it were not for hysteresis, the magnetization current and magnetism would reach zero simultaneously.

Experiments show that the energy loss due to hysteresis is proportional to the area of the hysteresis loop. Hence, it is customary in such cases where alternating current transformers and alternating current electromagnets are used to choose a kind of steel which tests show have a thin loop; that is, in which the distance DG in the hysteresis loop is relatively small.

Magnetic Circuits

Because magnetic flux of lines of force form closed loops, the path that the flux loops follow is called the magnetic circuit. Electrical circuits and magnetic circuits have many points of similarity. The force which produces a flow of electrons in an electrical circuit is the electromotive force. In the magnetic circuit the force which produces the flux is called the *magnetomotive force* (*m.m.f.*). Similarly, just as the resistance opposes the flow of current in an electrical circuit, so *reluctance* opposes the magnetic flux in a magnetic circuit. Similarly, just as *conductance* indicates the ease with which electrical current flows, so *permeability* indicates the ease with which magnetic lines of force flow in a magnetic circuit.

No name has been given to the unit of reluctance in a magnet circuit. Quantitatively, one unit of reluctance is the reluctance of a magnetic circuit one cm long and one sq cm in cross section with unit permeability. Mathematically, the unit of reluctance is expressed by the equation;

$$R = \frac{l}{\mu A}$$

where l represents length in centimeters, A the cross-sectional area in sq. cm, and μ the permeability.

Much in the same manner that Ohm's law expresses the relation between current, voltage and resistance, the expression,

$$\phi = \frac{mmf}{R}$$

gives the relation between magnetomotive force, flux and reluctance where ϕ is flux in maxwells, mmf the magnetomotive force in *gilberts* (a gilbert is the mmf required to establish a flux of one maxwell in a magnetic circuit in which the reluctance is one unit), and R the reluctance.

Force on a Conductor in a Magnetic Field

When a current-carrying conductor is located in a magnetic field, the interaction of the field about the conductor and the magnetic field exerts a force upon the conductor. By tests, it can be proved that this force is proportional to the flux density, the current, and the length of the conductor. Mathematically, this is expressed by the equation,

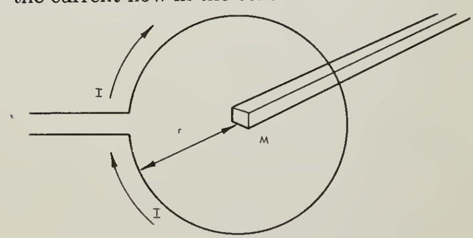
$$F = \frac{BIL}{10}$$

where F is the force in dynes, B the flux density per square centimeter, I the current in amperes, and L the length of the conductor in centimeters.

This is how the formula is derived. Assume that an isolated magnetic pole with strength M is located at the center of a coil with N turns, a radius r, and current I amperes. The force which acts on the magnetic pole is expressed by the equation,

$$F = MH$$

where H is the field strength which results from the current flow in the coil.



Force on Unit Magnetic
Pole at Center of Conducting Loop

Previously, it was shown that H is equal to $\frac{2\pi NI}{10r}$. By substituting this value for H in the preceding equation, it becomes,

$$F = M \frac{(2\pi NI)}{10r}$$

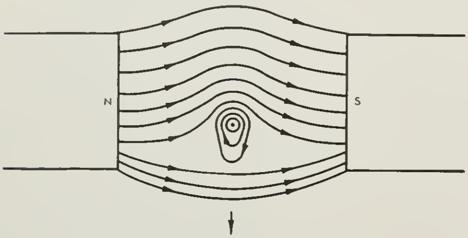
By multiplying both numerator and denominator by r, the equation becomes,

$$F = \frac{M}{r^2} \bullet \frac{I}{10} \bullet 2\pi rN$$

where $\frac{M}{r^2}$ is the flux density of the magnetic field at r centimeters from the pole of M strength, and the quantity $2\pi rN$ is the length of the conductor.

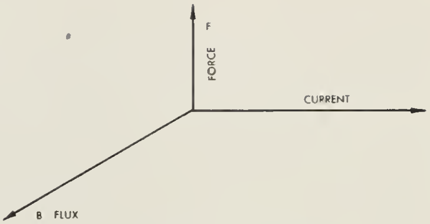
Since flux density is also expressed as $\mu H = B$ (see discussion of permeability on page 2-21), you can replace $\frac{M}{r^2}$ in the equation above by B. Then by substituting the single quantity L for the length of the conductor $2\pi rN$, the equation expressing the force exerted on a conductor located in a magnetic field is

$$F = \frac{BIL}{10}$$



Composite Magnetic Field about Conductor Located Between Magnetic Poles

The principle of magnetism which makes possible the operation of electric motors is that when a current-carrying conductor is placed in the field produced by two magnetic poles, the field about the conductor and the field between the poles react to produce a force which causes the conductor to move either upward or downward. For example, notice the illustration showing the composite of the field about a conductor and the field between two poles. Since the electrons in the conductor flow toward you, according to the left-hand rule, the field about the conductor is clockwise. Notice that when this is the case, the conductor field has the same direction at the top, as the lines of force do between the two magnets. Under this condition, the two fields reinforce each other at the top and oppose each other at the bottom. Where these fields travel in opposite directions, they counteract each other. In other words, there is a strong force at the top of the conductor, and a weak force at the bottom. A rule of magnetism is that a current-carrying conductor which is located in a magnetic field is always pushed away from the point of stronger force. Therefore, the direction of motion of the conductor illustrated is downward. Had current flowed away from you in the conductor, the two fields would have strengthened each other at the bottom and weakened each other at the top and the conductor would have been pushed upward.



Direction of Force and Conductor

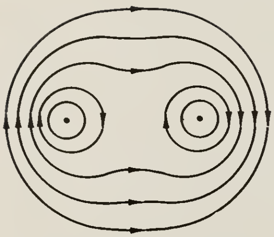
Another fact of interest in connection with a current-carrying conductor located in a magnet field is the relationship between force, flux, and current flow. When the conductor is at right angles to the field, force, current and flux are all mutually perpendicular as shown by the vectors directly above. Mathematically, their relationship is expressed by the formula $F = \frac{BIL}{10}$, which was stated on a previous page and which is often referred to as Amperes Law. Had the conductor not been at a right angle, but at some angle ϕ with respect to the field, the vectorial length of the conductor would be effectively $L \sin \phi$ and the equation would read

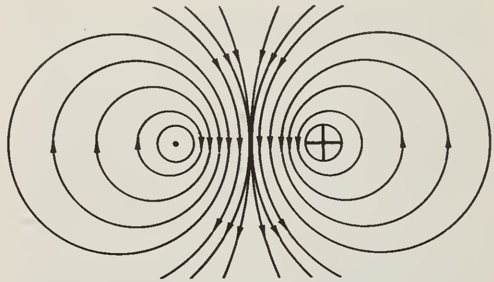
$$F = BIL \sin \phi$$

Whenever two current-carrying conductors are located near each other, as illustrated on the next page, the magnetic fields about the conductors produce a force which exerts itself on each conductor. If electron flow in the two conductors is in the same direction, the magnetic fields surrounding each conductor cancel between the conductors and strengthen outside the conductors. This force which results is called a force of attraction. When the current flow in the conductors is in the opposite direction, the fields strengthen between the conductors and weaken outside. This force is one of repulsion.

INDUCTION

In the year 1831, Michael Faraday performed a simple experiment with a coil and a permanent magnet and discovered that a galvanometer connected to a coil would deflect in one or the other direction, depending on whether the magnet was being thrust into, or being removed from the coil. At that time, it was known that a battery, the only known source of emf, caused current to flow through a circuit. Thus, Faraday





Field Surrounding Two Adjacent Conductors

concluded that electrons flow through the coil when there is relative motion between the coil and *magnetic field* of the magnet, and that an emf is being induced in the coil. Later experimentation showed that the magnitude of the induced emf (which is indicated by the range of swing of the galvanometer pointer) depends on the number of turns of the coil, and the rate of change of the magnetic field through the coil; and that the polarity of the induced emf depends on whether the flux is increasing or decreasing through the coil.

Quantitative Law of Induced EMF

In the preceding paragraph, you saw that a current is produced while a magnetic field is changing and that an induced emf produces this current. Actually, the induced emf is the primary effect and the current only the secondary. The law which expresses the magnitude of the induced emf is the following: Whenever the magnetic flux through a coil changes, an emf is induced in the coil equal to the rate of change of the flux multiplied by the number of turns in the coil.

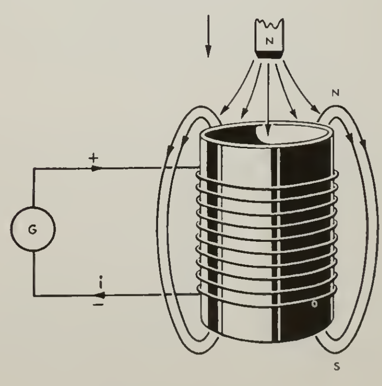
In the form of an equation, the law reads,

$$E = \frac{-N\Delta\phi}{\Delta T} \times 10^{-8}$$

where E is the induced voltage, N the number of turns, $\Delta\phi$ the change in flux, and ΔT is the change in time. The ratio $\Delta\phi/\Delta T$ represents the rate of change of flux with time. The term 10^{-8} converts the rate of change of flux in lines per second into induced emf in volts.

The negative sign in the equation indicates that the polarity of the voltage is such as to cause a current to flow (if the circuit is a closed one) in such a direction as to oppose the motion of the magnetic field that is inducing the voltage. To make this clearer, imagine a wire coil which is wound in the form of a solenoid and is placed so that its turns are in a horizontal plane. Across the coil either a galvonometer or a resistor is connected so that there will be a current flow when a voltage is induced. When the north end of the magnet is thrust into the center of the coil, a voltage is induced, and the current which arises from this voltage will flow in such a direction as to cause the top end of the coil to appear as a north pole and thereby oppose the downward direction of the magnet.

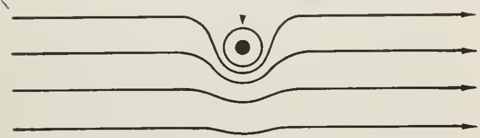
This is illustrated in the diagram below. As shown, the north end of the magnet is moving toward the coil and electrons are flowing in the direction indicated by the i arrow. So far as the external circuit is concerned, the polarity of the induced emf is as shown, the top lead being positive and the bottom lead negative.



Inducing Voltage in Solenoid by Moving Magnet

On giving it some thought, it is clear that the induced polarity of the coil must be the same as that of the approaching pole. Otherwise, it would be necessary only to start a magnetic pole into a coil, and upon its inducing an opposite pole in the coil, the magnet would be drawn in, inducing additional emf without work. This would provide an unlimited source of energy without doing any work, a condition which is contrary to the law of the conservation of energy.

Next, consider a wire moving downward with its length perpendicular to a magnetic field, which runs from the left to the right. Think of the lines as being distorted as shown below.



Cutting of the lines of flux will induce an emf along the wire so that, if current could flow, it would flow into the page. A simple rule which determines the *direction of electron flow* through the wire, is to grasp the wire with the left hand so that the fingers point in the way the distorted lines go around the wire; the thumb, which lies along the wire, will point in the direction of the electron flow.

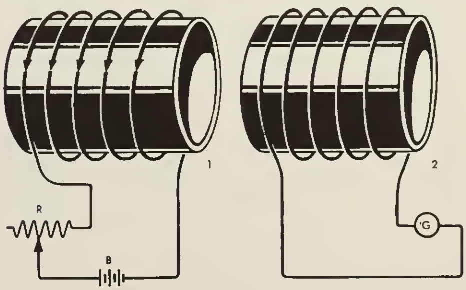
Lenz's Law

H. Lenz, a Russian, in 1834, summed up induction in the form of a law which is used in many cases to predict direction of induced current and voltages. This law, which is known as Lenz's law, is as follows: Whenever an induced current is produced by any motion, current will flow in a direction such that mechanical forces will be produced which oppose the motion.

Mutual Induction

In an electrical circuit, a change of current is always accompanied by a change in the magnetic field surrounding the circuit. If the current is increasing, the field is said to be *expanding*, that is, its intensity at any particular point is increasing. On the other hand, if the current is decreasing, the field is said to be *collapsing* or decreasing in intensity. When a conductor is placed within the magnetic field of a circuit in which the expanding or collapsing lines of force cut the conductor, a voltage will be induced in it.

In this connection, consider coils 1 and 2 shown in the next illustration to be side by side, and assume that battery B is connected to coil 1 and that galvanometer G (an instrument for indicating the amount and direction of current flow) is connected to coil 2. As long as the current flowing in coil 1 is steady, its magnetic field will likewise be steady, and there will be no voltage induced in coil 2. However, if you change the variable resistance R, there will be a change in the current in coil 1, and as the result, the flux through coil 2 will change and there will be a deflection on the galvanometer. Obviously, this deflection is due to current flow resulting from a voltage induced in coil 2. This action in which a change in current in one coil results in a voltage being induced in the other is called *mutual induction.*



Mutual Induction

If in a short period of time, Δt, the current in coil 1 in the illustration changes the amount, Δi, and causes a change $\Delta\phi$ in the flux through coil 2, the voltage induced in coil 2 is equal to,

$$e_2 = -N\frac{\Delta\phi}{\Delta t} \times 10^{-8}$$

When a change in current flow in a coil causes a change in flux in an adjacent coil, the change in flux is proportional to the change in current. Therefore the change in flux $\Delta\phi$ in coil 2 is proportional to the change in current, Δi, in coil 1. This proportion is expressed as,

$$\frac{\Delta\phi}{\Delta t} \propto \frac{\Delta i}{\Delta t}$$

Similarly, the induced voltage is proportional to the rate of change of current in coil. Thus the voltage induced in coil 2 is equal to,

$$e_2 = -M\frac{\Delta i}{\Delta t}$$

where M is a factor which depends on the magnetic coupling between the two coils and is called mutual inductance.

The formula indicates that the more rapid the change of current in coil 1, the greater the induced voltage. In tests, it can be shown that if coils 1 and 2 are interchanged, the voltage induced in coil 1 by a change of current in coil 2 is equal to,

$$e_1 = -M\frac{\Delta i}{\Delta t}$$

Mutual inductance is represented by the symbol M and is measured in henrys. A henry is defined as follows: Two coils have a mutual inductance of one henry if a change in current at the rate of one ampere per second in one coil results in a voltage of one volt being induced in the other.

When a soft iron core is inserted between two coils, the mutual inductance between them is greatly increased. Inductance is proportional to the number of lines of force, and since soft iron can conduct many lines of force, the inductance of the coil is greatly increased. However, the inductance can only be increased to an amount represented by the permeability of the soft iron. Since the permeability of iron depends on the field which in turn depends on the current in the coil, the inductance of an iron core coil is not constant over large changes of current.

The direction of the current resulting from the voltage induced in a coil can be determined by Lenz's law. Refer back to the illustration for a moment. An increase in the current in coil 1 is equivalent to moving it toward coil 2 and would induce a voltage causing the current in coil 2 to be opposite to the current in coil 1. Decreasing the current in coil 1 will cause current in coil 2 to be in the same direction as in coil 1.

You have just seen that a change in current flowing in one coil results in a voltage being induced in a second coil nearby. At the same time that this action occurred, the flux through the coil in which the change of current took place also changed and a voltage was induced in it. This action in which a change in current in a coil results in a voltage being induced in the coil by its changing magnetic field is called self induction. This voltage is also proportional to the rate at which the current is changing and is equal to

$$e_t = -L\frac{\Delta i}{\Delta t}$$

where L is a factor called *self inductance* and, like M in the formula for mutual inductance, is dependent upon the magnetic characteristics of the coil.

The minus sign in the formula indicates that the induced voltage, which is called the *counter emf* or back emf, is of such a polarity as to oppose the change of current that caused it.

The unit of self inductance (usually shortened to *inductance*) is the henry. A henry is defined as the inductance of a coil in which a voltage of one volt is induced by a current change of one ampere in one second.

The inductance of a coil of many turns is much greater than the inductance of a coil of a single turn since the induced voltage is dependent not only on the change in flux, but also on the number of turns through which the flux passes. Inserting an iron core in a coil greatly increases the inductance. However, the increase is not constant over a wide range of current flows.

When two coils are placed close together, the relation between the mutual inductance, M, of two coils and their individual inductances, L_1 and L_2, is equal to

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

where k has a value between zero and one and is called the coefficient of coupling.

If *all* the flux produced by a current in one coil links *all* the turns of the other, k is equal to one.

On comparing the formulas

$$e_2 = -N\frac{\Delta\phi}{\Delta t}\times 10^{-8} \text{ and } e_2 = -L\frac{\Delta i}{\Delta t}$$

which respectively give the voltage induced in a coil by a changing flux and by a change in current, you can see that

$$L\frac{\Delta i}{\Delta t} = N\frac{\Delta\phi}{\Delta t}\times 10^{-8}$$

$$L = N\frac{\Delta\phi}{\Delta i}\times 10^{-8}$$

Then assuming that the flux and current start from zero, $\Delta\phi$ becomes ϕ when Δi reaches i, and self inductance is given by the equation,

$$L = \frac{N\phi}{i}\times 10^{-8}$$

The product $N\phi$ is referred to as flux linkages, since the flux loops ϕ are linked by N turns. This formula defines a henry in a different form. According to it, the henry is the inductance of a circuit which produces 100 million flux linkages per ampere of current in that circuit.

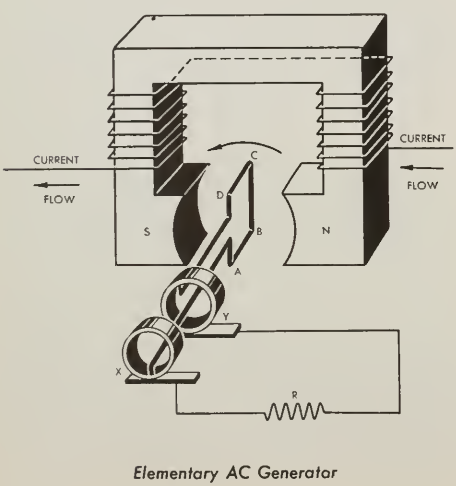
GENERATORS AND MOTORS

GENERATORS

Whenever a conductor moves in a magnetic field in such a way as to cut across lines of force, first in one direction and then in the other, an alternating voltage is induced in the conductor and current arising from this voltage flows first one way and then in the other. A current which reverses direction periodically is called an *alternating current*, and is abbreviated AC. On the other hand, a current which flows continuously in the same direction is called direct current, or DC. A generator which produces an alternating current and voltage is called an AC generator. One which produces DC is called a DC generator. Essentially, the two types are alike, their principal difference being in the method in which the generated energy is taken from the generator. Energy is taken from an AC generator by slip rings, and from a DC generator by commutators as explained later. Internally, both produce an electrical voltage in the same manner.

AC Generator

The elementary AC generator shown on this page consists of a conductor formed into the loop ABCD and located in the magnetic field between the poles of an electromagnet. The two ends of the loop connect to slip rings *X* and *Y* and they in turn make contact with two brushes which are connected to the resistor *R*.



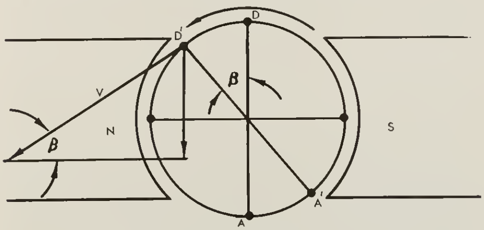
Elementary AC Generator

When the loop rotates about an axis in the plane of the coil and perpendicular to the field, it cuts the flux first in one direction and then in the other. At the instant when the loop is in the position illustrated, the coil sides *AB* and *CD* are moving parallel to the field and cut no flux. Therefore there is no voltage induced in the loop at this position. As the coil turns in a counterclockwise direction, side *AB* moves up and *CD* moves downward through the flux, and a voltage is induced in *A* toward *B* and in *C* toward *D*. These two voltages add together in series and make brush *X* positive and brush *Y* negative. The radial portions of the loop do not cut across flux and hence there is no voltage induced in them. The voltage that is thus induced in the loop causes current flow in the resistor *R* from *Y* to *X*. This current keeps increasing in magnitude and reaches maximum when the coil is horizontal to the lines of flux, at which time the loop is moving perpendicular to the field and is cutting the greatest number of lines. As the coil turns further, the induced voltage and current decrease until they reach zero where the loop is again in a vertical position. During the other half revolution an equal voltage is produced except that its polarity is reversed since then both *AB* and *DC* cut flux in the opposite direction. Due to the reversed polarity of the induced voltage, the current flow through *R* is from *X* to *Y*.

To investigate how the emf (or current) varies from moment to moment, consider the coil shown in cross section in the next illustration. *A* and *D* are the ends of the two conductors facing you. The coil is revolving at constant speed V in a uniform magnetic field. As was explained before, at the position *AD*, the voltage induced in the coil is zero. As the coil passes the axis of the poles, NS, the induced voltage reaches a maximum value E_m. At some intermediate position, $A'D'$, the coil makes an angle β with the starting position *AD* and the induced voltage there has a value between 0 and E_m. The value of this voltage becomes apparent on resolving the linear velocity, *V*, of the conductor into its horizontal and vertical components. The vertical component, $V \sin \beta$, represents the effective component since it is the one perpendicular to the flux. It is seen that the ratio of the induced voltage in the position $A'D'$ to the maximum induced voltage E_m is the sine of the angle β. If you let e represent the instantaneous value, then $e = E_m \sin \beta$. If t is the time in which a coil turns through the angle β, then the angular velocity ω of the coil is equal to β/t. Another way of expressing this is $\beta = \omega t$. As an equation, the relationship between the induced voltage, the maximum induced voltage and the angular velocity of the coil can be expressed as,

$$e = E_m \sin \omega t$$

The conclusion is that a coil rotating at a uniform rate in a uniform magnetic field generates a sine wave of voltage.



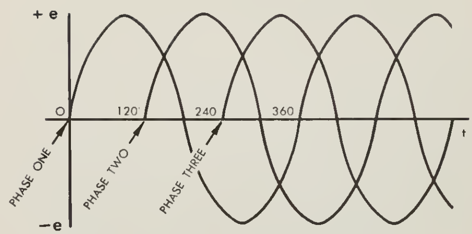
Coil Rotating in a Uniform Magnetic Field

Practical AC Generator

The elementary AC generator just described can be expanded into a practical AC generator usually by increasing the number of conductors in the armature, and by increasing the number of pairs of poles.

As previously mentioned, the principal difference between an AC and DC generator is the method of connection to the external circuit, that is, the AC generator is connected to the external circuit by slip rings, while the DC generator is connected by commutators. Another difference is that since the output current of AC generator sets up a magnetic field which reverses several times per second, it is not possible to use the generator output to excite the field coils. Therefore an AC generator is separately excited, usually by means of a smaller DC generator which is located or geared to the same shaft as the AC generator. (Field excitation is discussed under DC generators.)

An AC generator which has a single set of coils all of which are connected in series so that the emf's are additive is called a *single-phase* generator and its output is called *single-phase* current. If the coils in a generator are divided into a number of sets spaced and arranged so that the voltage output of each set of coils reaches a maximum point (or a minimum point) at different positions in the loop (armature) rotation, it is called a *polyphase* generator. For example, one

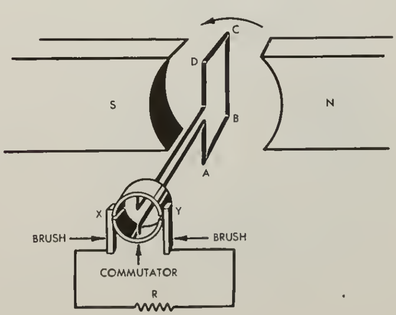


Output of Three-Phase Generator

type of polyphase generator which is used frequently in heavy power equipment has three sets of coils. This generator is called a *three phase generator*. Because heavy current is taken from the armature windings, and because four to six brushes are needed, the armature in this generator is usually stationary and the field is rotatable. It produces three phases of voltage with but one set of brushes. The sine waves of voltage differ in phase from each other by a third of a cycle (120°).

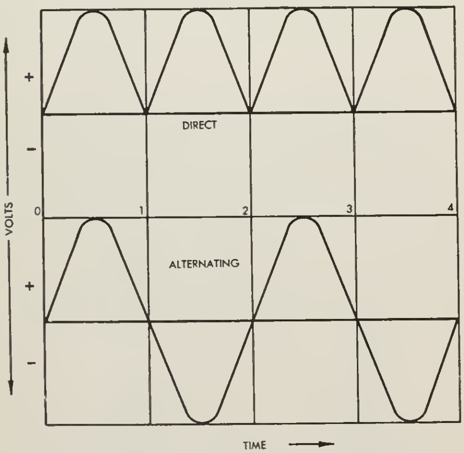
Direct Current Generator

To convert a simple alternating current generator into a direct current generator, it is necessary only to replace the slip rings of the AC generator by a commutator. A commutator is a ring cut into two halves insulated from each other. Each half is connected to one end of the generator coil. It serves to reverse the connections of the loop to the external circuit, by means of the brushes, at the instant the polarity of the induced voltage in the loop coil reverses.



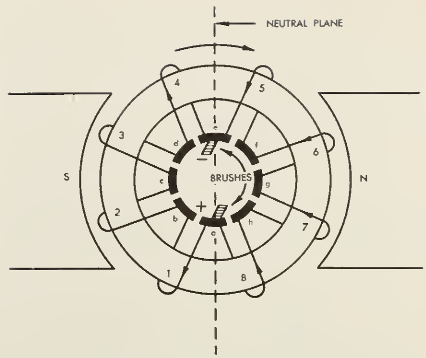
Elementary DC Generator

To understand the action of the commutator, refer to the waveforms shown below and the elementary DC generator on the preceding page. Starting from the position of the loop shown in the elementary DC generator, the action resulting during the first-half revolution of the loop is identical to that of an AC generator inasmuch as both generate a positive lobe of voltage as you can see on examining both waveforms from 0 to 1. During this time the conductor *AB* is connected to brush *Y* and *CD* to *X*. At the proper moment, the commutator interchanges the generator connections *AB* to *X* and *CD* to *Y*, so that during the next half of the revolution of the loop (1 to 2), the generator produces another positive lobe instead of a negative lobe as the AC generator does. Thus, you can conclude that the polarity of the brushes remains the same, *Y* negative and *X* positive throughout the entire revolution in the DC generator, and that the current flow through *R* is always from Y to X. In a simple one-loop DC generator as this one, the output is not pure DC but DC of the pulsating variety as shown by the DC waveform.



Waveforms of Simple DC and AC Generators

A generator employing a one turn loop for the armature produces only a small voltage output. To increase the output and at the same time make it steady instead of pulsating, several coils are equally spaced and connected in series around the armature. Each coil is connected to special connectors on the commutator called commutator *segments*. The number of commutator segments is equal to the number of coils. Each coil is connected to two commutator segments as shown in the above diagram. One end of coil 1 goes to segment a and the other to b and so on around the armature. The two brushes are located so that the contact with the commutator is in a plane exactly perpendicular to the magnetic lines of force. This plane is called the *neutral plane*. In the neutral plane no voltage is induced in windings since they move parallel to the lines of force. Immediately upon emerging from this plane, the windings cut the lines in reverse direction, and the voltages induced in them changes polarity. At the instant shown, coils 1, 2, 3, and 4 are cutting across the field in an upward direction inducing an emf to cause electron flow from segment a to segment b, from b to c, from c to d, and from d to e. Since these all tend to produce flow in the same direction, the voltages are additive, and the amplitude of the resulting voltage is increased and the amount of variation (pulsating) in the output is small.



DC Generator with Several Coils

The output variation of a DC generator is called ripple. The frequency of the ripple depends on the number of coils and the speed of rotation. The greater the number of coils per pole, the smaller the amplitude of the ripple voltage.

In connection with ripple, notice the output voltages of the DC generator for coils 1, 2, 3, and 4 at the top of the next page. There is a voltage induced in each coil. The combined output of these voltages is shown by the waveform just above the individual voltages of each coil.

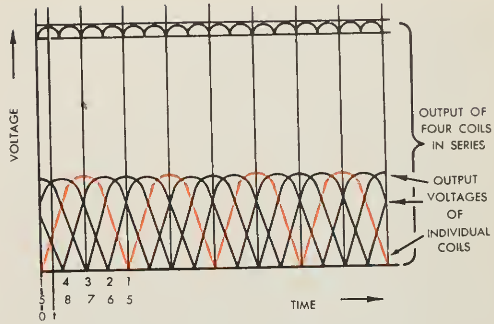
The red lined waveform is the output of a single coil (coil 1). The ripple frequency in this case is low, being only four cycles per second. But in the case of the output of all four coils (1, 2, 3, and 4) in series the ripple frequency is four times as high—that is, 16 cycles per second and approaches pure DC. The letter t represents the angular position occupied by the armature in the DC generator.

The question arises as to what is happening in coils 5, 6, 7, and 8 at the same time. They are cutting flux in a downward direction and have induced in them a voltage that cause the electron flow to be in the opposite direction from the commutator segment a to segment h, from h to g, from g to f, and from f to e. The net result is that between the positive brush and the negative brush there are two parallel paths, abcde and ahgfe. The output voltage at any instant is the same on both sides of the neutral plane and is the sum of the instantaneous voltages in the four coils on that side of the plane at the time. As the armature rotates each coil changes twice per revolution from one side of the neutral plane to the other and the current through it reverses.

In the simple generator shown on the preceding page, the armature is the ring type. This type is not as widely used as another called the drum type. The main difference is that in the ring type the windings are placed on an iron ring and the windings located on the inner side of the ring do not cut flux and consequently have no emf induced in them. In the drum type the windings are laid in slots in a drum-shaped iron core and each winding is joined to one nearly opposite it by an end connection. By this method all windings cut flux and have voltage induced in them. However, the operation of DC generator with a drum-type armature is essentially the same as the one using a ring-wound armature.

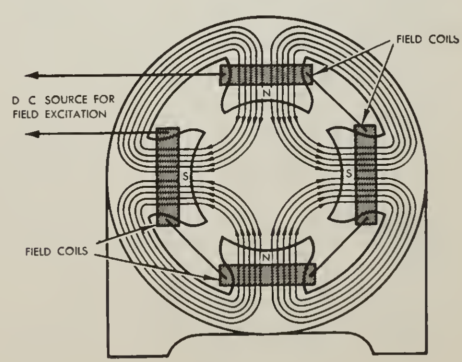
To make the magnetic field more uniform and stronger, a generator in common use has more than one pair of poles and uses electromagnets rather than permanent magnets to produce the field. The illustration to the right shows a four-pole field structure along with the resultant magnetic field. Electromagnets consist of coils which are wound on a soft iron core. By varying the applied voltage, it is possible to vary the strength of the field of the electromagnet and thereby control the output of the generator.

Usually, the number of brushes in a generator is equal to the number of poles. Each brush is located approximately halfway between each pair of poles. The positive brushes are connected together to form the positive output and the negative brushes to form the negative output. This arrangement in a four-pole machine provides four parallel paths through which the armature current can flow and also makes the internal resistance of the generator low. The voltage generated is the voltage of any one of the parallel paths and the current capacity is equal to the sum of the current in the four paths.



Effect of Coils in Series on Output Voltage

The voltage of a generator can be computed by multiplying the average voltage induced in a single conductor by the number of conductors in any one of the parallel paths. If the armature is driven at a rate of N revolutions per second and there are P poles with a flux of ϕ per pole, each conductor cuts ϕPN lines per second and the voltage induced in it is $\phi PN \times 10^{-8}$ volts. If z is the total number of conductors, and there are



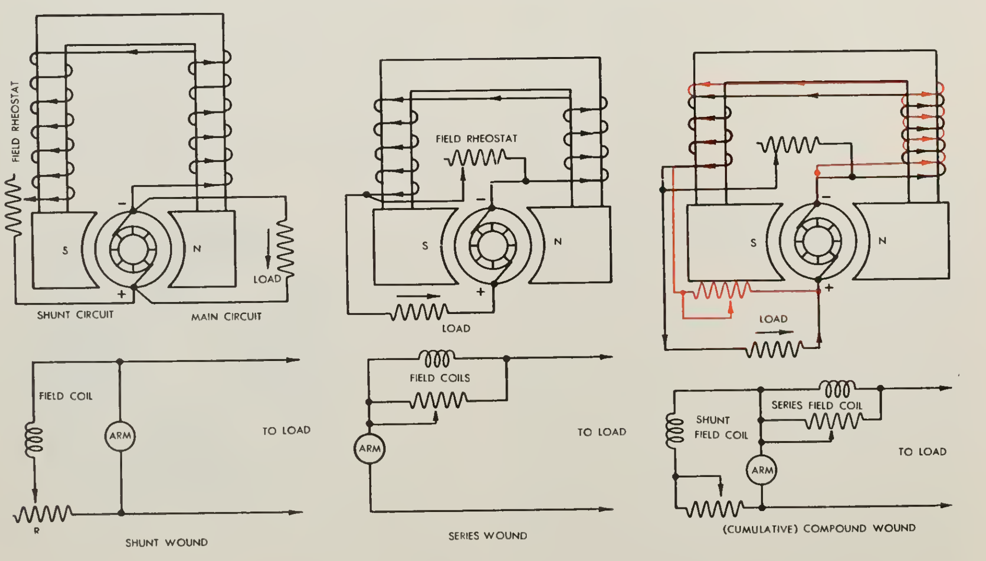
Field Structure of Four-Pole Generator p paths, the number of conductors per path is z/p. It follows that the average output of the generator in volts can be expressed,

$$E = \frac{\phi P N z}{p \times 10^8}$$

In any given generator the quantities, P, z, p are fixed by the manufacturers. This leaves only ϕ and N as the variable quantities by which it is possible to control the output voltage. The usual practice is to keep N constant by driving the generator with a fixed-speed engine and to control the output voltage by controlling the flux. Usually the flux is controlled by controlling the current in the field coils in a generator.

Field excitation. In nearly all practical generators and motors, the magnetic field is developed by electromagnets. The principal reason for using electromagnets is that they produce stronger magnetic fields and can be controlled much more readily than permanent magnets. Whether the machine is a generator or a motor, a magnetic field is necessary to develop a voltage in a generator or a torque in a motor. The field is produced by current flow in the field windings. This current can be produced by the generator itself, or it can be supplied by an external voltage. A generator which supplies its own field current is called a *self-excited* generator. Self excitation is possible if the core in a generator is slightly magnetized by residual magnetism when the generator starts rotating. The residual magnetism will generate a small voltage, and the current resulting from this voltage will flow through the field coils and increase the magnetic field. As the armature continues to rotate the small voltage is increased. This is a cumulative process and continues until the output voltage reaches normal. When an external DC voltage source is used to excite the generator, the source of voltage is usually another DC generator which is connected to the field windings. In this method of excitation the field is constant regardless of whether the armature turns or not.

Classification of generators. Self excited generators are classed according to the type of field connection they employ. There are three general types of fields—*series*, *shunt*, and *compound*. Compound generators are further classified as *cumulative compound*, and *differential compound*. In cumulative compound generators, the series and shunt fields are wound in such a direction as to *aid* each other; in the differential compound, the two fields are connected so as to oppose each other.



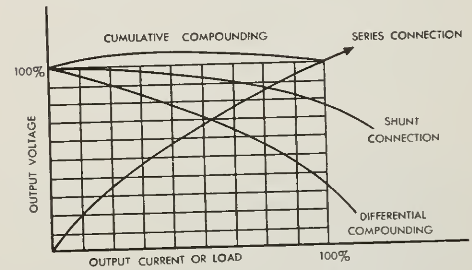
Types of Generators

Series Wound. In the series wound generator, the field coils are connected in series with the armature. Current that flows in the load also flows through the field coils. This type of machine uses very low resistance field coils, consisting of a few turns of large wire. It is characterized by very low voltage output at no load, and by an output voltage that increases as the load increases. Under no load condition, the current which flows in the load, and in turn in the generator, is small. Since a small current means a small magnetic field, only a small voltage is induced in the armature. When the ohmic resistance of the load drops to a smaller value, the load is said to be *increased.* Under this condition a higher current flows through the low resistance load, the magnetic field becomes stronger, and the output voltage increases. The voltage output increases as the load on the generator increases. This generator is seldom used because a variable output voltage normally is not desirable.

Shunt Wound. In a shunt-wound generator the field coils contain high-resistance windings, and are connected in parallel with the armature terminals. They consist of many turns of small wire, and produce maximum voltage output under no load conditions. The voltage drops with load for two reasons. First all the increased load current flows through the armature. With the armature resistance (*R*) constant, the increased current makes the *IR* drop greater. In other words the internal resistance of the armature decreases the output voltage. The second reason is a result of the first. The lowered voltage allows the field current to decrease, the magnetic field decreases, and the output voltage drops a little more. The output voltage seldom drops below 75% of the no-load voltage.

Cumulative Compound. In a cumulative-compound generator, both series and shunt fields are employed simultaneously. The shunt field is usually designed the stronger of the two fields, but both contribute to the magnetic field. When the load increases, the armature voltage decreases and the voltage applied to the shunt field decreases. The increased load current flowing through the series field increases its field. By proportioning the two fields so that the series fields just compensate the undesirable effects of the armature and shunt fields, the output voltage may be kept constant.

Differential-Compound. In a differential-compound generator, the two fields are wound so that they oppose each other—that is, the magnetic field in the series field is in opposite direction to that of the shunt field. For any increase in load, there is an increase in the strength of the series field and because its direction is opposite to the shunt field, the total field strength is decreased. Therefore, the output voltage decreases as the load increases. Differential compound generators are used in such equipment as welding generators, where the output terminals are sometimes short circuited.

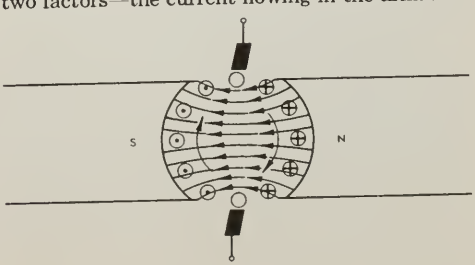


Load Characteristics of DC Generators

DIRECT CURRENT MOTOR

A DC motor depends for operation on the fact that a conductor-carrying current located in a magnetic field has a force exerted on it. In the illustration below, note the armature-conductors within the field of a two-pole motor. The commutator (not shown) makes the current flow away from you (into the page) in the right-hand group of conductors and toward you (out of the page) in the left-hand conductors. Because of the direction of current flow, the right-hand conductors have a downward force exerted upon them, and the left-hand conductors have an upward force exerted on them. As a result, the armature rotates in a clockwise direction.

The rotational force which turns the armature is called *torque.* Torque in a motor depends on two factors—the current flowing in the armature



DC Motor Principle

conductors, and the flux density of the field, or the total flux from each pole face. Mathematically, this relationship is expressed as,

$$T \propto Ia\ (armature\ current) \times \phi$$

When the conductors of the armature of a motor are being rotated so as to cut lines of force, a voltage which is called back emf is induced in them. It is called back emf because according to Lenz's law, the voltage induced in a conductor has a polarity so that the current arising from the induced voltage flows through the conductor in the direction opposite that of the current which induced the voltage. The amount which this back emf limits the armature current can be found by the equation,

$$I_a = \frac{V_t - E_c}{R_a}$$

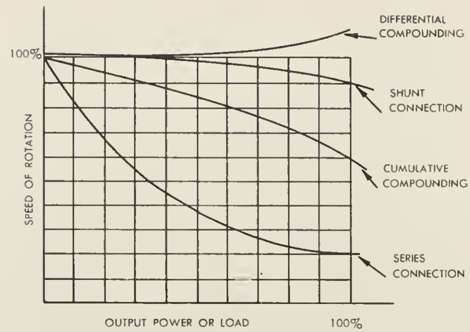
where V_t is the voltage applied to the armature terminals, E_c the counter emf, and R_a the resistance of the armature.

DC Motor Types and Characteristics

The circuits and characteristic of generators are also applicable to motors. This is true because a generator will run as a motor, and convert electrical energy into mechanical energy, provided the proper voltage is applied to the terminals in reverse. This voltage must be the same as that generated when the machine is rotated at the proper speed. DC motors, like DC generators, are classified as shunt, series, and compound motors. In addition there are two types of compound motors—cumulative and differential.

Shunt. In a shunt motor, the field windings are connected in parallel with the armature and have fairly high resistance. Since both armature voltage and field voltage are constant in shunt connections, the speed of a shunt motor is nearly constant for all values of load. A shunt motor is characterized by fairly high starting torque and current.

Series. Since the armature and field coil are connected in series in the series motor, the same current flows through both. This results in a widely variable field flux, a widely variable speed depending on the load, and a very high starting torque with moderately high starting current. Since field flux decreases as the load decreases, a series motor may run excessively fast under no load condition. Series motors are used in streetcars, elevators, and in automobiles as starting motors.



Load Characteristics of DC Motors

Cumulative-Compound. Cumulative-compound motors employ both series and shunt windings. The windings are connected so that the series winding aids the shunt field. Characteristics of this motor lie somewhere between those of a series and shunt machine. By properly proportioning the two fields, the speed of the motor can be made relatively constant under varying load conditions.

Differential-Compound. In a differential-compound motor the fields are arranged so that the series field opposes the main shunt field. This weakens the main field and tends to increase the speed of the motor as the load is increased. Differential compounding is not used extensively inasmuch as it produces instability, especially when the series field is very strong.

Relation of Speed and Load in DC Shunt Motors

When the load placed on a motor is very small, the motor speeds up until the induced armature voltage is almost exactly equal to the applied terminal voltage. It can never reach a speed where these two voltages are exactly equal, because the armature current will reach zero at that point. Even when a motor is carrying a heavy load, the back emf will be much greater than the voltage drop due to the resistance in armature. Hence, speed depends on the amount of flux in the main field, plus the back emf which the machine must produce.

Two methods are available for varying speed. They are the change of armature voltage and the control of main field flux.

Speed variation, by armature voltage control, is sometimes accomplished by placing a variable resistor in series with the armature. Increasing this resistance decreases armature voltage and motor speed. This method is not widely used because it increases losses and therefore lowers efficiency. The most common method of speed variation is control of the main field flux. From the relationship between speed, flux, and back emf, it can be proved that *decreasing* flux causes *increase* in speed, while *increasing flux* causes *decrease in speed.* While this may seem inconsistent at first, it is not difficult to explain. This is what happens. As explained before, torque, and hence speed, is proportional to the product of armature current and flux. Similarly, as discussed earlier, armature current is equal to $I_a = \frac{V_t - E_a}{R_c}$. Now since the back emf is proportional to the flux and represents by far the greater part of the terminal voltage V_t, I_a is affected much more by a small change in flux than the flux itself changes. This means that an increase, for example, of one percent in flux will result in a decrease of several percent in the armature current. Hence, the torque or speed is decreased by an increase of flux and vice versa.

Control of flux in a motor is normally obtained by placing a rheostat in series with the field coils. Increasing the resistance of this rheostat decreases flux and increases speed of the motor while decreasing series resistance increases flux and decreases speed.

TERMS USED IN AC

Previously you learned that the output voltage of an AC generator can be expressed in two ways. One is graphically by a sine wave as shown in the illustration at the bottom of this page. The other is by the equation,

$$e = E_m \sin \omega t$$

where e equals the voltage, E_m the maximum value of generated voltage, and ωt the angular velocity multiplied by the time.

When a generator produces an AC voltage, the current arising from it varies in step with the voltage. Like the voltage, the current can be represented graphically by a sine wave and algebraically by the following equation,

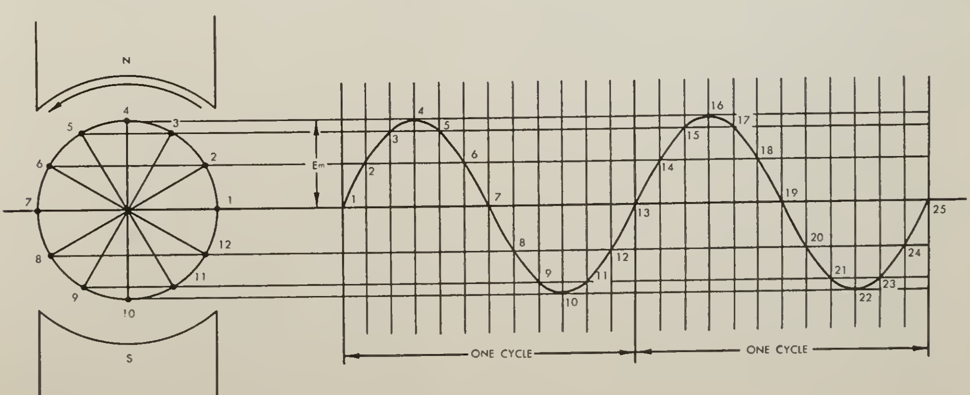
$$i = I_m \sin \omega t$$

where i equals the current, I_m the maximum value of generated current, and ωt the angular velocity multiplied by time.

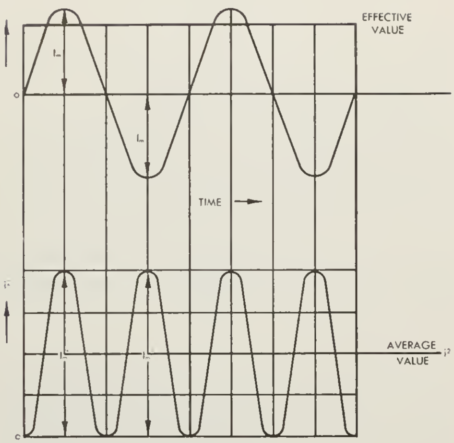
Frequency and Cycle

During the time the coil in a generator rotates through 360°, that is, one complete revolution, the output voltage goes through one complete *cycle.* During one cycle, the voltage increases from 0 to positive E_m in one direction, decreases to 0, increases in the opposite direction to negative E_m, and then decreases to 0. The first 180° (one-half of the voltage cycle) is called the *positive* alternation and the last 180°, from 180° to 360°, the *negative* alternation. The value of the E_m voltage at 90° is called the *amplitude* or *peak* voltage. The time required for a positive and a negative alternation is called the *period*, and the number of complete cycles per second is called the *frequency* of the sine wave. When the angular velocity ω at which the coil rotates is expressed in radians per second, the mathematical relation between ω and f is given by the equation

$$\omega = 2\pi f$$



Sine Wave of Voltage



Determining Effective Value of Current

Effective Value of AC

Since a sine wave of AC current (or voltage) varies continually between zero and maximum (or peak) values, first in one direction, then in the other, the question of numerical values arises. The value of AC most commonly used is *effective value.* The effective value of alternating current is the amount of alternating current which produces the same heating effect as an equal amount of direct current. In other words, one ampere (effective value) of AC will produce the same amount of heat in a given conductor in a given time as one ampere of DC. Since the heating effect of a quantity of current is proportional to the square of the current, it is possible to calculate the effective value of alternating current by squaring all the ordinates in a sine wave, taking the average of these values, and then extracting the square root. The effective value, thus being the root of the mean (average) square of the instantaneous currents, is also known as the *root-mean-square* or *RMS* value.

To understand the meaning of effective current, study the above diagram. In it, the instantaneous values of i are plotted in the upper curve and the corresponding values of i^2 are plotted in the lower curve. The i^2 curve is also a sine curve. It has twice the frequency of the i curve and varies above and below a new axis. The ordinate of this axis is the average of the i^2 values and the square root of this ordinate is the *RMS*, or effective value, of current. Since the i^2 curve varies uniformly from 0 to Im^2, its average value is $\frac{1}{2}$ Im^2. Thus the *RMS* value is $\sqrt{Im^2/2}$ or $\frac{Im\sqrt{2}}{2}$, which simplified is equal to 0.707 I_m. The *RMS*, or effective value, is commonly represented by *I*. Hence, the mathematical relation between effective and maximum (or peak values) of current can be expressed as $I = 0.707\ I_m$.

The relation between maximum (or peak) values of emf and effective emf is the same as the relation between peak and effective current. Thus,

$$E = 0.707\ E_m$$

The reciprocal of 0.707 is 1.414. Therefore,

$$E_m = 1.414\ E \text{ and}$$
$$I_m = 1.414\ I$$

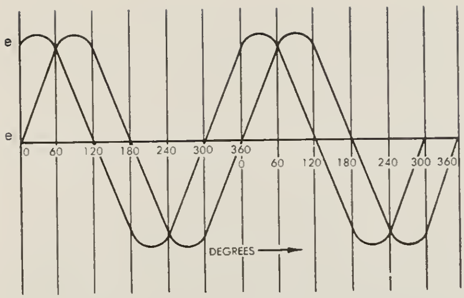
Average Value of AC

Another value in a sine wave sometimes used is the average value of the ordinates during the positive alternation. Average value is not as widely used as the *RMS* value, but in some instances is more a descriptive value of current or voltage. However, you can always assume that when an AC voltage is described by a value but is not designated *RMS*, peak or average, *RMS* value is meant. For example, the 110 volt AC which is used for lighting and other purposes in your home means 110 volts *RMS*. The mathematical relation between I_{av}, I_m, I_{av}, and I is given by the formula $I_{av} = 0.636 I_m = 0.9 I$. Similarly, the relation between E_{av}, E_m, E_{av}, and E is given by the formula $E_{av} = 0.636 E_m = 0.9 E$.

Phase

Phase is the fraction of a cycle that has elapsed since a voltage or a current has passed through a given value. Usually, this value is the zero value. Referring to the illustration on page 2-35 showing the sine wave of voltage and taking point 1 as the starting point or zero phase, you can see that the phase at point 2 is 30°, at point 3 it is 60°, at point 4 it is 90°, and so on throughout one complete cycle until point 13 where the phase is 360° or 0°.

A term more commonly used than phase is *phase difference.* Phase difference can be used to describe two voltages having the same frequency which pass through zero values at different instants. In the illustration marked phase relations, the degrees along the axis indicate the phases



Phase Relations

of the voltage e_1 and e_2 at any instant. At the 0° position, e_2 equals 0. On the other hand, e_1 passes through the zero value at 120° and 300°, both of which are 60° *ahead* of e_2 in time. Thus the voltage e_1 is said to *lead* e_2 by 60 electrical degrees. Another way of saying this is that e_2 *lags* e_1 by 60 electrical degrees.

The concept of phase difference is also used to compare two different currents or a current and a voltage. The expression *in phase* means that the phase difference between two currents or a current and a voltage is zero degrees, and *out of phase* means that their phase difference is some amount other than zero.

One fact to remember is that the phase angle of a voltage is not necessarily the same as the angle of rotation of the armature in a generator. If the generator has more than one pair of poles, the output voltage may go through a corresponding number of cycles for each revolution of the armature, depending on the number and placement of the brushes. This gives rise to the term *electrical degrees*, a term which is used to distinguish between phase angle and the angle of rotation of an armature.

AC CIRCUIT COMPONENTS

Resistance

Resistance, you recall from the previous discussion of DC, is the property by which a conductor opposes the flow of current in it. In alternating current circuits, the resistance of a conductor opposes alternating current in the same way that it opposes direct current. In fact, except at extremely high frequencies, the DC resistance of a given conductor is the same as its AC resistance. However, at ultra-high frequencies, current has a tendency to flow only in the surface of a conductor. This tendency is known as *skin effect* and causes the resistance of the conductor to increase. The higher the frequency, the more pronounced becomes skin effect. Because of it, many conductors of high frequency current in radar equipment are silver plated on the outside to reduce the resistance of the surface layer.

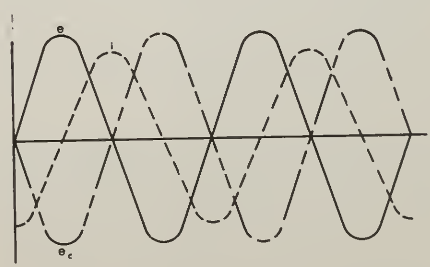
Inductance

In the discussion of induction you learned that a coil opposes any change in the current which flows through it by building up a counter emf and that this counter emf is an induced voltage, which is equal to $e_i = -L\frac{di}{dt}$, where e_i is the counter emf, L the inductance in henrys, di the change in current, and dt the change in time. The term $\frac{di}{dt}$ is the rate of change in current with respect to time.

In alternating current, the instantaneous value of i is equal to $I_m \sin \omega t$. If you substitute this value for i in the above equation and take the derivative, it becomes

$$e_i = -\omega L I_m \cos \omega t,$$

This is the equation for the instantaneous value of the alternating voltage. It is also the equation of a cosine curve, a curve which has the same shape as a sine wave curve but differs in phase from it by 90°. That this phase difference exists is apparent, if you recall that it is the *rate of change* of current that determines the value of the counter emf. The counter emf therefore becomes maximum not at the time of maximum current but at the time the current is changing most rapidly, that is, at the time when i is zero. By Lenz's law the counter emf is in such a direction as to oppose the change in current. Hence, if i is increasing, the counter emf will be in the opposite direction to the current.



Voltage and Current Relations in an Inductance

When i is decreasing the direction of the emf is the same as that of the current. The counter emf (e_c) lags the current (i) by 90 degrees.

In an AC circuit with both inductance and resistance, you can resolve the emf applied to the circuit into two components—one which sends the current i through the resistance of the circuit (the IR drop) and the other which overcomes the opposition of the counter emf (equal to e_c but *opposite in phase*). The component which overcomes e_c is represented by the sine wave e. This wave passes through zero (or maximum) *ahead* of the current and e is said to *lead* i by 90 degrees.

INDUCTIVE REACTANCE. Inductance in a circuit not only makes the current lag the voltage, but makes the magnitude of the current smaller. This choking down effect which results from the opposition to current flow caused by the inductance is known as *inductive reactance* and is given by the equation,

$$X_L = 2\pi f L$$

where X_L is the inductive reactance in ohms, f the frequency in cycles, and L the inductance in henrys.

Here is the derivation of the inductive reactance formula. In a circuit in which the resistance is completely ignored and in which the only opposition to current flow is the reactance of the inductance, the instantaneous voltage is equal to

$$e = -e_c = \omega L I_m \operatorname{Cos} \omega t$$

This equation is true for any value of ωt. Therefore, it is true when ωt equals zero degrees. At that instant the cos ωt equals 1 and e is at its maximum value E_m. Thus, by substituting these values in the equation, it becomes,

$$E_m = \omega L I_m$$

or by division,

$$\omega L = \frac{E_m}{I_m}$$

As $\frac{E_m}{I_m} = \frac{E}{I}$, ωL can be expressed as equal to

$$\omega L = \frac{E}{I}$$

This quantity has the characteristics of an opposition to current flow since it is equal to voltage divided by current. In this circuit, the opposition is the inductive reactance which is designated at X_L. As shown previously, $\omega = 2\pi f$. Therefore, the inductive reactance in the circuit is,

$$X_L = 2\pi f L$$

When inductances are connected in series and are not close enough to be in the magnetic field of each other, the inductances as well as their inductive reactances add like resistances connected in series. Thus, in a series circuit the sum of the inductive reactances is expressed by the equation,

$$X_{Lt} = X_{L1} + X_{L2} + X_{L3} + \text{— — — —}$$

and the sum of the inductances by the equation,

$$L_t = L_1 + L_2 + L_3 + \text{— — — —}$$

When inductances are connected in parallel, their inductances and the inductive reactances add by the sum of the reciprocals method, like resistances connected in parallel. Thus, in a parallel circuit, the sum of the inductive reactances is expressed by the equation,

$$X_{Lt} = \frac{1}{\frac{1}{X_{L1}} + \frac{1}{X_{L2}} + \frac{1}{X_{L3}} + \text{— — — —}}$$

and the sum of the inductances, by the equation,

$$L_t = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \text{— — — —}}$$

When mutual coupling exists between two inductances, the equations become somewhat more involved. For example, in the case of two inductances L_1 and L_2 with a mutual inductance of M, the total reactance is equal to,

$$X_t = X_{L1} + X_{L2} \pm X_{2m}$$

The plus or minus sign preceding X_{2m} is a necessary part of the equation since the two coils may be located in such a way that the fields either aid or oppose each other. When the fields are aiding, X_{2m} is preceded by the plus sign; when they are opposing, it is preceded by the negative sign.

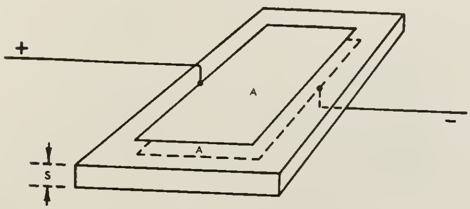
Capacitance

A condenser is a circuit element in which charges of electricity can be stored. It is composed of two conductors, usually flat plates separated by an insulator or dielectric. A condenser can be charged by connecting it to a source of DC such as a battery. Due to the battery voltage, electrons will be removed from one plate and will flow through the external circuit to the other plate. The charge built up results from the deficiency of electrons on one plate and the abundance of electrons on the other. After a period of time depending upon the capacity of the condenser and the resistance of the conductors, the condenser becomes fully charged, that is, the potential difference between the plates, equals the voltage of the battery. If a battery of higher voltage were connected, the condenser would take on an additional charge. For any given condenser, there is a definite ratio of charge to the voltage. Mathematically, it is expressed,

$$C = \frac{Q}{V}$$

where C is the capacitance of the condenser, V the voltage, and Q the charge in coulombs.

Capacitance is measured in farads. A condenser has a capacitance of one farad when one coulomb of electricity charges it one volt. For practical uses, the farad is much too large a unit. More commonly used, are the microfarad (μf) which is equal to one millionth of a farad, and the micromicrofarad ($\mu\mu$f) which is equal to one millionth-millionth of a farad.



Parallel Plate Condenser

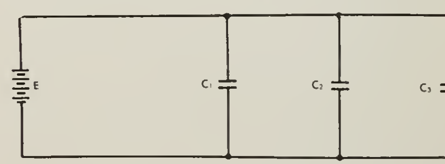
In its simplest form, a condenser consists of two parallel plates as above. The value of the capacitance of a condenser depends upon the area of the plates, the distance between them and the dielectric. A given charging voltage will, if the area of the condenser plates is increased, move an increased number of electrons to the negative plate before the voltage across the condenser equals the charging voltage. In this condition, the electrons spread over a wider area, and more therefore are required to produce a certain voltage. Whenever, the distance between the plates is increased, the repelling force between the positive plate and the negative plate is less, and fewer electrons are forced away from the positive plate and fewer are deposited on the negative plate. Accordingly, since there is less movement of electrons, the capacitance is less. Finally the dielectric determines the capacitance. The dielectric is rated according to its dielectric constant. The dielectric constant of air is taken as the basis and is equal to one.

In insulating materials, it is usually much higher. The higher the dielectric constant of the dielectric, other factors being equal, the greater the capacitance of the condenser. In the charging process of a condenser, the electron orbits in the atoms of the dielectric become distorted. The energy required to distort the electron orbits is supplied by the charging source. When this source is removed, the condenser starts to discharge, and the orbits spring back to their normal shape and release energy. Since this release of energy aids the normal discharge of the condenser, it represents an additional source of stored energy. The higher the dielectric constant, the greater the stored energy, for dielectrics with high constants are harder to distort than those with lower constants. Thus, capacitance varies with the dielectric. The relation between the capacitance, the area of the plates, the distance between the plates and the dielectric constant is expressed mathematically, as

$$C = 0.2246 \frac{KA}{S}$$

where C is the capacitance in micromicrofarads, K, the dielectric constant, A, the area of the plates in square inches, and S the distance between the plates in inches.

When condensers are connected in parallel, the total capacitance is the sum of the individual capacitances. Assuming each condenser in the illustration just below has a capacitance of 0.001 mf, then the total capacitance is 0.003 mf. This is easy to understand if you keep in mind that when



Condensers in Parallel

you connect condensers in parallel, there is an increase in plate area. Let Q_1, Q_2 and Q_3 represent the charge on C_1, C_2 and C_3 respectively. Since the condensers are connected in parallel, the same voltage E is across each.

Thus, if the charges of each condenser are added,

$$Q_t = Q_1 + Q_2 + Q_3$$

The charge of a condenser equals the charging voltage times the capacitance, or

$$Q = CE$$

Therefore, by substitution,

$$Q_t = C_1E + C_2E + C_3E$$

Similarly, $Q_t = C_tE$, *hence*

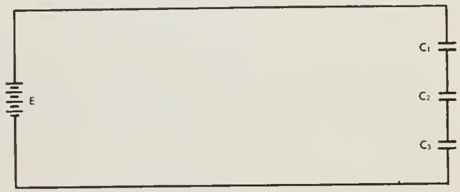
$$C_tE = C_1E + C_2E + C_3E$$

By dividing both sides by E, you get the formula for condensers in parallel,

$$C_t = C_1 + C_2 + C_3$$

Sometimes, condensers are connected in series to lessen the voltage across each. In a series circuit, the same current flows through all elements. Therefore, when condensers are connected in series; the charges on each condenser due to current flow are equal and the total charge can be expressed,

$$Q_t = Q_1 = Q_2 = Q_3$$



Condensers in Series

Since the voltage drop around a series circuit equals the applied voltage, the voltages across the condensers equal the applied voltage. Thus,

$$E_t = E_1 + {}_2E + E_3$$

Capacitances in series combine by the reciprocal of the sum of reciprocals like resistances or inductances in parallel.

$$\text{Thus } \frac{1}{C_t} = \frac{E_t}{Q_t} = \frac{E_1+E_2+E_3}{Q_t} = \frac{E_1}{Q_1} + \frac{E_2}{Q_2} + \frac{E_3}{Q_3}$$

$$\text{But, } \frac{E^1}{Q_1} = \frac{1}{C_1}, \frac{E^2}{Q_2} = \frac{1}{C_2}, \frac{E^3}{Q_3} = \frac{1}{C_3}$$

$$\text{Therefore, } \frac{1}{C_t} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

$$\text{From this, } C_t = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}}$$

The value of i at any instant during the period of charge of a condenser is the rate of charge and is equal to $i = \frac{dQ}{dt}$ where dQ equals the change in charge and dt the change in time.

The rate of charge is C times the rate of change of voltage, since Q equals CV and C is constant. Thus, the rate of charge is also equal to,

$$i = C\frac{dv}{dt}$$

Capacitive Reactance

When a condenser is connected in a circuit to which a source of alternating voltage is applied, the condenser will be charged first in one direction and then in the other and the electrons in the rest of the circuit will flow to and fro as an alternating current at the frequency of the applied alternating voltage. Although the current in the circuit appears to flow through the condenser, it actually does not flow through it at all. Yet the effect is the same as if current were passing through the condenser. An AC ammeter placed in the circuit would show a steady deflection just as if the condenser were not in the circuit. However, the amount of current flow is less when the condenser is in the circuit than when it is out of the circuit. This indicates that the condenser offers opposition to the flow of current. This opposition, which is called *capacitive reactance*, is given by the equation,

$$X_c = \frac{1}{2\pi fc}$$

where X_c is the capacitive reactance in ohms, f the frequency in cycles, and c the capacitance in farads.

This formula is derived in much the same way as the formula for inductive reactance.

As previously stated, the charging current in a capacitive circuit is equal to

$$i = c\frac{dv}{dt} \qquad (1)$$

As v is an alternating voltage, the voltage in the circuit is equal to,

$$v = v_m \sin \omega t \qquad (2)$$

Taking the derivative of v with respect to t produces the equation,

$$\frac{dv}{dt} = \omega\, v_m \cos \omega t \qquad (3)$$

This is the equation of a cosine wave which has the same waveshape as a sine wave except that it (cosine wave) leads by 90°.

Combining equations (1) and (3) produces the equation,

$$i = \omega c\, v_m \cos \omega t \qquad (4)$$

The current i is maximum when the cos ωt equals I. Thus by substituting in equation (4),

$$I_m = \omega c\, v_m \qquad (5)$$

or by division,

$$\frac{v_m}{I_m} = \frac{1}{\omega c} \qquad (6)$$

Since the ratio of the peak voltage to the peak current is equal to the ratio of the effective values,

$$\frac{v}{I} = \frac{1}{\omega c} \qquad (7)$$

The quantity $\frac{v}{I}$, the ratio of the voltage to current, represents the opposition to current flow in the circuit and is designated as X_c. The quantity ω equals $2\pi f$. Therefore, by substituting in equation (7), the capacitive reactance in the circuit equals,

$$X_c = \frac{1}{2\pi fc}$$

From the discussion, it is apparent that there is a difference in nature of inductive and capacitive reactances. Inductive reactance causes current to lag the voltage while capacitive reactance causes current to lead the voltage. This is another way of saying that an inductance opposes a change in current and a capacitance opposes a change in voltage. Still another difference is that inductive reactance is directly proportional to the inductance and frequency while capacitive reactance is inversely proportional to capacitance and frequency.

PRACTICAL COILS AND CONDENSERS

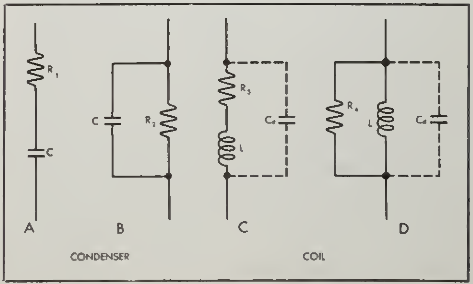
All electrical networks may be reduced to three circuit components—resistors, inductors and capacitors. In practice none of the three are perfect. All coils possess a certain amount of resistance, even though the material used in them is of the best quality. Not only do coils have some resistance, but they also have a certain amount of capacitance between the turns of wire, a condition which must be taken into consideration, especially at the higher frequencies. Similarly, resistors cannot be perfect, for they inherently possess a certain amount of inductance. Even straight wires have inductance, another condition which must be taken into account at the higher radio frequencies.

In the discussion of DC circuits, coils, or condensers were not mentioned. The reason for this is that a coil in a DC circuit behaves like a *resistor* and a condenser acts like an *open* circuit, *after* the current in the circuit has reached a *steady state*. For example, if a condenser, coil and resistor are connected in series and a steady source of voltage, such as furnished by a battery, is suddenly impressed on the combination by closing a switch, there is a relatively short period, called the *transient* period, in which a current does flow before coming to a stop. After the current stops, the condenser acts like an open circuit, and consequently there can be no further current flow.

In AC circuits, all three types of circuit elements have one thing in common; they all *oppose* the flow of current. This does not mean that all elements act alike, for each opposes current flow in a different manner. However, the *opposition* offered by each is measured in ohms and each can be treated graphically as vector quantities.

Another factor to consider in connection with resistors, coils, and condensers is that neither a perfect coil nor a condenser dissipates energy. A resistor is the only one of the three which does. It does so in the form of heat. A coil, to be sure, does store up energy in its magnetic field but gives this energy back to the circuit when the field collapses. The same is true of the condenser, except that in it the energy is stored in the dielectric field between the condenser plates. Nevertheless, there is no such thing as a perfect coil or condenser, for *practical coils and condensers dissipate energy* because of the *resistance* inherent in them.

For some purposes, the resistance of the coil or condenser must not be neglected; while for many other purposes, it may be completely ignored. In circuit analysis a practical condenser may be considered a perfect condenser in series with a small resistor, or in shunt with a large resistor. A practical coil may be considered a pure inductance in series with a low resistance, or in parallel with a high resistance. However, in the case of the coil, it is also necessary to consider the capacitance (distributed capacitance) which exists across its windings. The equivalent circuit illustrations directly below show (A) the equivalent circuit of an imperfect condenser with a perfect condenser in series with a small resistor R_1 and (B) an alternate equivalent circuit in which a perfect condenser is in parallel with the large resistor R_2. The equivalent circuits for a coil (C and D) show L in series with R_3 and shunted by C_d and L in parallel with R_4 and C_d. R_3 is small compared to R_4. C_d represents the distributed capacitance between windings.



Equivalent Circuits for Practical Condensers and Coils

Dielectrics

Impregnated paper, mica, oil, oxides, and air are the dielectrics most frequently used in condensers in radio and radar. Although a paper dielectric has the advantage of low cost, it is characterized by appreciable losses because of dielectric hysteresis when high frequency AC is applied to it. These losses are similar to magnetic hysteresis and manifest themselves in the form of heat generated within the dielectric itself. They cause imperfect condenser action with a consequent lowered efficiency.

On the other hand, air is free from dielectric hysteresis. Losses in air condensers are chiefly due to the solid dielectric material used in mounting the plates and to the charging currents, which cause a small I^2R loss in the plates.

The dielectric strength of practically all insulating materials diminishes with an increase in temperature so that the condenser may break down in continuous service under voltages which could be withstood indefinitely in intermittent service. A condenser with a wax-impregnated dielectric seems to undergo an aging process after a period of months, so that a breakdown finally occurs. Oil-impregnated paper is considered superior to wax-impregnated paper, particularly at high voltages.

Variable Condensers

Variable condensers, which employ air dielectrics, are used to adjust the resonant frequency of tuned circuits in radar and radio sets, where a continuously variable capacitance within limits, having small losses, is required. The angle of rotation of the shaft determines the capacitance of the condenser. By using semi-circular rotor and stator plates, it is possible to obtain a capacitance which is approximately proportional to the angle of rotation. Condensers using this type of construction are called straight-line capacity (SLC) condensers. Straight-line capacity condensers are rarely used in receivers designed for broadcast purposes since the various stations are separated by equal frequency intervals. Using SLC in broadcast receivers would result in crowding the higher frequency stations on the tuning dial of the condenser. The shape of the plates may be modified so that the capacitance is proportional to the square of the angle of rotation. When this is done, the condensers are called straight-line-wave (SLW) condensers.

The ideal condenser for broadcast reception is one in which the resonant frequency (with a given inductance) is proportional to the angle of rotation. This type condenser is called a straight-line-frequency (SLF) condenser. In it, the rotor plates are scimitar-shape, being rather long and pivoted near one end. Because of the shape of the rotor plates, the SLF condenser requires more space than other types, a fact which is considered a disadvantage.

Electrolytic Condenser

An electrolytic condenser operates by virtue of the fact that when two aluminum plates are immersed in a suitable electrolyte, such as a solution of ammonium borate or sodium phosphate, and connected to a source of direct current, a thin insulative film of oxide forms on the positive plate. This film acts as a dielectric, separating the two plates which form the condenser.

The thickness of the film in the electrolytic condenser depends on how large a voltage is used in the forming process, higher voltages causing thicker films. The film gradually disintegrates when the forming voltage is removed and forms again when the voltage is reapplied. During the reforming of the film, a large leakage current flows, but it soon drops to a normal value. For example, for a condenser with a working voltage of about 450 volts, the leakage current is about 200 microamperes per microfarad while the film is forming. Over a period of several hundred hours of continuous operation, the leakage current gradually drops to only a few microamperes per microfarad.

When an electrolytic condenser is operated for a considerable period of time at a voltage appreciably lower than the forming voltage, the thickness of the film decreases and the capacitance of the condenser rises. To prevent the film from completely disintegrating and ruining the condenser, it is necessary to observe polarity when connecting an electrolytic condenser into a circuit—that is, the positive plate of the condenser must be connected to the positive side of the circuit. This is because electrolytic condensers offer high resistance to current passing in one direction, but very low resistance to current in the other direction. This unidirectional-conduction characteristic makes them unusable in alternating current circuits. They are used chiefly in rectifier filter circuits where their comparatively large leakage losses are of little consequence.

Because of the extreme thinness of the film (dielectric), it is possible to get very high capacitance in only a fraction of the space required by paper dielectric condensers. Depending on the forming voltage, capacitances ranging from 1 to 1.5 mfd per square inch of area may be obtained with electrolytic condensers.

The maximum continuous working potential of electrolytic condensers is about 450 volts. Voltages greatly in excess of this value will puncture the film. However, on reducing the voltage, the film is restored. Thus, the condenser is said to be self-healing. Higher voltage operation is obtainable when two or more condensers are connected in series. However, it usually is necessary to shunt each unit with a high resistance to insure that the voltage divides equally across the condensers. Otherwise, the voltage will divide in accordance with the insulation resistance of the condensers and, while high in value, may differ appreciably in each condenser.

In recent years, electrolytic condensers have been developed which use a paste for the electrolyte. These condensers are called dry electrolytic condensers. In some types, a layer of gauze, or some other absorbent material, is saturated with the electrolyte and then rolled up between the two aluminum foils. Glycerin is usually mixed with the electrolyte because of its ability to absorb moisture.

Dry electrolytic condensers are very compact and are available in very large values of capacitance for low voltage use. For example, one commercial dry electrolytic, which is rated at 6000 mfd at a working voltage of 150 volts, has a volume of only 100 cubic inches. Those with working voltages of 450 volts have about a microfarad per cubic inch of volume, or about 15% of the space required by paper condensers for the same voltage. The life of a dry electrolytic condenser depends considerably upon the conditions of use.

Coils

Because of their property of inductance, coils of varied forms are employed in radio and radar. Inductance, remember, is the property of a circuit which causes it to oppose any change in current flowing through it. Mathematically, it is equal to the number of lines of flux linkage per ampere divided by the current producing the flux, that is,

$$L = \frac{N\phi}{I} \times 10^{-8} \text{ henries}$$

where N is the number of turns in the coil, and ϕ is the magnetic flux linking the coil and produced by the current I. The term 10^{-8} converts the inductance into henries.

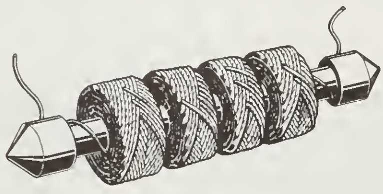
This formula applies only to coils using air cores. When a core, iron, for example, is inserted into the coil (or in cases where the permeability of a core in a coil is changed), the inductance of the coil changes. The inductance of a coil with a core is equal to

$$L = \mu Lo$$

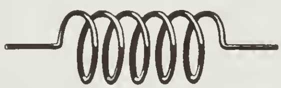
where μ is the permeability of the core and Lo is the inductance of the coil without the core.

In radio, the use of iron core coils is limited chiefly to audio frequency devices, such as interstage transformers and choke coils. Furthermore, their use is limited to frequencies below 150 kc unless especially prepared iron is used in the core. This special preparation consists of iron dust which is mixed with a suitable binder. The binder serves to insulate the particles of iron from each other. These special cores are common in radar receiver circuits.

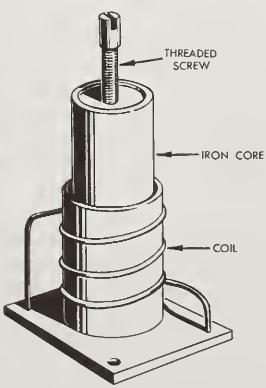
At high frequencies the chief difficulty with iron cores is increased eddy current losses. Eddy current losses in iron cores are proportional to the square of the frequency and to the square of thickness of the laminations in the core and become very great at high frequencies. These eddy currents in the core also induce currents of their own which act as a shield and prevent the main magnetic flux from penetrating very deeply into the laminations or particles of iron in the core and thus causes the effective permeability of the iron to decrease with increased frequency. This is what happens. In a particle of iron through which alternating current flows, there are eddy currents induced. These currents flow around the outer periphery of the iron particle in a direction in opposition to the flux which produced them. Since the demagnetizing effect of eddy current is greatest along the center line of the iron particle, the flux density there is minimum and the flux at the surface is maximum. Since the permeability of iron is the ratio of the flux density to the magnetizing force, this constitutes a reduction in the permeability of the iron.



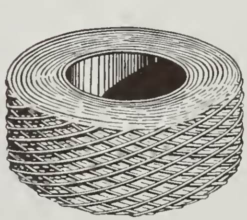
LOW FREQUENCY RF CHOKE COIL



HIGH FREQUENCY CHOKE COIL



PERMIABILITY TUNED COIL



HONEYCOMB COIL

Various Types of Coils

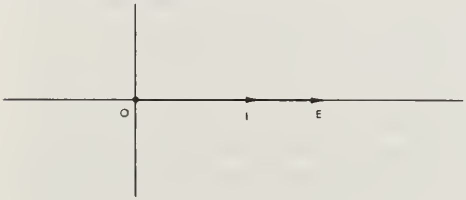
Eddy currents can be reduced by using sheet steel rolled thinner than 0.001 inch. Since it is very difficult to roll steel this thin, a still better method is to divide the core material (as explained previously) into fine particles and then mix the particles with an insulating compound. This effectively reduces the eddy current losses and at the same time increases flux penetration and therefore increases the effective permeability of the iron in the core.

Distributed capacitance. As pointed out previously, in any coil there is a small amount of capacitance between each turn and every other turn, and from each turn to ground. These individual capacitances, which are called stray capacitances, are equivalent to small condensers shunted across the terminals of the coil. Coils which are used in resonant circuits at the higher radio frequencies are nearly always single-layer coils, since they have low distributed capacitance at these frequencies. Multilayer coils are seldom employed at these frequencies because of high distributed capacitance unless they are *bank* or *honeycomb* wound. In a multilayer coil, wound in the ordinary fashion, the voltage between the first turn of one layer and the last turn of the next layer is quite high. This high voltage causes a large value of *displacement* current to flow through the capacitance between these turns and greatly increases the distributed capacitance of the coil. However, if the coil is bank wound, then the distributed capacitance will be considerably below that of a similar layer-wound coil having an equal number of turns. In the honeycomb-wound coil, the layers are wound in such a way that there is a space between adjacent turns and the wires of one layer cross those of the layer beneath at an angle. This reduces the area of contacts, and consequently lowers the capacitance between the layers. The distributed capacitance is less than would be obtained with the ordinary form of layer winding of similar dimensions. Radio-frequency choke coils commonly use honeycomb windings.

SERIES AC CIRCUITS

Resistance Only

When an AC voltage is applied to a series AC circuit containing resistance only, the current and voltage in the circuit are in phase. One method of showing the in-phase relationship is by sine waves. The current in the circuit is sinusoidal and will be maximum at the same time as the voltage is maximum. Another method of showing the same relationship is by a vector diagram. The usefulness of the vector method of representing phase of currents and voltages will become more apparent later in this manual. In the vector diagram just below, the current and voltage vectors coincide, showing that they are in phase, and differ only in amplitude. (Refer again to Chapter 1 for a discussion of vectors and their manipulation.)



Vector Representation of Current and Voltage in a Resistive Circuit

Since a resistor presents the same opposition to the flow of alternating current as it does to direct current, current flow in an AC circuit containing resistance only is equal to the voltage divided by the impedance. This relationship is expressed mathematically as,

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

where I equals the current in amperes, E the voltage, and Z the impedance in ohms.

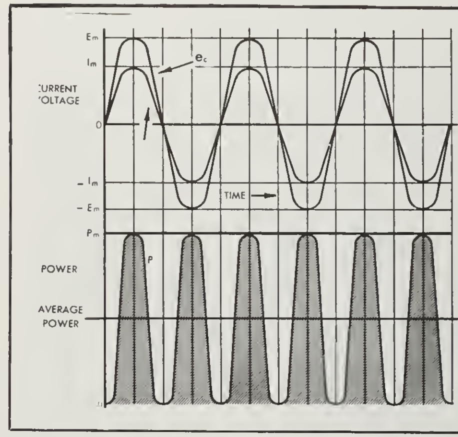
Impedance is a term embracing the total opposition to current flow in an AC circuit. In a circuit containing resistance only, resistance and impedance mean the same thing. In some AC circuits, impedance includes not only resistance but also reactance, both capacitive and inductive reactance. (The relation between Z, R and X_L or X_c is to be taken up shortly.)

The formula which gives the relation between current, voltage, and impedance in an AC circuit is called Ohm's law for AC. This formula is based on effective values, but is equally true for maximum, average, or instantaneous values of current and voltage.

Power in a resistive series AC circuit is equal to the product of the effective voltage by the effective current and is equal to,

$$P = EI$$

To understand how this formula is derived, refer to the following illustration showing the relation between the current and voltage and the instantaneous power resulting from them in a series resistive AC circuit.



Current, Voltage and Power in a Resistive Circuit

The power at any instant is the product of the current and voltage at that instant ($P = e \times i$). At the instant when $e = E_m$, i also is I_m and $P_m = E_m \times I_m$. When e is zero, i is zero and the power is zero. Then as e increases in a negative direction, i increases negatively and their product increases in a positive direction (the product of two negative numbers is positive). Thus P varies between $P_m = E_m \times I_m$ and *zero* at a frequency twice that of the voltage current. The line drawn through the center of the power sine wave shows the average power. To determine the average power mathematically, it is necessary to perform the following operation:

$$P_{av} = \frac{P_m}{2} = \frac{E_m I_m}{2} = \frac{I_m}{\sqrt{2}} \times \frac{E_m}{\sqrt{2}}$$

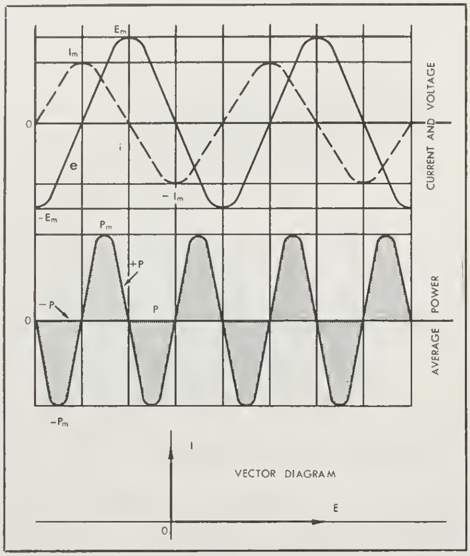
The values represented by $\frac{E_m}{\sqrt{2}}$ and $\frac{I_m}{\sqrt{2}}$, are the effective values of voltages and current respectively and when multiplied together give the average power in a resistive circuit, which as stated previously is,

$$P = E \times I$$

and means the power being consumed in the form of heat as all is in the positive direction.

Capacitance Only

When an AC voltage is applied to a series AC circuit containing capacitance only, the capacitor opposes the flow of current in such a way that current leads the voltage by 90°, as you can see by the sine curves and the vector diagrams on the next page. At each point of the sine waves for current and voltage, the current is 90°



Current, Voltage and Power in a Capacitive Circuit

ahead of the current. In the vectorial representation, the voltage E is the reference vector. I is drawn upward at 0, which in vector representations is called the leading position. It could have been drawn downward at 90°. However, the practice is to draw the leading component upward. One point to notice here which has a bearing on power, as you will see later, is that the product of E and I is alternately positive and negative.

Current leads voltage in a capacitive circuit because the capacity of the capacitor, as you saw earlier, opposes any change in voltage. The opposition which the capacitor offers to the flow of current is called capacitive reactance. It is indicated as X_c and like resistance is measured in ohms by the Ohm's law formula,

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

where I equals the current in amperes, E the emf in volts and X_c the capacitive reactance in ohms.

One point to notice here that the impedance Z in a capacitive circuit is represented by X_c, the capacitive reactance. Thus, so far, you have seen that impedance in an AC circuit includes resistance and capacitive reactance.

In resistive circuits you learned that the power consumed in the circuit is expressed by the equation, $P = E \times I$. This power is the true power in the circuit and is the power in any circuit in which the current and voltage are in phase. For that matter, the formula is the same power formula you applied earlier in DC circuits. However, in AC circuits in which current and voltage are not in phase, as is the case when there is capacitance (or inductance) in the circuit, the true power is not equal to the product of $E \times I$. What happens is this. The reactive parts of the circuits, either the capacitance or the inductance or both, return some energy back to the circuit instead of converting it to heat. Therefore, power in a reactive type circuit depends not only on the voltage and current but also upon the reactive components of the circuit. In a circuit containing a pure capacitance, that is, one in which there is no resistance at all, no power is consumed.

To understand how capacitance affects power in a reactive circuit, refer to the power diagram in the illustration. When P in the power diagram is positive it means that the condenser is taking power from the circuit and when P is negative, the condenser is giving power back to the circuit. Since the positive power lobes are symmetrical with the negative power lobes, the power returned to the circuit is equivalent to the power taken from it and the average power is zero. In other words, a pure capacitance does not consume power.

Inductance Only

In a series AC circuit containing only inductance, the current and voltage, like those in a capacitive circuit, are not in phase. Because the inductance opposes any change in current, the voltage across the inductance leads the current by 90°—that is, the voltage is maximum 90° ahead of the current.

By reversing E and I in vector diagram of the capacitive circuit, you get the vector representation of the inductive circuit. Power in an inductive circuit resulting from the out-of-phase relationship between the current and voltage, like that in a capacitive circuit, is also zero, since power is delivered back to the circuit by the inductance, assuming the circuit is purely inductive.

Circuit Containing Resistance and Inductance

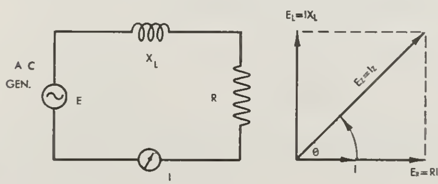
In a series AC circuit which contains both resistance and inductance, the voltage across the inductance leads the current by 90% and the voltage across the resistance is in phase with the current. Because of the phase difference caused by the inductive component, the total impedance in the circuit is not equal to the arithmetic sum of the ohmic values of the resistance and the inductance, but is equal to their vector sums and is expressed by the formula,

$$Z = \sqrt{R^2 + X_L^2}$$

Similarly, the total voltage drop in the circuit is the vector sum of the individual voltage drops and is equal to

$$E = \sqrt{E^2_R + E_L^2}$$

To understand how these formulas are derived, it is necessary to establish certain geometrical relationships by drawing a vector diagram involving the circuit components. Directly below is a circuit with a resistance R in series with an inductive reactance, and a vector diagram showing the voltage relationship of the circuit components. In the vector diagram, the current I is the reference vector or the 0° line.



Circuit and Vector Diagram with Inductance and Resistance

Since the voltage across R is in phase with the current, it is drawn to coincide with I. Since the voltage across the inductance is equal to $I \times X_L$ and since it leads the current I by 90°, it is drawn perpendicular to E_R. The resultant of these two emf's in series, according to the parallelogram of forces method, is their vector sum. The diagonal E_z of the parallelogram is the total voltage drop in the circuit and, by Kirchoff's laws, is equal to the applied voltage E. The magnitude of E can be obtained from the right triangle whose two legs are E_R and E_L and hypotenuse E_z is equal to E. Thus,

$$E = \sqrt{E_R^2 + E_L^2}$$

Since the voltage drop across E_R is equal to IR and the drop across L is equal to IX_L, it is evident that the total voltage drop is given by the equation,

$$E = \sqrt{(IR)^2 + (IX_L)^2}$$

By factoring,

$$E = \sqrt{I^2 \, (R^2 + X_L^2)}$$

By taking root,

$$E = I\sqrt{R^2 + X_L^2}$$

But since $E = IZ$

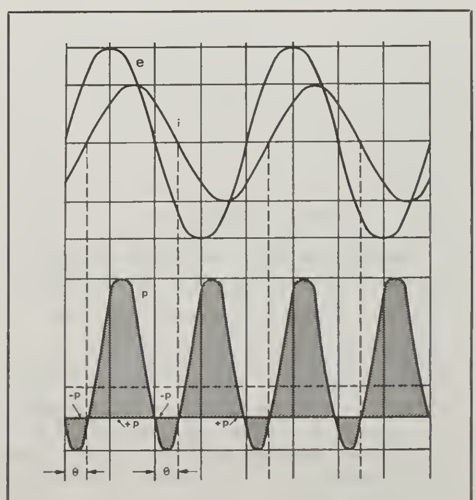
Therefore, the total impedance Z is

$$Z = \sqrt{R^2 + X_L^2}$$

In determining a vector quantity, it is necessary to consider phase as well as magnitude. Therefore, in addition to magnitude, the usual designation of voltage (or impedance) includes the phase angle θ between it and the quantity compared with it. One method of expressing the phase angle vectorially is in the polar form. By this method the applied voltage in a circuit is represented as $E\underline{/\theta^\circ}$ and the impedance as $Z\underline{/\theta^\circ}$.

A vector can also be expressed in the rectangular form by giving it horizontal and vertical components. From Chapter 1, you recall that you designated the horizontal component as the real component and the vertical component as the imaginary component and that in order to indicate the 90° phase difference between them, you used the notation j. Thus, instead of the polar form $E\underline{/\theta^\circ}$, you can write the applied voltage in rectangular form as $E_R + jE_L$, and instead of the polar form $Z\underline{/\theta^\circ}$, you can write the total impedance in the rectangular form $R + jX_L$. These forms do not imply that the terms jE_L and jX_L are only imaginary, for they are as real as E_R and R. This system of notation is merely a convenient method for handling these quantities.

The illustration just below shows the current, voltage and power relations in an AC circuit which contains resistance and inductance con-



Current, Voltage and Power Relations

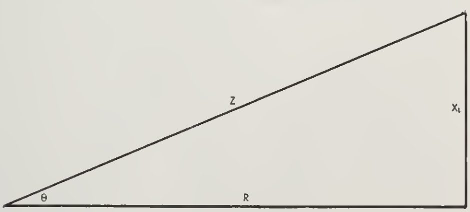
nected in series. Since the voltage leads the current by θ degrees, there is a period of θ degrees during each power cycle where the product of e and i is negative—that is, when power is being returned to the circuit. Note that then the negative power lobe is considerably smaller than the positive lobe. Hence, you see that the true power is neither $E \times I$ nor zero. It is not equal to zero, for you can see that since the negative power lobe is smaller than the positive lobe, the average of the two is not zero. It is not equal to $E \times I$, for the product of these quantities is not equal to I^2R, which is the power dissipated by the resistor. Remember the only element that actually consumes power in a circuit is resistance. In this circuit the product of E and I is the apparent power. The ratio of true power to apparent power is called the power factor and is expressed as,

$$p.f. = \frac{P \quad (\text{true power})}{ExI \ (\text{apparent power})}$$

Since $E = IZ$ and the true power P is equal to $P = I^2R$, the power factor ratio can also be written,

$$p.f. = \frac{I^2R}{I^2Z} = \frac{R}{Z}$$

Earlier you saw that it was possible to express the relation of the quantities Z, X_L, R by the expression, $Z = R + jX_L$. In a graphical representation of this relationship, which is called an *impedance triangle*, you can see that $R = Z \cos \theta$.



The Impedance Triangle

On substituting this value for R in the power factor formula, you can obtain the formula p.f. = $\frac{Z \cos \theta}{Z} = \cos \theta$. This formula stated in words is as follows: The power factor of the circuit is the cosine of the phase angle between the current and voltage in the circuit. Thus, the power factor is a number, the minimum value of which is zero and the maximum value of which is one. In terms of this definition, the power in a reactive circuit then is equal to,

$$P = EI \cos \theta$$

Circuit Containing Resistance and Capacitance

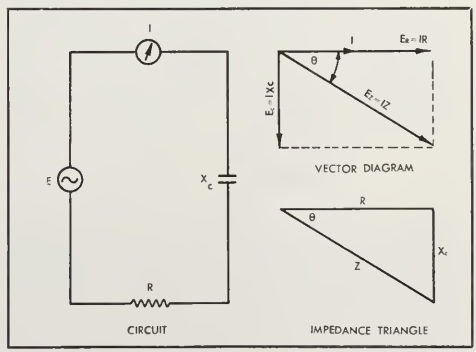
In a circuit containing resistance and capacitance, the ohmic values of the resistance and the reactance of the capacitance, like the values of the resistance and the reactance in a resistive inductive circuit, do not add arithmetically but vectorially. In a circuit containing resistance and capacitance connected in series, the total opposition to current flow is equal to

$$Z = \sqrt{R^2 + X_c^2}$$

and the effective voltage resulting from the out of phase relationship is equal to

$$E = \sqrt{E_R^2 + E_c^2}$$

Both of these formulas are easily proved by the impedance triangle and the parallelogram force method of vector representation, both of which are shown in the illustration just below. The method is similar to that employed in resistance-inductive circuits.



Resistance and Capacitance in Series AC Circuit

When the vector quantities just illustrated are referred to a system of coordinates, the effective voltage in a resistive-capacitive circuit can be expressed as,

$$E_Z\underline{/-\theta^\circ} = E_R - jE_c$$

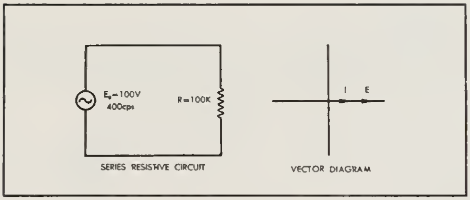
and the total impedance as,

$$Z\underline{/-\theta} = R - jX_c$$

For finding power in resistive-capacitive circuits, use the power formulas developed in the resistive-inductive circuits. This is possible since the cosine of a negative angle is equal to the cosine of the corresponding positive angle.

Examples

To understand the principles discussed in connection with AC circuits, study the solution of the following circuit problems.



Problem 1

In the above circuit, using the values indicated, find the current flow and power dissipated across the resistor.

Solution:

Since the load in the circuit is a resistance, the current flow is in phase with the voltage output of the generator.

By Ohm's law, the current is,

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

$$I = \frac{100}{100 \times 10^3}$$

$$I = 1 \text{ ma}$$

To find the power dissipated across the resistor, substitute in the power formula,

$$P = EI \cos\theta$$

$$P = (100)\ (.001)\ \cos\theta°$$

Since in a resistive circuit the phase angle equals 0°

$$\text{Cos } \theta = \text{Cos } 0°$$

From tables

$$\text{Cos } 0° = 1$$

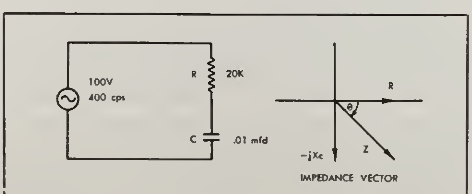
Therefore, the power is,

$$P = (100)\ (.001)\ 1$$

$$P = .1 \text{ watt } (p.f = \cos\theta = \cos 0° = 1.0)$$

Problem 2

When a condenser or an inductor is placed in series with a resistor in an AC circuit, the current is not in phase with the applied voltage. In the circuit below, the reactance of the .01 condenser at 400 cycles is $\frac{1}{2\pi fC}$ or $39.8\ K$, and the impedance of the load Z is $20\ K - j39.8K$. Find the current and voltage across $R(E_R)$ and the voltage across $C(E_C)$.



Solution:

You can solve for the current either by the rectangular form or the polar form of impedance. Method 1 uses the rectangular form and method 2 the polar form.

Method I

$$I = E/Z$$

$$= \frac{100}{20{,}000 - j39{,}800}$$

$$= \frac{100\ (20{,}000 + j39{,}800)}{(20{,}000 - j39{,}800)\ (20{,}000 + j39{,}800)}$$

$$= \frac{(2 + j3.98)\ 10^6}{4 \times 10^8 + j7.95 \times 10^8 - j7.95 \times 10^8 - j^2 15.8 \times 10^8}$$

$$= \frac{(2 + j3.98)\ 10^6}{19.8 \times 10^8}$$

$$= \frac{(2 + j3.98)\ 10^{-2}}{19.8}$$

$$I = (.1005 + j.205)\ 10^{-2}$$

The current I is,

$$I = 2.28\ \underline{/63.9°} \text{ milliamperes}$$

Method II

$$Z = 20K - j39.8K = 44.5K\ \underline{/-63.3°}$$

$$I = E/Z$$

$$I = \frac{100}{44.5K\ \underline{/-63.3°}}$$

The current I is,

$$I = 2.25\ \underline{/63.3°} \text{ milliamperes}$$

On comparing the two methods, it is obvious that the solution by polar form is much simpler than the solution by rectangular form. For this reason, the polar form is used in this manual, except in cases where vectors are to be added or subtracted.

From the value of current obtained, ($I = 2.25\ \underline{/63.3°}$), you can find the voltage across the resistor as follows:

$$I = 2.25\ \underline{/63.3°} \text{ milliamperes}$$

$$I = 2.25\ \underline{/63.3°} \times 10^{-3} \text{ amperes}$$

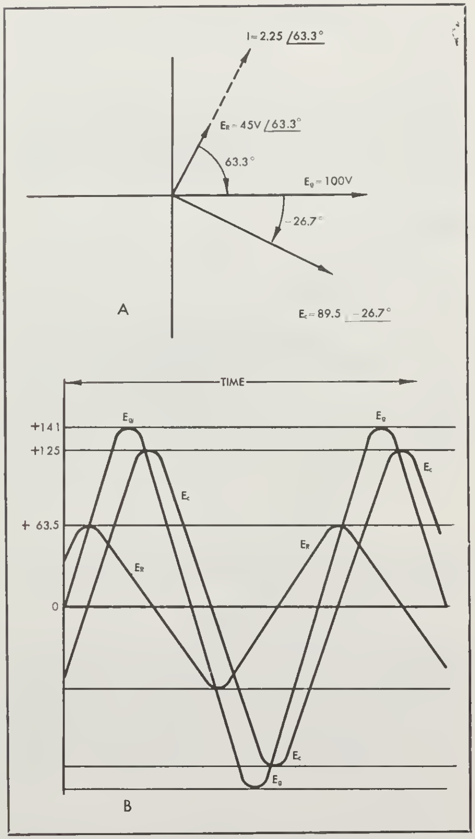
$$R = 20K = 20{,}000 \text{ ohms}$$

$$E_R = IR = 2.25\ \underline{/63.3°} \times 10^{-3} \times 20{,}000$$

Hence, the voltage across R is,

$$E_R = 45\ \underline{/63.3°} \text{ volts.}$$

The answer means that the voltage across the resistor is in phase with the current through the resistor, but leads the applied voltage by 63.3°.



Graphical and Vectorial Results of Circuit Components

To find the voltage across the condenser, use Ohm's law and solve as follows:

$E_C = IX_C$

$= 2.25 \angle 63.3° \text{ ma} \times 39.8 \angle -90° K$

The voltage across the condenser is,

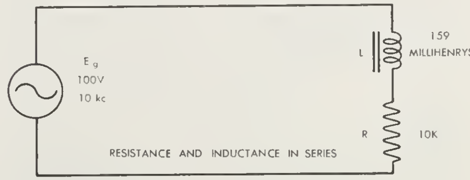
$E_C = 89.5 \angle -26.7° \text{ volts } (I_c = 2.25 \angle 63.3°)$

Diagram A directly above shows the relationship of the circuit components vectorially, and diagram B shows them graphically.

The apparent power delivered to the circuit is I^2Z or 224 milliwatts, while the true power consumed is I^2R or 101 milliwatts. The power factor $\frac{tp}{ap}$ is equal to 101/224 or .451. By the power formula $p.f. = \cos \theta$, it is equal to cos 63.3° or .449. The difference in results is due to the different mathematical approach.

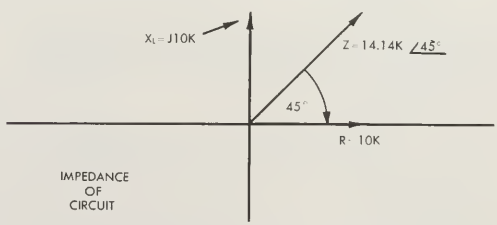
Problem 3.

In the circuit just below, using the values indicated, find the voltage (E_L) across the inductor and the voltage (E_R) across the resistor.



Solution:

The current which flows through a pure inductor lags the voltage across it by 90°. Further, the reactance of the inductor is directly proportional to the frequency and is equal to $2\pi fL$. Thus, by using the values in the circuit, the impedance of the inductor is $2\pi \times .159 \times 10{,}000$ or 10,000 ohms, and the impedance of the circuit (see the vector diagram just below) is equal to $10{,}000 + j10{,}000$ or $14.14 \times 10^3 \angle 45°$.



First find the current flow as follows:

$I = E/Z$

$= \dfrac{100 \angle 0° \text{ volts}}{14.14 \angle 45° \text{ kilohms}}$

$= 7.07 \angle -45° \text{ milliamperes}$

By Ohm's law,

$E_L = IX_L$

$= 7.07 \angle -45° \times 10^{-3} \times 10 \angle 90° \times 10^3$

The voltage across L is,

$E_L = 70.7 \angle 45° \text{ volts}$

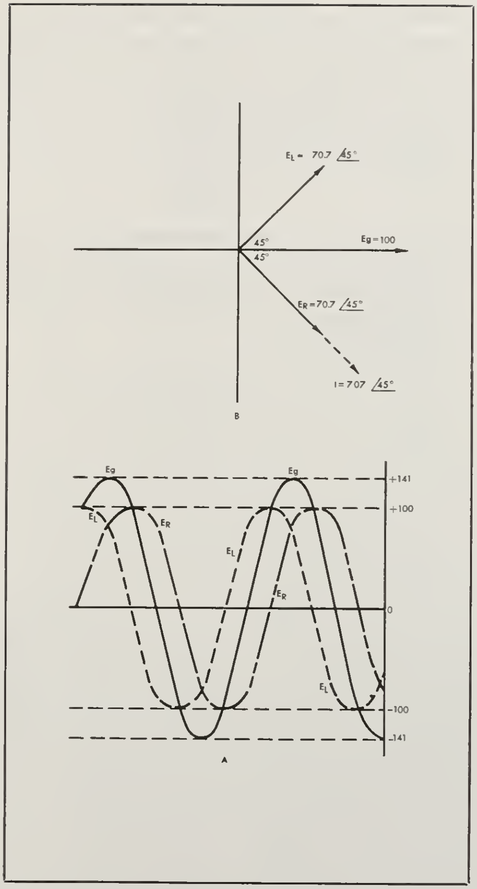
$E_R = IR = 7.07 \angle -45° \times 10^{-3} \times 10 \times 10^3$

The voltage across R is,

$E_r = 70.7 \angle -45° \text{ volts.}$

Diagram A following shows a graphical representation of the quantities just calculated and diagram B shows them vectorially.

From the preceding calculations and discussion, it is apparent that the voltage across the inductor leads the generator voltage by 45°, and the voltage across the resistor lags the generator



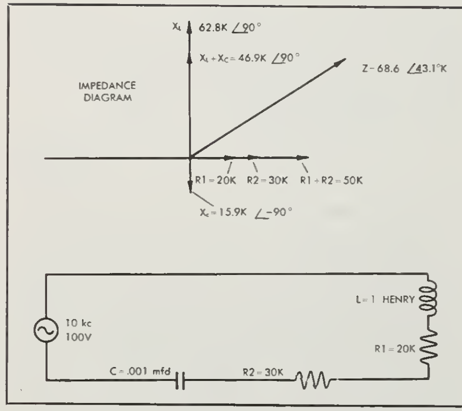
Graphical and Vectorial Results of Circuit Quantities

voltage by 45°. Notice also that the voltage across the resistor lags the inductor voltage by 90°, and the total current flow through the circuit lags the applied voltage by 45°.

Problem 4

In all cases of series AC circuits, the impedance into which the generator is working is equal to the vector sum of the reactances and resistances around the loop. To understand the mathematics involved, study the following problem and the impedance diagram.

Assume a generator with an output of 100 v at 10 kc is working into a load in the following circuit, find the impedance into which the generator works.



R, L, and C Series Circuit and Impedance Diagram

Solution:

$X_L = 62{,}800$ *ohms*

$X_C = 15{,}900$ *ohms*

The total impedance (Z) of the circuit is equal to,

$Z = X_L + R_1 + R_2 + X_C$

Substituting,

$Z = [j62.8 + 20 + 30 + (-j15.9)] \times 10^3$

The total impedance is,

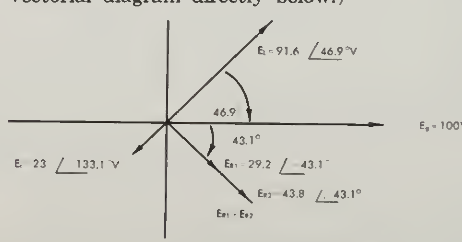
$Z = (50 + j46.9) \times 10^3$ *ohms*

In polar form Z is,

$Z = 68.6 \times 10^3 \angle 43.1^\circ$ *ohms.*

This circuit is said to be inductive, since the inductive reactance in it is greater than the capacitive reactance. In this type circuit, you can expect to find the following conditions: *First,* the current will lag the generator voltage; *second,* the voltage across the condenser will be 180° out of phase with the voltage across the inductor; and *third,* the voltage across the resistors will lag the generator voltage by the same angle as the current.

The following gives the solution of the remaining circuit components in the circuit. (Notice that the values are shown vectorially in the vectorial diagram directly below.)



Vector Diagram of Circuit Quantities

$$\text{The current, } I = \frac{100}{(68.6 \angle 43.1^\circ)\ 10^3} = 1.46 \angle -43.1^\circ \text{ ma}$$

$$E_C = 1.46 \times 10^{-3} \angle -43.1^\circ \times 15.9 \times 10^3 \angle -90^\circ = 23.2 \angle -133.1^\circ \text{ volts}$$

$$E_L = 1.46 \times 10^{-3} \angle -43.1^\circ \times 62.8 \times 10^3 \angle 90^\circ = 91.6 \angle 46.9^\circ \text{ volts}$$

$$E_{R1} = 1.46 \times 10^{-3} \angle -43.1^\circ \times 20 \times 10^3 = 29.2 \angle -43.1^\circ \text{ volts}$$

$$R_{R2} = 1.46 \times 10^{-3} \angle -43.1^\circ \times 30 \times 10^3 = 43.8 \angle -43.1^\circ \text{ volts}$$

Series Resonance

As explained in the foregoing discussion and examples, you saw that in a series circuit containing both inductive and capacitive reactances, the effective reactance is their difference and the circuit may be capacitive or inductive depending upon which is the larger. This fact expressed mathematically is $Z = R + j(X_L - X_C)$. Now since X_C decreases and X_L increases with an increase in frequency, there is a frequency at which a given coil and condenser will have equal reactances. This frequency is called the resonant frequency. The condition that exists in the series circuit when $X_C = X_L$ is known as series resonance. At series resonance the j term drops out of the impedance formula and the impedance is the resistance only. Since there is zero reactance, the current is in phase with the voltage and the power factor is one. Furthermore, since the only impedance is the resistance, the current becomes maximum at resonance.

In a series resonant circuit, the frequency at which the circuit is resonant is,

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where f is the frequency in cycles, L the inductance in henrys, and C the capacitance in farads.

The following is the derivation of the series resonance formula:

At series resonance,

$$X_L = X_C$$

But $X_L = 2\pi f L$ *and* $X_C = \dfrac{1}{2\pi f C}$

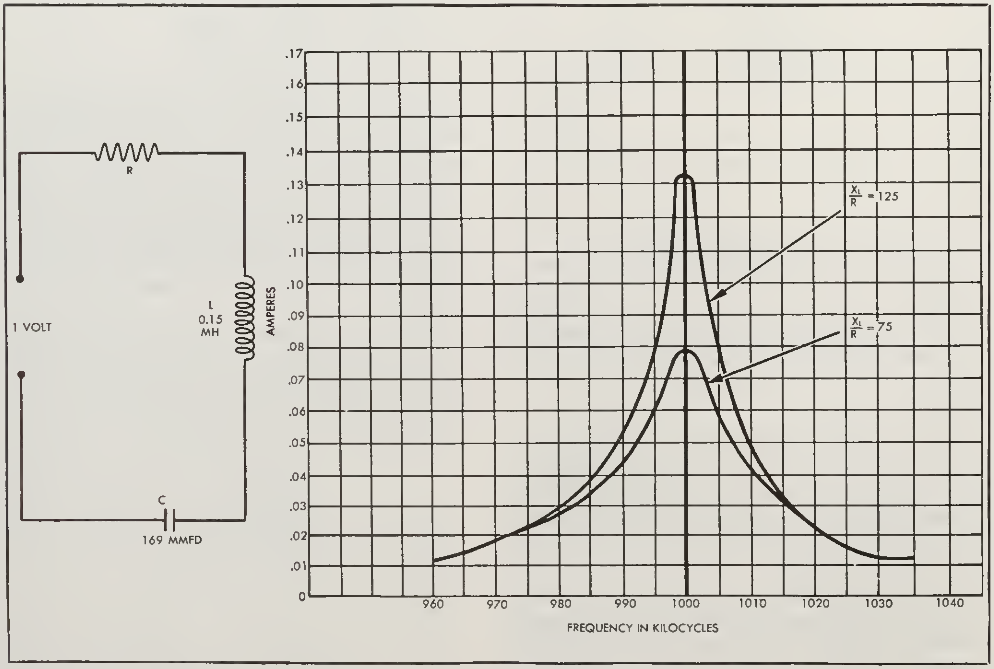
Therefore $2\pi f L = \dfrac{1}{2\pi f C}$

or

$$f^2 = \frac{1}{4\pi^2 LC}$$

Taking the square root,

$$f = \frac{1}{2\pi\sqrt{LC}}$$



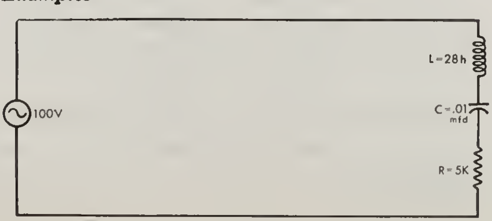
Series Resonant Circuit

At frequencies above resonance the inductive reactance in a circuit is greater than the capacitive reactance and the circuit is said to be *inductive.* Below resonance, the capacitive reactance is greater than the inductive reactance and the circuit is *capacitive.* The preceding illustration shows a circuit which is resonant at 1000 kilocycles, and the graph of the current versus frequency of the applied voltage. Below 1000 kc the capacitive reactance increases and the current drops off rapidly. Above resonance, the current decreases due to the increase in inductive reactance. In practical circuits the resistance R is low. Although it appears as a separate component in the circuit, it really represents the resistance of the inductor and the connector leads. Therefore very high currents flow at resonance in a series resonant circuit. At frequencies which differ appreciably from the resonant frequency, current flow is practically independent of the resistance, being determined chiefly by the reactance.

In the series resonant circuit shown in the illustration at the bottom of the preceding page, the reactances X_L and X_C are each equal to 940 ohms and the current flow is approximately 0.133 amperes. Hence, the voltage across both E_L and E_C is equal to 125 volts. The drop across the resistance R is only one volt, the applied voltage in the circuit. The voltage E_C and E_L are 180° out of phase and cancel each other when adding the voltages in the circuit. The ratio of E_L (or E_C) to E_R is 125 and is known as the Q of the circuit. Since R represents the resistance in the circuit, being mostly that of the coil, you will usually find the term Q applied to the coil in a circuit. The Q of a coil is the ratio of X_L to R at resonance. The second curve shown is for a lower Q circuit. The Q in it is 75. The lower Q of the second curve is due to a greater R in the circuit.

SERIES RESONANCE CIRCUIT PROBLEMS. To understand the principles of series resonance, study the following examples.

Examples



Problem 1. In the circuit just shown E_g is equal to 100 volts. Using the values indicated, (a) find the resonant frequency (f_r), and I, E_L, E_C, and E_R, at resonance, and (b) show the magnitude of I, E_L, E_C, and E_R both graphically and vectorially.

Solution:

a. *Solving for the values required*

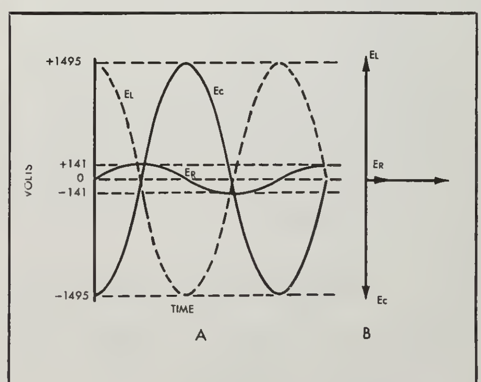
1. *Substituting the values given for this problem in the equation,* $fr = \dfrac{1}{2\pi\sqrt{LC}}$

$$fr = \frac{.159}{\sqrt{28\times.01\times10^{-6}}} = \frac{.159\times10^{4}}{\sqrt{28}} = 300 \text{ cps}$$

2. *By previous formulas,*

$$X_L = 53K$$
$$X_c = 53K$$
$$Z = 5K + j53K - j53K \text{ or } 5K$$
$$Z = 5K \angle 0^\circ$$
$$I = \frac{100}{5K\angle 0^\circ} = 20 \text{ ma } \angle 0^\circ$$
$$E_C = (53\times10^3 \angle -90^\circ)\times20\times10^{-3} \angle 0^\circ$$
$$= 1060 \angle -90^\circ \text{ volts}$$
$$E_L = (53\times10^3 \angle 90^\circ)\times20\times10^{-3} \angle 0^\circ$$
$$= 1060 \angle 90^\circ \text{ volts}$$
$$E_R = 5\times10^3\times20\times10^{-3} \angle 0^\circ = 100 \angle 0^\circ \text{ volts}$$

b. *Showing graph and vector diagram.*

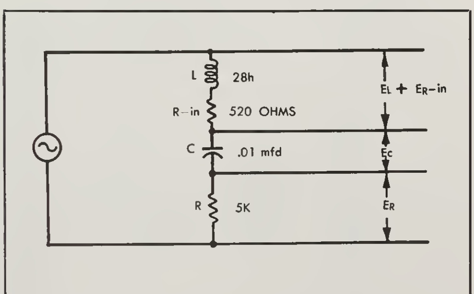


Graph and Vector Diagram for Problem 1

Draw the graph and vector diagram as shown in the above illustration. From the graph of the voltages across the circuit elements at resonance at A, you can see that since E_c and E_L are equal and 180° out of phase; at any time t, $E_L + E_c = 0$. Another point to notice is that E_c and E_L are much greater than the applied voltage. This is significant since a condenser with a peak voltage rating of 141 volts would break down if it were placed in this circuit, even though the applied voltage were only 100 volts. A further consideration is that the voltage across the resistance in this circuit is equal to the applied voltage.

Problem 2. In the first problem it was assumed that the inductor had no resistance at all. This is not true, for the wire with which the conductor is wound always has some resistance. For accurate computations it is necessary to consider the ohmic resistance of the wire itself in an inductor. It is considered as in series with the inductor and is handled in the same manner as an external resistance. When this resistance is considered, it is impossible to measure the inductor voltage (E_L), since the voltage across the inductor is actually $E_{R\text{-}IN} + {}_jE_L$, where $E_{R\text{-}IN}$ is the voltage drop that results from the internal resistance of L.

In understanding more completely the effect of the internal resistance of a coil on circuit operation, solve for the values required in the previous problem when the coil is considered to have an internal resistance of 520 ohms as illustrated directly below.



Circuit with R of Coil Included

Solution: From the previous problem,

$$f_r = 300 \text{ cps}$$
$$X_L = 53K \text{ ohms}$$
$$X_c = 53K \text{ ohms}$$

Including the internal resistance of the coil, the impedance of this circuit is,

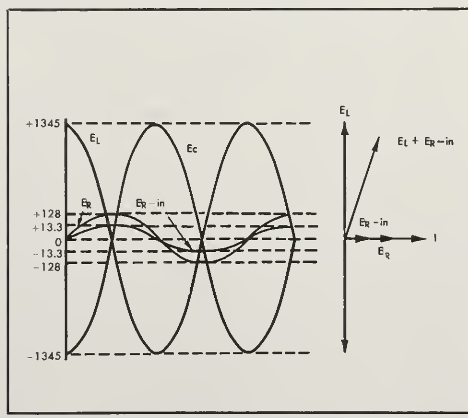
$$Z = j53K + 0.520K - j53K + 5K = 5.52K$$

Then finding the other required values at resonant circuit condition,

$$I = \frac{100}{5.52K} = 18 \text{ ma}$$
$$E_L = 954 \angle 90° \text{ volts}$$
$$E_c = 954 \angle -90° \text{ volts}$$
$$E_{R\text{-}IN} = 9.36 \text{ volts}$$
$$E_R = 90.64 \text{ volts}$$

In this circuit, the real voltage across L is equal to $E_{R\text{-}IN} + {}_jE_L$ or $9.36 \angle 0° + 954 \angle 90°$ which is equal to $954.04 \angle 89.4°$ or practically the same as E_L.

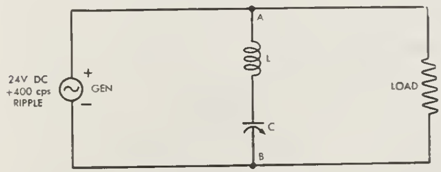
In comparing the results obtained with those in the previous problem, notice that the internal resistance actually lowers the voltage developed across $E_L + E_c$. Also notice that the current which was 20 ma in the first problem becomes 18 ma. You can see just how the internal resistance affects this circuit by studying the graph and vector diagram just below.



Graph and Vector Diagram for Problem 2

Continuing with the effect of the internal resistance, it is well to note that if this resistance were the total resistance in this series-resonant circuit, the current would be equal to 100/.520K, or 192 ma. Further, if an inductor having the same inductance but wound with larger wire having less resistance, an even greater current would flow at resonance. In a circuit, the ratio of the opposition which inductor furnishes to current flow—that is, its inductive reactance—to its internal resistance as you learned earlier is the Q of the inductor—that is, $Q = \frac{X_L}{R}$. The Q of the coil of the circuit in this problem is equal to $\frac{53{,}000}{520}$ or 102.

High Q coils are preferred to low Q coils, especially at resonance. In resonant circuits, high Q coils permit high current to flow to the load since there is very little power lost because of the small resistance of the windings in the coil. Remember the lower the internal resistance, the greater the Q.

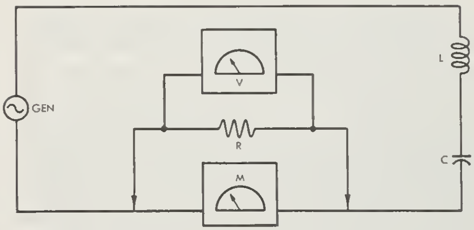


Series-Resonant Filter

Use of Series Resonant Circuits

One use of a series-resonant circuit is to filter out unwanted frequencies. Here is an example. In the above illustration, the generator is producing 24 volts DC with a 400 cps ripple. If C were tuned so that $X_L = X_C$ at 400 cps, the 400 cps ripple then would be effectively shorted out. This is because the impedance to the 400 cps voltage between points A and B consists only of the resistance of the inductor.

Another use of a series-resonant circuit is to measure frequency. For example, if the resonant circuit C in the illustration at the right and top is calibrated in cycles per second and M is an AC current-measuring device, you can use C to measure the frequency of the voltage produced by the generator. This is done by tuning C to a maximum current indication on M, and then reading the frequency on the calibrated dial of C. If the Q of L is high, and if M is an ammeter, it is possible then to obtain an accurate frequency reading, since the curve of I with reference to f will be quite sharp. However, it does have the undesirable effect of shorting the generator and causing an excessive current flow in the circuit. To avoid this you may replace M by a resistor across which there is an AC voltmeter. At resonance then, IR will be maximum and the voltmeter will indicate a peak reading. This arrangement does not give a very sharp indication of the resonant frequency, but does prevent excessively high current from flowing in the circuit.



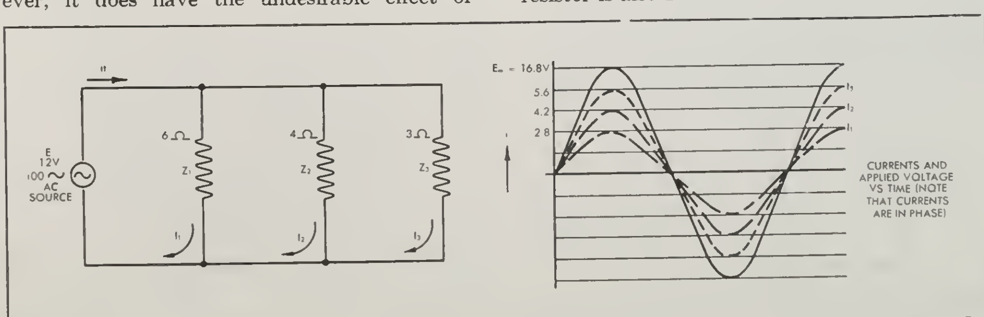
Series-Resonant Circuit as Frequency Indicator

PARALLEL AC CIRCUITS

Resistive Circuits

When resistors all in parallel are connected across an AC voltage, the circuit behaves like a parallel DC circuit, and the circuit problems involved are essentially like those in DC parallel circuits. For example, in the circuit illustrated below, using the values indicated, find the values of current through—and the voltage across—each of the three parallel resistors. Also find the total impedance.

In this problem both Ohm's and Kirchoff's laws are applicable. By using one of Kirchoff's laws which states that the voltages across elements connected in parallel are equal, it follows that since the voltage of the source in the problem is 12 volts, then the voltage across each resistor is also 12 volts.



Parallel Resistive Circuit

Another one of Kirchoff's laws states that the sum of the currents leaving a junction equals the current entering that junction. Therefore, the total current in this circuit equals the sum of the branch currents I_1, I_2, and I_3. But since you do not know the currents in the branches, it is then necessary to find these currents first in order to find the total current. To find them, substitute the voltage which you have already found for each branch and the resistance of each resistor indicated in the circuit in the Ohm's law formula,

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

In branch 1,

$$I_1=\frac{E}{Z_1}=\frac{12}{6}=2 \text{ amperes}$$

In branch 2,

$$I_2=\frac{E}{Z_2}=\frac{12}{4}=3 \text{ amperes}$$

In branch 3,

$$I_3=\frac{E}{Z_3}=\frac{12}{3}=4 \text{ amperes}$$

Then according to Kirchoff's law, the total current I_t is equal to $I_t=I_1+I_2+I_3$. Thus,

$$I_t=2+3+4=9 \text{ amperes}$$

Then by the Ohm's law formula, the total impedance Z_t is equal to $Z_t=\frac{E}{I_t}$. Thus,

$$Z_t=\frac{12}{9}=1.33 \text{ ohms}$$

Another way to find the total impedance and total current in this problem is to obtain the total resistance by the parallel resistance formula first, and then the total current directly by Ohm's law. Here is the solution.

According to the parallel resistance formula, the total impedance of resistors in parallel is equal to,

$$Z_t=\frac{1}{\frac{1}{Z_1}+\frac{1}{Z_2}+\frac{1}{Z_3}}$$

Then by substituting the indicated values of resistances in this equation,

$$Z_t=\frac{1}{\frac{1}{6}+\frac{1}{4}+\frac{1}{3}}$$

Multiplying by $\frac{12}{12}$

$$Z_t=\frac{12}{2+3+4}$$

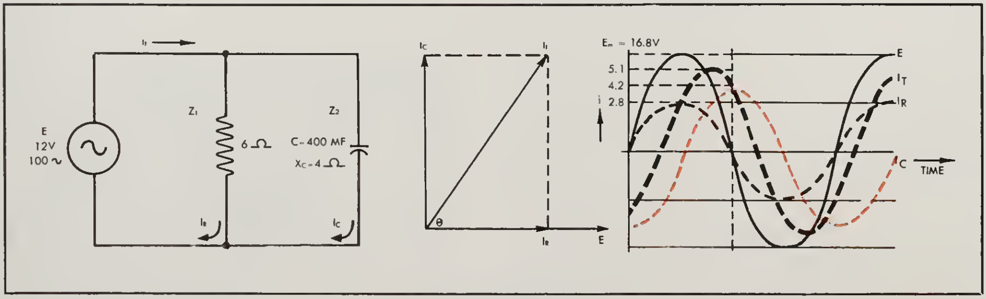
$$Z_t=\frac{12}{9}=1.33 \text{ ohms}$$

Now by substituting in the Ohm's law formula, $I_t=\frac{E}{Z_t}$,

$$I_t=\frac{12}{1.33}=9 \text{ amperes}$$

Resistance and Capacitance

When a resistor and a capacitor are connected in parallel across an AC voltage source, the capacitor introduces capacitive reactance into the circuit as it does in a series AC circuit but with a somewhat different action. Here is what happens. As in any parallel circuit, the voltages across the resistive branch and the capacitive branch here are equal to each other and equal to the applied voltage. The current in the resistive branch is in phase with the applied voltage and also equal to the applied voltage divided by the resistance of the resistor. But the current and voltage through the capacitive branch are not in phase. This results from the fact, as you recall, that current leads voltage in a purely capacitive circuit by 90°. The best way to understand how these currents and voltages act is by use of vector diagrams and sine waves. In this connection, first refer to the vector diagram in the illustration just below. Notice first that



the vector diagram differs from that for the series resistance-capacitive circuit shown before in this chapter. The most important difference results from the fact that in a parallel circuit, the voltage is the same in all parts of the circuit. The voltage, therefore, is the reference vector and not the current as in a series circuit. All other vectors are drawn with reference to the voltage vector. The voltage vector might be labeled E_a, E_R, or E_c. This is because all these voltages are equal and in phase. The current flowing in the resistor is in phase with the voltage applied to it. Thus, it is drawn to coincide with the voltage vector. But since the current through the capacitor leads the applied voltage by 90°, it is drawn perpendicularly upward from the voltage vector as shown. Now if you construct a parallelogram as shown with these vectors, you can find the total current in the circuit. This is represented by the diagonal of the parallelogram. An important point to remember is that in a parallel AC circuit, the total current is not the algebraic but the *vector* sum of the branch currents.

In the vector diagram, also notice the total current vector I_t. It makes an angle of 56° with the voltage vector and has a magnitude (length) about 30% more than I_c and twice I_R. If you mentally let I_t be $\frac{12 \text{ volts}}{6 \text{ ohms}}$, or 2 amperes, the total current would be 4 amperes. The 56° angle tells you the lead effect caused by the capacitor causes the total current to lead the applied voltage by 56°.

In referring to the sine waves graph, you can see that it provides another way of finding the total current. The dashed lines represent the branch current and the black line the total current. This method requires accurate construction.

Example

Problem. Assume the resistance in the circuit in this illustration is equal to 6 ohms and the capacitor has a capacitance of 400 μf, find the following quantities:

1. Branch Current

Solution. First, find the capacitive reactance of the capacitor. This is equal to

$$X_c = \frac{1}{2\pi fc} = 4 \text{ ohms}$$

Now find the current in each branch by the Ohm's law formula,

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

$$\text{In the } I_R \text{ branch, } I_R = \frac{E}{Z_1} = \frac{E}{R} = \frac{12}{6} = 2 \text{ amperes}$$

$$\text{In the } I_c \text{ branch, } I_c = \frac{E}{Z_2} = \frac{E}{X_c} = \frac{12}{4} = 3 \text{ amperes}$$

2. The Total Current

Solution. To find the total current, add the currents I_R and I_c vectorially. The result is a parallelogram like that shown on page 2-56 in which the diagonal is vectorially equal to the sum of the branch currents. Notice that the diagonal is also the hypotenuse of a right triangle. Therefore, by the geometrical theorem, which states that the hypotenuse equals the sum of the squares of the legs of a right triangle, you get the equation which shows that the total current,

$$I_t = \sqrt{I_R^2 + I_c^2}$$

Thus, by substituting in this equation, you obtain,

$$I_t = \sqrt{2^2 + 3^2} = \sqrt{4 + 9} = \sqrt{13} = 3.6 \text{ amperes}$$

3. Phase Angle

Solution. Before finding the phase angle, it is well to note that the dotted lines on each side of I_t in the vector diagram on page 2-56 represent only the absolute values of vector length. This is the length without regard to the phase angle. In practice, certain measuring instruments record only the absolute value of the quantity being measured. In reactive circuits the voltmeter, for example, which indirectly measures voltage by measuring current flow, indicates only absolute values and tells you nothing about the phase angle.

Trigonometry provides the usual method for obtaining the phase angle. Either of the two trigonometric functions—*sine* or *cosine*—can be used. The sine of an angle is the ratio of the *adjacent side* to the *hypotenuse*, and the cosine, the ratio of the *opposite side* to the *hypotenuse*. Here is how the cosine function is used.

In the vector triangle shown on page 2-56, I_R and I_t form two sides of a right triangle and also include the phase angle θ. According to the definition of a cosine,

$$\text{Cos } \theta = \frac{I_R}{I_t}$$

Simplifying,

$$\theta = \text{Cos}^{-1} \frac{I_R}{I_t}$$

Substituting for I_R and I_t

$$\theta = \text{Cos}^{-1} \frac{2}{3.6} = \text{Cos}^{-1} .555$$

From trigonometric tables, the phase angle θ is,

$$\theta = 56°$$

To find the phase angle θ by the sine function, solve for θ in the equation, $\text{Sin } \theta = \frac{I_c}{I_t}$

$$\theta = Sin^{-1} \frac{I_c}{I_t}$$

Substituting,

$$\theta = Sin^{-1} \frac{3}{3.6} = Sin^{-1} .833$$

From tables, $\theta = 56°$

4. *Total Impedance*

Solution. In finding the total impedance, substitute the total current and applied voltage in the Ohm's law formula, $Z_t = \frac{E}{I_t}$.

Thus, $Z_t = \frac{12}{3.6} = 3.33 \text{ ohms}$

5. *Power Factor*

Solution. When calculating the power factor, keep in mind that the current in the circuit is out of phase with the voltage and that the resistor alone dissipates power. In an AC circuit that contains capacitance and resistance connected in parallel, the power factor equals the ratio of the energy dissipated to the total power delivered to the circuit. Since the ratio of the current in the resistance to the total current is equal to the ratio of the energy dissipated to the total power delivered to the circuit, the power factor, p.f., then is equal to,

$$p.f. = (Cos\ \theta) = \frac{I_R}{I_t}$$

By substituting,

$$p.f. = \frac{2}{3.6} = .555 \text{ or } 55.5\%$$

6. *Power*

Solution. As mentioned before, a resistor is the only element in a circuit which actually consumes power. Of course a capacitor takes power from the circuit during one alternation, but it returns it again to the circuit during the next alternation. Therefore, over a complete cycle, a capacitor dissipates no power.

In resistive-capacitive circuits there are several methods for determining power consumption.

The method you use depends on the quantities whose values you know. One method is to find the apparent power in the circuit, and then multiply it by the power factor. The following describes this method. In this problem, you have already found that the power factor is .555 or 55.5%. Therefore, it is only necessary to find the apparent power. Apparent power, if you recall, is the total current multiplied by the applied voltage— that is, $P_a = EI$. The total current in the circuit (see page 2-57) is 3.6 amperes. Thus, substituting in the formula, $P_a = EI$, you obtain,

$$P_a = 12 \times 3.6 = 43.2 \text{ watts}$$

Then multiplying this result by the power factor, the true power is,

$$P = 43.2 \times .555 = 24 \text{ watts}$$

Here is another method. The true power dissipated in a circuit—that by the resistance—is equal to the square of the current through the resistor multiplied by ohmic resistance of the resistor—or mathematically,

$$P = I^2R$$

Therefore, substituting for I_R (see page 2-57) and for R (see circuit on page 2-56),

$$P = 2^2 \times 6 = 4 \times 6 = 24 \text{ watts}$$

A variation of this formula is the formula,

$$P = \frac{E^2}{R}.$$

By substituting the voltage applied to the resistance and the ohmic value of the resistor, you likewise find that the true power is,

$$P = \frac{12^2}{6} = \frac{144}{6} = 24 \text{ watts}$$

7. *Rectangular Notation*

Solution. Solving this problem is easy enough by using simple mathematics, but you will often come across complicated problems where it is easier to use rectangular or polar notation. Here is how you can solve for some of the quantities in this problem by rectangular notation.

a. *Total Impedance*

When the resistor and capacitor are referred to rectangular coordinates, they are equal respectively to,

$$Z_1 = 6 + j0$$
$$Z_2 = 0 - j4$$

By substituting in the formula for resistances in parallel and simplifying,

$$Z_t = \frac{Z_1Z_2}{Z_1 + Z_2}$$

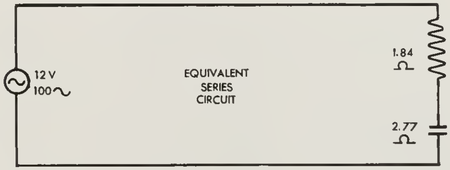
$$Z_t = \frac{(6+j0)(0-j4)}{(6+j0)+(0-j4)} = \frac{-j24}{6-j4}$$

By rationalizing the denominator,

$$Z_t = \frac{(-j24)\ (6+j4)}{(6-j4)\ (6+j4)} = \frac{96-j144}{52} = 1.84 - j2.77$$

This answer is of little value unless you intend to make additional computations. What you really want to know is the absolute value of Z_t—that is, its value in ohms. This is what you do. First, notice that the answer has the form, $R\text{-}jx$, where R is the amount of resistance and $-x$

is the amount of reactance in an equivalent series circuit. The negative sign before the x indicates that the reactance is capacitive. Thus, in this condition, this parallel circuit will have the same effect on current flow as the equivalent circuit illustrated directly below.



Since absolute value is not concerned with phase, you can drop the j and substitute 1.84 for R and 2.77 for the reactance X_c in the formula,

$$Z_t = \sqrt{R^2 + Xc^2}$$

and solve. This gives,

$$Z_t = \sqrt{1.84^2 + 2.77^2} = \sqrt{3.3856 + 7.6729}$$

$$Z_t = \sqrt{11.0585} = 3.33 \text{ ohms}$$

b. Phase Angle

In the rectangular form, the expression for the phase angle is $\tan\theta = \frac{X}{R}$

Therefore, by substitution,

$$\tan\theta = \frac{2.77}{1.84}$$

By simplifying,

$$\theta = \tan^{-1}\frac{2.77}{1.84} = \tan^{-1} 1.5$$

From tables,

$$\theta = 56^\circ$$

c. Power Factor

The power factor is equal to the cosine of θ. Since the cosine of θ equals $\frac{R}{Z_t}$ and this in turn equals $\frac{R}{\sqrt{R^2 + X_c^2}}$, therefore,

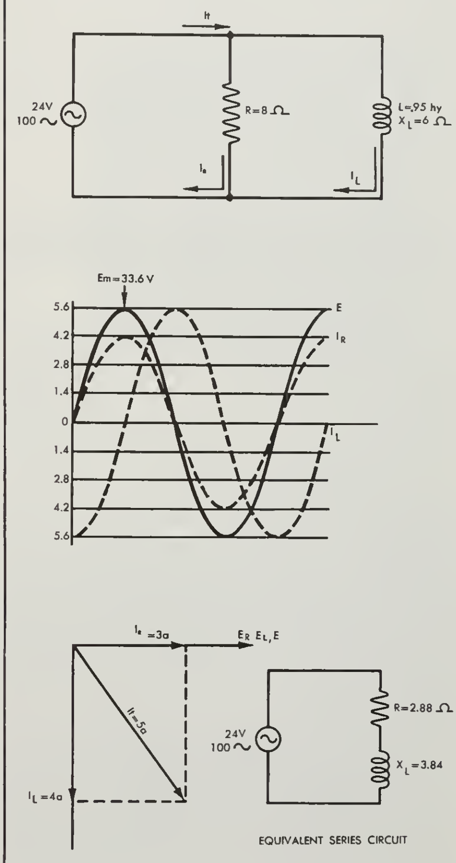
$$p.f. = \cos\theta = \frac{R}{\sqrt{R^2 + X_c^2}}$$

By substituting,

$$p.f. = \frac{1.84}{3.33} = .555 \text{ or } 55.5\%$$

Circuits Containing Inductance and Resistance

Now consider an AC circuit in which inductance and resistance are connected in parallel. In studying this circuit, refer to the inductive-resistance circuit illustrated to the right and find the following circuit quantities.



Inductive-Resistive Circuit

CURRENT. In finding the total current in the circuit, first draw the vector diagram as illustrated and from it derive the formula for total current. In constructing the vector diagram, it will be necessary to find the currents. Here is how it is done.

Since the voltage across all circuit elements is the same—that is, the applied voltage $E_a = E_R = E_L$, the voltage becomes the reference vector. Since current, I_R, (the resistor current) is in phase with the applied voltage E_R, it must be drawn to coincide with the voltage vector as shown. But since the current flowing through the inductance

L lags the applied voltage by 90°, it must be drawn downward at a 90° angle as shown. Before you can actually draw these vectors, you must first determine their magnitude. To do this, first note that the total current divides, part going through the resistance, R, and part through the inductance, L. Then find the current through each by

Ohm's law, $I = \frac{E}{Z}$.

Substituting in this formula the values indicated in the circuit, the current in the resistive branch is equal to,

$$I_R = \frac{24}{8} = 3 \text{ amperes}$$

In finding the current through the inductive branch, it is first necessary to find the inductive reactance of that branch by use of the formula, $X_L = 2\pi f L$.

Substituting the values given in this formula,

$$X_L = 6.28 \times 10^2 \times .95 = 6 \text{ ohms}$$

Then by Ohm's law

$$I_L = \frac{24}{6} = 4 \text{ amperes}$$

As illustrated, when the vectors are completed into a parallelogram, the diagonal is the vector sum of the branch currents and also the hypotenuse of a right triangle. Since the legs of this triangle are I_R and I_L respectively, you can form the equation,

$$I_t = \sqrt{I_R^2 + I_L^2}$$

Thus, in substituting the values for I_R and I_L,

$$I_t = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5 \text{ amperes}$$

TOTAL IMPEDANCE. In finding the total impedance, substitute the values for I_t and applied voltage in the Ohm's law for AC, $Z = \frac{E}{I_t}$. Thus,

$$Z = \frac{24}{5} = 4.8 \text{ ohms}$$

PHASE ANGLE. In finding the phase angle, substitute the values for I_R and I_t in the formula $\text{Cos } \theta = \frac{I_R}{I_t}$. Thus,

$$\text{Cos } \theta = \frac{3}{5}$$

$$\theta = \text{Cos}^{-1} \frac{3}{5}$$

$$\theta = \text{Cos}^{-1} .60$$

From tables,

$$\theta = 53°$$

POWER. In finding the actual power consumed, multiply the apparent power (EI_t) by the power factor.

The power factor is equal to Cos θ. From the previous problem θ was found to equal 53°. Thus, from trigonometric tables, Cos 53° = .6.

In substituting for the values E, I_t and $\text{Cos } \theta$ in the equation, $P = EI_t \text{Cos } \theta$,

$$P \text{ (actual)} = 24 \times 5 \times .6 = 72 \text{ watts}$$

RECTANGULAR NOTATION. The following are the calculations of the preceding circuit values by rectangular notation. In these calculations notice that the resistive branch is expressed as 8+j0, the inductive branch as 0+j6. (Inductive qualities are signed positive. Remember from the discussion of the resistive capacitive circuit that the capacitive component is signed negative.)

In finding total impedance, substitute the required values in the equation, $Z_t = \frac{Z_R Z_L}{Z_R + Z_L}$.

Thus,

$$Z_t = \frac{(8+j0)\ (0+j6)}{(8+j0)+(0+j6)} = \frac{j48}{8+j6}$$

Rationalizing the denominator,

$$Z_t = \frac{(j48)\ (8-j6)}{(8+j6)\ (8-j6)} = \frac{288+j384}{100} = 2.88+j3.84$$

Changing answer to absolute value of Z_t by using equation, $Z_t = \sqrt{R^2 + X_c^2}$

$$Z_t = \sqrt{2.88^2 + 3.84^2} = \sqrt{8.29 + 14.77} = \sqrt{23.06}$$

$$Z_t = 4.8 \text{ ohms}$$

Using $\tan\theta = \frac{X}{R}$, the expression for phase angle in rectangular form for finding the phase angle θ,

$$\tan\theta = \frac{3.84}{2.88}$$

transposing,

$$\theta = \tan^{-1} \frac{3.84}{2.88} = \tan^{-1} 1.33$$

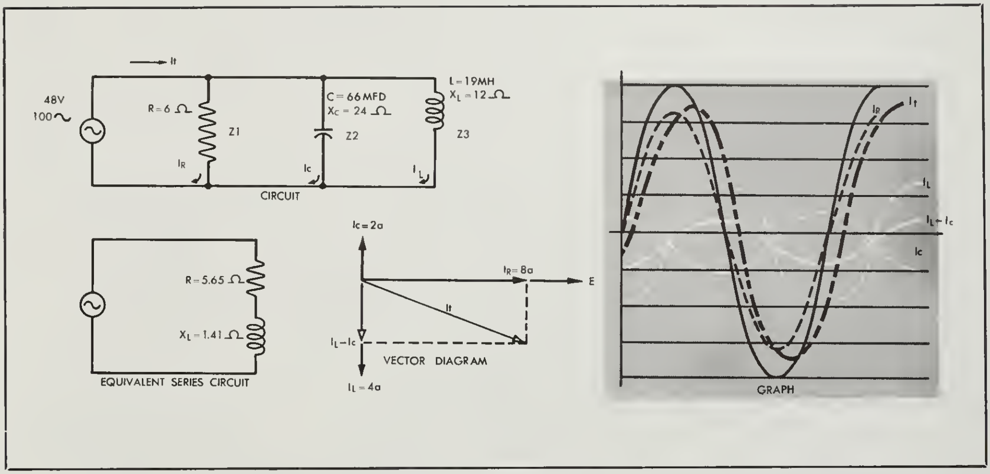
From trigonometric tables,

$$\theta = 53°$$

The rectangular form of the total impedance, 2.88+j3.84 is in the form $R+jX$. This form indicates that the series circuit equivalent to this circuit will contain a resistance of 2.88 ohms and a reactance equal to 3.84 ohms.

Circuits Containing Inductance and Capacitance

When both inductance and capacitance exist in combination with resistance in parallel circuits, the computations are somewhat lengthy, but not any more difficult than those in circuits which contain only resistance and capacitance. In the circuit illustrated on the next page find the following quantities:



Inductive-Capacitive Circuit

Total current. In this circuit, the total current is the vector sum of the individual branch currents. Other current characteristics are the following: The current in the resistive branch is in phase with the applied voltage and thus is indicative of the true power consumption. The inductive current lags the applied voltage by 90°, and the capacitive current leads it by 90°. Since the effect of these currents is exactly opposite, they cancel each other. In finding the branch currents, substitute the values, $R=6$ ohms, $X_c=24$ ohms, $X_L=12$ ohms, in the Ohm's law formula, Thus

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

$$I_R=\frac{48}{6}=8 \text{ amperes}$$

$$I_c=\frac{48}{24}=2 \text{ amperes}$$

$$I_L=\frac{48}{12}=4 \text{ amperes}$$

In this circuit, as in the previous circuit, the voltage is the common factor for the resistance, inductance, and capacitance. Therefore, in drawing the vector diagram, first lay out the voltage to scale as a horizontal vector to the right. Since the current in the resistance is in phase with the voltage applied to it, lay out the current in the resistance along the voltage vector. Now, lay out the current through the capacitance. Since the current through this element leads the impressed voltage by 90°, draw the condenser current (I_c) perpendicularly upward from the starting point on the voltage vector. Next, construct the vector for the current flow through the inductance. As you recall, the current flowing in a pure inductance lags the voltage 90°. Therefore, draw this vector (I_L) perpendicularly downward from the starting point of the voltage vector.

Since the current flowing through the condenser is 180° out of phase with the current flowing through the inductor, you can subtract these two values numerically. In this circuit the inductor current is greater than the condenser current. Thus, the resultant of these two currents, I_L-I_c, is a current of smaller value but in a lagging position with reference to the applied voltage. The next step is to obtain the resultant between the current (I_L-I_c) and the vector of the resistor current (I_R). As you learned before, this is done by completing the parallelogram between the (I_L-I_c) vector and the I_R vector. The diagonal of this parallelogram is represented by I_T and its scalar length is the value of the total current flowing in the circuit. The angular displacement between this current vector I_T and the voltage vector E is the phase angle. The cosine of this angle—or expressed in another way, $\frac{I_R}{I_T}$ is the power factor.

As you can see, the diagonal of this parallelogram is the hypotenuse of a right triangle of which the two sides are I_R and I_L-I_c. Therefore,

$$I_T=\sqrt{I_R^2+(I_L-I_c)^2}$$

On substituting the previously found values for the currents in this formula, you obtain,

$$I_T = \sqrt{8^2 + (4-2)^2}$$
$$= \sqrt{64+4}$$
$$= \sqrt{68}$$
$$I_T = 8.25 \text{ amperes}$$

Of course, there are circuits with resistance, capacitance, and inductance in which the current flowing through the capacitance is greater than the current flowing through the inductance. This shifts the parallelogram above the voltage vector. (In the circuit just described, notice that the parallelogram is below the voltage vector.) So far as the formula for the total current is concerned, it makes no difference which current—the inductive or the capacitive—is the larger, for the square of a negative number has a positive sign.

PHASE ANGLE. In finding the phase angle, substitute the previously found values for I_R and I_T in the equation,

$$\text{Phase angle } \theta = \cos^{-1}\frac{I_R}{I_T}$$

$$\text{Thus,} \quad \theta = \cos^{-1}\frac{8.00}{8.25} = \cos^{-1} \ 969$$

From trigonometric tables,

$$\theta = 14^\circ$$

POWER. You can find the power (true) consumed in the circuit by multiplying the product of the applied voltage and the total current by the power factor. Thus, in substituting the cos θ from the previous problem (phase angle) and the values of E and I_T in the formula, $P = EI \text{ Cos } \theta$. Thus, Power (true) = 48×8.25×.9697 = 384 watts.

A simple method of checking this answer is to multiply the voltage across the resistor by the current flow through it—that is, by performing the operations indicated in the formula, $P = EI_R$. Thus,

$$P \text{ (true)} = 48 \times 8 = 384 \text{ watts}$$

TOTAL IMPEDANCE. In solving for the total impedance, substitute the previously indicated and calculated values in the Ohm's law formula, $Z_t = \frac{E}{I_c}$. Thus,

$$Z_t = \frac{48}{8.25} = 5.81 \text{ ohms}$$

The vector for the total impedance shows it to be largely resistive, but since there is some inductive reactance, the current slightly lags the voltage.

RECTANGULAR NOTATION. Rectangular notation permits you to find the total impedance directly. In other words, it is not necessary to find current and then to substitute in Ohm's law. This is the solution for Z_t by rectangular notation:

In the rectangular form the impedances of the branches, Z_1, Z_2, and Z_3 are 6, $-j24$, and $j12$ respectively.

Substituting these values in the formula for impedances connected in parallel,

$$\frac{1}{Z_t} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3} \text{ gives,}$$

$$\frac{1}{Z_t} = \frac{1}{6} + \frac{1}{-j24} + \frac{1}{j12}$$

$$= \frac{4 + j1 - j2}{24} = \frac{4 - j1}{24}$$

Rationalizing,

$$Z_t = \frac{24}{4 - j1} \times \frac{4 + j1}{4 + j1}$$

$$= \frac{96 + j24}{17}$$

$$= 5.65 + j1.41$$

Changing this answer to absolute value,

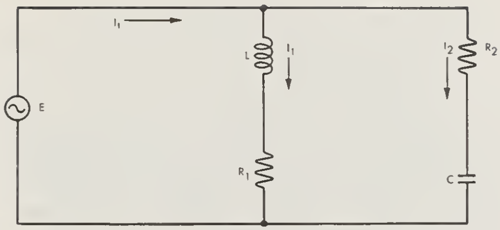
$$Z_t = \sqrt{(5.65)^2 + (1.41)^2}$$
$$= \sqrt{31.9 + 2.0}$$
$$= \sqrt{33.9}$$
$$Z_t = 5.81 \text{ ohms}$$

You can draw the equivalent series circuit from the rectangular form of Z_t as shown on page 2-61. It contains 5.65 ohms of resistance and 1.41 ohm of reactance. The sign in front of the reactive component ($j1.41$) is positive, indicating that the reactance is inductive.

Complex Circuit Combination

Another type of parallel circuit to consider, is one containing an inductive-resistive branch, and a capacitive-resistive branch. In studying this circuit, find the circuit quantities, Z_t, I_t, I_1, I_2, E_{R1}, E_{R2}, E_c and E_L, and draw the vector diagrams, using the following circuit and the following given values:

$$E = 100 \text{ volts}$$
$$f = 500 \text{ kc}$$
$$L = 9.55 \text{ mh}$$
$$C = 63.3 \text{ mmfd}$$
$$R_1 = 15K$$
$$R_2 = 12K$$



Solving for Z_t. First find the reactances of L and C.

$$X_L = 6.28 \times 5 \times 10^5 \times 9.55 \times 10^{-3} = 30K$$

$$X_c = \frac{1}{6.28 \times 5 \times 10^5 \times 63.6 \times 10^{-12}} = 5K$$

Now find the impedances of branches 1 and 2,

$$Z_1 = 15K + j30K = 33.6K \angle 63.5^\circ$$

$$Z_2 = 12K - j5K = 13K \angle -22.6^\circ$$

Then by substituting these values in the equation, $Z_t = \dfrac{Z_1 Z_2}{Z_1 + Z_2}$

$$Z_t = \frac{(33.6K \angle 63.5^\circ)\ (13K \angle -22.6^\circ)}{(15 + j30 + 12 - j5)K}$$

$$= \frac{436K \angle 40.9^\circ}{27 + j25} = \frac{436K \angle 40.9^\circ}{36.8 \angle 42.7^\circ}$$

$$Z_t = 11.85K \angle -1.9^\circ$$

Solving for the currents, I_t, I_1, and I_2.

$$I_t = \frac{E}{Z_t} = \frac{100 \angle 0^\circ}{11.85K \angle -1.9^\circ} = 8.43 \angle 1.9^\circ \text{ ma}$$

$$I_1 = \frac{E}{Z_1} = \frac{100 \angle 0^\circ}{33.6K \angle 63.5^\circ} = 2.98 \angle -63.5^\circ \text{ ma}$$

$$I_2 = \frac{E}{Z_2} = \frac{100 \angle 0^\circ}{13K \angle -22.6^\circ} = 7.8 \angle 22.6^\circ \text{ ma}$$

Solving for the voltage drops, E_{R1}, E_L, E_{R2}, and E_c.

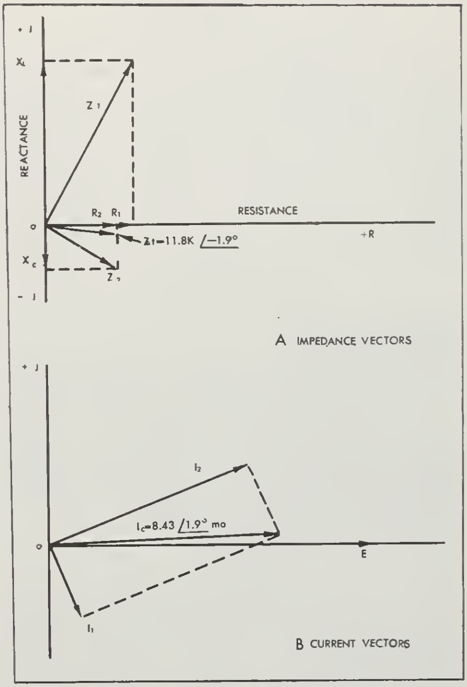
$$E_{R1} = I_1 R_1 = 2.98 \angle -63.5^\circ \text{ ma} \times 15K = 44.6 \angle -63.5^\circ \text{ volts}$$

$$E_L = I_L X_L = 2.98 \angle -63.5^\circ \text{ ma} \times 30K \angle 90^\circ = 89.4 \angle 26.5^\circ \text{ volts}$$

$$E_{R2} = I_2 R_2 = 7.8 \angle 22.6^\circ \text{ ma} \times 12K = 93.6 \angle 22.6^\circ \text{ volts}$$

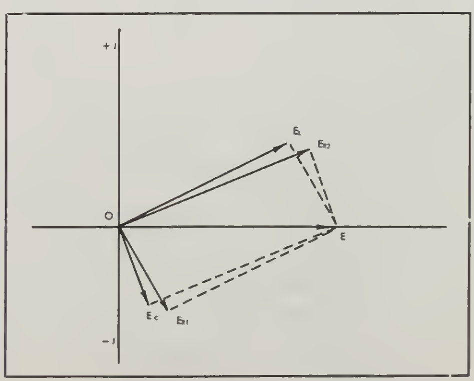
$$E_c = I_2 X_c = 7.8 \angle 22.6^\circ \text{ ma} \times 5K \angle -90^\circ = 39 \angle -67.4^\circ \text{ volts}$$

Now draw the vector diagrams as illustrated at the upper right. In diagram A, notice that the total impedance vector lies slightly below the resistance axis. This indicates there is a small amount of capacitive reactance associated with the total impedance. You therefore can expect the total current to lead the applied voltage by a small angle. If the applied voltage were plotted on the same graph, it would lie as shown in diagram B.

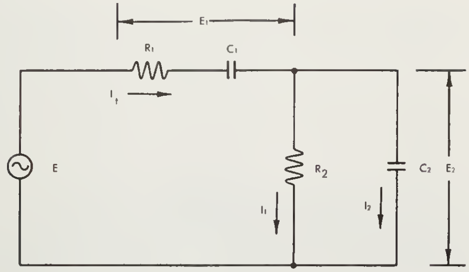


Impedance and Current Vectors

On referring to the voltage vector diagram just below, notice the vector sum of $E_L + E_{R1}$ and $E_c + E_{R2}$ equals the applied voltage. Notice also that E_L and E_{R1} are 90° out of phase. This same out-of-phase relationship also exists between E_c and E_{R2}.



Voltage Vectors



Series-Parallel Circuit

Series-Parallel Circuits

The following is the solution of the AC series-parallel circuit illustrated above. This circuit is part of a Wien-bridge oscillator circuit, a circuit of which the theory of operation is discussed later in this manual. With reference to it here, find the circuit quantities, Z_t, I_t, I_1, I_2, E_1, and E_2, using the following given values:

$E = 100$ *volts*
$f = 2000$ *cps*
$R_1 = R_2 = 90K$
$C_1 = C_2 = 0.001$ *mfd*

SOLVING FOR THE TOTAL IMPEDANCE Z_t. First, find X_{c1} and X_{c2},

$$X_{C1} \text{ and } X_{C2} = \frac{1}{6.28 \times 2 \times 10^3 \times 10^{-3} \times 10^{-6}} = 79.5K$$

Then find Z_3 (impedance of R_1 and C_1 in series),

$$Z_3 = 90K - j79.5K = 120K \angle -41.4^\circ$$

Then find the total impedance, Z_t,

$$Z_t = 90K - j79.5K + \frac{(90K)(79.5K \angle -90^\circ)}{90K - j79.5K}$$

$$= 90K - j79.5K + \frac{7155K \angle -90^\circ}{120 \angle -41.4^\circ}$$

$$= 90K - j79.5K + 59.6K \angle -48.6^\circ$$

$$= 90K - j79.5K + 39.4K - j44.8K$$

$$Z_t = 129.4K - j124\ 3K$$

Thus,

$$Z_t = 179K \angle -44^\circ$$

FINDING I, I_1, AND I_2,

$$I_t = \frac{E}{Z_t} = \frac{100 \angle 0^\circ}{179K \angle -44^\circ} = 0.559 \angle 44^\circ \text{ ma}$$

$$I_1 = \frac{I_t \times (-jX_{c2})}{R_2 - jX_{c2}} = \frac{0.559 \angle 44^\circ \text{ ma} \times 79.5K \angle -90^\circ}{90K - j79.5K}$$

$$= \frac{0.559 \angle 44^\circ \text{ ma} \times 79.5K \angle -90^\circ}{120K \angle -41.4^\circ}$$

$$= 0.370 \angle -4.6^\circ \text{ ma}$$

$$I_2 = I_t \times \frac{R_2}{R_2 - jX_{c2}} = \frac{0.559 \angle 44^\circ \text{ ma} \times 90K}{120K \angle -41.4^\circ}$$

$$= 0.418 \angle 85.4^\circ \text{ ma}$$

SOLVING FOR VOLTAGE DROPS E_1 AND E_2.

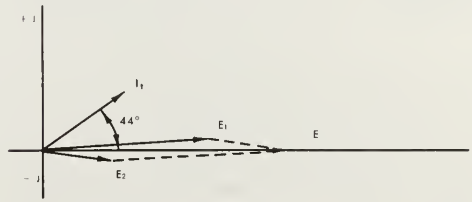
$$E_1 = I_t Z_3 = 0.559 \text{ ma} \angle 44^\circ \times 120K \angle -41.4^\circ$$

$$E_1 = 67 \angle 2.6^\circ \text{ volts}$$

$$E_2 = I_1 R_2 = I_2 X_{c2} = 0.370 \text{ ma} \angle -4.6^\circ \times 90K$$

$$E_2 = 33.3 \angle -4.6^\circ \text{ volts}$$

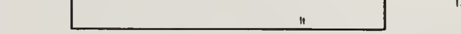
The vector diagram just below shows these circuit quantities. Notice that E_1 and E_2 are almost in phase with the applied voltage, E. At the same time, the line current is almost 45° out of phase with the generator voltage.



Vector Diagram

Parallel Resonance

A parallel-resonant circuit is one of the most widely used devices in electronic receivers, transmitters and frequency-measuring devices. Unlike a series-resonant circuit, a parallel-resonant circuit is characterized by high impedance and minimum current at resonance. The following discussion of parallel resonance is based on the parallel-resonant circuit shown in the illustration directly below.



Parallel Resonant Circuit

ANALYSIS. In the circuit illustrated the following notation is used:

E_G = *generator voltage.*
R_G = *generator resistance in ohms.*
I_t = *total current flow in amperes.*
I_C = *current through condenser in amperes.*
I_L = *current through inductor in amperes.*
R_L = *internal resistance of inductor in ohms.*
L = *inductance of inductor in henries.*
C = *capacitance of capacitor in farads*

An important consideration in the study of parallel resonance is the circuit condition which produces resonance. In this connection there are three definitions of the frequency at which a tank circuit is resonant. Based on the circuit illustrated on page 2-64, they are the following:

1. *The frequency of E_G at which I_t is in phase with E_G. (This is the condition of unity power factor.)*

2. *The frequency of E_G at which X_L equals X_C.*

3. *The frequency of E_G at which the impedance of the tank is maximum.*

These definitions lead to resonant frequencies that differ by as little as one per cent when the circuit Q is at all large. For all practical purposes, the resonant frequency of a parallel circuit is the frequency that satisfies the relation,

$$\omega L = 1/(\omega C),$$

or in simpler terms, the relation,

$$\text{resonant frequency} = 1/(2\pi\sqrt{LC})$$

Another useful concept in the study of parallel resonance is the expression for the impedance of the tank,

$$Z_{AB} = \frac{-jX_C(R_L+jX_L)}{R_L+jX_L-jX_C} = \frac{X_LX_C-jX_CR_L}{R_L+jX_L-jX_C}$$

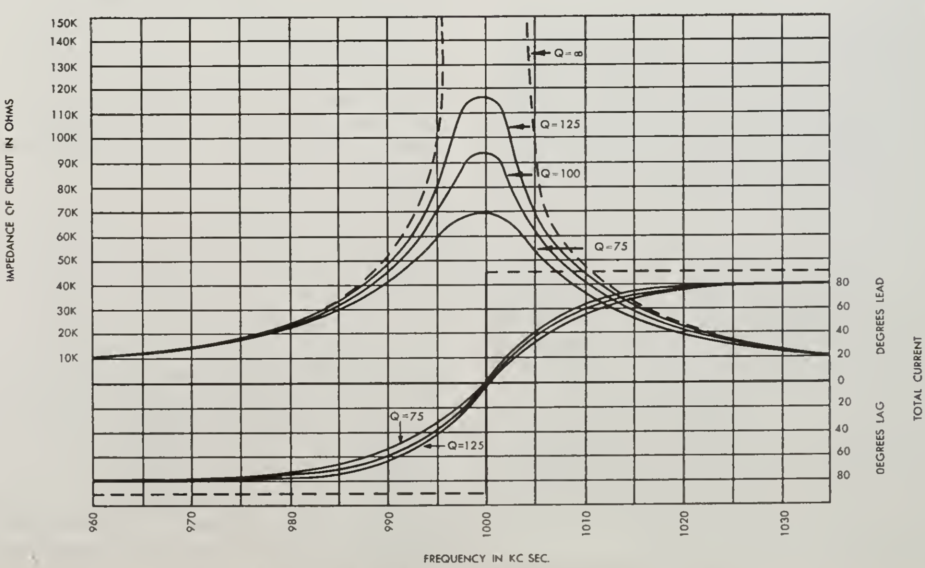
From the second definition of resonant frequency, you can see that when the frequency of E_G is such that $X_L=X_C$, then the expression for the impedance of the tank is,

$$Z_{AB} = \frac{X_L^2-jX_CR_L}{R_L} = \frac{X_L^2}{R_L} - jX_C$$

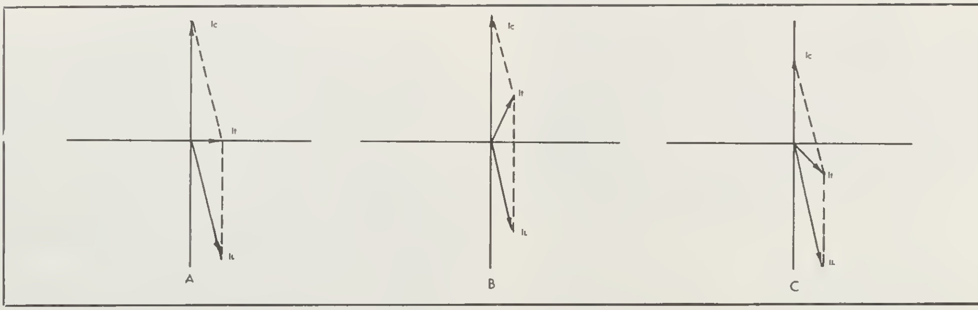
From this expression, you can see that as R_L approaches zero, Z_{AB} approaches infinity. Since $X_L/R_L=Q$, you can write the relationship $Z_{AB}=X_L^2/R_L-jX_C$ as QX_L-jX_C. Thus when the Q of the circuit is reasonably high, QX_L becomes much greater than $-jX_C$ and since arc tan (X_C/QX_L), or arc tan $(1/Q)$ is approximately equal to $90/0°$, the impedance approaches pure resistance.

From the previous paragraphs, it is apparent that the three definitions are all identical when R_L is 0, and Q is infinitely large. Furthermore for all practical purposes they are identical when the inductor L has a reasonably high Q.

In the tank circuit shown on page 2-64, L equals .15 millihenrys, C equals 169 mmfds, and R_L equals 0. If you first were to calculate the absolute impedance of the tank for frequencies ranging from 960 kc through resonance (1000 kc) and to 1030 kc, and then plot a graph of the impedance against frequency, you would obtain the dotted curve shown directly below. This graph is important for the portion showing phase relations shows the following rule to be true: At frequencies below resonance, the impedance Z_{AB} is inductive; and, at frequencies above resonance, Z_{AB} is capacitive.



Magnitude and Phase Angle of Impedance as a Function of Frequency



Currents At, Above, and Below Resonance

Again referring to the graph, as R_L increases from zero to 7.5 ohms, Q decreases to 125. Likewise when $R = 9.4$ ohms, $Q = 100$; when $R = 12.5$ ohms, $Q = 75$. The other curves on this graph are obtained by letting $Q = 125$, 100, and 75, respectively and using the procedure explained before. Notice in the chart that, as the Q decreases, the sharpness of the curve decreases.

Also notice that, as Q decreases, the angle of lead or lag decreases for any one frequency except that of resonance. For example, at 1005 kc, the circuit with a Q of 125 would have an impedance of $70K/-50°$. Similarly the circuit with a Q of 75 would have an impedance of

$$53K/-40°.$$

CURRENT AND VOLTAGE RELATIONSHIPS. As previously shown, the tank offers a pure resistance load to the generator of approximately QX_L or $Q\omega L$ at resonance. The total impedance of the parallel resonant circuit shown on page 2-64 is equal to $R_G + QX_L$ when the frequency of E_G is such that the tank is resonant. Since the load for E_G is pure resistance, I_t is in phase with E_G.

Since the tank acts as pure resistance, the voltage across the tank, E_C, is in phase with E_G, but the current through the condenser will lead the voltage E_C by 90° and the current through the inductor will lag the voltage E_L by an angle whose tangent is QX_L/R_L. This angle approaches −90° as R_L approaches zero.

CURRENTS AT RESONANCE. The vector diagram above at A shows the phase and amplitude relationships of I_C, I_L, and I_t in a parallel-resonant circuit at resonance. From it you can see that I_C and I_L are much larger than I_t and are approximately 180° out of phase.

FREQUENCIES ABOVE RESONANCE. At frequencies above resonance, the reactance X_C decreases and X_L increases. This causes an increase in current flow through C and a decrease in current through L. Therefore, the total current now leads the applied voltage, and the tank acts capacitively as shown vectorially at B.

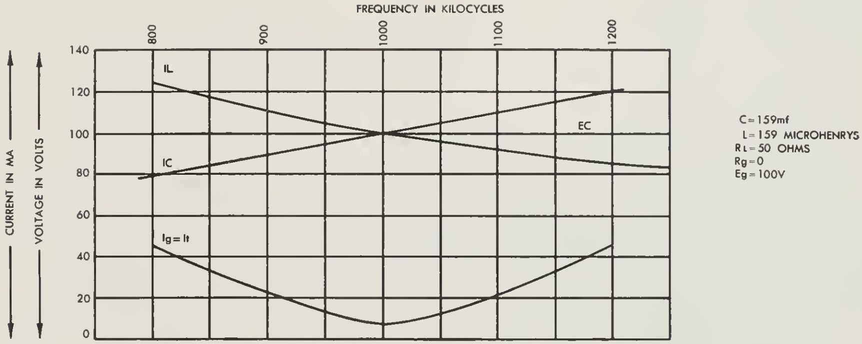
FREQUENCIES BELOW RESONANCE. At frequencies below resonance, the reactance of C increases while that of L decreases. This causes a decrease in current flow through C and an increase in current through L. Since the total current then lags the applied voltage, the tank acts inductively as shown at C.

CURRENTS AND VOLTAGES IN A RESONANT CIRCUIT. If E_G in the parallel-resonant circuit illustrated on page 2-64 is constant and R_G equals 0, the voltage across the tank, E_C, then is constant at all frequencies. But if the frequency of the generator is changed in steps from below resonance to above resonance, I_C will increase while I_L will decrease. As you can see in the graph at the top on the next page, I_t will be minimum at resonance, since at resonance the impedance is maximum.

Also on the next page notice the graph labeled, Effect of R_g on Currents and Voltages. This graph shows the relationship of the quantities I_c, I_L, I_t, and E_c when $R_g = 20{,}000$ ohms.

Notice in this graph that the voltage across the tank becomes maximum at resonance. This is explained by the fact that, at resonance, the impedance of the tank is maximum. At frequencies above and below resonance, the impedance of the tank goes down, I_t increases, and a large portion of E_G is dropped across R_G.

Also notice in the same graph that, at any one frequency below resonance, I_L is greater than I_C while, at any one frequency above resonance, I_C is greater than I_L.

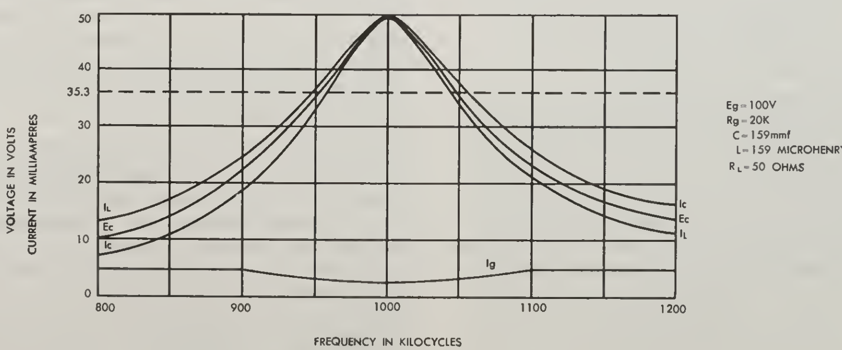


Currents and Voltages in a Circuit near its Resonant Frequency

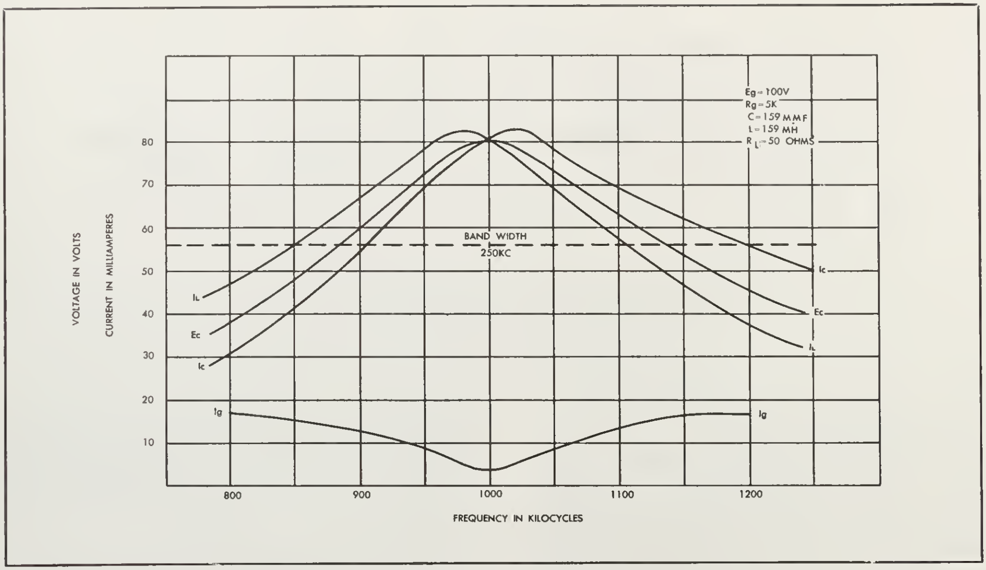
Bandwidth and Selectivity

One characteristic which makes a parallel resonant circuit very useful in radio circuits is its response to the resonant frequency and its discrimination against frequencies off-resonance. In this connection, notice how the currents I_c and I_L and the voltage E_c in the graph directly below, peak at the resonant frequency and drop off rapidly on either side of resonance. The rapidity at which the response drops off is the figure of merit of a resonant circuit and is called the *selectivity* or the *sharpness* of resonance. An arbitrary standard has been set up to compare the selectivity of resonant circuits. This standard uses as reference points the two frequencies on each side of resonance at which the circulating tank current is .707 times its maximum value. The band of frequencies between these points is called the *bandwidth*. Since the value .707 is the square root of two, the power ($P = I^2R$) in the circuit is one-half maximum when the tank current is .707 times its maximum value. Therefore, the two frequencies are called the *half-power* points. From this it follows that selectivity is the ratio of the bandwidth to the resonant frequency, that is, $\frac{\Delta f}{f_r}$.

In many radio circuits, the chief interest is not current flow in the circuit, but the voltage developed across the tank circuit. For example, assume a resonant circuit in which voltage is applied to the grid of a vacuum tube in a following stage (one which takes very little power from the resonant circuit). Since near resonance the reactance of the condenser in the tank circuit does not vary much, the graph of the condenser voltage is practically like that of the condenser current. Therefore, you can consider the bandwidth as a band of frequencies between the values of E_c that equal .707 times maximum value. Notice that the half-power points are 950 kc and 1050 kc and that the bandwidth is 100 kc.



Effect of Rg on Currents and Voltages



Effect of R_g on Bandwidth

Determining factors in parallel resonance. To determine the factors which influence the selectivity of a parallel resonant circuit, convert the parallel combination of R_g, R_L, and X_L illustrated on page 2-64 into an equivalent series impedance, and find the Q of this combination when forming a series resonant circuit with C. This gives,

$$Q = \frac{R_g X_L - R_L X_L}{R_g R_L + X_L^2}$$

Since

$$\frac{\Delta f}{f_r} = \frac{1}{Q} \text{ and } R_L X_L \ll R_g X_L$$

Then,

$$\frac{\Delta f}{f_r} = \frac{R_g R_L + X_L^2}{R_g X_L} = \frac{R_L}{X_L} + \frac{X_L}{R_g} = \frac{1}{Q} + \frac{X_L}{R_g},$$

where Q_L is that of the coil alone.

This expression shows that bandwidth and selectivity are primarily affected by the Q of the tank circuit and the generator resistance. Specifically, the greater the Q of the circuit, the sharper the resonance and the more selective the circuit. R_g must be large in comparison to X_L if the circuit is to be selective. The greater the R_g, the greater the selectivity. To understand this point, notice the I_g, I_c, I_l, and E_c curves for the circuit illustrated above in which R_g is 5K rather than 20K. In this case the bandwidth is 250 kc. In radio, the generator is usually a vacuum tube having a high internal resistance. Thus, the selectivity is quite good when a parallel resonant circuit with high Q is used. This also shows why series resonant circuits are little used. These circuits are very unselective since their Q is lowered considerably by the series resistance of the vacuum tube.

Distributed capacity in an inductor. In an inductor without a shunt capacitor, there is one particular frequency at which the inductor resonates. This is due to the distributed capacitance between the windings of the inductor. In circuit operation, you must consider distributed capacitance as in parallel with the inductor and thus as forming a resonant circuit. Since such a resonant circuit is undesirable, various methods are employed in winding coils to minimize their distributed capacitance.

Resonance at radar frequencies. Since radar operates at micro-wave frequencies, it is necessary that inductances and capacitances in resonant circuits be very small. A resonant circuit in radar uses a horseshoe-shaped bar or a dough-nut-shaped cavity instead of a condenser and a coil. However, the principles of resonance explained before are applicable to radar, no matter how high the frequencies or what physical form the resonant circuit may take.

AC NETWORK THEOREMS

While Ohm's law and Kirchoff's law are the basic laws for solving network problems (combinations of circuit elements), there are other methods which are much more direct and shorter. Often Ohm's law or Kirchoff's law leads to round-about solutions that necessitate your finding quantities in which you are not directly interested. For example, in the wien-bridge circuit previously solved in this chapter, it was necessary to find the total current before calculating the total impedance. In such circuits it would be much easier if you were able to find the impedance directly. One method which makes it possible to find impedance directly uses T and Pi networks transformations. These transformations enable you to find the impedance of the network by converting it to an equivalent impedance. Another method which permits direct calculation is Thevenin's theorem, which enables you to find current in a particular circuit directly. A third method, the maximum power transfer theorem, shows the relation which exists between the load impedance and the generator impedance during the most efficient operation.

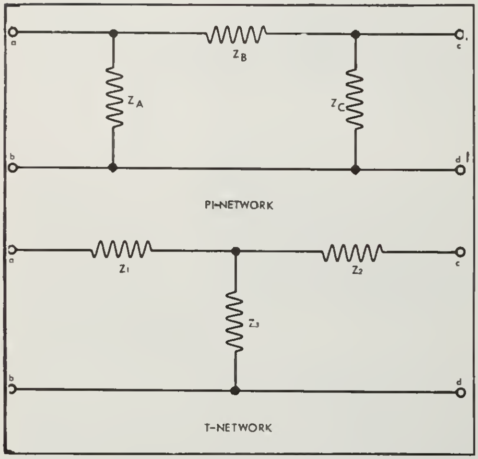
The only limitation to the validity of each of these methods is that the impedances of the network be *linear* and *bilateral*. A linear impedance is one which is independent of the amount of current flowing through it, and a bilateral impedance is one which conducts electricity well in both directions. The latter eliminates such elements as vacuum tubes, copper-oxide rectifier units, and saturable inductors.

Equivalence of T- and Pi-Networks

T- and Pi-networks get their names from their forms. In power systems, networks similar to these networks are called *WYE* and *Delta* connections.

A basic consideration in the equivalence of T- and Pi-networks is the theorem which states that in any network at any single frequency, a three-element T-structure can be interchanged with a three-element Pi-structure, provided certain relations are maintained. The following discusses these relations.

In this connection, suppose the Pi-network shown above is given, and it is necessary to find the equivalent T-network. The first consideration is the requirements for equivalence of these networks. In order for them to be equivalent, the impedance must be the same for both, looking into terminals a, b, c, and d in pairs. Equating



Pi- and T-Networks

the impedances, looking into terminals ab, ac, and cd in one network, to the corresponding impedances of the other, while all other terminals are disconnected produces the following equations:

Looking into ab, $$Z_1+Z_3=\frac{Z_A(Z_B+Z_C)}{Z_A+Z_B+Z_C} \quad (1)$$

Looking into oc, $$Z_1+Z_2=\frac{Z_B(Z_A+Z_C)}{Z_A+Z_B+Z_C} \quad (2)$$

Looking into cd, $$Z_2+Z_3=\frac{Z_C(Z_A+Z_B)}{Z_A+Z_B+Z_C} \quad (3)$$

You may solve these equations for Z_1, Z_2, and Z_3 in terms of Z_A, Z_B, and Z_C, or vice versa. Notice that only three equations are needed since there are three unknowns to find. Now let $Z_A+Z_B+Z_C=S$. Then to solve for Z_1, subtract equation (3) from equation (1) as follows:

$$Z_1-Z_2=\frac{Z_AZ_B+Z_AZ_C-Z_BZ_C-Z_AZ_C}{S}$$

$$Z_1-Z_2=\frac{Z_AZ_B-Z_BZ_C}{S} \quad (4)$$

To equation (4), add equation (2):

$$2Z_1=\frac{Z_AZ_B+Z_BZ_C+Z_AZ_B-Z_BZ_C}{S}=\frac{2Z_AZ_B}{S} \quad (5)$$

$$Z_1=\frac{Z_AZ_B}{S}=\frac{Z_AZ_B}{Z_A+Z_B+Z_C} \quad (6)$$

Going through the same procedure for the other combinations of equations (1), (2), and (3) gives,

$$Z_2=\frac{Z_BZ_C}{Z_A+Z_B+Z_C} \quad (7)$$

$$Z_3=\frac{Z_AZ_C}{Z_A+Z_B+Z_C} \quad (8)$$

In obtaining the equations for a T-to Pi-transformation, multiply equations (6) and (7), equations (7) and (8), and equations (6) and (8) as follows:

$$Z_1Z_2 = \frac{Z_AZ_CZ_B^2}{S^2} \qquad (9)$$

$$Z_2Z_3 = \frac{Z_AZ_BZ_C^2}{S^2} \qquad (10)$$

$$Z_1Z_3 = \frac{Z_BZ_CZ_A^2}{S^2} \qquad (11)$$

Adding these three equations and factoring the right-hand side gives,

$$Z_1Z_2+Z_2Z_3+Z_1Z_3 = \frac{Z_AZ_CZ_B^2+Z_AZ_BZ_C^2+Z_BZ_CZ_A^2}{S^2}$$

$$= \frac{Z_AZ_BZ_C(Z_A+Z_B+Z_C)}{S^2}$$

$$= \frac{Z_AZ_BZ_C}{S} \qquad (12)$$

Now, by making use of equation (7) in equation (12) produces the equation,

$$Z_1Z_2+Z_2Z_3+Z_1Z_3 = Z_2Z_A$$

Solving for Z_A,

$$Z_A = \frac{Z_1Z_2+Z_2Z_3+Z_1Z_3}{Z_2} \qquad (13)$$

In a like manner, by using equations (8) and (6), Z_B and Z_C equal respectively,

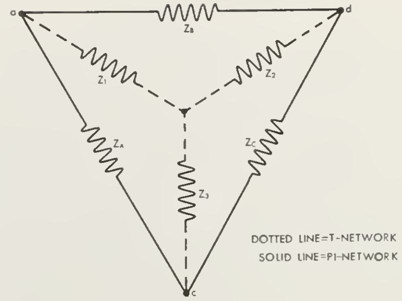
$$Z_B = \frac{Z_1Z_2+Z_2Z_3+Z_1Z_3}{Z_3} \qquad (14)$$

$$Z_C = \frac{Z_1Z_2+Z_2Z_3+Z_1Z_3}{Z_1} \qquad (15)$$

From the preceding information, it is possible to make the transformation in either direction. You can prove that the networks on the preceding page are equivalent by connecting a generator across terminals ab, attaching a load Z_R across terminals cd and calculating current flow through the load in each case. This shows that the currents are equal and proves that the networks are equivalent.

The preceding transformation formulas are not easily remembered. Therefore it is well to fix the processes of transformation in your mind by rules. For this purpose, the T- and Pi-networks are redrawn as shown above at the right. The rules are as follows:

RULE 1. PI-TO-T-TRANSFORMATION. To find any of the elements of the T-network, multiply the *two adjacent impedances* of the Pi-network and divide by the *sum of the impedances* of the Pi-network. For example, suppose you want to find Z_2. Then divide the product of Z_BZ_C by $Z_A+Z_B+Z_C$. The result is Z_2. Use similar procedure for finding Z_1 and Z_3.

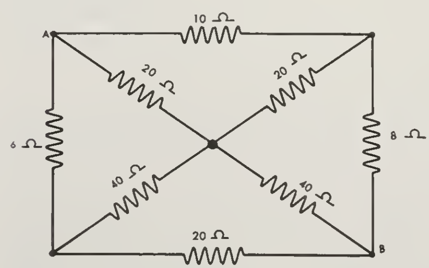


T- and Pi-Networks Redrawn

RULE 2. T-TO-PI TRANSFORMATION. To find any of the elements of the Pi-network, find the *sum of the impedance products* of the T-elements two at a time and *divide* by the impedance element of the T-network which is *opposite* the desired Pi-element. For example, suppose you desire to find Z_C when Z_1, Z_2, and Z_3 are known. To do this, first, form the sum $Z_1Z_2+Z_2Z_3+Z_1Z_3$ and then divide by Z_1, which is opposite Z_C. To obtain Z_A and Z_B follow a similar procedure.

Example

Problem. In the network of resistors illustrated just below, using the values indicated, find the impedance between points AB.



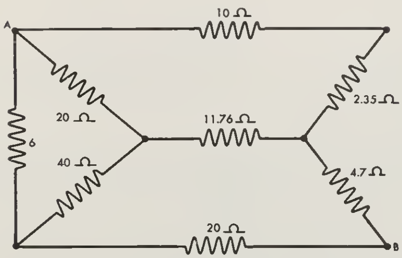
Circuit for Problem

Solution. As there are many methods for solving this problem, it is arbitrary as to which three elements you transform first, or which kind of transformation you make. But, as you gain experience, you will find that by proper selection, you can reduce the number of transformations necessary to reduce the circuit to series and parallel circuits. In this case, transform the Pi-network made up of the 20-, 40- and 8-ohm resistors to a T-network. Applying rule 1, produces the following facts:

$$\frac{20\times 8}{68}=2.35 \text{ ohms}$$

$$\frac{40\times 8}{68}=4.7 \text{ ohms}$$

$$\frac{20\times 40}{68}=11.76 \text{ ohms}$$



First Equivalent Circuit

After transformation the result is as illustrated by the first equivalent circuit above but it still does not lend itself to immediate calculation of the impedance between points AB. Therefore, transform the Pi-network consisting of the 6-, 40- and 20-ohm resistors into a T-network. To do this perform the following calculations:

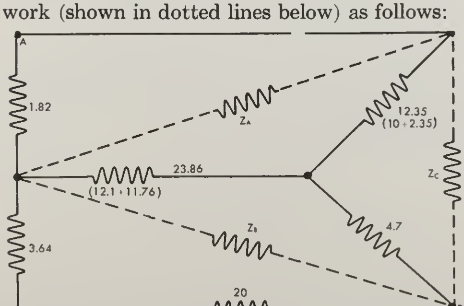
$$\frac{6\times 20}{66}=1.82 \text{ ohms}$$

$$\frac{6\times 40}{66}=3.64 \text{ ohms}$$

$$\frac{20\times 40}{66}=12.1 \text{ ohms}$$

The resulting equivalent circuit is as shown below in solid lines (the 12.1 ohms has been combined with the 11.76 ohms from the circuit shown above).

Transform the T-network composed of 23.86 ohms, 12.35 ohms, and 4.7 ohms into a Pi-network (shown in dotted lines below) as follows:



Second Equivalent Circuit

$$Z_A=\frac{12.35(4.7)+4.7(23.86)+23.86(12.35)}{4.7}$$

$$=\frac{58.0+112.1+295.0}{4.7}=\frac{465.1}{4.7}=99.0 \text{ ohms}$$

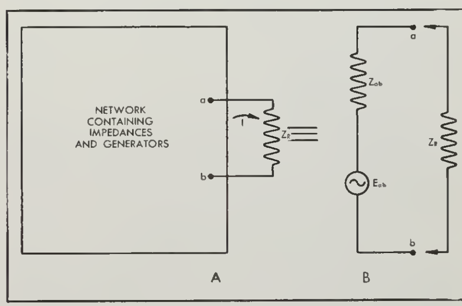
$$Z_B=\frac{465.1}{12.35}=37.7 \text{ ohms}$$

$$Z_C=\frac{465.1}{23.86}=19.5 \text{ ohms}$$

Combining 1.82 ohms which is in parallel with Z_A yields 1.79 ohms. Z_B in parallel with 23.64 ohms gives 14.5 ohms. These two parallel circuits are in series with each other and are also in parallel with Z_C. Therefore, the impedance between AB is 8.86 ohms.

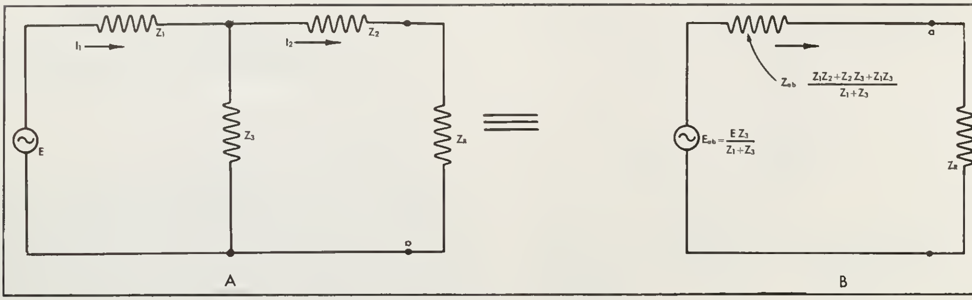
Thevenin's Theorem

Thevenin's theorem states that the current in any impedance Z_R connected to two terminals of a network is the same as if Z_R were connected to a simple generator whose generated voltage is the open-circuited voltage at the terminals of the network, and whose impedance is the impedance of the network looking back from the terminals, with all generators replaced by impedances equal to the internal impedances of these generators.



Illustrating Thevenin's Theorem

In the diagrammatic representation of Thevenin's theorem just above at A, E_{ab} is the voltage measured at terminals ab with Z_R removed, and Z_{ab} is the impedance measured back from the terminals *ab*. In measuring this impedance, assume that the generators are replaced by impedances equal to their internal impedances. When the load, Z_R is connected into the equivalent generator at B, the current that flows through Z_R is the same as that which flowed when Z_R was connected to the original terminals ab. By Thevenin's theorem, the network shown at A can be reduced by T- and Pi-transformations to a single T- or Pi-network and a single generator as shown at A in the next illustration.



Equivalent Circuits Based on Thevenin

Proof of theorem. The proof of Thevenin's theorem is based on showing that the current flowing through Z_R in circuit B in the above illustration is the same as the current which flows through Z_R when it is connected to the equivalent generator.

From the circuit at A in the same illustration, the current I_2 is equal to,

$$I_2 = I_1 \frac{Z_3}{Z_3 + Z_2 + Z_R} \quad (1)$$

The total current flow I_1 through the generator E is equal to,

$$I_1 = \frac{E}{Z_1 + \dfrac{Z_3(Z_2+Z_R)}{Z_3+Z_2+Z_R}} \quad (2)$$

Substituting equation (2) in equation (1) gives I_2 in terms of the impedances and the generator voltage. Thus,

$$I_2 = \frac{EZ_3}{Z_1Z_2 + Z_2Z_3 + Z_1Z_3 + Z_R(Z_1+Z_3)} \quad (3)$$

Now find the equivalent generator according to Thevenin's theorem in the following manner: First, remove Z_R from the terminals ab. This makes the open-circuited voltage equal to,

$$E_{ab} = E\frac{Z_3}{Z_1+Z_3} \quad (4)$$

This is the voltage of the equivalent generator. Now find the impedance of the equivalent generator by replacing the generator illustrated above at A with its internal impedance (which is zero in this case) and measuring the impedance, looking in terminals ab. This impedance is equal to

$$Z_{ab} = Z_2 + \frac{Z_1Z_3}{Z_1+Z_3}$$

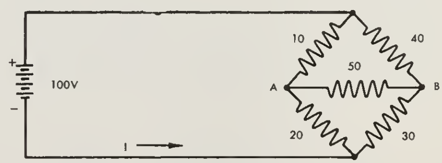
$$Z_{ab} = \frac{Z_1Z_2 + Z_2Z_3 + Z_1Z_3}{Z_1+Z_3} \quad (5)$$

After finding the equivalent generator by Thevenin's theorem, connect Z_R to this generator, and calculate the current flow through it by Ohm's law as follows:

$$I = \frac{\dfrac{EZ_3}{Z_1+Z_3}}{\dfrac{Z_1Z_2+Z_2Z_3+Z_1Z_3+Z_R}{Z_1+Z_3}}$$

$$I = \frac{EZ_3}{Z_1Z_2+Z_2Z_3+Z_1Z_3+Z_R(Z_1+Z_3)} \quad (6)$$

On comparing equations (3) and (6), you can see that they are equal, thus proving Thevenin's theorem.



Example

Problem. In the bridge circuit just above using the values indicated, find the current flow through the 50-ohm resistor assuming that the internal resistance of the battery is zero.

Solution:

1. First, find the open-circuited voltage between points AB.

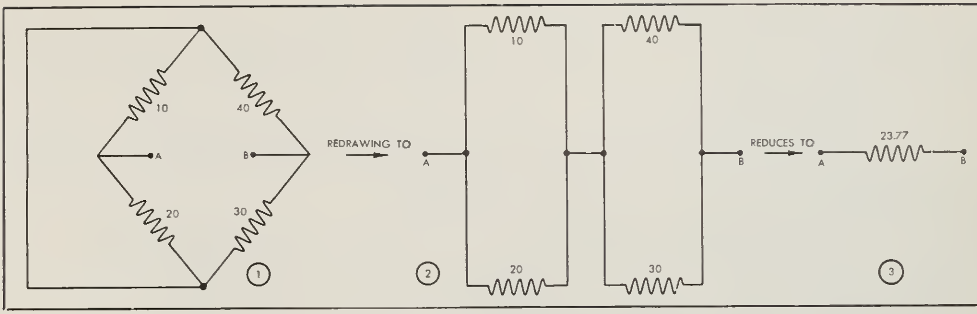
The drop across the 20-ohm resistor

$$= 100 \times \frac{20}{30} = 66.67 \text{ volts}$$

The drop across the 30-ohm resistor

$$= 100 \times \frac{30}{70} = 42.8 \text{ volts}$$

The drop between AB = 66.67 − 42.8 = 23.87 volts. This also is the voltage of the equivalent generator.



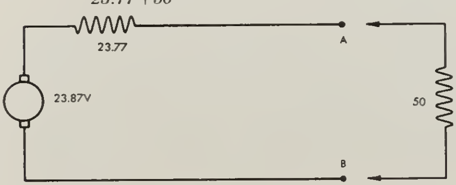
Equivalent Circuits of Bridge Circuit

2. Next, find the impedance looking into points AB (with the 50-ohm resistor disconnected as shown at 1 above) with the generator replaced with its internal impedance (0 ohms). Redrawing the circuit of 1 produces the parallel combinations shown at 2 of which the series resistance equals 23.77 ohms (shown at 3).

3. Set up the equivalent generator circuit illustrated directly below. The voltage of this equivalent generator is the open-circuited voltage at points AB and its internal impedance is the impedance looking in at points AB.

4. Lastly, connect the 50-ohm resistor to the terminals of the equivalent generator and calculate the current flow as shown directly below. Thus,

$$\frac{23.87}{23.77+50}=0.324 \text{ amperes}$$



Thus, the current flow through a resistor can be calculated without setting up simultaneous equations whose solution, at best, is tedious. It is suggested that you study this example until you thoroughly understand each step. The four steps outlined are general and apply to all similar problems.

Maximum Power Transfer Theorem

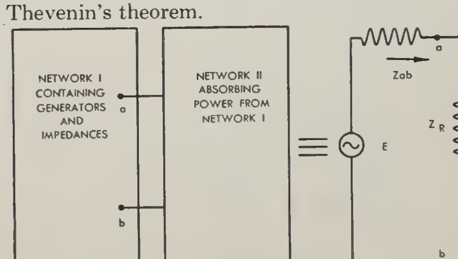
The maximum power transfer theorem gives the conditions under which maximum power may be delivered to a load. A given emf source with a certain internal impedance does not furnish the same power or energy to different loads. It so happens that only when the impedance of the emf source is equal to that of the load is maximum energy delivered to the load. Even then the efficiency of transfer is only 50%. In communication circuits where the *total* power available is small, it is very important to match the impedance of the load to the source of power. Radio and radar, for example, deal with relatively weak signals; it is, therefore, important to pass as much of the signal as possible from one circuit to another.

For a generator which feeds through a network to supply maximum power to an impedance Z_R at the terminals ab, Z_R must comply with the conditions required in the following cases:

1. When the phase angle of Z_R is unrestricted, Z_R must be the conjugate of the impedance Z_{ab} looking into the network at the terminals ab.

2. When the phase angle of Z_R is fixed, then the absolute values of Z_R and Z_{ab} must be equal.

The illustration below indicates that any network consisting of two or more generators that supply power to another network may be reduced to an equivalent circuit containing only a single generator of certain internal impedance Z_{ab} feeding a load Z_R. This is simply the application of Thevenin's theorem.



Maximum Power Transfer Theorem

The following summarizes the principal points of the maximum power transfer theorem:

1. When Z_{ab} is purely resistive, then Z_R must be entirely resistive and equal to Z_{ab} for maximum power transfer. Thus,

$$Z_{ab} = R_{ab} = Z_R = R_R$$

2. When Z_{ab} is complex,

$$Z_{ab} = R_{ab} - jX_{ab}$$

Thus, Z_R must be complex and the conjugate of Z_{ab}. In other words,

$$Z_R = R_R + jX_R$$

where $R_R = R_{ab}$ and $X_{ab} = X_R$ for maximum power transfer.

3. When it is not possible to vary Z_R, both as to its resistive component or its reactive component, the transfer of power will be best when the absolute value of Z_{ab} and Z_R are equal. In the language of mathematical symbols, the best transfer of power takes place when

$$|Z_{ab}| = |Z_R|$$

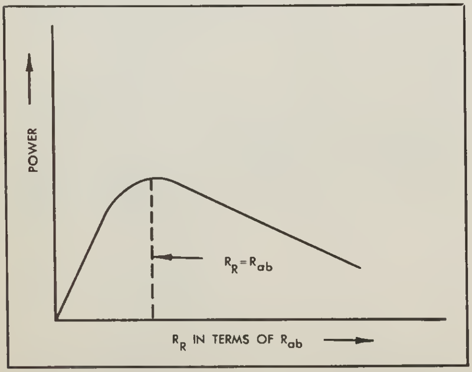
Proof of theorem: The rigorous proof of maximum power transfer requires differential calculus, the application of which is beyond the scope of this course. However, the following elementary presentation should be helpful. In this connection consider the condition in which Z_{ab} and Z_R are pure resistances. Then from the preceding illustration, the current through R_R equals

$$I = \frac{E}{R_{ab} + R_R} \qquad (1)$$

and the power in the load is

$$P = I^2 R_R = \frac{E^2 R_R}{(R_{ab} + R_R)^2} \qquad (2)$$

To determine for what value of R_R the power, as given by equation (2), is a maximum, it is necessary to make a graph of power versus R_R in terms of R_{ab} as shown directly below.



Power Delivered to R_R as R_R is Varied

From this graph you can see that the power to the load increases until the load resistance is equal to the internal resistance of the generator, and then begins to decrease gradually to zero.

Now, consider the case in which Z_{ab} and Z_R have reactance as well as resistance. In this case the absolute value of current is expressed by the equation,

$$|I| = \frac{E}{Z_{ab} + Z_R} = \frac{E}{\sqrt{(R_{ab} + R_R)^2 + (X_{ab} + X_R)^2}} \qquad (3)$$

and the power delivered is equal to,

$$P = I^2 R_R = \frac{E^2 R_R}{(R_{ab} + R_R)^2 + (X_{ab} + X_R)^2} \qquad (4)$$

From inspection it is clear that, as far as X_R is concerned, the power in equation (4) is a maximum when $X_{ab} = -X_R$. That is, when Z_{ab} is inductive, Z_R must be capacitive and vice versa. Now when this condition is fulfilled, equation (4) reduces to equation (2). Therefore, $R_R = R_{ab}$ is the value of R_R which makes the power a maximum. Thus, the circuit is series resonant, and the only opposition to the current is the total resistance of Z_R and Z_{ab}. At maximum power transfer, the total resistance is $2R_{ab}$.

Example.

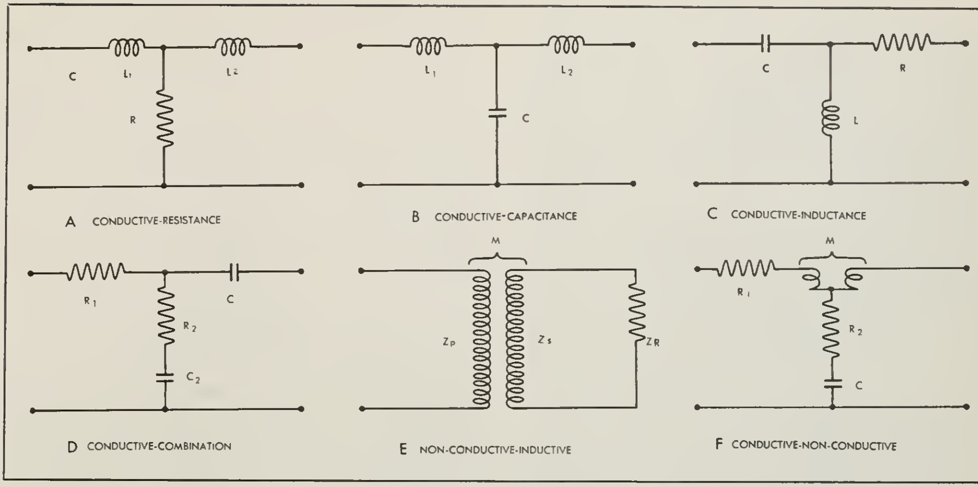
Problem. Using the values indicated in the bridge circuit shown on page 2-72, find the value of the resistor across the points ab which will absorb the greatest amount of power from the 100-volt battery.

Solution. By applying Thevenin's theorem, you can find the equivalent generator circuit as shown at top of page 2-73. Since the internal impedance of the equivalent generator is 23.77 ohms, the resistor which must be placed across AB must have the same value in order to absorb the greatest amount of power from the battery.

COUPLED CIRCUITS

Interconnected circuits and nearby circuits which transfer electrical energy from one circuit to the other are called *coupled circuits*. There are three general types of coupling—conductive coupling, nonconductive coupling, and combination coupling. Combination coupling includes both conductive and nonconductive coupling.

In the illustration showing various types of coupling, notice that coupling between circuits can be effected by resistors, coils, or capacitors, either singly or in combination. At A, B, C, and D are examples of conductive coupling. At E is an



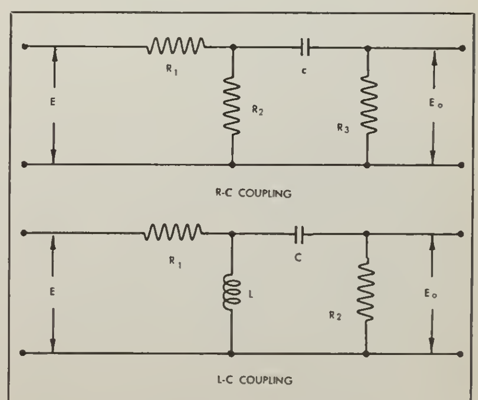
Various Types of Coupling

example of nonconductive coupling. Usually, this type of coupling is called *indirect* coupling, because the circuit elements are not connected. Other names for indirect or nonconductive coupling are *inductive*, *magnetic*, or *transformer* coupling. At F is an example of a conductive-nonconductive or combination type coupling.

This manual deals only with two types of coupling conductive and transformer, since they are the ones most widely used in radio and radar. For the same reason, only two types of conductive coupling are discussed. They are conductive-resistance coupling, better known as resistance-capacitance or *RC* coupling, and conductive-capacitance coupling, which is commonly referred to as inductance-capacitance or *LC* coupling.

RC Coupling

The circuit labeled RC coupling is typical of networks used for connecting two amplifier stages in radio receivers and in other equipment employing RC coupled amplifiers. An important consideration in RC coupled circuits is the comparison between the relative magnitude and phase of the output voltage, E_0 (the voltage across R_3), and the impressed voltage, E. When you know the circuit constants, the frequency, and the magnitude of E, you can calculate the voltage, E_0, by means of the complex algebra. However, before discussing the calculations involved, it is well to observe the following facts with reference to this circuit. The higher the frequency of the impressed voltage, the smaller the reactance of C and the more nearly the circuit is like a series-parallel resistive combination. On the other hand, the lower the frequency, the greater the reactance of C. This causes a smaller proportion of the voltage across R_2 to appear across R_3 as the output voltage, E_0, and a greater phase difference to exist between the voltages, E and E_0, with E_0 leading. You can see, therefore, that there is a wide variation in the magnitude of the output voltage and the phase relation between the input and output voltages. In practical circuits which employ RC coupling, the



RC and LC Coupled Circuits

range of variation is further widened by small distributed capacitances called shunt capacitances, which appear across R_2 and R_3. (Shunt capacitances are considered in detail in a later chapter.) At the higher frequencies, these capacitances become increasingly important because they lower the output voltage and cause E_o to lag behind E.

The following is the derivation of the general equation for the voltage across R_3 (E_o).

By letting the parallel combination R_2, C, and R_3 be represented by Z, Z equals,

$$Z = \frac{R_2\left(R_3 - j\frac{1}{\omega c}\right)}{R_2 + R_3 - j\frac{1}{\omega c}}$$

The voltage across R_2 equals,

$$E_2 = \frac{EZ}{R_1 + Z} \qquad (1)$$

The output voltage E_o equals,

$$E_o = \frac{E_2R_3}{R_3 - j\frac{1}{\omega c}} = \frac{EZR_3}{(R_1+Z)\left(R_3 - j\frac{1}{\omega c}\right)} \qquad (2)$$

Substituting for Z in equation (2) gives,

$$E_o = \frac{E\left[\dfrac{R_2\left(R_3 - j\frac{1}{\omega c}\right)}{R_2 + R_3 - j\frac{1}{\omega c}}\right]R_3}{\left[R_1 + \dfrac{R_2\left(R_3 - j\frac{1}{\omega c}\right)}{R_2 + R_3 - j\frac{1}{\omega c}}\right]\left(R_3 - j\frac{1}{\omega c}\right)} \qquad (3a)$$

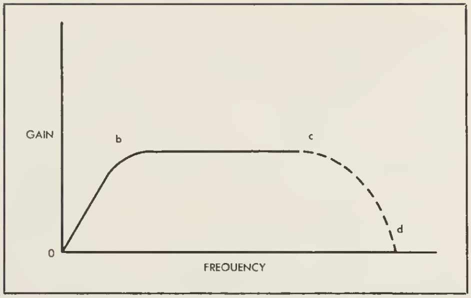
$$E_o = \frac{ER_2R_3}{R_1R_2 + R_1R_3 + R_2R_3 - j\frac{1}{\omega c}\,(R_1 + R_2)}$$

Forming the ratio $\frac{E_o}{E}$ and calling it the gain, then from equation (3a),

$$Gain = \frac{E_o}{E} = \frac{R_2R_3}{\sqrt{(R_1R_2 + R_1R_3 + R_2R_3)^2 + \left(\frac{R_1 + R_2}{\omega c}\right)^2}} \qquad (3b)$$

Close examination shows that equation (3a) or (3b) bears out the qualitative statements made on page 2-75 with reference to RC coupled circuits. Making the ω term large, virtually eliminates the j component in these equations and leaves in them only resistive terms. In this condition, the gain approaches unity for all values of R_2 and R_3 which are large with respect to R_1. In practice, however, the size of R_2 is limited to about 100K and R_3 to about 1 megohm.

Notice in the chart showing the approximate frequency response for RC coupled circuits that



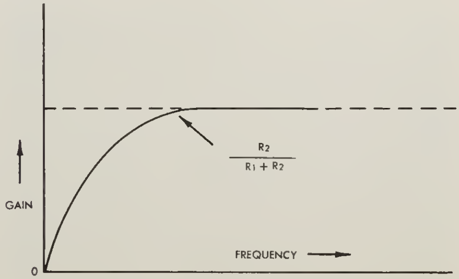
Approximate Frequency Response for RC Coupled Circuits

as the frequency increases from zero to mid-frequency range, gain increases and then levels off at a point at which capacitive reactance is negligible. The dotted portion of the curve from c to d represents the dropping off of gain. In RC coupled amplifier circuits, this drop is due to the shunting effect caused by distributed capacitances in the vacuum tubes and wires.

LC Coupling

Sometimes LC coupling is employed to couple amplifier stages. With reference to the LC coupling circuit shown on page 2-75, coil L may employ either an iron core or an air core, depending on whether the signals being amplified are in the audio range (30-15,000 cps) or in the radio frequency range. In practical circuits, L not only has some resistance, but also is shunted by distributed capacitances in the coil windings and by the plate-to-cathode capacitance of the preceding tube.

An important fact to consider in LC coupling is the effect that high- and low-frequency voltage sources have on the output voltage. Remember that at high frequencies the reactance of a coil is high while that of a condenser is low. Now for



Approximate Frequency Response for LC Circuits practical purposes you can consider that L and R_2 in the LC coupled circuit are in parallel. The voltage output then will equal approximately the voltage across the coil. Thus, the higher the frequency, the greater the magnitude of E_o. Further, the voltage E_o will lead E. At lower frequencies, the coil has small reactance, but the condenser has a large reactance. Therefore, the voltage across the coil will equal the voltage across the voltage divider, C and R_2. In this case, voltage E_o leads the input voltage E as before, but by a larger angle.

With this information in mind, you can obtain the output voltage in terms of the circuit constants for the high frequency case in the following manner:

Since the reactance of the condenser is negligible at these frequencies, you can treat L and R_2 as in parallel. Therefore, if you let Z represent the impedance of L and R_2 parallel, Z is equal to

$$Z = \frac{(j\,\omega\,L)R_2}{R_2 + j\,\omega\,L}$$

Then E_o equals,

$$E_o = \frac{EZ}{R_1 + Z} = \frac{E(j\omega L)R_2}{(R_2 + j\omega L)\left[R_1 + \frac{(j\omega L)R_2}{R_2 + j\omega L}\right]} \qquad (1)$$

$$= \frac{j\omega L R_2 E}{R_1 R_2 + j\omega L R_1 + j\omega L R_2}$$

$$= \frac{j\omega L R_2 E}{R_1 R_2 + j\omega L(R_1 + R_2)}$$

The absolute value of the gain $\frac{E_O}{E}$ equals

$$\frac{E_o}{E} = \frac{\omega L R_2}{\sqrt{(R_1 R_2)^2 + \omega^2 L^2 (R_1 + R_2)^2}} \qquad (2)$$

If the second term under the square root sign in equation (2) is much larger than $(R_1 R_2)^2$, then the gain equals,

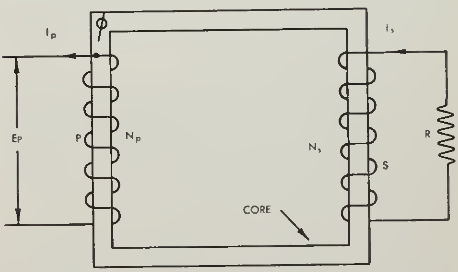
$$\frac{R_2}{R_1 + R_2}$$

an expression which shows that the gain approaches unity when $R_2 > R_1$. In the rough graph of frequency response on the preceding page, notice that gain approaches $\frac{R_2}{R_1 + R_2}$. This value in general is close to unity, since for practical circuits R_2 is large with respect to R_1. The shape of the response curve depends on the relative values L, C, R_1 and R_2. In actual cases, the curve may have points of maxima and minima because of the resonance of the coil with its distributed capacitance.

Transformer Coupling

A transformer consists of two coils which couple energy from one to the other by mutual inductance. Depending on the frequency of operation and the particular piece of equipment in which it is installed, a transformer may employ an iron, magnetic alloy, or an air core. The coil that is connected to the AC source is called the *primary* winding, and the coil that is connected to the load is called the *secondary* winding. The relative number of turns of the windings depends on whether the voltage induced in the secondary is to be larger or smaller than the voltage applied to the primary; the actual number of primary turns depends on the magnetic properties of the core and the frequency.

In both radio and radar, transformers generally are classed as power, audio, and radio frequency transformers. Power and audio transformers employ an iron core, or some alloy which has the desired magnetic properties. In ordinary broadcast receivers, the RF transformers may or may not contain iron cores. In radar sets, powered iron cores are commonly employed in transformers.



Simple Transformer

The simple transformer shown is a core type transformer in which the primary is wound on one leg of the core, and the secondary on the opposite leg. To reduce eddy current losses, the transformer core usually is constructed of laminated sheets instead of a solid piece.

When the primary current flows momentarily in the direction of the arrow labeled Ip, the magnetic flux ϕ flows in the direction indicated. To determine this direction, use the left-hand rule, which states that when the fingers of the left hand are placed so that they encircle the primary winding and point in the direction of electron flow, the thumb will point to the north pole of the electromagnet and the direction of flux flow.

According to Lenz's law, whenever flux (ϕ) is varying and links a secondary winding in a transformer, a voltage is induced with a polarity that opposes the primary flux when current flows in the secondary. The current in the secondary flows in the direction represented by the arrow labeled Is in the illustration, and the flux set up by the secondary current, opposes the initial flux produced by current flow in the primary winding.

Since a transformer requires a varying flux for operation, obviously it cannot be used with a DC source of voltage. When a steady voltage is suddenly connected to the primary, a voltage is induced in the secondary for an *instant*, but none at all after this transient period, for then current is not changing through the primary.

An iron core transformer is a very efficient piece of electrical equipment. Any time power is supplied to the primary, a magnetic field is set up. It links the secondary windings and induces a voltage in it and causes current to flow in the load. In other words, power is transferred to the secondary. This transfer of power is greatest when all the flux set up by the primary links all the turns of the secondary.

Because of the large permeability of iron (and some of its alloys), its reluctance to the path of the magnetic field in a transformer is much smaller than air, and the amount of flux lost (leakage) to air therefore is small. This accounts for efficiencies as high as 99% in iron-core transformers. The effectiveness of the coupling between the primary and the secondary is expressed in terms of the coefficient of coupling K. The coefficient of coupling is defined as the ratio of flux that links the two coils to the flux created by the current flow through the primary.

DERIVATION OF TRANSFORMER FORMULAS. The preceding paragraphs dealt with the theory and operation of transformers. Now it is well to study the following formulas for an ideal transformer (one assumed to have no flux leakage and no power loss in either the primary or the secondary).

For an ideal transformer, the formula for the instantaneous voltage which is induced across the primary by the changing flux is,

$$e_p = -N_p \frac{\Delta\phi}{\Delta t} \times 10^{-8} \text{ volts} \qquad (1)$$

As the same flux links the secondary coil, the voltage induced in the secondary is,

$$e_s = N_s \frac{\Delta\phi}{\Delta t} \times 10^{-8} \text{ volts}$$

As dividing equation (1) by equation (2) gives,

$$\frac{e_p}{e_s} = \frac{N_p}{N_s} \qquad (3)$$

and since this relation is true for any instant, it is also true when the voltages are maximum. Thus,

$$\frac{E_{pm}}{E_{sm}} = \frac{N_p}{N_s} \qquad (3a)$$

where E_{pm} is the maximum primary induced voltage and E_{sm} is the corresponding secondary voltage.

Since $\frac{Emax}{2} = Eeff$, you can divide the numerator and the denominator of the left-hand member of equation (3) by $\sqrt{2}$.

This gives,

$$\frac{E_{pm}/\sqrt{2}}{E_{sm}/\sqrt{2}} = \frac{E_p}{E_s} = \frac{N_p}{N_s} \qquad (3b)$$

where E_p and E_s represent effective values of primary and secondary voltage respectively.

Equation (3b) is the expression of the relationship between the number of turns of the primary and secondary and the corresponding voltages in a transformer.

The power input at any instant in an ideal transformer equals the power output at that instant. That is,

$$i_p e_p = i_s e_s \qquad (4)$$

where i_p and i_s represent the instantaneous values of the primary and secondary currents.

Since this relationship is true for all values, it follows that it is true for maximum and for the effective values of voltage and current. Thus,

$$\frac{E_p}{E_s} = \frac{I_s}{I_p} \qquad (5)$$

Then from equation (3B) and equation (5), it follows that,

$$\frac{I_s}{I_p} = \frac{N_p}{N_s}. \qquad (6)$$

According to the maximum power transfer theorem discussed before in this chapter, it is necessary to match the impedance of two circuits to obtain the greatest transfer of energy. The simplest and most common circuit used for impedance matching is the transformer. There are a number of mathematical relationships between primary and secondary impedances and the corresponding turns in a transformer.

The impedance looking into the primary of a transformer is

$$Z_p = \frac{E_p}{I_p} \qquad (7)$$

The terms E_p and I_p are given by equations (3b) and (6) respectively. Substituting for these values in equation (7) and noting that E_s/E_s is Z_s, gives,

$$Z_p = \frac{E_p}{I_p} = \frac{E_s N_p/N_s}{I_s N_s/N_p} = Z_s \frac{N_p^2}{N_s^2} \qquad (8)$$

$$\text{or,} \quad \frac{Z_p}{Z_s} = \left(\frac{N_p}{N_s}\right)^2$$

From this equation, it is apparent that for a transformer with unity coupling, the impedance ratio is equal to the square of the turns ratio. For example, if R in the transformer shown 2-77 is 100 ohms, N_p is 100 turns, and N_s is 1000 turns, then the impedance looking into the primary is equal to 1 ohm. Another way of expressing this result is to say that the impedance *reflected* into the primary is 1 ohm.

The following shows that the phase angle between the primary voltage and the primary current in a transformer is equal to the phase angle between the secondary voltage and the secondary current:

The average power in the primary circuit is,

$$P_p = E_p I_p \cos \theta_p$$

The average power in the secondary is,

$$P_s = E_s I_s \cos \theta_s$$

where θ_p and θ_s are the phase angles of the primary and the secondary circuits.

Then, since $P_s = P_p$,

$$E_p I_p \cos \theta_p = E_s I_s \cos \theta_s$$

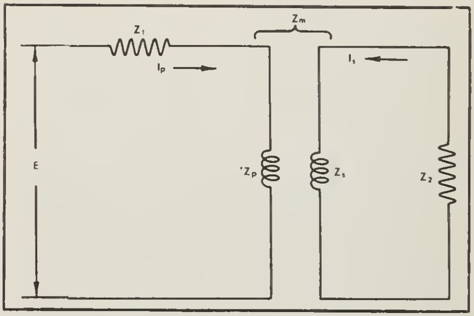
But from equation (5), $E_p I_p = E_s I_s$,

$$\text{Therefore, } \cos \theta_p = \cos \theta_s \qquad (9)$$
$$\text{and } \theta_p = \theta_s$$

So far, only an ideal transformer having unity coupling has been discussed.

Another important concept to consider is the general transformer circuit. In the following discussion, equations for calculating primary current, secondary current, input impedance, and reflected impedance are derived by Ohm's and Kirchoff's laws. In addition, the concept of *mutual impedance* is discussed.

In the typical transformer circuit shown above, the load (Z_2) is connected across the secondary and the impedance (Z_1) is in series with the primary. These impedances may be either resistive, reactive, or any combination of these elements. Z_p is the impedance looking into the primary



Typical Transformer Circuit

with the secondary open-circuited, and Z_s is the impedance of the secondary with the primary open. The mutual impedance is defined as $Z_m = j\omega M$ where M is the mutual inductance between the primary and the secondary.

Applying Kirchoff's laws first to the primary and then to the secondary gives,

$$E = I_p Z_1 + I_p Z_p + I_s Z_m \qquad (10)$$
$$O = I_s Z_2 + I_s Z_s + I_p Z_m \qquad (11)$$

The term $I_s Z_m$ is the voltage drop in the primary due to current flow in the secondary and the mutual impedance Z_m, while $I_p Z_m$ is the voltage drop in the secondary due to I_p in the primary.

In the two equations, (10) and (11) I_p and I_s are considered as unknowns. Solving them for I_p and I_s simultaneously give the equations,

$$I_p = \frac{E(Z_2 + Z_s)}{(Z_1 + Z_p)(Z_2 + Z_s) - Z_m^2} \qquad (12)$$

$$I_s = \frac{-EZ_m}{(Z_1 + Z_p)(Z_2 + Z_s) - Z_m^2} \qquad (13)$$

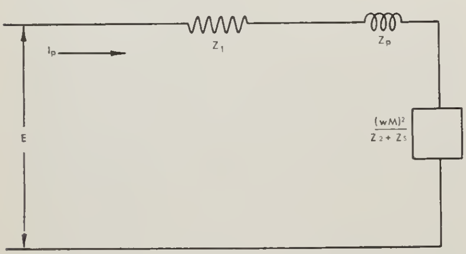
An important fact to understand is that these equations are *vector equations* and it therefore is necessary to use complex expressions to represent the various impedances. The negative sign of equation (13) indicates that the secondary current is 180° out of phase with the primary current.

In finding the impedance looking into the primary circuit from the source of applied voltage E, divide E by I_p. Thus,

$$Z = \frac{E}{I_p} = \frac{(Z_1 + Z_p)(Z_2 + Z_s) - Z_m^2}{Z_2 + Z_s} \qquad (14)$$
$$= Z_\cdot + Z_p - \frac{Z_m^2}{Z_2 + Z_s}$$

In this equation the term $\frac{Z_m^2}{Z_2 + Z_s}$ is the impedance reflected into the primary by the secondary. In generalized impedances, the sign preceding this term must always be negative and the denominator is usually written $R \pm jx$. The latter is expressed this way so that the reflected impedance will contain, (in addition to a resistive term of positive sign) a reactive term of which the sign depends upon the nature of the total reactance in the secondary circuit. The sign of this reflected reactance is always opposite to that of the total secondary reactance. In other words, when Z_2+Z_s is inductive, the reflected term is of the nature of $-jx$ and reduces the inductive reactance looking into the primary. With perfect coupling and with Z_2 as a pure resistance, this reflected term is just large enough to cancel completely the inductive reactance of the primary. The impedance looking into the primary, therefore, is in the nature of a pure resistance.

The preceding equations are applicable to all transformers and are frequently used in radar and radio. In power transformers, the coefficient of coupling is nearly equal to unity and produces a ratio of transformation that is very nearly equal to the turns ratio. In radio and radar circuits, particularly where the primary and secondary are tuned to resonance, the coefficient of coupling may be very small. As leakage losses therefore are often 95% or more, the effective ratio of transformation differs greatly from the ratio of turns on the transformer. This difference makes it better to analyze coupled circuits in radio and radar in terms of mutual inductance rather than in terms of turns ratio.



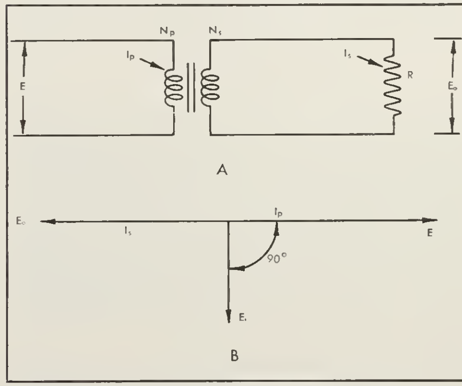
Equivalent Circuit Referred to the Primary

With equation (14) it is possible to construct the equivalent circuit of the typical transformer as illustrated above. In this case the impedance looking into the primary is the sum of three quantities—Z_1, Z_p, and the reflected term,

$$-Z_m^2/(Z_2+Z_s).$$

Voltage-Current Phase Relationship in Transformers

IRON CORE TRANSFORMERS WITH RESISTIVE LOADS. When a pure resistive load is connected to an iron core transformer, the impedance looking into the primary is equal to R/a^2 where $a=Ns/Np$. Since the circuit is purely resistive, the applied voltage E, and the current I_p are

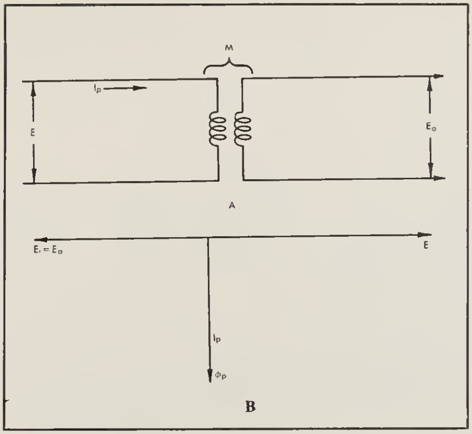


Iron Core Transformer with Resistive Load and Vector Diagram

in phase as shown above at B. The sinusoidally varying primary current Ip induces in the secondary the voltage E_i which lags the primary current by 90°. Since a load is connected, current can flow in the secondary, and since in actual transformers, the inductance is large in comparison to the resistive load, the current in the secondary lags the induced voltage by almost 90°. But since the secondary current also flows through R, the voltage (E_0) across R is in phase with the secondary current. This puts the voltage E_0, 180° out of phase with the primary voltage.

While in ideal or perfect iron core transformers, the impedance of the primary and secondary are infinitely large, in actual transformers, however, this is not the case. Therefore vector diagrams are to be considered only as approximations.

AIR-CORE TRANSFORMERS WITH SECONDARY OPEN. In the circuit of an air-core transformer with secondary open-circuited at A on the next page, assume that a sinusoidal current is flowing through the primary coils. Since there is coupling between primary and secondary, there is a voltage induced in the secondary, but since the secondary is open there is no current flow. This means that there is no reflection of impedance into the primary.



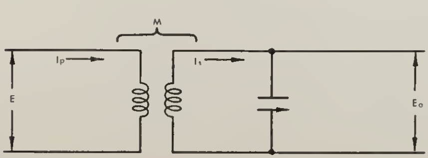
Air Core Transformer with Open Secondary and Vector Diagram

In the vector diagram for this circuit you can see that the current flowing in the primary lags the applied voltage E by 90°. Also the flux about the primary is in phase with the current and induces a voltage e_i in the secondary that lags the primary current by 90°. Since no current flows in the secondary, the only voltage across it is the induced voltage (E_0) which is 180° out of phase with E.

In practical circuits of this type, the windings have some resistance which tends to increase the phase lag between E_0 and E. Therefore consider the vector diagram shown above at B as representing the ideal case, but a good enough approximation for most practical circuits.

Air-Core Transformers with Secondary Tuned

In an air-core transformer with a tuned secondary, coupling is assumed to be small enough so that the reflected impedance into the primary is small. This means that the primary voltage looks into an inductive circuit and that E and I_p therefore are 90° out of phase as shown in the

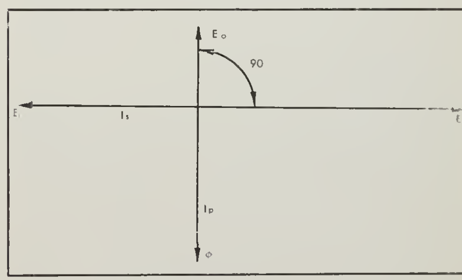


Air Core Transformer with Secondary Tuned

diagram below. The changing flux in the primary induces in the secondary, voltage (E_i) which lags the flux by 90°. Current then flows in the secondary, but it is in phase with E_i as the circuit is tuned. In other words, so far as the induced voltage is concerned, it is working into a *series* resonant circuit which consists of the secondary inductance and the condenser across the secondary. The output voltage E_o also is the voltage across the condenser. Since the phase relation of current and voltage for condensers is 90° with current leading the voltage, E_0 lags the secondary current (I_s) 90°. This places E_0 90° *ahead* of E.

This phase relationship between primary voltage and voltage across the condenser is used in radar altimeter sets to obtain a circular trace on a cathode ray tube.

The same phase relationships are obtainable when the primary as well as the secondary is tuned. In this case the vector diagram would be exactly like the diagram shown directly below.



Vector Diagram for Air Core Transformer with Secondary Tuned

Other phase relationships between E and E_0 exist when either the frequency of the applied voltages or the size of the tuning condenser are changed. For example, when the frequency is changed, the phase relationships between E, I_p, and E_i are still like the relationships shown above. When the frequency is below resonance, the induced voltage E_i in the secondary causes current to flow as before. But the secondary *appears* as a series circuit in which capacitive reactance predominates. The secondary current I_s therefore leads the induced voltage E_i and the voltage across the condenser E_0 lags I_s by 90°. This results in E and E_o being out of phase by more than 90° with E_o leading. In a like manner, when the input frequency is greater than the resonant frequency, the phase angle between E and E_o is less than 90° with E_o leading E.

Power Losses in Transformers

There are two kinds of poser losses in transformers—*copper loss* and *iron loss*. Copper loss is due to resistance in the primary and secondary windings and mathematically is equal to,

$$P_c = I_p^2 R_p + I_s^2 R_s$$

Iron loss is due to hysteresis and eddy currents. Hysteresis, as you recall, is the lagging of the magnetic flux (produced in a core) behind the magnetizing force. Energy is needed to overcome the frictional opposition which the molecules composing the magnetic material display to the orientation of the small elemental magnets twice during each cycle of alternating current or voltage. This energy must be supplied by the source of voltage and therefore constitutes a loss in power.

In advanced texts it is proved that hysteresis loss equals

$$P_h = kB^n fV10^{-7} \text{ watts}$$

where k and n are constants, depending on the kind and treatment of the core material; B is the maximum value of the flux density used, f is the cycles per second, and V is volume in cubic centimeters of the core.

Eddy currents are produced by the changing flux density inside the core material. The changing flux density induces an emf not only in the windings, but in the core itself. The induced emf's in turn produce extraneous currents which are called eddy currents. Eddy currents cause excessive heating of the core and effectively shield a winding from the inductive influence of the changing flux. They therefore represent a power loss which in the interest of efficiency should be reduced to a minimum.

Eddy current loss is equal to,

$$P_e = 1.6 \times 10^{-11} (fLB)^2 V \text{ watts}$$

where L is the thickness of the core plates. (For the meaning of the other terms refer to the equation for hysteresis loss).

Note in the equation that eddy-current losses are proportional to the square of the thickness of the core plates. This means that the thinner the laminations, the smaller the eddy-current loss, and explains why transformers are laminated.

The efficiency of a transformer depends on the amount of copper loss (Pc), hysteresis loss (Ph), and eddy-current loss (P_e). The smaller the sum of these losses, the more efficient the transformer. There are lower resistance materials usable for transformer windings but copper has the least resistantce commensurate with cost. Loss due to hysteresis is reduced by the use of core material which has a small hysteresis loop. The most practical and effective way to reduce eddy-current loss, as pointed out before, is to laminate the core material.

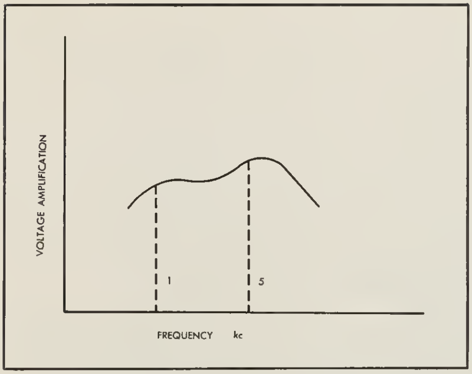
Types of Transformers

POWER TRANSFORMERS. The theory of operation and the formulas for power transformers already have been discussed to some extent. The following are some of the facts about the physical construction, voltages employed, and uses of power transformers.

The physical size of a transformer depends on the amount of power that it is to handle, the number of secondary windings that are needed, and the frequency at which it is to operate, and so on. Typical secondary voltages in transformers in conventional power supplies, such as those used in home broadcast receivers and radar receivers, are 350, 700, 6.3, and 5 volts. The higher voltages are rectified by vacuum tubes and then used as plate voltages for the various amplifier stages, while the small voltages are used directly to supply heater and filament power to vacuum tubes. In radar transmitters, transformers step pulse voltages up to as high as 12 to 16 kv. Physically, the size of these transformers is not as large as you might expect since they are not required to handle power continuously. In practice, a radar transmitter is operated only for extremely short periods of time, for example, ½ to 1-microsecond at intervals of several hundred microseconds. In home receivers, the usual input voltage to transformers ranges from 110 to 120 v at 60 cps. For airborne radar equipment, AC voltages vary from about 80 to 110 v at frequencies ranging from 400 to 1600 cps.

AUDIO TRANSFORMERS. Audio transformers transfer voltages over a wide range of audio frequencies, rather than at a single frequency like the power transformer. They have cores of iron or alloy, and are designed to carry a limited amount of direct current in the primary winding without affecting the audio frequency.

Since the ideal AF transformer has a flat frequency response, it can step-up or step-down voltages having different frequencies by the same amount. Practical, inexpensive AF transformers are designed with nearly a flat response for the mid-frequency range (100-4,000 cps), while at lower and higher frequencies their output voltages begin to fall off. Above 4 or 5 kc, a resonance



Typical Response Curve for Audio Transformers

peak appears because of distributed capacitance in the transformer windings and the input capacitance of vacuum tubes to which it is connected.

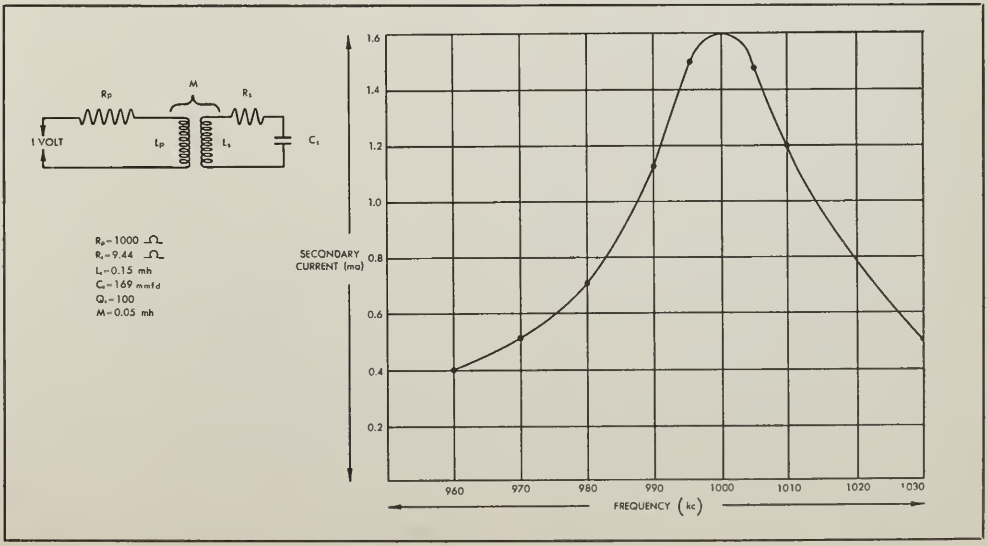
Audio frequency transformers are constructed more or less like power transformers. However, greater care is exercised in the choice of core material and special precautions are taken to reduce capacitances between the primary and secondary windings as well as between the turns of each winding.

Other characteristics of AF transformers are coefficients of couplings ranging from 0.99 to about 0.999, a primary inductance of 50 to 100 henrys, and a secondary inductance that is greater than the primary inductance by the square of the turns ratio. (The typical turns ratio is 3:1.) The formulas for power transformers are applicable to audio transformers over the flat portion of the frequency range.

Audio-frequency transformers are used in public address systems and in the audio stages of broadcast and communications receivers for matching the output of the last stage of this equipment to the loud speakers.

Tuned transformers. A tuned transformer is a radio frequency transformer in which a condenser is placed across either the primary or the secondary or both. Transformers are tuned whenever they are required to display a frequency discriminator effect—that is, pass a certain frequency or a band of frequencies. The type of response curve desired determines whether both the primary and secondary are tuned.

Below is a transformer which has an untuned primary and a tuned secondary. It is typical of the transformers employed in transformer-coupled radio-frequency amplifiers. Usually R_p is the plate resistance of a vacuum tube and is very large in comparison to the inductive reactance of



Response Curve with Secondary Tuned

the primary. The curve shown is the graph of secondary current versus frequency of the voltage applied to the primary. This curve has the same shape as an ordinary resonance curve, but has a lower Q than the secondary circuit curve alone.

The equation for the coupled impedance which explains the electrical characteristics of a transformer with an untuned primary and tuned secondary is,

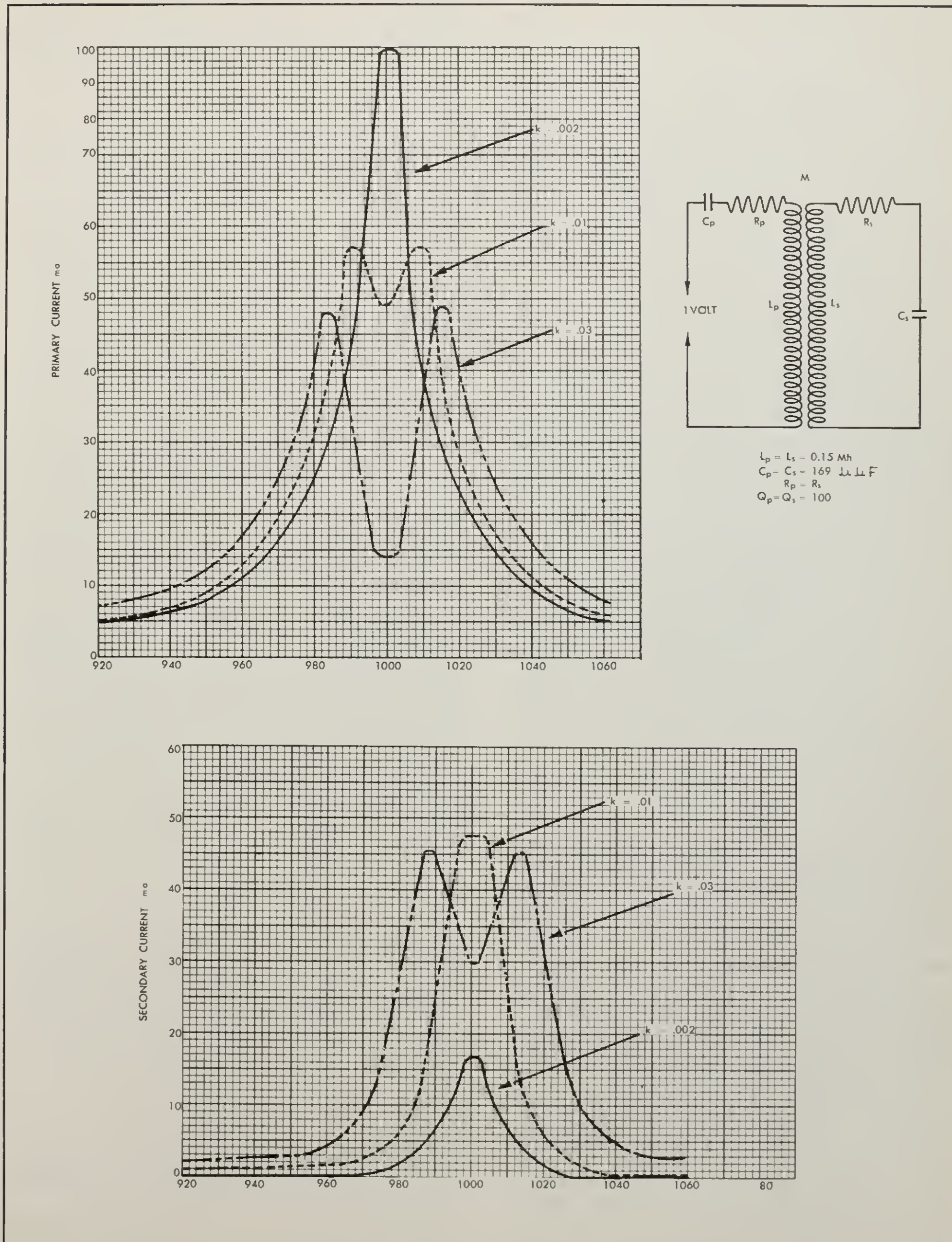
$$\text{coupled } Z = \frac{Z_m^2}{Z_2 + Z_s} = \frac{(\omega M)^2}{R_s + j\left(\omega Ls - \frac{1}{\omega Cs}\right)}$$

Within a limited frequency range, the numerator $(\omega M)^2$ in this equation does not vary appreciably. The denominator, $R_s + j\left(\omega Ls - \frac{1}{\omega Cs}\right)$ represents the series impedance of the secondary. This is the form of the impedance of a parallel resonant circuit $\left(Z_t = \frac{(\omega L)^2}{Z_s}\right)$. Thus, the coupled impedance of a tuned secondary circuit varies with frequency according to the same general law as does the parallel impedance of the secondary circuit but the magnitude of the curve depends upon the mutual inductance M.

When both the primary and the secondary of a transformer are tuned, the behavior of the transformer depends on the coefficient of coupling between the primary and the secondary. Notice on the next page, the illustration of a circuit with a transformer having a tuned primary and secondary circuit, and the curves showing the primary and secondary current for three values of coupling. When the coefficient of coupling is small ($k = 0.002$), the primary current curve is practically the same curve that would be obtained if the tuned circuit were considered alone, and the secondary current is small and more peaked than if the secondary were considered alone. As the coefficient of coupling is increased, the curve of primary current becomes broader. This is caused by the reduction in primary current at resonance and the increase in current at frequencies slightly off resonance. Meanwhile, the secondary current peak becomes larger and somewhat broader. The broadening continues until the resistance which is coupled from the secondary into the primary is equal to the primary resistance. This is the condition of critical coupling and causes the secondary current to reach its maximum value as shown by the curve labeled $k = 0.01$. The curve of secondary current is now somewhat broader than the resonance curve of the secondary alone. At the same time the primary current curve has two peaks, since it is a little larger just below and above resonance than at resonance. When the coupling becomes greater than critical, for example, when $k = 0.03$, the double peaks of primary current become more prominent and farther off resonance. Besides, the secondary current begins to show double peaks. Notice that the peaks of primary current are decreased, but that the peaks of secondary current are almost the same as for critical coupling.

The double humps of the current curves is explained by the nature of coupled impedance. In understanding the explanation, first consider the total primary impedance. It consists of the actual impedance of the primary and the impedance coupled from the secondary. In the case of a tuned secondary and untuned primary, coupled impedance was proved to be essentially that of a parallel resonance curve. Therefore at resonance it is maximum and pure resistance; below resonance it is inductive; and above resonance, capacitive. Now when the coupled impedance is added to the self-impedance of the primary, the effective primary resistance at resonance increases above the value it would have if there had been no secondary. This results in less primary current. Furthermore, the greater the degree of coupling, the greater the reflected and effective resistance, and the lower the current. At frequencies below resonance the self-impedance of the primary is capacitive and the reflected impedance is inductive. The effective series impedance, therefore, is decreased, and the current reaches a maximum value at a frequency below resonance where greatest cancellation of reactance occurs. Above resonance a similar situation exists, except that the self impedance is inductive and the coupled impedance is capacitive.

The amount of current in the secondary depends on the secondary impedance and the voltage induced in the secondary by the primary current. The induced *voltage* curve has approximately the same form as the primary current curve. This is because it is equal to ωMIp, and ωM does not vary appreciably over the small range of frequencies. Thus, the secondary current curve is almost exactly the same as the product of the shape of the primary current curve and the shape of the resonance curve (current) of the secondary when each is considered separately. Since the latter curve is sharply



Response Curve with Primary and Secondary Tuned peaked, the secondary current curve is more peaked than primary current curve.

The value of M in these circuits is usually equal to, or slightly greater than, the critical value. Besides having a flatter top, the sides of the curve fall off more rapidly than with a single tuned circuit. This makes the discriminator effect between the pass and other frequencies much better than with a single tuned circuit. Such a circuit therefore has good selectivity. Tuned transformers are used in IF amplifier stages in radio and radar receivers.

FACTORS AFFECTING BAND WIDTH OF TUNED TRANSFORMERS. The shape of a selectivity curve depends not only on the degree of coupling between primary and secondary, but also on the Q's and the resonant frequency of the primary and the secondary. It was pointed out before that for any given double-tuned circuit, optimum (or critical) coupling produces the greatest amount of current flow in the secondary and the greatest voltage output. The shape of this curve, as well as the magnitude of the current flow in the secondary, depends on the primary and secondary being tuned to the same resonant frequency and the Q of each circuit. The higher the Q, the greater the selectivity—that is, the steeper the slope of the sides of the curve, or in other words, the narrower the bandpass. The Q of the primary or the secondary depends, among other things, on the resistance of the wire with which it is wound and whether or not a resistor is connected in parallel with the tuning condenser. Circuits with resistors across transformer windings are said to be *loaded*. These resistors reflect a certain amount of resistance in series with the condenser and therefore tend to lower the effective Q of the circuit. In some applications, where a wide band of frequencies need to be passed, resistors are connected across the transformer windings.

It can be proved with critical coupling between the primary and secondary that $\omega M = \sqrt{R_1 R_2}$, where R_1 and R_2 are the resistances of the primary and secondary, respectively.

The expression for the coefficient at critical coupling is $k = \frac{1}{\sqrt{Q_1 Q_2}}$ where Q_1 and Q_2 represent the Q's of the primary and secondary.

When the primary and secondary resonate to the same frequency and are coupled loosely, the primary current vs frequency curve closely approximates the series-resonant curve for the primary when it is considered alone. As coupling is increased, the primary current curve broadens and peak current values decrease. With the passing of the point of critical coupling, double peaks and further broadening occur.

So far as the secondary is concerned, current flow is small when the coefficient of coupling is small and current variation with respect to frequency closely approximates the resonance curves for the primary and secondary circuits, when each is taken separately. As coupling is increased, the secondary current flow increases and the current curve broadens. As the coefficient of coupling exceeds the critical value of the current curve, the peaks become progressively more pronounced and farther apart as the coupling is increased.

When the Q's of the primary and the secondary are equal, the response curves are symmetrical with respect to a mean frequency. However, when the primary and secondary are tuned to frequencies which differ slightly, the effect is as if the coefficient of coupling were increased and the two circuits were tuned to the same frequency. In other words, detuning the circuits when the Q's are equal is equivalent, so far as shape is concerned, to increasing the effective coupling. The only difference is that detuning causes the absolute magnitude of the response curve to be less than when the same shape is obtained without detuning. When the Q's differ, the curves are not symmetrical regardless of whether the two circuits are tuned to the same frequency or not. The low frequency peaks will be less than the high-frequency peaks when the secondary is tuned to a higher frequency and its Q is *higher* than the primary, or when the secondary is tuned to a *lower* frequency and its Q is *lower* than the primary. Otherwise the high-frequency peak will be depressed. The position of the two resonant peaks for two slightly (less than critical) coupled-tuned circuits of relatively high Q and tuned to the same frequency can be found by the equation,

$$f = \frac{f_o}{\sqrt{1 \pm k}}$$

where f_o is the frequency to which the primary and secondary are tuned, and k is the coefficient of coupling.

The negative sign in the denominator is used for finding the position of the upper resonant peak; the positive sign for the lower.

CHAPTER 3

Measuring Instruments

The solution of many problems involving circuit quantities such as voltage, current, and resistance were taken up in the preceding chapter. However, there you did not learn how these quantities are measured, either for obtaining the data for problems or for checking the results of the computations. For the purpose of measuring electrical quantities, there are many useful instruments, among which the most common are the ammeter, the voltmeter, and the ohmmeter.

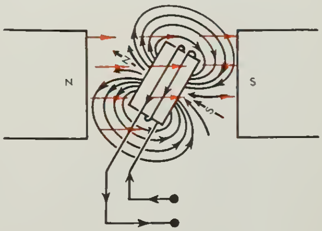
This chapter deals with the principles, construction, and application of instruments commonly used for making DC and AC measurements, instruments for measuring resistances and reactances, combination meters, the vacuum tube voltmeter, and the oscilloscope (an instrument for observing and studying alternating current wave forms). Not only will you learn how measuring instruments perform their function, but also what precautions you must observe in using them.

DC INSTRUMENTS

The three principal DC measuring instruments are the ammeter (for current), the voltmeter (for voltage), and the ohmmeter (for resistance). Basically, both the ammeter and the voltmeter are current measuring instruments, their principal difference being only in the method in which they are connected in a circuit. Ammeters are connected in series in the circuit in which current is being measured, and voltmeters in parallel with the circuit element across which the voltage is being measured. While an ohmmeter is basically also a current measuring instrument, it differs from an ammeter and a voltmeter in that it provides its own source of power and contains other auxiliary circuits.

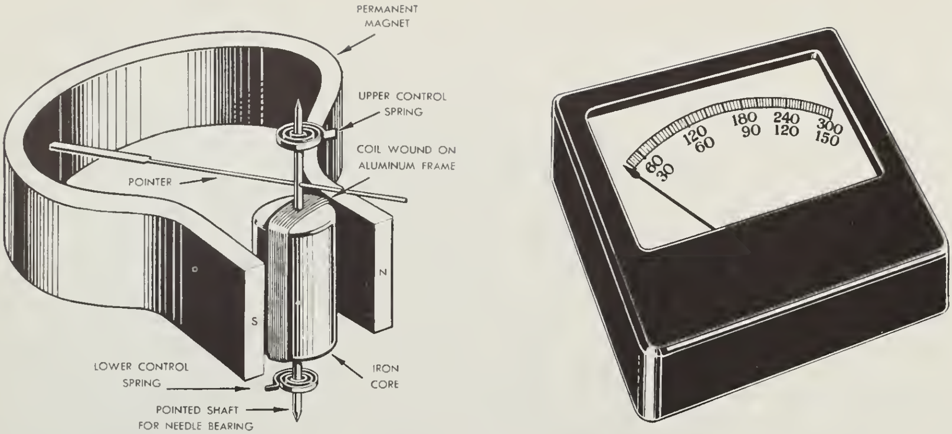
The Ammeter

A DC ammeter is an instrument designed for measuring the amount of direct current flowing in an electrical circuit. In the diagram of the simple ammeter, note that the main parts of the instrument are a coil and a permanent magnet.



Simple Ammeter

The coil is pivoted so that it is able to rotate back and forth within the magnetic field set up by the magnet. When the coil is connected in a circuit, current flows through the coil in the direction indicated by the arrows and sets up a magnetic field within the coil. This field has the same polarity as the adjacent poles of the magnet. Interaction of this field and the field produced by the magnet sets up forces which cause the coil to rotate to a position at which the two magnetic fields are aligned. This rotation is proportional to the interaction between the like poles of the coil and magnet. Accordingly, since the force or torque which causes the rotation is a function of the force of interaction, it is proportional to the amount of current flow in the coil. Thus a pointer moving across a graduated scale and connected to the coil will indicate the amount of current flowing in the circuit in which the ammeter is connected.



Components of Practical Ammeter

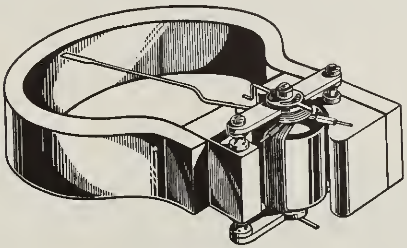
The pole pieces of the permanent magnet in a practical ammeter are curved in shape. Located within these pole pieces is a fixed cylindrical iron core. Although the illustration just above indicates wide spacing between the coil and magnet, in an actual meter there is only enough space for the coil to rotate within the gap of the permanent magnet without striking the pole pieces. The iron core insures that the magnetic field is uniform in strength and constant in direction. Any torque sufficient to overcome the inertia and friction of the moving parts causes rotation until the fields are aligned. However, such uncontrolled rotation would produce inaccurate current indications. Therefore, the mechanism includes a special spring one end of which is fixed, and the other end of which is attached to the movable coil. The spring exerts a tension proportional to the amount of rotation. By controlling the rotation of the coil, it insures accurate meter readings.

With the addition of an indicating pointer and a graduated scale, the ammeter just described becomes a complete current measuring instrument except for one thing—a damping device. Without a damping device, the pointer would swing back and forth several times at the indicated reading on the scale before coming to rest. Generally, ammeters are damped by winding the movable coil on a light-weight aluminum frame. Aluminum is used since its low weight keeps inertia at a minimum. The frame itself is a closed circuit. When the current being measured causes the coil to swing, it also swings the frame, causing it to cut the magnetic lines of force set up by the permanent magnet and inducing in it small magnetic fields which are opposite in direction to the field produced by the magnet itself. These fields oppose the action of the field of the magnet and when the torque of coil and the controlling force of the spring reach a position of balance, these opposing fields in the frame act as a brake to reduce the swing of the moving parts beyond this position. As a result, the needle comes to rest after only one or two oscillations.

To keep inertia low, light weight material is used in the moving parts of an ammeter. To keep weight down, and at the same time to insure that there is sufficient torque, the coil is wound of many turns of small wire. Usually, the pointer is made of very thin aluminum. However, any light weight material is satisfactory. To reduce friction the pivot on which the shaft carrying the coil rotates is made of a very hard metal, and is set in highly polished jewel bearings.

D'ARSONVAL MOVEMENT. The type of movement just described is the D'Arsonval type, a type movement which practically all DC meters employ. In the D'Arsonval-type meter, the moving element may be connected in such a way as to form either a voltmeter or an ammeter. For use as an ammeter, the moving coil usually has connected across it a shunt which carries most of the current flowing through the meter. For use as a voltmeter, the coil has a resistance connected in series with it, so that the meter itself will take very little current from the circuit being measured. The following are the essential parts of a D'Arsonval movement meter:

1. Stationary element (permanent magnet)
2. Moving element (rotating coil)
3. Controlling element (spiral springs)
4. Mountings and bearings
5. Damping device
6. Pointer and scale

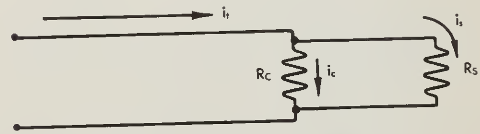


D'Arsonval Movement

Meter sensitivity. The amount of current which causes the indicator on an ammeter to deflect full scale is known as *meter sensitivity*. Thus meter sensitivity is expressed in terms of the current required for the full scale deflection. The smaller the current required, the more sensitive the meter. For measuring current in electronic equipment, ammeters with a sensitivity better than 0.1 amperes are commonly employed. Not too uncommon are meters with one milliampere and even 100 microampere sensitivity.

Ammeter shunts. To measure larger amounts of current than the coil itself can safely carry, a resistance is connected in parallel with the coil. This causes the current being measured to divide itself between the coil and the resistor, a small portion flowing through the coil, the remainder through the parallel resistor, called the *meter shunt*. The shunt may be built into the meter or it may be mounted externally. Meters designed to measure several ranges of current, employ a shunt for each range. The shunts are mounted on a common terminal board and are connected to a multi-pole switch. Setting the switch to the desired range connects the proper shunt resistor into the meter circuit. Shunt resistors usually contain only a fraction of an ohm resistance and consist of a few inches of a metal alloy having a low temperature coefficient. The alloy is drawn into a wire and is wound around a piece of mica or fiber and mounted on a terminal board.

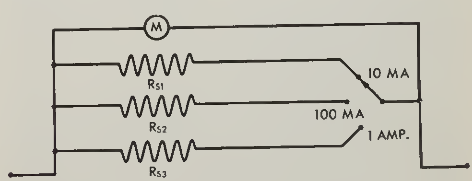
Since current through the two parallel branches divides in a ratio inversely proportional to the resistance of each branch, it is possible to calculate the current flowing through the coil as well as the total current flow in a circuit in which current is being measured. In the circuit just below, for example, if $R_s=5R_c$ and $I_c=0.5$ ma, find the current flowing in the shunt (I_s) and the total current flowing in the circuit. Since the



Shunt Resistor (R_S) Increases Current Range

resistance of the coil is five times that of the shunt, and since the current divides in inverse proportions to the resistance, five times as much current will flow through the shunt as through the coil. Therefore, the current through the shunt, $I_s=5(0.5$ ma$)$ or 2.5 ma. The total current in the circuit is equal to I_c+I_s. Thus, the total current, $I_t=0.5+2.5$ or 3 ma.

Calculations such as those involved in this example will enable you to determine the size of the shunt required to extend the range of an ammeter. Although in actual practice, it is seldom necessary to perform these calculations, it is well to know how meters having the same movement are constructed with different ranges, and how a range switch on the same meter changes the current range of the meter. The following example illustrates the method for computing values of shunts required in a multi-range meter having a single meter movement.



Computing Values of Shunts

Example

Problem

Assume that you desire to construct the ammeter shown with the ranges 0-10 ma, 0-0.1 amp, and 0-1.0 amp. If the meter sensitivity is 1 ma and the coil resistance is 75 ohms, find R_{s1}, R_{s2}, and R_{s3}.

Solution

1. Range 1-10 ma.

Since a full scale reading on the meter will indicate 10 ma, then 9 ma must flow through the shunt and 1 ma through the coil.

Therefore, substituting in the equation—

$$\frac{R_{s1}}{R_c}=\frac{I_c}{I_s}$$

$$\frac{R_{s1}}{75}=\frac{1}{9}$$

$$R_{s1}=8.33 \text{ ohms}$$

2. Range 0-0.1 amps.

R_{s2} must carry 99 ma (0.1 amp—1 ma)

Therefore substituting,

$$\frac{R_{s2}}{R_c}=\frac{I_c}{I_s}$$

$$\frac{R_{s2}}{75}=\frac{1}{99}$$

$$R_{s2}=\frac{75}{99}=.758 \text{ ohms}$$

3. Range 0-1.0 amps.

R_{s3} must carry 999 ma (1.0 amps—1 ma)

Therefore,

$$\frac{R_{s3}}{R_c}=\frac{I_c}{I_s}$$

$$\frac{R_{s3}}{75}=\frac{1}{999}$$

$$R_{s3}=\frac{75}{999}=.0751 \text{ ohms}$$

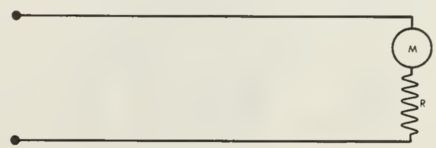
PRECAUTIONS IN USING AMMETERS. When using ammeters, observe the following precautions:

1. Always connect an ammeter in series with the element through which the current flow is to be measured.

2. Never connect an ammeter across a source of voltage, such as a battery or generator. Remember that the resistance of an ammeter, particularly on the higher ranges, is extremely low and that any voltage, even a volt or so, may cause very high current to flow through the meter causing damage to it.

3. Use a range large enough to keep the deflection less than full scale. Before measuring a current, form some idea of its magnitude. Then switch to a large enough scale or start with the highest range and work down until you reach the appropriate scale. Most accurate readings are obtained at approximately half-scale deflection. Many a milliammeter has been ruined by attempting to measure amperes. Therefore be sure to read the lettering either on the dial or on the switch positions, and choose the proper scale before connecting it in the circuit.

4. Observe proper polarity in connecting the meter in the circuit. Current must flow through the coil in a definite direction in order to move the indicator needle up-scale. Current reversal because of incorrect connection in the circuit results in a reversed meter deflection and frequently causes bending of the meter needle. You can avoid improper meter connections by observing the polarity markings on the meter and by remembering that the *black* meter leads are the *negative* leads and the *red* the *positive* leads.

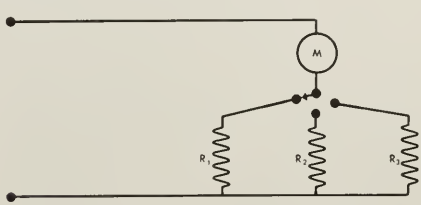


Voltmeter Circuit

The Voltmeter

Previously it was stated that the D'Arsonval meter movement can be used either as an ammeter or as a voltmeter. Thus, you can measure voltage with the ammeter just described by placing a resistance in series with the coil and measuring current flowing through it. In other words, a voltmeter is simply a current-measuring instrument, designed to indicate voltage through measurement of the current flow across a resistance of known value. Various voltage ranges may be obtained by adding various resistors called *multipliers* in series with the coil.

The following example illustrates how you can calculate the values of multipliers needed to extend the range of a voltmeter.



Computing Values of Multipliers

Example

Problem

What size multipliers does a voltmeter having a 0-1 ma meter movement with 75 ohms resistance require for ranges of 0-10 v, 0-100 v, and 0-500 v?

Solution

1. Multiplier R_1

Since 1 ma causes a full scale deflection, the resistance R_1+75 must be such that the voltage drop across it is 10 v when 1 ma flows.

Thus:

$$R_1+75=\frac{10}{.001}=10000 \text{ ohms}$$

$$R_1=9925 \text{ ohms}$$

2. Multiplier R_2

Similarly, $$R_2+75=\frac{100}{.001}=100{,}000 \text{ ohms}$$

$$R_2=99{,}925$$

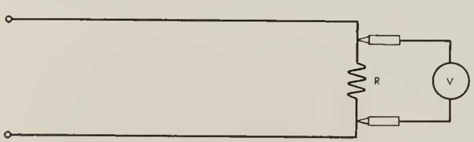
3. Multiplier R_3

Similarly, $$R_3+75=\frac{500}{.001}=500{,}000 \text{ ohms}$$

$$R_3=499{,}925 \text{ ohms.}$$

SENSITIVITY OF VOLTMETER. The sensitivity of a voltmeter is measured in *ohms-per-volt* and is determined by dividing the *resistance* of the meter and the multiplier by the full-scale reading in *volts*. It is just another way of stating what current will cause full-scale deflection. A voltmeter should have very high resistance so that it will draw very little current and affect the circuit as little as possible during voltage measurements. Sensitivity, therefore, is an indication of the measuring quality of a voltmeter. Generally a meter with a 1000 ohms per volt is satisfactory for most uses. For circuits with high resistances, however, a meter with greater sensitivity is required for accuracy.

PRECAUTIONS IN USING A VOLTMETER. When using a voltmeter, observe the following precautions:



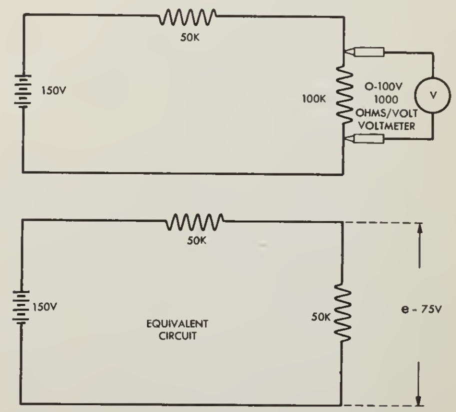
1. Always connect a voltmeter in parallel across the portion of the circuit in which voltage is being measured as shown above.

2. Use a range large enough to insure less than full-scale deflection.

3. Observe the proper polarity in connecting the voltmeter across the circuit.

SENSITIVITY AND LOADING. The accuracy of a reading indicated by a voltmeter depends upon the sensitivity of the meter. As previously mentioned, when a voltmeter having a low resistance, that is one with low sensitivity, is connected into a high resistance circuit, the voltage indicated by the meter is somewhat in error depending on the meter sensitivity. This error in reading is due to the *loading* effect which the meter has on the circuit in which voltage is being measured.

To understand just how a meter affects or loads a circuit, refer to the diagram of the series circuit below. According to the values indicated, the total current flow in the circuit is equal to $\frac{150}{100{,}000+50{,}000}$, or 0.001 amperes. The applied



Loading Circuit with Voltmeter

voltage divides across the two resistors—$0.001\times 100{,}000$, or 100 volts, across the 100,000-ohm resistor, and $0.001\times 50{,}000$, or 50 volts, across the 50,000-ohm resistor.

When you connect a voltmeter across the 100,000 ohm resistor, you naturally expect the meter to read 100 volts. However, depending upon the sensitivity of the meter, it may give readings other than the 100 volts you expect. For example, suppose you are using a voltmeter with a sensitivity rating of 1,000 ohms-per-volt on the 0-100 volt scale. The resistance of the meter on this scale equals 1000 ohms$\times$100, or 100,000 ohms. (To find the resistance of a voltmeter, multiply the sensitivity rating by the voltage.) When you connect this voltmeter across the 100,000 ohm resistor in the circuit, you put the meter in parallel with this resistor, and the circuit effectively becomes as shown in the equivalent circuit. As you can see the meter is in parallel with the 100,000 ohm resistor, and since the meter resistance is equal to the resistance of the circuit resistor, the effective resistance is 50,000 ohms. In other words the 100,000 ohm circuit resistor is effectively replaced in the circuit by an equivalent 50,000 ohm resistor.

The total circuit resistance is no longer 150,000 ohms, but 100,000 ohms, and the total current flow is no longer 0.001 amperes but $\frac{150}{100,000}$ or 0.0015 amperes. The voltage across the meter and the circuit resistor now equals the product of the total current and the 50,000 ohms (which represents the equivalent resistance of the meter and the 50,000 ohm resistor). Therefore, the meter reads only $.0015 \times 50,000$, or 75 volts. Thus, the loading effect of the meter on the circuit has reduced the voltage across the resistor from 100 volts to 75 volts.

Obviously, this error in the reading is due to the loading effect the meter has on the circuit. When the meter is removed, the voltage across the 100,000 ohm resistor again becomes 100 volts.

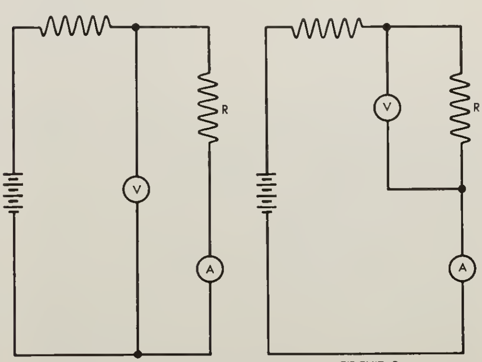
You can avoid loading a circuit by using a voltmeter in which the resistance is large compared to the circuit element across which voltage is being measured. If a voltmeter with such high sensitivity is not available, you can improve accuracy by using a higher-range scale on the voltmeter you have. Thus, for example, if you had used the 0-500 volt scale on the meter, the voltage across the resistor would have read approximately 91 volts.

On the 0-500 volt scale, the resistance of the meter equals $1,000 \times 500$ or 500,000 ohms. When the meter is put across the circuit, the parallel resistance of the circuit resistor and the meter equals $\frac{100,000 \times 500,000}{100,000 \times 500,000}$, or about 83,000 ohms. The total resistance in the circuit is then 83,000 plus 50,000, or 1,331,000 ohms. The total current flow is $\frac{150}{133,000}$ or approximately 0.0011 amperes. Thus, the voltage across the meter and the circuit resistor is equal to $.0011 \times 83,000$ or approximately 91 volts, which is much closer to accuracy than the reading obtained on the 0-100 volt scale.

Loading a circuit by a voltmeter explains why the same meter gives different reading on different scales when it is placed across the same circuit element. You saw this proved in the example where the 0-500 volt scale gave a much more accurate reading than the 0-100 volt scale on the same meter.

Although the highest range of a voltmeter gives readings more nearly equal to the actual voltage, it is not always desirable to use the highest range in all cases. You will find it much more difficult to read higher-range scales accurately than the lower-range scales where the numbers are less crowded.

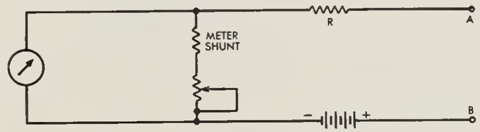
If you wish to use a voltmeter and an ammeter in the circuit at the same time, there are two possible ways to connect them, but each produces an error. In the first case, as you can see directly below, the voltmeter reads not only the voltage drop across the resistor R but also that across the ammeter. In the second case, the ammeter reads not only the current flowing through the resistor R, but that through the voltmeter as well. Both methods further emphasize the fact that the resistance of an ammeter must be very low to keep the voltage drop across it small and that the resistance of a voltmeter must be great to keep the current flowing through it small. The correct connection of the circuit is the one leading to the least error. It depends on the relative values of resistance. If the resistance of the circuit is small, approaching the resistance of the ammeter, use the second circuit as it leads to the least error. If the resistance of the circuit is large compared to the voltmeter resistance, use the first circuit. For intermediate values of circuit resistance, either circuit connection is satisfactory. Of course you may take the measurements by connecting one instrument at a time, but this practice does not necessarily increase the accuracy.



Measuring Current and Voltage Simultaneously

The Ohmmeter

Although you can determine the resistance of a circuit element by first measuring the current flowing through it with an ammeter, then the voltage drop across it with a voltmeter, and finally applying Ohm's law, it is much more practical to use an ohmmeter with which you can read resistance directly from scale.



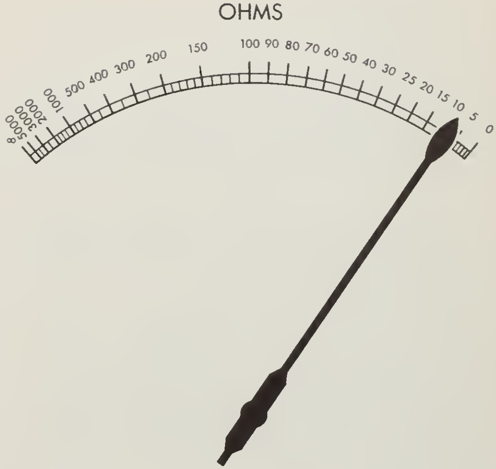
Ohmmeter Circuit

Ohmmeters possess a number of features not found in ammeters and voltmeters. An ohmmeter must supply its own power. Usually, this is supplied by a battery of known voltage. In an ohmmeter there are two resistors, one fixed, and one variable. The fixed resistor, marked R in the illustration, is of such a value that when the points A and B shown are shorted, the meter will read full scale. The variable resistor is connected in parallel with the meter and is called the *zero adjustment*. It serves to compensate for changes in the battery voltage. Usually, it is adjusted by a control knob on the meter panel.

If zero resistance is connected between points A and B in the ohmmeter, the meter pointer will deflect full scale. If a resistance equal to R is connected between these points, the deflection will be half scale, and if a resistance equal to 2R is connected, the deflection will be one-third scale. The difference in deflection indicates that the upper end of the scale reads low resistance.

The scales on ohmmeters are not linear but are similar to a scale of reciprocals. At the lower end (high resistance end), the calibrations are crowded, making accurate readings there difficult. For this reason, it is best to select a scale in which the indicator will fall in the upper portion.

A typical ohmmeter employs more than one scale. Additional scales are made possible by the use of various values of R and battery voltages. On some meters there is a special scale for reading very low resistances. This is made possible by connecting an unknown resistance *in parallel* with a known resistance. Readings of resistance in this case appear on a special scale called *low ohms*.



Typical Ohmmeter Scale

The ohmmeter is not as accurate a measuring device as the ammeter or the voltmeter because of the circuit associated with it. Therefore do not expect to be able to read resistance with greater than five to ten percent accuracy. While there are instruments which read the resistance of an element with very great accuracy, they usually involve some sort of bridge network and are much more involved in use. For the work you will do, the ohmmeter is accurate enough.

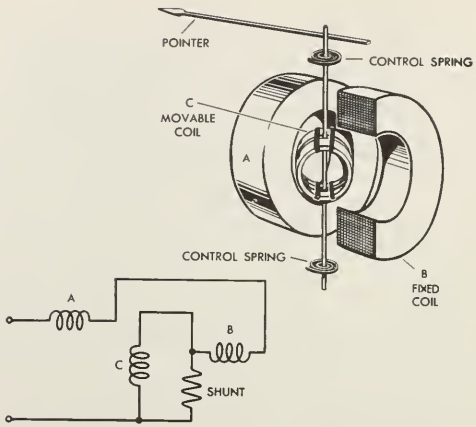
In addition to measuring the resistance, the ohmmeter is a very useful instrument for checking continuity in a circuit. Often, when trouble shooting electronic circuits, or wiring a circuit, you cannot readily make visual inspections of all parts of the current path. Therefore, it is not always apparent whether a circuit is complete or whether current might be flowing in the wrong part of the circuit because of contact with adjacent circuits. The best method of checking a circuit under these conditions is to send a current through it. If the conductor makes a complete circuit, current will flow through the circuit. The ohmmeter is the ideal instrument for checking circuits in this manner. It provides the power to send the current through and the meter to indicate whether the current is flowing. To make the check, first study the circuit diagram, and then check the corresponding parts of the circuit itself with the ohmmeter. It will indicate perfect conduction, partial conduction (resistance), or no conduction at all.

USING AN OHMMETER. When using an ohmmeter, proceed as follows:

1. Choose a scale which you think will contain the resistance of the element that you are measuring. In general, use a scale in which the reading will fall in the upper half of the scale (near full scale deflection). Reading the scale is easier there and gives greater accuracy.

2. Short the leads together and set the meter to read zero ohms by adjusting the zero adjustment. If you change scales at any time, remember to readjust to zero ohms.

3. Connect the unknown resistance between the test leads and read its resistance from the scale. Never attempt to measure resistance in a circuit while it is connected to a source of voltage. Disconnect at least one end of the element being measured to avoid reading the resistance of parallel paths.



Electrodynomometer Type Meter

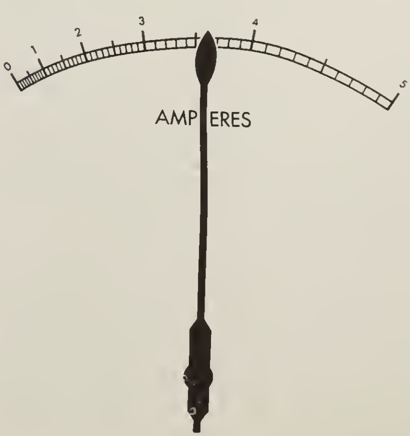
AC INSTRUMENTS

The D'Arsonval meter is not suitable for measuring alternating current unless it is modified to include a device (called a rectifier) for changing the periodically reversing AC into a unidirectional current. Recall that the current through the coil must flow in a definite direction. On page 3-10 is an explanation of the rectifier-type voltmeter. The same device could be used as an ammeter by use of a shunt instead of a multiplier. Other types of meters for the measurement of AC include the electrodynamometer, moving iron vane, and thermocouple.

Electrodynamometer Ammeter

For alternating current measurements, a very accurate and sensitive type of movement is the electrodynamometer. In this meter, the permanent magnet is replaced by an electromagnet. Instead of one coil, there are two coils, the moving coil (C) and the fixed coil (A and B). The fixed coil is split into two electrically continuous parts. The moving coil is located between the parts of the split coil. Any current which flows in the moving coil likewise flows in the fixed coil. Interaction between the fields about the moving coil and the fixed coil causes the moving coil to rotate. When the current being measured reverses direction, the fields about both the fixed coil and the moving coil reverse direction simultaneously, leaving the torque about the moving coil unchanged. A pointer connected to the coil and moving across a calibrated scale gives indications of the current flowing in the circuit being measured.

In a practical AC ammeter of the electrodynamometer type, the stationary coil is wound of heavy wire, enabling it to carry heavy current. The moving coil is small, since weight and inertia are important considerations in the design of moving parts. To permit it to handle the same current as the fixed coil, there is a shunt resistor connected across the moving coil.

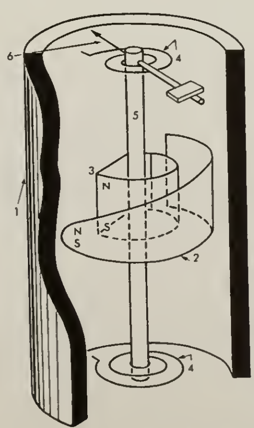


Ammeter Scale

In an electrodynamometer-type meter, the fields about the moving and fixed coils are proportional to the current flowing in them, and the force acting upon the moving coil is proportional to the product of the two fields. Deflection of the indicator, therefore, is proportional to the square of the current. For this reason, the scale of the meter is non-linear, but is crowded at low values and spread out at high values of each range. Most accurate results are obtained at near full scale deflection. Therefore, always choose a range which gives the greatest readable deflection.

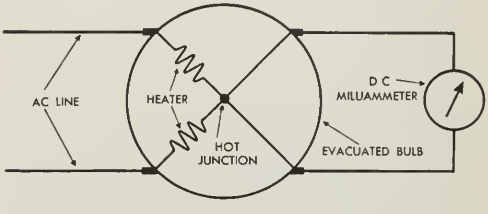
Moving Iron Vane Ammeter

Another meter type suitable for measuring AC current flow is the moving iron vane ammeter. In the moving iron vane meter illustrated, repulsion occurs between the two strips of soft iron, ② and ③. One strip is fixed, and one is movable. Both are cylindrical in shape and are oriented within the cylindrical coil ①. Through this coil flows the current which is being measured. This current magnetizes both strips, setting up like magnetic poles in the upper edges as well as in the lower edges. Since like poles are opposite, the two strips repel each other. Because of the shape of the fixed strip, the resulting repulsion causes the movable vane to rotate about its axis ⑤. The spiral springs ④ resist this rotation and cause it to be proportional to the torque. The pointer ⑥ connected to the movable vane moves across a graduated scale indicating the amount of current flow in the coil.



Moving Iron Vane Type Meter

The coil in the moving vane meter is wound of heavy wire and is capable of carrying more current than the electrodynamometer-type meter. Additional ranges are made possible by the use of shunts. Usually, the scale in this type of meter is like that in the electrodynamometer; that is, its deflection is not linear but is proportional to the square of the current.



Thermocouple Ammeter

Thermocouple Ammeter

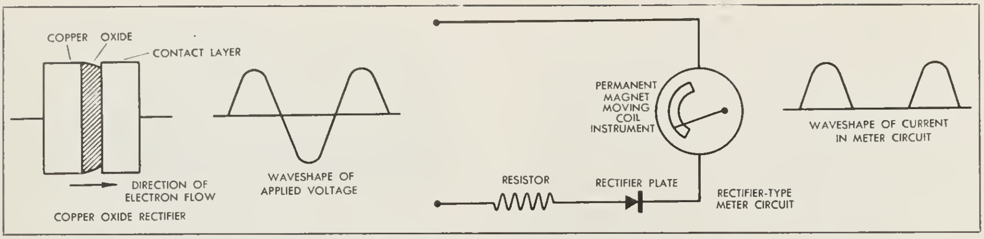
A type of AC ammeter that is especially useful for measuring high frequencies is the thermocouple meter. This type meter depends on the heating effect rather than the magnetic effect of current flow. It is based on the fact that whenever the junction of two wires of unlike materials such as iron and copper-nickel alloy is heated, a voltage is produced. Such a combination of different metals is called a thermocouple. The heat which produces the voltage is the result of the current flow through the thermocouple.

In a thermocouple meter, a sensitive DC milliammeter connected to the unheated (cold) ends of a thermocouple will deflect in an amount proportional to the difference in temperature between the unheated ends and the junction (hot end). Usually, the thermocouple is heated by AC passing through a high resistance wire, called the heater element. Since the temperature of the heater is proportional to the square of the current flow, the meter scale usually is like that used in electrodynamometer meters.

The thermocouple ammeter reads either direct current or alternating current of any frequency. Its chief use is for making measurements in radio frequency circuits.

AC Voltmeter

Each of the meters just described may be used as a voltmeter when suitable multipliers are added in series with them. The size of the multiplier is determined in the same way as with DC voltmeters.

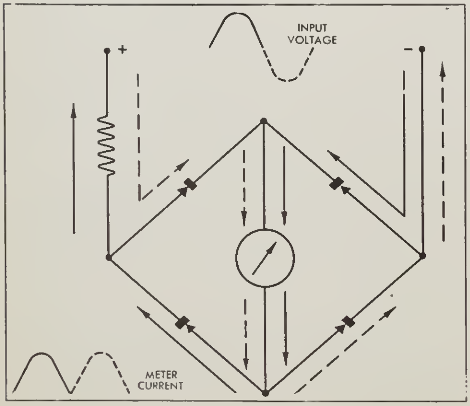


Half-Wave Copper-oxide Rectifier Meter Circuit

Rectifier-Type Meter

Another type of AC voltmeter widely used for measuring AC voltages is the rectifier-type meter. This meter consists of a DC meter with a D'Arsonval movement connected in series with a circuit element called a *rectifier*, which is designed to pass current in only one direction. Usually, the rectifier is of the copper-oxide type, which consists of one or more copper-oxide discs alternately separated by a copper disc and fastened together in a single unit. Since current flows more readily from copper to copper oxide, than from copper oxide to copper, the rectifier will permit current to flow in only one direction. In some rectifier meters a selenium rectifier is used instead of a copper-oxide rectifier. However, the principle of operation is the same.

A copper-oxide rectifier is incapable of withstanding high voltages. Whenever too high a voltage is applied across it, arcing takes place between the oxide and copper plates of the rectifier. Because of this inability of a single rectifier to withstand a high voltage, several rectifiers are connected into a circuit arrangement called a *bridge circuit*. In the bridge circuit shown below, four rectifiers are connected together. This arrangement causes voltage across each rectifier to be only half as great as it would be if only a single rectifier were used and also makes possible full wave rectification.



Bridge Rectifier Meter Circuit

Current flow through a rectifier is not steady but is characterized by many alternations per second. Although the indicator in the rectifier meter cannot follow these rapid changes in current flow, it will indicate some value which has a definite ratio to the applied voltage. For this reason it is possible to calibrate the scale on the rectifier-type meter directly in voltage. Actually the deflection of pointer is proportional to the average value of the AC voltage. But, since there is a definite ratio between average and RMS valued of a sine wave, the scale may be calibrated in RMS values as is the case in other types of AC meters. One point to remember, though, is that the readings are accurate only when the voltage being measured is a sine wave of voltage.

Observe the following precautions in the care and use of the rectified meter:

1. Do not subject the rectifier to high temperatures or chemical vapors. Certain chemicals such as sulphuric-acid vapors condense on the rectifier, forming a path for current flow in both directions.

2. Keep in mind that the accuracy is usually not better than 5% of the full scale reading.

3. Never remove the load from a bridge rectifier without first removing the applied voltage. Removing the load subjects the rectifier to the full applied voltage, and since rectifiers are not constructed to withstand more than a few volts, arcing will occur and puncture the oxide.

MEASUREMENT OF INDUCTANCE AND CAPACITANCE

A circuit similar to the ohmmeter circuit previously described can be arranged for measuring reactances if you replace the battery by a known source of AC voltage, and the resistor by a reactive circuit element such as a condenser or an inductor. Current flow through the known and unknown reactances will give an indication of the reactance in the circuit being measured. If the frequency of the AC voltage source is constant, such as 60 cps, the scale can be graduated to read capacitance directly. An inductance is measured by using the meter in the same way as for measuring capacitance, except that a conversion chart or graph is used in conjunction with the meter. Inductance measurements are quite inaccurate in the case of inductors with large resistance.

An instrument of this type does not possess a high degree of accuracy as a measuring device. However, it is quite useful, providing you exercise some common sense in interpreting results. For making precise measurements of capacitance or inductance, a somewhat more elaborate circuit is required. Most precision-reactance measuring instruments employ some type of bridge arrangement, but these instruments are primarily for laboratory use and are not very practical for use as field-test equipment.

In measuring capacitance or inductance, first short the leads together and *zero* the meter as you do in measuring resistance. Then connect the capacitance or inductance element being measured between the leads and read the reactance from the calibrated dial. Make no measurements of circuit elements while voltage is applied. To obtain correct indications, be sure to disconnect one terminal from the circuit to avoid reading the reactance of parallel paths.

THE VACUUM TUBE VOLTMETER

It is desirable that voltmeters for use in servicing many electronic circuits consume practically no power, for the loss of even a small amount of power seriously disturbs many circuits being tested. The vacuum tube voltmeter meets this requirement and has a widespread use in the servicing of electronic equipment.

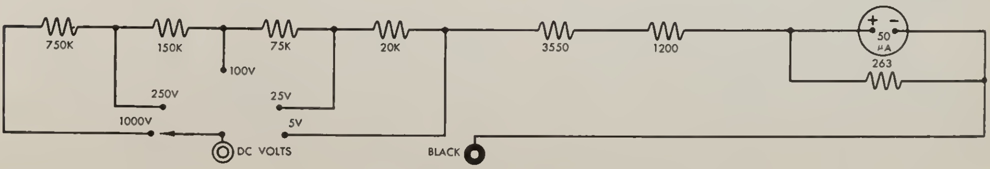
Essentially, this instrument consists of one or more vacuum tubes, a DC milliammeter, a voltage supply, and various circuit components such as resistors, capacitors, and switches. The chief advantages of the vacuum tube voltmeter are little or no power taken from the circuits being tested, large deflections for small input voltage variations, and high degree of accuracy in measuring circuits with high impedance and low power.

Either DC or AC voltages can be measured with the vacuum tube voltmeter. It operates by virtue of the fact that a change in grid voltage causes a change in plate current. When the meter is used as a DC measuring instrument, the vacuum tube in the meter acts as a DC amplifier; when it is used as an AC measuring instrument, the vacuum tube acts as a rectifier. In either case, the current flowing in the plate circuit is proportional to the voltage applied to the grid of the tube, and as indicated on the milliammeter, calibrated to read voltage.

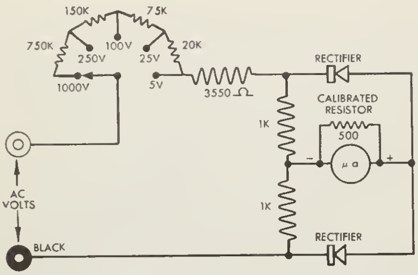
TEST UNIT 1-176

An instrument in wide use for making various electrical measurements is the Test Unit I-176. This unit is a combination meter comprising a single meter movement and a number of associated circuits. Switches on the meter connect various resistors, shunts, and other circuit components in such an arrangement that, depending upon the position of the switches, the I-176 can be used to measure AC voltage and current, DC voltage and current, and resistance. The following are types of measurements and ranges available in Test Unit I-176:

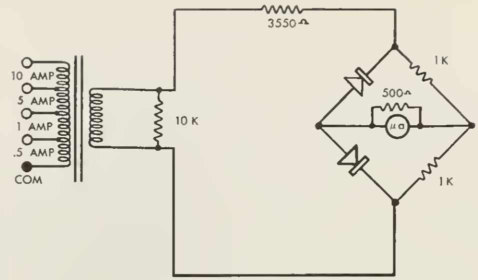
1. DC voltages 0-5 v, 0-25, 0-100 v, and 0-250 v. These ranges are provided by means of the multipliers shown in the circuit below



Simplified 1000 Ω/V DC Voltmeter Circuit



Simplified AC Voltmeter Circuit



Simplified AC Ammeter Circuit

The range selector switch connects the correct multiplier into the meter circuit. There is a selection between two sensitivity ratings—1000 ohms per volt and 20,000 ohms per volt. The selection is made by the function switch. On the 20,000 ohm per volt position the 263-ohmmeter shunt is out of the circuit.

An additional DC voltage range is available. This range is 0-5000 v. It has a sensitivity rating of 20,000 ohms per volt and requires special test leads which have a resistance of 50 megohms.

2. AC voltages 0-5 v, 0-25 v, 0-100 v, 0-250 v, and 0-1000 v. All these ranges have a sensitivity rating of 1000 ohms per volt. Except for the 1200-ohm resistor, this position of the switch connects the same multipliers as the DC voltage position.

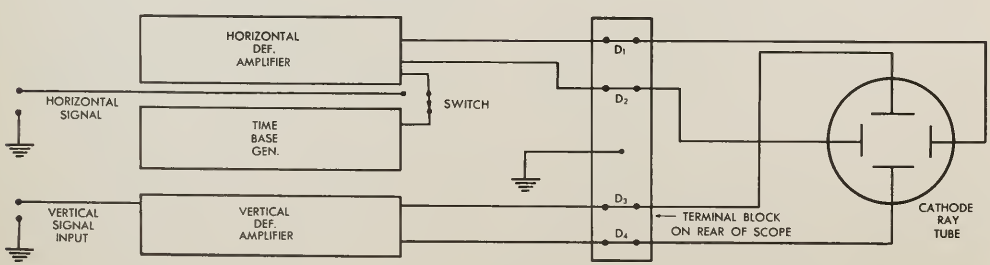
3. Direct current 0-50 ma, 0-1 ma, 0-10 ma, 0-100 ma, 0-500 ma, 0-1 a, and 0-5 a. Various shunts are placed into the meter circuit by the function selector switch, the range selector switch, and by plugging the test leads into the separate test jacks.

4. Alternating current 0-500 ma, 0-1 a, 0-5 a, 0-10 a. Plugging the test leads into the jack corresponding to the desired range completes that range circuit. Each jack connects electrically to taps on the transformer primary. The secondary circuit of this transformer uses the same rectifier and meter used on the AC voltage position.

5. Resistance 0-1000 ohms, 0-100,000 ohms and 0-10 megohms. The desired range is selected by the range selector switch and by plugging the test leads into the jack corresponding to that range. Power is supplied by either a 1.5 v or a 22.5 v battery depending on the equipment.

THE OSCILLOSCOPE

The oscilloscope is an instrument consisting of a cathode-ray tube and associated circuits for use in viewing wave shapes of voltages or currents. The cathode-ray tube, which is discussed in detail later, consists of three parts—an electron gun for supplying a stream of electrons in the form of a beam, deflection plates (coils) for changing the direction of the electron beam a small amount, and a screen covered with a material which gives off light when struck by the stream of electrons directed at it by the gun.



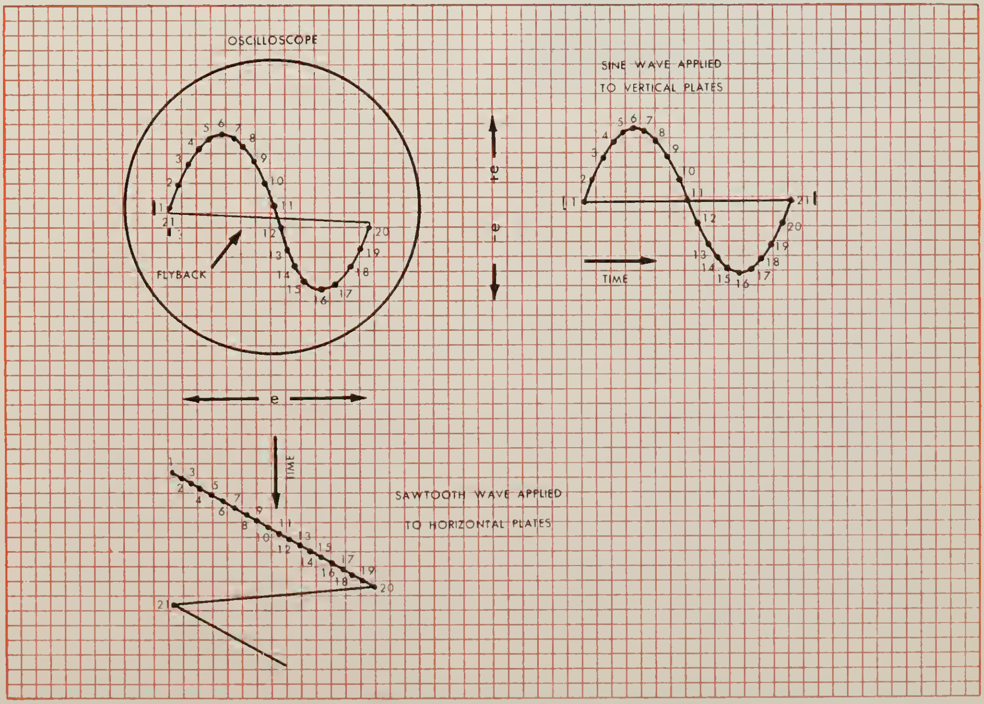
Block Diagram of Oscilloscope

The associated circuits in the oscilloscope include a time-base generator which causes the electron beam to move from the left to right side of the screen at a uniform rate and then return hurriedly to the left side where it begins another sweep across the screen. This action is accomplished by generating a voltage that rises at a uniform rate to a certain value and then quickly drops back to its starting value. A wave shape such as this is called a *saw-tooth wave.*

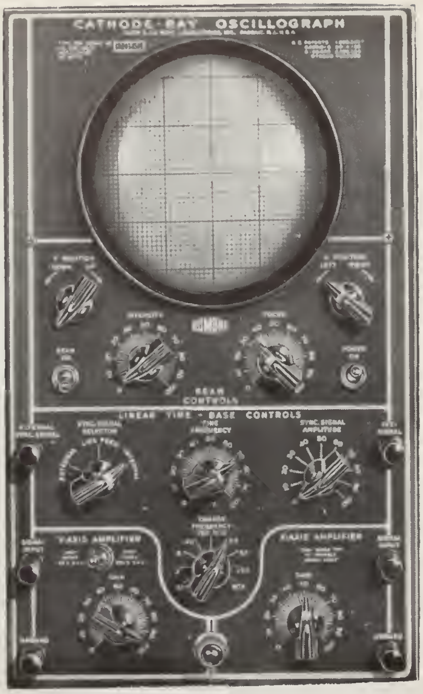
The saw-tooth voltage wave is applied to the horizontal deflection plates, where it causes the electron stream to change direction. Since negative voltages repel and positive voltages attract electrons, the gradual rise in voltage causes the left plate to become increasingly negative and the right plate increasingly positive and thereby causes the spot to move across the screen. The quick drop of the voltage back to its starting value returns the spot from right to left in a very short time, called the *flyback time.*

The saw-tooth voltage is generated by the time base generator, and this voltage or some other desired external voltage for horizontal deflection applied at the terminals marked *horizontal signal input,* is fed into the horizontal deflection amplifier which increases the amplitude to that needed for a trace of the desired length. The voltage to be observed is applied to the *vertical signal* terminals and is amplified to the desired value before being applied to the vertical deflection plates. With voltages thus applied simultaneously to the vertical plates and to the horizontal plates, the deflection of the beam is the resultant of the two forces.

You can see how this works out by studying the illustration below which shows how a sine wave is reproduced on the screen when a saw-tooth voltage is applied to the horizontal plates and a sine wave voltage to the vertical plates. The resultant is a single cycle of sine wave on the screen provided that the duration of one cycle of the saw-tooth voltage is the same as one cycle of the sine wave voltage.



Development of Sine Wave on Face of Oscilloscope



DuMont 208 Oscillograph Controls

Getting two cycles of the sine wave on the screen requires that the period of the saw-tooth voltage be twice that of the sine wave; that is, its frequency must be one-half of the sine wave frequency. Thus, you can see that the sweep generator must be variable in frequency to permit observation of other than one cycle or, for that matter, waves of any frequency other than that of the saw-tooth voltage.

DuMont Type 208 Oscillograph

The DuMont 208 Oscillograph is typical of the oscilloscopes employed in radar. There are others in common use, but this one serves to illustrate the kind and purposes of the controls. In this instrument the *horizontal* axis is called the *X* axis, and the *vertical* axis the *Y* axis.

POWER ON SWITCH. The POWER ON switch applies 115 volts at 60 cycles per second to the oscillograph. The pilot light goes on when power is applied.

BEAM ON SWITCH. The BEAM ON switch turns the beam off and on. The Off position provides for stand-by operation, during which the set operates in all respects except that there is no spot on the screen. In this condition, the set is ready for operation without delay for a usual warm-up period, but the life of the screen is not being needlessly shortened when the oscilloscope is not in actual use.

INTENSITY CONTROL. The INTENSITY control determines the amount of beam current. With this control, keep the intensity setting no higher than necessary for convenience of use, in order to conserve tube life. Do not permit a sharply focused spot or line to remain stationary on the screen at high intensity.

FOCUS CONTROL. The FOCUS control sets the potential of the focusing electrode of the cathode ray tube gun. Generally, there is a setting for optimum focus at each intensity level. This makes the picture on the screen clear cut.

POSITION CONTROL. The X POSITION and the Y POSITION controls permit adjustment of the position of the trace along the X axis and Y axis respectively. They serve to *center* the picture.

LINEAR TIME-BASE CONTROLS. The group of controls marked LINEAR TIME-BASE CONTROLS governs the operation of the gas discharge tube which serves as the oscillator in the linear time-base circuit.

The COARSE FREQUENCY control determines the range of the sweep frequencies which can be selected with the FINE FREQUENCY control. This latter control also stabilizes the pattern on the screen after the proper range of sweep frequency has been selected with the COARSE FREQUENCY control. In this equipment, the repetition frequency rate of the linear time base is continuously variable from 2 to more than 50,000 cycles per second. The ranges which are marked on the front panel, however, serve as a guide only. They are not exact frequency calibrations.

The OFF position of the COARSE FREQUENCY control turns off the sweep circuit oscillator and connects the input circuit of the X-axis amplifier to the X-axis signal input terminal post.

SYNCHRONIZING SIGNAL SELECTOR CONTROL. The SYNC SIGNAL SELECTOR control determines the source of the signal with which the linear time-base repetition frequency is synchronized. The signal sources for synchronizing the linear time-base repetition frequency may be External, Line Frequency, or Internal.

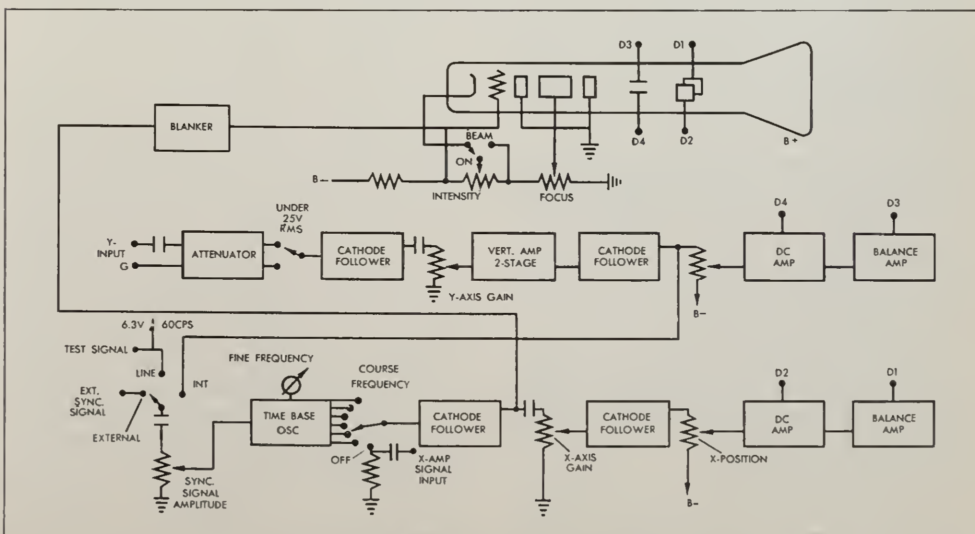
In the EXT position, the switch synchronizes the discharge tube with a signal connected between ground and the EXTERNAL SYNC SIGNAL terminal post. In general, this signal is different than that amplified in the Y-axis amplifier.

When you throw the switch to the LINE FREQ position, you can synchronize the sweep oscillator to the frequency of the power line supplying power to the instrument. This position is often useful when employing the line frequency as a standard frequency source.

When you throw the switch to the INT position, a signal having the same phase and wave form as that applied to the Y-axis amplifier signal input terminal post can be used to synchronize the linear time base. Since this source is available only when it is possible to amplify the signals, connect the synchronizing signal to the external synchronizing signal terminal post when direct deflection plate connections are used. Under such conditions, you may have to employ an auxiliary impedance-matching vacuum tube circuit to prevent excessive loading of high-impedance signal sources by the synchronizing circuit.

SYNCHRONIZING SIGNAL AMPLITUDE CONTROL. The SYNC SIGNAL AMPLITUDE control determines the size of the synchronizing signal applied to the grid of the relaxation oscillator which produces the sweep oscillation frequency. Use the minimum setting of this control at all times, since too large a synchronizing signal will distort the output wave form of the sweep oscillator and introduce nonlinearity. After getting the pattern nearly at stability by means of the FINE FREQUENCY control alone, advance the SYNC SIGNAL AMPLITUDE control just far enough from zero to prevent drifting of the pattern. When a signal other than that from the power line or from the signal being amplified in the Y-axis amplifier is to be used for sweep circuit synchronization, connect it to the external synchronizing signal terminal post. Under such conditions, throw the SYNC SIGNAL SELECTOR switch to EXT.

Y-AXIS AMPLIFIER CONTROL. The Y-AXIS AMPLIFIER controls consist of the SIGNAL INPUT terminal post, the Input Amplitude Selector switch, and the amplifier GAIN control. The input circuit to the amplifier presents a constant resistance of 2 megohms and a shunt capacitance of approximately 30 micro-microfarads in either position of the Amplitude Selector switch to which it is capacitively connected. This control is followed by a two-stage resistance-capacitance-coupled amplifier which connects to the positioning circuit and to the final direct-coupled balanced deflection amplifier stage. Connect the signal (generally the unknown signal) used to provide deflection along the Y, or vertical axis between ground and the signal input terminal



Functional Block Diagram of DuMont Oscillograph post. Signals up to 250 volts RMS amplitude may be connected to these terminals directly; signals of greater amplitude should be applied through an external attenuator. Since the signal is applied through a vacuum-tube amplifier, the DC potential of the input signal is removed. Consequently, when you want to study direct current signals, combinations of direct and alternating current signals in which you desire to preserve DC relationship, or high frequencies above one megacycle, connect them directly to the deflection plate terminal posts at the rear of the oscilloscope.

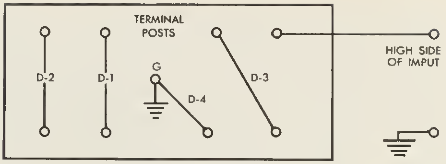
Y-AXIS AMPLIFIER INPUT ATTENUATOR CONTROL. The Y-axis Amplifier Input Attenuator is a two-position switch marked with the maximum signal voltage permissible for application directly to the signal input terminal. When no component of the signal has an amplitude greater than the peak amplitude corresponding to 25 volts RMS, throw the Attenuator switch to the position marked INPUT UNDER 25 VOLTS RMS. For inputs corresponding to maximum of 250 volts RMS, or for signals of unknown amplitude, throw the switch to the position marked INPUT UNDER 250 VOLTS RMS. Greater signal amplitudes require external attenuation.

Y-AXIS AMPLIFIER GAIN CONTROL. The Y-axis Amplifier GAIN control is a continuously variable low-impedance attenuator following the input coupling stage. Keep the setting low enough to prevent overload of the input grid circuit. Overload is indicated by a change of wave shape at high gain settings. The setting of the Y-axis Amplifier GAIN control determines the amplitude of deflection along the Y axis.

X-AXIS AMPLIFIER SIGNAL INPUT TERMINAL POST. The X-axis Amplifier SIGNAL INPUT terminal post is used when an external signal is to be amplified for deflection along the X, or horizontal axis. Do not permit the maximum signal amplitude to exceed 35 volts. For greater signal amplitudes use external attenuation.

X-AXIS AMPLIFIER GAIN CONTROL. The X-axis Amplifier GAIN control is a continuously variable low-impedance attenuator located immediately after the impedance-transforming input stage. It determines the amplitude of deflection along the X, or horizontal, axis.

TEST SIGNAL TERMINAL POST. The TEST SIGNAL terminal post provides a convenient source of signal for test purposes at the same frequency as the power line.



Direct Connections to Plates

REAR TERMINAL POSTS. The terminal posts on the rear of the oscilloscope may be used to apply signals directly to the deflection plates. Altogether there are nine terminals. Note how they are arranged in the illustration just above. Terminal posts D-3 and terminal posts D-4 connect to the vertical deflection plates. The lower connection of D-3 goes to the amplifiers and the input terminal on the front of the oscilloscope. The upper post connects to the deflection plates. The arrangement is the same for the other pair of posts. A direct connection between the upper and lower terminals is made when the signal is applied to the plates directly. This allows you to retain control of the vertical centering at the front panel. Terminal posts D-1 and D-2 connect to horizontal deflection plates. Use them the same way as D-3 and D-4.

Other Uses

In addition to its use in observing wave shapes, the oscillograph is used for measuring voltages, currents, frequency, and phase relations.

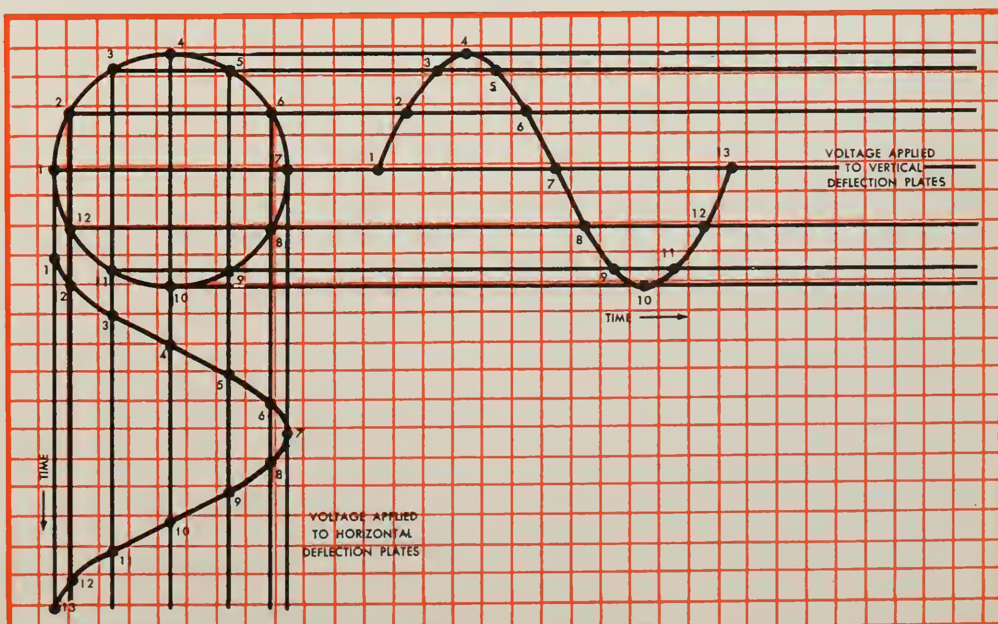
MEASURING VOLTAGES AND CURRENTS. A voltage is measured by comparing the deflection it produces with the deflection produced by a known voltage. Do not change the vertical gain control while comparing the deflections. Since deflection is proportional to voltage, this is a very good method for measuring peak voltage of odd wave shapes (other than sine waves). First, apply the known voltage and then adjust the vertical gain control to give as much amplitude as possible without letting the image extend beyond the flat surface of the screen. Measure the peak amplitude accurately. Do this with a ruler and dividers or by counting squares on the shield in front of the screen. Apply the unknown voltages and take a similar measurement. By comparing deflections, you get the ratio of the peak voltages. If both are sine wave voltages, the RMS values have the same ratio. If either is nonsinusoidal, the only comparison possible is the peak values. The meters you have encountered thus far do not read peak values, but you know that the peak value of a sine wave is 1.414 times the RMS value. However that is not true for other wave shapes.

Measure currents by passing the current through a known resistance, and measuring the resulting voltage drop in the manner described for measuring voltages.

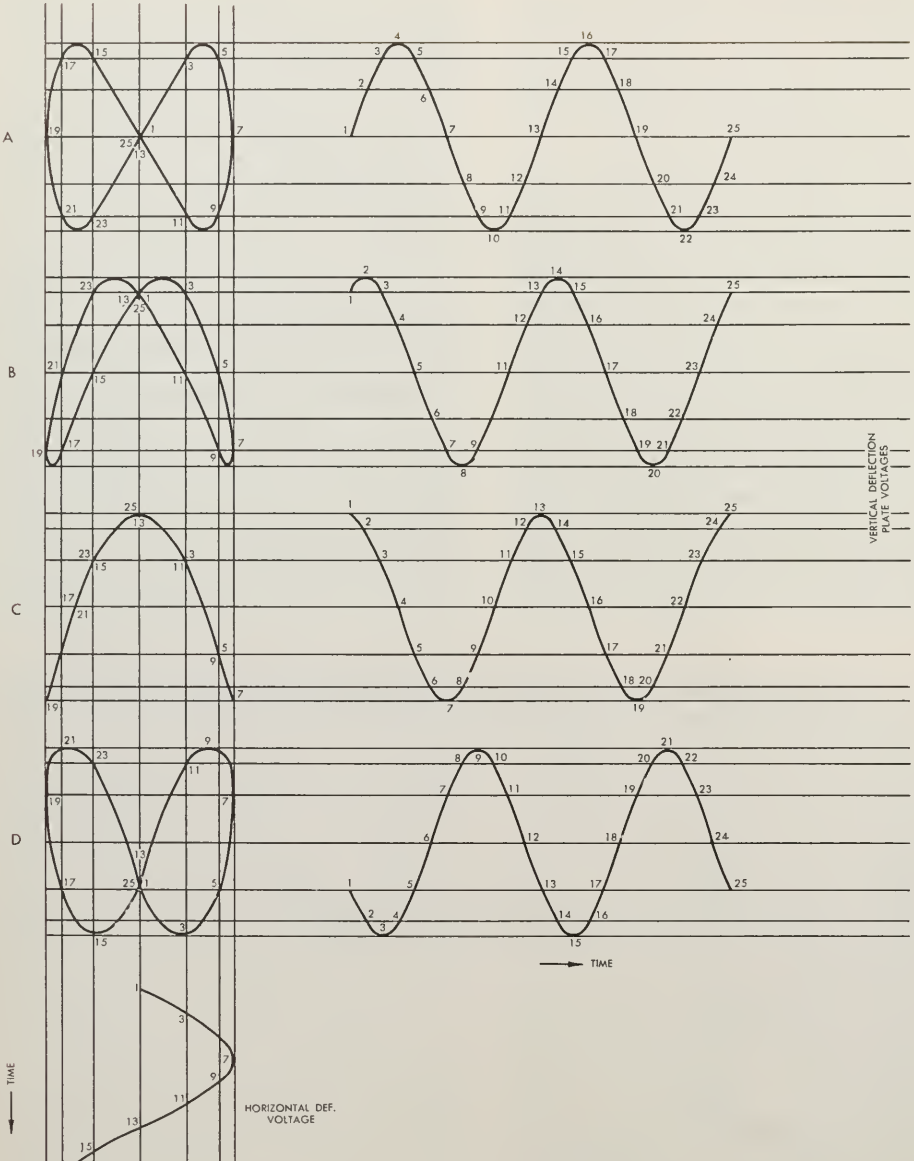
MEASURING FREQUENCY. Frequency measurement may be made by either of two methods. The first method is to compare the unknown frequency with a known frequency by applying them one at a time to the vertical deflection plates and then comparing the number of cycles which appear on the screen. Do not change the sweep frequency controls during the comparison. Unless the two frequencies have an integral ratio, this method is not entirely successful. For example, suppose that when a known frequency of 1000 cps is applied to the oscillograph four cycles appear on the screen, and when an unknown frequency is applied, five cycles appear. If you calculate the unknown frequency by the relation $\frac{f}{1000}=\frac{4}{5}$, you get f=1250 cycles per second. Suppose the unknown frequency were some figure such as 1223. In this case you can only approximate its frequency since changing the sweep frequency slightly will only tell you if the unknown frequency is above or below 1250. If you have to increase the sweep frequency to get five cycles, then the unknown frequency is above 1250 cps. In using this method of approximating frequency, place the SYNC SIGNAL AMPLITUDE control in its minimum (CCW) position for best results.

A second method for determining the frequency of a sine wave is to apply the unknown frequency to the vertical deflection plates and a known frequency sine wave to the horizontal deflection plates. The resulting figure on the screen is called a *Lissajous figure*, and from it you can determine the unknown frequency. The simplest type of Lissajous Figure is a circle. A circle can be produced by two voltages differing 90° in phase having the same frequency and amplitude. The illustration below shows the generation of a circular pattern by two sine waves 90° out of phase.

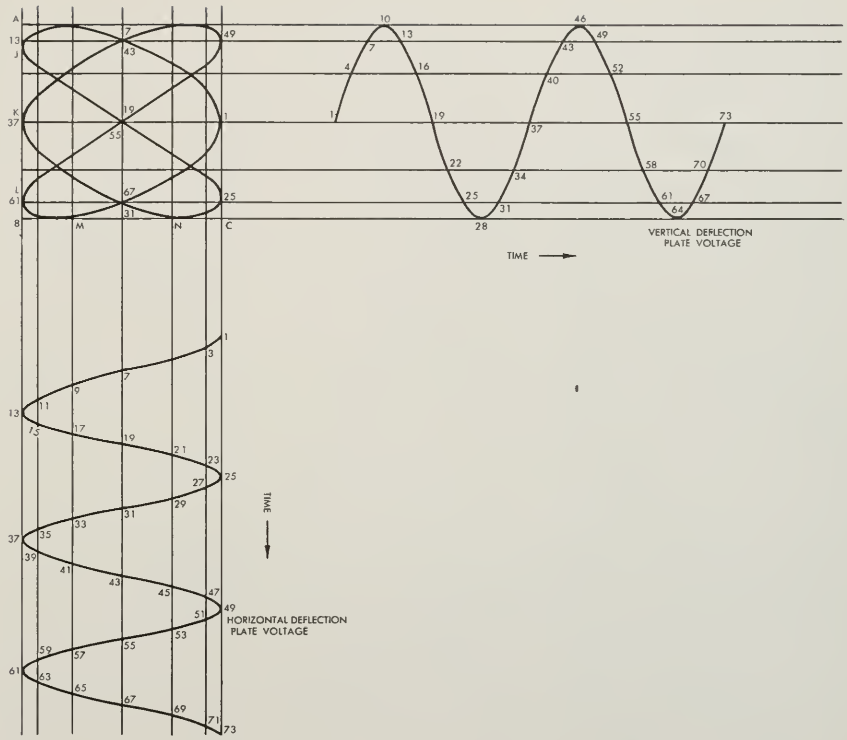
Figures resulting from the application of one frequency to the vertical deflection plates and



Generation of Circular Pattern by two Sine Waves 90° Out of Phase



Lissajous Figures for 1:2 Frequency Ratio



Lissajous Figures for 2:3 Frequency Ratio

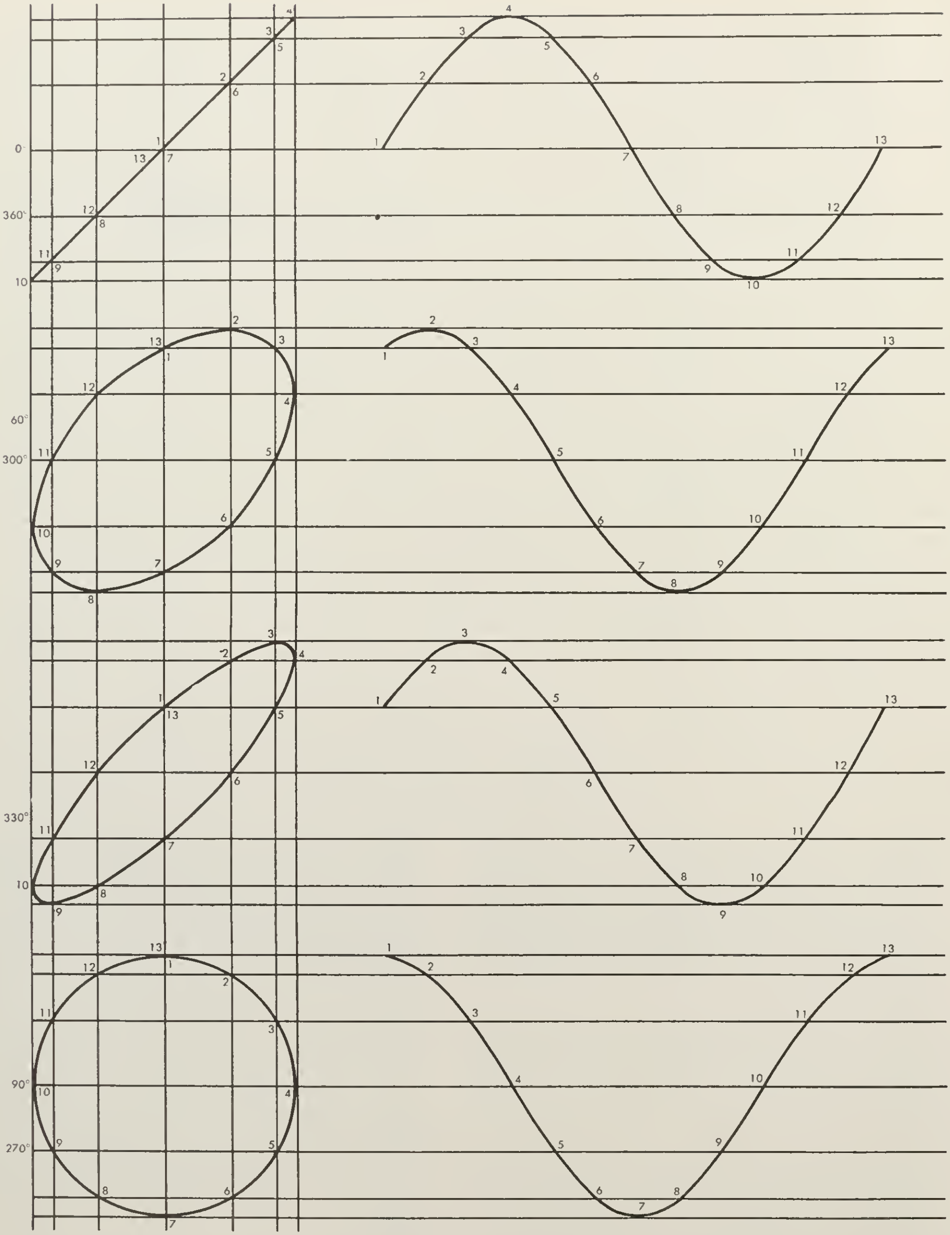
another frequency half the value of the first to the horizontal plates are shown in the illustration labeled Lissajous Figures for 1:2 Frequency Ratio. It makes no difference what the actual frequencies are as long as they have a ratio of 1:2. Note that the phase angle difference alters the pictures of the wave shapes.

Two voltages in phase appear on the screen as shown at A. If the phase angle is 90°, the loops are closed as shown at C. Phase angles greater than 180° appear inverted as shown at C.

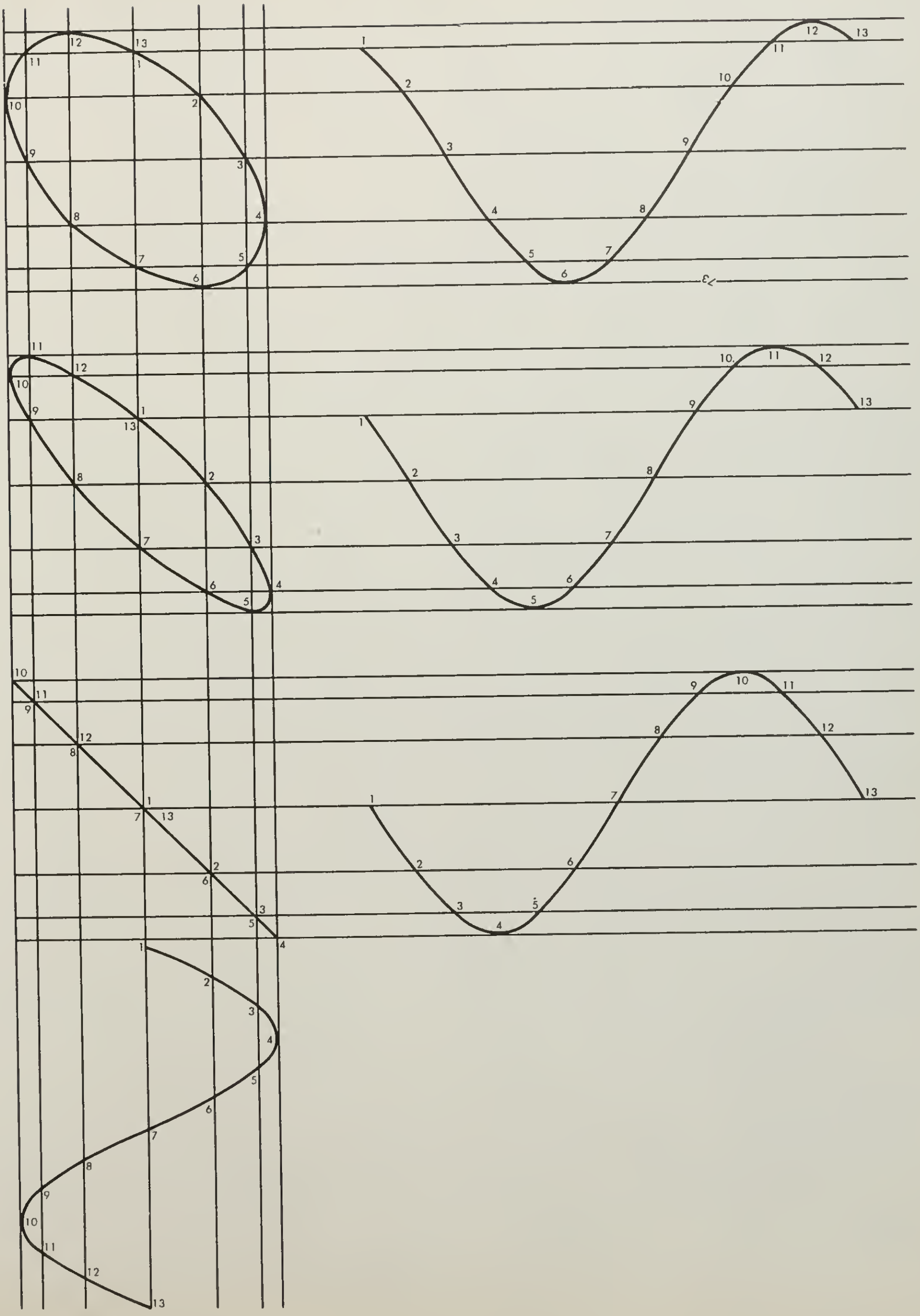
Another common Lissajous figure is the one showing the result of applying voltages having frequencies in a ratio of 2:3 to the horizontal and vertical deflection plates respectively.

As you can see it is possible to draw a great number of pictures for different frequency ratios, but that would be quite a lengthy process. Instead, examine this illustration along with those shown previously and see whether any general rule will work for determining the frequency ratio from a pattern. In this connection note in the illustration of the 2:3 ratio figure that the pattern touches the horizontal line BC at the two points M and N, and that the vertical line AB at the three points J, K, L. From these observations you can form the following general rule for determining the frequency ratio from a pattern: The ratio of the number of points at which the pattern touches the vertical line to the number of points at which it touches the horizontal line is the same as the ratio the frequency applied to the horizontal deflection plates bears to the frequency applied to the vertical deflection plates.

In applying this rule to the C figure in the 1:2 ratio Lissajous figures, don't be confused by the fact that the pattern *appears* to touch the top



Lissajous Figures Indicating Phase Difference Between Sine Waves of Same Frequency


1
2
3
4
5
6
7
8
9
10
11
12
13

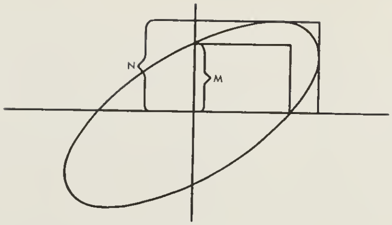
horizontal line only once. Remember that the spot moves through point 1 twice. Another fact to know in applying the rule is that it holds good for any combination of vertical and horizontal lines providing they intersect the pattern.

For accurate determination of frequencies by the Lissajous figures, you need a calibrated variable frequency generator. Apply the output of this generator to one set of deflection plates and the frequency being measured to the other deflection plates. Then determine the frequency ratio from the relationship given in the preceding rule. Read the frequency of the generator and use the ratio to determine the unknown frequency.

DETERMINING PHASE RELATIONS OF VOLTAGES. Lissajous figures may also be used to determine the phase relations of two voltages. The resulting pattern when the same frequency is applied to both sets of plates is a circle only if the two voltages are 90° apart in phase and are of the same amplitude.

In the illustrations on the preceding two pages, note the result when the phase angle is other than 90°. You can determine the approximate phase angle by comparing the patterns resulting from unknown voltages to these patterns.

A more exact method for determining the phase angle is measuring the intercept *m* on the Y-axis, and the maximum deflection *n* on the X-axis. Exercise care both in centering the figure, and in measuring the quantities *m* and *n*. The phase angle is arc sin (m/n).



Phase Angle by m and n

You can also approximate the phase angle by applying the voltages one at a time to the vertical plates while applying the sweep generator output to the horizontal plates. First, synchronize the sweep externally with some voltage having the same frequency as the voltages being observed. Then apply one voltage, note the position on the screen of a peak (or zero point) and measure the length of a complete cycle (or half cycle). Now apply the second voltage and measure the distance the peak (or zero point) is shifted. By comparing the shift to the length of a cycle you can determine the phase angle. For example, suppose, you find that a cycle covers 20 cm on the screen and that the peak of the second voltage is 3 cm at the right of where the peak of the first voltage was located. Since 20 cm corresponds to one cycle or 360°, 1 cm then corresponds to 18°. Therefore, 3 cm means a phase difference of 54°. Since the peak of the second was to the right of the peak of the first, it occurred later in time, and thus the second voltage lags the first by 54°.

CHAPTER 4

Vacuum Tubes and Power Supplies

No doubt the most important single device in radio is the vacuum tube, for without it modern radio science would be impossible. The first receivers used no vacuum tubes. However, modern radio and radar equipment employs a wide variety of vacuum tubes. Most equipment today is multi-tube, employing in some cases hundreds of tubes.

This chapter discusses the various types of vacuum tubes commonly used in radio and electronic circuits. You will learn about tubes with only two elements, tubes with three or more elements, and specially designed tubes, called special purpose tubes. In addition, you will learn about power supplies which furnish energy for the operation of these tubes. What you learn about vacuum tubes and power supplies will largely determine how much you will be able to understand about radio and radar circuits.

EMISSION OF ELECTRONS

In 1883, Edison, experimenting with an evacuated tube containing a heated filament and a metal plate which was near, but not touching the filament, discovered that current flowed between the filament and the metal plate when the plate was more positive than the filament, but that it did not flow when the plate was negative with respect to the filament. Edison thus discovered *thermal emission*, the process by which vacuum tubes obtain their supply of electrons from the cathode.

Thermal Emission

This process of thermal emission, which is known as the *Edison effect*, may be explained by the electron theory of matter, which assumes that the outer electrons in the atoms of metals are very loosely bound to the nucleus. When the filament in an evacuated tube is connected to a heating element, the free electrons move along the filament and collide with other free electrons. The collisions between the electrons generate heat. They also give some of the moving electrons enough energy to overcome the attractive force within the wire and break out from the surface of the wire. These electrons are called *emitted electrons*. The filament which emits the electrons is called the *cathode*. When a plate placed near the emitter has a voltage applied to it making it more positive than the emitter, the plate attracts the negative electrons, causing a current to flow from the filament to the plate, as you can prove by connecting a milliammeter in the plate circuit. If the plate is connected to the negative side of the voltage, making the plate more negative than the filament, there will be no current flow, as the electrons emitted from the heated filament will be repelled by the negative plate.

The escape of electrons from the surface of a metal is analogous to the escape of molecules from a liquid during evaporation. In liquids, the molecules evaporated are the molecules that possess enough energy to overcome the force at the surface which tends to keep them within the liquid. In the case of heated metals, the process is similar and, in a sense, can be considered the evaporation of electrons from a heated surface.

The number of electrons emitted per unit area of an emitting substance is expressed mathematically by the equation,

$$I = AT^2\epsilon^{-\frac{b}{T}}$$

The symbols in this equation stand for the following:

1. *The current I.* I is the current in amperes per square centimeter.

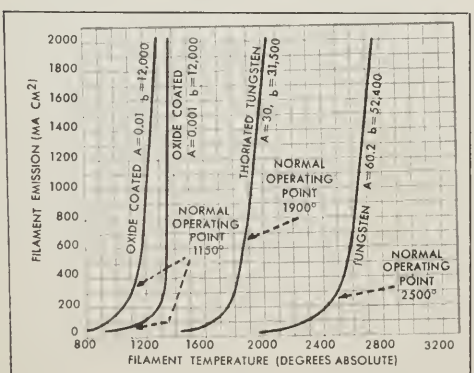
2. *The constant A, the figure of merit.* The value of A varies with different types of emitters. For tungsten, its value is 60.2.

3. *The absolute temperature T of the emitting material.* Absolute temperature is temperature measured by the centigrade scale. It uses −273.1C, the temperature at which there is no electron movement, as its zero point, and is expressed as degrees Kelvin (°K).

4. *The quantity b.* Quantity b represents the work the electron must do to escape through the surface of the metal.

Although you will have little occasion to use the equation, it is important in that it brings out some facts pertinent to the phenomena of thermionic emission. From the equation you can see that I (current per sq. centimeters), which represents the electronic emission of a metal, depends largely on the temperature and on the quantity b. Since these values appear in the exponent, any change in either of them will greatly change the exponent and it in turn the value of I. The value of A is of secondary importance, for any change in its value, even though a large one, is offset by a small change in either T or b.

The chart below shows electron emission for three different types of emitters. Letter A is the constant, the value of which depends upon the type of emitter. The letter b represents work or energy required for the electrons to break through the surface. The filament temperature is given in degrees Kelvin. The temperature at which emission becomes appreciable is labeled normal operating point.



Electron Emission of Different Emitters

Common substances suitable for use as thermionic emitters in vacuum tubes are tungsten, thoriated tungsten, and oxide coated materials. Oxide coated emitters emit at much lower temperatures than either tungsten or thoriated tungsten. For tungsten to be a satisfactory emitter, it must be operated at very high temperatures. In spite of the large amount of power required to operate tungsten emitters, due chiefly to the large value of b, it is widely used as an emitter in a large number of high power vacuum tubes because of its durability. Tungsten covered with thorium emits at temperatures appreciably lower than pure tungsten.

Directly and Indirectly Heated Tubes

Emitters are heated either by passing a current through the emitter or by a heating element located near the emitter. A vacuum tube in which the emitter is heated directly is called a *directly heated* tube, and the emitter in it is known as the *filament*. A tube in which the emitter is heated by a heating element is an *indirectly heated* tube, and the emitter unit is called the *cathode*. The heating element is known as the *heater*. Either DC or AC can be used to supply the current for heating in either a directly or an indirectly heated tube. However, tubes directly heated with AC generate an objectionable AC hum in the form of AC variations in plate current.

Filament Voltages

The voltage and current required for heating the emitter to the proper temperature varies considerably in different tubes. Any tube manual will give you the filament voltage and current rating of any particular tube. Current ratings vary considerably for different tubes even with the same voltage ratings.

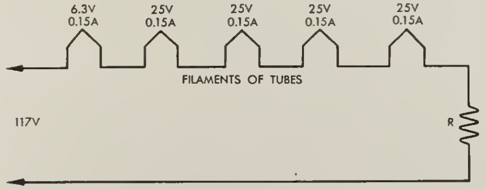
Common filament voltage for tubes are 1.4, 2.0, 2.5, 5.0, 6.3, 12.6, 25, 50, 70, and 117 volts. Tubes commonly used are the 6.3 and 12.6 volt tubes. A 1.4 volt tube is designed for use where voltage is supplied by a dry-cell battery. A 117 volt tube can be heated directly from a 117 volt AC line. Tubes generally are referred to by their filament voltages. Thus, in the 6.3 volt tube, or in the 117 volt tube, the voltages stated are the voltages required for filament operation.

Connecting Filaments

Filaments may be connected in parallel when they have the same voltage rating. This is a common practice in receivers employing a power transformer for stepping down the line voltage to suitable voltages for operation of the receiver. For example, a power transformer which receives voltage from a 117 AC power source may have three secondary windings, one to supply voltages for the plates of the tubes, one to supply voltage for the rectifier filaments, and one to supply 6.3 volts for the filaments of tubes other than the rectifier tubes. All these 6.3 volt tubes can be connected in parallel across the 6.3 secondary winding. The total current drain through this winding is equal to the sum of the current ratings of the individual tubes connected across it.

AC-DC receivers, types of receivers which operate either from AC or DC, do not employ power transformers but obtain power directly from the line. In connecting filaments in these receivers, connect them in series across the power source (line). However, for this connection to work satisfactorily, be sure that the combined voltage rating of all the tubes equals the power supply voltage. If this is not the case, then drop the applied voltage to the combined voltage rating of all the tubes by connecting a resistor in series with the power source. Furthermore, when connecting the filaments in series, be sure that all tubes have equal current ratings. If any tube has a rating lower than the others, connect a resistor in parallel across this tube. This resistor carries additional current, and makes the current rating of this tube equivalent to that of the other tubes.

The following examples illustrate typical conditions that you will encounter in connecting filaments of vacuum tubes:



Example 1.

Problem. In the circuit shown above, 5 filaments are connected in series. Find the resistance and the power rating of the resistor, R, using the values indicated.

Solution. The sum of the voltage ratings of the 5 tubes equals 106.3 volts. Subtracting this voltage from the voltage of the power source, 117 volts, gives 10.7 volts, which is the voltage drop across resistor R. Then according to Ohm's law, The resistance equals

$$R = \frac{E}{I},$$

$$R = \frac{10.7}{0.15} = 71 \text{ ohms.}$$

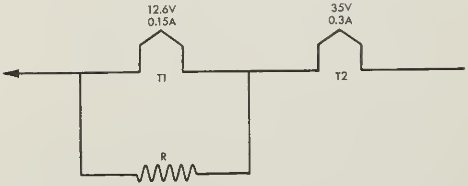
Using the power formula,

$$P = EI,$$

$$P = 10.7 \times 0.15$$

The power equals

$$P = 1.605 \text{ watts}$$



Example 2.

Problem. In the circuit shown above, two filaments are connected in series. T_2 requires 0.3 amperes, but T_1 needs only 0.15 amperes. Find the value of R and its wattage rating.

Solution. If 0.3 ampere is required for T_2 and T_1 needs only 0.15 ampere, the other 0.15 ampere must flow through R. By Ohm's law,

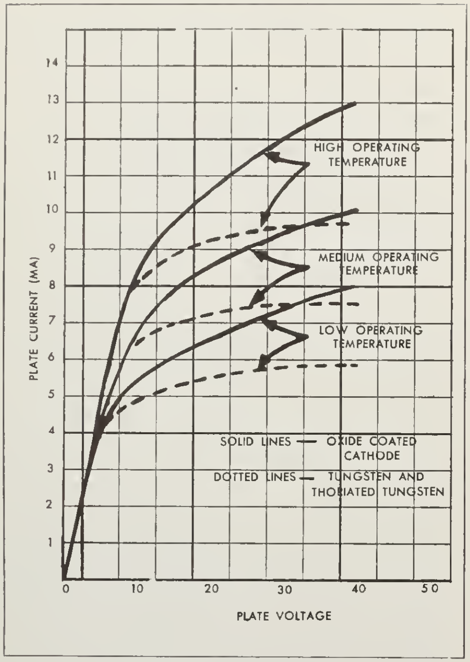
$$R = \frac{E}{I} = \frac{12.6}{0.15} = 84 \text{ ohms}$$

Finding wattage rating of R,

$$P = 12.6 \times 0.15 = 1.89 \text{ watts}$$

DIODES

The simplest form of radio tube contains two electrodes—a cathode, and a plate. This type of tube is called a *diode.* In the diode, positive voltage is supplied by a suitable electrical source connected between the plate terminal and the cathode terminal. When a positive voltage is applied to the plate, electrons will flow from the cathode to the plate and return to the cathode through the external plate battery circuit. If a negative voltage is applied to the plate, the free electrons in the space surrounding the cathode will be forced back to the cathode and no plate current will flow. Thus, a diode permits the electrons to flow from its cathode to its plate, but not from the plate to the cathode. When an alternating voltage is applied to the plate, the plate will alternately become positive and negative. Plate current will flow only during the time when the plate is positive and since it flows through the



Diode Ep-Ip Characteristics

tube in only one direction, it is said to be *rectified.*

Diode rectifiers are employed in AC operated receivers and transmitters to convert AC into DC for supplying the plate, screen, and grid bias voltages for all other tubes in the receiver or transmitter. They are also used as detectors or demodulators in receivers. Rectifiers and detectors are discussed in detail later.

The relation between plate current and plate voltage in a diode is given in the above chart. It shows the results obtained in a typical diode operated at several cathode temperatures. Note that at high plate voltages, plate current (the number of electrons attracted to the plate) is determined largely by cathode temperatures, being practically independent of plate voltage, and that at low voltage, plate current depends entirely on plate voltage and is independent of cathode temperature.

Tube action at high plate voltages results from the fact that the plate attracts electrons from the cathode as rapidly as they are emitted. Under these conditions, increasing the plate voltage makes no difference, for the plate current is already equal to the total emission of the cathode, and the tube is said to be operating at *emission saturation.* When the cathode temperature is raised, however, the cathode emits electrons at a faster rate, emission saturation occurs at higher plate voltage, and a higher current flows.

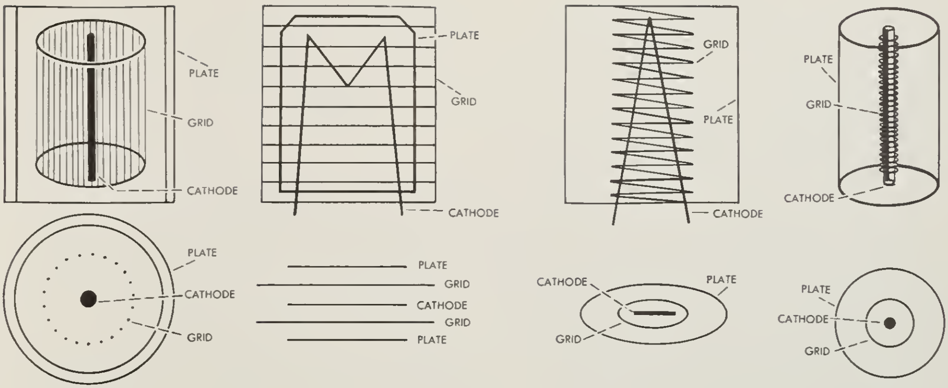
At low plate voltages, electrons are emitted by the cathode faster than they are attracted to the plate. Electrons, therefore, tend to remain in the space between the cathode and plate and produce a negative charge (called *space charge*) which exerts a repelling force upon other electrons being emitted from the cathode. At any instant, the number of electrons traveling between the cathode and plate cannot be greater than the number required to neutralize the attraction of the plate voltage. All electrons in excess of this number are repelled back to the cathode. Thus, plate current is independent of cathode emission as long as the negative space charge exists.

At higher plate voltages, the plate attracts more of the electrons emitted from the cathode and reduces the tendency for electrons to remain in the space charge region and to repel others. The higher the plate voltage, the greater the proportion of electrons attracted and the less the space charge.

Beyond a certain plate voltage, however, emission saturation occurs, as explained previously, and additional plate voltage has little effect in increasing plate current. The point where no more current flows when the plate voltage is increased is called the *saturation point.* For any further increase in current, there must be an increase in cathode temperature.

TRIODES

A triode is a three-element vacuum tube, containing a grid, a cathode, and a plate. The grid is usually an open coil or mesh of fine wire and is placed between the cathode and plate. The turns of wire in the grid are spaced far enough apart so that the passage of electrons from the cathode to the plate is virtually unobstructed. The number of electrons that pass from the cathode to the plate constitute plate current flow in the tube. If the grid is made more negative than the cathode, some of the electrons emitted by the cathode will be repelled by the grid and, therefore, will not reach the plate; plate current accordingly will be decreased. It follows that as the



Typical Triode Structures

grid voltage is made increasingly negative, fewer and fewer electrons will reach the plate. Eventually, when the grid is made negative enough with respect to the cathode, no electrons will reach the plate and plate current will be cut off entirely. The point at which plate current ceases to flow is called *cut-off*, and the negative grid voltage required to stop the plate current is called *cut-off voltage.*

Going in the opposite direction from cut-off, plate current will increase as you make the grid less negative. When the voltage on the grid becomes zero, it will have no effect on the passage of electrons to the plate. If the grid is made positive with respect to the cathode, it will aid in drawing electrons from the space charge between the plate and grid, and there will be an increase in plate current. However, not all of these electrons will go to the plate. The grid itself will act as an anode (plate) and a current, called *grid current*, will flow through it. If the grid is made positive enough, it will attract the majority of the total number of electrons emitted by the cathode since it is closer to the cathode (and space charge) than the plate.

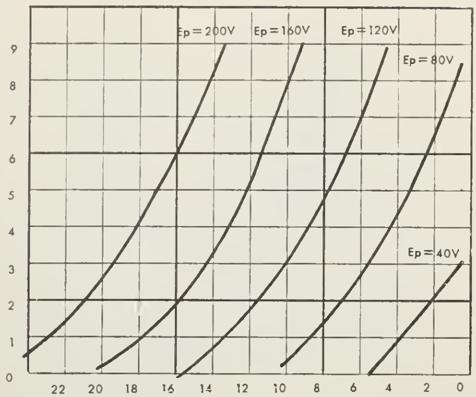
Because of this closer proximity to the cathode, a comparatively small change in the voltage on the grid in a triode has the same effect on plate current flow as a much larger change in the plate voltage. It is this feature that gives the triode its chief property—*amplification.*

Here is how amplification is achieved. A small alternating voltage applied to the grid causes comparatively large variations in plate current at the frequency of the applied voltage. If a resistor, called the *load resistor*, is placed in series with the plate circuit, the varying plate current, in flowing through this resistor, will develop a voltage which has a frequency equal to that of the AC voltage applied to the grid. However, the magnitude of the voltage across the load resistor will be much greater than that of the voltage applied to the grid. Since voltage appearing in the output is greater than the voltage applied to the grid, it is said to be *amplified.*

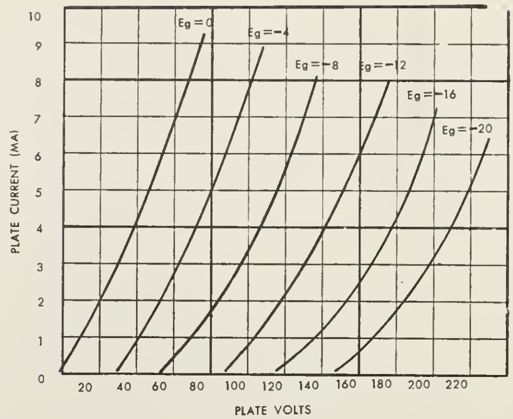
Characteristic Curves of Triodes

The behavior of a triode is best described by means of curves, called *characteristic curves.* These curves are obtained by keeping one of the three quantities, grid voltage, plate voltage, and plate current, constant and observing the relationship between the other two. Since any of the three quantities may be kept constant, there are altogether three families of curves.

In the three families of curves shown on the next page, each curve is obtained by applying specific voltages to the grid and plate and by observing the plate current. The Eg-Ip curves are obtained by keeping the plate voltage (Ep) constant and plotting the plate current (Ip) against the grid voltage (Eg). The Ep-Ip curves are obtained by holding the grid voltage (Eg) constant and plotting the plate voltage (Ep) against the plate current (Ip). The third family, the Ep-Eg curves, result from keeping the plate current constant, and plotting the grid voltage (Eg) against the plate voltage. The most commonly used curves are the Ep-Ip curves. You can find these curves in any tube manual.

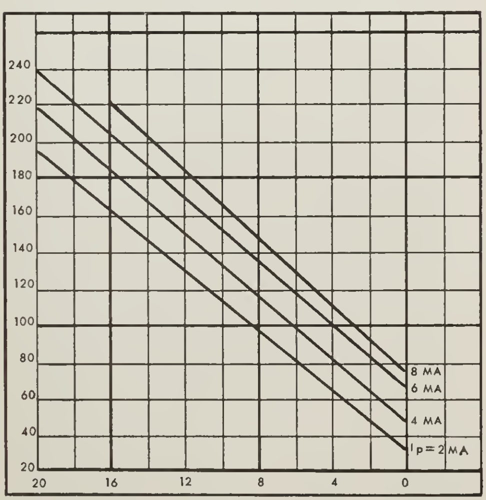


Eg-Ip Curves of Triode



Ep-Ip Curves of Triode

Curves of the type given are called *static curves*, since they are obtained by using DC voltages under non-operating conditions. Characteristic curves obtained under operating conditions are called *dynamic curves.* Dynamic curves are useful only when circuit elements are operated under the same conditions that the curves were obtained. Static curves, although obtained under non-operating conditions, will give you considerable information relative to the tube even under operating (dynamic) conditions. For this reason they are the ones usually given.



Ep-Eg Curves of Triode

AMPLIFICATION FACTOR. The ratio of the plate voltage change required for a given change in plate current to the change in grid voltage which will produce the same change in plate current is called the *amplification factor* of a tube. It is commonly expressed by the Greek letter μ (MU) and expressed mathematically,

$$\mu = \frac{\Delta ep}{\Delta eg} \quad (\text{When } Ip = \text{constant})$$

where Δep is change in Ep, and Δeg is change in Eg.

In triodes, the amplification factor varies from 3 to about 100.

PLATE RESISTANCE. The ratio of a small change in plate voltage to the change in plate current it produces when the grid voltage is constant is called the dynamic, or AC plate, resistance of a tube and is represented by Rp. The formula for plate resistance is,

$$Rp = \frac{\Delta ep}{\Delta ip} \quad (\text{When } Eg = \text{constant})$$

Plate resistance is important in determining the performance of a tube in a given circuit since it forms a voltage divider with the load resistance. This divides the voltage across the tube in a ratio depending on the value of the plate and load resistance. The larger the load resistance with respect to the plate resistance, the smaller the ratio and the greater the amplification of the circuit. Any decrease in this ratio results in an increase in amplification; any increase results in a decrease in the amplification.

TRANSCONDUCTANCE. Another important tube constant is called transconductance and is represented by gm. Transconductance tells how much plate current change is caused by a small change in grid voltage. The formula for this constant is,

$$gm = \frac{\Delta\, ip}{\Delta\, eg} \quad (When\ Ep = constant)$$

The unit of transconductance is the *mho.* Transconductance bears a mathematical relation to the amplification factor and plate resistance of a tube and is also expressed in terms of these constants as,

$$\mu = gm\ rp$$

TUBE CONSTANTS AND CHARACTERISTIC CURVES. The tube constants μ, Rp, and Gm may be determined with reasonable accuracy from Ep-Ip curves. To understand how they can be determined in this way, refer to the Ep-Ip curves shown on page 4-6. On these curves, note that when Eg changes from −4 to −8 volts with Ip remaining at 5 ma, Ep changes from 88 volts to 125 volts. Thus,

$$\mu = \frac{88-125}{(-4)-(-8)} = \frac{-37}{4} = 9\frac{1}{4}$$

According to a tube manual, the amplification factor for the tube for which the curves are drawn is 9. Thus for practical purposes the curves provide values which give results reasonably correct. In general the answer you obtain from the curves will vary somewhat, depending upon the values of Eg and Ip.

By assuming a value for Eg and then comparing the change in Ep with the change in Ip, you can approximate rp. For example, assume Eg is equal to −8. When Ip changes from 4 ma to 5 ma, Ep changes from 112 volts to 125 volts. Hence,

$$rp = \frac{125-112}{.005-.004} = \frac{13}{.001} = 13{,}000\ ohms$$

Mathematically, the ratio expressed by the plate resistance is the slope of the curve at a particular point. On the same curve or on different curves, it has different values at different points. In other words, plate resistance depends, to some extent, on the conditions under which the circuit is operated.

When finding the gm using the same Ep-Ip curves illustrated previously, assume that Ep is equal to 100 volts. Thus when Eg changes from −4 volts to −8 volts, Ip changes from 6.4 ma to 2.8 ma according to the curves. Therefore by substituting in the equation for transconductance,

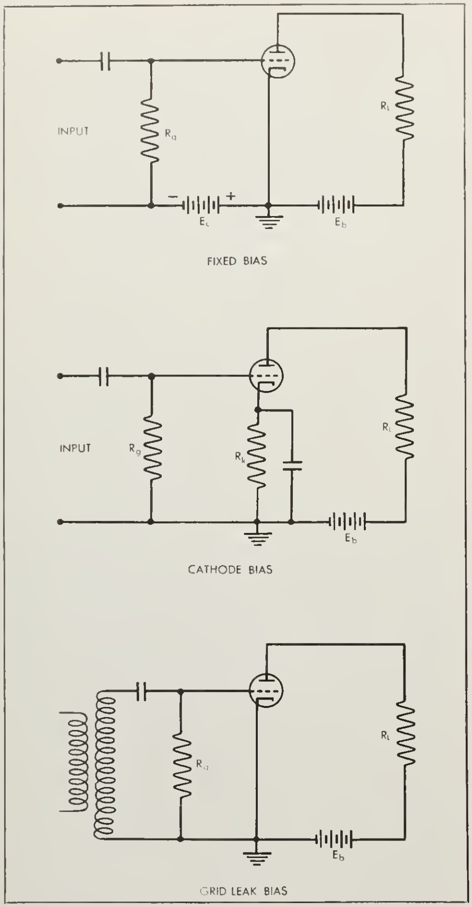
$$gm = \frac{.0064-.0028}{(-4)(-8)} = \frac{.0036}{4} = .0009\ mhos,\ or\ 900\ micromhos.$$

Bias

A sufficiently negative DC voltage added to an AC voltage will make the AC voltage entirely negative. This means that if an AC voltage applied to the grid of a tube is made negative by the addition of a negative DC voltage, the grid will not conduct current as is the case when the grid is positive with respect to the cathode but will make the plate current vary in proportion to changes in grid voltage. The negative DC voltage applied to the grid, which makes the applied voltage negative, is called *bias voltage.* Normally, bias is applied directly and continuously to the grid, and not to the applied voltage before it reaches the grid. Bias must not be too high, or the grid voltage will exceed the cut-off value and introduce *distortion*, a deviation from true reproduction of grid voltage which is discussed later. There are three general types of bias employed to furnish grid-bias voltage for vacuum tubes. They are *fixed bias*, *cathode bias*, and *grid leak bias.*

FIXED BIAS. Fixed bias is usually supplied by a battery, called a C battery. However, a battery often is inconvenient and is usually avoided unless there are no other practical sources of power. Another method of providing fixed bias is the voltage divider system. Negative voltage is taken from a voltage divider in the power source of the tube.

CATHODE BIAS. One of the simplest methods of supplying a steady negative voltage on the grid of a tube is by cathode bias. This method consists of placing a resistor between the cathode and ground, and applying to the grid the voltage drop developed by the plate current flowing through this resistor. To keep this bias voltage applied to the grid constant and free from the effect of the AC variations in plate current, a



Types of Bias

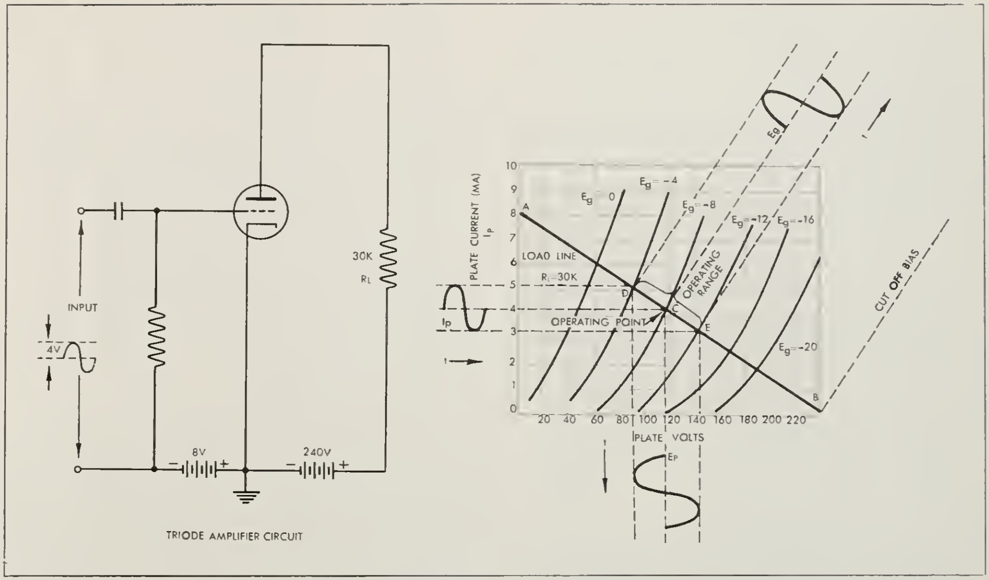
condenser is connected in parallel with the cathode biasing resistor. This condenser charges to the average value of the cathode-to-ground voltage, and the cathode voltage cannot change unless the charge on the condenser increases or decreases. However, since the condenser is quite large and the resistance through which it charges or discharges is likewise fairly large, the time of charge or discharge is too great for the cathode voltage to be affected by the high-frequency variations in the plate current produced by the signal voltage applied to the grid. Hence the cathode voltage is held constant by the condenser action.

The condenser connected across the cathode resistor is called a *bypass condenser*. In the cathode bias circuit the grid is connected to ground through the grid resistor Rg. As no DC current flows through the grid resistor, the grid is at ground potential. The cathode is positive with respect to ground because plate current flows through the resistor between it and ground. Thus the cathode is more positive than the grid. In other words, the grid is negative with respect to the cathode by the cathode-to-ground voltage.

GRID LEAK BIAS. Grid leak bias makes use of the flow of grid current during a portion of the input cycle to develop negative voltage for biasing the grid. The circuit is usually like the one at the left. When the input voltage across the transformer secondary is positive, grid current flows and the condenser charges. When the secondary voltage becomes zero or negative, the condenser discharges through the closed path of the secondary and resistor Rg. Since this resistor is quite large, the condenser discharges slowly during the entire negative portion of the input cycle. The discharge current through Rg is in such a direction as to make the grid end negative with respect to ground. Hence the average grid potential is negative, provided the grid becomes positive for at least a small portion of each cycle. The cathode is at ground potential; hence the grid-to-cathode potential is negative.

Load Lines

You can apply characteristic curves in analyzing an amplifier circuit as shown in the illustration in connection with the triode circuit and the Ep-Ip characteristic curves on the next page. The straight line drawn across the Ep-Ip curves is called the *load line*. It is the result of plotting the values of plate current and plate voltage in the triode circuit and connecting all of the plotted points by a straight line. One extremity of the load line is the place where the plate current is zero and where the voltage drop across the load resistor is zero. The other extremity is the point where the plate current is absolutely unlimited by the tube—that is, the place where the tube acts like a short circuit. Current flow at this point is limited only by the load resistor and is obtainable by Ohm's law. When the tube is shorted, its plate voltage is zero. The slope of the load line is determined by the load resistor, hence the name *load line*. The greater the load resistance, the less the slope. The load line across the Ep-Ip curves is based



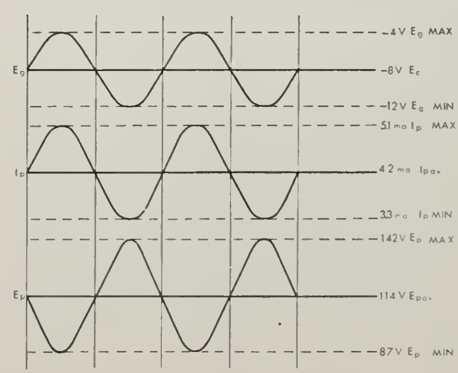
Analyzing the Operation of a Triode

on the triode circuit shown, and the curves are likewise based on this same circuit.

Here is how the line is located. Suppose, in the triode circuit, that the plate current drops to zero; the voltage on the plate will then be 240 volts. Thus one end of the load line is located at a point where Ip = 0 and Ep = 240. When the tube is shorted, the plate voltage is zero and the current is the supply voltage divided by the load resistance $\frac{240}{30K}$ or 8 ma. At this point Ep = 0, and Ip = 8 ma. This locates the other end of the load line. The load line is useful in that if one of the quantities Ep, Eg, or Ip is given, you can determine the value of the other two with the load line.

In the circuit of the triode, the battery bias is 8 volts. With no signal applied to the grid you can determine the static Ep and Ip from the intersection (C) of the −8 volt curve and the load line. The point C, called the operating point, corresponds to Ip = 4.2 ma and Ep = 114 volts. Suppose now you apply an AC voltage of 4 volts peak amplitude to the grid. Since the grid is biased at −8 volts, this 4 volt signal will cause the grid voltage to vary between −4 volts and −12 volts. The two curves Eg = −4 and Eg = −12 intersect the load line at points D and E, which correspond to Ep = 87 volts and 142 volts and Ip = 5.1 ma and 3.3 ma respectively. The portion of the load line between D and E is called the operating range since the normal voltage variations during operation stay within these limits.

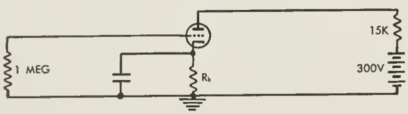
The waveshapes of the plate voltage and current and the grid voltage of the triode curves are reproduced below in time reference to each other. They point out that the plate voltage



Phase Relation of Eg, Ip, and Ep

variation is 180° out of phase with the variation in grid voltage. The grid voltage variation, eg, is the applied AC signal. It has a peak amplitude of 4 volts. The variation in plate voltage, ep, is the AC component of Ep and has a peak amplitude of 27 volts. The amplification of the stage is 27 volts divided by 4 volts, or 6.75. The reason that the amplification of the stage is less than the amplification factor of the tube (9 in this tube) is that a voltage divider is formed by rp and R_1 and only the voltage across R_1 is usable as an output voltage. This is discussed in more detail in a later chapter.

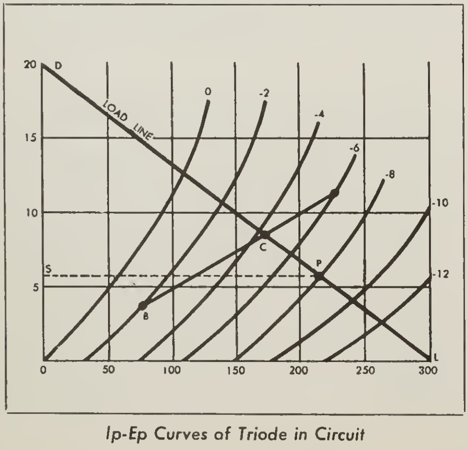
The following problems illustrate further uses of load lines:



Problem 1: Find the value of Rk that will give 8 volts cathode bias in the circuit above.

Solution: Using the family of Ip-Ep curves shown below, locate the intersection of the Ep-Ip curve marked −8 volts and the load line. This gives point P, which is projected on the vertical axis (Point S) to determine Ip, which turns out to be 5.5 ma. Determine the value of Rk by Ohm's law. Thus,

$$Rk = Ek/Ip$$
$$= 8/5.5 \times 10^{-3}$$
$$= 1450 \text{ ohms.}$$



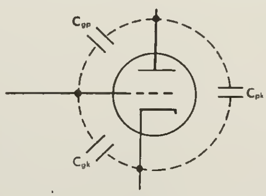
Ip-Ep Curves of Triode in Circuit

Problem 2. Using the same circuit, find the cathode bias if Rk = 500 ohms.

Solution: Assume that Ip = 4 ma; then the drop across Rk = $500 \times 4 \times 10^{-3} = 2$ volts. Locate the point on the Eg = −2 curve which is opposite 4ma on the vertical axis. This gives point B. Assume the Ip = 12 ma, and locate a corresponding point in a similar manner (point A). Connect these two points. This line will intersect the load line in some point C. The point C lies between the Ep-Ip curves marked −4 and −6. The approximate value is determined by proportional distance from the two curves. In this case, the value is about −4.3 volts. This means the cathode is 4.3 volts more positive than the grid.

Interelectrode Capacitance

You are familiar with the fact that two conductors separated by a dielectric form a condenser, the capacitance of which depends upon the size of the conductors, the distance between them, and the dielectric. In a vacuum tube, the capacitance between the electrodes, called interelectrode capacitance, connects the electrodes together capacitatively, as shown in the illustration below. The largest capacitance usually



Interelectrode Capacitance of Triode

exists between the plate and cathode because of the large area involved. Cgk is the grid-to-cathode capacitance, and Cgp is the grid-to-plate capacitance. Though these capacitances are small, being measured in micromicrofarads, they couple the elements of the tube. At low frequencies, the coupling is negligible but, with increase in frequencies, it becomes increasingly objectionable. The grid-to-plate capacitance has the most disturbing effect on tube operation since it couples energy from grid circuit to plate circuit and vice versa.

TETRODES

A tetrode is a four-electrode vacuum tube. Essentially, it is a triode with the addition of a screen grid located between the control grid and the plate. The addition of the screen grid materially reduces the grid-to-plate capacitance, making the tetrode more useful for higher frequency work than the triode. The voltage applied to the screen grid is positive, usually somewhat less positive than that at the plate, and creates an electrostatic field within the tube. Because this electrostatic field is closer to the control grid than that of the plate, it exerts a greater influence than that of the plate on the number of electrons drawn from the space charge. Furthermore, since the screen voltage is held constant, the varying voltage at the plate only varies the electrostatic field between the screen and plate but does not affect electrons near the control grid. In general, increasing screen grid voltage increases plate current. While the screen grid, being positive, shares electrons with the plate, the construction of this screen grid is such that the majority of electrons pass through and on to the plate. The division of the electrons depends upon the voltages applied to the electrodes. The higher the screen voltage, the more of the total number of electrons the screen takes. Up to a certain limit this increases the plate current, too, because the screen draws more electrons out of the space charge and increases the total amount. But when the maximum number of electrons obtainable from the space charge are being obtained, the total current does not increase with higher screen voltages. Then the screen and plate have to divide the available electrons between them. At this point, higher screen voltage increases the screen current at the expense of the plate current, and plate current actually decreases with an increase of screen voltage. This action is called the *negative resistance* characteristic of the tetrode tube.

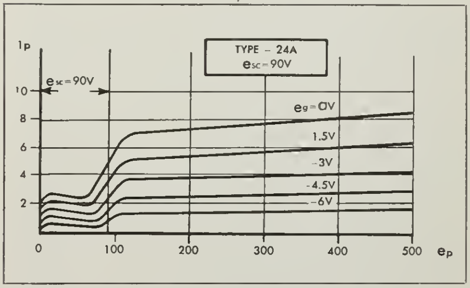
Secondary Emission

An effect called *secondary emission* is produced when electrons striking the plate or screen grid with sufficient force to knock other electrons from it. These electrons, thus freed, travel to the electrode exerting the greater force on them—that is, the electrode with the more positive potential. Flow of electrons from screen grid to plate due to secondary emission is negligible since they are liberated on the side away from the plate and must pass through the screen to get to the plate. On the other hand, the flow of electrons from the plate to the screen grid may become appreciable when the screen grid is more positive than the plate.

At very low plate voltages, the number of secondary electrons produced at the plate is small and there is a tendency for space charge electrons to accumulate in a cloud in front of the plate and turn back some of the arriving electrons. This space charge is just like the one at the cathode, but here the electrons turned back are attracted to the screen grid. Under these conditions, the plate current depends upon the plate voltage and is much less than the total space current.

Characteristics of Tetrode

When the plate voltage is less than the screen grid voltage, but not low enough to cause a space charge to form in front of it, the plate current decreases as the plate voltage increases (see the characteristic curves of tetrode tubes). This



Typical Ip-Ep Tetrode Characteristics

is another way of producing a negative resistance characteristic. While the number of primary electrons that the plate receives is independent of the plate voltage (dependent upon screen grid and control grid potentials), the number of electrons lost to the screen through secondary emission increases as the plate voltage increases since electrons strike the plate with greater force. When the tube is used as a negative resistance device, it is called a dynatron—a device sometimes used as an oscillator. The transition from this region of decreasing plate current with increasing plate voltage to the region where the plate current becomes practically independent of the plate voltage occurs when the screen and plate are at approximately the same voltage.

When the plate voltage is appreciably more positive than the screen grid, the plate retains the primary electrons it receives and in addition receives some secondary electrons from the screen grid. This results in the plate current being very nearly equal to the total space current and substantially independent of the plate voltage since the total space current depends primarily upon the potentials of the screen grid and control grid.

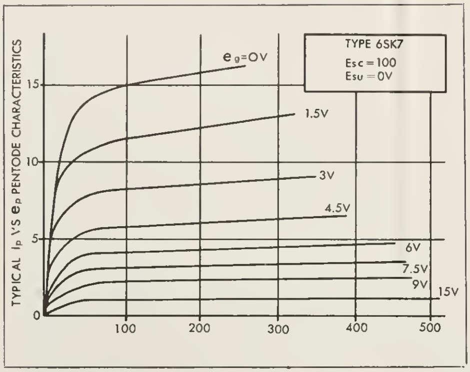
To overcome the effects of secondary emission, the tetrode usually operates at a point where the plate voltage is high. In this condition the tetrode has high plate resistance and amplification factors ranging from 500 to 800. The transconductance is about the same as that of a triode operating with similar plate current. The plate-to-grid capacitance, however, is lower than in a triode.

The high amplification factor and lower plate-to-grid capacitance make the tetrode particularly useful as an amplifier of radio frequency voltages.

PENTODES

To overcome the effects of secondary emission, there are tubes in which a third grid, called the *suppressor grid,* is added between the screen grid and the plate. This type of tube is known as a *pentode,* or five-electrode tube. The suppressor grid is usually connected directly to the cathode in such a way that it repels back into the plate the secondary electrons emitted from it. The suppressor grid does not materially affect the flow of primary electrons to the plate since their flow depends almost entirely upon the screen grid and control grid voltages.

Although the screen grid in either the tetrode or pentode greatly reduces the effect of the plate voltage upon plate current flow, the control grid, because of its closeness to the cathode in the space charge region, still is able to control the total space current flow, and consequently the plate current. For this reason the transconductance of a pentode (or a tetrode) is of approximately the same value as that of a triode of similar structure. On the other hand, since the plate voltage has relatively little effect upon the plate current flow in a pentode, the values of Rp and μ are very high. This is obvious if you remember the definitions of these tube constants. The amplification factor of a pentode is in the neighborhood of 100 to 1500, and the plate resistance from one-half to one megohm. The high plate resistance makes it difficult for a pentode to utilize much of the advantage of its high amplification factor. Therefore, pentode amplifiers using resistance capacitance coupling have gains of only 100 to 200.



Characteristic Curves of a Pentode

Pentodes Characteristics

A typical set of characteristic curves is shown just above. The fact that the plate voltage is relatively ineffective in controlling the plate current is shown by the small slope of the curves beyond the point where the plate voltage is high enough to keep the electrons in the space between screen grid and plate from being attracted back to the screen grid. The plate voltage at which this occurs is less than the screen grid potential since the electrons enter the space with sufficient velocity to cause them to travel on to the plate in spite of the higher screen grid voltage.

Pentodes are used mostly as radio frequency voltage amplifiers. They are also suitable as power amplifiers, having greater power sensitivity than triodes. Power sensitivity is the ratio of output power to AC grid voltage. For a triode to develop the same power output as a pentode, the amplification factor of the triode has to be quite low and, because of this low amplification factor, a larger input voltage would be required to obtain power output comparable to that of a pentode.

SPECIAL PURPOSE TUBES

In addition to the types of tubes discussed thus far, there are tubes of special design intended to give optimum performance in a particular application or to combine in one envelope functions normally requiring two or more tubes. Tubes of this type are called *special purpose tubes.* The most commonly used special purpose tubes are beam power tubes, remote cut-off tubes, multipurpose tubes, UHF tubes, gas filled tubes, and cathode ray tubes.

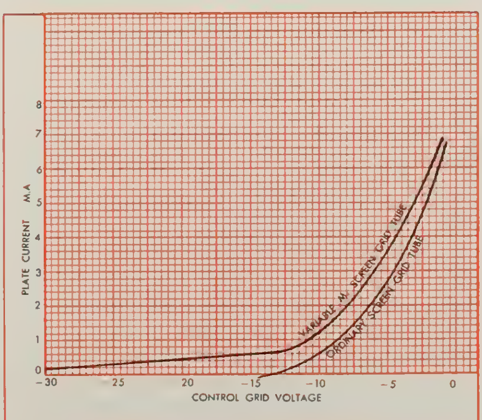
Beam Power Tube

A beam power tube is a special type of tetrode with characteristics similar to those of a pentode power amplifier. Although a beam power tube has no suppressor grid, the effect of one is obtained by the creation of a space charge between the screen grid and plate. This is accomplished by appropriate spacing of the screen and plate and by concentrating electron flow into well-defined beams by means of beam-forming plates and specially aligned grid wires. The control and screen grid wires are arranged in line toward the plate in such a way that electrons can pass between them without changing direction. This creates a space charge that causes secondary electrons from the plate to be repelled back to the plate without adversely affecting the normal flow of electrons from the cathode to the plate.

The characteristic curves for beam power tubes differ from pentode curves in that the transition between the region where the plate current is practically independent of the plate voltage and to the one where the plate current is controlled by plate voltage is more abrupt than in the pentode and occurs at a lower value of plate voltage.

Remote Cut-Off Tube

A remote cut-off tube is a tetrode or pentode voltage amplifier tube constructed in such a way that space current trails off gradually at highly negative control grid voltages rather than quickly at a well defined cut-off point. Other names for remote cut-off tubes are variable-mu and super-control tubes. The control grid structure is spaced non-uniformly, giving a different amplification factor and cut-off voltages at different parts of the grid structure. This means that when the grid goes negative enough to cause current through part of the grid to cease it may still flow through some other portion which requires a more negative voltage to cut off the tube. For this reason the plate current flow tapers off slowly, and requires a much more negative voltage than an ordinary tube for the plate current to reach zero. Remote cut-off tubes are used in circuits employing automatic volume control and in circuits in which the gain is controlled by varying the bias on one or more tubes.



Variable-Mu Tube vs. Ordinary Tube

Multi-Purpose Tubes

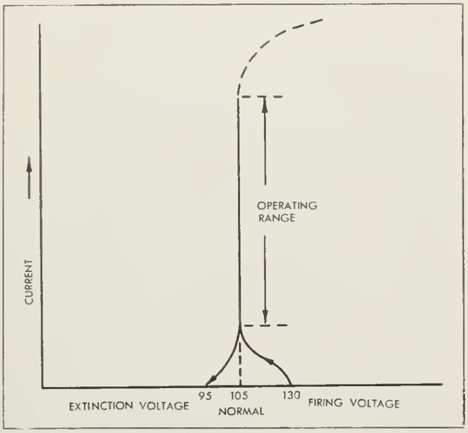
A multi-purpose tube is one which contains two or more tubes in the same envelope. Frequently you will find two diodes or two triodes located in the same envelope. Such tubes are known as duplex diodes or duplex triodes. Some other multi-purpose tubes in common use are duplex-diode-triodes, twin pentodes, and diode-heptode tubes.

UHF Tubes

UHF (ultra-high-frequency) tubes are especially constructed for operation at higher frequencies, where it is necessary to reduce interelectrode capacitances and to cut down transit time (the time electrons require to travel from the cathode to the plate). This is done by making the electrodes quite small and spacing them very closely together. Because of these construction features, the power handling ability of this type of tube is somewhat less than that of tubes used at lower frequencies. In UHF tubes, there is frequently no tube base. Connections to the electrodes are made through pins which protrude through the envelope in a way which keeps the leads short and minimizes capacitance between them. Three special types of UHF tubes, the *acorn tube*, *door-knob tube*, and the *lighthouse* tube, are so named because of their shapes and sizes. Acorn and door-knob tubes are available in diodes, triodes, and pentodes. The lighthouse tube is designed to fit directly into the end of the concentric tubing used to form the tuned or tank circuit in UHF circuits. By directly connecting the tube in this manner, losses due to connecting wires are eliminated.

Gas-Filled Tubes

Although the presence of gas in a high vacuum tube is objectionable, there are special purpose



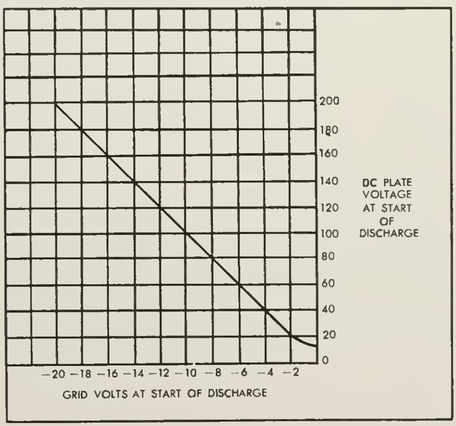
Characteristics of Gas-Filled Diode VR 105-30

tubes which are gas filled. Common gasses for this purpose are neon, nitrogen, argon, and mercury vapor. The gas filled tube is characterized by the fact that the voltages applied to the electrodes have little effect on the current flow once they reach a certain minimum value. This is due to the fact that the gas becomes *ionized*—that is, the molecules of gas lose electrons through collision with other electrons in the interelectrode space and become positively charged ions which drift toward the cathode and attract more electrons out of the cathode, thus increasing the current flow. There is a certain minimum voltage difference between plate and cathode which will cause the electrons to move with sufficient velocity to produce ionization. This is called the ionization potential or firing voltage. Once the tube is ionized, the plate voltage can be reduced somewhat without affecting the flow of electrons. There is a minimum value of plate voltage necessary to maintain ionization. This voltage is called the extinguishing or de-ionizing voltage.

When a gas-filled tube conducts current, it behaves like a low resistance which varies in value in such a way that the voltage developed across it remains practically constant over a moderately wide range of current. This characteristic makes gas filled tubes useful as voltage regulator tubes, mercury vapor rectifiers, protective devices, and thyratrons.

A *voltage regulator* tube is a tube especially designed to keep the voltage at a practically constant value regardless of voltage regulation in the power supply or variation in load current. An example of a voltage regulator tube is VR105-30. The first number in the tube designation indicates the value of the voltage that the tube maintains constant; the second number is the maximum amount of permissible tube current. The use of the voltage regulator tube is described later in this chapter in voltage regulator circuits.

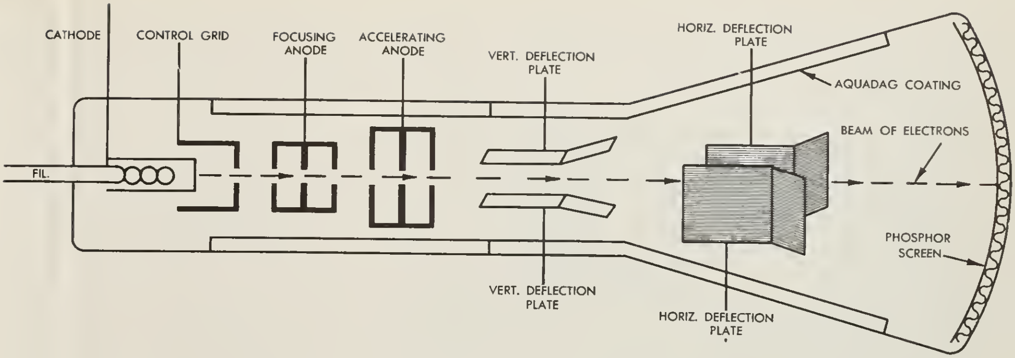
Another type of gas filled tube in wide use is the *mercury vapor tube*. This tube contains a small amount of mercury which vaporizes when the cathode is heated and which ionizes when plate voltage is applied. The plate-to-cathode voltage drop in the mercury vapor tube is practically constant at 15 volts regardless of the amount of plate current flow. Mercury vapor diodes are widely used as rectifiers in power supplies where considerable current is drawn.



Characteristics of Thyratron

A *thyratron* is a gas filled triode or tetrode in which the grid controls the ionization voltage. Tube 884, a triode, is a widely used thyratron. The chart above shows its characteristics. Note the plate voltages necessary to produce ionization for various values of grid voltage. In industry this tube has many applications; in radar its principal use is as a saw tooth generator to provide sweep voltages for cathode ray tubes.

Gas filled diodes are also used in radar equipment as protective devices to keep receivers from being over-loaded during transmission. The protective tube and circuit in radar equipment is called the T-R box.



Construction of Cathode Ray Tube

A Cathode Ray Tube

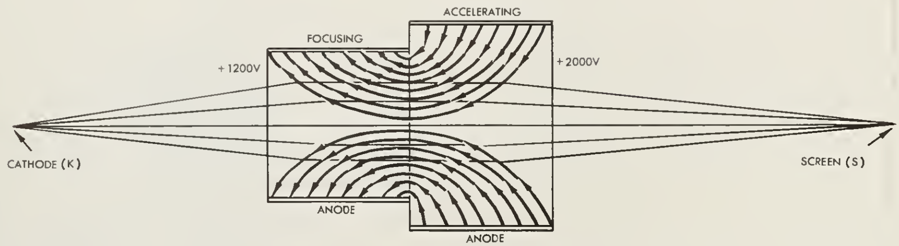
The *cathode ray tube* (CRT) is a tube especially designed to measure electrical phenomena which cannot be studied by other means. Not only is the CRT the heart of the oscillograph, but it is widely used in radar equipment to display visually information obtained from a transmitter or a receiver.

The CRT contains four essential parts; an evacuated glass envelope, an electron gun for producing a stream of electrons, a means of deflecting the electron stream, and a screen which transforms the electrical energy of the electron stream into light. The working parts of the CRT are enclosed in a glass envelope having a high vacuum in order to permit the beam to be sharp and to preserve the filament. Commercial tubes are of various sizes, having screens varying from one inch in diameter, for use in radio servicing, to giant tubes of 20 or more inches in diameter. CRT tubes for most airborne radar equipment are 3, 4, or 5 inches in diameter. Construction of the glass envelope is difficult, because of atmospheric pressure. On the face of a five-inch tube the pressure is almost 300 pounds. In most equipment there is a transparent shield between the operator and the face of the CRT. The screen serves to protect the operator and also to provide a place for cutting calibration marks.

The electron gun, which virtually shoots electrons at the screen in the CRT, consists of a filament, a cathode, a control grid, a focusing anode, and an accelerating anode. In the illustration, note these parts as well as their relative positions within the tube. Connections to the various electrodes are brought out through pins in the base of the tube rather than as shown. In many tubes the cathode is connected internally to the filament. The filaments are heated, usually by an alternating current furnished by a separate filament transformer. The cathode is a nickel cylinder, the end of which is coated with oxides of barium and strontium so that it emits electrons freely in the desired direction.

The accelerating anode is a cylinder. Inside it is a diaphragm which has a small opening in its center. Since the accelerating anode is highly positive, it will attract the electrons emitted by the cathode. The voltages applied to the accelerating anode range from 250 volts to as high as 10,000 volts. Three to five inch tubes commonly use 1500 to 2000 volts. This high voltage causes the electrons to travel at high velocity, which is proportional to the square root of the voltage difference and may be of the order of 10,000 miles per second for a 1000-volt potential difference. While the major portion of the electrons in the beam are attracted to and captured by the accelerating anode, many pass through the opening in the diaphragm and proceed on to the screen.

The grid is cylindrical in shape and partially surrounds the cathode. It serves to control the space current flowing from the cathode to the other electrodes, as in other vacuum tubes. The control grid is operated at a more negative potential than the cathode. If made negative enough it will cut off the flow of electrons. As stated earlier in the discussion of the oscillograph in the preceding chapter, the grid voltage control is called the *intensity control*. In radar equipment it has the same name.



Focusing Action

You can see that if some means is not provided for focusing, the electrons leaving the cathode will be emitted in a cone and, due to mutual repulsion, will be still further diffused and strike the screen in a scattered mass having a fuzzy or feathered appearance. The focusing and accelerating anodes make up an electron *lens*, which concentrates the electron beam and makes the image on the screen clear cut. To understand how this focusing action takes place, refer to the illustration labeled focusing action. The focusing anode has a potential of 1200 volts and the accelerating anode 2000 volts. Due to the 800 volts difference of voltage, there is a strong electrostatic field present in the region between these anodes. The strength of this field can be varied by changing the focusing anode voltage. The electrostatic field is represented by the curved lines. The electrons passing through this field will have a force acting upon them tending to cause them to follow the lines of force (in the opposite direction to the one indicated by the arrow points, since the direction of an electrostatic field is defined as the direction in which a unit *positive* charge would be acted upon). An electron entering the lens has two forces acting upon it, one force being due to the acceleration given it by the attraction of the accelerating anode and the other force due to the electrostatic field between the two anodes. The tendency is for the electron to be deflected and, instead of traveling in a straight line, to move in a direction tangent to the lines of force. The amount of deflection from its original path depends upon its velocity and the relative curvature of the lines of force. This curvature in turn depends upon the difference in potential between the anodes.

All electrons passing through the lens at a given angle tend to come together at a point called the focal point. This focus is comparable to the focus of a camera lens. You can see now why it is desirable to use small openings in diaphragms in the cylindrical elements. This insures that only those electrons which leave the cathode (K) at a small angle from the longitudinal axis of the tube will go through these openings, and consequently the stream of electrons reaching the lens will not be widely diffused and can be focused into a well defined beam. By changing the potential applied to the focusing anode, the position of the focal point can be changed, and correct adjustment will occur at the screen (S).

The purpose of the screen in the CRT is to make the electron beam visible. The screen, composed of a semi-transparent substance known as *phosphor*, is located on the inside face of the tube. When the electron stream strikes the screen, the latter gives off light, the color of which depends upon the composition of the phosphor. A commonly used screen coating for commercial oscillographs is Willemite zinc silicate, which gives off a green light. Other screen coatings are phosphorescent zinc sulphide, which produces yellow light, and calcium sulphide, which produces a white light shaded with green. An important consideration in the choice of phosphors is its persistency. Persistency determines the length of time that the screen will continue to glow after the beam of electrons has been removed or moved to a new location. If the screen has short, or low, persistency, this period of time is 0.1 second or less; if persistency is long, or high, the period is 1 second or longer. Between these two limits the persistency is medium. The

Willemite screen has low persistency. Those used in radar may be of any one of the types given depending upon the use of the equipment.

You can readily see that some means must be provided for removing the electrons from the screen; otherwise the negative charge on the screen would build up such a point that no more electrons could reach it. The usual method for removing the electrons is to place a conducting coating of aquadag all along the inside of the tube, except the face, and to connect the coating to the cathode. Secondary emission from the screen, which is sufficient to prevent serious effects from a negative voltage there, is then collected by the aquadag coating and returned to the cathode.

The remaining part of the CRT is a means of deflecting the electron beam. The two types of deflection are *electrostatic* and *electromagnetic*. The type of tube shown on page 4-15 is designed for electrostatic deflection. Two pairs of deflection plates set at right angles to each other are used and the electron beam passes between them. The plates are called the horizontal and vertical deflection plates according to the direction in which they deflect the beam, and *not* according to their positions in the tube. Suppose the upper vertical deflection plate has a positive DC potential applied to it, and that the lower vertical plate has a negative DC potential applied to it. The electrons in the beam will then be attracted by the upper plate and repelled by the lower plate, with the result that the electron beam is deflected upward. If the lower plate were made positive and the upper plate negative, the beam would be deflected downward. If an AC voltage were applied to the plates, the beam would be deflected up and down alternately at the frequency of the applied AC voltage. This would appear as a vertical line on the face of the scope unless the frequency were very low. The amount of deflection depends upon the voltages applied to the plates.

The deflection sensitivity of a CRT is the amount of deflection of the spot on the screen for a given voltage on the deflection plate. The deflection sensitivity is commonly about 50 volts per inch, which means that a positive voltage of 50 volts applied to the upper plate will cause the spot to be deflected upward one inch, or that an AC voltage which has a peak amplitude of 50 volts will, if applied to the vertical deflection plates, cause a deflection of two inches (one inch up and one down).

The second set of plates, called the horizontal deflection plates, function in the same way as the vertical plates. A positive voltage applied to the right horizontal plate will move the beam to the right; one applied to the left plate will move it to the left, and so on. You have already seen in the preceding chapter how a saw-tooth voltage applied to the horizontal plates gives a horizontal trace which can be used as a linear time base.

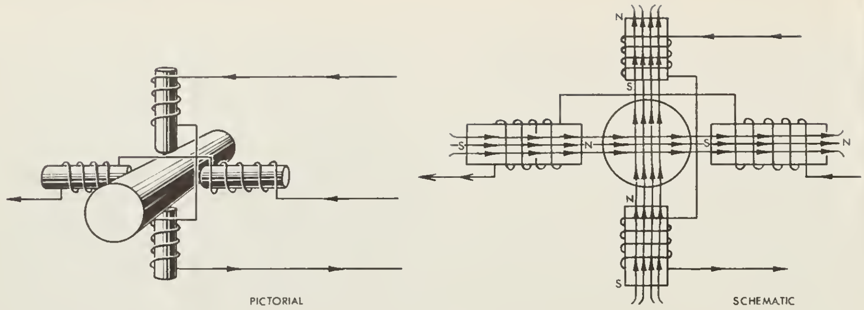
The deflection sensitivity of a tube is affected by several factors. First, the distance between the plates and the screen determines the swing of the beam for a given deflection angle. When the screen is farthest from the plates, the arc subtended by the deflection angle is greatest and the tube is most sensitive. When the vertical plates are farther from the screen than the horizontal plates, the sensitivity is greatest in the vertical direction.

A second factor affecting the deflection sensitivity is the voltage applied to the final anode. The speed of the electrons in the beam is determined by the anode voltage. The faster the electrons travel, the greater force it takes to deflect them and the shorter the time they require to pass between the plates. Hence higher voltage reduces the sensitivity.

Since both of the factors mentioned above are more or less fixed in any given tube, it is possible to state what the deflection intensity is in both directions at the time of manufacture.

Electromagnetic deflection is used in many cases where it is impossible to get a suitable electrostatic deflecting voltage. In electromagnetic deflection the deflecting force is due to the magnetic field set up within the tube by the set of coils arranged around the neck of the tube. One of the big advantages of this kind of deflection is the fact that the tube construction is simpler, as there are no deflection plates within the tube. Focusing can also be done electromagnetically, which still further simplifies the tube construction.

If the two sets of coils are placed at right angles (see illustration marked electromagnetic deflection), the two magnetic fields are perpendicular to each other and the strength of each field is proportional to the current flowing through the set of coils producing it. You are familiar with the fact that a current-carrying conductor located in

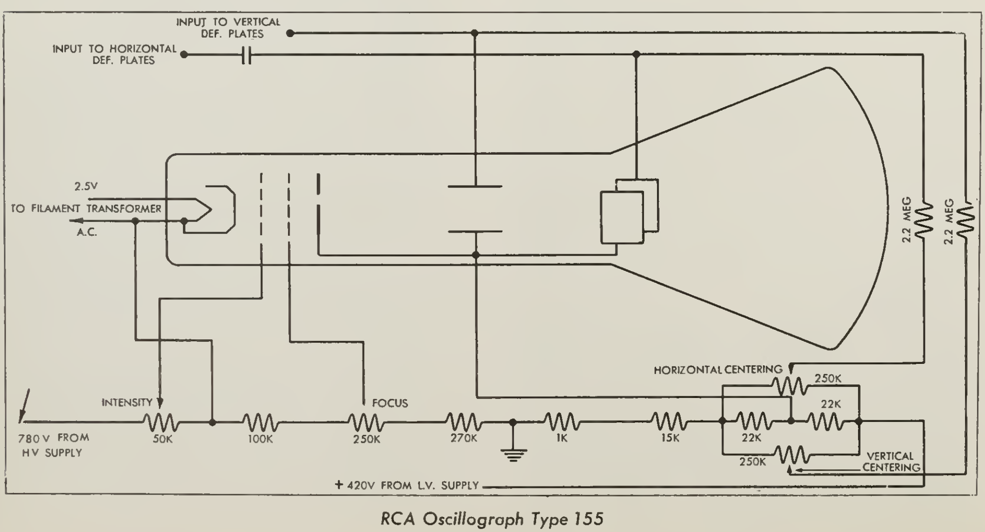


Electromagnetic Deflection

a magnetic field has a force exerted upon it which is at right angles to the field and also at right angles to the direction of electron flow. The stream of electrons is affected in the same way, for as the electron stream enters the field it is deflected in a direction perpendicular to the field in an amount proportional to the field strength. Deflection in a vertical direction is produced by the coils located on either side of the tube, and deflection in a horizontal direction by the coils above and below the neck of the tube. Deflection sensitivity is usually expressed in inches per ampere turns.

Electromagnetic focusing is obtained by winding a coil around the neck of the tube in such a way that the magnetic field is lengthwise with the tube. This means that electrons traveling down the axis of the tube are unaffected, but those traveling at a small angle have a force applied to them perpendicular to the magnetic field tending to bring the electrons back into a line of travel which focuses the beam at the screen. Electromagnetic deflection and focusing are widely used in radar equipment but only occasionally in oscillographs.

Circuits in the RCA Oscillograph type 155 are typical of circuits used in conjunction with a CRT. Although there are more complicated applications, all use essentially the same principle. Voltages for operation of the RCA Oscillograph



RCA Oscillograph Type 155

illustrated are obtained from two rectifier circuits, one called the low voltage rectifier and the other the high voltage rectifier. The output voltages of these rectifiers are approximately 420 volts positive and 780 volts negative, respectively. The cathode voltage is fixed at negative 720 volts. The grid voltage is variable between negative 720 volts and negative 780 volts, and acts as an *intensity control*. The position of the movable top of the 250-kilohm potentiometer labeled *focus* controls the focusing anode voltage. The focus control varies the anode voltage between 320 volts negative and 590 volts negative. The accelerating anode is kept at a fixed value of approximately 275 volts. The voltage between the accelerating anode and the focusing anode ranges from 595 volts to 865 volts, giving ample control over the focusing action.

The lower vertical deflection plate and the left hand horizontal deflection plate connect internally to the accelerating anode and are maintained therefore at a voltage of 275 volts positive. The movable tap on the 250 K potentiometer labeled *vertical centering* gives voltages between 120 volts and 415 volts. This range is from 155 volts below to 140 volts above the voltage of the lower plate. This makes it possible to center the beam vertically at any spot on the face of the CRT. The right hand horizontal deflection plate connects to a similar potentiometer marked *horizontal centering* which functions in a like manner for positioning the spot horizontally. The sweep voltage is applied to the right hand horizontal plate through the coupling condenser at the top of the diaphragm. This voltage may come from the saw-tooth generator in the oscillograph through the horizontal amplifier or from an external sweep circuit either through the horizontal amplifier or directly to the plate. The signal input to the upper vertical plate is applied through the coupling condensers and may come through the vertical amplifier or be applied directly to the plate. The two 2.2 meg. resistors serve as decoupling resistors to keep the signals from being attenuated in the voltage divider system.

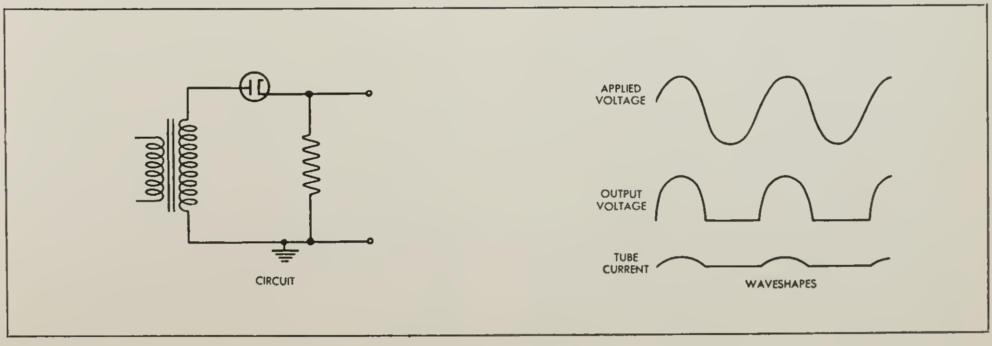
POWER SOURCES FOR OPERATING VACUUM TUBES

As previously pointed out, vacuum tubes require voltages of various values for their filaments, screens, and plate circuits. Except for the filaments which can be heated by either DC or AC, these voltages are DC. The most convenient source of DC voltage for vacuum tubes is a rectifier. A rectifier is a device which changes AC into DC by permitting current either to flow in only one direction or to flow more readily in one direction than in the other. Devices commonly used for changing AC into DC for use in radio and radar circuits are vacuum tube rectifiers, metallic oxide rectifiers, and crystal rectifiers, which are explained later.

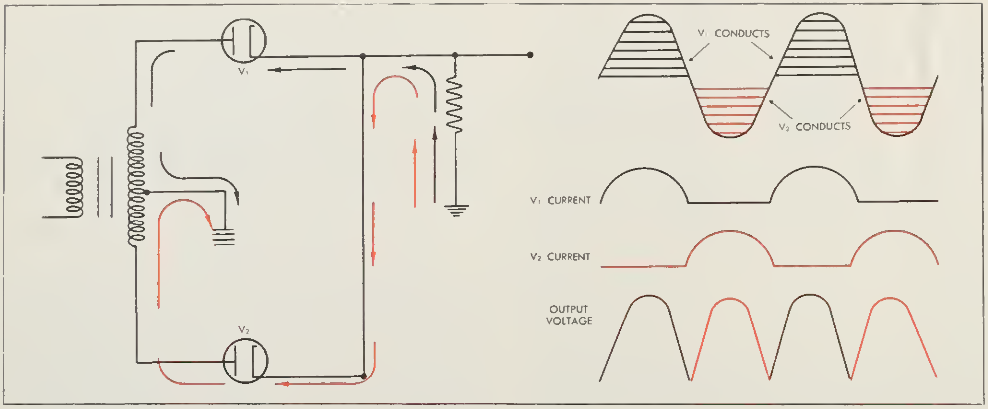
Vacuum Tube Rectifier

The rectifier tube is a diode type tube. Rectifier tubes with one plate and one cathode are called *half-wave rectifiers* since current flows only during one-half of the AC cycle. When two plates and one or more cathodes are used in the same tube, current can flow on both halves of the AC cycle. Tubes of this type are called *full-wave rectifiers.*

HALF-WAVE RECTIFIER. A half-wave rectifier conducts only during the half of the input cycle that makes the plate more positive than the cathode. The output voltage is a reproduction of



Half-Wave Rectifier



Full-Wave Rectifier

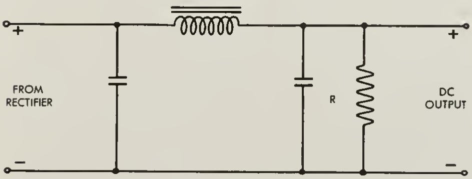
the positive alternation of the voltage across the secondary of the input transformer except for a low voltage drop across the tube which acts as a low resistance when the tube is conducting. The output voltage is a pulsating positive voltage that has an average value equal to one-half of the average value of the secondary voltage or 0.318 of the peak value of the secondary voltage. The efficiency of a half-wave rectifier is low. The voltage drops to a low value if a load requiring much current is connected to the tube. Hence, its use is limited to places where voltage is the primary concern and current drain is low. This makes it a suitable power supply circuit for anode (plate) voltages in the oscilloscope.

FULL-WAVE RECTIFIER. A full-wave rectifier conducts on both halves of the input voltage in such a way that current always flows through the load resistor in the same direction. Note in the illustration of the full-wave rectifier that the secondary is grounded at the center and the opposite ends are connected to the plates of two diodes. These two diodes conduct alternately since at any given instant one plate is positive and the other negative, while a half cycle later the voltages are reversed. The result is that, in the output, both halves of the input are reproduced. The output voltage has an average value of twice that of a half-wave rectifier with the same input. This makes the full-wave rectifier a more efficient power source than a half-wave rectifier. It is capable of standing higher current drain without excessive drop in voltage than a half-wave. The tube used must be capable of withstanding a voltage twice the peak of the secondary voltage applied to each tube. This is the voltage applied between plate and cathode, with cathode more positive than the plate, when the tube is not conducting. When V_1 is conducting its cathode is positive almost to the peak value of the secondary voltages and the cathode of V_2 is at the same potential. But the plate of V_2 is very negative at the same instant so that the cathode-to-plate potential is approximately twice the peak of the secondary voltage applied to each tube. This is called the inverse voltage and must be taken into consideration when designing a full-wave rectifier circuit.

Filter Circuits

The output voltages in the rectifier circuits just discussed are not suitable for supplying DC to most vacuum tubes because of their pulsating voltage output. The variations in the direct voltage are known as *ripple* voltage. Ripple voltage may be thought of as an alternating voltage superimposed upon direct voltage. Actually, the ripple voltage is composed of a fundamental frequency and a number of harmonic frequencies (harmonic frequencies are multiples of the fundamental frequency). The fundamental frequency is equal to that of the input voltage if the rectifier is half-wave or twice the input frequency if the rectifier is full-wave. The amplitude (magnitude) of the fundamental component is greater than that of a harmonic. For power purposes the ripple voltage must be reduced to a low percent of the DC output voltage. A device for eliminating ripple is called a filter.

FILTER. A filter usually consists of condensers connected in parallel and inductors connected in series with the load. A condenser opposes any change in voltage across its terminals (across the load) by storing up energy in its electrostatic field whenever the voltage tends to rise, and converting the stored energy back into voltage whenever the load voltage tends to decrease. An inductor opposes a change in current through it (through the load) by storing up energy in its electromagnetic field when the current tends to increase and by taking energy from the magnetic field to maintain the flow when the voltage tends to decrease. Another way of looking at the action of a filter circuit is to consider that the condenser forms a low impedance path to ground for the ripple voltage and offers a very high impedance to the DC, whereas the inductor offers a low impedance to DC and a very high impedance to the ripple voltage. Either way you look at it, it leads to the conclusion that there is a reduction of the amplitude of the ripple voltage without seriously affecting the DC voltage.



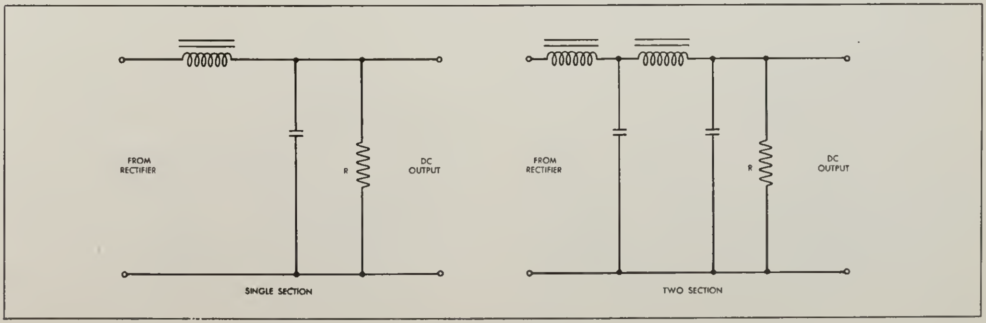
Capacitance Input Filter

CAPACITANCE INPUT FILTER. One type of filter is the capacitance input Pi type filter. A capacitance input filter is characterized by high output voltage at low current drain, the voltage being almost equal to the peak value of the secondary voltage applied to each tube. However, the voltage falls rather rapidly as the current drain increases, reaching as low as ninety percent of the RMS value when the full rated current is drawn. This fall of voltage under load conditions is referred to as poor regulation.

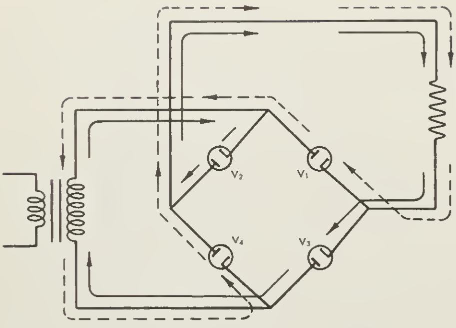
An important feature of this filter is the bleeder resistor, which is used chiefly to discharge the condensers when the equipment is turned off. A bleeder resistor should draw 10 percent or less of the rated current output of the power supply.

INDUCTANCE INPUT FILTER. Another type of filter is the inductance input filter. When the tube draws no current, an inductance input filter keeps the voltage at practically the same value as a capacitance input filter. When the tube draws a small amount of current, the voltage drops rapidly to a value which remains fairly constant over a wide range of drains. Regulation is good after the first drop, so if the bleeder resistor is small enough to insure drawing sufficient current to operate at the beginning of the linear portion of the voltage-current curve, this type filter provides very satisfactory results under changing load conditions.

There are many other types of filters but the action of each is much the same. Sometimes the inductor is replaced by a less expensive resistor, but a resistor has the disadvantage of offering as much impedance to DC as to the ripple voltage. In some applications, resonant circuits are used as filters. A series resonant circuit in parallel with the load and a parallel resonant circuit in series with the load form a low impedance path to ground for the ripple frequency and a high impedance path to block its passage through the load.



Inductance Input Filter



Bridge Rectifier

Bridge Rectifier

The bridge-type rectifier shown directly above is a full-wave rectifier capable of withstanding a higher inverse voltage than a two-tube full-wave rectifier. This results from the fact that the inverse voltage appears across two tubes in series and is therefore only half as great per tube. By tracing current flow you can see that V_1 and V_4 are in series, and V_2 and V_3 are in series. At the same time the series arrangement in this type circuit offers disadvantages. Instead of the drop across one tube there are two drops to be deducted from the output voltage. Another disadvantage is that the circuit requires three filament transformer windings since only V_1 and V_3 have the same voltages on the filaments. The main point in favor of the use of this circuit is the fact that the secondary of the transformer does not have to be center tapped and the output voltage is twice as great as it would be if the same transformer were used in a conventional full-wave rectifier.

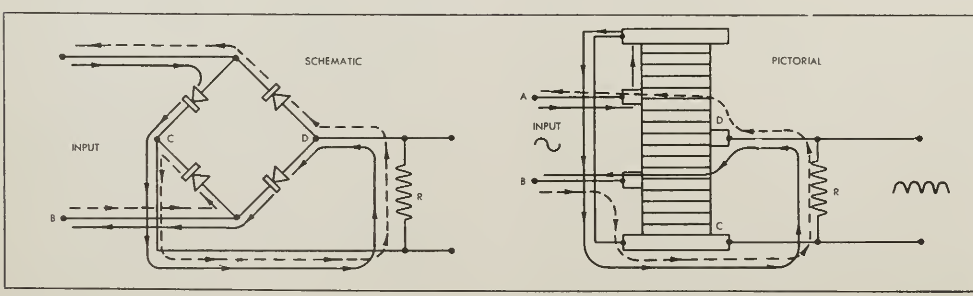
Metallic Oxide Rectifiers

The principal metallic oxide rectifiers are the copper oxide and the selenium rectifiers. Both are rectifiers by virtue of the fact that they permit current to flow more readily in one direction than in the other. A copper oxide rectifier consists of a thin film of copper oxide on a copper plate. A selenium rectifier consists of a prepared film of selenium on a metallic substance such as iron. A selenium rectifier has a somewhat lower resistance than the copper oxide rectifier and is therefore more efficient since it can pass higher current. In each case, election flow is principally from the metal to the film. This type rectifier cannot stand high voltages since the film is very thin. When several such sections are built up in series to form a stack, the breakdown voltage is increased. To reduce the danger of breakdown due to inverse voltage, these rectifiers are usually arranged in a bridge circuit as shown below.

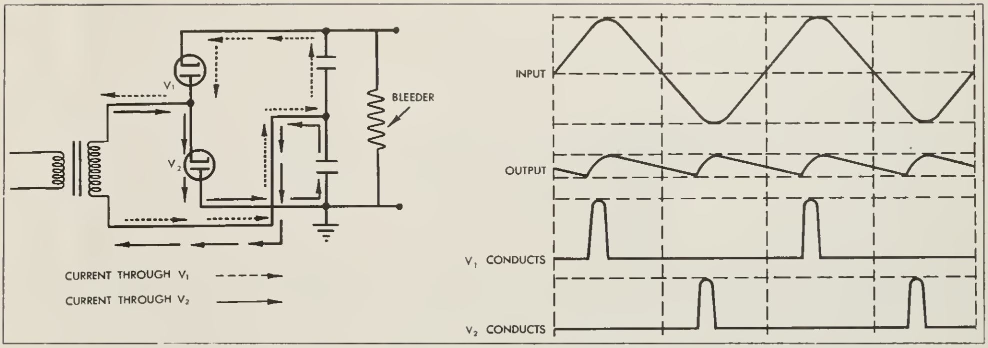
Voltage Doubler

A voltage doubler is a circuit arrangement capable of delivering a DC voltage that is equal to twice the peak voltage of the AC line from which the circuit receives its energy. It consists of two rectifier tubes and a circuit which connects the output of the two tubes in series. The voltage doubler is a good source of power in places where high-voltage with low-current drain is required.

An elementary circuit using two half-wave rectifiers is shown in the illustration on the next page. To learn its operation, first consider the upper half. When plate V_1 is positive, current flows through the tube and charges the upper condenser. Next, consider the lower half. When plate V_2 is positive, current flows through the tube and charges the lower condenser. Since the tubes are connected opposite to



Selenium Bridge Rectifier



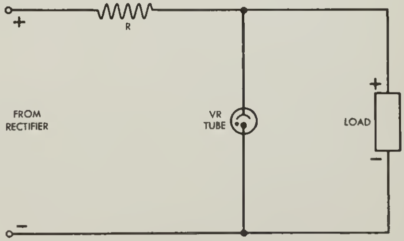
Voltage Doubler

each other across the applied AC, the two condensers charge in the same direction. If no current is taken from them by the load, each will charge to the peak value of the voltage across the secondary. Since the condensers are effectively in series, the voltage across both condensers (the output voltage of the doubler) is approximately twice the peak value or 2.82 times the RMS value of the secondary voltage. To discharge the condensers when the equipment is off, there is a bleeder resistor connected across the output. It is made large to keep the condensers from discharging appreciably between charging alternations. Due to the discharge of the condensers through the bleeder, there is a small amount of ripple. Each tube conducts each time its plate is more positive than the condenser voltage, and therefore conduction is both small and short in duration. Since current drawn from a voltage doubler power source depends upon the discharge of the condensers, obviously the amount of current that can be drawn is limited. During discharge the voltage drops rather rapidly and the ripple becomes appreciable.

Voltage Regulator Circuits

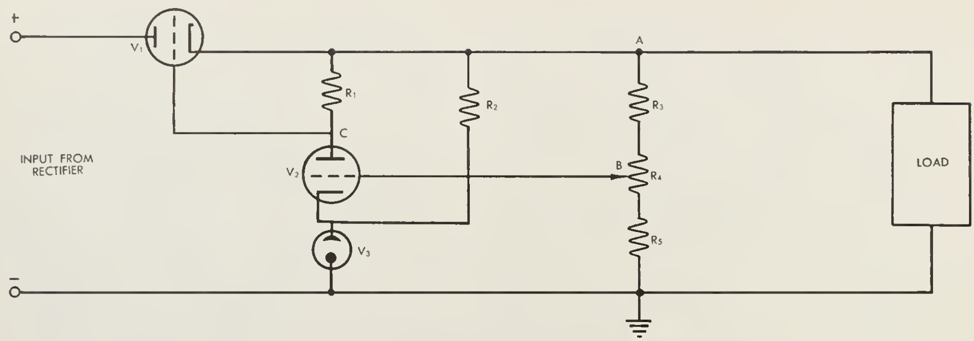
As operation of radar circuits requires very stable voltages for optimum results, it is well to consider some of the circuits designed to maintain the output voltage of a power source constant at a predetermined level in spite of varying input voltage or varying load conditions. There are other electronic circuits and mechanical devices in use for this purpose, but the ones discussed here exemplify the types you will encounter.

Simple regulator circuit. The circuit at the right shows a simple voltage regulator circuit. Note that it uses a VR tube, which was discussed earlier in this chapter. Remember, too, that the VR tube is characterized by the fact that over a wide range of current flow there is no change in the voltage drop across it. In operation, a VR 150-30 tube draws current up to 30 ma and maintains the output voltage at 150 volts. If the input to the regulator were to rise, more current would flow through the VR tube and there would be a greater voltage drop across the resistor R. This voltage drop would subtract from the rising voltage and tend to maintain the voltage across the load and across the VR tube constant. On the other hand, if the input voltage falls, the VR tube will conduct less current, the drop across R will be less, and, as before, the voltage across the load



Simple Voltage Regulator

will remain at the desired level. When the load draws more current, the voltage tends to drop because of voltage drop across R. The VR tube compensates for this by conducting less, decreasing the total current, and, as a result, maintaining the load voltage constant. Regulation with this circuit is satisfactory as long as the current conducted remains between 5 and 30 milliamperes.



Electronic Voltage Regulator

Electronic voltage regulator circuit. The electronic voltage regulator shown is capable of very close regulation at a level which may be set by varying the potentiometer setting. This circuit contains a vacuum tube (V_1) in series with the load. The voltage across the load is regulated by controlling the conduction of V_1—that is, making V_1 act as a variable resistor that automatically adjusts itself to the correct value. V_3 is a VR tube, the purpose of which is to maintain the cathode of V_2 at a fixed positive potential. The voltage divider system, composed of R_3, R_4, and R_5, is arranged so that the variable arm of R_4 can be adjusted to a positive voltage sufficiently low to bias V_2 for operation in the linear portion of the Eg-Ip curve. R_1 is the load resistor of V_2, which is connected in series with the VR tube, V_3. The purpose of this resistor is to absorb any changes of voltage so that the cathode of V_2 will remain at a fixed potential.

If the output voltage (point A) tends to rise due to an increase in input voltage or to a decrease in current drawn by the load, the voltage at point B will rise by any amount dependent upon the resistances from A to ground and from B to ground. The voltage at B is the voltage applied to the grid of V_1 and determines how much V_1 will conduct. When V_1 conducts more, there is an increase in current through the load resistor R_1 and consequently a fall in voltage at point C. This voltage is applied to the grid of V_1, and causes it to conduct less, counteracting the tendency of the rise in output voltage which set off this action. If the output voltage should tend to fall, the voltage at B drops, causing a rise in voltage at C, which causes V_1 to conduct more, resulting in a greater output voltage. It has already been pointed out that V_2 must be biased to operate in the linear portion of the characteristic curve. This is also true of V_1, the bias of which is the IR drop across R_1.

The setting of B determines the bias of V_2 and the consequent current flow. This current flow which passes through R_1 determines the bias on V_1. The current flow through V_1 determines the load voltage. Hence, the output voltage is adjustable, within limits, by setting the movable tap, point B, of potentiometer R_4. This circuit, with some refinements, is used in the power sources of several radar equipments. Due to the great amount of current flow required in some sets, you may find two or more tubes in parallel to serve the purpose of V_1. You can see that the rectifier and filter circuit must be designed for greater current capacity than the load is expected to draw since the regulator draws current. To illustrate, there is current flow through R_1, V_2, and V_3; a second path consists of R_2 and V_3; and a third path includes R_3, R_4, and R_5. This extra current drain is the price you have to pay for the regulation of the output voltage within very close limits.

CHAPTER 5

Principles of Radar

Radar circuits are no more complicated than other electrical circuits. They are nothing more than new combinations of circuits which have been used for many years in other instruments. Up to this point, the manual has described these standard circuits. The rest of the manual, however, deals with these same circuits in new combinations—combinations called *radar* circuits.

The purpose of this chapter is to prepare you for the discussion of the radar circuits in the following chapters. It discusses the basic principles of radar, the components of radar sets, and in general gives you a basis for understanding individual radar circuits in terms of a complete radar set.

DEFINITION

The word *Radar* is a contraction of the expression *RA*dio *D*etection *A*nd *R*anging. The italicized capitalized letters, as you can see, form the word *Radar*. Radar is thus an application of radio principles to detect objects that cannot be observed visually and to determine the direction, range, and elevation.

PRINCIPLES OF OPERATION

You know from experience that when you make a loud noise, such as a shout or a gunshot, you often hear an echo of that noise. Sometimes you hear several echoes from the same noise, some of them apparently coming from different directions. Without looking, you know that there is a building, hill, grove of trees, or some other object nearby in the direction of the first echo, and another similar object still further away in the direction of the second echo. Obviously, the loud noise consisted of sound waves, which started out at your location and, on striking the object, were reflected. As these reflected sound waves move by you, you hear them. According to the science of physics, sound travels at a constant rate. Therefore, it took a certain amount of time for the sound waves to go to the object and return to you as an echo. And still more time would have been required had the object been farther away. Furthermore, you are able to tell the direction of the object which caused the echo because of your ability to hear. With this faculty, your brain tells you that the object creating the echo is to the right, left, or in front of you.

In the same way, radar depends on creating and picking up an echo—a radio echo. Radar sets send out short, strong bursts of radio energy. Many of these bursts of energy strike objects in the vicinity and are reflected back to the site of the radar set. When the time required for the energy to go to the object and to return is carefully measured and translated into distance in miles or feet, it is possible to determine the distance of the object causing the echo. In early radar sets, two antennas were used to compare the strength of the reflected energy. From this comparison it was possible to determine the direction from which the echo came. For the same purpose, modern sets employ various types of directional antennas. Thus Radar, like the method you use in determining the direction and distance of an object causing a sound echo, determines the distance and direction of an object creating a radio echo.

In radar terminology, the reflected energy is called the *radar echo* or simply the echo. The object, principally due to military influence, is called the *target.*

Modern radar technique is highly developed. Sets are unbelievably accurate; in many cases they are almost completely automatic. Yet, regardless of their stage of development, all depend upon this simple principle of creating and detecting a radio echo.

RADAR SYSTEMS

Essentially, a radar system consists of a transmitter which sends out radio signals, a receiver which is located at the same site, and an indicator which gives a visual indication of echoes returned by a target.

When a radio signal which has a constant frequency is emitted by a radio transmitter, radio waves, in a manner similar to light and sound waves, travel out in all directions and are reflected by any object that they strike. On striking the object, components of the wave are reflected and likewise travel in all directions. Some of the reflected waves return to the site of the transmitter originating them, where they are picked up by the receiver, providing it is tuned to the correct frequency.

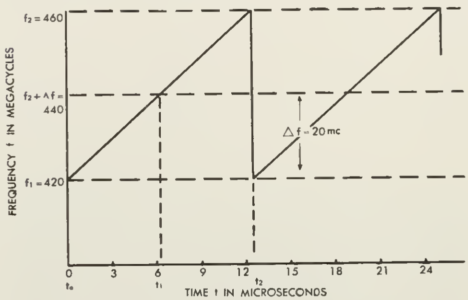
This is a simple explanation. Naturally, there are some problems which must be solved. First, there is the problem of having a powerful transmitter near the receiver and operating on the same frequency. In order for the receiver to detect the reflected signal, and thus indicate the presence of the target, the transmitter signal must be prevented from affecting the receiver.

The second problem is measuring distance. If a continuous constant frequency signal were transmitted, it would be impossible for the receiver to distinguish between various echoes at different distances, for they would all be alike. Thus, some means must be provided to eliminate the problems involved in using constant frequency signals.

There are several systems which satisfactorily solve these problems. They are the frequency modulation system, the frequency shift system, and the pulse modulation system. The last system is the most important and is the only one discussed in detail in this manual.

Frequency Modulation System

Because each cycle of a frequency modulated wave differs by a small increase in frequency from the others of that wave, a frequency modulation system makes it possible to identify each cycle of a wave transmitted and to recognize it from all others when it returns to the receiver. By designing a transmitter which produces a signal which regularly changes over a known range of frequencies, it is possible to identify any particular reflected signal cycle.



Frequency Modulation Chart

An example of a frequency modulated signal is plotted against time in the frequency modulation chart. As shown, the 420 mc frequency increases linearly to 460 mc and then quickly drops to 420 mc again. When the frequency drops to 420 mc, the frequency cycle starts over again.

Since the frequency regularly changes 40 mc with respect to time, you can use its value at any time during its cycle as the basis for computing the time elapsed after the start of the frequency cycle. For example, at time t_0 the transmitter sends a 420 mc signal toward an object. It strikes the object and returns to the receiver at time t_1 when the transmitter is sending out a new frequency of 440 mc. Thus, since f changes from 420 mc to 460 mc in 12.4 microseconds, the time required for the 420 mc signal to change from 420 to 440 mc is $\frac{20}{40}$ of 12.4 or 6.2 microseconds. Radar waves strike and return from an object one mile away in 12.4 microseconds. Therefore, the distance of the object in the example is $\frac{6.2}{12.4}$ or $\frac{1}{2}$ mile.

In the frequency modulation system, two separate signals are fed to the receiver at the same time. For example, at t_1, the 440 mc transmitted signal and the 420 mc reflected signal reach the receiver simultaneously. When these two signals are mixed in the receiver, a beat note results. The frequency of the beat note varies directly with the distance to the object, increasing as the distance increases. A device that measures frequency can be calibrated to indicate range (distance to object).

Frequency-Shift System

Another system for making the received echo change continuously in frequency in order to locate remote objects is the frequency shift system. This system is based on the *Doppler effect*, a familiar example of which is the change in pitch of an automobile horn as the car approaches. Radio waves act in the same way. If the source of radio energy (in the case of radar, an object from which radio waves are reflected) is moving rapidly, the frequency of the radiated energy (the echo) changes. In radar, employing the Doppler effect to locate objects is based on the transmission of continuous or unmodulated radio waves toward the object. At the receiver the frequency of the waves reflected by the object is changed provided the object moves toward or away from the receiving point. Cross-wise movement (which would describe a circle around the receiving point) would not change the frequency. The amount of change is proportional to the speed at which the object is moving toward or away from the receiving point. Since the receiver is located near by the transmitter, it receives a signal from both the transmitter and the remote object. When the object is moving toward or away from it, the receiver receives a frequency from the remote object that is different from the transmitter frequency. The detector in the receiver responds to the difference in frequency. If the object is not moving or if it is moving crosswise to a radius drawn through it, the returning frequency is the same as the transmitter frequency, and the detector response is zero. It is therefore impossible by the frequency shift method to detect objects which are not approaching or moving away from the receiver.

The frequency shift principle is sometimes used in conjunction with a pulsed radar set to eliminate echoes from stationary objects. For example, at air bases in mountainous areas, the mountains cause radar echoes so strong that the weaker echoes from aircraft flying in the vicinity are obscured on the radar screen. By applying the frequency shift principle, the moving objects can be differentiated from stationary objects. The echoes from the stationary objects can be eliminated, and the radar operator is able to see only the flying aircraft. In practice, the radar set itself uses the pulse system and the frequency shift detector device is a supplementary component attached to the radar set. This device is called a *moving target indicator.*

Pulse System

The pulse system of detection is employed in almost all radar sets. In this system, the transmitter is turned on for short periods and off for long periods. During the period when the transmitter is turned on, it transmits a short burst of energy called a *pulse.* When a pulse strikes any object, part of the reflected energy is returned to the receiver, where it is displayed on the screen of a cathode ray tube. (The cathode ray tube is a device capable of measuring periods of time as short as one-millionth of a second.) Since the transmitter is turned off after each pulse, it does not interfere with the receiver (as would be the case if a constant signal were used).

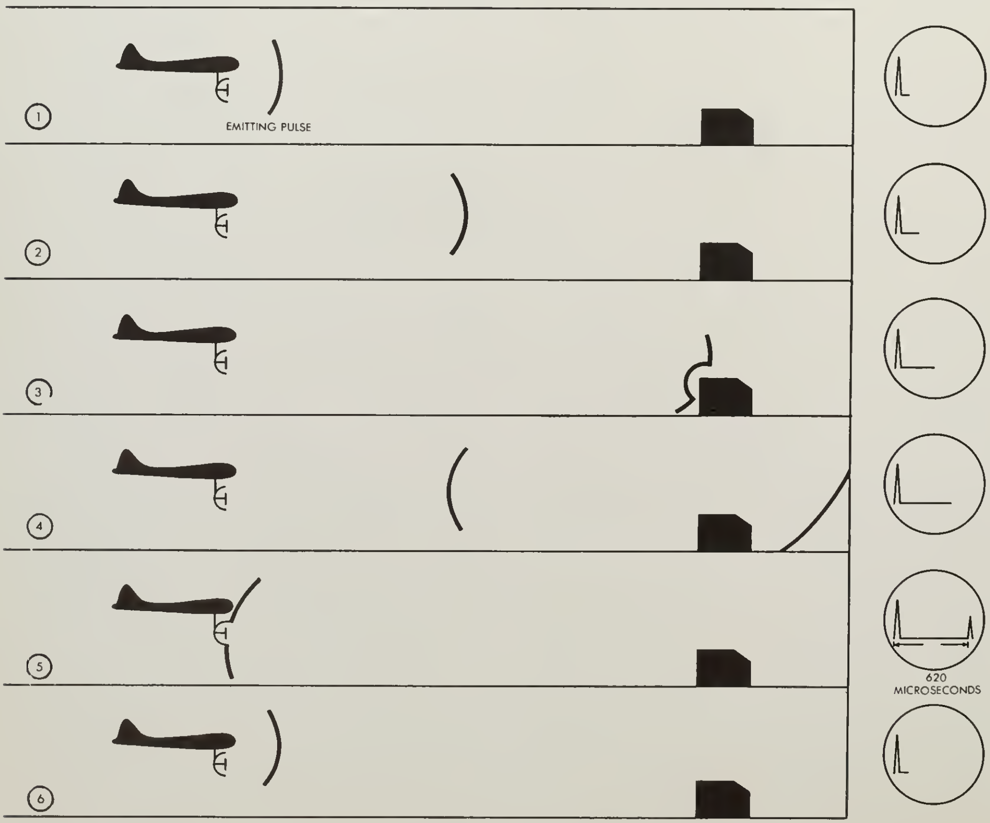
Complete location of an object in space by radar pulses depends upon two factors—the range or the distance of the target, and the direction, including both the azimuth and elevation directions of the target.

DETERMINING RANGE. The successful use of a pulse-modulated radar set depends primarily on its ability to measure distance in terms of time. When radio energy is radiated into space, it continues to travel with a constant velocity. Its velocity is that of light, or about 186,000 statute miles per second or 162,000 nautical miles per second. In more useful terms, radio waves travel a nautical mile in 6.2 microseconds. They travel a *radar mile*, that is, a go and return mile, in 12.4 microseconds.

This constant velocity is used in radar to determine the distance or range of a target by measuring the time required for a pulse to travel to a target and return. Suppose, for example, a pulse of radio energy is transmitted toward a target some distance away and the radar echo returns after 620 microseconds. Energy moves a mile and back or a radar mile in 12.4 microseconds. Therefore, the distance of the target is $\frac{620}{12.4}$ or 50 miles.

In order to use the time-range relationship, a time-measuring device must be used. The cathode ray tube is useful for this purpose since it responds to changes as rapid as one microsecond apart. A time base is provided by using a linear sweep to produce a known rate of motion of an electron beam across the screen of the cathode ray tube.

The formation of the time base is shown in the illustration below. In ① a radar pulse is leaving the airplane. At the time the pulse is radiated, the spot on the screen of the cathode ray tube is deflected vertically for a brief instant, then it continues across the screen to the right. In ② and ③ the pulse is traveling toward the target and the spot is moving across the screen. When the pulse strikes the target, there is no deflection, since energy is at the target itself. In ④ the reflected pulse is returning. In ⑤ the reflected energy has returned to the receiver and there is a second vertical deflection of the spot on the right side of the cathode ray screen. The distance between the two upward deflections serve as the basis for determining the range of the target from the radar antenna. Assume, for example, that the set is designed so the spot moves across the cathode ray tube in 700 microseconds. The spot was almost to the end before the echo arrived, its position indicating that it took 620 microseconds to reach that point. Since radar waves travel one radar mile in 12.4 microseconds, the range of the target in the illustration is $\frac{620}{12.4}$ or 50 miles away. The last picture shows another pulse being emitted and the start of the formation of a new time base.



Formation of a Time Base

DETERMINING DIRECTION. Two dimensions must be considered in determining the direction of a target. One is azimuth, which is the relative horizontal direction of the target with respect to some direction reference expressed in degrees. For example, this direction may be expressed with reference to true north if the radar set is a ground installation or with reference to the heading of the airplane if the set is airborne. The other dimension is elevation, which, like azimuth, can also be expressed in degrees. Elevation expresses the angular degrees that the target is above or below the radar set.

The determination of azimuth and elevation depend upon the directional characteristics of antennas and antenna arrays. Antennas for performing these functions are discussed later in this manual.

BASIC ELEMENTS OF PULSE SYSTEMS OF RADAR

The basic elements in a typical pulse radar system are the timer, modulator, antenna, receiver, indicator, and transmitter.

Timer

The timer, or synchronizer, is the heart of all pulse radar systems. Its function is to insure that all circuits connected with the radar system operate in a definite time relationship with each other, and that the interval between pulses is of the proper length. The timer may be a separate unit by itself, or it may be included in the transmitter.

The Modulator

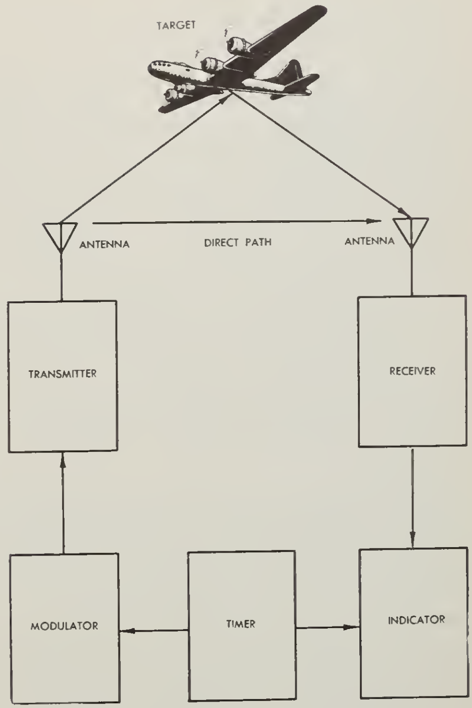
The modulator is usually a source of power for the transmitter. It is controlled by the pulse from the timer. It sometimes is called the keyer.

The Transmitter

The transmitter provides RF energy at an extremely high power for a very short time. The frequency must be extremely high to get many cycles into the short pulse.

The Antenna

The antenna is very directional in nature because it must obtain the angles of elevation and bearing of the target. To obtain this directivity at centimeter wavelengths, ordinary dipole antennas are used in conjunction with parabolic reflectors. Usually, in order to save space and weight, the same antenna is used for both transmitting and receiving. When this system is used, some kind of switching device is required for connecting it to the transmitter when a pulse is being radiated, and to the receiver during the interval between pulses. Since the antenna only "sees" in one direction, it is usually rotated or moved about to cover the area around the radar set. This is called *searching*. The presence of targets in the area is established by this searching.



Elements in Pulse Radar System

Receiver

The receiver in radar equipment is primarily a super-heterodyne receiver. It is usually quite sensitive. When pulsed operation is employed, it must be capable of accepting signals in a bandwidth of one to ten megacycles.

Indicators

The indicators present visually all the necessary information to locate the target on the indicator screen. The method of presenting the data depends on the purpose of the radar set. Since the spot "scans" the indicator screen to present the data, the method of presentation is often referred to as the type of *scan*.

The following are several of the common types of scan used in radar receivers:

Type A scan, in which the spot maintains a constant intensity, starting the instant a pulse of energy is radiated by the transmitter and travelling at a constant speed across the face of the indicator. When the spot reaches the right side of the indicator, it is blanked out, jumps quickly back to the left side, and repeats the process. Receiver echoes cause a vertical deflection of the spot, roughly proportional to the strength of the received signal. The horizontal distance between transmitted pulse and echo represents distance to the target. Although the principle function of this scan is to determine the distance of the object, it is possible to obtain a rough approximation of direction by rotating the antenna until the maximum echo is received.

Type B scan, which presents the bearing and range of reflecting objects as abscissa and ordinate, respectively. In this system, a highly directional antenna system is rotated about a vertical axis. This causes the radiated beam to cover a horizontal plane and gives the spot on the screen a horizontal motion which corresponds to at least a part of the angle of rotation of the antenna system. In the absence of other deflection, the scanning spot describes a bright horizontal line across the lower portion of the indicator. This line represents the transmitted pulses and is the so-called "base-line" of the pattern. A uniform vertical motion from bottom to top of the screen is also given to the scanning spot, each vertical line being synchronized with a transmitter pulse for indication of range. The repetitive vertical sweep is very much more rapid than the horizontal sweep, and the spot is maintained at low intensity. When an echo is received, the signal is impressed on the control grid of the indicator, causing a bright spot to appear on the screen. The position of this spot to the right or left of the center line of the screen indicates the azimuth of the target (its angle to the right or left of the radar set). The height of the spot above the base line indicates range or distance of a target.

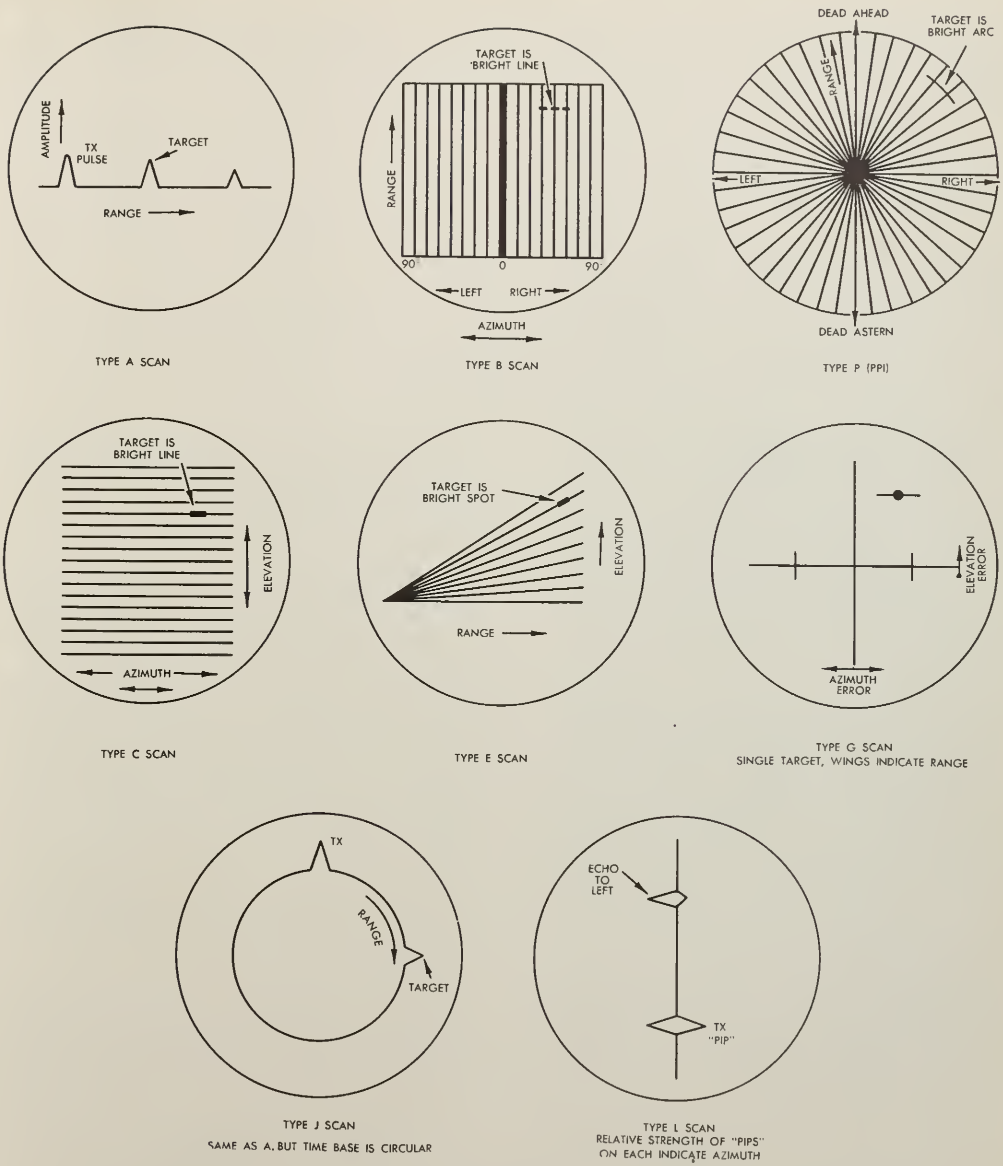
The PPI (PLAN POSITION INDICATOR), which is another type of scan for presenting range and bearing (direction) information. You can think of the PPI scan as a modified type of B-scan, in which rectangular coordinates are replaced by Polar coordinates. The antenna generally is rotated uniformly about the vertical axis so that searching is accomplished in a horizontal plane. The beam is usually narrow in azimuth and broad in elevation, and large numbers of pulses are transmitted for each rotation of the antenna. As each pulse is transmitted, the unintensified spot starts from the center of the indicator and moves toward the edge along a radial line. Upon reaching the edge of the indicator, the spot quickly jumps back to the center and begins another trace as soon as the next pulse is transmitted. As the antenna rotates, the path of the spot rotates around the center of the indicator screen so that the angle of the radial line, on which the spot appears, indicates the azimuth of the antenna beam, and distance out from the center of the indicator indicates the range. When an echo is received, the intensity of the spot is increased considerably, and a bright spot remains at that point on the screen, even after the scanning spot has passed it. Thus it is possible with this scan to produce a map of the territory surrounding the observing station on the indicator tube. This type of scan is useful when the radar set is used as an aid to navigation.

Type C scan, in which the echo appears as a bright spot with the azimuth angle as the horizontal coordinate and the elevation angle as the vertical coordinate. This type of scan has been used by night fighters to aid in following an enemy plane.

Type E scan, a modification of the B scan on which an echo appears as a bright spot with the range as the horizontal coordinate and the elevation as the vertical coordinate. This type of scan is used in directing planes in making a blind landing. The approaching planes must follow a certain angular line to reach the touchdown point at the left of the screen.

Type G scan. On this scan only the echo appears. It is a bright spot on which wings grow as the distance to the target is diminished. Azimuth appears as the horizontal coordinate and elevation as the vertical. This type of scan is used for pointing guns by hand. Centering the spot indicates correct aim. When the wings grow to a certain length, the range is correct for a volley from the guns.

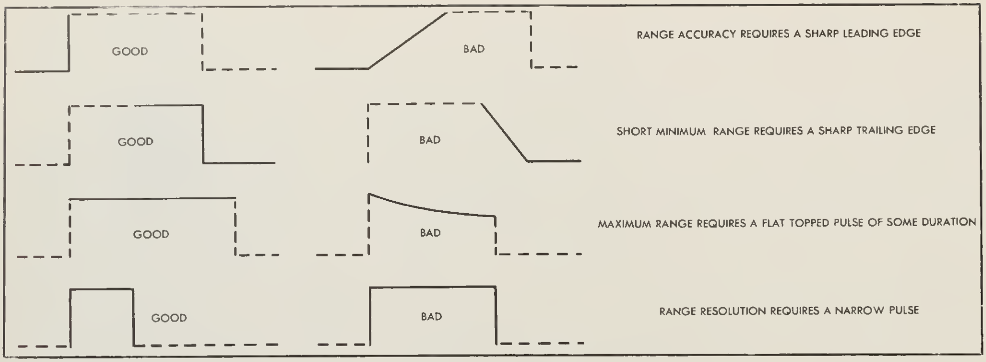
Type J scan, a modification of the A scan in which the spot rotates in a circle near the edge of the CRT face. An echo appears as a deflec-



Types of Scans

tion from the circle. As the distance changes, the deflection moves around the circle like the pointer of the aneroid altimeter. Its chief use is as a radar altimeter.

Type L scan, another variation of the A scan which indicates azimuth by comparing signals from the left and right antennas. It is used in homing operations. The ratio of the signal amplitudes from the two antennas indicates the error in homing.



Pulse Characteristics

PULSE SHAPE, LENGTH, AND REPETITION FREQUENCY

The effectiveness of a radar set on a particular mission is increased by designing it so the transmitted pulse has the proper characteristics. The most important characteristics of a radar pulse are its shape, length, and repetition frequency.

Pulse Shape

The shape of the pulse determines range accuracy, minimum and maximum range, and resolution or definition of target.

RANGE ACCURACY. Range accuracy depends to a large degree on the leading edge of the pulse. A pulse with a sharp leading edge provides a definite time from which to measure the start of the pulse and a better picture on the indicator scope than one produced by a pulse with a sloping leading edge.

MINIMUM RANGE. Detection of objects at the minimum range of the set requires a pulse which has a sharp trailing edge. Sloping trailing edges increase the width of a pulse and consequently cause inaccuracy in range since the edge of the pulse may itself cover several range miles.

MAXIMUM RANGE. To achieve maximum range, the pulse must be flattopped and be sufficiently long in duration. A flat top in a pulse means that the energy in the pulse is constant for the duration of the pulse and all cycles reflected will have the greatest possible strength. Long pulses mean more cycles to strike the target and be reflected.

RANGE RESOLUTION. Range resolution, the ability to indicate very small differences in range where the targets are in the same direction, requires narrow pulses.

Pulse Length

Pulse length determines minimum range of equipment. The length of the pulse length must be such that the transmitter is shut off by the time the received echo returns from the target. Therefore, the shorter the pulse, the closer to the set echoes can be received. On the other hand, long pulses provide greater maximum range since a long pulse provides more energy for the target to reflect. Long pulses also make possible a reduction in the bandwidth of the receiver. This results in less noise and an extension of range. Pulse durations used in radar equipment vary from as little as 0.25 microseconds to as much as 50 microseconds with the most common systems employing pulse durations ranging from one to ten microseconds.

Pulse Repetition Frequency

The pulse repetition frequency rate (PRF) largely determines the maximum range, and to some degree, the accuracy of a radar set. As you can readily see, the actual time elapsing between the beginning of one pulse and the beginning of the next, called the pulse repetition period, is the reciprocal of the pulse repetition frequency. Thus, for example, if the pulse repetition frequency is 400 cycles per second, the pulse repetition period is 1/400 seconds, or 2500 microseconds. When the frequency is too high—that is, when the period between pulses is too short—the echo from the farthest target may return to the receiver after the transmitter has emitted another pulse, making it impossible to tell whether the observed pulse is the echo of the pulse just transmitted or the echo of the preceding pulse. Such a condition is referred to as range ambiguity.

Although the pulse repetition rate must be kept low enough to realize the required maximum range, it must also be kept high enough to avoid some of the pitfalls a single pulse might encounter. If a single pulse were sent out by a transmitter, atmospheric conditions might attenuate it, the target might not reflect it properly, or moving parts—such as a propeller—might throw it out of phase or change its shape. When it did return, you might blink your eyes just when it would be displayed on the screen and you would not see it at all. Thus, information derived from a single pulse would be highly unreliable. But by sending many pulses, one after another, many good ones will return. The equipment will integrate or sum up the good points of all the pulses and present you with a good clear reliable picture. Equipment is therefore designed in such a way that many pulses, ten or more, are received from a single object. In this way, effects of fading are somewhat reduced. At short ranges, such as those used in gun-laying sets, the repetition frequency is increased in order to insure accurate short-range measurements. Many echoes are received from one target, and the integrating effect is increased. This makes possible more accurate range and bearing determination, highly necessary in gun-laying equipment.

Another factor affecting the pulse repetition frequency is the rate of angular motion of the antenna in searching extensive regions. If the antenna is moved through too large an angle between pulses, not only will the number of pulses-per-target be too low, but there may even be areas in which targets may exist without their being detected. In this connection, still another factor to consider is the sharpness of the antenna beam. A sharp beam must be pulsed more often than a wide beam to avoid skipping over targets.

Tactical employment of a radar set determines, to a large degree, the pulse repetition frequency to be used. Long-range search sets, covering distances to 150 miles, require a pulse rate slow enough to allow echoes from targets at the maximum range to return to the receiver before the transmitter is again pulsed. Higher pulse rates are used in aircraft interception sets where the maximum range is around 10 miles.

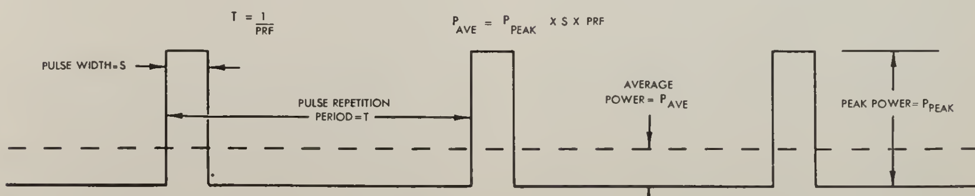
Power Requirements in Pulse System

The maximum range which a radar set can attain obviously depends largely on the power output of the set. Enough power must be radiated so that at the maximum range, the received echo signal will have a power level at least equal to the electronic noise level at the receiver.

Radar systems have been developed to the degree where they transmit the largest peak power ever radiated by any type radio transmitting equipment. This peak power is sometimes as high as 5 megawatts, and seldom less than 20 kilowatts.

In considering power requirements, you must distinguish between two kinds of power output—*peak power*, which is the power during a pulse, and *average power*, which is the average power over the pulse repetition period. While the peak power is very large, the average power may be small, because of the great difference between pulse duration and pulse interval.

The ratio of the pulse length to the pulse repetition period is known as the duty cycle. To convert peak power into average power, multiply the peak power by the duty cycle. These terms are defined in the illustration below. For the pulse waves shown, the relation between the peak



Definition of Pulse Characteristics

↑ Increase → No Change ↓ Decrease	PRF	PULSE WIDTH	PEAK POWER	AVERAGE POWER	MAXIMUM RANGE
	(↑)	→	→	↑	↓
	→	→	(↑)	↑	↑
	→	(↑)	→	↑	↑
	→	(↓)	(↑)	→	→

Circle denotes the elements (characteristics) which are varied. For example, if the PRF is increased, then the average power must be increased to maintain the same peak power. This results in less maximum range (less listening time).

Effect of Varying Pulse Characteristics

power and the average power is expressed by the equation,

$$P_{ave} = P_{peak} \times S \times PRF$$

where S is the pulse length in seconds.

The amount of power that a transmitter must radiate depends first on the minimum amount of power to which the receiver will respond, and second, on the elements which govern attenuation of power to and from the target. Such elements are moisture in the atmosphere, clouds, and weatherproof covering over the antennas.

The effect of varying the pulse characteristics is summarized in the table above.

USES OF RADAR

The following are some of the uses of radar.

1. Aircraft Warning (AW), or Early Warning (EW). Reporting the presence of targets at long range to a central reporting station, and continuing the search for other targets.

2. Ground Controlled Interception (GCI). Directing fighter planes from a ground installation to attack positions against enemy bomber formations.

3. Aircraft Interception (AI). Directing fighter planes from an airborne installation to attack positions against enemy aircraft. These planes were later called *Night Fighters*.

4. Identification, Friend or Foe (IFF). Providing immediate identification of planes and ships.

5. Air to Surface Vessels (ASV), or sea search. Detecting enemy shipping at distances up to 100 miles or more from an airborne installation.

6. Bombing Through Overcast (BTO). Directing bombing of enemy targets through overcast with a high degree of accuracy.

7. Ground Controlled Approach (GCA). Directing planes to land through zero visibility.

8. Tail Warning (TW). Detecting the approach of planes from the rear and protecting the tail position of the plane.

9. Beacons, or navigational aids. Providing a coded indication of bearing and distance to a known point.

10. Gun-laying Radar (AGL). Sighting the guns aboard an aircraft. Radar continuously gives the present position of the enemy aircraft. The radar data is fed to a computer which predicts place where the enemy aircraft will be when the projectile arrives. This prediction is used to aim the guns and the operator can fire at any time.

CHAPTER 6

Timing Circuits

Timing or control is the function of the majority of the circuits in radar. Circuits in a radar set accomplish this function by producing a variety of voltage waveforms, such as square waves, saw-tooth waves, trapezoidal waves, rectangular waves, brief rectangular pulses, and sharp peaks. In sound and in radio, tubes operate within the limits for which they are designed, but in radar timing circuits, they often are violently overdriven, frequently operating at points which range from well in the grid current region to far beyond cut-off. Although all these circuits are broadly classified as timing circuits, the specific function of any individual circuit might be timing, waveshaping, or wave generating. This chapter discusses the principles of a variety of timing circuits, and gives you a foundation for understanding the operation of specific circuits in radar transmitters and receivers.

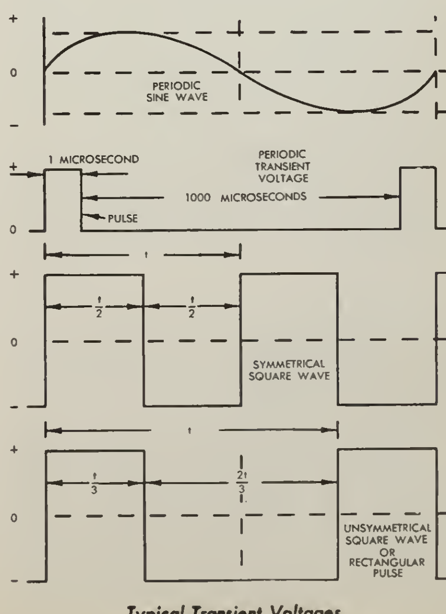
TRANSIENT AND NON-SINUSOIDAL VOLTAGES

Most of the waveshapes in radar timing circuits are produced by combinations of vacuum tubes and resistors or combinations of vacuum tubes and resistors, condensers and inductors. For example, the charging and discharging times of condensers and inductors make these elements extremely useful in timing circuits because of their effect on transient and non-sinusoidal voltages, the voltages widely used in radar.

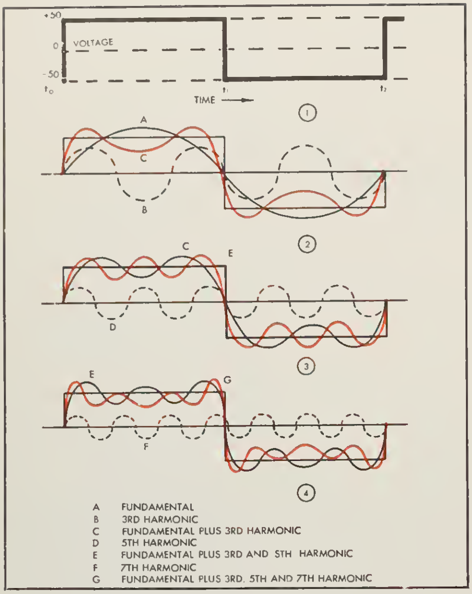
A transient non-sinusoidal voltage is a brief voltage change called a pulse which occurs once and does not recur again for a comparatively long time. Such a voltage change is never sinusoidal but is usually periodic, that is, the change repeats itself at regular intervals. Examples of typical transient voltages are compared to a periodic sine wave in the illustration below.

Analysis

For the purpose of circuit design, there are two ways of analyzing transient and non-sinusoidal voltages. One is to consider that the transient waveshape is a rapid change of voltage, which is followed after a certain interval by another and similar change. The other is to assume that the waveshape is the algebraic sum of a large number of sine waves having different frequen-



Typical Transient Voltages



Composition of Symmetrical Square Wave

cies and amplitudes. The second method of analysis is more useful in most cases. In the design of an amplifier, for example, this type of analysis is necessary because the amplifier cannot handle a transient without distortion unless it is capable of handling all the sine wave frequencies making up the transient.

Transient voltages have many different and complex shapes. The more commonly used transient waves in radar are the square wave, the sawtooth or triangular wave, and the peaked wave.

Square Waves

The above illustration at ① shows a square wave in which voltage is plotted against time. This wave is called *symmetrical* because the alternations are identical.

In terms of the first method of analysis stated in the preceding paragraph, you can consider the square wave in the illustration as a voltage which remains unchanged at minus 50 volts until time t_1, when it suddenly changes to plus 50 volts; it remains at this value until time t_2, then suddenly drops to minus 50 volts and remains at this value until time t_3, and so on.

In terms of the sine wave method, you can analyze the square wave by determining what sine waves are required to reproduce it. To reproduce a symmetrical square wave, you start with a sine wave, which is equal to the square wave repetition frequency, and add to it the odd harmonics of this frequency as shown graphically in the illustration at ②, ③, and ④. Diagram ② shows the waveshape C formed by adding the fundamental frequency A and its third harmonic B which has an amplitude $1/3$ that of the fundamental. Notice that the resultant wave already slightly resembles a square wave as you can see by the square wave superimposed on the diagram. Diagram ③ shows the result of adding the 5th harmonic of the fundamental frequency to the resultant wave C. Notice in the resultant wave E that the corners are much sharper and the top somewhat flatter.

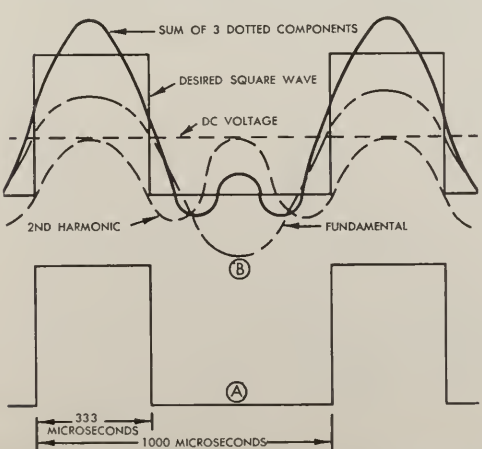
In diagram ④, the 7th harmonic at $1/7$ amplitude is added to form the resultant curve G. This curve is fairly smooth across the top and fairly sharp at the corners. Adding the 9th harmonic at $1/9$ amplitude, the 11th harmonic at $1/11$ amplitude, etc. would further sharpen the corners and flatten the top of the wave.

The assumption that a square wave contains a number of frequencies, although it is actually just a change of voltage, is useful in studying the square wave in circuits. For example, according to the frequency method of analysis the higher harmonics are responsible for the sharpness in the corners of square wave. Therefore, rounding off the corners of the square wave can be explained as discriminating against high frequencies by the circuit. Furthermore, since this method of analysis reduces a square wave to terms of sine waves of different frequencies, it is possible for engineers to design radar circuits by using conventional sine wave engineering practices.

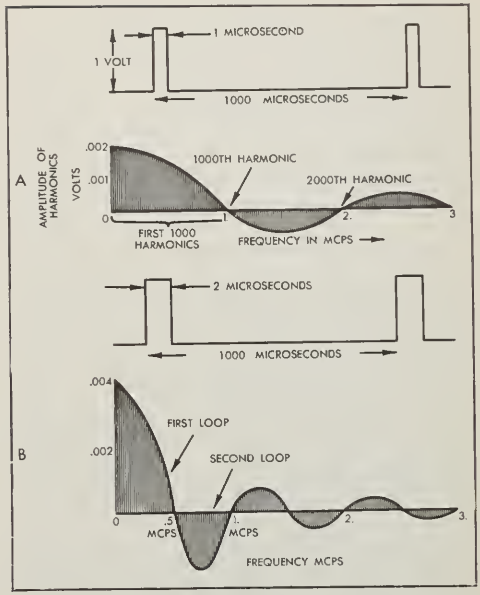
Most square waves employed in radar circuits consist of short pulses separated by long time intervals, as shown by diagram A in the next illustration. Such pulses are constructed in the same way as square waves, but in addition to the sine waves of the fundamental pulse frequency and many harmonics frequencies of fractional amplitudes, a small DC voltage is required. In diagram A, the duration of the pulse is one-third of the repetition time which, in this case, is 1000 times per second. To reconstruct this pulse, add a DC voltage equal to one-third of the desired pulse amplitude to a 1000 cps sine wave having an amplitude ⅔ of the pulse amplitude, a 2000 cps sine wave having an amplitude ⅓ of the pulse amplitude, etc. The solid curve in diagram B represents the sum of all these voltages. Notice that there is a rough resemblance between the outline of the solid curve and the desired square wave. If you add all the harmonics up to about the 15th harmonic (15,000 cps), in the proper phase and amplitude, the resultant waveshape will be quite square.

An important fact to consider is the number of harmonics to include in forming the waveshape. The rule to follow is that the maximum number varies inversely with the pulse width. This means the wider the pulse, the lower the harmonics, and the narrower the pulse, the greater the number of harmonics. For example, in this illustration, the width of the pulse is such that the first 15 harmonics can be included.

In the illustration above, a much narrower pulse is shown—1-microsecond square wave having a repeating frequency of 1000 cps. This pulse requires a great many more harmonics, as you can see from diagram A. Here, the amplitude of each harmonic is plotted on the Y axis, and its frequency is plotted on the X axis. The tops of all the vertical lines are connected by a curve which forms a sine wave. It is impossible to show all the frequencies, since 1000 harmonics actually occur between zero and 1 megacycle.



Composition of Square Wave with Short Duration Pulses

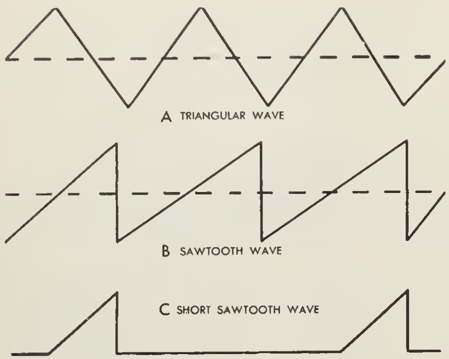


Harmonic Content of Short Square Pulses

When the pulse amplitude, as shown in the illustration, is one volt, the damped sine wave produced by connecting the amplitude of the harmonics is less than .002 volts. From the 1000th to the 2000th harmonic, it is 180 degrees out of phase with the lower frequencies. The relative importance of the various frequencies depends upon the area under the sine wave curve, varying with the size of the area. Notice that the largest area occurs between zero and 1 mcps. The significance of this is that any circuit which does not discriminate against any frequency from 1000 cps to 1 mcps will pass this pulse with little distortion.

Diagram B in the same illustration shows that doubling the pulse width to two microseconds causes the amplitude of the harmonic frequencies to drop to zero at a half megacycle. Thus, any circuit designed for operating with a two microsecond width will not discriminate against any frequency from 1000 cps to 500,000 cps.

All this leads to the conclusion that there is a certain relationship between the highest frequency that a circuit can handle and the width of the pulse. For example, in the case of the 1-microsecond pulse, the highest frequency was 1 mcps, and in the case of the 2-microsecond



Triangular and Sawtooth Waves

pulse, ½ mcps. The rule is that frequency varies inversely with pulse width. Mathematically this is expressed as,

$$F_h = \frac{1}{PW}$$

where F_h is the highest frequency for the circuit, and PW the pulse width in microseconds when the pulse width is much smaller than the pulse repetition period.

Note that the pulse repetition frequency is not considered in the equation, since it is quite evident also by transient analysis that the pulse recurrence frequency has nothing to do with the harmonic frequencies in the pulse. For example, if the repetition frequency in diagram B changed to 500 cps, there would be twice as many harmonic frequencies in the first loop, but the loop still would pass through zero at .5 mcps.

The lowest frequency that this circuit must be able to handle is the fundamental or the pulse repetition frequency. However, several of the low frequency harmonics can be removed from the curves in diagrams A and B without noticeably reducing the areas. Thus a circuit may be designed for a 1-microsecond pulse and handle a frequency band as narrow as 10 kcps to 1 mcps.

Triangular and Sawtooth Waves

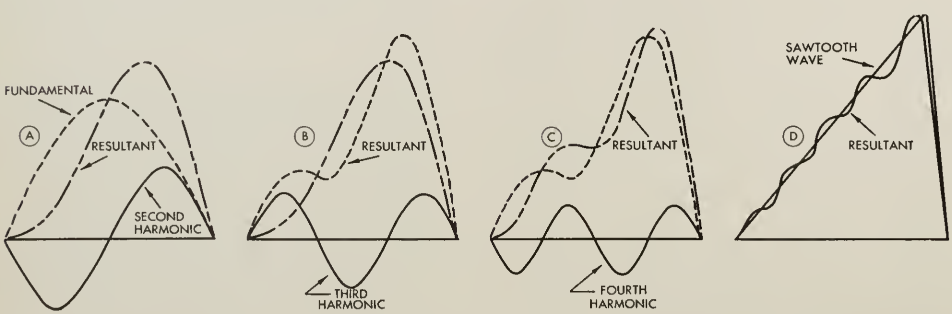
Triangular and sawtooth waves, like square waves, can be constructed with a series of sine waves. Circuits designed to handle the short sawtooth wave shown at C in the illustration at the left must have about twice the band-width as a circuit required for a square pulse of the same dimensions.

COMPOSITION. A sawtooth wave contains both odd and even harmonics. The wave in diagram D below, for example, contains all the harmonics up to the 7th. Notice how closely this wave resembles the sawtooth wave superimposed on it.

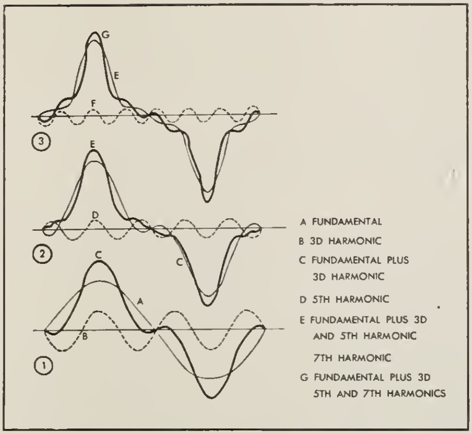
From the viewpoint of a transient voltage, you can consider the sawtooth voltage wave a slow linear voltage change from an initial value to a certain higher value, followed immediately by a sudden drop back to the initial value. The wave may repeat immediately or repeat after an interval of time has passed.

Peaked Wave

A peaked wave is another wave which can be constructed with a number of sine waves. The illustration on the next page shows how the



Composition of Sawtooth Wave



Composition of Peaked Wave

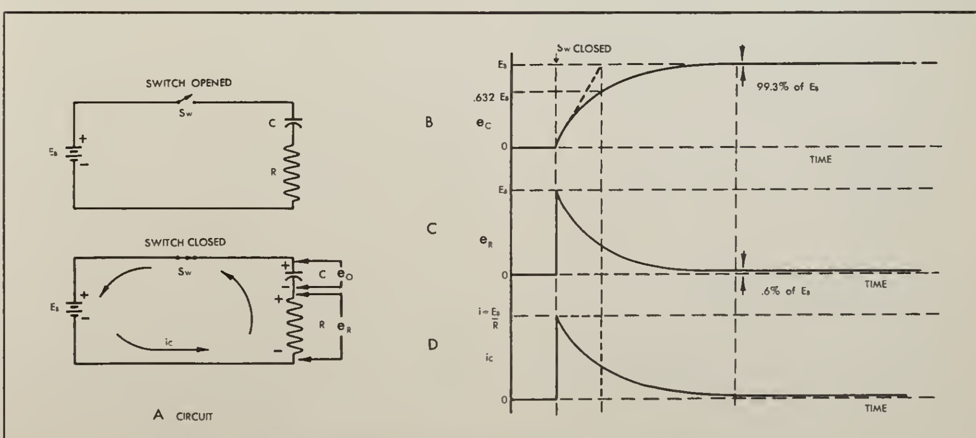
addition of successive odd harmonics to a fundamental sine wave produces a peaked wave. First, the 3rd harmonic, a wave having a lower amplitude, is added to the fundamental. This produces a wave C with a higher peak and steeper sides than the fundamental. Next, the 5th harmonic is added to wave C. This produces a wave E in which the peak is further heightened and the sides further sloped. This process is carried further as illustrated with the addition of each harmonic producing a wave with higher and higher peaks and steeper and steeper sides.

TRANSIENT VOLTAGES IN RC CIRCUITS

The characteristics of the wave forms required by radar, such as slope, duration, or repetition frequency, are determined largely by controlling the variation of a voltage with respect to time. Since the charging or discharging of a condenser through a resistor requires time, resistance-capacitance circuits are commonly used in radar.

To learn how a resistor-condenser circuit affects a waveshape applied to it, it is well to investigate first the voltage and current changes which occur when a DC voltage is applied to the circuit. Most RC circuits and the voltages applied to them are reducible to a simple equivalent circuit containing a battery, condenser, and resistor such as shown at diagram A in the illustration below.

The diagram at A shows a simple series circuit consisting of a battery, a switch, a condenser, and a resistor. According to condenser theory, the condenser C will offer infinite resistance to the direct current flow in the circuit, since it charges to a voltage equal and opposite to the battery voltage E_b when the switch is closed. When the condenser is charged, the battery voltage is cancelled completely, leaving no voltage to cause a current to flow through the resistor. However, before the current does reach zero, there is time consumed. This consumption of time provides the time element required in the formation of many radar waveshapes.



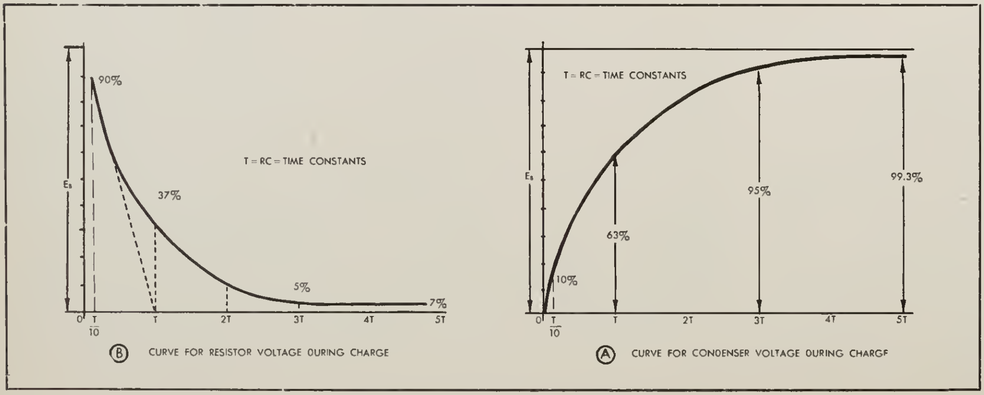
Current and Voltage During Condenser Charge

To understand how a condenser affects the time delay in a circuit, first consider what happens in the condenser itself and how it is affected by the rest of the circuit. Assume that before you close switch sw, the condenser is completely discharged—that is, there are an equal number of electrons on each of its plates. When you close the switch, the battery voltage E_B immediately forces electrons into the condenser plate connected to the negative terminal of the battery, placing a negative charge on it. This creates a stress on the dielectric which, in turn, forces electrons away from the other plate and makes its charge positive. It might seem that if it were not for the limiting effect of the series resistor, the number of electrons that would accumulate on the condenser plates would be infinitely high because, upon the closing of the switch, the current in the circuit would immediately jump from zero to a value determined only by the resistor. (That is, $i = \frac{E_B}{R}$ since the uncharged condenser offers no opposition to current flow). The fact is, however, that the condenser itself offers opposition to the current. Just as soon as a few electrons accumulate on the condenser plate, there is a like number forced away from the other plate. The net result is a small voltage difference across the condenser, the plate with more electrons being the negative point of voltage, and the plate with less electrons being the positive point of voltage. The polarity of the circuit then becomes as shown in diagram A. Notice here that the condenser is charged to a voltage opposite in polarity to the battery voltage. Thus, in a short time after you close the switch, the movement of electrons—that is, current flow—does not remain at its initial value, but drops to a lower value because the voltage across the condenser subtracts from the charging voltage. Since the voltage applied to the condenser is thus effectively reduced, it charges at a lower rate.

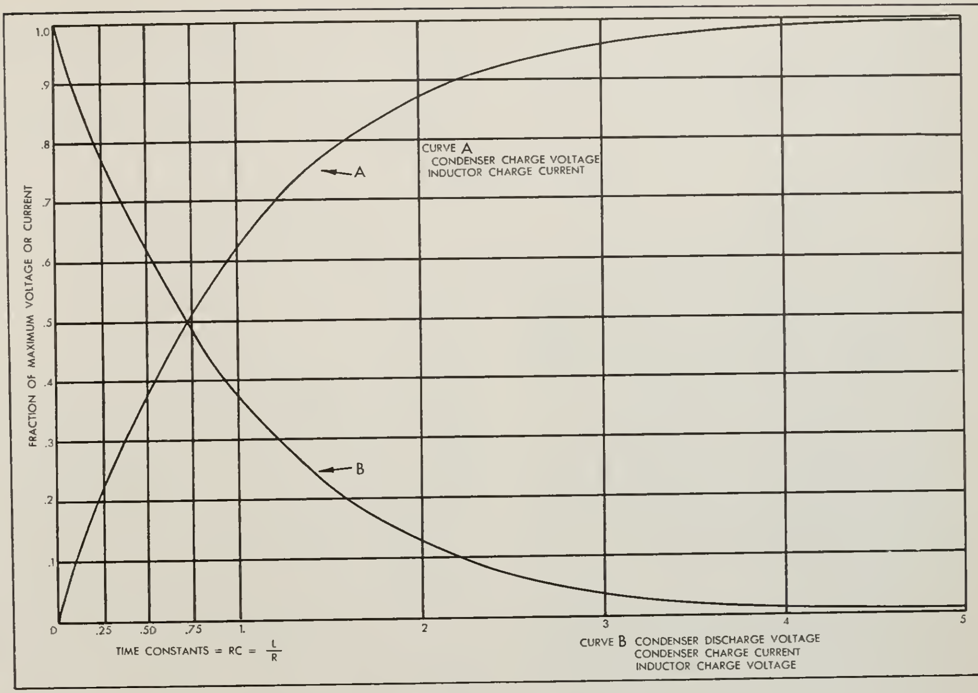
Now consider the action resulting during the further charge at a slow rate up to the point of so-called complete charge. (Theoretically the condenser never completely charges to 100% charge.) When the switch was first closed as shown in diagram A, the condenser charged rapidly to 63.2% of the applied voltage. At this point the current flow in the circuit was 36.8% of its initial value, and the resistor voltage 36.8% of the applied voltage. With the continuous increase in the voltage difference across the condenser and its subtraction from the voltage applied to it, the charge *rate* tapers off, but the charge itself rises exponentially to the point where the condenser is considered charged.

Notice the voltage and current curves shown at B, C, and D. Had the condenser charged at its original rate, its charge curve would be the dotted line at B. The slope of this condenser charge curve is expressed as E_B/RC.

The term RC in this ratio is called the time constant (t.c.) of an RC circuit. It is the time required to charge a condenser to 63.2% of its final voltage. To find the time constant in seconds, multiply R, the resistance in ohms by C, the capacity in farads.



Exponential Curves for Rough Approximations



Universal Time Constant Chart

The condenser voltage curve at B and the instantaneous current curve at D are exponential curves. An exponential curve is the graph of an equation in which a variable appears as an exponent.

The instantaneous resistor voltage curve at C can be found from the instantaneous current values by using the Ohm's law formula, E = IR. Since R is constant, E, the voltage across the resistor is directly proportional to the current through it. Therefore, its curve has the same exponential slope as the instantaneous current curve.

Quantitative Values of Condenser and Resistor Voltages

There are three methods of determining the instantaneous values of voltage across condensers and resistors in RC circuits. The first is a rough estimate from the exponential curve with the values at four main points known. The second is a close approximation from a very accurate exponential curve on a universal time constant chart. The third involves solving the equation for the curve with the quantities for the RC problem inserted in the equation.

ROUGH APPROXIMATION CURVES. The voltage or current at any instant during the charging time of a condenser can be approximated from the two curves shown on the preceding page. These curves employ only rough values, but it is well that you memorize them, since they occur frequently. They are also useful for approximating current at any instant during the charge of a condenser. To find current, substitute the total current flow in the circuit for E_B on the curves.

UNIVERSAL TIME CONSTANT CURVES. A universal time constant chart, such as the one shown above, provides a much more precise method for determining the voltage and the current at any instant during the charge of a condenser than the method just described. The chart shows a number of time constants plotted against the fraction of maximum current or voltage. In learning to use this chart study the following examples:

Examples

Problem 1. In the circuit to the right, assume that C equals 0.001 mf and that R equals 100 K. Find the condenser voltage 200 microseconds after the switch is closed.

Solution:

a. Find the time constant.

$$t.c. = RC$$
$$t.c. = 100{,}000 \times .001$$
$$= 100 \text{ microseconds}$$

b. Since 200 microseconds is equivalent to 2 time constants, enter the universal time constant chart at 2 along the bottom and proceed up to curve A. (This curve represents the condenser charge voltage.) At curve A, proceed horizontally to the vertical scale and read 0.86. This represents the fraction of maximum voltage.

c. Thus, the condenser voltage is about 86% of maximum. With 200 volts as the applied voltage, the condenser voltage equals 200 × .86 or 172 volts.

Problem 2. In the same circuit find the resistor voltage 200 microseconds after the switch is closed.

Solution.

Using Kirchoff's Law.

From the previous example, the condenser voltage equals 172. Since the circuit is a series circuit, the resistor voltage is the difference between the applied voltage and the condenser voltage. Thus, it is 200 − 172 or 28 volts.

Solution.

Using the universal time chart.

Knowing that the resistor voltage starts high and then drops off you use curve B. Since there are two time constants involved, go up from 2 to curve B and then across to the vertical scale. At the intersection you find 0.14. Since the applied voltage is 200 volts, the resistor voltage must be 200 × 0.14 or 28 volts.

Solution.

Using universal chart, find current.

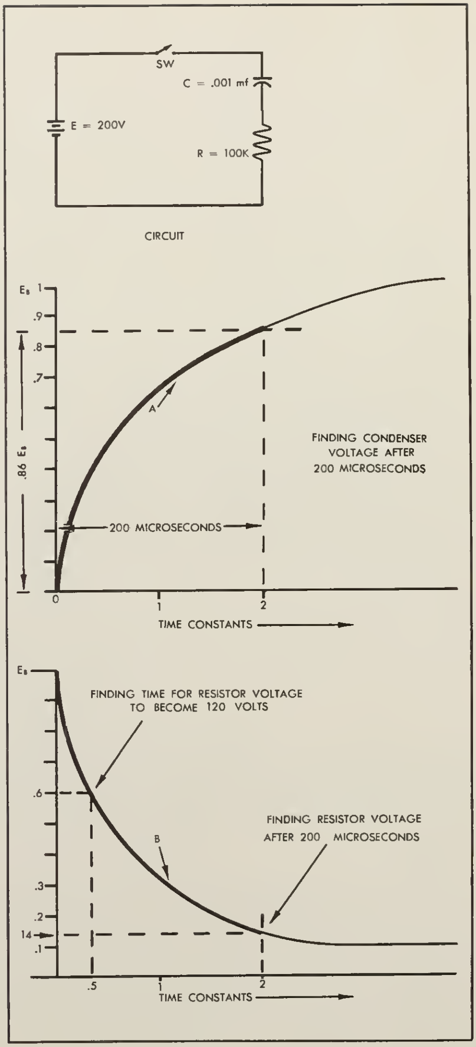
By Ohm's law the maximum current is $\frac{200}{100K}$ *or 2 ma.*

At two time constants the current is .14 of maximum, according to curve B and the scale at the left. Therefore, it equals .14 × 2 ma or .28 ma. By Ohm's law, E = IR. The instantaneous voltage is .28 ma × 100 K or 28 volts.

Problem 3. In the same circuit find how long it takes for the resistor voltage to drop to 120 volts.

Solution.

120 volts is $\frac{120}{200}$ *or .6 of 200 volts. Find .6 on the vertical scale, go to curve B, thence down to .5 time constants. If one time constant equals 100 microseconds the time required for the resistor voltage to drop to 120 volts must equal .5 × 100 or 50 microseconds.*



Using Universal Time Constant Chart

NOTE

Any value of voltage and any size of resistor or condenser can be used with the universal time constant chart. Time can be long or short, providing all units have the same time constant. In the chart shown, after 5 time constants the condenser is considered to be fully charged, and the resistor voltage and the circuit current are considered to be zero.

Exponential equation. Another method of determining values of voltage across a condenser and a resistor in a RC circuit is by use of the exponential equation for the instantaneous voltage curves. The exponential equation for the resistor voltage curve B shown at the right is,

$$e_R = E_B \epsilon^{\frac{-t}{RC}}$$

where e_R *= instantaneous resistor voltage*
E_B *= the applied voltage*
ϵ *= the base of natural logarithms, 2.718*
t *= time in seconds*
R *= resistance in ohms*
C *= capacity in farads*

This equation leads to much more precise results than a universal time constant resistor voltage curve. For example, from the time constant curve you find that the resistor voltage reaches 37% of the applied voltage after one time constant. For purpose of comparison, determine the voltage across the same resistor by the exponential equation just given $\left(e_R = E_B \epsilon^{\frac{-t}{RC}}\right)$

Since the time constant equals 1, $e_R = E_B \epsilon^{-1}$

From tables of hyperbolic functions,
ϵ^{-1} *equals 0.3679*

Therefore, substituting,
$$e_R = E_B \times 0.3679$$

NOTE

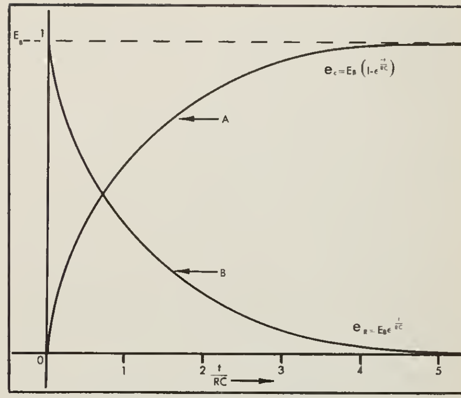
Finding the value of $\epsilon^{\frac{-t}{RC}}$ *requires the use of tables of hyperbolic functions of which an extract is shown. Thus, in finding the value of* $\epsilon^{\frac{-t}{RC}}$ *in the example, substitute 1, which is the time constant, for* $\frac{-t}{RC}$. *The expression then becomes* ϵ^{-1}. *Referring to the table, notice that when x (the exponent) equals 1, the function* ϵ^{-1} *equals 0.3679 as is indicated by the arrows.*

According to the universal chart curve, the voltage of e_R was indicated as approximately 37%. However, the equation and the table show it to be exactly 36.79% of the applied voltage at the end of one time constant. Therefore, for greatest accuracy, use the exponential equation method.

The equation for curve A, the instantaneous condenser voltage charge curve is,

$$e_C = E_B \left(1 - \epsilon^{\frac{-t}{RC}}\right)$$

where e_C = condenser voltage at any time, t.



Exponential Equations of Voltage Curves

Examples

To understand the use of these exponential equations, study the following examples and their solutions:
Problem 1. In the circuit on the preceding page, find the condenser voltage 200 microseconds after the switch is closed.

Solution

$$e_C = E_B\ (1 - \epsilon^{-t/RC})$$

Substituting the values in the problem,

$$e_C = 200\ \left(1 - \epsilon^{\frac{-2\times10^{-4}}{10^{-9}\times10^{5}}}\right)$$
$$= 200\ (1 - \epsilon^{-2})$$

From the tables, $\epsilon^{-2} = 0.1353$

Substituting,

$$e_C = 200(1 - 0.1353)$$
$$= 200 \times 0.8647 = 172.94 \text{ volts}$$

x	FUNCTION	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.9	ϵ^{x}	2.4596								
	ϵ^{-x}	0.4066								
1.0	ϵ^{x}	2.7183								
	ϵ^{-x}	0.3679								
↑	↑	↑								
1.1	ϵ^{x}	3.0042								
	ϵ^{-x}	0.3329								
1.2										

Extract from Tables of Hyperbolic Functions

Problem 2. Using the same circuit and circuit values, find the resistor voltage 200 microseconds after the switch is closed.

Solution

$e_R = E_B \epsilon^{-t\ RC}$

Substituting the values in this problem,

$e_R = 200\ \epsilon^{\frac{-2\times10^{-4}}{10^{-9}\times10^{5}}}$

$= 200\ \epsilon^{-2}$

From the tables, $\epsilon^{-2} = 0.1353$ *(from table)*

Substituting,

$e_R = 200 - 0.1353 = 27.06$ *volts*

Problem 3. Find how long it will take for the resistor voltage to drop to 120 volts. Assume that $e_R = 120$ *volts.*

Solution

$e_R = E_B \epsilon^{-t\ RC}$

Transposing, $\frac{e_R}{E_B} = \epsilon^{-t\ RC}$

Substituting, $\frac{120}{200} = \epsilon^{-t\ RC}$, *or* $.6 = \epsilon^{-t\ RC}$

From the tables $\epsilon^{-x} = .6$ *when* $x = .51$ *(x represents the exponent of* ϵ*)*

Therefore

$\epsilon^{-t/RC} = -.51$

$-t/RC = .51$

Changing signs and transposing:

$t = .51\ RC$

$= .51 \times 10^{5} \times 10^{-9}$

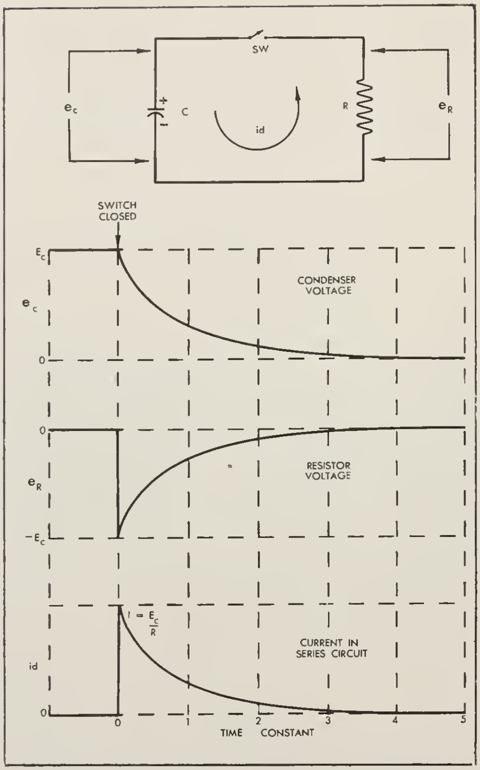
$= .51 \times 10^{-4}$

$= 51$ *microseconds*

You can find the values for ϵ^{-x} on the more complete slide rules by placing the hairline on the ratio t/RC on the A scale, and reading the value $\epsilon^{-t/RC}$ on the LL00 or LL0 scales. The LL00 scale reads values for $\epsilon^{-t/RC}$ ratios between .1 and 10. To familiarize yourself with the meaning of the slide rules values, it is a good idea at first to check the slide rule against the tables or the universal curve.

Condenser Discharge Through Resistance

The discharge of a condenser through a resistor follows the same exponential curve as the charge through a resistor as you can see in the simple circuit above. When the condenser is charged to a voltage of E_c, there is an excess of electrons on one plate of the condenser. This unbalanced condition causes a potential difference which exists as long as the switch is open (assuming the condenser is perfect). The portions of the curves indicating open switch condition show that the condenser voltage is E_c, the current in the circuit is zero, and the resistor voltage is also zero.



Current and Voltage During Condenser Discharge

When the switch is closed, the electrons on the negative plate immediately surge through the circuit toward the other condenser plate, building up a current which is limited to a value determined by Ohm's law—that is, a value equal to E_c/R. This is what happens. The initial current causes a voltage drop across the resistor equal to the condenser voltage. But as soon as current flows, the number of excess electrons causing the current flow decreases. This causes a voltage decrease which in turn decreases the current. The decrease is not linear, however. The high initial current decreases the voltage rapidly, but subsequent lower voltages cause lower currents, which removes the charge at a lower rate. The discharge becomes slower and slower until at the end of five time constants, the current is less than 1% of the initial value. The voltage causing this low current is also extremely low because the condenser is 99.4% discharged.

Theoretically, the time required to discharge the condenser is infinitely long, but for practical purposes, five time constants may be considered as the time required for complete discharge.

SOLVING CONDENSER DISCHARGE PROBLEMS. To determine instantaneous values of current and voltage in RC circuits during condenser discharge, use the same methods that you used previously in the charging circuit problems. One fact of significance in discharging circuits is that the voltages across the resistor and the condenser follow the same curve inasmuch as they both start at high values and decrease exponentially to zero. Therefore, Curve B of the Universal Time Constant Chart represents the instantaneous values of voltage both across the condenser and the resistor during condenser discharge. The only requirement for using it for both voltages is assigning e_R a polarity opposite that of e_C. Remember that the voltage across the resistor is not only equal but opposite that across the condenser when the condenser is discharging. This is as it should be according to Kirchoff's law, which states that the sum of the voltages around a series circuit is zero.

As just mentioned, you can use any of the three methods given earlier to solve for the voltage and current values—that is, you can use the curves which roughly approximate voltages and currents, the universal curves, or the exponential equations.

The following problems are solved by all three methods. They are first solved by the less precise approximation and universal time constant curves, and then again by exponential equations. Note that the results obtained by the exponential equations differ slightly from the answers obtained by use of the curves. The differences are due, of course, to the greater degree of accuracy obtainable by the equation method.

Examples

Problem 1. Assume that in the adjacent circuit, the condenser is charged to 200 volts when the switch is closed. Find the condenser voltage 300 microseconds after the switch is closed.

Solution

First find time constant by the equation, t.c. = RC. Substituting the values indicated in circuit,

$$t.c. = 100\ K \times .001\ mf$$

$$t.c. = 10^5 \times 10^{-9} = 10^{-4}$$

Therefore t.c. = 100 microseconds

Since the time constant is 100 microseconds and since the problem requires 300 microseconds, the condenser discharges for 3 time constants. Turning to the rough chart B on page 6-6 or diagram B below, refer to the heavy lined curve. There you see that at 3 time constants the condenser has discharged to 5% of E_C or .05 x 200, or 10 volts. Thus, after 300 microseconds, the condenser voltage is 10 volts.

Problem 2. Find the current 200 microseconds after the switch is closed.

Solution

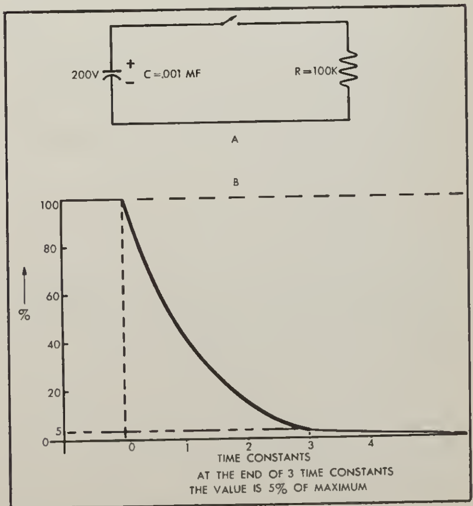
The current decreases from an initial value of 2 ma (calculated by Ohm's law, $\frac{E}{R} = \frac{200V}{100K}$*). The time of discharging is 200 microseconds, or 2 time constants. Referring to the universal time constant curve B on page 6-7, the value of current at 2 time constants is 14% of its maximum value. Since 14% of 2 ma is .28 ma, the current after 200 ma is .28 ma.*

Problem 3. Find the resistor voltage after one time constant, using time constant curve B.

Solution

After 1 time constant, the voltage drops to 37% of its maximum value. Since the resistor voltage starts at 200 volts and drops to zero when the condenser is completely discharged, it is 37% of 200 volts, or 74 volts, 1 time constant after discharge starts.

Notice here that curve B, which is the condenser discharge voltage curve, also is used for voltage decrease across the resistor during discharge.



Condenser Discharge Problem

In problems concerning RC voltage values, the following exponential equations apply as indicated:

1. *Instantaneous resistor voltage:* $e_R = E_c \epsilon^{\frac{-t}{RC}}$
2. *Instantaneous condenser voltage:* $e_C = E_c \epsilon^{\frac{-t}{RC}}$
3. *Instantaneous discharge current:* $i_d = \frac{E_c}{R} \epsilon^{\frac{-t}{RC}}$

In each equation, E_c *represents the initial condenser charge voltage.*

Examples

Problem 1. Find e_c *when t equals 300 microseconds.*

Solution

$$e_c = E_c \epsilon^{-t/RC}$$
$$e_C = 200 \epsilon^{\frac{-3 \times 10^{-4}}{10^5 \times 10^{-9}}}$$
$$e_C = 200 \epsilon^{-3}$$
$$\epsilon^{-3} = .0498 \text{ (from tables)}$$
$$e_C = 200 \times .0498$$
$$e_C = 9.960 \text{ volts.}$$

Problem 2. Find i_d *when t = 200 microseconds.*

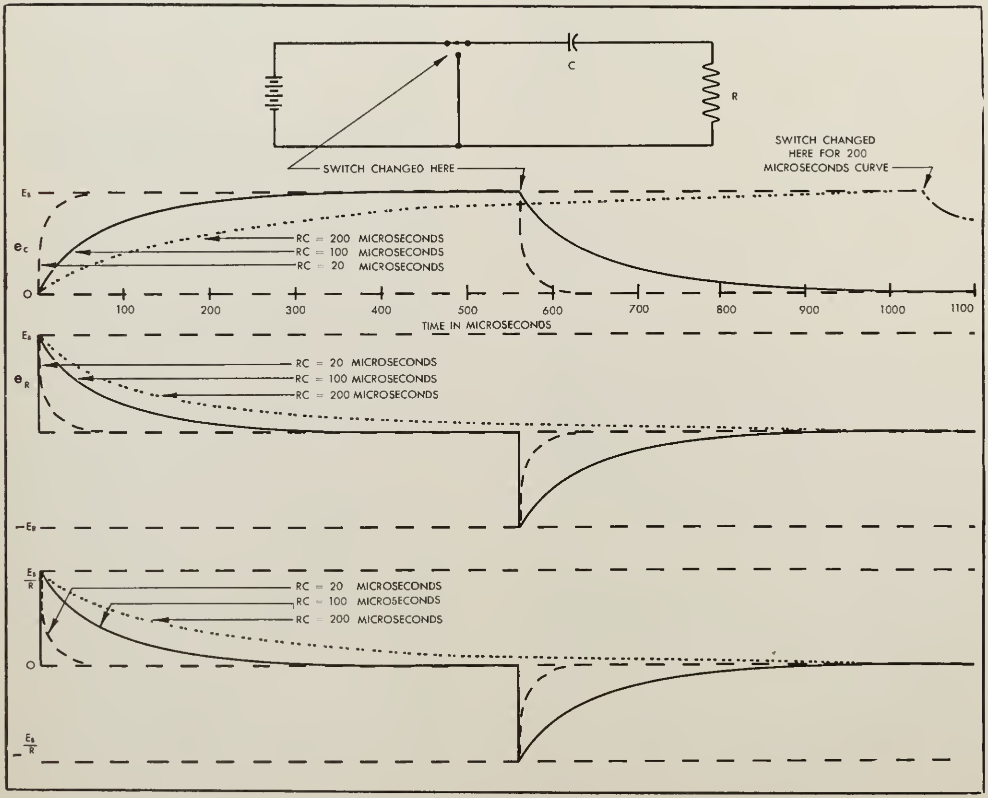
Solution

$$i_d = \frac{E_c}{R} \epsilon^{-t/RC}$$
$$= \frac{200}{10^5} \epsilon^{\frac{-2 \times 10^{-4}}{10^5 \times 10^{-9}}}$$
$$= .002 \epsilon^{-2}$$
$$\epsilon^{-2} = 0.135$$
$$i_d = .002 \times .135$$
$$= .27 \text{ ma}$$

Problem 3. Find e_R *when t = RC (time constant = 1)*

Solution

$$e_R = E_c \epsilon^{-t/RC}$$
if $t = RC$ *then* $t/RC = 1$
$$e_R = 200 \epsilon^{-1}$$
$$\epsilon^{-1} = .368 \text{ (from table)}$$
$$e_R = 200 \times .368 = 73.6 \text{ volts}$$



Effect of Varying R and C in RC Circuit

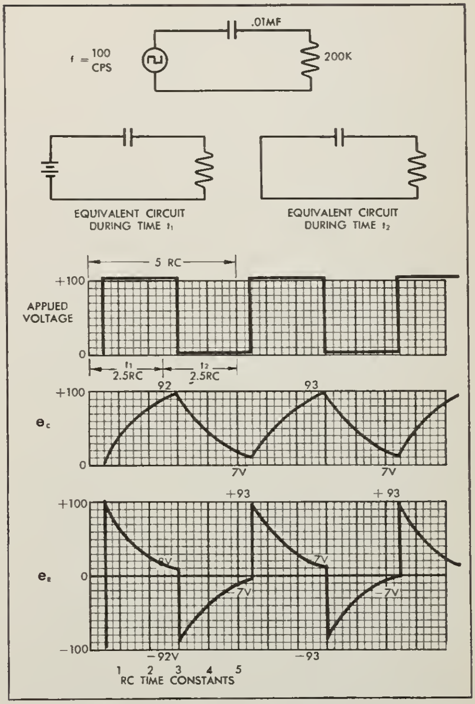
Effect of Varying R and C

Thus far, the RC curves illustrated were drawn separately but on the preceding page three curves are superimposed on one another for the purpose of showing what effect changing the time constant (RC) has on the waveshape produced by an RC circuit. Each curve represents an RC circuit having the same applied voltage but a different time constant. Naturally, in all RC circuits, values reach maximum (or zero) after 5 time constants. However, since the duration of each time constant is different, the curves differ in duration. The curve labeled 20 microseconds, for example, is produced by a small RC, that is, a small condenser or resistor. A small condenser has the property of quick charging, and a small resistor connected to it permits a high charging current to flow, further facilitating rapid charging. Due to the small RC, this curve reaches its maximum limit quickly or, in approximately 100 microseconds.

Larger RC values produce longer time constants. A larger condenser requires more current to charge it completely, while a larger resistor limits the current more and thereby increases the time required for charging. Thus, the curve labeled 200 microseconds has a long time constant. Notice that 1000 microseconds are required to fully charge this condenser.

Notice that the e_R and i curves both follow the condenser charge curve very closely. The first half of each of these curves represents the condenser charge, and the last half the condenser discharge. The difference between the two halves is due to the position of the switch.

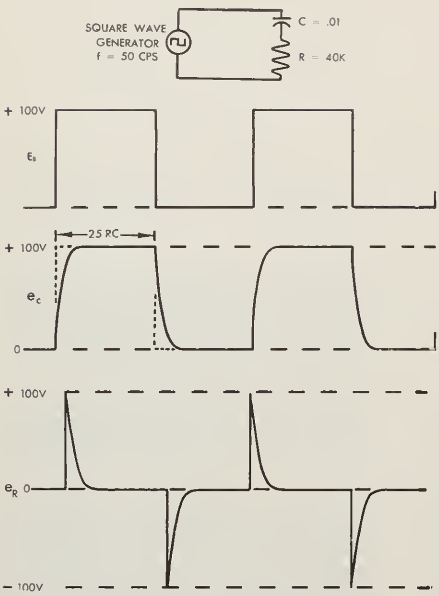
Effect of a medium time constant. Since most RC circuits in radar sets are associated with high speed timing devices, the voltages generally applied to an RC circuit are rapidly repeating square wave voltages rather than DC voltages. The illustration above shows the condenser and resistor voltages resulting from applying a square wave having a period of five time constants to the circuit at the top. For the purpose of analyzing the circuit, the square wave generator is redrawn into two equivalent circuits. They make it easier for you to understand the action of the generator, providing that you regard the square wave first as a constant voltage lasting for a definite time, and then as a zero voltage lasting for an equal time. Since the constant voltage equals the battery voltage during the first half of the



Effect of Medium Time Constant on Square Wave Input

square wave, the equivalent circuit for this time period (t_1) shows the battery voltage. During the second half of the square wave, or the zero voltage period, the condenser discharge current flows through the generator. It is therefore not necessary to include the battery in the (time t_2) equivalent circuit. The battery, however, must be replaced by whatever resistance the discharge current encounters in the generator, which in the case being considered is zero resistance.

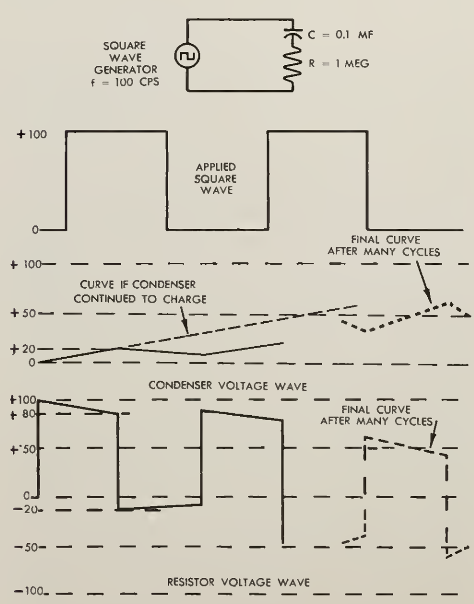
In contrast with the time constants shown in previous illustrations, the time constant used in this illustration is called a medium time constant. Actually any distinction between long, short, or medium time constants is arbitrary, but in general the practice is to consider that a circuit has a long time constant when the RC product is equal to or greater than ten times the period of the applied voltage, a short time constant when the RC product is equal to or less than one tenth of the period of the applied wave and a medium time constant when the RC product lies between these two limits.



Waveshapes for Short Time Constant RC Circuit

Effect of a short time constant. Consider now the effect of a short time constant on an RC circuit. The square wave shown above is applied to a resistance-capacitance combination with a time constant equal to .02 of a frequency cycle. Thus, in 20% of a half cycle, five time constants will occur. If you assume as before that the square wave is a DC voltage applied for a short time and then shorted out, you can apply the rules you used previously. These stated that the charging rate for a condenser with a short time constant is very high at first and then drops slowly to zero. Similarly, its discharge rate is very high to start with and then gradually decreases to zero. The early rise of the condenser voltage to the full applied voltage and its quick drop to zero cause the condenser waveshape to resemble the square wave input. On the resistor, the short duration of the high charging and discharging currents have a quite different effect. The rapid rise and drop of the resistor voltage cause the resistor waveshape voltage to be peaked at each square wave change over (change from maximum voltage to zero or from zero to maximum). These sharp peaks are excellent for initiating action in another circuit, as, for example, in a radar transmitter which is designed to transmit a one microsecond pulse whenever it is triggered by a positive voltage. The sharp positive pulses shown in the illustration may be used for this purpose since they occur at precisely the changeover time in the square wave. Referring again to the illustration, notice that the universal exponential curve occurs over and over in the waveshapes. You can readily determine the time and amplitude of each curve by inserting the values for each curve from the universal time constant curve.

Effect of a long time constant. In the illustration showing the waveshapes produced by an RC with a long time constant, the time constant is ten times as long as the period. On referring to the universal curve, you will see that the condenser in the circuit will charge almost linearly for about 20% of a time constant. This causes the condenser curve in the case illustrated here to be linear during the entire half cycle. The dotted curve shows how the condenser would have continued to charge if the applied voltage had not been removed. Note how the voltages in the circuit illustrated comply with Kirchoff's law which states that the sum of volt-

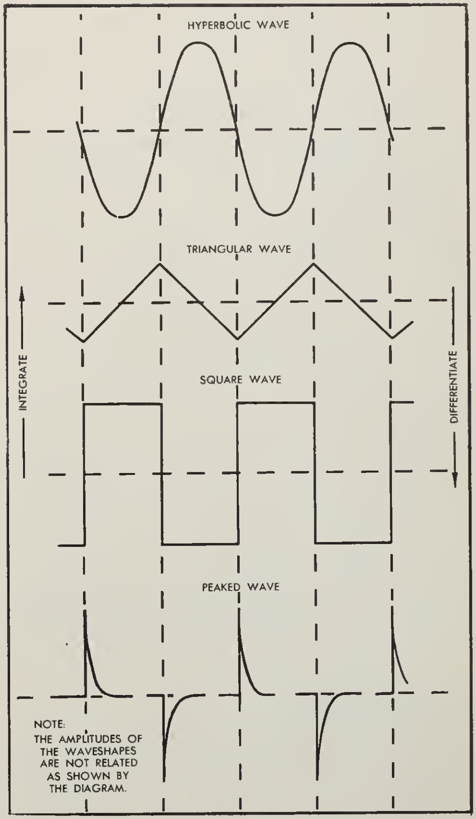


Waveshapes for Long Time Constant RC Circuit ages around the circuit must equal the applied voltage. During the first half cycle, the applied voltage is 100 volts. Although the resistor and condenser voltages are constantly changing, their sum at any one instant is also 100 volts. Thus, at the middle of the first half cycle, for example, the condenser voltage is 10 volts, and the resistor voltage is 90 volts, making a total of 100 volts and fulfilling Kirchoff's law. During the other half cycle, the applied voltage from the generator is zero. In this case the only voltage in the circuit is the condenser voltage, which starts out at 20 volts and discharges 20%. As a result, the resistor voltage becomes minus 20 volts making all voltages around the circuit equal to zero and again fulfilling Kirchoff's law.

The applied voltage in the illustration showing waveshapes produced by long time constants may either be positive voltages or voltages having a 50-volt DC component plus a great number of AC harmonics. In the latter case, the action of the condenser is to charge to the DC value after a number of cycles, this action being noticeable only during the first few cycles. Since a condenser does not pass a DC voltage, the resistor voltage appears as pure AC as soon as the condenser is completely charged. You can see this from the resistor voltage wave in the illustration. Such a curve thus formed varies an equal amount on each side of the zero axis. Another important fact about this curve is that since each half cycle is a replica of the exponential curve, you can determine instantaneous values of voltages either by the exponential equations or the universal time constant chart, both of which were given earlier in this chapter.

Differentiation and Integration

The process of changing a square wave into a peaked wave is called *differentiation*. When the square wave is changed into a triangular wave, the process is called *integrating*. The illustration shows both differentiated and integrated waveshapes each of which may be expressed algebraically. When the algebraic expression representing the wave is differentiated, there results a new expression, which is the expression for the next lower curve shown. When the expression is integrated, the new expression is the same as the expression for the next higher curve. Notice that the top curve is not a sine wave but a series of hyperbolic curves which form a periodic wave shape. Differentiating or integrating a sine wave by an RC circuit produces another sine wave having a different amplitude and phase, but with the same sinusoidal waveshape.

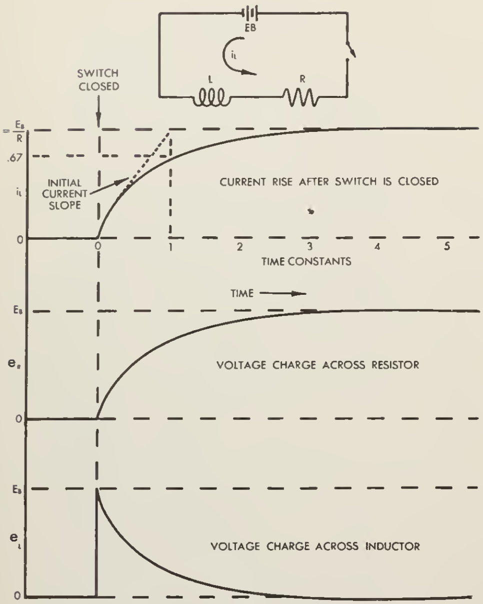


Differentiation and Integration

In an RC integrating circuit (one that produces an output which is the integral of the input), the time constant must be very long and the output must be taken across the condenser. In RC differentiating circuits, the time constant must be very short and the output must be taken across the resistance.

TRANSIENT VOLTAGES IN LR CIRCUITS

The operation of LR circuits is very similar to that of the RC circuits just discussed since the inductor current in an inductor-resistor circuit rises exactly like condenser voltage in an RC circuit. With slight modifications, all the waveshapes obtainable in an RC circuit are also obtainable in an LR circuit.



Current and Voltage Curves in LR Charging Circuit

LR Charging Circuits

The illustration shows typical current and voltage curves produced in LR circuits. In the LR circuit shown, inductance L is assumed to be perfect, that is, it offers no resistance to DC. When the switch is open, there is no voltage across the circuit and because of this, the current and voltage are zero. When you close the switch, the battery voltage E_b is applied across the resistor R and the inductance L. Current attempts to flow, but the inductor opposes this current by building up a back EMF, that, at the initial instant, exactly equals the input voltage E_B. Since no current can flow under this condition, there is no voltage across the resistor R. The two bottom curves thus show that at the instant the switch is closed in the circuit, all the voltage is impressed across L and no voltage across R.

As current starts to flow, the voltage e_R appears across R, and the voltage e_L across L is reduced by the amount of e_R at that instant. A reduced voltage across the inductor means a less rapid increase in i_L and thus a less rapid increase in the resistor voltage. The e_L curve shows that e_L finally becomes zero when the current stops increasing while the e_R curve shows that e_R builds up gradually to the input voltage as the charging current rises. Under a state of steady conditions, the resistor is the only factor which limits the magnitude of the current.

Suppose you examine the three curves closely. Notice in the i_L (instantaneous current) curve that the current starts from zero and increases rapidly at first and then tapers off to practically a zero rate of increase at the end of 5 time constants. According to Ohm's law, the instantaneous values of the voltage across the resistor are directly proportional to the current flowing through it. Therefore, the e_R curve is similar to the i_L curve. Since the sum of voltages around a series circuit (the circuit illustrated is a series LR circuit) equals zero, and since the applied voltage is constant, the inductor voltage e_L is always equal to the difference between E_B and e_R. The inductor voltage equals E_B when the resistor e_R equals zero, and equals zero when e_R is at maximum value. If you were to reverse the polarity of the e_L curve, you may also consider it is representing the back EMF in the inductor. (The polarity of back EMF is always opposite that of the applied voltage.) At first, the back EMF is high since the current tends to jump to a value of E_B/R. This large back EMF opposes the applied voltage and keeps the current small. In order for the back EMF to be sustained at its initial high value, the current must continue to increase at the same rate as initially (dotted line in i_L curve). This is not possible, however, for as the current increases, the resistor voltage increases in proportion. The increase in resistor voltage effectively reduces the inductor voltage. Hence, the back EMF decreases while current and resistor voltage increase. All three curves change along exponential curves as shown in the above illustration.

LR Time Constant

The meaning of the time constant in an LR circuit is analogous to the meaning of the time constant in an RC circuit. An LR time constant is defined as the time required for the current through an inductor to increase to 63.2% of its maximum value. The formula for the LR time constant is

$$\text{t. c.} = \frac{L}{R}$$

where t represents the time constant in seconds, L the inductance in henrys, and R the resistance in ohms.

The ratio $\frac{L}{R}$ also represents the time required for the resistor voltage to equal 63.2% of the applied voltage, and the time required for the inductor voltage to equal 36.8% of the applied voltage.

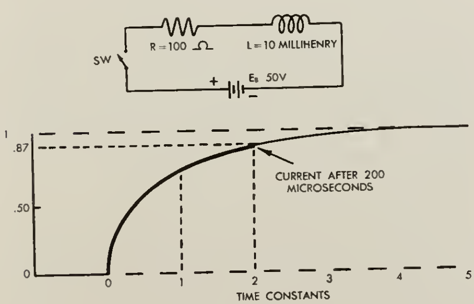
LR Charging Problems

The shapes of the curves for current and voltage in the LR circuits are like those in the RC circuit. Therefore you can use the rough approximation curves, the universal time constant curves, and the exponential equation for determining instantaneous values of current and voltage.

Study the following sample problems to learn how to find instantaneous values of current and voltages in LR circuits. These problems are solved first by the curve chart method and then by exponential equations.

Examples

SOLVING BY USE OF CURVES.



Problem 1. Using the values indicated in the above circuit, find the amount of current 200 microseconds after the switch is closed.

Solution

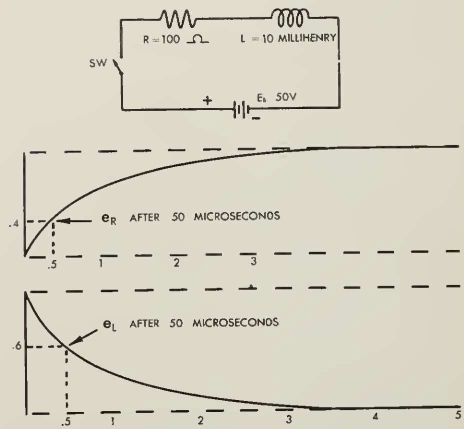
Find the time constant by the formula, $t.c. = \frac{L}{R}$

$$t.c. = \frac{.01 \text{ henry}}{100 \text{ ohms}}$$
$$= 10^{-2} \times 10^{-2}$$
$$= 10^{-4}$$
$$t.c. \text{ (time constant)} = 100 \text{ microseconds}$$

Since the time constant is 100 microseconds, and the time involved is 200 microseconds, the charging action last 2 time constants. From the rough approximation curves on page 6-6 or the universal time constant chart on page 6-7, you find that the current in the inductor is 87% of maximum at 2 time constants. Calculating the maximum current by the formula, $I = \frac{E_R}{R}$,

$$I = \frac{50}{100} = 0.5 \text{ amperes.}$$

Since the maximum current is 0.5 amperes, it will be .85 of .5 or .425 amperes at the end of 200 microseconds.



Problem 2. Using the same circuit repeated above and the same time constant, find e_R *and* e_L, *50 microseconds after the switch is closed.*

Solution

The rough approximation curves do not contain a 50 microsecond time. Therefore, refer to the universal time constant chart. (Curve A is the resistor voltage curve; B the inductor voltage during charge.)

From curves A and B

$e_R = .4$ *of maximum voltage*

$e_L = .6$ *of maximum voltage*

Therefore, $e_R = .4 \times 50$ *volts*

$= 20$ *volts*

$e_L = .6 \times 50$ *volts*

$= 30$ *volts*

SOLVING BY USE OF EXPONENTIAL EQUATIONS. The following are the LR charging circuit exponential equations.

1. *Instantaneous current:* $i_L = \frac{E_B}{R}\left(1 - \epsilon^{\frac{-tR}{L}}\right)$
2. *Instantaneous resistor voltage:* $e_R = E_B\left(1 - \epsilon^{\frac{-tR}{L}}\right)$
3. *Instantaneous inductor voltage:* $e_L = E_B \times \epsilon^{\frac{-tR}{L}}$

Problem 1. Using the same circuit, find i_L *when* $t =$ *200 microseconds.*

Solution

$$i_L = \frac{E_B}{R}\left(1-\epsilon^{\frac{-tR}{L}}\right)$$

$$i_L = \frac{50}{100}\left(1-\epsilon^{\frac{-200\times10^{-6}\times10^{2}}{10^{-2}}}\right)$$

$$i_L = .5\,(1-\epsilon^{-2})$$

From tables, $\epsilon^{-2} = .135$

Therefore, $i_L = .5\,(1-.135)$

$$i_L = .5\,(.865)$$

$$i_L = .4325 \text{ amperes}$$

Problem 2. Find e_R *when* $t=50$ *microseconds.*

Solution

$$e_R = E_B\left(1-\epsilon^{\frac{-tR}{L}}\right)$$

$$e_R = 50\left(1-\epsilon^{\frac{-50\times10^{-6}\times10^{2}}{10^{2}}}\right)$$

$$e_R = 50\,(1-\epsilon^{-.5})$$

$$\epsilon^{-.5} = .607 \text{ (from tables)}$$

$$e_R = 50\,(1-.607)$$

$$e_R = 19.6 \text{ volts}$$

Problem 3. Find e_L *when* $t=50$ *microseconds.*

Solution

$$E_B = e_R + e_L$$

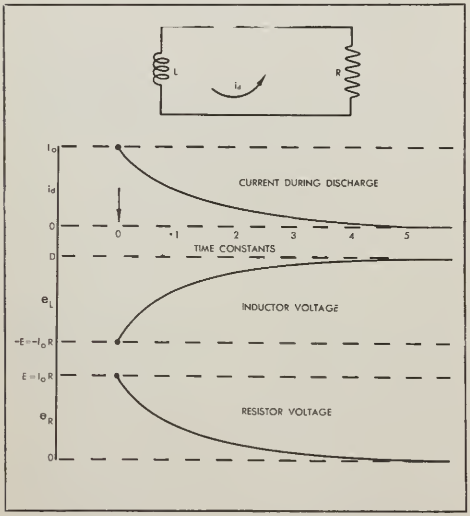
$$e_L = E_B - e_R$$

$$= 50 - 19.6$$

$$= 30.4 \text{ volts}$$

LR Discharging

The curves shown in the illustration below are formed when the source of energy is removed from a series LR circuit, and the energy stored in the magnetic field is returned to the circuit by the collapse of the field. At the initial instant when you open the switch, the resistor voltage e_R attempts to decrease to zero, but the inductance of the inductor resists any change in current and tends to maintain the current flow. This action of the inductor occurs because opening the switch removes the applied voltage, causes the current flow in the circuit to tend to cease, and starts the collapse of the magnetic field about the inductor. This collapse induces a voltage in the inductor, which causes the current to flow in the same direction that it did when the circuit was charged. Now the self induced voltage is the source voltage in the circuit and its polarity is opposite. As the discharge current, i_d, begins to decrease, the voltage across R decreases proportionately.



Voltage and Current Curves During Collapse of Inductor Field

Since $e_R+e_L=0$, as previously mentioned in LR charging, the value of e_L also decreases. This change continues as shown in the e_R and e_L curves until the e_R and e_L voltages are both zero and no current is flowing. The discharge current at the end of one time constant is 36.8% of maximum.

LR DISCHARGING PROBLEMS. Study the following sample problems to learn how to find the instantaneous values of current and voltage in LR discharging circuits. The methods employed are like those used previously in RC discharge circuits. The formula for the LR time constant previously discussed applies here too.

Examples

SOLVING BY USE OF CURVES.

Problem 1. Find the current 150 microseconds after the current has started to decrease in the next circuit.

Solution

Finding the time constant,

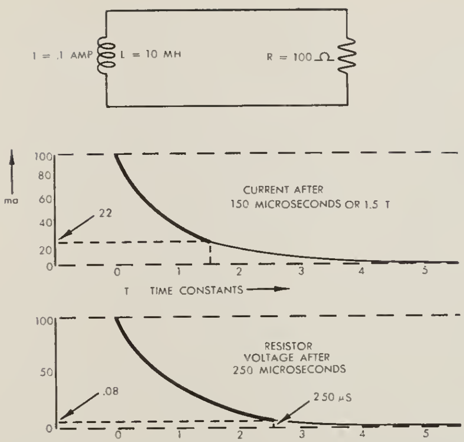
$$T = \frac{L}{R} = \frac{.01}{100} = 10^{-4} = 100 \text{ microseconds}$$

Since the time involved is 150 microseconds, the number of time constants is 1.5. Then, since the current starts high and decreases to zero, refer to curve B in universal time constant chart. This curve shows that the current drops to 22% of its initial value after 1.5 time constants. The initial current is 100 ma; therefore, the current after 150 microseconds is equal to $100\times.22$*, or 22 ma.*

Problem 2. Using the same circuit, find the resistor and inductor voltages 250 microseconds after the current starts decreasing.

Solution

The time given, 250 microseconds, constitutes 2.5 time constants. The universal time constant chart shows



that after 2.5 time constants, the voltage drops to 8% of its initial value. Therefore, since the maximum voltage, $E_R = IR = .1$ ma $\times 100$ ohms, or 10 volts, the voltage e_R, after 2.5 time constants, equals $10 \times .08$, or .8 volts.

In finding the inductor voltage, utilize the fact that the inductor voltage is the source voltage in an LR discharging circuit and, at all times, is equal to the voltage across the resistor. Therefore, since the resistor voltage, e_R, is .8 volts after 2.5 time constants, the inductor voltage, e_L, equals .8 volts.

Solving by use of exponential equations.

Use the following exponential equations for solving discharge LR circuit problems:

1. *Instantaneous current:* $i_d = I_o\, \epsilon^{\frac{-tR}{L}}$
2. *Instantaneous inductor voltage:* $e_L = I_o R\, \epsilon^{\frac{-tR}{L}}$
3. *Instantaneous resistor voltage:* $e_R = I_o R\, \epsilon^{\frac{-tR}{L}}$

The symbols in these equations stand for the following:

i_d = *Current at any time t during discharge.*

I_o = *Initial current at start of discharge.*

e_L = *Instantaneous voltage across inductor.*

e_R = *Instantaneous voltage across resistor.*

The other symbols have the same meaning as in exponential equations previously given.

Problem 1. Using the circuit shown with the preceding sample problems, find e_R when t is 250 microseconds.

Solution

$$e_R = I_o R \epsilon^{-tR\ L}$$

$$e_R = 10^{-1} \times 10^2\ \epsilon^{\frac{-250 \times 10^{-6} \times 10^2}{10^{-2}}}$$

$$e_R = 10^{-1} \times 10^2 \times \epsilon^{-2.5}$$

$$e_R = 10\ \epsilon^{-2.5}$$

$$\epsilon^{-2.5} = 0.0821 \text{ (from table)}$$

$$e_R = 10 \times .0821$$

$$e_R = .821 \text{ volts.}$$

Problem 2. Using the same circuit, find out when the voltage across the inductor will equal 3 volts.

Solution

When working a problem of this type, first make a rough estimate of the result, in order to check the answer you obtain by the equation. For example, in this sample problem, the initial voltage e_L is I_oR, or $.1 \times 100$, or 10 volts. In one time constant, the inductor voltage in an RL discharge circuit drops to 37% of its initial value. Three volts is 30% of 10 volts and one time constant equals 100 microseconds. Therefore, the voltage in this problem must drop to 3 volts in slightly more than one time constant, that is, 100 microseconds.

Use the equation, $e_L = I_o R\ \epsilon^{\frac{-tR}{L}}$

Dividing by I_o R gives: $\frac{e_L}{I_o R} = \epsilon^{\frac{-tR}{L}}$

In this equation substituting the values given for R and L in the circuit

$$\frac{3}{10^{-1} \times 10^2} = \epsilon^{\frac{-t \times 100}{10^{-2}}}$$

$$\frac{3}{10} = \epsilon^{\frac{-t \times 100}{10^{-2}}}$$

$$.3 = \epsilon^{\frac{-t \times 10^2}{10^{-2}}}$$

$$.3 = \epsilon^{-t \times 10^4}$$

In the tables, $\epsilon^{-x} = .3$ *when* $-x = -1.2$; *therefore*

$$\epsilon^{-t \times 10^4} = \epsilon^{-1.2}$$

$$-t \times 10^4 = -1.2$$

$$t = 1.2 \times 10^4$$

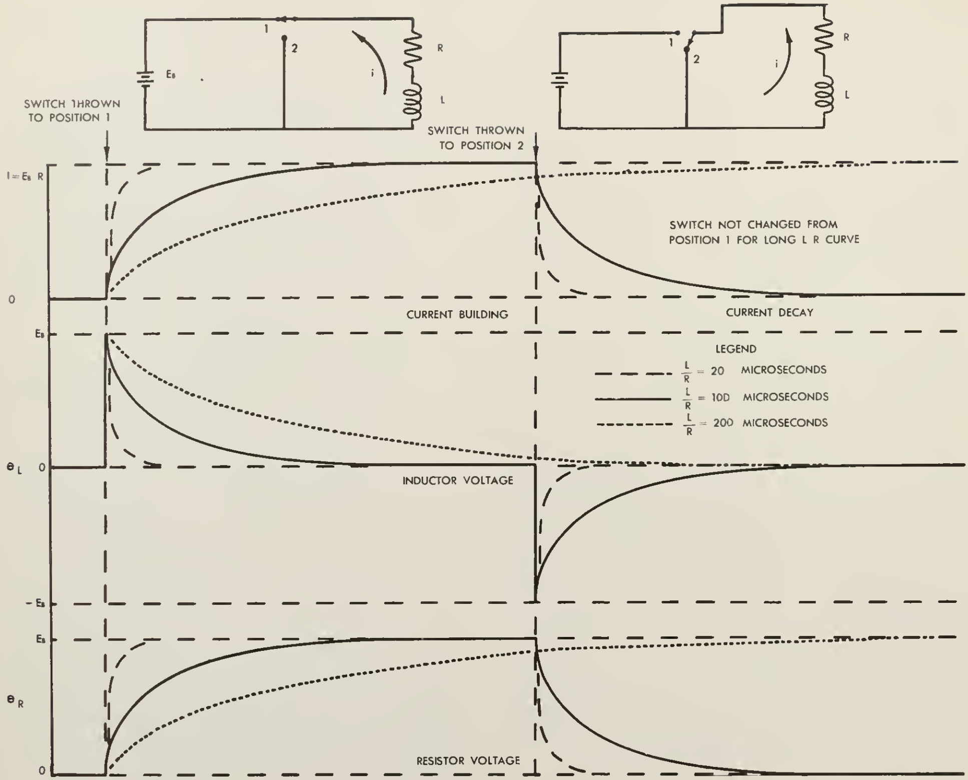
$$t = 120 \text{ microseconds}$$

Since one time constant equals 100 microseconds, 1.2 x 100 or 120 microseconds are required for the inductor voltage to drop from 10 volts to 3 volts.

Effect of Varying R and L

As shown on page 6-20, varying the size of R and L in an LR circuit changes the charge and discharge curves. The changes result from the fact that any change of either R or L changes the LR time constant. The solid-line curves shown represent the instantaneous values of current and voltage during charge and discharge when the time constant is 100 microseconds.

Decreasing the time constant sharpens the slope of both the charge and discharge curves. Notice, for example, the dashed-line curve. It represents the effect of reducing the time constant in the circuit to 20 microseconds. A time constant may be reduced either by decreasing L or by increasing R. Decreasing L decreases the opposition to current flow change and permits current to build up and decay quickly. Increasing R does not affect the inductance, but it does reduce the maximum current flow. Since

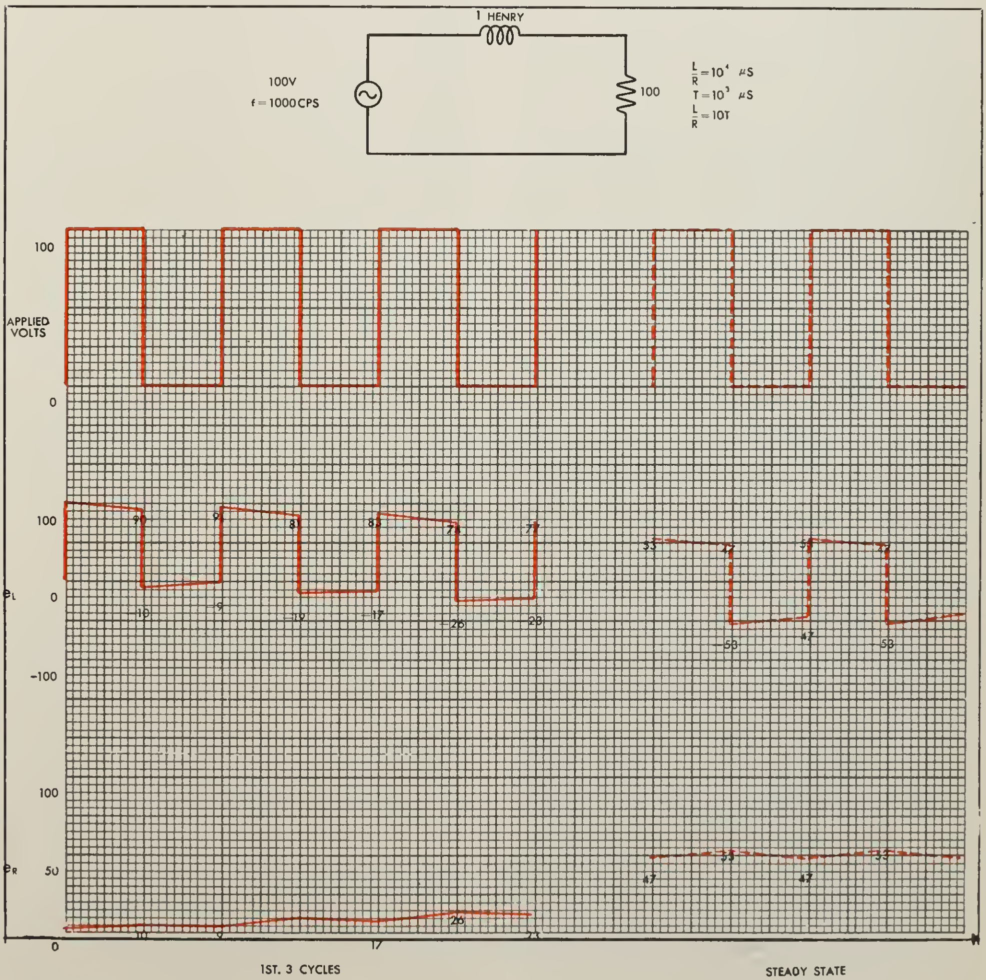


Effect of Varying R and L

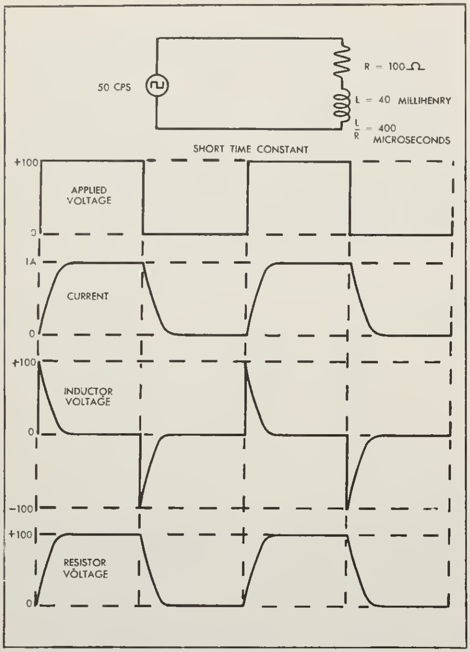
the maximum value is lower, the current reaches its maximum value more quickly. Because the current builds up very rapidly in a circuit with a short time constant, the voltage across the resistor (e_R), which keeps in step with the current, similarly builds up rapidly. The inductor voltage (e_L), which drops just as fast as the resistor voltage increases, likewise forms a sharp peak. When current is decaying, the resistor voltage and the inductor voltage similarly decay much more rapidly than in a circuit with a longer time constant.

Increasing a time constant produces curves with gradual slopes. For example, notice the dotted-line curve. It represents the curve formed by a 200 microsecond time constant. The rise of current is slower than before because the inductor furnishes more opposition to the flow of current. The effect is noticeable also in the other 200 microsecond curves, where the fall of voltage across the inductor is very slow, and the rise of voltage across the resistor requires appreciably more time than with a 100 microsecond time constant. Similarly, the decay of current takes longer since a large inductor stores more energy than a small one, and consumes more time in dissipating it through the resistor. If the time constant were lengthened by decreasing R, the same relationships would hold true, because the smaller resistor would dissipate the energy stored in the inductor more slowly. You can see why by analyzing the basic concepts, energy dissipated is power, and power is I^2R. Therefore, if the current (I) is constant and if the resistance (R) is decreased, the power dissipation is less at any single instant. This means that if the rate of dissipation is decreased by using a smaller resistor, a longer time is required to dissipate the energy stored in the inductor.

Long LR time constant. The waveshapes illustrated below are those produced when a square wave is applied to an LR circuit with a long time constant. The time constant used is extremely long, being ten times the period. One important fact to observe here, other than the waveshapes, is how the inductor voltages change polarities during the time that the inductor acts as the source of voltage in the circuit. During current build-up, the end of the inductor that current enters may be called the negative end, for during this time the inductor acts just like any resistance or, for that matter, like any other load. However, when the source of voltage is removed, the inductor becomes a source, and like a battery, the current leaves the negative end and enters the positive end. Since current flows in the same direction during either build-up or decay, the polarity of voltage actually reverses at each half cycle.



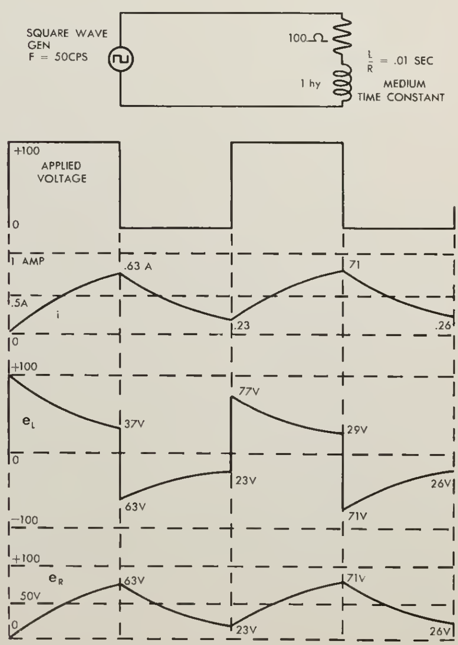
Effect of Long LR Time Constant



Effect of Short LR Time Constant

Short LR time constant. The waveshapes illustrated above are due largely to using a smaller inductor. The small inductor allows the current to reach a maximum value very quickly and thus causes the resistor voltage to be nearly a replica of the input voltage. Another result is that the inductor voltage is differentiated into a sharp peaked wave.

Medium LR time constant. The waveshapes produced by applying a square wave to a medium LR time constant circuit with the values indicated are shown at the right. The time constant employed is equal to the time duration of one-half cycle of the input square wave. Therefore, the current reaches .63 amperes, or 63% of its maximum charge value during the first half cycle. During the second half cycle, when the applied voltage is zero and the inductor furnishes the current, the current drops to .23 amperes or 37% of .63 amperes. Since the applied voltage varies between 0 and 100 volts, there exists a 50-volt DC component in the waveshape. This DC component causes the current during charge and discharge to vary about an average value of 50 volts divided by 100 ohms, or .5 amperes. However, because of the inductance in the circuit, the current does not reach this average value immediately but approaches it gradually. It does this by starting each current half-cycle a little higher. For example, the average value in the second half cycle is .42 amperes; the current range is .63 amperes to .23 amperes. In the third half cycle it is .47, the range being .23 to .71, and in the fourth, the average is about .49 with a range of .71 to .26 amperes. During the first positive half-cycle, the inductor voltages are entirely positive. With succeeding cycles of the input voltage, these voltages work down until they center around the zero voltage line; that is, about as much of the inductor voltage is above zero as is below. The e_R voltage wave increases in value with each half cycle of the applied voltage thus indicating that a DC voltage component exists in the square wave. Notice that at the end of the first half cycle the resistor voltage is 63 volts and at the beginning of the second half cycle it is 71 volts. After 5 half cycles of the input voltage, the waveshape repeats at the values in the last half cycle shown.



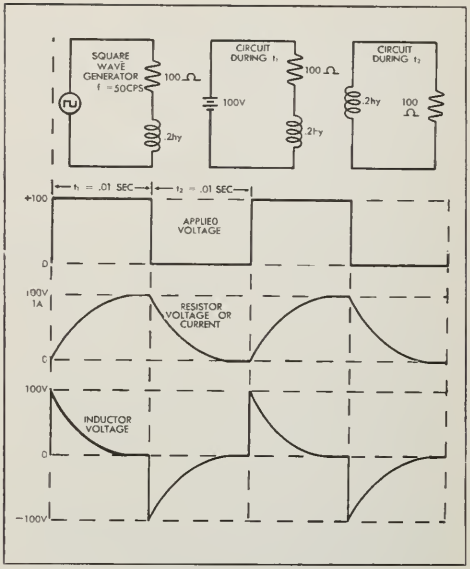
Effect of Medium LR Time Constant

Frequently, the applied voltage does not contain a DC component at all, but is a pure square wave. In that case waves like those shown in the bottom illustration are produced. The current and voltage waves produced when the pure square wave is applied to the LR circuit shown are identical to those produced when the square wave has a DC component, but the resistor voltages are 50 volts less than their original value.

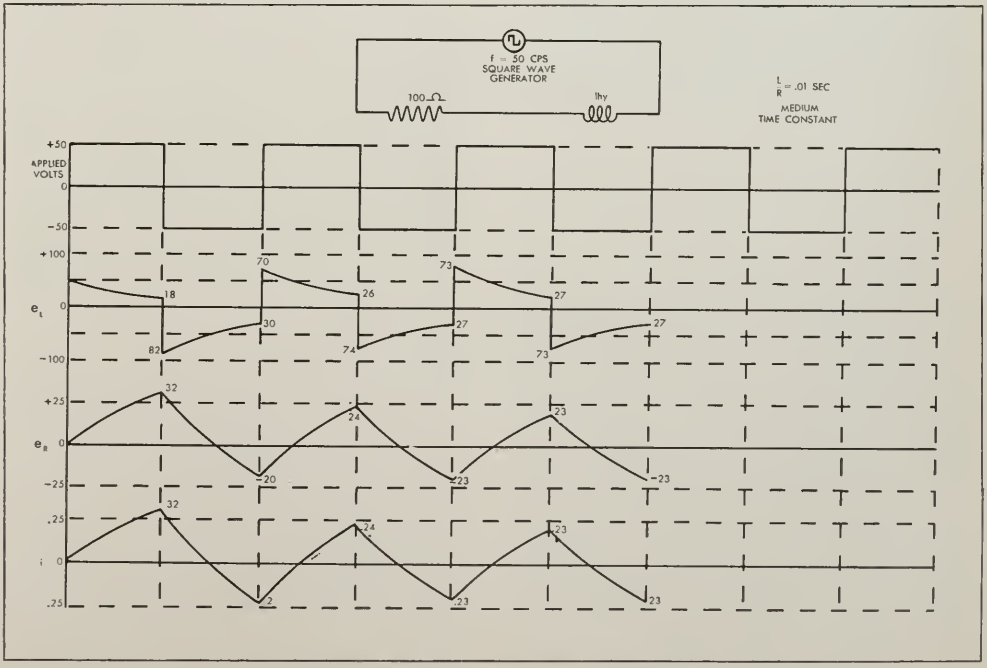
Quantitative Values of Current and Voltage

An easy means of determining the voltage or current any time a square wave is applied to an LR circuit is to reduce the circuit to a simple DC circuit. For example, the adjacent LR circuit can be represented by two equivalent circuit—one (t_1) representing the .01 seconds of the square wave when 100 volts is applied, and the other, (t_2) the .01 seconds when zero voltage is applied.

During t_1 the circuit may be reduced to an LR circuit in which the generator is replaced by a battery with a voltage equal to the generator voltage. This is proper since the current rises



Equivalence in LR Circuit



LR Waves When Input Contains no DC Component from zero when a battery is used just like it does when the square wave voltage is applied. Therefore, if you know the time the voltage is applied and the values of L and R, you can compute the value of current at any time.

During time t_2 you can reduce the LR circuits shown to a circuit containing a resistor and an inductor in which current is flowing.

To understand how to apply these equivalent circuits in finding either current or voltage at a fixed time, study the following example:

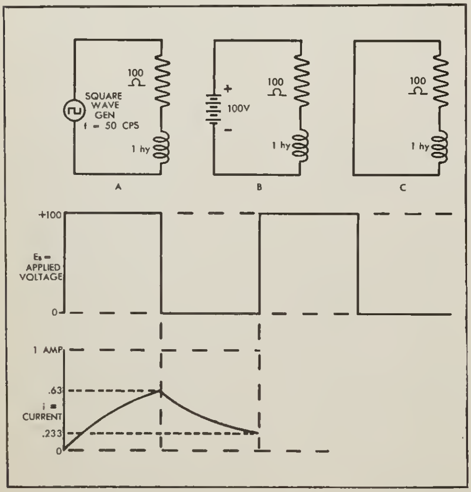
Problem. In the circuit below, find the current at the end of each of the first two time constants.

Solution

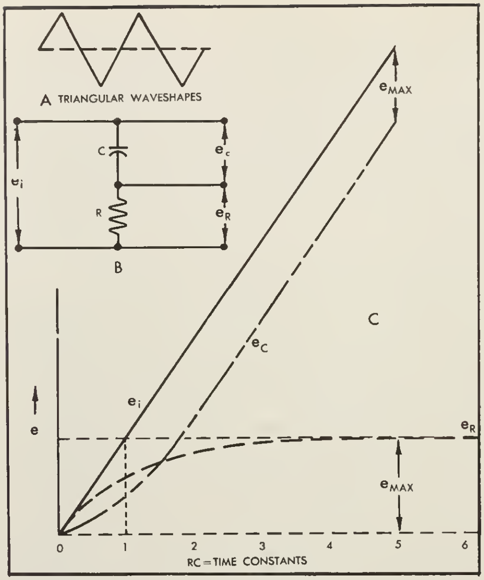
When the square wave generator is producing the positive half cycle, it applies a constant 100 volts to the circuit at A. (Diagram B shows the equivalent circuit.)

During this time current rises toward its maximum $\frac{E}{R}$ *value of 1 ampere. Since the time constant is .01 seconds, and since the time of a half cycle is also .01 second, the 100 volts is applied for one time constant. Therefore, at the end of one time constant, the current equals 63% of maximum, or .63 amperes.*

Diagram C shows the circuit condition when the applied voltage is zero. Initially the current is .63 amperes, as just determined. The time of this half-cycle is also .01 seconds, and likewise one time constant. Since current in an LR circuit drops to 37% of its initial value in one time constant, it here equals 37% of .63 amperes or .233 amperes at the end of the second time constant.



Circuit with Voltage and Current Curves



Response of RC Circuit to Triangular Wave Input

TRIANGULAR WAVES IN RC AND LR CIRCUITS

All the voltages thus far described in connection with RC and LR circuits were voltages of sudden changes, such as a voltage which changes to full magnitude when a switch is closed or such as the voltages of the steep wavefront of a square wave. However, not all voltages applied to RC and LR circuits are sudden change voltages. For example, triangular or sawtooth waveshape voltages, such as those shown at A, are commonly used in radar circuits. It is important that you know the effect that these waveshapes have on long and short time constant RC and LR circuits.

Response to Triangular Wave Input

For ease of understanding the effect of triangular waves in LR and RC circuits, first examine the first rise in voltage when this wave is applied to the RC circuit at B. Diagram C shows the voltage wave enlarged several times for easier study. Notice that the voltage starts at zero and keeps on increasing at a constant rate. The triangular waveshape is thus different than that of a square wave, in which the voltage rises quickly, then levels off. Likewise, the effect of a triangular wave on the condenser is different than that of a square wave. With a square wave, the condenser charges exponentially to the value at which the square wave levels off, but with a triangular wave the voltage increases as the condenser charges toward it (note the triangular shape of the wave), and the condenser must continually charge toward a new higher value. Initially when the voltage e_i is applied to the circuit at B, the current rises with the voltage because the condenser is initially uncharged. This current stacks electrons on the condenser plate, and it charges to a voltage that opposes the applied voltage. This should decrease the current, but since the input voltage is continually increasing, the opposing voltage is more than overcome by the applied voltage, and the current actually increases, but less rapidly than before. As the current increase continues, the condenser becomes charged to a higher voltage and the rate of current increase falls off some more until finally the current is high enough to raise the condenser voltage at the same rate the input voltage waveshape increases. At this point, the current becomes constant, and the condenser voltage slope is essentially the same as the input voltage slope. At this time the constant current through the resistor causes a constant resistor voltage.

Since current increases from zero to a constant value in an RC circuit, its rate of increase is exponential. You can use the curve for e_r as the current curve. Since both the current and voltage are exponential, the current and e_r at the end of one time constant will be 63.2% of the voltage applied at that time. After 5 time constants, the current will level off to a value equal to the voltage applied at the end of the first time constant—that is, at the value e_{max} shown in the illustration.

The maximum value of resistor voltage (e_{max}) is a function of the time constant and the rate of rise. It is expressed by the equation,

$$e_{max} = \frac{de_i}{dt} RC$$

In this equation, e_{max} *= Maximum resistor voltage*

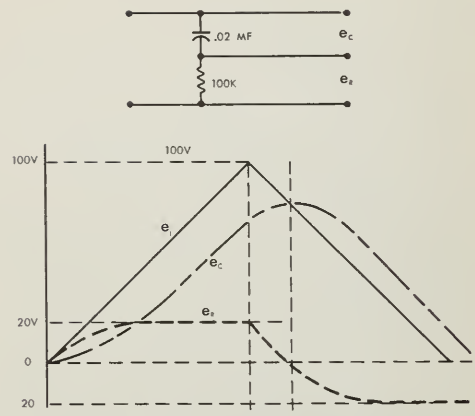
R = Resistance in ohms

C = Capacity in farads

$\frac{de_i}{dt}$ *= Rate of change of input voltage in volts per sec. (The symbol d indicates the amount of change.)*

Example

Problem. Assuming that the triangular waveshape illustrated has a frequency of 50 cps and is applied to the circuit above, find the maximum resistor voltage.



Condenser and Resistor Voltages with Triangular Wave Input

Solution

Inspection shows that the change of the input voltage e_i *is 100 volts during one-half cycle. You can represent this change in input voltage as* de_i*. Since the frequency is 50 cps, one cycle is* $\frac{1}{50}$ *or .02 second. Therefore the half cycle is .01 second long. You can represent the change in time as dt. Therefore, dt, during the 100 volt rise is .01 second.*

Substituting in equation,

$$e_{max} = \frac{de_i}{dt} RC$$

$$= \frac{100 \times 10^5 \times .02 \times 10^{-6}}{.01}$$

$$= \frac{10^2 \times 10^5 \times 2 \times 10^{-2} \times 10^{-6}}{10^{-2}}$$

$$= 2 \times 10$$

= 20 volts.

The condenser continues to charge at the same rate until the rise of input voltage ceases. As the voltage abruptly starts decreasing during the second half of the cycle, the charging rate of the condenser decreases until the voltage drops to the same value as the condenser voltage. At this point, the current is zero. As the applied voltage continues to decrease, it falls below the condenser voltage, and the condenser starts discharging. This reverses the current, causing a negative voltage across the resistor. Again, the change-over cannot occur any faster than the condenser can discharge, so the current, and consequently the resistor voltage change is exponential. The decrease in voltage is at the same rate (de_i/dt); so the negative voltage is also 20 volts.

Effect of Varying the Time Constant Ratio

The effect of varying the time constant is indicated by the illustration. Diagram A shows the waveshape of the previous problem repeated over several cycles. As previously mentioned in connection with square waveshapes, the condenser waveshape resembles the input when the RC is short and the resistor waveshape resembles it when the RC is long. This is true for triangular waveshapes also. In diagram B, the RC is twice the time of the applied voltage; so the resistor waveshape very closely follows the input. If the RC is shortened to equal the time the voltage is increasing, the resistor voltage reaches 63% of the applied voltage after one time constant. However, if the RC is shortened to one-fifth of the time the voltage is increasing, the condenser charging rate quickly reaches the rate of applied voltage, after which the resistor voltage is constant. Notice in diagram A that this condition causes the resistor voltage to approach a square wave. This is differentiation, which was described earlier. When a triangular wave is integrated by use of a long time constant, the condenser waveshapes will resemble a hyperbolic waveshape (see diagram B).

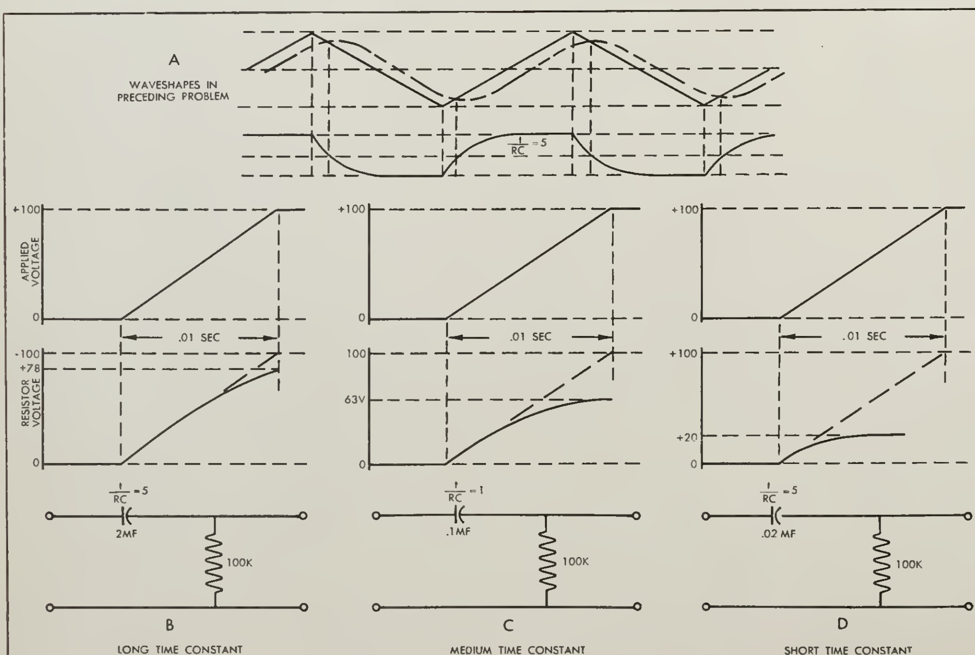
An LR circuit has the same effect on triangular waveshapes as an RC circuit provided you observe the following changes:

First, the resistor waveshape of the RC circuit occurs across the inductor while in the LR circuit the condenser waveshape occurs across the resistor. Second, the time constant is L/R instead of R x C. Third, the term de_i/dt must be replaced by di/dt to obtain the final voltage across the inductor.

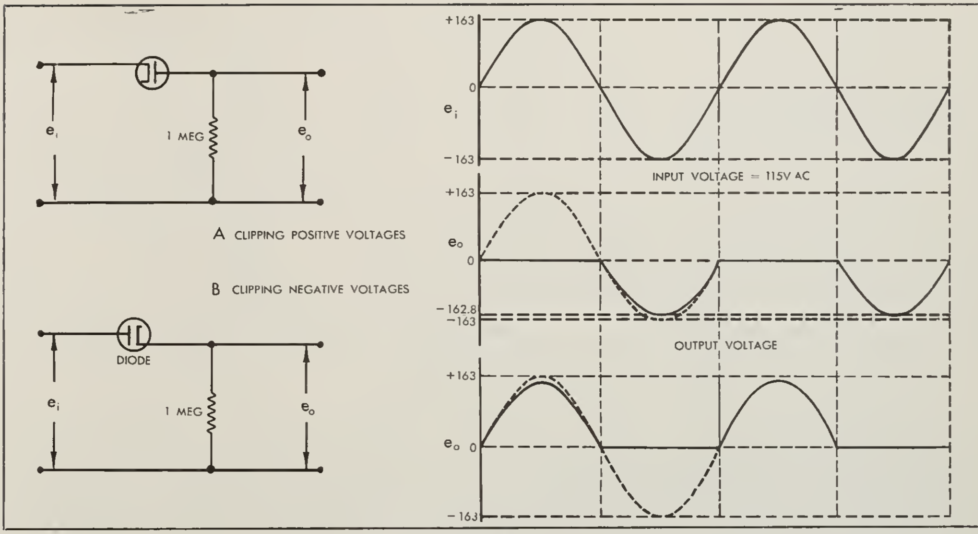
When a triangular voltage waveshape is applied to an LR circuit, the current increases at a constant rate. This, in turn, causes the inductor voltage to become constant after 5 time constants. You can express the maximum inductor voltage by the equation,

$$E_L = di/dt \frac{L}{R}$$

where E_L is the maximum inductor voltage, di/dt the rate of change of current, L the inductance in henrys, and R the resistance in ohms.



Varying the Time Constant with Triangular Wave Input



Series Diode Limiters

LIMITING CIRCUITS

Limiting circuits are circuits which remove either one extremity or the other of an input wave. Tubes which perform this function are referred to as *limiters* or *clippers*.

Limiters are useful in a variety of ways. They are applicable in waveshaping circuits where it is desired to square off the extremities of the input signal. A sine wave can be converted into a rectangular wave by a limiter circuit. A peaked wave may be applied to a limiter to eliminate either the positive or the negative peaks from the output. Limiters are commonly used to prevent a voltage from swinging too far in either the positive or negative directions.

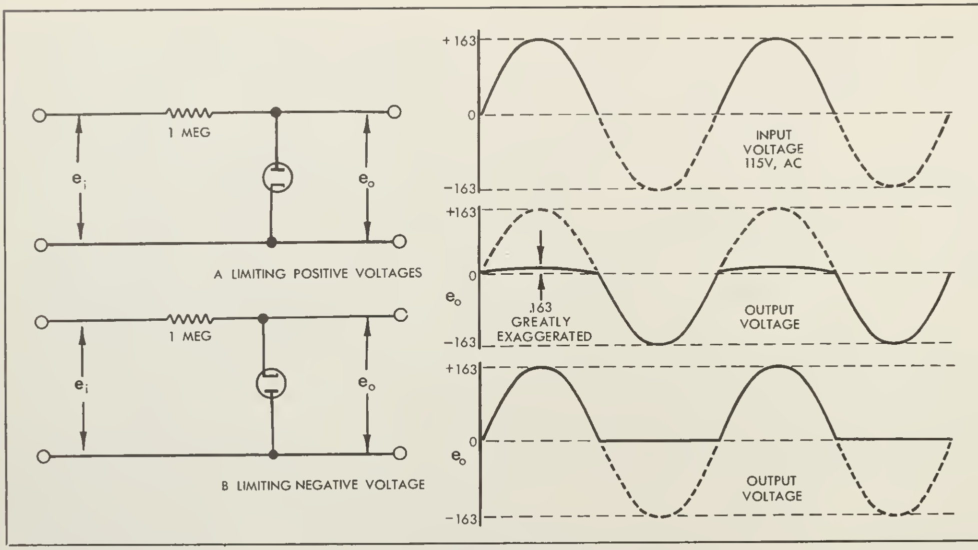
Series Diode Limiters

Diodes are very useful for limiting since they conduct current only when the plate is positive with respect to the cathode. Diagram A shows a series-connected diode (one in which load is connected in series with the tube) that is used to limit the positive input cycle of a sine wave. The input voltage is e_i and the output is e_o. When any input voltage, say 115 volts (163 volts peak value), is applied to the input terminals of the diode, the output follows the input only when the negative input cycle is applied.

When the positive cycle is applied, the cathode becomes positive with respect to the plate and consequently the tube can not conduct current. Since there is no current flow, there is no voltage developed across the 1-megohm output resistor. Therefore, the positive input cycle is limited to zero in the output.

When the negative input cycle is applied, the plate becomes positive with respect to the cathode and, since the plate is at the proper potential, it conducts current. This current flows through the circuit, developing a voltage across the internal resistance of the tube and the load resistor. These two resistances act as a voltage divider and divide the applied voltage, with $\frac{1}{1000}$ of it appearing across the tube (assuming the internal resistance of the tube is 1000 ohms), and with $\frac{999}{1000}$ of it across the load resistor. Since the applied voltage is 163 volts peak value, about .163 volts appears across the tube, and the remainder, or about 162 volts, across the load resistor. Thus, except for the small drop across the tube, the output is nearly equal to the positive input. Thus, the diode has limited or clipped off the positive input cycle, and has virtually reproduced the negative input cycle both in shape and magnitude.

When the input terminals are reversed as shown at diagram B, the diode limits the nega-



Shunt Diode Limiters

tive cycle of the input sine wave because during that cycle the plate is negative with reference to the cathode. During positive cycles, the tube can conduct because its plate is positive with respect to the cathode, and it develops a voltage across the resistor which follows the positive cycle of the input and is equal to it except for the small drop across the tube itself.

Shunt Diode Limiters

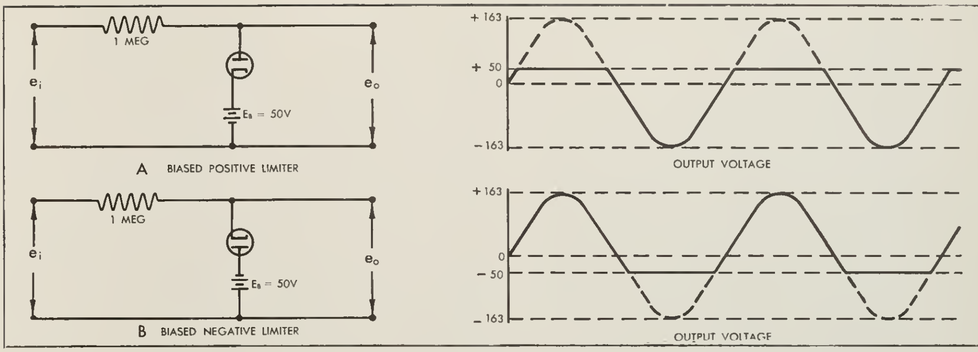
Another method of using a diode as a limiter is connecting it in parallel with the load resistor. In diagram A, the diode is connected to limit the positive input cycle. When the positive cycle of the input voltage is applied to the plate, the diode can conduct. Therefore, as in the case of the series diode limiter, the current flows through a voltage divider consisting of the one megohm resistor and the internal plate resistance. This voltage divider divides the applied voltage at a ratio of one thousand to one million. One thousandth of the applied voltage, or .163 volts appears across the tube. The remainder, about 162 volts, appears across the resistor. However, since the output terminals here are across the tube, the output is only .163 volts. Thus, during the time when the input cycle is positive, the output is limited, or clipped, to practically zero voltage.

During the time the negative cycle of the input voltage is applied to the tube, the plate is negative with respect to the cathode and current does not flow. Therefore there is no voltage drop across the resistor. However, since the tube appears as an open circuit, the applied voltage appears across its output terminals. Thus, the output voltage follows and is equal to the input voltage during the time the negative cycle is applied to the tube.

The negative input cycle may be removed by connecting the tube as shown at diagram B. When connected in this manner, the tube conducts when the cathode is negative and virtually all the applied voltage appears across the resistor and a very small amount across the tube. Therefore, since the output is taken from across the tube, the negative input cycle has been practically reduced or limited to zero.

The input voltage can be limited to some value other than zero by maintaining the plate or cathode at that voltage by means of a battery or a biasing resistor. The two limiting circuits at the top of the next page—one for limiting the positive input cycle and the other for limiting the negative input cycle—both employ a battery to supply the biasing voltage.

Circuit A is designed to limit the swing of the positive input cycle to +50 volts. The input

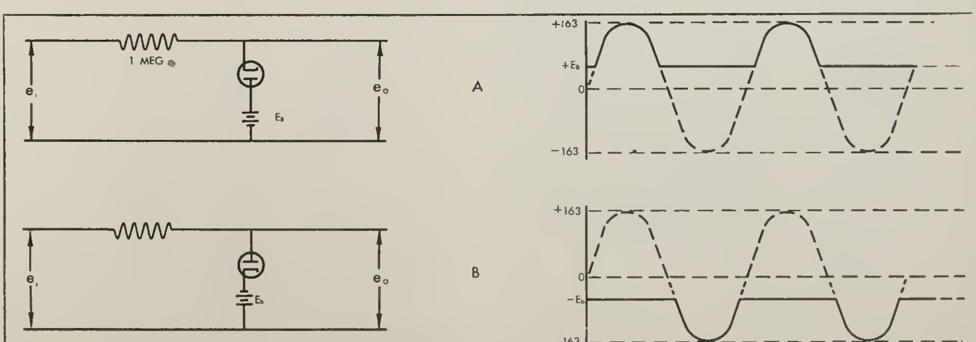


Limiting to Other than Zero

voltage e_i is 115 volts effective value and 163 volts peak value. The battery connected in the cathode circuit maintains the cathode 50 volts positive with respect to the plate when the tube is not conducting. As long as the input voltage is less than 50 volts, the tube does not conduct, but just as soon as the input exceeds this amount, current starts to flow and effectively connects the upper output terminal of the circuit to the positive terminal of the battery. Therefore during the portion of the positive input cycle when the voltage exceeds the 50 volts on the cathode, the output voltage equals 50 volts, and the difference between the input voltage and this voltage (neglecting the drop across the internal resistance of the tube) appears across the 1 megohm resistor.

In diagram B, the battery E_B is so connected in the diode plate circuit as to make the plate 50 volts negative with respect to the cathode. As long as the input is positive or less negative than E_B the diode acts like an open circuit, and the voltage across the output is equal to the input. When the input becomes more negative than E_B, that is, when the cathode is negative with respect to the plate, the diode conducts current and connects the upper output terminal to the negative terminal of the battery. Therefore, during this part of the input cycle—50 volts appears in the output and the rest of the 163 volts input, neglecting the small drop across the tube, appears across the 1 megohm resistor.

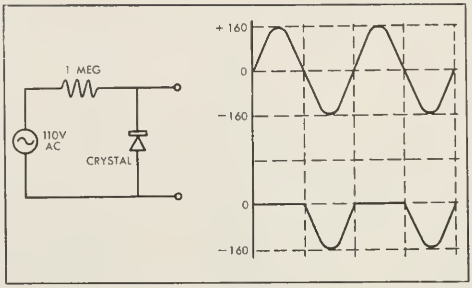
Shunt diodes may be also used to limit the amount to which the input voltage can drop. In other words, only the peaks of waveform are reproduced in the output. In circuit A below, the diode conducts during the entire part of the input waveform that is below the positive voltage of the battery. The output voltage under



Maintaining Voltage above Certain Level

this condition varies between the positive level of the battery voltage and the positive extremity of the input waveform.

In circuit B the entire portion of the input waveform above the negative potential of the battery causes the diode to conduct, thus producing an output voltage which varies between the negative level of E_B and the negative extremity of the input. In both cases, circuits A and B, the difference between the value of E_B and the applied voltage, during the time the diode conducts, is represented by the voltage drop across the 1 megohm resistor.



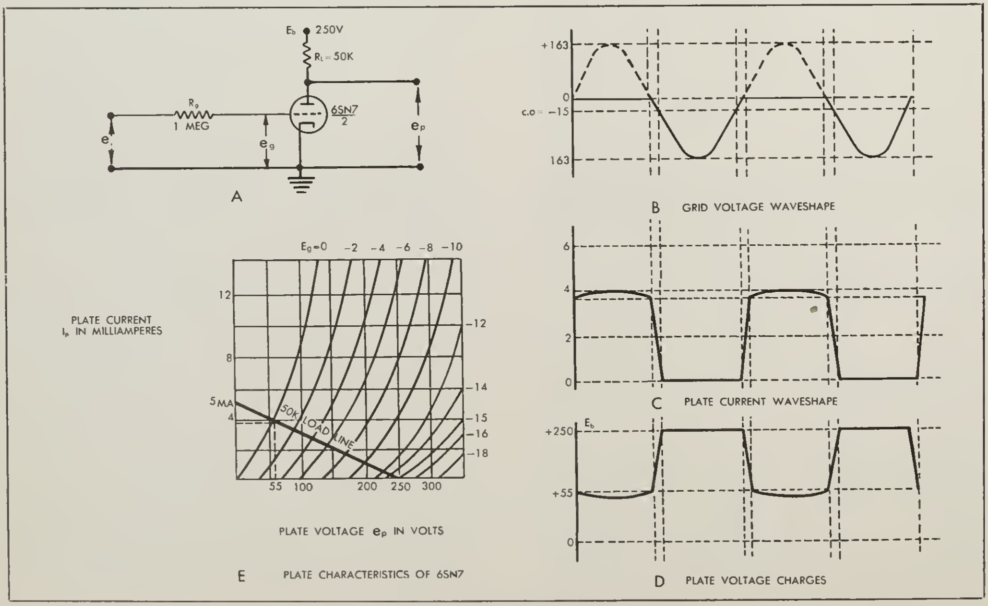
Positive Crystal Limiter

Crystal Diode Limiter

Although most radar circuits use diode vacuum tubes in limiter circuits, the crystal diode or silicon crystal rectifier is becoming increasingly popular for limiting. The crystal diode usually consists of a small piece of silicon and a pointed tungsten wire contactor. It operates by virtue of the fact that it passes little current from the crystal to the wire, because of high resistance, but permits current to flow readily from the wire to the crystal when an AC voltage is applied to its input terminals. Since, like a diode vacuum tube, it conducts more current in one direction than in the other, the crystal assembly often is called a crystal diode. Not only is the crystal almost as efficient as the vacuum tube, but requires no heater voltage and is entirely free of interference from the power frequency. It may be used in any of the diode limiter circuits previously described. The one illustrated is for use in limiting or clipping the positive cycle of the input waveform.

Grid Current Limiting

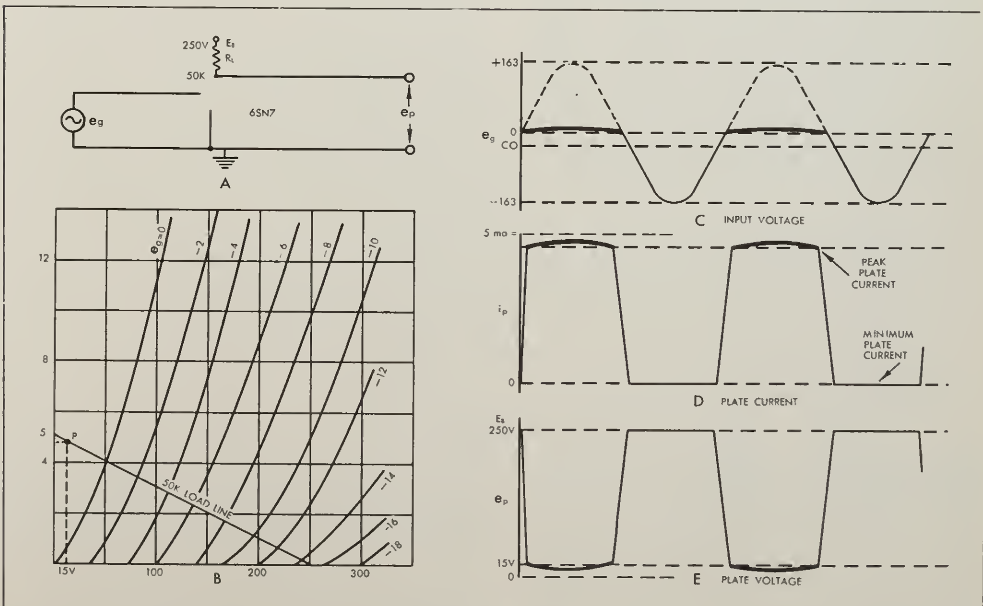
When a large limiting resistor is connected in series with the grid of a triode, tetrode, or pen-



Grid Current and Grid Cutoff Limiting

tode tube, the grid-cathode circuit may be used as a limiter circuit in exactly the same manner as the plate-cathode circuit of the shunt-diode limiter. What this means is that when the grid is positive with respect to the cathode the grid will attract electrons just as the plate in a diode when it is positive with respect to the cathode. The limiter circuit on page 6-30 is held normally at zero bias. During the positive portion of the input signal, the grid tries to swing positive. Grid current flows through the resistor R_g, developing an IR drop across it with a polarity which opposes the positive input voltage. Since the input voltage must equal the sum of the drop across R_g and the grid to cathode voltage, the larger R_g is with respect to the cathode-grid resistance, the nearer the voltage on the grid is limited to that of the cathode. In the circuit shown, the grid is never driven more than a fraction of a volt positive by the positive input cycle. This means that all voltages in the positive direction of the input signal are leveled off at zero.

If the input voltage is great enough in the negative direction to exceed the grid voltage required to cut off plate current, limiting occurs. The numerical values are inserted to enable you to calculate the output. As before, the input voltage is a 115-volt sine wave and as previously mentioned the 1-megohm limiting resistor prevents the grid voltage from exceeding more than a fraction of a volt in the positive direction. The grid voltage, however, can swing the full distance in the negative direction since no current flows from grid to cathode to reduce the applied voltage as when the positive signal is applied. To observe the effect of the grid voltage variation, study the load line on the characteristic curves. When the grid voltage is zero, the plate current is 4.0 ma and the instantaneous plate-to-cathode voltage e_p is 55 volts. As the voltage swings in the negative direction, the plate current decreases and when the grid is −15 volts the plate current is zero. The plate current cannot decrease below zero as the grid is made more negative, and therefore both the plate current and its resultant plate voltage becomes constant. This condition prevails until the grid voltage again becomes less than −15 volts. With the applied waveshape indicated, the grid voltage exceeds −15 volts for most of the half cycle. The plate voltage



Saturation Limiting

variation shows a rapid change to the maximum voltage of 250 volts, and remains at that value for most of the half cycle. The plate voltage change is essentially a square wave except for a slight rounding which occurs because the grid draws current when it is driven positive a fraction of a volt. This voltage change is amplified in the plate circuit and rounds off the square wave.

The action just described is perhaps more correctly defined as squaring the waveshape since the plate waveshape rather than the input waveshape is limited.

Saturation Limiting

A grid limiting resistor is not absolutely necessary for limiting the positive half of an input cycle. If the grid is made positive by an input voltage, there is a saturation point beyond which the plate current will not increase. This point is indicated by a dot on the load line at P in the illustration on page 6-31. That is the point at which the plate resistance of the tube is assumed to be zero. The maximum plate current obtainable then depends on the size of the plate load resistor. Actually, there is always some resistance to electron flow in the tube, no matter how positive the grid becomes, and the current will stop increasing below the saturation value. Limiting occurs when plate current and plate voltages stop changing. This occurs at some positive value of grid voltage. Therefore saturation limiting occurs at a higher point on the grid voltage swing than when a limiting resistor is used. The curves also show limiting by plate current cutoff.

Double Diode Limiting

When a pair of biased diodes are connected as below, both the negative and positive half cycles of the input voltage can be limited, leaving a waveshape that is fairly square. The voltage at which limiting occurs in each direction may be varied by varying the bias voltage on the diode. In the circuit shown, tube V_1 limits the positive half cycle and V_2 limits the negative half cycle.

The Squarer and Peaker Circuit

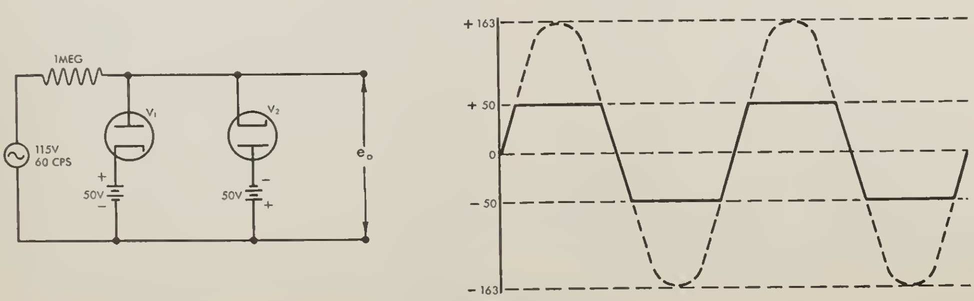
The squarer and peaker circuit uses the principles discussed in the preceding limiter circuits. Each part of this circuit is used in several radar circuits, and the entire circuit is used in some radar sets.

Notice the simple squarer and peaker circuit at A on the next page. The first tube is an overdriven amplifier which squares the input sine waveshape and amplifies it according to principles of triode limiting. The .002 mf condenser and 500 K resistor form a short time constant, which peaks the square wave. The grid of the second triode is biased so that only a positive or negative amplified peak appears in the output. Because of this limiting action it is called a limiter or clipper stage.

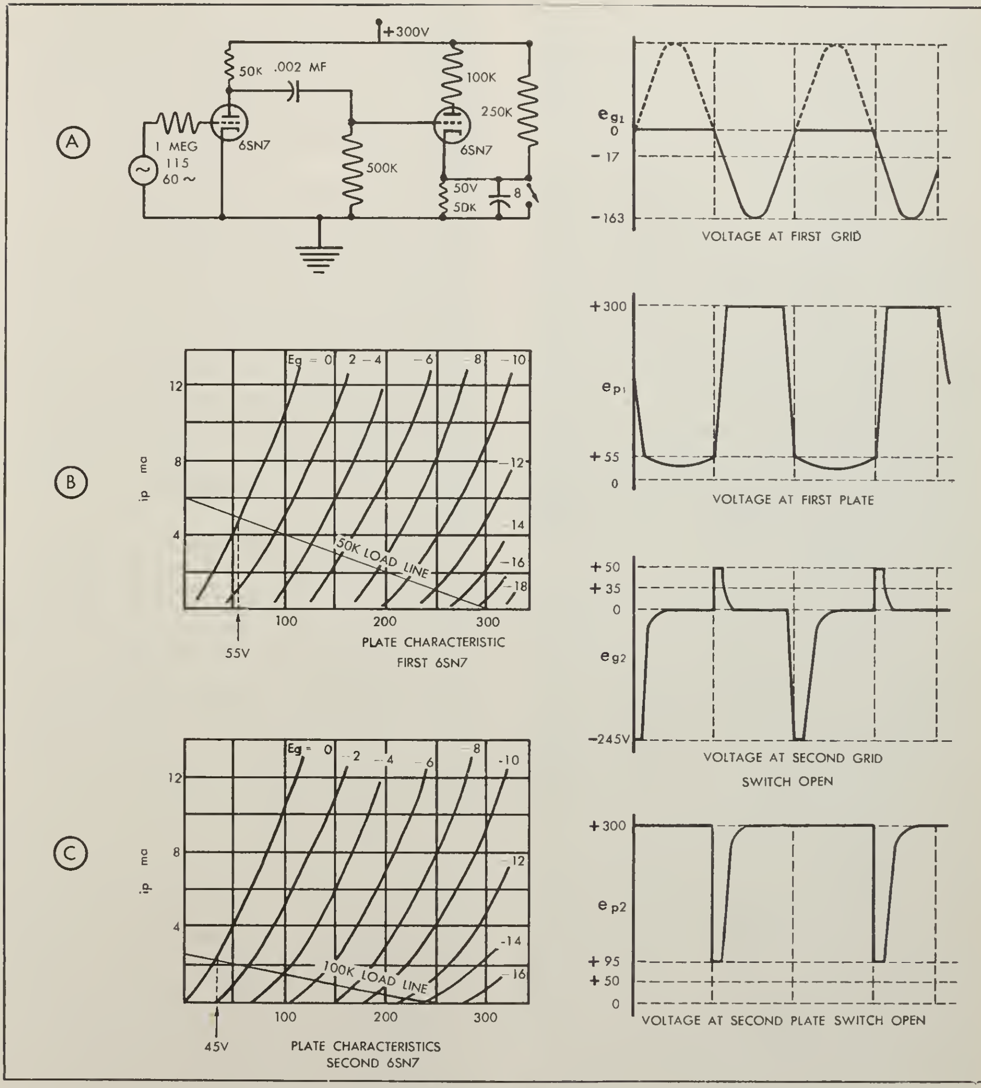
Circuit Analysis

In analyzing the peaker circuit, first inspect the circuit diagram. The first stage is a triode which looks like an amplifier stage except that it is not biased. However, there is a one-megohm grid limiting resistor in its grid circuit. This stage applies its output through capacitive coupling to the second triode, a stage which uses a 100 K load resistor and fixed bias of 50 volts. Bias in this stage is provided by a 250 K and 50 K voltage divider across the positive 300-volt supply.

Now consider the operation of the peaker. As the input voltage goes from zero to a positive



Double Diode Limiter



Squarer and Peaker Circuit

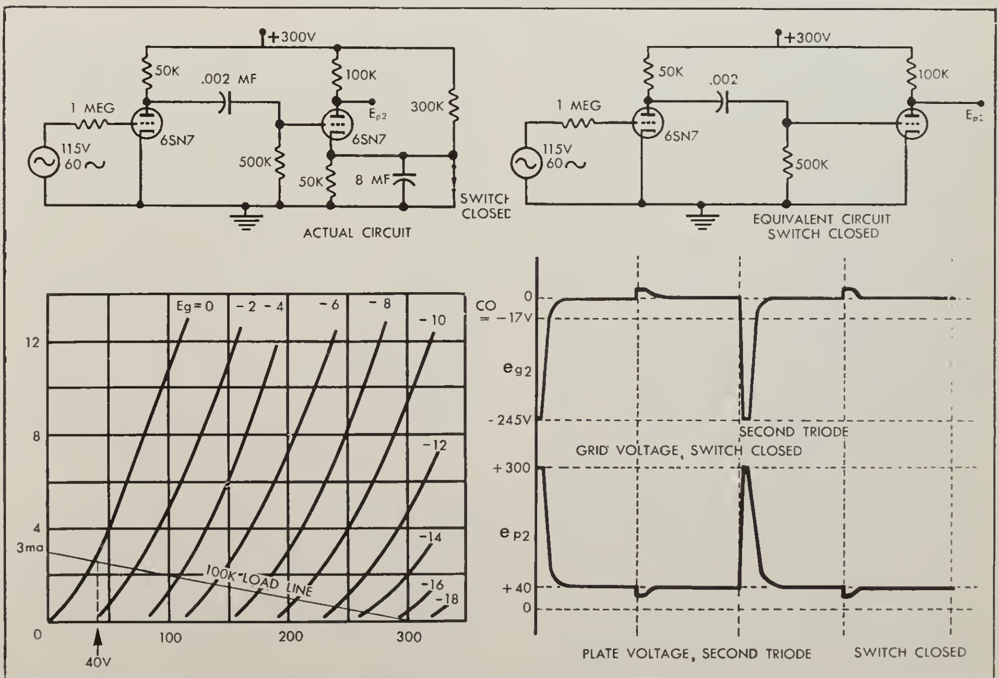
voltage, the grid of the first triode is made positive with respect to its cathode. Immediately, grid current flows through the cathode-to-grid resistance of the tube and the one-megohm resistor. The internal resistance of the cathode to grid path is less than a thousand ohms; therefore, the voltage developed by grid current divides at a ratio of 1000 to 1 across the tube and the plate load resistor. At the positive peak of the input voltage peak (163 volts), the grid voltage is about a tenth of a volt. The grid waveshape shows this voltage as a flat line at the zero level. (The dotted line is the applied voltage; the solid line the actual grid voltage.) The plate voltage of first triode during this time is rather low because the plate current is unlimited by the grid. To determine its exact value, construct a load line on the characteristic curves as follows: First, calculate the maximum current by dividing the plate supply of 300 volts by the load resistance of 50K. Thus, the maximum current is 300 50K, or 6 ma. Draw a line across the curves from the 300 volt point on the bottom to the 6 ma point along the current scale. This line is the load line. The grid voltage is practically zero in the circuit. Therefore, finding the plate voltage for zero grid voltage is the first use of the load line. At the intersection of the zero grid voltage curve and the load line, drop down to the plate voltage scale at the bottom. The point shows the plate voltage is 55 volts. During the positive swing of applied voltage, the grid actually goes slightly positive. This is shown by amplified curve of the plate voltage change which is shown below the grid voltage curve. At the start of the half cycle, the curve is at 55 volts, but drops below that slightly as the grid goes slightly positive, then returns to 55 volts at the end of the half cycle.

Next, consider the input voltage during the negative half cycle. First, however, remember that the grid does not draw current when it is negative. Therefore when no current flows through the one-megohm resistor, there will be no voltage drop across it. Thus, the full input voltage will appear at the grid during the entire negative half cycle as is shown in the grid voltage curve. On glancing at the load line, notice that it crosses the zero current axis at the end of the minus 17-volt grid curve. Thus, when the grid voltage reaches minus 17 volts, the current will have decreased to zero. The grid curves shows that this value is reached very quickly during the negative half cycle. The grid voltage continues to swing with the applied voltage through minus 163 and back to zero. All this time the grid voltage is more than 17 volts negative, and the current remains at zero. As the grid voltage starts to become negative, the decrease of current through the 50 K load resistor causes the plate voltage to increase toward the 300 volt supply voltage. The change is very rapid because the grid voltage crosses the cut-off value in less than a 1/100 of the cycle. Therefore, the plate voltage rise to plus 300 volts is almost vertical. The plate voltage remains at plus 300 volts during the entire time the grid voltage is below the cut-off value on the curve. As it comes above cut-off, the plate voltage again drops to 55 volts. And again, the grid voltage change from minus 17 to zero is very steep, and accordingly the plate voltage change is exceedingly fast. The long period of time that the grid is below cut-off and the rapid changes cause a square wave to be developed at the plate of the first triode.

The square wave output of the first stage is applied to an RC circuit consisting of the .002 MF condenser and 500 K resistor in the grid of the second tube. The voltage across the resistor becomes the grid voltage. At the start of the first half cycle the square wave is negative. Since the condenser is initially uncharged, the application of this negative-going voltage to it immediately causes current to flow through the RC circuit. With no drop across the uncharged condenser, the full voltage appears across the resistor. The third curve labeled e_{g2} shows the entire voltage change occurring at the second grid.

Suppose for a moment you examine the conditions in the second tube when this grid voltage is applied to it. The cathode is 50 volts positive with respect to ground. The actual plate voltage is 300 minus the 50 volts on the cathode, or 250 volts. Notice the load line on the lower set of characteristic curves. This line is drawn from the 250-volt supply voltage point to the 250/100 K, or 2.5 ma current point. On referring to this line, notice that the plate current becomes zero when the grid voltage is minus 15 volts. With 50 volts bias on this tube, there will be no current until more than plus 35 volts is applied to the grid. Therefore, the negative swing of 245 volts applied to the grid will not change the plate current from its zero value. But also note that this waveshape is not like the square wave applied because the RC time constant is short. The total time the voltage is applied is 1/120 or .00833 second, while the RC is .002 mf x .5 megohm or .001 seconds, a ratio of about 8 to 1. In other words the condenser is small and becomes charged during the very first portion of the square wave. As it becomes charged the current through the 500 K resistor gets lower and lower until it drops to zero after 5 time constants or .005 seconds. The resistor voltage stays in step with this decreasing current so the voltage drops exponentially to zero. The grid voltage stays at zero for the duration of the half cycle or until the change over of the square wave occurs.

At the end of the first half cycle the voltage at the first triode changes 245 volts in the positive direction. Again this change immediately appears across the resistor. Nothing happens, however, until the voltage reaches 35 volts. Then currents starts to flow in the tube. As the voltage increases from 35 volts to 50 volts, the actual grid-to-cathode voltage at the second tube rises from minus 15 volts to zero. Now the plate waveshape for this second tube can be examined. During the entire previous half cycle, the tube was cut off and the plate voltage remained at plus 300 volts. But now current is flowing, and the voltage drops to that value of plate voltage for zero grid voltage. This value can be determined from the load line again. Find the intersection of the zero grid voltage line with the load line and proceed downward to the plate voltage scale. It reads 45 volts in this curve. This might seem inconsistent inasmuch as the plate voltage curve at C indicates that the plate voltage is 95 volts. But remember that the voltage is being indicated with respect to ground. Therefore the 45 volts across the tube must be added to the bias voltage to get the total voltage from plate to ground. Now return to the second grid. As the square wave gets above 50 volts, the grid becomes positive and starts to draw current. So you see that the positive-going voltage on the second grid levels off at zero. Only a small part of the remaining voltage rise of 245 volts ever appears across the grid to cathode. This is due to the fact that the grid current flows through the .002 mf condenser and 50 K load resistor, and this combination of the 50 K resistor and the high reactance of the small capacitor acts as a grid limiter to divide the total voltage. The small grid-to-cathode resistance of about 500 ohms will have a very small voltage drop across it. By the way, the 500 K resistor is still in the circuit but so much of the current flows through the grid itself that the 500 K resistor can be practically disregarded during this time. The low resistance and small capacitor form a very short RC circuit, so the capacitor will be charged in a hurry. The voltage

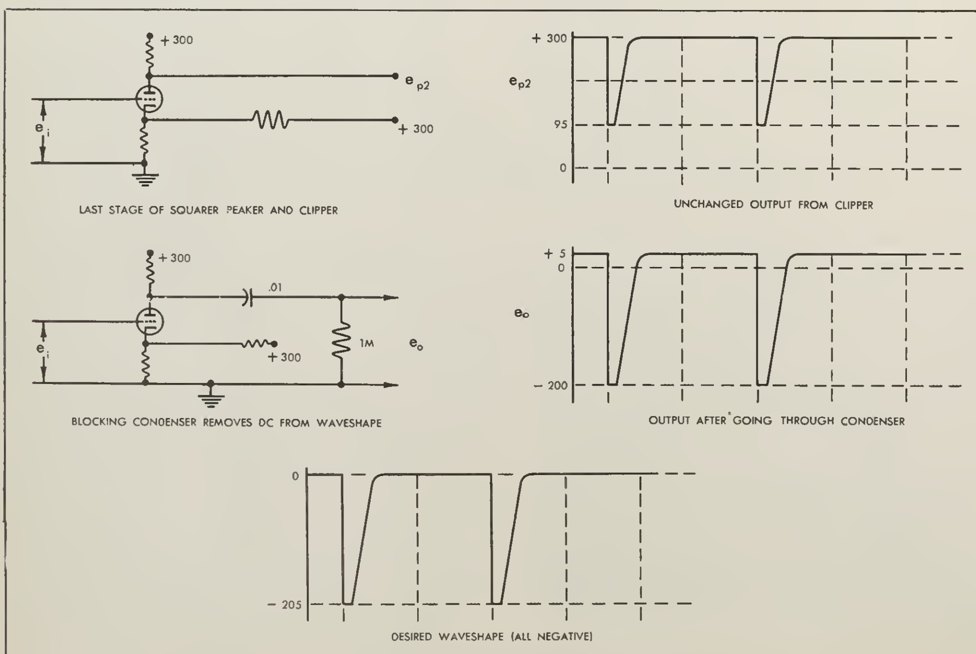


Squarer and Peaker with Switch Closed to form Negative Peak at the grid then drops back to zero with respect to ground—which is 50 volts negative with respect to the cathode. So the triode conducts only during the brief period of time during which the grid is above cut-off. The resultant plate voltage drop is also very brief. The plate voltage stays at plus 300 volts most of the time. A negative pulse is thus generated for each cycle of the original input sine wave. Its amplitude is 300-95 or 205 volts.

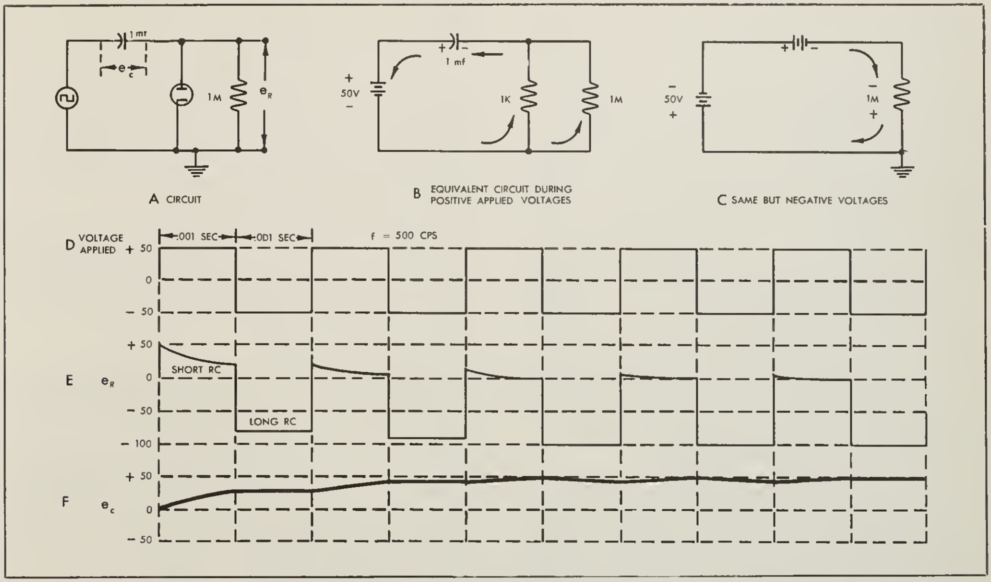
When the switch is closed this circuit will generate a positive pulse for each cycle of sine wave. As shown on page 6-35, the switch is closed in the circuit at the left. When the switch is closed, it shorts out the 50 K bias resistor and its parallel condenser, and connects the 300 K resistor to ground. The three circuit elements become ineffective and an equivalent circuit may be drawn by omitting these elements. The circuit at right is the equivalent circuit. Note that now there is no bias on the second triode, and that the full 300 volts supply voltage is across the tube and its load resistor. To evaluate the current and voltages, it is necessary to draw a new load line shown in the characteristic curves. This line extends from 300 volts at the bottom to 300 100 K or 3 ma at the side. With no bias, the grid voltage is zero, and the plate voltage under this condition is 40 volts.

On the negative half cycles the square wave will be peaked in the same manner as before. But on the positive peaks, grid current flow occurs at zero instead of the plus 50 volts as before. Similarly, as in the previous case, the high impedance circuit of the .002 mf condenser and 50 K load resistor prevents the grid from getting more than a fraction of a volt positive during positive half cycles.

The grid remains at zero value except when the input voltage is negative-going. Therefore at zero grid voltage where the plate current is high, the plate voltage which varies inversely with the plate current remains at a constant value of 40 volts. At the close of the first half cycle, the voltage is negative-going. Here the grid voltage changes rapidly from zero to the cut-off value, and the plate voltage rises sharply from 40 volts to plus 300 volts. The grid



Typical Need for Restorer



Direct Current Restorer

voltage becomes negative by 245 volts, and the resulting plate voltage levels off at the supply voltage level of 300 volts. However, the rapid charge of the .002 mf capacitor causes the grid voltage to rise immediately to the cut-off value and beyond. As the grid voltage continues to rise to zero, the plate voltage drops back down to 40 volts. If you study the plate voltage waveshape, you notice the fairly constant low plate voltage with positive peaks with each cycle of input voltage. The positive peak voltage at the second grid only raises the grid a fraction of a volt above zero, thus causing the plate voltage to dip slightly at each of these points. Thus you see, with this circuit it is possible to obtain either positive or negative peaked pulses by the flip of the switch.

DIRECT CURRENT RESTORERS

A *direct current restorer*, or clamping circuit as it is also called, is a circuit which shifts a waveshape so that it is all above or all below a certain voltage—often zero voltage. It is also called a *clamping circuit* because it clamps the top or bottom of a voltage waveshape at a certain voltage, which may be zero, any positive voltage, or any negative voltage, depending on the design of the circuit.

Negative Direct Current Restorer

A typical example of the need for a clamping circuit is the waveshape output produced by the squarer and peaker circuit (page 6-36) which shows a negative peak output. Although this voltage is called a negative peak voltage, it actually is all positive, the peak merely being less positive than most of the waveshape. To make it a truly negative peak voltage, all the positive voltage must be removed. This is partly accomplished by a blocking condenser, which removes all the DC component. With no DC component removed, part of the waveshape is positive and part is negative. For the waveshape to be all negative, a DC negative voltage must be added to it. Adding the negative DC voltage shifts the whole waveshape downward with respect to the zero axis, creating the desired waveshape as shown in the illustration. The action is accomplished by the clamping circuit which *restores* enough of the DC component to make the waveshape all negative.

Circuit analysis. To analyze a simple DCR (direct current restorer) in detail, study the elementary circuit illustrated above at A. Judging by its action, it actually consists of two time constants, a short one during one half cycle and a long one during the other half cycle.

The applied voltage used in the circuit just mentioned is a square wave with a frequency of 500 cps and with a variation from plus 50 volts to minus 50 volts in amplitude or 100 volts peak-to-peak. Starting from the beginning, assume that the condenser is uncharged. When the first positive half cycle is applied, the condenser will be charged to a constant 50 volts for .001 second and this positive voltage will cause the diode to conduct. Since the diode has an average internal impedance of a thousand ohms, the 50-volt potential will charge the condenser through a resistance of 1000 ohms in parallel with 1 megohm, or 999 ohms. The time constant is about 1000 ohms x 1 mf or .001 seconds. Due to these conditions the condenser will charge 63% during the half cycle, or to about 31 volts. The equivalent circuit at B for this half cycle shows the action just described.

During the negative half cycle a negative 50 volts is applied to the circuit. The condenser is charged to 31 volts and both the applied voltage and the condenser cause current to flow in the resistor. During this cycle the diode plate is negative and does not conduct. The RC is 1 megohm x 1 mf or one second. Notice in the diagram that current from the discharge of the condenser and current from the source both are in the same direction. Since current in the same direction develops additive voltages, the voltage applied to the circuit equals 50+31 or 81 volts. The time constant is now equal to 1000 times the half cycle. Therefore the condenser will discharge only .1% during the half cycle.

On the next cycle the condenser starts with almost a 31 volts charge. This 31 volts opposes the applied voltage and leaves only 50−31 or 19 volts to force current through the circuit. Since the time constant is short when the tube conducts the condenser will charge 63% of the 19, or 12 volts. Adding this to the 31 volts already on the condenser makes the charge at the start of the next half cycle 43 volts. During the next half cycle, the discharge is slight, and the voltage at the source and the voltage across the condenser cause a constant 50+43 or 93 volts to appear across the resistor. The condenser continues to be charged to 63% of the remaining voltage until its charge is practically 50 volts. At this time each positive half cycle will cause practically zero voltage across the resistor while each negative half cycle will cause 50+50 or 100 volts. In effect, the input voltage is half negative and half positive while the output (resistor) voltage is all negative.

Because of the preceding action the entire waveshape has shifted downward until the top is at the zero axis. Electrically you can say that a 50 volt DC component has been inserted. If the waveshape had been all negative at one time in the circuit, the restorer can be said to have restored it to negative.

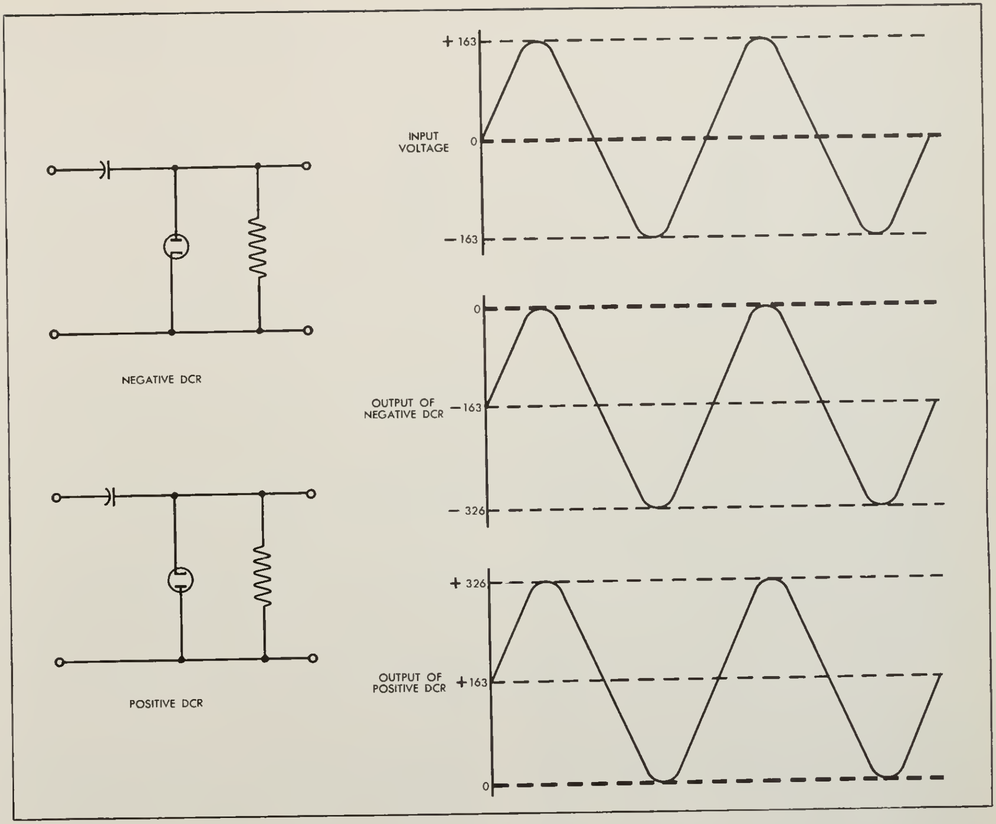
Actually, there is a tiny bit of the waveshape left above the zero axis. This bit is the fraction of a volt required to recharge the condenser after each half cycle. With proper circuit design this can be made negligible.

The circuit described is a *self-adjusting* circuit, in that the top of the waveshape is clamped at zero regardless of the amplitude of the source voltage. When the amplitude is greater than shown, the condenser charge becomes greater. When the amplitude decreases, the condenser is not recharged as high during the charging time, and is allowed to discharge to the new lower voltage required.

VARYING THE R AND C. When the condenser in a restorer circuit is made smaller, the charging rate is increased. However, the charging time is usually very short anyway and is of little consequence. But, on the other hand, the discharge time becomes faster. This is an advantage when the amplitude of the input waveshape is suddenly reduced and a disadvantage when a constant input amplitude is present. As the condenser discharges a lot with each cycle, a large part of the waveshape is shifted to the wrong side of the zero axis for recharging. A further result is that the recharging becomes a distortion of the original waveshape, a condition which usually is not desirable.

Increasing the resistor size increases the discharge time causing less distortion during the cycle, but slowing the readjustment if a sudden drop in input amplitude occurs. The charge time (due to a sudden increase in input voltage) is unaffected because the diode resistance carries most of the current since the resistor current is too small to effect the charging time.

Circuit designs often incorporate a compromise between a short RC for following changes in amplitude and a long RC for least distortion. An important point to remember is that a DCR (Direct Current Restorer) will work with no resistor at all, even though most designs include a resistor in parallel with the diode.



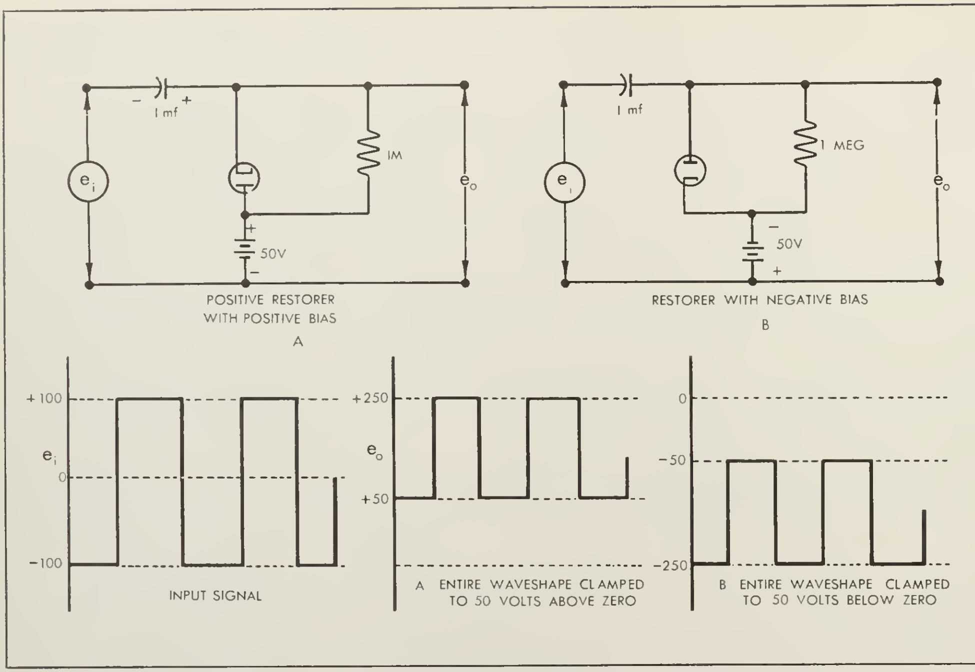
Positive DCR

The Positive DCR

The DCR circuit just discussed may be called a negative DCR because it restores the waveshape in the negative direction. To shift the waveshape in the positive direction simply reverse the diode connections as shown in the illustration of a positive DCR circuit. Here the input waveshape is a sine wave, but the action is the same as that in a negative DCR.

Biased Restorers

A biased restorer clamps a waveshape at an axis other than zero by means of inserting a DC bias of a voltage and polarity equal to the voltage at which clamping is desired. The only change in the circuit from the restorers just taken up is the insertion of a DC potential in series with the diode. The circuit at A on the next page is a positive clamping circuit; that shifts the entire waveshape to the positive side of zero. In operation, the DC source inserts an additional positive 50 volts, which shifts the entire waveshape to 50 volts above zero. Thus, the waveshape which originally varied from minus 100 volts to plus 100 volts, now varies from plus 50 volts to plus 250 volts. Electronically, the shift is due to the DC source charging the condenser to the DC potential, so that when the tube is nonconducting, the condenser has an extra voltage on it, in addition to that from the input voltage. For example, in this case, the square wave would charge the condenser to half its peak-to-peak value, or 100 volts. This voltage is sufficient to clamp all the square wave in the output above the zero axis. Since the battery will add a 50 volt charge to the condenser, the total charge will be 150 volts.



Biased Restorers

The square wave input varies around this value in the output or between 150+100 and 150−100 which is 250 to 50 volts.

Restoring in a Grid Circuit

Restoring can occur in a grid circuit due to the fact that the grid and cathode of a triode, in the absence of fixed bias, will form a diode and act as the clamping circuit. In the restorer circuit on the next page, grid current flows during positive half cycles, and charges the condenser to a value which clamps the positive extremity of the input signal at zero. Except for the small positive voltage required to recharge the condenser with each cycle, the signal at the grid will be all negative. The charge on the condenser becomes the grid bias voltage. This method for providing bias has the advantage of automatic adjustment to any amplitude of input signal. This means that, as in diode DCR's, a larger signal will cause the condenser to charge to a higher voltage, and reduction in signal amplitude will discharge the condenser accordingly. In every case, the positive part of the signal is maintained at the zero axis. Note the input and grid waveshapes in the illustration. The average bias (condenser charge) for producing these waveshapes is minus 20 volts. The effect of a reduction in amplitude is shown in the last waveshape. The top is still clamped at zero. This clamping action does not change with frequency.

In radio circuits, grid clamping is called grid-leak bias, a type of bias which is generally used in Class C RF amplifiers in communications transmitters.

TIME BASE GENERATORS

A time base generator circuit is a circuit that generates the voltage which causes the spot to move across the CRT screen at a constant rate. The distance along the trace formed by the moving spot is the basis of establishing time intervals. When recurrent voltage changes are superimposed on the trace, it is possible to measure both time and frequency of a radio signal, and in radar to measure the distance of targets. Commercial test oscillographs employ a thyratron tube in the time base generator circuit because of the simple way in which the frequency can be changed. In radar sets the time base generator circuit usually employs a high vacuum tube.

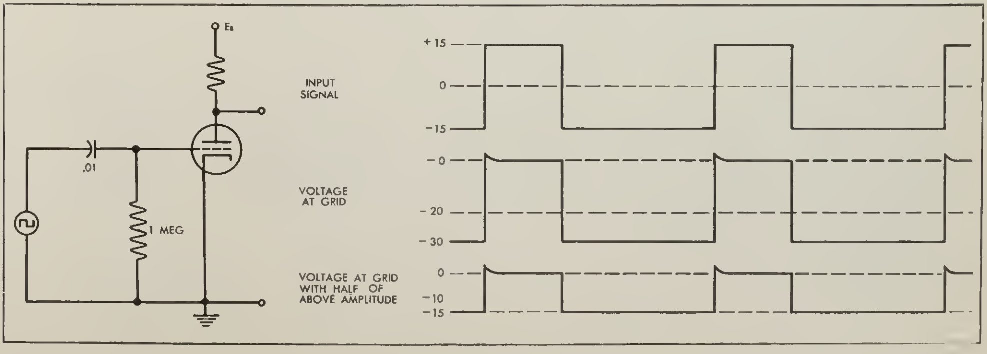
The Thyratron Sweep Generator

The movement of the spot in a CTR may be caused either by a voltage change in the deflection plates or by a current change in the deflection coils. The chief requirement is that the voltage (or current) must change linearly with time.

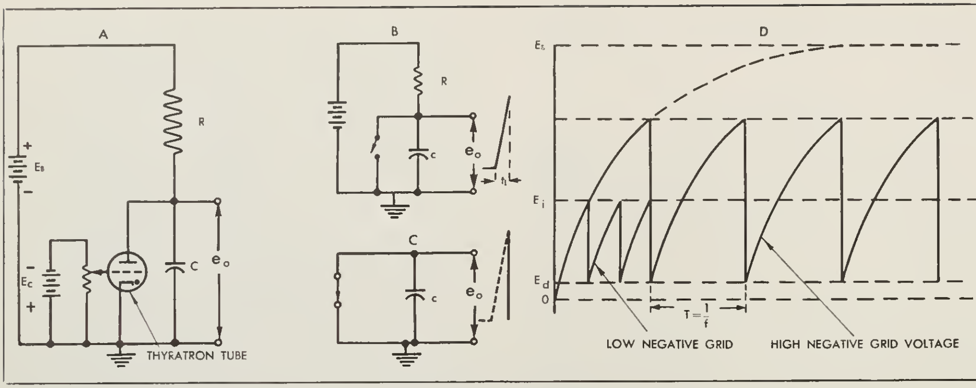
Although both the RC and LR circuits consume time during a voltage change, the RC circuit is the one usually preferred as a means of generating a changing voltage. As previously mentioned, an RC circuit contains a DC voltage source, a series resistor and a condenser. When those components are connected, the condenser will start to charge but since the resistor limits the current, the charge will not be instantaneous, but at a rate which is exponential. The charging curve is quite linear in the very beginning, and the condenser charging process is usually stopped before the slope of the charging curve changes appreciably. After the charging stops, the condenser discharges, and the events just described repeat and the condenser recharges.

Most time base generators use this very same charging and discharging process. Any circuit differences that exist result largely from the methods employed for high speed switching of the condenser from charge to discharge. One device quick enough for this purpose is the vacuum tube. The circuit at A on page 6-42 shows a gas triode or thyratron, a special type of vacuum tube, employed for providing the necessary switching operation since this tube has certain special characteristics which make switching occur automatically. In the circuit the tube is in parallel with the condenser. When the circuit is turned on, there is no charge on the condenser, and the tube acts like an open circuit due to the negative cut-off voltage on the control electrode, or grid. The condenser charges along the first curve at D. As the voltage increases, a potential is reached where the voltage across the tube is too high for the control electrode to keep the tube cut-off and a small plate current starts to flow. This plate current ionizes the gas in the tube, and before the voltage can increase any further a sudden high plate current flows. The plate voltage for the tube is provided by the charged condenser. Forcing a current through the tube causes discharge of the condenser. Below a certain potential, the tube does not remain ionized, and when the condenser is discharged down to this potential, the gas in the tube deionizes, and the grid regains control of plate current. At this point it becomes an open circuit again and permits the condenser to recharge. For further study refer to the equivalent circuits; the charge of the condenser at B and the circuit during discharge at C.

The curve for several charges and discharges at D shows that the voltage increases to the ionizing potential Ei, and then drops very abruptly to Ed which is the deionizing potential. The slow rise is due to the size of R and C. The rapid drop in voltage forms an RC curve, but the resistance of the ionized tube is so small that the condenser is discharged almost immediately, causing the curve to be practically vertical. The



Restoring in the Grid Circuit

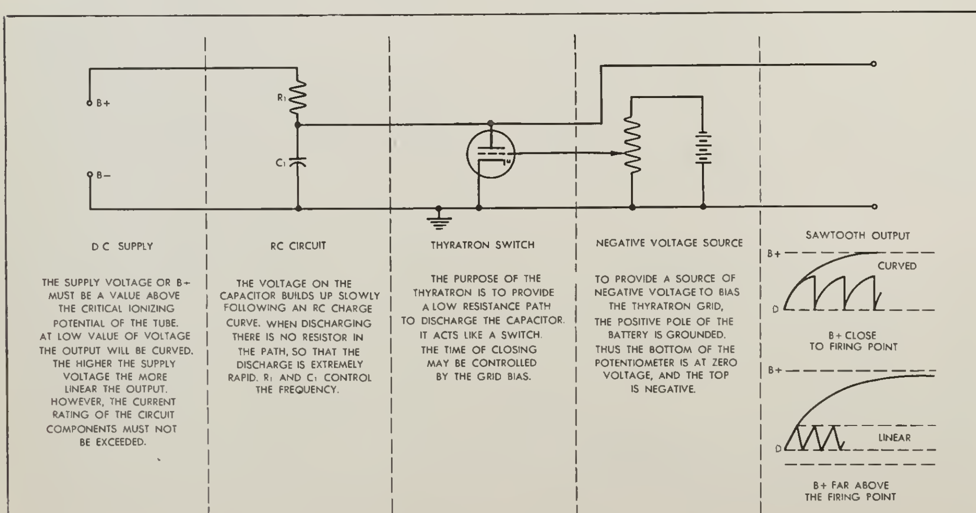


Thyrotron Sweep Generator

voltage at which ionization occurs depends on the amount of voltage on the grid of the thyratron. If the grid is made less negative, a lower plate voltage will cause ionization as you see in the second curve at D. When ionization occurs earlier the amplitude of the waveshape may decrease on one hand, but on the other, since the discharge time is the same, the wave will repeat more often, that is, its frequency will become higher. Another result is that the charging curve is much more linear, that is, its slope is more constant.

Two facts of interest in thyratrons are that the deionization potential depends entirely on the gas pressure within the thyratron envelope and that deionization always occur at the same potential regardless of the grid voltage or the applied plate potential. For example, the 884 tube, one customarily used in time base generator circuits, deionizes at about 20 volts.

The circuit which was described is called a free-running circuit, that is, when a DC potential is applied to it, it generates one time



Functions of Circuit Components

base after another indefinitely. There are other circuits which are discussed later that must be driven. These generate only one time base and then wait for the driving voltage to start the charging process which generates the next time base.

CALCULATING FREQUENCY OF OSCILLATION. To calculate the frequency of oscillation of a thyratron sweep generator there are few factors which you must know. To get acquainted with these factors note the illustration which shows a practical thyratron sweep generator circuit. The applied voltage is 250 volts. The 50 K resistor in the plate circuit provides a minimum resistance in series with the tube. The 10 K resistor in the grid circuit limits the grid current to a safe value. To see the details of this circuit's operation, use the values just given in a problem in which the bias is set at minus 8 volts and the resistor equals 316 K. The curve at B shows the relation of the firing potential to the grid bias. Thus, when the grid voltage is minus 8 volts, the tube ionizes when the plate-to-cathode voltage reaches 75 volts. Using this information, you can plot a charging curve like that shown at C. This shows that during the first cycle, the condenser voltage starts from zero, but on the second and succeeding cycles, it always starts charging from 20 volts, the deionization potential. This charge cancels 20 volts of the 250 volts plate potential. At 75 volts, the thyratron ionizes, stopping the condenser charge. With 230 volts applied, therefore, one cycle has a duration equal to the time required to charge the condenser from 20 to 75 volts, or 55 volts. To determine what fraction of full charge is represented by 55 volts, use the universal time constant chart curve. From it you learn that this value is 55/230, or .239. Therefore, the condenser is charged 23.9% before ionization. Using curve A in a universal time constant chart curve, find .239 of full charge, then move across to the curve, and down to .26 time constants. One time constant is RC, or 366 K x .01, or .00366 seconds. Therefore, the time required for the charge is .26 x .00366, or .000952 seconds. Since frequency is the reciprocal of the time of one cycle, F = 1/.000952, or 1050 cps.

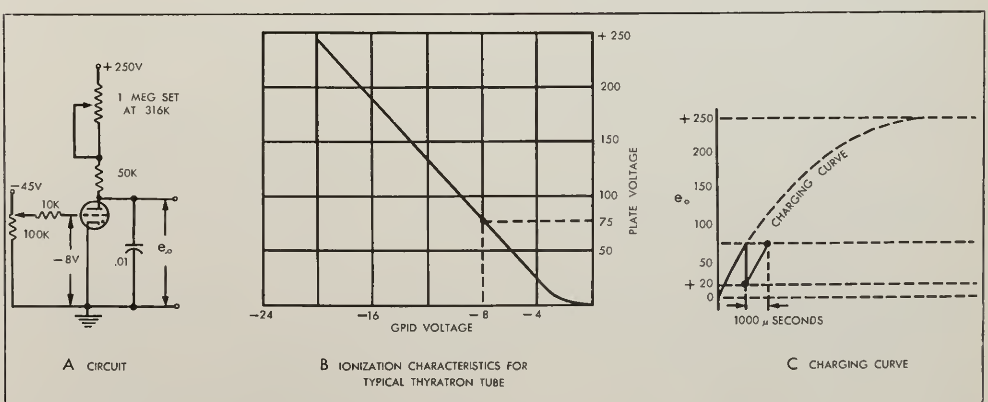
For greater accuracy, find t by the exponential equation, $e_c = E_b\,(1-\epsilon^{-t/RC})$. Since the condenser starts charging from 20 volts toward 250 volts, the curve is the same as from zero to 230 volts. This makes the E_B in the equation equal to 230 volts. The condenser charges from 20 to 75 volts, or to 55 volts of the 230. Therefore e_c is 55 volts. Substituting,

$$55 = 230\,(1-\epsilon^{-t/3.66\times10^{5}\times10^{-8}})$$

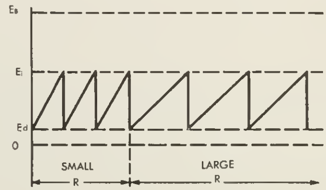
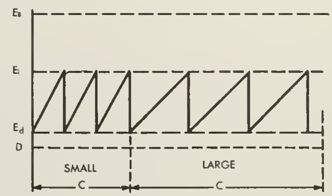
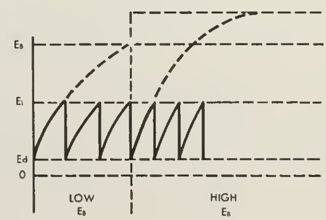
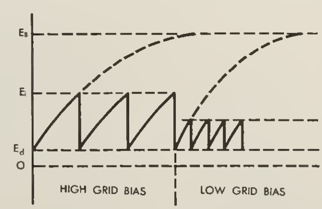
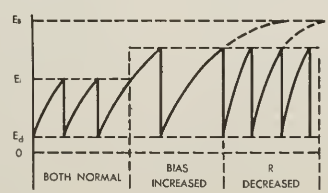
$$.273 = \frac{t}{3.66\times10^{-3}}$$

$$t = 10^{-3} \text{ seconds}$$

$$F = \frac{1}{t} = \frac{1}{10^{-3}} = 1000 \text{ cps}$$



Practical Thyratron Sweep Generator Circuit

ELEMENT VARIED	DURATION OF ONE CYCLE	FREQUENCY	AMPLITUDE	LINEARITY	EFFECT ON WAVESHAPE
INCREASING R	LONGER	LOWER	UNCHANGED	UNCHANGED BECAUSE C STILL CHARGES TO SAME PERCENTAGE OF E_B	
INCREASING C	LONGER	LOWER	UNCHANGED	UNCHANGED	
INCREASING E_B	SHORTER	HIGHER	UNCHANGED	IMPROVED BECAUSE LESS OF CHARGING CURVE IS BEING USED	
DECREASING NEGATIVE VOLTAGE ON GRID	SHORTER	HIGHER	DECREASED BECAUSE IONIZATION POTENTIAL IS LOWER	IMPROVED BECAUSE A SMALLER PART OF THE START OF THE CHARGING CURVE IS USED	
INCREASING THE BIAS FOR GREATER AMPLITUDE AND DECREASING R TO RETAIN ORIGINAL FREQUENCY	UNCHANGED	UNCHANGED	INCREASED BECAUSE IONIZATION POTENTIAL IS HIGHER	DECREASED BECAUSE GREATER PART OF CHARGING CURVE IS USED	

Effect of Varying Elements in Thyratron Sweep Generator

The time base voltage repeats 1000 times per second. To calculate frequency directly, use the following equation:

$$F = \frac{1}{RC \log_\epsilon \frac{(E_B - E_i)}{(E_B - E_d)}}$$

Where,

F = *Frequency in cycles per second*

RC = *Time constant*

E_B = *DC applied voltage*

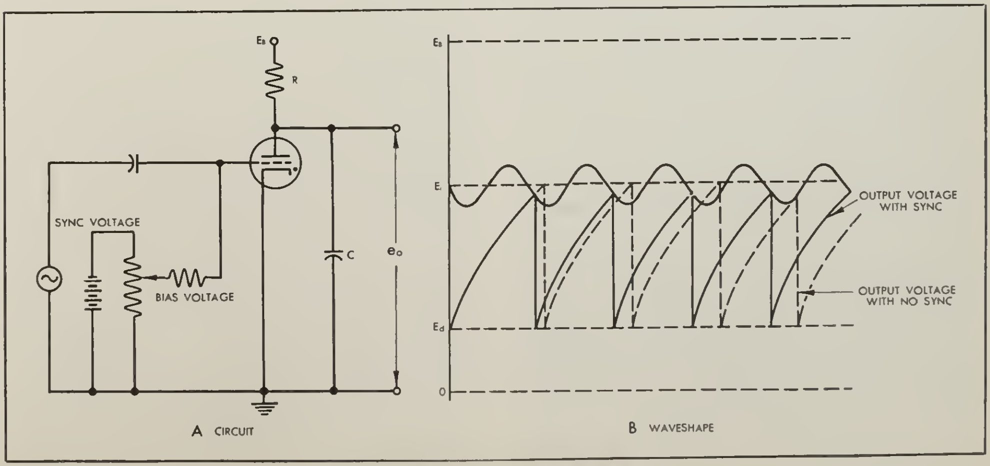
E_i = *Ionization potential*

E_d = *De-ionization potential*

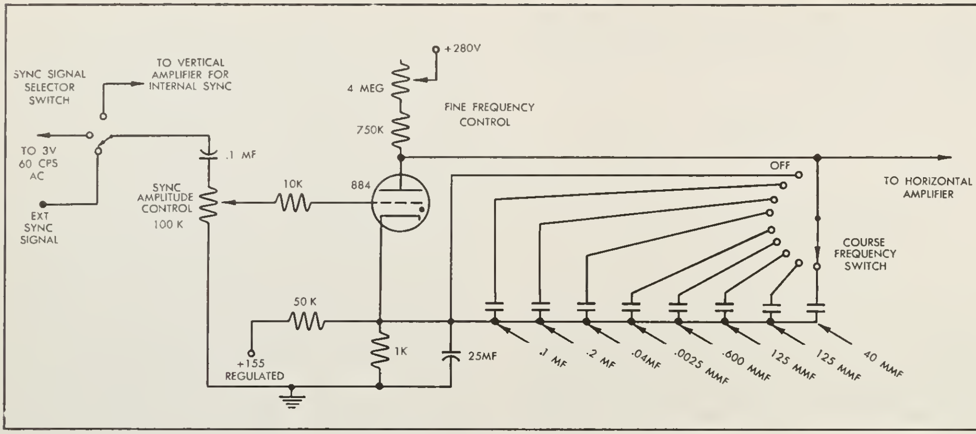
VARIATION OF CIRCUIT ELEMENTS. Any variation in the circuit constants affects frequency, linearity, and amplitude. The effect of varying circuit elements in a thyratron sweep generator are summarized in the chart on the preceding page.

Employment of a thyratron as a time base generator in a commercial oscilloscope, requires a frequency that is variable. In these sets coarse frequency changes are provided by means of switching condensers and fine frequency adjustment by means of varying the plate resistance and holding the grid bias constant. The usable frequency range extends from a few cycles to 50 kcps in the commercial oscilloscope. The upper frequency limit is 50 kcps because the discharge time at 50 kcps is equal to the charging time, and half of any signal being investigated will be lost during the flyback (discharge) time.

SYNCHRONIZATION. In addition to all the factors just mentioned, the temperature of the gas, the slightly variable ionization potential, and the load on the circuit also affect frequency. Because of these factors, the circuit is usually not stable enough to maintain an adequately constant frequency by itself and must be synchronized with an AC voltage that has a constant frequency. This synchronizing voltage may be any voltage that has a frequency near (usually a little above) the multiple, submultiple, or fundamental frequency of the thyratron generator. Due to the synchronization voltage, the thyratron tube ionizes slightly ahead of normal time, and the condenser discharges immediately and starts a new cycle. Diagram A below shows a typical synchronization circuit in which a sine wave is used to synchronize the generator in the grid circuit. The sine wave varies the grid voltage and this in turn the ionization potential. The waveshape at B shows how the ionization potential varies at a slightly higher frequency than the natural free-running frequency of the generator. The dotted sawtooth waveshape results from a constant ionization potential. At the time the condenser voltage is about to reach the normal value, the ionization potential drops. When the condenser voltage reaches the ionization potential, the ionization occurs, starting a new cycle immediately. This



Synchronization with Sine Wave



DuMont 208 Oscilloscope Thyratron Time Base Generator

same action recurs with each cycle and causes the time base frequency to "lock-in" step with the synchronizing frequency. Often, in the use of the commercial oscilloscope, the waveshape being displayed on the CRT is used for synchronization. This is done so that the waveshape and time base will occur simultaneously, and the pattern will be stationary. If there is no synchronization, a difference in frequency will exist and the pattern displayed on the screen will move forward or backward along the time base.

DUMONT 208 OSCILLOSCOPE. A typical set using a thyratron time base generator is the DuMont 208 oscilloscope. The tube is a type 884 thyratron. Fixed bias is provided in the cathode by a voltage divider from B+. The bias is held constant at 3.1 volts when the tube is nonconducting. Synchronization may be switched in from three sources—the signal being displayed, the 60 cycle power line frequency or from a terminal post to which any other signal can be connected. The 100K potentiometer provides for varying the amplitude of the synchronization voltage applied to the grid. One of the eight condensers when connected in the RC circuit provide for the course frequency adjustment. The 4-megohm potentiometer provides for fine tuning in that it can be used to set the frequency at exactly the desired value. Maximum linearity is obtained by keeping the amplitude of the output wave low. The desired amplitude is produced by the amplifiers. The output signal from this stage goes to a video amplifier.

The Hard Tube Sweep Generator

Because a thyratron is not readily adaptable for producing the time base with the spaced intervals required by radar equipment, the high vacuum or hard tube is used instead. A hard tube is one that is highly evacuated, that is, one from which as much gas as possible is removed. In contrast, a *soft* tube is one with an incomplete vacuum, that is, one in which some gas remains in the envelope. A hard tube serves as a switch for discharging the condenser but is not automatic and must be operated by an external circuit.

At A on the next page is the circuit of a hard tube sweep generator. At B, the top waveshape shows the waveshape of the voltage that is usually used to operate the circuit. Normally, this voltage will hold the grid of the tube at zero. The tube in the circuit conducts a relatively large current. The load line shown at C shows this current to be nearly 2 ma. According to the load line chart, the plate voltage due to this current is 25 volts. A time base is formed whenever a radar pulse is transmitted. Simultaneously with the transmitted pulse, the negative square wave input voltage drives the grid far below cut-off. As a result, the plate voltage tends to rise immediately to 250 volts, but since it cannot rise faster than the condenser can charge, its rise is exponential. If a long time constant is used, the condenser can only reach a small percentage of full charge before the grid cut-off voltage is removed, and tube current discharges the condenser back down to 25 volts. The long time interval between transmitter pulses and the short time constant formed by the low tube resistance insure condenser discharge to 25 volts before the next cut-off voltage occurs.

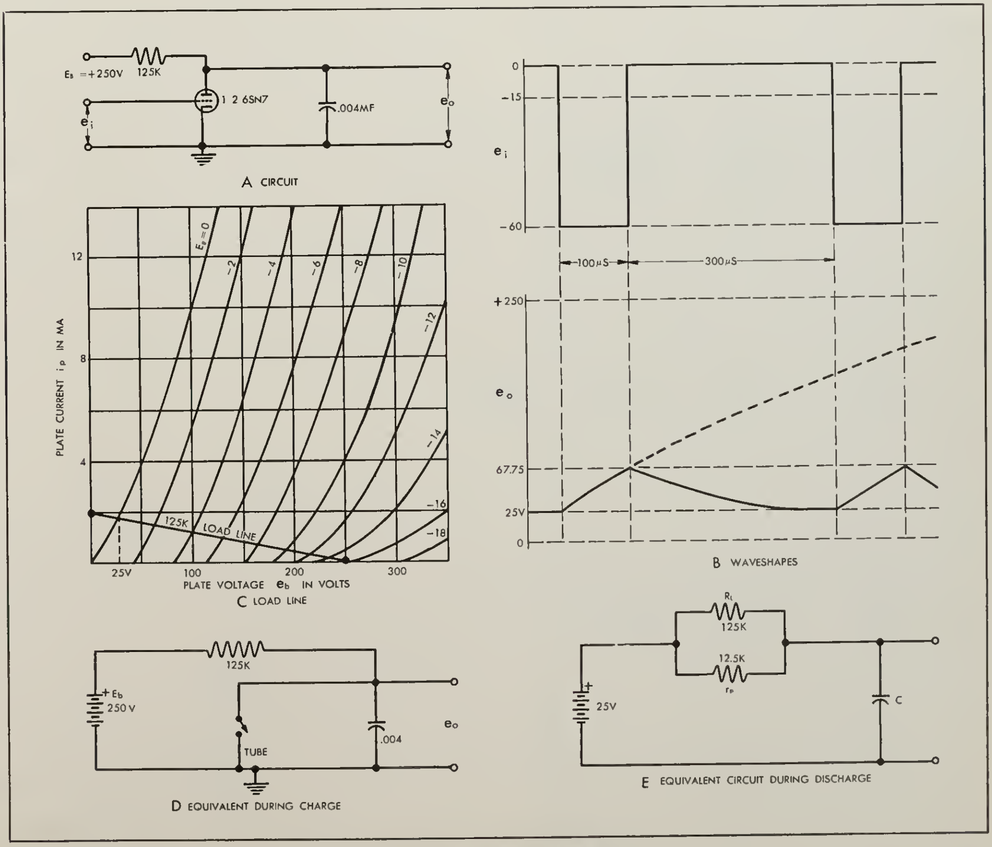
You can readily calculate the amplitude and linearity of a time base such as the one formed in the preceding discussion in the following manner. First, draw the load line for the tube. The load line shows that the voltage at the plate is 25 volts when the grid voltage is zero, and that cut-off for the tube is −16 volts when 250 volts is applied to the plate. Other factors to consider are that the −16 volts cut-off bias is easily exceeded by the −60 volt of the square pulse, which is 100 microseconds in duration (represented as time t in the calculations), and that the applied voltage to the condenser is 250−25 or 225 volts. To find the condenser voltage (e_c) after 100 microseconds, perform the following operation.

$$e_c = 25 + E_B\ (1 - \epsilon^{-t/RC})$$

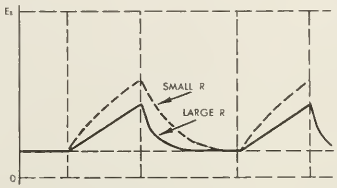
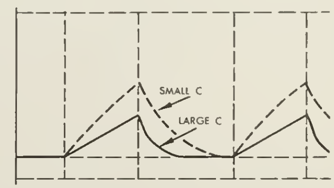
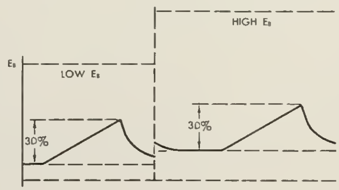
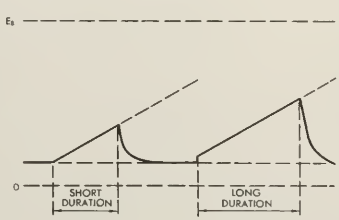
$$e_c = 25 + 225\ (1 - \epsilon^{-10^{-4}/1.25\times10^{5}\times4\times10^{-9}})$$

$$e_c = 67.75 \text{ volts}$$

The equivalent circuit during the charge is shown at D. Here the tube acts as an open switch. If the rise continued at the initial rate, the curve would be perfectly linear. At the initial rate, the voltage would rise 225 volts during the first time constant. Since it is allowed to charge for .2 time constants, it would have risen 225 x .2 or 45 volts after this time. However, the actual rise is 42.75 volts (67.75−25). Thus, the actual curve varies from the linear curve by 2.25 volts.



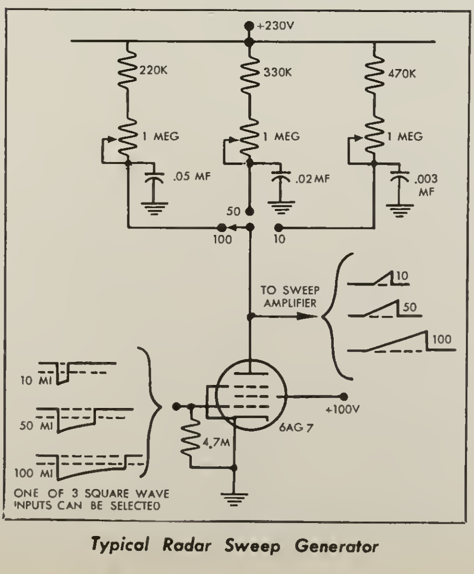
Hard Tube Sweep Generator

ELEMENT VARIED	EFFECT ON AMPLITUDE	EFFECT ON LINEARITY	EFFECT ON DURATION	WAVESHAPE
R IS INCREASED	DECREASED. BECAUSE CONDENSER CHARGES SLOWER AND IS STOPPED BEFORE IT GETS AS HIGH	IMPROVED. BECAUSE LOWER AMPLITUDE MEANS LESS OF CHARGING CURVE IS USED	NO CHANGE. DURATION DETERMINED BY INPUT WAVESHAPE.	
C IS INCREASED	DECREASED. CONDENSER CHARGES SLOWER, AND IS STOPPED BEFORE IT GETS AS HIGH	IMPROVED. LOWER AMPLITUDE MEANS LESS OF CHARGING CURVE IS USED.	NO CHANGE. DURATION DETERMINED BY INPUT WAVESHAPE.	
E_B IS INCREASED	INCREASED. CONDENSER CHARGES TO SAME PERCENTAGE WHICH IS A HIGHER VOLTAGE	NO CHANGE. CONDENSER CHARGES TO SAME PERCENTAGE AS BEFORE	NO CHANGE.	
DURATION OF INPUT CUT-OFF VOLTAGE INCREASED	INCREASED. CONDENSER CHARGES LONGER AND SO TO A HIGHER VOLTAGE	DECREASED. MORE OF CHARGING CURVE IS USED.	INCREASED BECAUSE INPUT DURATION IS INCREASED	
AMPLITUDE OF INPUT SIGNAL INCREASED	NO CHANGE IF SIGNAL WAS ALREADY CUTTING OFF TUBE	NO CHANGE IF SIGNAL WAS ALREADY CUTTING TUBE OFF	NO CHANGE	NO CHANGE IN PLATE WAVESHAPE.

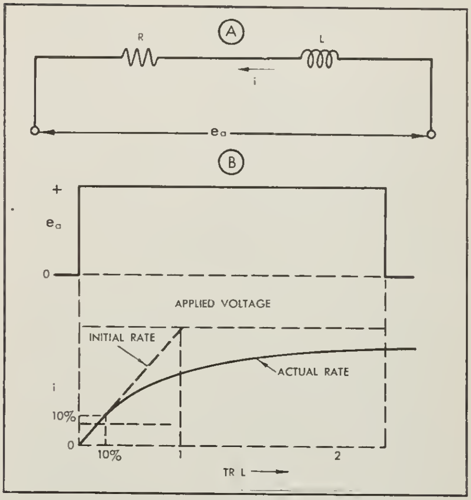
Effect of Varying Elements in Hard Tube Sweep Generator

When the grid voltage returns to zero, the equivalent circuit becomes as shown at E, and the following action occurs. The condenser discharges from 67.75 volts to 25 volts through the internal resistance of the conducting tube and the load resistor. At zero grid voltage, the plate voltage is 25 volts, and plate current approximately 2 ma. Although this value of plate current holds only when the grid voltage is zero, it is used here because the grid voltage remains constant at zero until the next cycle of negative voltage arrives. The resistance of the tube in parallel with 125K makes a total resistance of 11.35K. The time constant of this RC is 11.35 x .004 mf, or 45.5 microseconds. Because there are 300 microseconds between pulses, there is more than ample time for the condenser to discharge to 25 volts.

Typical radar sweep generator. As shown below, the typical sweep generator circuit in a radar set has a negative square wave applied to the grid of the tube. The duration of the square wave is adjusted to exactly the time required for the radar wave to travel 10, 50, or 100 miles. When the range switch is on the 10 miles position, the input voltage is the 10 mile square wave, and the .003 mf condenser is switched in the plate circuit. For the other ranges, larger condensers are switched into the plate current, and the input square wave is longer.



Typical Radar Sweep Generator



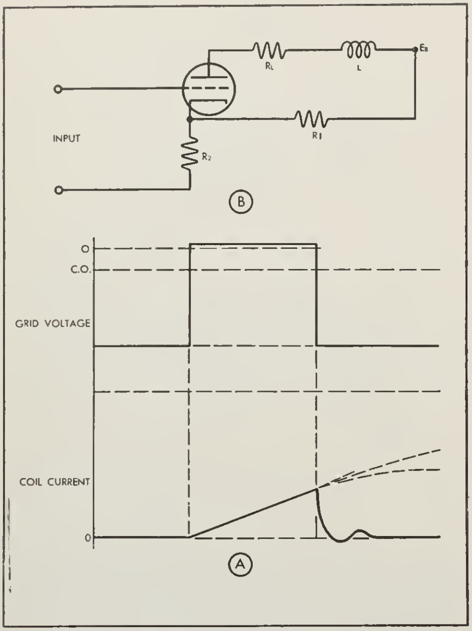
Square Wave Applied to LR Circuit

The one-megohm potentiometers in the plate circuit varies the amplitude of the output sawtooth wave. They are adjusted so that the sweep line on the CRT screen will extend from the top to bottom.

Trapezoidal Waveshape Generator

The sawtooth sweep generator previously described is suitable for any electrostatically deflected CRT. However, you will find that the majority of the larger radar sets employ magnetic deflection instead in cathode ray tube circuits in which sweep voltages are applied to coils. In long range sweeps, the sawtooth voltage is adequate for a linear sweep, but on short ranges, a special sweep voltage with a trapezoidal wave form must be applied to overcome the effects of the LR formed by the sweep coils.

The deflection of the spot in a magnetic CRT depends upon a change in magnetic field strength. Since this magnetic field strength varies directly with the current in the deflection coils, any linear deflection of the spot requires a linear variation of the current through the deflection coil. A fairly linear current change occurs whenever a square wave is applied to an inductor if the first ten per cent of the time constant is used. As shown above at A, square wave e_a is applied to an LR circuit. (Remember that all deflection coil circuits are LR circuits because the coil wire and associated components introduce resistance in series with the actual inductance of the coil).



Linear Rise Due to Long Time Constant

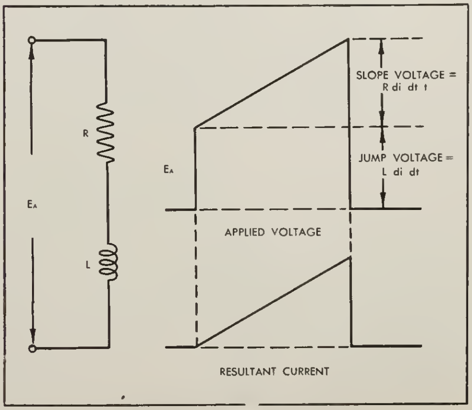
In this LR circuit the current change through the coil increases rapidly at first but then decreases when the voltage drop across the resistor decreases the voltage applied to the coil. During the first 10% of the rise, the increase in current is quite constant.

Making the L/R much longer than the time of the applied voltage produces a reasonable linear sweep. Referring to the simple circuit above at B utilizing a square wave input, you see a tube that is normally cut off by fixed bias. When the sweep is ready to start, the grid voltage is brought up to zero with respect to the cathode by a positive square wave. Plate current flows, increasing at a rate determined by the inductance and resistance. This resistance is the plate resistance of the tube and the resistance of the coil. The time constant of the LR circuit formed is long, and therefore the current rise is slow. The pulse on the grid ends before the current becomes very high, thus keeping the current increase in the linear portion of the curve. When the tube is cut off, the magnetic field collapses, causing an oscillatory transient. The transients can be prevented by a change in circuit or made harmless by blanking the spot during this time.

It is difficult to keep the series resistance small because the plate resistance of the vacuum tube involved is never small. Therefore shorter time constants (larger R in equation L/R) are accepted, and the applied voltage is changed to compensate for the large R. Referring to B on page 6-49, notice that if there were no resistance in this circuit, the current would increase indefinitely at the initial rate along the broken curve. This is because the fundamental relationship between the voltage and current in an inductor is $E = L\, di/dt$, where E is the applied voltage, L the inductance, di the change in current, and dt the change in time. L and E are constant and time changes linearly; therefore the current i must change linearly to make the equation true. The rate decreases along the exponential curve because the voltage drop across the resistor, due to this same current through it, subtracts from the applied voltage to leave less voltage across the inductor. Less voltage causes the current to increase at a lower rate.

The effect of the resistor may be eliminated if the applied voltage is increased at the same rate that the resistor voltage increases. Under these conditions, the subtraction of the resistor voltage maintains a constant voltage across the inductor, and the current increases at a constant rate.

TRAPEZOIDAL WAVESHAPES. The applied waveshape just described is called a trapezoidal wave. In the waveshape shown below. the initial instantaneous rise is called the *jump* voltage. It is this voltage which starts the current flow in



Trapezoidal Waveshape

the inductor. The slow increase is called the *slope* voltage. Its action is to continuously compensate for the drop across the resistor.

A quantitative analysis shows that there is a direct relation between the L/R, the jump voltage, and the slope voltage. In addition, since the trapezoidal voltage is produced with an RC circuit, there is a direct relationship between the L/R of the sweep amplifier and the RC of the sweep generator.

The jump voltage is the voltage which would be applied to a pure inductance. It is a constant value equal to L di/dt, where L is the inductive reactance and di/dt is the current change in amperes per second. As the current change also occurs in the resistor, you can use the Ohm's law formula E = IR to determine the voltage across the resistor. The facts are that R is constant and I is the same continuously changing current that is flowing through the inductor. Thus according to Ohm's law the slope of the resistor voltage becomes R di/dt.

Important facts to remember are that in the determination of the waveshape required for any sweep coil, the ratio of the jump voltage to the slope voltage is the most essential quantity and that the jump to slope ratio is equal to the time constant of the LR circuit. This relationship expressed mathematically reads,

$$\frac{jump}{slope} = \frac{L\ di/dt}{R\ di/dt} = \frac{L}{R}$$

To learn to use this relationship, study the following sample problems:

Examples

Problem 1. Assume the CRT and coils in the CRT assembly have been assembled and that as part of a radar set, a linear time base for a 50-nautical-mile range is required—that is, the spot must swing across the scope at a constant rate in 620 microseconds (the time of 50 radar miles of travel). Experiments with DC current indicate that a current of 100 ma causes the spot to move the required distance. Therefore, for a linear time base that covers the screen, the current in the coils must change at a constant rate from zero to 100 ma in 620 microseconds. The required current waveshape is shown at C. Further measurements indicate that the total resistance in the coil windings is 400 ohms and the inductance is 100 millihenries. The equivalent circuit of the coils is shown at B. What jump and what slope voltage is required to cause a perfectly linear sweep on the CRT?

Solution

The voltage across the inductor, E_L, is L di/dt.

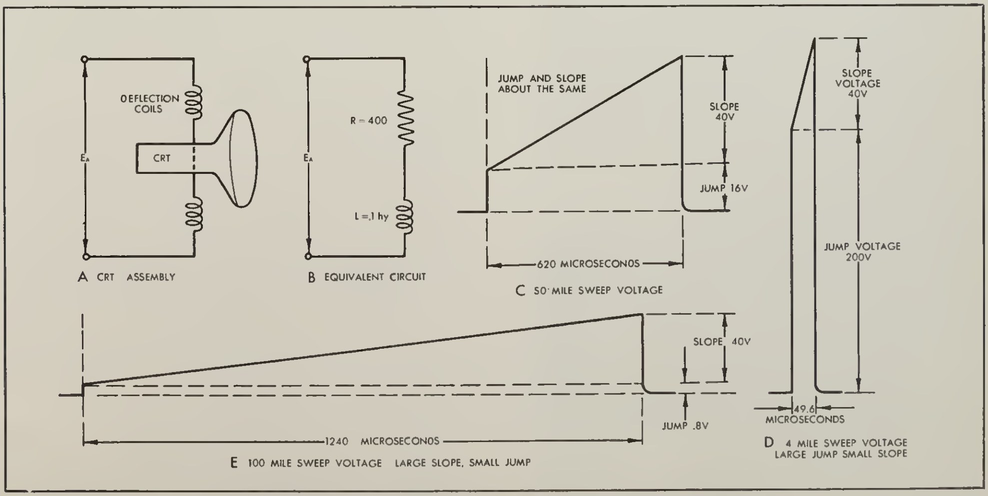
Therefore, $$E_L = \frac{10^{-1} \times 10^{-1}}{620 \times 10^{-6}}$$

$$E_L = 16.1 \text{ volts}$$

The resistor voltage slope is R di/dt

Therefore, it equals $\frac{400 \times 10^{-1}}{620 \times 10^{-6}}$, *or 6450 volts per second.*

The change in voltage across the resistor is at a rate of 6450 volts per second. Since the time base only lasts 620 microseconds, the voltage change will be 620 millionths of this value, or $6450 \times 620 \times 10^{-6}$, or 40 volts.



CRT Assembly and Waveshapes for 4-, 50-, and 100-mile Ranges

You can easily check this result as follows: The current at the end of the time base is 100 ma and is the final resistor current. The resistance is 400 ohms, therefore $E_R = IR = .1 \times 400$ or 40 volts. The required input waveshape, drawn to scale, is shown at C.

Problem 2. Notice the difference in the waveshape when the same sweep coil is used for longer and shorter sweeps. In each case the final current is 100 ma for full scale deflection; therefore, the final resistor voltage is always 40 volts (100 ma through 400 ohms). Find the jump voltage for a 4-mile range.

Solution:

The sweep time is 12.4 microseconds per mile. Therefore for 4 miles, it is 4 times that, or 49.6 microseconds. Substituting in the equation $E_L = L\, di/dt$, you get $\frac{10^{-1} \times 10^{-1}}{49.6 \times 10^{-6}}$, or 200 volts.

Since the 100 ma is a smaller percentage of the final current when 200 volts is applied than when 16 volts is applied, the rise is more linear and the slope voltage is a smaller proportion of the total voltage. This condition approaches the applied square wave shown in the illustration on page 6-49. When the precentage of slope voltage becomes somewhat smaller, it may be safely omitted without serious non-linearity in current rise.

Problem 3. Calculate the applied waveshape for the 100-mile range and find the final slope voltage, when the final slope voltage is still 40 volts across the 400 ohm resistor and the time is 1240 microseconds.

Solution:

Find the jump voltage, E_L, by the equation $E_L = L di/dt$.

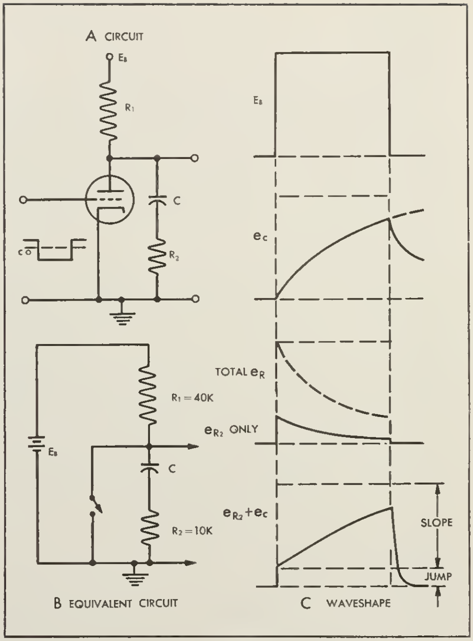
$$\text{Substituting } E_L = \frac{10^{-1} \times 10^{-1}}{1240 \times 10^{-6}} = .8 \text{ volt}$$

Since the current rise is very slow, only a small voltage is applied to the inductor. In fact, the jump voltage is so small that it may be omitted altogether on this long range without appreciable change of the waveshape. The applied waveshape becomes a sawtooth waveshape similar to that used for the electrostatic CRT.

The waveshapes for the 4-, 40-, and 100-mile ranges are drawn to scale at C, D, and E. Notice that the small jump voltage go with the long range and the large jump voltage go with the very short ranges.

Trapezoidal Sweep Generator

Where it is necessary to supply a trapezoidal waveshape to a sweep coil, a vacuum tube and RC circuit may be used to produce the desired waveshape. Diagram A on this page shows a simple vacuum tube circuit that will produce this kind of waveshape, and B the equivalent of the producing circuit. The triode serves only as a high speed switch. When the tube is cut off by the driving pulse, the plate voltage rises toward the applied voltage at the rate the condenser charges. However, some of the resistance of the RC circuit is between the plate and ground. The high initial charging current flows through this resistor and the condenser, raising the voltage immediately toward the applied voltage. In the curves, the condenser voltage and total resistor voltage curve are shown at C. If R_2 is one fourth the size of R_1, one fourth of the voltage will be across it at all times. The curve for the R_2 voltage is a miniature replica of the total resistor voltage. The output voltage is the sum of the R_2 voltage and the condenser voltage. This total voltage is the trapezoidal voltage required for the sweep coil circuit.



Trapezoidal Sweep Generator

As before, the two elements—jump and slope—form the waveshape in the sweep coil. The ratio of the jump to slope is L/R. To produce the same jump to slope ratio with an RC circuit, it is necessary that the RC circuit be designed so that R_2C is equal to the L/R of the sweep coil circuit. Mathematically, the relationship is expressed as $L/R = R_2C$.

This equation is derived as follows:

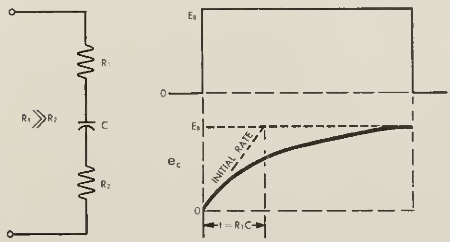
First, consider the jump voltage when the voltage E_B is applied to the circuit. The total voltage E_B appears across the resistors R_1 and R_2. The proportion of the voltage that appears across R_2 is expressed by the following equation:

$$E_{R2}=\frac{E_BR_2}{R_1+R_2}$$

If R_2 is much smaller than R_1, you can simplify the equation and get an approximate answer by dropping R_2 from the denominator. The equation now reads,

$$E_{R2}=\frac{E_BR_2}{R_1}$$

Second, consider the slope, being sure not to confuse slope with the final slope voltage. Slope is a rate of so many volts per second while final slope voltage is so many volts at the end of the rise. The initial charging rate of the condenser is such that it would charge the con-



Slope of Initial Charge

denser to the applied voltage in one time constant. If only a small part at the start of the curve is used, then the initial rate is the slope. Since the voltage increases at a rate of one E_B per time constant,

$$Slope=\frac{E_B}{(R_1+R_2)\ C}$$

When R_2 is quite small, you may disregard it for simplicity. Therefore, the slope equation reads,

$$Slope=\frac{E_B}{R_1C}$$

Using the simplified equation, you can express the relation of the jump to slope as follows:

$$\frac{jump}{slope}=\frac{\dfrac{E_BR_2}{R_1}}{\dfrac{E_B}{R_1C}}$$

$$=\frac{E_BR_2}{R_1}\times\frac{R_1C}{E_B}$$

$$\frac{jump}{slope}=R_2C$$

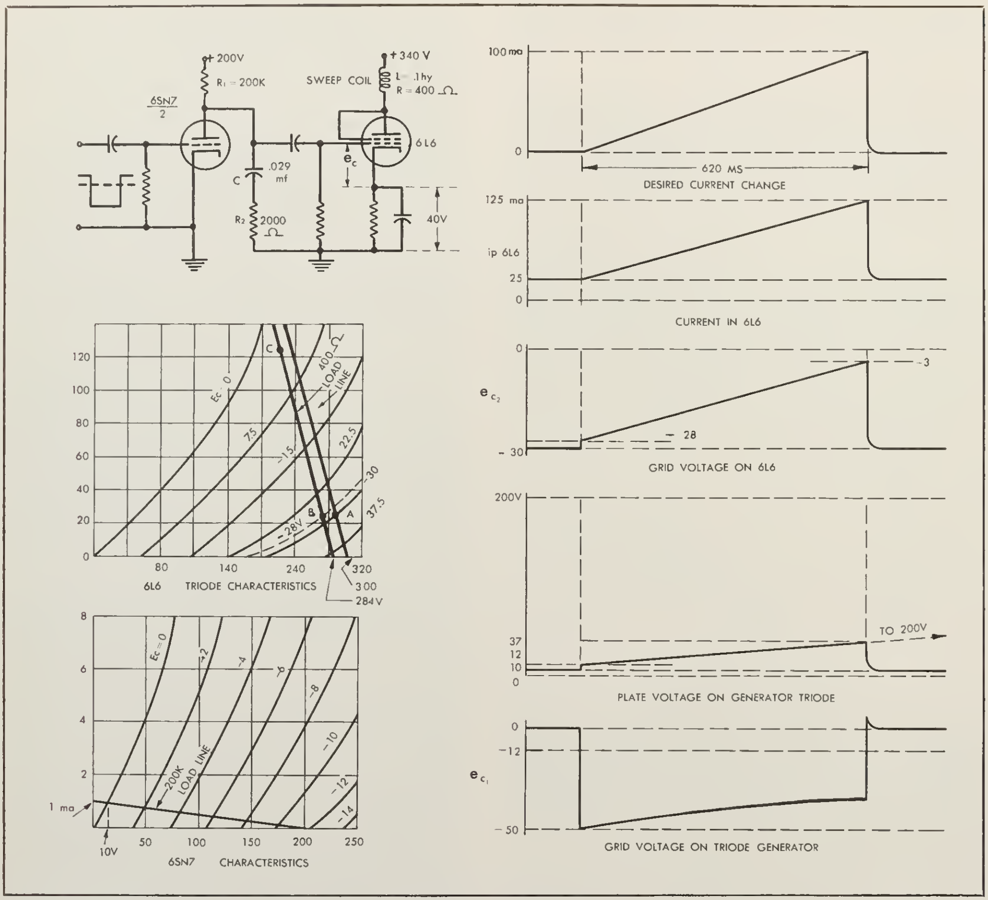
You know this is an approximation, but you can see that the jump to slope ratio is nearly equal to the product of R_2 times C. For linear current increase, therefore, the time constant of the deflection coil circuit (L R) should equal R_2C.

Design considerations in a trapezoidal generator. In the practical utilization of the principles just described, it is usually necessary to make the sweep coil a part of the plate circuit of a power amplifier tube in order to obtain the high currents with the desired waveshape. The trapezoidal waveshape is then generated at a low level and amplified by the power amplifier tube.

To see how the details are worked out, study the example of a complete generator and sweep coil circuit on page 6-54. The quantitative values worked out are for the coil used for producing the 50-mile sweep voltage described earlier. In the design of this circuit, you will find it convenient to start with the sweep coil and work backwards to the sweep generator. Therefore, the waveshapes which are shown and the analysis which follows are in reverse order.

The 50-mile sweep voltage coil has an inductance of 100 millihenries and a resistance of 400 ohms. A linear time base requires a linear current rise of 100 ma in 620 microseconds, which can be accomplished with a jump voltage of 16 volts (from previous calculations). Since this applied voltage is from the plate circuit and is due to grid voltage change, which is amplified in the tube, the required jump in grid voltage is only a fraction of the plate voltage change.

In the circuit illustrated, the sweep coil is in the plate circuit of a 6L6 power amplifier. The 100 ma current change actually starts from 25 ma instead of zero to take advantage of the more linear part of the tube characteristics. Notice the 6L6 tube characteristics. The load line is drawn for the 400-ohm resistance of the coil. The voltage on the grid between sweep voltages (first tube conducting) is plus 10 volts, but 40 volts bias makes the grid-to-cathode voltage minus 30 volts. Therefore the waveshape starts forming at point A on the load line. Then the sudden grid voltage change causes a jump of 16 volts in the plate circuit. The current increase in the coil generates 16 volts of back E. M. F., reducing the effective plate voltage by 16 volts. Thus a new load line is drawn from 300−16 or 284 volts. The grid voltage must change to point B on the load line for a 16 volts change in the plate. The current has just started increas-



Complete Generator and Sweep Coil Circuit

ing above 25 ma, so point B is at 25 ma plate current and 284 volts plate voltage. The grid voltage change is from −30 to −28 volts as read from these characteristic curves. Therefore, the jump required at the grid of the 6L6 is 2 volts.

After the jump, the slope occurs. The current must rise at a constant rate to 125 ma. At 125 ma the grid voltage, as read from the characteristic curves is −3 volts. This is point C on the characteristic curves. The final slope amplitude is 28−3 or 25 volts.

Note the difference in waveshape between the earlier circuit at C on page 6-51 and the grid voltage curve for the 6L6 above. The slope component is many times as large as the jump component at the grid while the slope and jump are almost the same as the circuit on page 6-51. The difference is due to the plate resistance of the 6L6 tube being in series with the coil. The additional resistance requires a much larger slope component.

The voltage change across the resistor R_2 and condenser C in the circuit must be the same as the required grid voltage change. The sweep generator components must be selected for this waveshape.

First consider the plate voltage required on the 6SN7. The slope must be linear, which means the charge must be limited to the first 10% of the charging curve. The voltage slope component has a final amplitude of 25 volts. The slope starts from 10 volts. If 100 volts is used as the applied voltage, 25 volts being 25% of 100 volts, which would be too nonlinear. Using 200 volts, and subtracting the 10 volts of original charge (plate voltage when the 6SN7 is conducting), 25 volts is 25/190 or 13.3% of full charge. This will be acceptable in this case because a higher voltage will raise the plate voltage, with the 6SN7 conducting, to more than 10 volts, which would require increasing of the 6L6 bias.

With 200 volts plate voltage, a load resistor must be selected which will hold the plate voltage at 10 volts while the 6SN7 is conducting. As shown by the load line on the 6SN7 characteristic curves, a 200 K resistor will accomplish this.

Next determine the condenser size. First, determine the condenser voltage by the formula:

$$e_c = E_B\,(1 - \epsilon^{-t/RC})$$

and then using the following given information:

$$e_c = 25 \text{ volts}$$
$$E_B = 200 - 10 = 190 \text{ volts}$$
$$t = 620 \text{ microseconds}$$
$$R = 200\text{ K}$$

Substitute in the equation,

$$1 - \frac{e_c}{E_B} = \epsilon^{-t/RC}$$

$$1 - \frac{25}{190} = \epsilon^{-620\times10^{-6}/2\times10^{5}C}$$

Therefore,

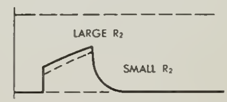
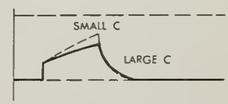
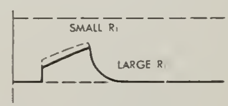
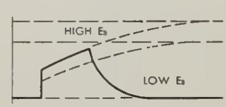
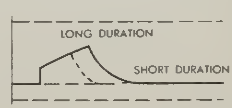
$$C = .029 \text{ mf}$$

The last item, the size of R_2, provides the jump voltage, which is 2 volts. You can calculate R_2 by transposing in the equation,

$$\text{jump} = \frac{E_B R_2}{R_1}$$

This gives, $R_2 = \dfrac{\text{jump} \times R_1}{E_B}$

Substituting, $R_2 = \dfrac{2v \times 200K}{200v} = 2000 \text{ ohms}$

CIRCUIT CHANGE	EFFECT ON JUMP VOLTAGE	EFFECT ON SLOPE AMPLITUDE	EFFECT ON SLOPE LINEARITY	EFFECT ON DURATION	WAVESHAPE
INCREASE R_2	INCREASED BECAUSE RATIO R_2/R_1 IS INCREASED	NO CHANGE RC IS STILL APPROXIMATELY THE SAME	NO CHANGE RC IS UNCHARGED	NO CHANGE DURATION IS DETERMINED BY GRID PULSE	
INCREASE C	NO CHANGE RATIO R_2/R_1 IS UNCHANGED	SLOPE IS LESS FINAL AMPLITUDE LESS BECAUSE RC IS LONGER, CHANGE IS SMALLER PERCENTAGE	IMPROVED BECAUSE SMALLER PART OF CHARGE CURVE IS USED	NO CHANGE	
INCREASE R_1	SMALLER RATIO R_2/R_1 IS LESS	GREATER BECAUSE RC IS SHORTER, C CHANGES TO HIGHER VOLTAGE	DECREASED BECAUSE GREATER PART OF CHARGING CURVE USED	NO CHANGE	
INCREASE E_B (PLATE SUPPLY VOLTAGE)	INCREASED DUE TO GREATER APPLIED VOLTAGE	INCREASED DUE TO HIGHER VOLTAGE CHARGE LIMIT	NO CHANGE SAME RC SAME DURATION SAME PERCENTAGE	NO CHANGE	
DURATION OF DRIVING PULSE INCREASED	NO CHANGE SAME APPLIED VOLTAGE, SAME R_2/R_1 RATIO	NO CHANGE SAME APPLIED VOLTAGE, SAME RC	DECREASED BECAUSE GREATER PART OF CHARGE CURVE USED	INCREASED BECAUSE INPUT PULSE IS LONGER	

Effect of Varying Elements in a Trapezoidal Sweep Generator

The pulse required to cut off the 6SN7 according to the characteristic curves must be more than 12 volts negative. The negative square wave (shown as e_{C1}), having an amplitude of 50 volts, is more than ample for this purpose.

In the cathode ray tube deflection coil circuit, a separate coil may be provided which will cancel the magnetic field caused by the residual 25 ma current in the 6L6 tube. This causes the magnetic field strength to be zero between sweeps even though 25 ma is flowing in the sweep coil.

In the chart on page 6-55 which summarizes the effect of varying the sizes of the component parts of the trapezoidal sweep generator, notice that varying R_1, the plate load resistor, affects the overall amplitude of the waveshape but has little effect on the jump-to-slope ratio. Varying the condenser size changes the slope only while varying R_2 changes the jump voltage only.

SQUARE WAVE GENERATORS

The circuits discussed thus far may be broadly classed as waveshaping circuits. Another class of circuits to consider is waveshape generating circuits. These circuits either produce a specified waveshape over and over again as long as power is supplied; or produce one wave when pulsed, and remain quiescent until another pulse is applied to the circuit; then they produce another wave. The output waveshape is the same regardless of the waveshape or amplitude of the initiating or "trigger" pulse. Circuits of this type are called "multivibrators" or "relaxation oscillators." Those which run continuously with only power applied are called *free-running* multivibrators. Those which must be started for each cycle are called *one-shot* or *start-stop* multivibrators. The frequency of free running multivibrators, as with other types of oscillators, is influenced by all voltages and component sizes associated with the circuits. Therefore, although such circuits have only fair frequency stability when free running, adequate stability is readily achieved from a reliable AC synchronizing voltage.

These circuits can produce sawtooth waveforms (thyratron sweep generator), triangular, or square waveforms. The square waveshape produced can be differentiated to produce accurately spaced peaked waveshapes. The duration and amplitude of any of the waveshapes

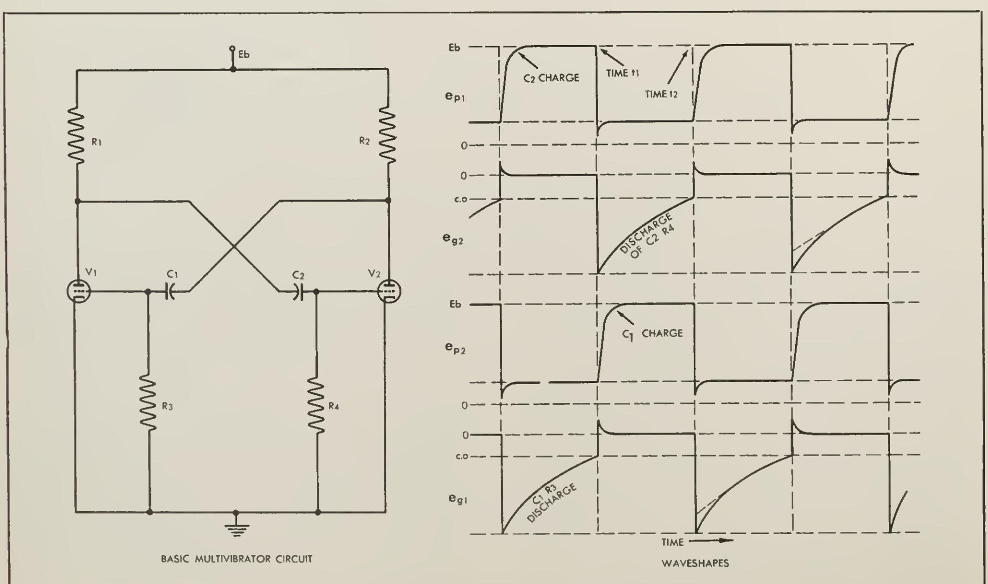


Plate Coupled Multivibrator

are readily varied. Time delayed pulses can be produced by triggering a multivibrator with a peaked wave, letting it go through a cycle and obtaining a new peaked wave somewhat later from the square wave producing it.

The Plate Coupled Multivibrator

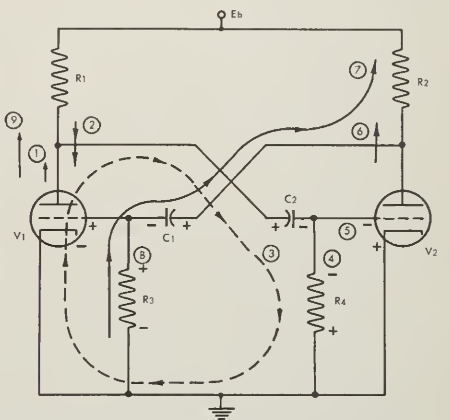
The plate coupled multivibrator is the simplest of the multivibrators. It produces almost a square wave when properly designed, and a perfect square wave when additions are made to the basic circuit. Note that basically this multivibrator consists of two triode amplifier stages, where the output of the second stage is introduced back into the input of the first. The phase inversion through the circuit is 360 degrees and any slight change in the first tube will be amplified through the circuit and cause a greater change in the same direction in the first tube. That is, feedback is "in phase" and the circuit will oscillate. There are no elements in the circuit to limit the feedback to any one frequency, so feedback occurs at all frequencies. Hence, a square wave, containing several hundred frequencies, is produced.

Looking at the action from the point of view of transients—that is, following each voltage change step by step—there are four different conditions that exist during one cycle of operation. They are as follows:

1. An extremely rapid change from V_1 conducting to V_2 conducting.
2. A long period during which V_1 is cut off and the circuit is relaxed.
3. A second violent change as V_1 conducts driving V_2 beyond cutoff.
4. A long period during which V_2 is cut off and the circuit is relaxed.

The cycle repeats as condition one follows condition four. The rapid changes just described are indicated by the vertical parts of the waveshapes at the left. These changes ordinarily occur in a fraction of a microsecond, and are so fast that the spot on a CRT becomes invisible when the waveshape is observed on an ordinary oscilloscope. The long relaxation periods are represented by the nearly constant voltages between switchover times.

OPERATION. When the filament voltages are on and the plate voltage is applied, current flows in both tubes. If the circuit components are the same in both tube circuits and both tubes are of the same type, the circuit is said to be

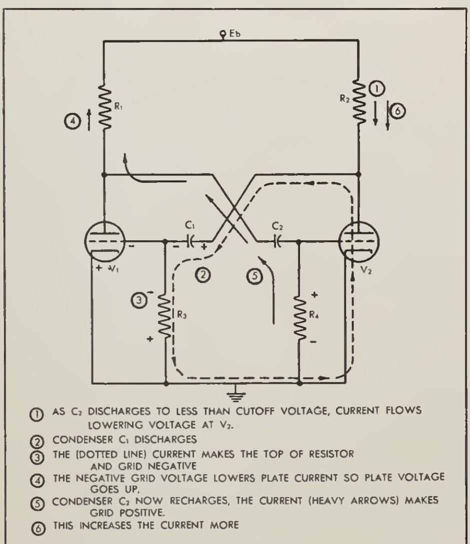


Sequence of Changes in Multivibrator.

symmetrical, and both currents should be equal. But it is impossible to obtain a perfect balance in the circuit, due to the nonuniformity of available parts and the irregular way in which emission occurs from the cathode. Any slight change in current in one tube that is not accompanied simultaneously by a similar change in the other tube, will start oscillation.

Assuming that a slight current increase has occurred in the tube V_1, the following series of events then occur in rapid order: The increase of current in V_1 increases the voltage drop across R_1 which lowers the plate-to-ground voltage at V_1. This lower voltage is now less than the charge on C_2, and the condenser C_2 discharges toward the new plate voltage. The discharge current flows through R_4 making the grid end negative. The negative grid of V_2 decreases the current on V_2. This reduces the drop across R_2, causing the plate to ground voltage at V_2 to be higher. Condenser C_1 is between this plate and ground and therefore must charge to the new higher voltage. The charging current flows through R_3 toward the plate supply voltage as shown by solid arrows on page 6-57. This current causes the top of R_3 to be positive, and therefore the grid of V_1 to be positive. The positive grid causes the original slight increase current to become a tremendous current increase. As the changes go around the circuit, a great voltage drop occurs at the plate of V_1, the grid of V_2 is made more negative, current in V_2 is decreased more, plate voltage at V_2 goes up, making the grid of V_1 more positive and further increasing the V_1 plate current.

This is a cumulative process which continues until the current in V_2 is decreased to zero by a grid voltage far below cut-off. With zero current through V_2 the plate voltage becomes equal to the supply voltage since there is no voltage drop across R_2. C_1 quickly charges to this value by grid current flow in V_1. After that, current through R_3 is zero; so the grid-to-cathode voltage is zero. The plate voltage of V_1 remains at a constant low value while the grid voltage is zero; so the condenser C_2 continues to discharge to the value. While discharging, the current (dotted lines) holds V_2 cut-off. Everything is at a standstill in the circuit except for the slow discharge of C_2. This corresponds to condition 4 given before.



Second Switching Action

The slow discharge of C_2 continues and the voltage across C_2 becomes lower and lower. The voltage across R_4 continuously equals the C_2 voltage in accordance with Kirchoff's voltage law, so the R_4 voltage decreases. Upon reaching a voltage just less than the cut-off voltage for V_2, a slight current starts flowing through V_2. This ends the period of inactivity in the circuit as swift changes follow this slight current.

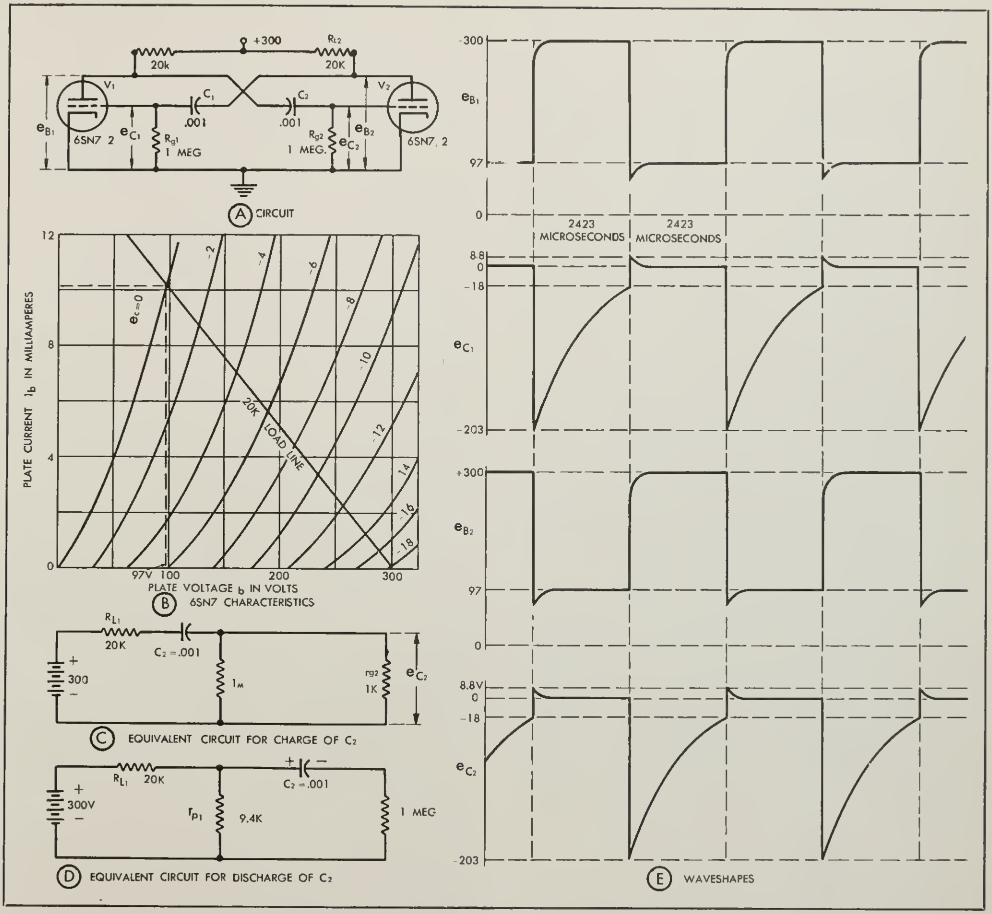
The current lowers the V_2 plate voltage. C_1 starts to discharge, making the grid of V_1 negative. The resultant decrease in plate current raises the V_1 plate voltage. C_2 now recharges to the new V_1 plate voltage. The C_2 charging current makes grid of V_2 positive and the original tiny current is increased tremendously. This regenerative process continues until V_1 is cut off by the discharge of C_1. At this point, V_1 plate voltage rises quickly to the plate supply voltage, C_2 charges swiftly to supply voltage through V_2 grid current, and the plate voltage of V_2 steadies at a low value. This ends the rapid changes during condition one.

The second inactive period occurs as V_1 is held nonconducting by the slow discharge of C_1. The discharge continues until the C_1 voltage is just below cut-off for V_1. This period is as described for condition 2.

Then the slight current in V_1 repeats the first rapid change described above. In the waveshapes in the basic circuit diagram on page 6-56, the long period between time t_1 and t_2 is the time that V_2 is cut off and the exponential grid voltage decreases of e_{g2} shows how long the condenser C_2 can hold V_2 nonconducting. Time t_2 is the rapid change just described while the half cycle following that is the time V_1 is cut off.

In summary, the operation of this circuit consists of long periods of time when one tube conducts a high current while the other tube is cut off, followed by an extremely rapid change to the other tube conducting and the first tube cut off.

The slow discharge of the condensers is due to the long time constant of the discharge circuit. In the waveshape in the basic circuit diagram on page 6-56, the grid voltage is shown as going positive for short periods of time. This occurs when the other plate is rising toward supply voltage. The condenser is recharging, making the top of grid resistor positive. The resultant



Quantitative Analysis of Plate Coupled MV

positive voltage on the grid makes the grid act as a diode plate, and part of the emitted electrons flow to the grid and into the condenser. This ready source of electrons charges the condenser quickly. That is, the grid-to-cathode equivalent resistance is very small, forming a short RC during the condenser charge. The positive going grid causes an extra dip in plate voltage, which makes the grid of the other tube somewhat more negative than normal. The effect if the grid did not become positive is shown by dotted lines in the waveshapes in the basic circuit diagram.

QUANTITATIVE ANALYSIS. It is fairly easy to calculate the frequency and approximate voltage throughout a given circuit of this type. Consider, for example, the circuit shown above at A. A plate voltage of 300 volts is applied to a 6SN7 dual triode with the circuit components shown. The plate voltage of V_1 varies between the value due to zero grid voltage and the value due to zero plate current. You can determine these values from a 20K load line on the 6SN7 characteristic curves. The load line crosses the zero grid voltage line at a plate voltage of 97 volts. At zero plate current, the plate voltage is 300 volts. From this information you can construct the plate voltage waveshape.

All this plate voltage change occurs across the grid resistor on V_2 at the first instant. The resistor voltage is the grid voltage for V_2; so the grid will swing negative by the same amount that the plate voltage drops. The plate voltage drops 300−97 or 203 volts. The grid is at zero when the plate voltage drop occurs, so the grid immediately goes 203 volts negative. Then the condenser starts discharging down to 97 volts. The discharge will be exponential. The equivalent circuit during the discharge is shown at D. The voltage across r_{p1} (the tube's plate resistance) is 97 volts and the condenser is discharging from 300 volts to 97 volts. The plate resistance r_{p1} is found by using the plate current and voltage when the grid voltage of V_1 is zero.

$$r_{p1} = e_b/i_b = 97/10.3 \text{ ma} = \text{about } 9.4K.$$

In parallel with this (through the battery) is the 20K plate load resistor. This parallel combination will have about 6K resistance in series with one megohm. For simplicity the 6K resistance can be disregarded. This will introduce an error of less than one per cent. The discharge time constant without the 6K resistor will be .001 mf x 1 meg = 1000 microseconds.

The discharge continues until the resistor (and grid) voltage at V_2 is just above cut-off. Cut-off voltage for this tube with 300 volts applied is −18 volts as read from the characteristic curves. The cut-off period is terminated when the voltage drops to 18/203 or 8.86%. The resistor voltage is down to 8.86% in 2.423 time constants. This is calculated from the exponential equation for resistor voltage during discharge,

$$e_R = E_B\, \epsilon^{-t/RC}$$
$$18 = 203\, \epsilon^{-t/10^6 \times .001 \times 10^{-6}}$$
$$.0886 = \epsilon^{-t/1000\ MS}$$
$$t = 1000 \times 2.423$$
$$t = 2423 \text{ microseconds}$$

The cut-off time for tube V_2 will be 2423 microseconds. Since the resistor and condenser values are exactly the same, tube V_1 will be cut off for 2423 microseconds during the other half cycle of operation. The grid and plate voltage values and curves will also be identical. If a difference exists in the circuit of each tube, separate calculations must be carried out for each half cycle. In this case, 2423+2423=4846 microseconds. The frequency at which square waves are produced is the reciprocal of the duration.

$$F = 1/4846 \times 10^{-6} = 206 \text{ cycles per second.}$$

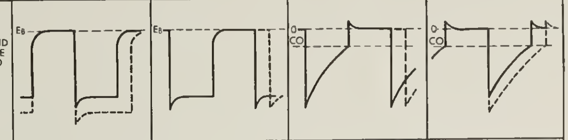
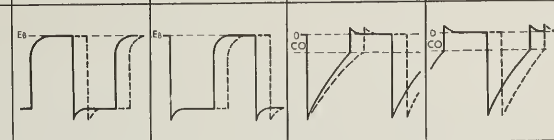
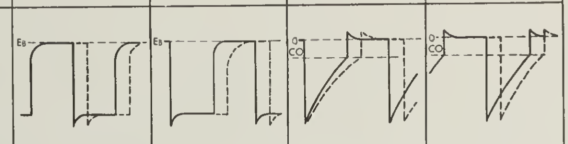
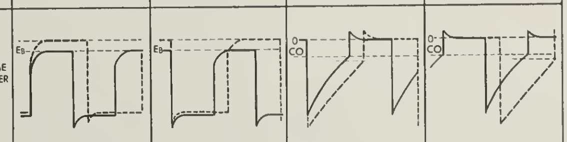
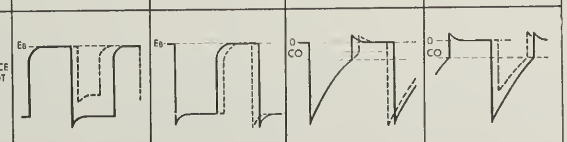
The duration and amplitude of the positive peak on the grid waveform can be readily calculated from the equivalent circuit shown at C in the illustration. The positive peak occurs as the condenser is charged to a higher plate voltage. The condenser charges through an R consisting of the plate load resistor for V_1 and the grid-cathode resistance of V_2. This is a total of 21 K. The one megohm resistor carries only .1% of the charging current and is therefore safely disregarded. With the .001 mf condenser, the time constant is .001 x 10^{-6} x 2.1 x 10^4, or 21 microseconds. If the charge requires 5 time constants, the positive peak will disappear after a hundred microseconds. The amplitude of this peak can be determined by using Ohm's law. The total voltage rise at the plate of V_1 is 203 volts. This raises the grid voltage from cut-off of −18 volts to some positive value. Since the first 18 volts of plate voltage rise change the grid voltage to zero, the remaining rise above zero is 203−18, or 185 volts. This causes grid current to flow and the voltage to divide across the series combination consisting of the grid-cathode resistance and the plate load resistance. As the grid-cathode resistance equals 1000 ohms, and the plate load resistance equals 20,000 ohms, 1/21 of the 185 volts, or 8.8 volts, appears across the grid-cathode resistance. Thus the positive peaks at the grid will be 8.8 volts above zero. Because the plate voltage cannot rise faster than the condenser can charge, the plate voltage rise is exponential over a period of 100 microseconds. This rounds off the leading edge of the plate waveshape. The positive grid at the other tube causes a dip in plate voltage at that tube.

The output from this circuit is normally taken from either plate although the grid voltage is sometimes useful.

In the chart showing the effects of varying the sizes of the component parts of this circuit, the essential data is given, but the detailed and less important effects are omitted because of a lack of space. The effects of changing R_{L2}, R_{g2}, C_2, and V_2 are not listed since these effects are similar to those that result from changing R_{L1}, R_{g1}, C_1, and V_1 respectively—except that each effect occurs with the other half cycle of operation.

Synchronization of the Multivibrator

Free-running multivibrators are generally not used as such in radar circuits, because their frequency stability is poor. To avoid this frequency instability, multivibrators are usually synchronized with another frequency which

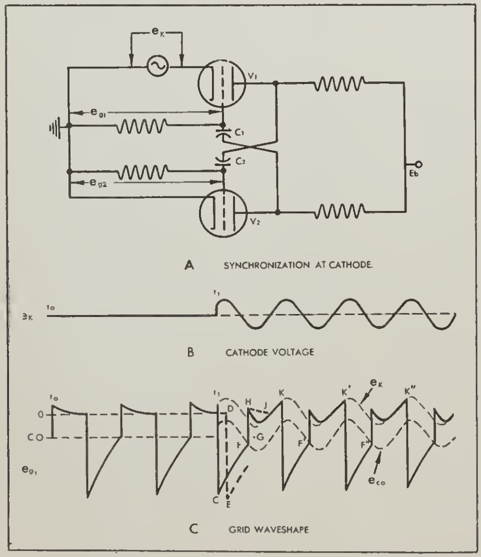
CHANGE	FREQUENCY	AMPLITUDE OF e_{b1}	DURATION e_{b1} AT E_b	AMPLITUDE e_{b2}	DURATION e_{b2} AT E_b	WAVESHAPE CHANGE (——— BEFORE CHANGE; - - - - - - AFTER INCREASE): e_{b1}, e_{b2}, e_{g1}, e_{g2}
R_{L1} INCREASED	DECREASED BECAUSE V_2 CUT OFF LONGER	INCREASED BECAUSE e_{b1} IS LOWER DURING ZERO GRID VOLTAGE TIMES	UNCHANGED	UNCHANGED	LONGER SINCE GRID IS MORE NEGATIVE AT START OF C.O TIME	
R_{g1} INCREASED	LOWER BECAUSE V_1 CUTOFF LONGER	UNCHANGED	LONGER SINCE TUBE IS CUT OFF LONGER	UNCHANGED	UNCHANGED	
C_1 INCREASED	LOWER BECAUSE V_1 CUT OFF LONGER	UNCHANGED	LONGER	UNCHANGED	UNCHANGED	
E_b INCREASED	LOWER BECAUSE BOTH GRIDS ARE DRIVEN FARTHER BELOW CUT OFF WHILE THE CUT OFF VALUE ONLY LOWERS SLIGHTLY	INCREASED BECAUSE E_b GOES UP A LOT WHILE THE e_{b1} VALUE FOR ZERO GRID VOLTAGE GOES UP ONLY SLIGHTLY	INCREASED BECAUSE V_1 IS CUT OFF LONGER	INCREASED AS E_b GOES UP A LOT WHILE THE e_{b2} VALUE FOR ZERO GRID VOLTAGE GOES UP ONLY SLIGHTLY	INCREASED BECAUSE V_2 CUT OFF LONGER	
V_1 REPLACED WITH TUBE OF HIGHER Mu	HIGHER	DECREASED BECAUSE HIGH r_p OF V_1 CAUSED e_{b1} FOR ZERO GRID VOLTAGE TO BE MUCH HIGHER	SLIGHTLY LONGER SINCE CUTOFF FOR V_1 IS LOWER	UNCHANGED	DECREASED CONSIDERABLY SINCE GRID OF V_2 IS NOT DRIVEN VERY FAR NEGATIVE	

Effects of Varying Elements in Multivibrator

forces the period of the multivibrator oscillation to be exactly the same as the period of the synchronizing frequency. Such a multivibrator is said to be *driven* by the synchronizing voltage.

SYNCHRONIZATION WITH A SINE WAVE. Although a waveform of almost any shape may be used for synchronization, a sine wave or a pulse is generally used. The circuit of a multivibrator which is synchronized by the injection of a sine wave voltage in the cathode circuit of one tube is shown below at A. The actual grid-to-cathode voltage of V_1, which is the voltage controlling the flow of plate current, is the difference between the grid-to-ground voltage e_{g1} and the cathode-to-ground voltage e_k. The source of sinusoidal voltage should have a low internal impedence, for otherwise the plate current of V_1 flowing through this source causes a voltage drop which alters the sinusoidal shape of the wave. Sometimes a low-voltage winding on a power-supply transformer, such as a filament winding, is used to supply this control voltage.

If the multivibrator at A in the illustration were balanced and running freely, the waveform of the voltage at the grid of V_1 would be that shown between times t_0 and t_1 at C in the illustration. Because there is no synchronizing voltage applied between these two times the cathode voltage will be constant at ground potential as shown at B. At time t_1, when the synchronizing voltage is suddenly applied, the cathode potential starts to vary sinusoidally. Although this variation does not affect the voltage on the grid with respect to ground, the grid-to-cathode potential contains this sinusoidal component of voltage. Therefore, the effective cut-off voltage of the tube varies sinusoidally about the normal value in phase with the synchronizing voltage on the cathode. The cathode voltage e_k and the effective cut-off voltage e_{c0}, shown at C, explain the synchronizing action by showing that the instant at which V_1 is made conducting occurs when the e_{g1} curve crosses the e_{c0} curve.



Synchronization with Sine Wave on Cathode

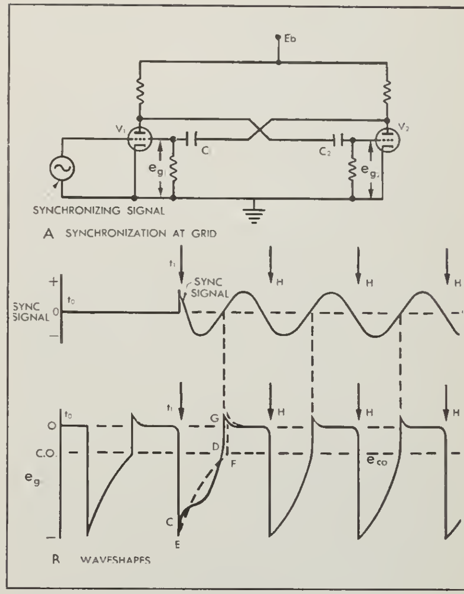
At time t_1 the voltage of the cathode starts to rise, decreasing the conduction of V_1. The positive-going voltage produced at the plate of V_1 initiates the switching action, and the tube is quickly cut off. Thus e_{g1} drops along t_1 instead of DE as it would in the free-running condition. Capacitor C_1 discharges exponentially along curve CFG. Since this curve intersects the e_{c0} curve at F, the switching action, by which V_1 is made conducting and V_2 is cut off, takes place at this time instead of at a short time later. Thus the switching takes place at time F instead of at G as would be the case in a free-running multivibrator. The switching drives the grid of V_1 positive, but the grid current drawn, quickly charges C_1 so that the grid returns to cathode potential. If e_{g1} followed curve HJ as it would if the multivibrator were free running, the grid would draw current because the synchronizing voltage causes the cathode to be negative relative to ground at this time. Therefore e_{g1} follows the cathode voltage along curve HK. When the cathode voltage begins to rise, the plate current through V_1 starts to decrease. By the time K the rise in voltage at the plate of V_1 resulting from this decrease in plate current has become large enough to drive V_2 into conduction, and the tubes are very rapidly switched.

The cycle of the multivibrator is forced to be somewhat shorter than the length of the free-running cycle by the action of the sine wave. Thus, switching in one direction occurs at points F, F′, F″, etc., and switching in the other direction occurs at K, K′, K″, etc. The period of the multivibrator, KK′ or F′F″, is seen to be equal to the period of the sine wave after the short transition interval, therefore, the multivibrator may be said to be synchronized. A sine wave can thus be used to control the frequency of a multivibrator. The synchronizing voltage can make the multivibrator operate either above or below its natural frequency. However, if an attempt is made to pull the multivibrator to higher and higher frequencies, a limit is reached beyond which the multivibrator synchronizes to one-half of the driving frequency. Similarly, the multivibrator may synchronize to one-third or a smaller fraction of the driving frequency and *frequency division* may be obtained.

When the sine wave synchronizing voltage is applied to the grid instead of the cathode of one of the tubes of the multivibrator as at A, the synchronizing voltage as shown adds directly to the grid voltage. Consequently, the grid-voltage waveform does not resemble that of a free-running multivibrator. Without the synchronizing signal, the multivibrator is free-running and the grid waveform appears as shown between t_0 and t_1 at B in the illustration. The synchronizing voltage is applied slightly before time t_1 in the random phase shown but, because of grid limiting, it has no effect until V_1 cuts off at its natural time t_1. Between C and D the synchronizing voltage adds to the normal discharge voltage which would rise along curve EF, and results in the distorted curve shown. Since the synchronizing frequency is higher than the natural frequency of the multivibrator, the distorted curve reaches cut-off sooner than curve EF.

When V_1 is conducting during the time between G and H, grid limiting occurs and prevents the synchronizing voltage from adding to the grid voltage. However, when the synchronizing signal drops below zero voltage and starts to decrease the voltage on the grid, the multivibrator regenerative action rapidly cuts off V_1. Since this action occurs at the same time, H, on each cycle of the synchronizing signal, the multivibrator is forced to operate at the synchronizing frequency. This latter action occurs in this way only when the synchronizing frequency is higher than the natural frequency of the multivibrator.

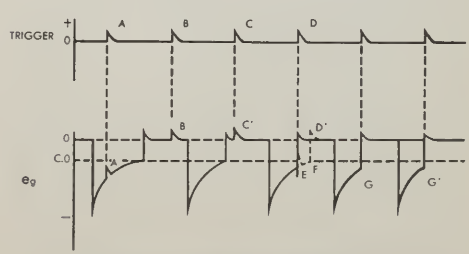
The first cycle of the synchronizing signal between t_1 and H will not, in general, be the same shape as the steady-state waveshape. The shape of this cycle depends on the time and phase at which the synchronizing signal is applied. The waveshapes at B show only one of many possible



Synchronization with Sine Wave on Grid

waveshapes for the first cycle. After one or two cycles, however, the multivibrator adjusts itself to a steady-state condition in which it operates so that the phase relation between the synchronizing signal and the multivibrator output is correct.

Synchronization with positive pulse. Although multivibrators can be synchronized with a sine-wave voltage, a more satisfactory method of synchronization is obtained by the use of short trigger pulses. These pulses may be either positive or negative. Note the effect of positive pulses on the multivibrator grid waveform. A positive pulse, such as at B or C, when applied to a tube that is already conducting, serves to cause only a momentary increase in current flow.



Synchronization with Positive Pulses on Grid

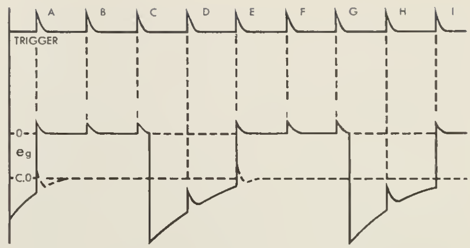
Thus, a positive trigger pulse applied to a conducting tube of a multivibrator has no effect on the action of the multivibrator.

However, when a positive trigger pulse is applied to a nonconducting tube and is of sufficient amplitude to raise the grid above cutoff, as the pulse at D, the tubes are switched as current starts to flow in the tube which was formerly cut off. If the trigger pulse occurs at a time such as A, the grid voltage will not rise to cut-off as shown at A′, and the switching action will not be started. *Thus a positive trigger pulse applied to a nonconducting tube in a multivibrator can cause switching action to take place only if the pulse is large enough to raise the grid above the cut-off voltage.*

With the exception of the trigger pulses, the grid waveform illustrated is that of a free-running multivibrator until time D. The positive pulses which occur at D drive the grid above cut-off, so that the cycle of the multivibrator is shortened by an amount equal to EF. In order for proper synchronization to take place, the period of the multivibrator must be greater than the interval between trigger pulses. Then the trigger pulses cause the multivibrator to switch earlier in the cycle than it would if free-running. As a result, the grid has not reached cut-off by the time G when the next trigger pulse is applied. Thus, the frequency of the multivibrator is forced to be the same as the repetition frequency of the trigger pulses, and the multivibrator will be switched consistently at time G after the transition is made from the free-running condition.

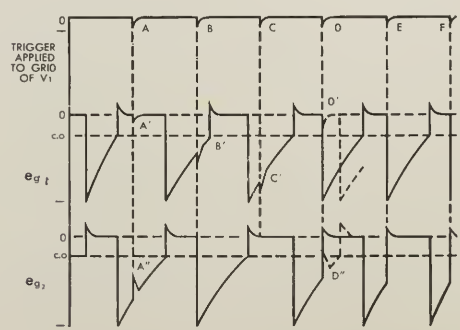
Synchronization with a submultiple of trigger frequency. A multivibrator may also synchronize to a submultiple of the trigger frequency. As shown, trigger A causes switching of the multivibrator but triggers B and C have no effect since they are applied to a conducting tube. Trigger D is applied to a nonconducting tube, but it is not large enough to start conduction so that the next pulse to cause a switching action is trigger pulse E. In the case shown, every fourth trigger pulse switches the multivibrator so that the repetition frequency of the multivibrator is one-fourth of the trigger pulses.

Synchronization with negative pulses. Negative pulses can be used to trigger a multivibrator as well as positive pulses. As shown, the first few cycles of the waveforms show the

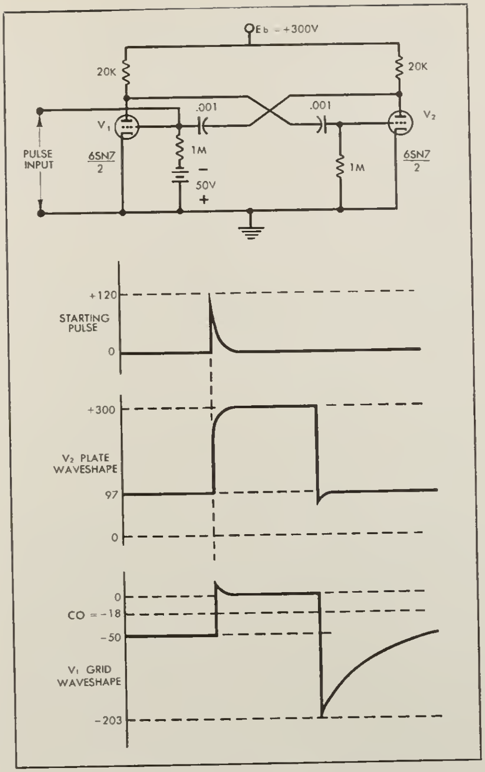


Synchronization at a Submultiple of Trigger Frequency

voltage variations that take place at the grids of a free-running multivibrator. As shown at B and B′ and at C and C′, *a negative trigger pulse applied to a nonconducting tube of a multivibrator has no effect on the operation of the multivibrator.* However, if the negative trigger pulse is applied to the conducting tube, that tube operates as a single-stage amplifier and applies a larger positive pulse to the grid of the non-conducting tube. The negative trigger pulse A reduces the grid voltage of V_1, as at A′, causing an increase in e_{g2}, as at A″. Since the change of e_{g2} is not sufficient to cause switching, trigger pulse A has no effect on the circuit. When a trigger pulse occurs in a later phase relative to the positive alternation of e_{g1}, as at D′, the amplified pulse applied to V_2, at D″ raises the grid above cut-off, exactly as if a positive pulse were applied directly. *Thus a negative trigger pulse applied to the conducting tube of a multivibrator can synchronize the multivibrator, provided that amplification by the conducting tube produces resultant positive pulse large enough to raise the grid voltage of the nonconducting tube*



Synchronization at Grid with Negative Pulses



One-Shot Plate-Coupled Multivibrator

to cut-off. Note that it is not necessary for the negative trigger to reduce the grid voltage of the conducting tube to cut-off.

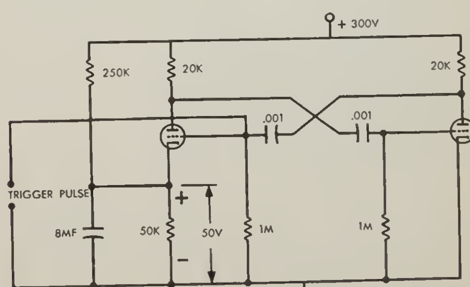
The synchronization voltage can be introduced at the plate as well as at the grid and cathode. The voltage has very little effect on the plate current of the tube to which it is introduced but has the usual effect on the *other* grid. The coupling condenser transfers the synchronizing voltage to the other grid. Therefore, introducing a pulse of a given polarity at the plate of one tube has the same effect at introducing the pulse with opposite polarity on the grid of the *same* tube.

The Driven Plate Coupled Multivibrator

You previously saw how the multivibrator circuit can be used either as a free-running or as a synchronized square wave generator. The circuit can also be used as a start-stop or one-shot multivibrator, if it is held normally inoperative and then a pulse applied to trigger or drive it into operation for one cycle and return to the inoperative condition.

The one-shot multivibrator circuit shown is the same circuit that you previously studied except that here the bias placed on the grid of V_1 is minus 50 volts. This bias keeps V_1 normally cut-off and permits V_2 to conduct continuously. However, the circuit will operate for one cycle if a voltage is applied to the grid with sufficient amplitude to raise the grid voltage above cut-off. Since cut-off is minus 18 volts, the required amplitude must exceed $50-18$, or 32 volts positive. Any voltage of more than 32 volts at the grid will cause V_1 to conduct. The resultant drop in plate voltage at V_1 is amplified around the circuit until V_2 is cut off, thus placing the plate of V_2 at supply voltage level. V_2 remains cut off for the usual time and then starts to conduct, driving the grid voltage at V_1 far below cut-off. Because the charge on the V_1 coupling condenser leaks off to minus 50 volts, which is still below cut-off, the cycle stops as V_1 can not conduct again. The illustration shows that a positive pulse is applied to the grid to start conduction. One positive square wave is generated at the plate of V_2 for the one pulse. Note in the grid waveshape at V_1 that the V_1 grid voltage starts and ends with minus 50 volts. In order for the circuit to produce another square wave, another pulse must be applied to the grid.

Other methods have been successfully used to keep the driven plate coupled multivibrator inoperative. The circuit below illustrates one method. It is the start-stop plate-coupled multivibrator. This circuit has an advantage over the circuit previously shown in that a separate



Start-Stop Plate-Coupled Multivibrator

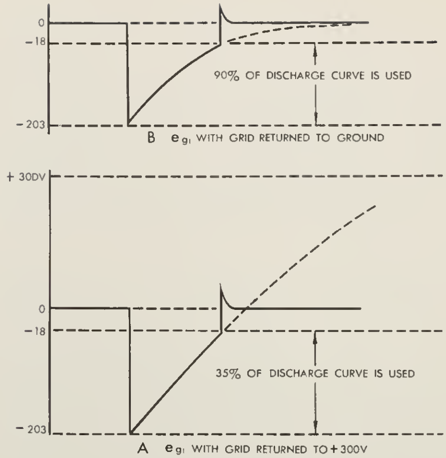
power supply is not required. It uses fixed bias by a voltage divider from plus 300 volts. With the values shown, the cathode will be 50 volts more positive than the grid. As cut-off is still at −18, 32 volts will, as before, be required to make the tube conduct. When conducting starts due to a pulse on the grid, the circuit will oscillate for one cycle, and then quit.

A negative pulse can be used to start a cycle in either of the last two circuits illustrated provided it is introduced at the plate of V_1 or the grid of V_2. This causes the negative pulse to be amplified and inverted by V_2 bringing the grid of V_1 above cutoff and thereby starting tube operation.

Improving the Basic Circuit

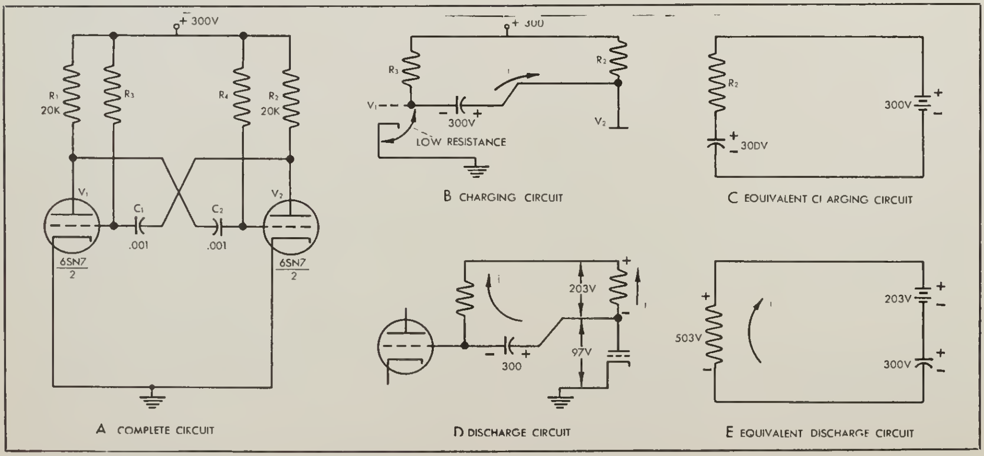
There are a number of slight changes that can be made in the basic plate-coupled multivibrator circuit to improve its stability and output waveshape. One is to return the grids to a positive voltage, and another is to use series limiting resisitors.

RETURNING THE GRIDS TO A POSITIVE VOLTAGE. Returning the grids to a positive voltage improves the stability of a plate-coupled multivibrator. In this arrangement, the grid leak resistors are often connected to the source of plate voltage, or the grids are returned to B plus. Under this condition the grid current flow causes the B plus potential to exist mostly across the resistor while the grid-to-cathode voltage remains within a fraction of a volt of zero. When



Improving Basic Circuit

the plate voltage of the other tube drops, the condenser discharges as usual, with a current that is high enough to drive the grid negative by the amount of the plate voltage drop. Due to this condition the exponential discharge of the condenser appears as though the condenser were charged to the plate voltage drop plus the plate voltage supply. The net result is that the curve shown above at A, resulting from the negative peak value to cut-off, is much more linear due to the fact that a smaller percentage of the discharge curve is used.

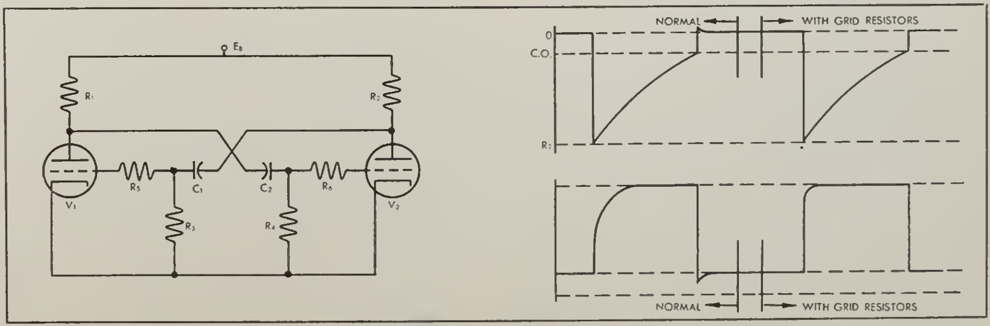


Effect of Returning Grid to B-plus

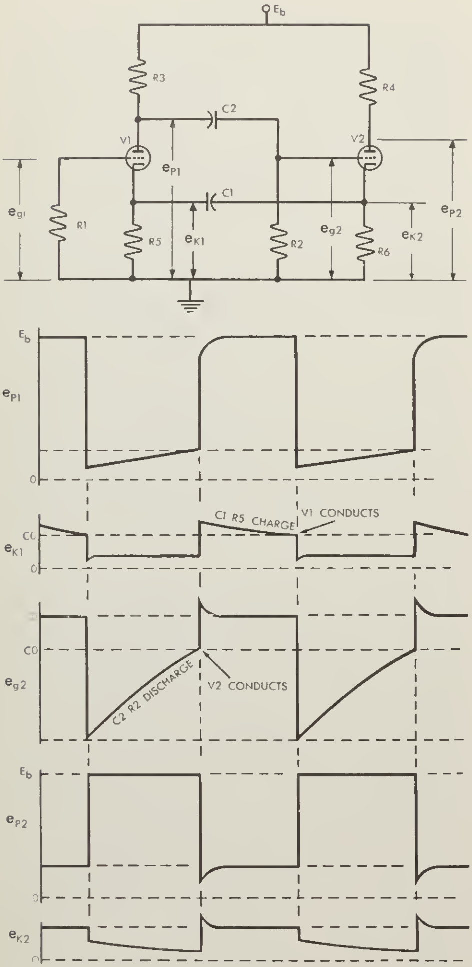
Returning the grids to a positive voltage produces an advantage in that the discharge curve strikes cut-off at a more vertical angle. This means that if variations of voltage or temperature in the circuit affects the relation of the waveshape to the cut-off value only a small change of time will occur. In contrast, study the grid voltage curve at B where the grid resistor is grounded. The discharge curve there is almost parallel at the cut-off value. In this case a small variation of the cut-off value will cause a large time change. In conclusion, then, returning the grids to B plus decreases the change in duration of pulses and the free running frequency for a given change in circuit constants.

You can further your understanding of the effect of returning the grid to B-plus by making quantitative analysis and studying equivalent circuits. At C, at bottom of page 6-66, notice the equivalent circuit during the time that V_1 is conducting. At B, the parts of the circuit involved during the charge of C_1 are shown. Since the grid-cathode resistance is very small, compared to the grid leak resistor, it can be represented as a direct connection, producing the equivalent circuit at C. Here R_2 is small and as a result the condenser will be fully charged before tube V_2 starts to conduct. Thus at the time V_1 is cut off, condenser C_1 starts with a charge of 300 volts. The discharge circuit for C_1 is shown at D. V_2 is conducting during this time, causing a voltage of 203 volts across the load resistor R_2. A closed circuit is formed as shown at E. The voltage across the load resistor may be considered as a source of voltage or battery which is of a polarity to aid the condenser in causing current flow. The total voltage across the resistor R_3 becomes the condenser charge plus the other voltage (300+203=503 volts). The grid end of resistor R_3 is 503 volts negative with respect to the end connected to 300 volts. As a result, the grid is 203 volts negative with respect to the cathode. Effectively, the condenser is charged by the plate voltage source. Since the polarity of the source is reversed, the plate voltage source will aid the discharge of the condenser. When a source is connected to a charged condenser with a polarity that discharges it, the condenser will completely discharge and recharge to the new voltage all in one long exponential curve. In this case, the curve will have the same amplitude as it would if it charged from zero to 503 volts. The resulting slope is as shown at B at top of page 6-66. Of course, the slope due to the more rapid discharge will go through zero sooner if the same RC is used. Therefore a longer RC must be used when grid resistors are returned to a positive voltage than is used when the resistors are returned to ground, for the same duration of output pulse.

USE OF SERIES LIMITING RESISTORS. The use of series limiting resistors in the grid circuits of this multivibrator increases the squareness of the waveshapes. In the circuit shown below, these resistors are R_5 and R_6. The first half cycle of the plate and grid waveshapes shown is without the grid-limiting resistors in the circuit, and the second half cycle is with R_5 and R_6 in the circuit. Note that the grid does not go positive when resistors are provided. When the other plate becomes more positive, most of the voltage change occurs across the grid-limiting resistor while only a small fraction of a volt exists between the grid and cathode, provided the grid-limiting resistor is of the order of 500K or more.



Using Series Limiting Resistors



Cathode Coupled Multivibrator

The plate that is going positive cannot normally rise in voltage any quicker than the coupling condenser can charge. So in the normal circuit, the plate waveshape resembles the coupling condenser charging curve. When grid-limiting resistors are inserted, the limiting resistor is in series with the low grid-cathode resistance. These two resistors and the grid leak resistor form a high resistance circuit for the condenser to charge through. The circuit is like the trapezoid sweep generator in that a "jump" voltage will occur across the limiting and grid leak resistor. Then the exponential rise will start from there. In the waveshape of this illustration, the jump is about 90% of the total amplitude, with only slight rounding of the leading edge of the pulse.

The Cathode Coupled Multivibrator

In the plate coupled multivibrator, feedback in the correct phase to sustain oscillation was obtained by connecting the plate of each tube to the grid of the other. Feedback in the correct phase can also be obtained by connecting the two triodes as shown in the cathode coupled multivibrator. The usual plate to grid connection in this case is made from V_1 to V_2, and the cathode of the second tube is capacitively coupled to the cathode of the first. If you follow through these changes, you can see that feedback is in phase. Assume that the current in V_1 is decreasing. As the plate becomes more positive, the V_2 grid will become more positive and current will increase in V_2. This makes the cathode of V_2 more positive. Since the cathode of V_1 is coupled to the V_2 cathode, cathode of V_1 will become more positive. Because a positive going cathode is like a negative going grid, the current in V_1 will decrease. Since the original change was a current decrease in V_1 the feedback is in phase to sustain oscillation.

Notice that in this multivibrator the cut-off time of the grid of the second tube is controlled by the discharge of C_2 through R_2, while the cut-off time for the first tube is controlled by the charge of the cathode coupling capacitor through the cathode resistor of the first tube.

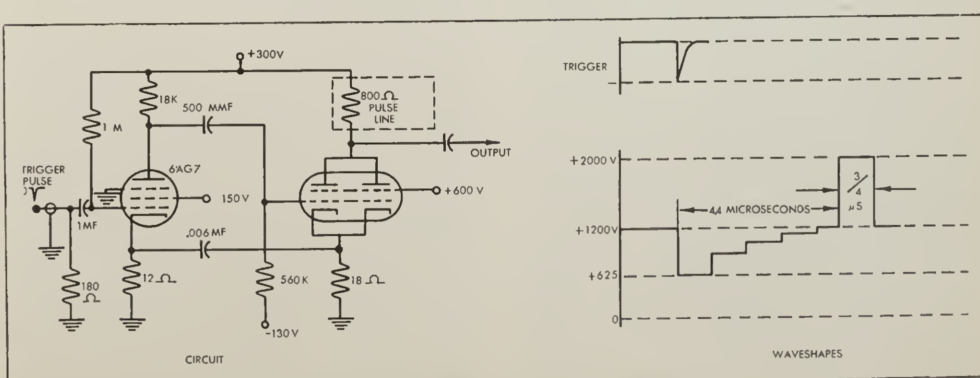
In order to understand the operation of this multivibrator, assume that V_1 is conducting and that V_2 has been cut off by an action similar to that just described. Capacitor C_2 discharges through R_2, R_p of V_1 and R_5, so that the voltage at the grid of V_2 decreases from a high negative potential toward ground as the discharge current through R_2 decreases. When e_{g2} rises to the cut-off voltage of V_2, the tube starts to draw current through R_6. The rising voltage produced across this resistor is coupled through C_1 to the cathode of V_1 since the voltage across the capacitor cannot change instantaneously. The positive-going voltage which is coupled to the cathode of V_1 adds to the existing voltage across R_5, reducing the flow of current in V_1. As i_{p1} decreases, the voltage at the plate of V_1 increases, and this positive-going voltage is coupled to the grid of V_2 to increase further the current in V_2. This action around the circuit is regenerative, and it ends with the current in V_1 reduced to zero and the current in V_2 at a maximum.

Capacitor C_1 is charging to the voltage existing across R_6 through the path R_5, R_p of V_2, and R_4 to the high-voltage supply. The charging current produces across R_5 a voltage which is more than sufficient to hold V_1 beyond cut-off. However, as the voltage across the capacitor rises, the magnitude of the charging current falls. Consequently, the voltage across R_5 decreases and at some time it will fall below the value required to cut off V_1. At this instant, a current starts to flow through V_1 lowering the voltage at the plate of the tube. This negative-going voltage is coupled through C_2 to the grid of V_2, causing i_{p2} to decrease. As a result of the reduction of i_{p2} and the decrease in e_{K2}, the voltage across R_5 decreases. Therefore, C_1 discharges through R_6, R_p of V_1, and R_3 in an effort to charge in the opposite direction to the voltage across R_5. The cumulative effect of the discharge of C_1 through R_6 and the negative-going voltage applied to the grid of V_2 combine to cut off V_2 almost instantaneously.

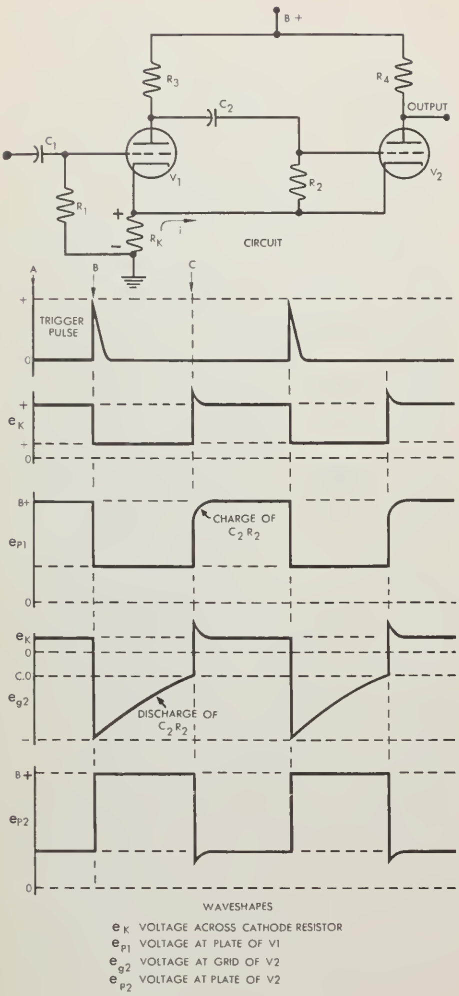
The waveshapes at the important points are shown in the illustration on page 6-68. While V_1 is cut off, a cathode voltage exists due to the charge of the cathode coupling condenser. The voltage is visible in the waveshape e_{k1}, as the high initial charging current causes a high cathode voltage. At a certain positive cathode voltage the grid-to-cathode voltage difference will be so great that no plate current will flow. The cathode voltage exceeds this value at first, but as the condenser charges, the resistor (R_5) voltage decreases exponentially until the cut-off value is reached. Upon dropping below this value, current flows in V_1, which by the drop in e_{p1} cuts off V_2. This discharge of C_2, for holding e_{g2} below cut-off, is the usual action as observed with the previous circuit. During the time V_2 is cut off, its plate voltage rises instantly to E_b, forming a square waveshape. There is no rounding of the corner because there is no condenser to charge. When V_2 is cut off, the cathode voltage tends to become zero, but it cannot be zero until the condenser is discharged. Therefore, the waveshape shows first a sudden drop as plate current ceases, then an exponential drop as the condenser discharges. The same discharge is indicated by the plate and cathode waveshape of V_1. The discharge is such that it makes the cathode of V_1 more negative, so the plate voltage drops below the zero grid voltage level, and a slight dip is seen in the grid voltage waveshape. As the discharge progresses, condenser voltage decreases and the effects of it decrease. The plate voltage gradually rises and a slight rise of cathode voltage is seen.

Generally speaking, increasing the size of any part in the circuit except R_1 will lower the frequency. The duration of each half cycle is dependent upon every part in the circuit. Changing any one value will affect the duration of both half cycles. Increasing the power supply voltage will also lower the frequency.

Because of the instability of this circuit when free running, it should be synchronized if it is to operate at a constant frequency or if its action is to be simultaneous with the action of some other circuit. Like the plate-coupled multivibrator, this circuit can be synchronized



Cathode Coupled MV in Radar Bombardment Set



Start-Stop Multivibrator

at either plate, grid, or cathode if the right polarity of synchronization voltage is used. If, for example, you should desire to start conduction in the second tube at a certain time, a positive voltage should be applied to the second grid, first plate, or first cathode at that time. A negative pulse is required at the first grid, second plate, or second cathode to do the same thing. This circuit can also be kept normally inoperative and be made to operate only once for each trigger pulse.

APPLICATION. A practical cathode-coupled multivibrator as used in a bombardment radar set is the one shown on the preceding page. The circuit is slightly simplified. Its purpose is to charge a pulse forming network, which then discharges, forming a positive square wave of exactly $^3/_4$ microsecond duration, at an amplitude of 800 volts. This pulse is amplified to 12,000 volts to become the plate voltage for the transmitter.

The circuit can be free running, but is normally a one-shot multivibrator because of a minus 130 volts that holds the second tube cut-off and a positive grid return that keeps the first tube conducting. When a negative trigger pulse is applied, it reduces the current in the first tube, after which regenerative amplification cuts the tube off completely, while the second tube conducts heavily. The time constants are such that the second tube conducts 4.4 microseconds, then is cut-off. A large amount of energy must be stored in the pulse forming network during this short time so a pair of very large tubes in parallel are used to provide a current of 2 amperes for the 4.4 microseconds. The network presents a resistive impedance to the circuit and is shown as an 800-ohm resistor in the diagram. With this high current, only 18 ohms of cathode resistance is required to cut-off the small pentode used as the first tube. The network discharges immediately, producing the pulse previously described. In some models a positive pulse follows the negative trigger pulse, which makes the switching action occur at exactly 4.4 microseconds after the start. The positive pulse is introduced at the grid of the first tube.

The Start-Stop Multivibrator

A widely used circuit, which never becomes free running and must always be triggered for each output pulse is the start-stop multivibrator. This circuit is often called a *flip-flop* from British slang or a *one-shot multivibrator*. As shown at the left, the circuit consists essentially of a two-stage resistance-capacitance-coupled amplifier with one tube cut off and the other conducting normally. The balanced condition of the circuit is established by the arrangement for biasing the tubes. The grid of V_2 is connected to its cathode through the resistor R_2, and the resultant voltage drop across R_k biases V_1 to cut-off. When V_2 is not conducting, V_1 cannot be cut off by the self-bias developed across R_k.

The following are the steps in the action of the start-stop multivibrator:

1. V_1 is cut off initially by the voltage drop produced across R_k by i_{p2}, the plate current of V_2 (time A).

2. V_2 is conducting heavily because its grid is at cathode potential (time A).

3. A positive trigger pulse (time B) sufficient in amplitude to raise the grid of V_1 above cut-off voltage, is impressed on the grid of V_1 through C_1.

4. V_1 begins to conduct and the voltage as its plate decreases. This decrease passes through C_2, as the voltage across a capacitor cannot be changed instantaneously, and appears on the grid of V_2 as a negative-going voltage.

5. The negative-going voltage on the grid of V_2 decreases i_{p2}.

6. The voltage drop across R_k decreases, allowing more current to flow in V_1.

7. The plate voltage of V_1 is still further decreased.

8. The grid of V_2 goes still more negative.

9. This action is repeated until V_2 is cut off and then V_1 starts conducting. The switch from V_1 to V_2 is practically instantaneous.

The circuit remains with V_1 conducting and V_2 cut off during the interval from B to C while C_2 discharges sufficiently toward the lowered value of plate voltage of V_1 to allow the grid of V_2 to rise from its lowest value to cut-off voltage.

10. V_2 begins to conduct (time C).

11. The plate current of V_2 flowing through R_k, raises the cathode voltage of V_1, thus reducing its plate current.

12. The decreased plate current of V_1 allows the plate voltage of V_1 to increase.

13. This increase is coupled to the grid of V_2, increasing still further its plate current.

14. The action described is repeated until V_1 is cut off and V_2 is conducting heavily. This action also is practically instantaneous.

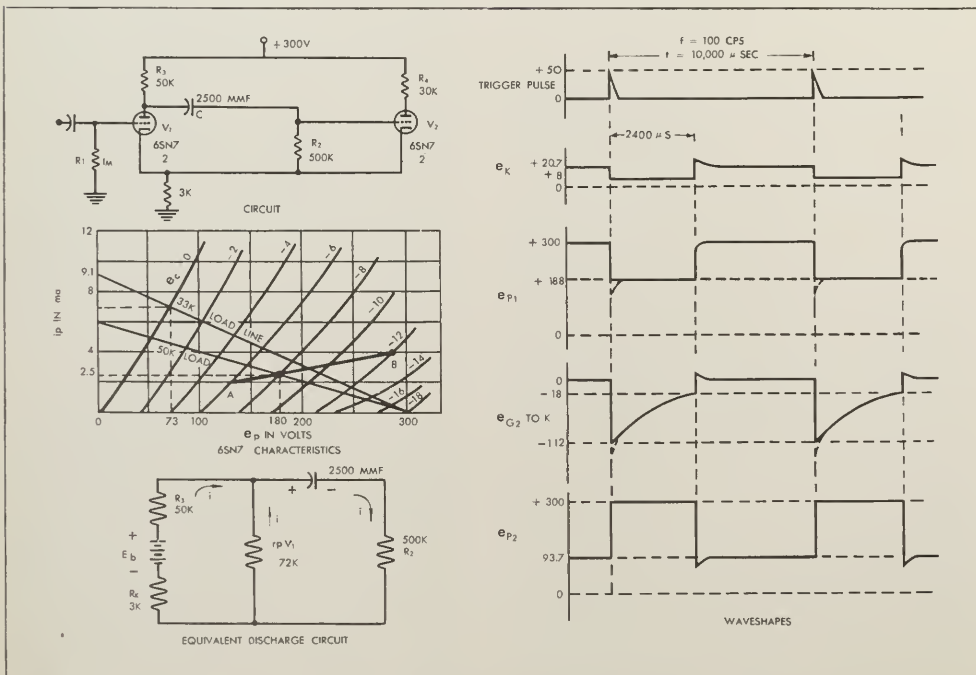
The circuit has now come back to its original balanced state and will remain so until another positive pulse arrives and causes V_1 to conduct. When applied to the input of the one-shot multivibrator, every positive trigger pulse which causes V_1 to conduct, results in a large positive-pulse output from the plate circuit of the second tube. The length of the positive output pulse produced at the plate of V_2 is controlled by the time constant of C_2 times R_2. If larger values of C_2 and R_2 are used, the length of the positive output is increased. A positive-output voltage pulse is produced for each positive-input trigger pulse.

QUANTITATIVE ANALYSIS. The quantitative determination of the amplitude and duration of the output pulse is not too difficult to accomplish. Study the illustration on the next page, as an example. The circuit shown is the most basic one. The waveshapes are similar to those in the illustration just preceding except that the voltages and time intervals have been determined and the waveshapes drawn to scale.

Start by finding the voltages in the circuit before the trigger pulse arrives. Check to see that V_1 is cut off with the 3K resistor. The circuit will not work if R_k is too small. V_2 is conducting due to zero bias. Now draw a load line on the characteristic curve to find the voltages. In drawing the load line, consider the cathode resistance. Any resistance in series with the tube will affect the current, but if the cathode resistor is quite small, its effect is negligible. You can obtain reasonably accurate results by disregarding the cathode resistance when it is less than 10% of the load resistance. When it is 10% or more, sketch the load line for the total cathode and load resistance. In the case of V_2 in this problem, the cathode resistance is exactly 10% of the load resistance; so the load line is drawn for 30K + 3K = 33K. Check the plate current and plate voltage at the intersection of the load line and zero grid bias. The plate voltage appears to be 73 volts and the current 6.9 ma. Knowing the current through R_k, find the voltage across it. ($E = IR$, $i_pR_k = .0069 \times 3000 = 20.7$ volts.) Cut-off is 18 volts, as read from the curves, so the first tube is actually cut off, and the circuit will function.

Now proceed with the calculations. The waveshape shown is the plate-to-ground voltage while the characteristic curves show plate-to-cathode voltage. So you must add the cathode voltage above ground to the plate voltage to find the actual plate-to-ground voltage. Since the cathode is 20.7 volts above ground, add this to 73 to get 93.7 volts at the plate of V_2. To simplify the picture, the waveshape at the grid of V_2 shows the voltage at the grid with respect to the *cathode* rather than ground. In the waveshape of cathode voltage shown, notice that before the trigger pulse arrives, the cathode voltage is 20.7 volts as previously calculated.

Now, determine the voltages in the circuit immediately after the trigger pulse and switch-over has occurred. During this time, the second tube is cut off so the plate voltage there is 300 volts. The first tube is conducting as much as self-bias with a 3K bias resistor will allow. To find this value, a load line and bias line are required. Since 3K is less than 10% of 50K, the load line can be drawn for 50K only. For the bias line, assume different values of current, find the bias, and plot the points on the characteristic curves. For example, if the current is 2 ma, the bias would be i_pR_k, or .002 x 3000, or 6 volts. To plot this, find the 2 ma current line and the 6 volts grid voltage curve (point A). Assume a second current of 4 ma. The bias would be 12 volts if 4 ma flowed through the 3K resistor. Mark the intersection of the 4 ma current line and 12 volts grid voltage line (point B). Draw a line through these points and the intersection of the bias line with load line is the operating point for the tube. According to the illustration, it appears that the bias or cathode voltage is 8 volts. The plate-to-cathode voltage is 180 volts. So the plate voltage has changed from 300 to 180+8 or 188 volts. The change is 112 volts in the negative direction. The condenser must discharge by this amount through the 500K grid resistor R_2. The new plate voltage at V_1 is shown on the E_{p1} curve. The grid voltage at V_2 goes from zero to negative 112 volts at the switch-over time, then decreases exponentially toward zero. This holds V_2 nonconducting until the voltage decreases below cut-off. From the curves, cut-off is −18 volts. In order to calculate the time of this exponential decrease, the RC must be known. C is given as 2500 MMF. The grid resistor of 500K can be considered as the total R for approximate calculations, but the current flows through the 500K resistor, then through two parallel paths in returning to the condenser. One path is through R_k, the power supply and R_3 in series. The other path is through the tube. Assuming the internal resistance of the power supply is zero, the first path has a



Quantitative Analysis of Start-Stop Multivibrator resistance of 53K. The other parallel path, through the tube, has a resistance which is somewhat variable due to bias change during the discharge, but is quite near the value determined by dividing the plate voltage by the plate current. From the curves, the plate voltage is 170 volts, the plate current 2.5 ma, both at the operating point with 8 volts bias. Since $r_p = e_{p1}/i_{p1}$, the tube resistance is 180/.0025, or 72K. This parallel combination (with a net resistance of 30.5K) is in series with the 500K resistor to form an equivalent circuit as shown in the equivalent discharge circuit. From this equivalent circuit you can see that the values for the equation to determine the time constant are as follows:

$$e_c = -18 \text{ volts}$$
$$E_B = -112 \text{ volts}$$
$$R = 530.5K$$
$$C = 2500 \text{ mmf}$$

Substituting in the exponential equation,

$$e_c = E_B\, \epsilon^{-t/RC},$$
$$-18 = -112\epsilon^{-t/530.5\times10^3\times2500\times10^{-12}}$$
$$t = 2400 \text{ microseconds}$$

The positive square wave at the second plate will have a duration of 2400 microseconds and its amplitude will be 300−93.7, or 206.3 volts.

The circuit switches over again at the end of 2400 microseconds. V_2 conducts, raising the cathode voltage to 20.7 volts, cutting off V_1. The plate voltage of V_1 goes to 300 volts as fast as the condenser can charge. The condenser charging current raises the grid of V_2 to a positive voltage momentarily and the cathode above 20.7 volts momentarily. This in turn causes a momentary dip in the plate voltage at V_2. The amplitude and duration of these momentary charges can be calculated by using the same procedure outlined for the plate-coupled multivibrator.

A trigger pulse of a large voltage will cause the dip shown by dotted lines in the waveshape of V_1 plate and V_2 grid. This dip is an amplified version of the trigger pulse and lasts no longer than the trigger pulse. It does not affect the discharge time of the condenser. If the pulse is of the minimum amplitude required to trigger this circuit, this dip will hardly be noticeable. Experience has shown that the trigger pulse should be of the lowest possible amplitude consistant with reliable operation.

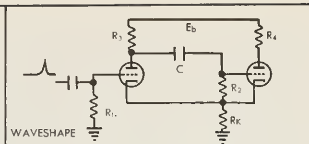
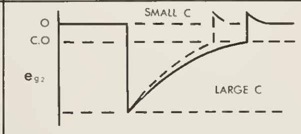
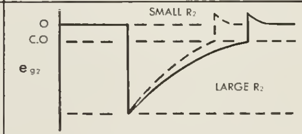
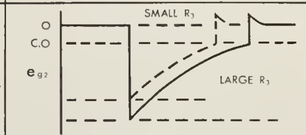
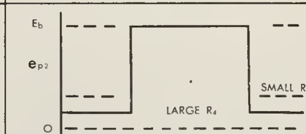
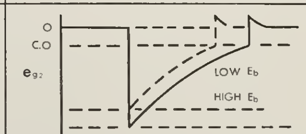
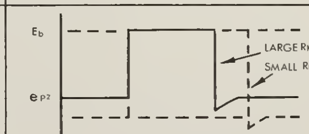
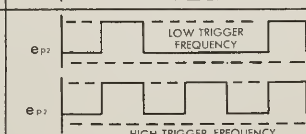
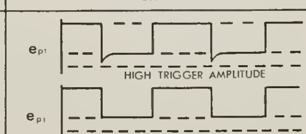
The duration of the output pulse can be varied from the time duration of the trigger pulse to the time between pulses. If attempts are made to make the pulse longer than the time between pulses, the second pulse may not start a new cycle, and the circuit would operate on every other trigger pulse.

Effect of varying circuit elements. The usual method of varying the output pulse width is to vary the size of the RC in the grid of V_2. The pulse width will increase with the resistance of R_2 and the capacity of C. Another result is that the amplitude of the output pulse will increase with an increase of the plate load resistor. Notice these effects in the chart on page 6-74 which summarizes the effects of varying the circuit elements in the start-stop multivibrator.

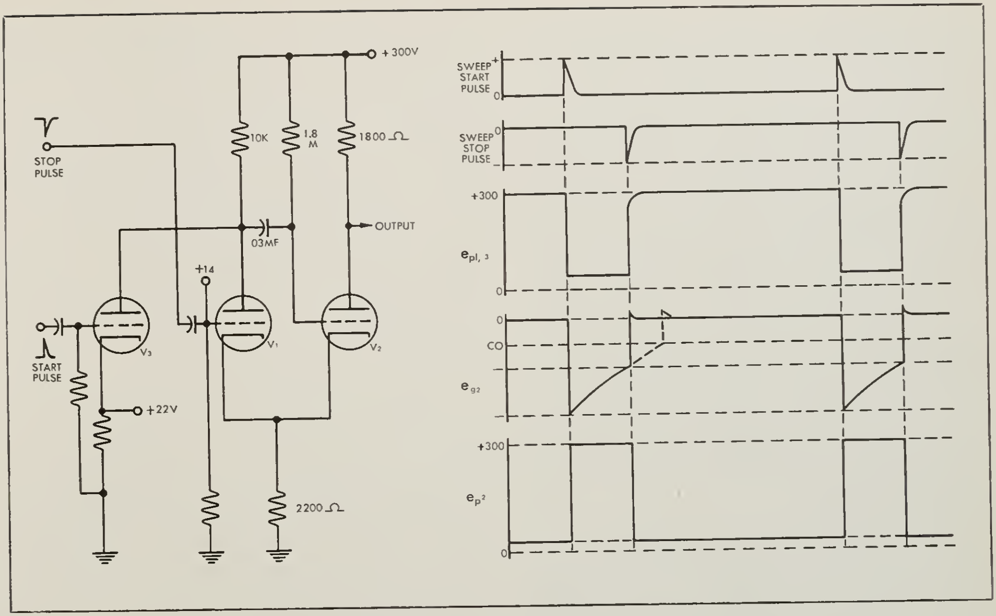
The stability of the pulse width from this circuit is improved by returning the grid of V_2 to B plus. The condenser discharges at a higher and more linear rate below the cut-off value, causing the slope to be steeper at cut-off.

Although the trigger pulse is applied to the first grid in this circuit, the circuit can also be triggered at the plate of the first tube or the grid of the second tube if a negative pulse is used. The action will be the same.

Application. The start-stop multivibrator on page 6-75 is used in a radar set to produce a pulse of measured duration which drives a sawtooth sweep generator. The basic circuit is modified several ways to produce this more efficient circuit. A third tube is incorporated in the circuit to isolate the multivibrator from the source of trigger voltage and to amplify and sharpen the pulse. This third tube (V_3 is biased beyond cut-off so that it acts as an open circuit at all times except when a trigger pulse is on the grid. The grid of first tube (V_1) has about +14 volts on it from the +300-volt voltage divider, while the grid of the second tube (V_2) is returned to plus 300 volts through a large resistor. The current through V_2 and the cathode resistor is high enough to cut off the first tube in spite of the positive voltage on V_1 grid. This positive voltage partly determines the duration of the output pulse. The circuit operates in this way —a positive trigger pulse, simultaneous with the transmitter pulse, is applied to the grid of V_3. This tube is called a clipper since 22 volts bias keeps it cut off and only the peak of the trigger pulse causes conduction. Due to amplification in V_3, a large negative going change appears in the plate circuit. This is coupled through the .03mf condenser to the grid of V_2. The current

CIRCUIT CHANGE.	EFFECT ON DURATION OF OUTPUT PULSE.	EFFECT ON AMPLITUDE OF OUTPUT PULSE.	EFFECT ON FREQUENCY OF PULSE REPETITION	EFFECT ON STABILITY.	
INCREASE C	INCREASED BECAUSE DISCHARGE LASTS LONGER	NO CHANGE.	NO CHANGE FREQUENCY GOVERNED BY TRIGGER FREQUENCY.	NO CHANGE SAME PERCENTAGE OF RC USED.	
INCREASE R_2	INCREASED DISCHARGE IS SLOWER	NO CHANGE.	NO CHANGE.	NO CHANGE. SAME PERCENTAGE OF RC USED.	
INCREASE R_3.	INCREASE BECAUSE AMPLITUDE OF e_{g2} IS INCREASED.	NO CHANGE.	NO CHANGE.	DECREASED BECAUSE DISCHARGE CURVE HAS LESS SLOPE AT CUTOFF.	
INCREASE R_4.	NO CHANGE DOES NOT APPRECIABLY EFFECT PRECEDING CIRCUIT	INCREASED SINCE C_{g2} IS LOWER BETWEEN TRIGGER PULSES	NO CHANGE.	NO CHANGE.	
INCREASE E_b	INCREASED BECAUSE V_1 PLATE SWING AND V_2 GRID VOLTAGE CHANGE IS GREATER	INCREASED DUE TO GREATER V_2 PLATE VOLTAGE CHANGE.	NO CHANGE	IMPROVED SLIGHTLY, SMALLER PERCENTAGE OF DISCHARGE CURVE USED.	
INCREASE R_K	SHORTER SINCE VOLTAGE CHANGE AT PLATE OF V_1 WILL BE DECREASED	NO CHANGE.	NO CHANGE	DECREASED BECAUSE VOLTAGE CHANGE AT PLATE OF V_2 WILL BE DECREASED	
INCREASE TRIGGER FREQUENCY	NO CHANGE. UP TO POINT WHERE DURATION IS TIME BETWEEN PULSES.	NO CHANGE.	INCREASED	NO CHANGE.	
INCREASE TRIGGER AMPLITUDE	NO CHANGE.	NO CHANGE.	NO CHANGE.	SMALL INCREASE WILL IMPROVE STABILITY IF ORIGINALLY TOO LOW	

Varying Elements in Start-Stop Multivibrator



Application of Start-Stop Multivibrator

in V_2 is reduced, V_1 conducts, the added drop cuts off V_2, and the discharging condenser keeps V_2 cut off. When the condenser discharges down to cut-off for V_2 this tube would conduct, ending the output pulse. However, this is the sweep circuit and the pulse should end when the spot reaches the range mark for maximum range. So a voltage is fed back to this multivibrator at the time the spot reaches the last range mark. This voltage is a negative pulse, applied to the grid of V_1. This reduces the current in V_1, and the small change is amplified around the circuit to cut off V_1 and cause V_2 to conduct again. Dotted lines in the V_2 grid waveshape show the normal changeover time, which is always later than the sweep-stop pulse. This system of stopping the sweep multivibrator with a voltage from the sweep deflection circuits simplifies the adjustments in the sweep circuit. When the range switch is changed to change the duration of the sweep, no part of this multivibrator is changed. But the speed of the spot movement is changed so it reaches the last range mark at a different time and the stop-pulse cuts off the multivibrator at a different time. Thus, the duration of the pulse is determined by the sweep circuit rather than the constants of the multivibrator circuit.

You recall that a hard tube sawtooth generator must be driven with a negative square wave at its grid. The positive square wave output from this multivibrator is amplified and inverted by a triode amplifier, then is applied to the sweep generator tube in the normal way.

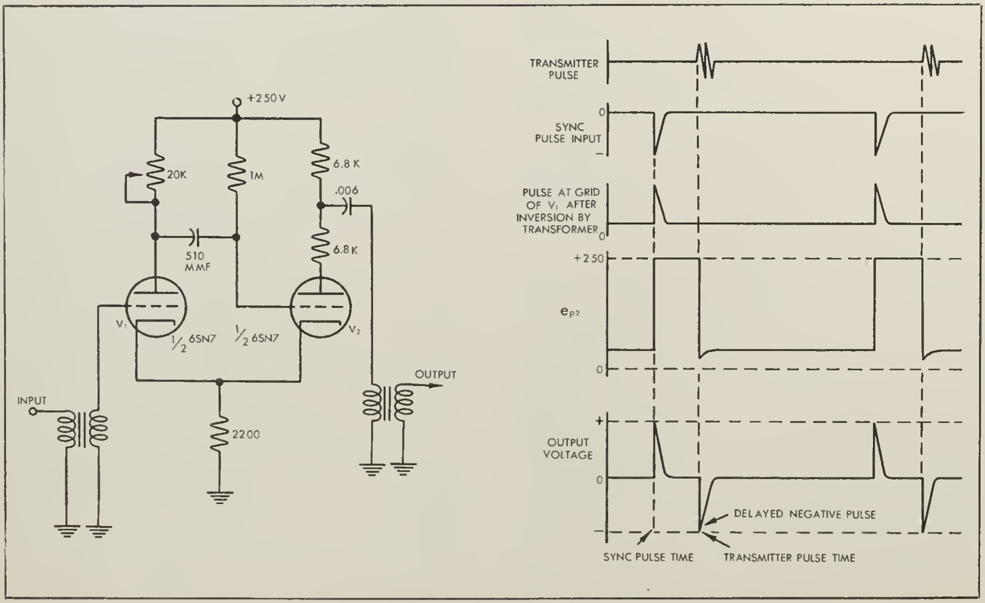
Use as delay circuit. In addition to producing pulses of a certain duration, the start-stop is sometimes used to "delay" a pulse. For example, in one radar set, the process of causing the transmitter to operate consumes about 5 microseconds, and the transmitter pulse occurs 5 microseconds after the original trigger pulse. While the sweep voltage in the indicator circuit must start at the same time as the transmitter pulse, the original pulse is used for the sweep voltage so that both the sweep and transmitter pulses occur at the same frequency. However, since the original synchronization pulse would start the time base 5 microseconds before the transmitter operates, it cannot be used directly. Instead, it must be "delayed" 5 microseconds. The start-stop multivibrator can introduce the delay.

The usual multivibrator circuit is used, with a transformer input. The input pulse is negative and occurs five microseconds before the transmitter signal. Due to the transformer connections, the polarity is reversed and the circuit is triggered by the resulting positive pulse. The usual square wave appears at the second plate, with the duration accurately measured by the RC circuit at the grid. The transformer features a short L R time constant, so the square wave is differentiated in the secondary. Again by selecting proper secondary connections, the second peak can be made negative. So the original negative peak is replaced by a similar negative peaked pulse which is delayed 5 microseconds. The delay need only be adjustable over a narrow range, so the plate load resistor of the first tube is variable. An ingenious built-in calibration indicator permits the operator to adjust this delay time while airborne. The adjustment is checked and readjusted, if necessary, just prior to the bomb run, when the radar set is used as the bomb sight.

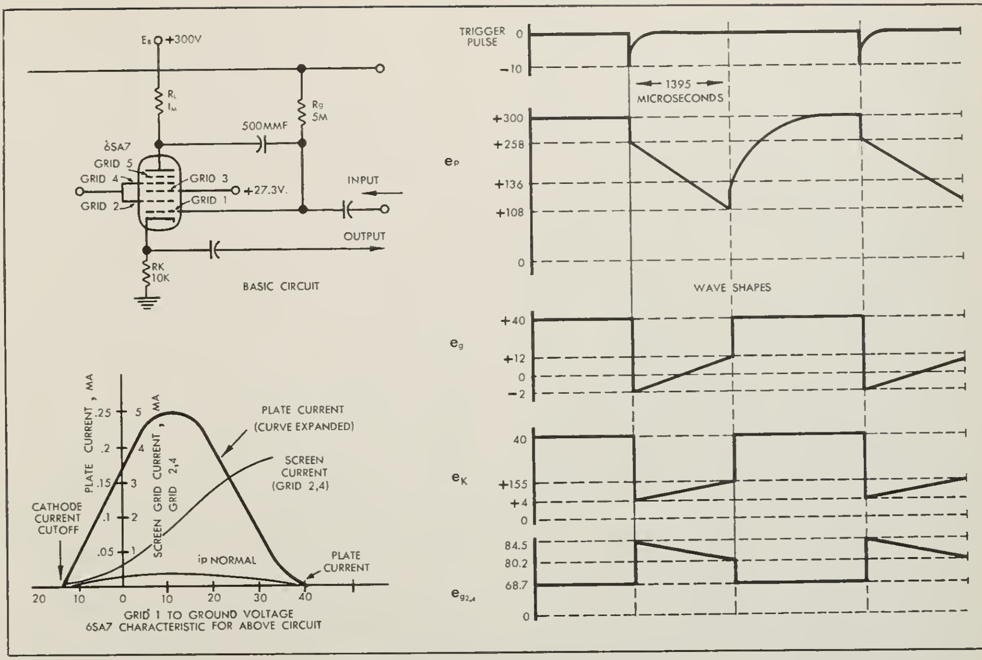
The Phantastron Circuit

The phantastron is another circuit which is used to delay a timing pulse. It is classed as a medium precision delay circuit. Its operation and output are very similar to those of the flip-flop circuit. Under power supply voltage changes it is quite stable because its operation depends on the fact that there is specific DC voltage relationship between the tube elements. Any variation of source voltage varies all these voltages in the same proportion, causing a minimum change in the voltage relationship since all voltages are supplied by voltage dividers across the same voltage source. The phantastron has the unique advantage over multivibrator circuits in that the pulse width, or delay, varies directly with one of the DC applied voltages. This control by a DC voltage, makes remote control feasible at any distance through unshielded power cables since there is no signal present in the control circuit.

Study the basic phantastron circuit on page 6-77. This circuit is never free running and is usually triggered by a negative pulse at the control grid (grid no. 1). Consider first the conditions before the trigger pulse arrives. Grids 2 and 4, which are tied together inside the tube, are connected to a voltage divider from +300 volts. The normal output of the divider is 86 volts, but with zero control grid voltage and 86 volts on grid 2, 4 ma of current will flow to grid 2. This additional current in the voltage divider



Start-Stop Multivibrator as Delay Circuit



Phantastron Circuit and Waveshapes

drops grid 2 to 68.7 volts. Grid 3 is connected to a voltage divider which sets its voltage at 27.3 volts. This grid does not draw current because it is negative with respect to the cathode. The current through the 10K cathode resistor is 4 ma. Since the cathode is 40 volts above ground, the grid 3 to cathode voltage is 27.3 − 40 = 12.7 volts. When grid 3 is more than 12 volts negative with respect to the cathode, it prevents electrons from passing it in their journey to the plate. Therefore, since the −12.7 volts is beyond plate current cut-off for grid 3, no electrons will reach the plate. Thus, the plate voltage is +300 volts. The grid is connected to E_B through a 5 megohm resistor. A very small grid current flows through the cathode resistor, but the grid voltage may be considered the same as the cathode voltage. In other words, the difference in potential between grid 1 and the cathode voltage is zero. Grid 5, the suppressor grid, is connected to the cathode and has no part in the operation of the circuit other than performing the normal function of a suppressor grid.

The next factor to consider is the similarity between the phantastron and the start-stop circuit. You can consider the 6SA7 phantastron tube as a tube within a tube. The cathode, grid 3, and plate are equivalent to the first triode in the start-stop. No plate current flows to the plate and the "first tube" is cut off. The cathode, grid 1 and grid 2 form another triode which is comparable to the second tube of the flip-flop. As mentioned previously, this triode is conducting. Consequently, any voltage change, which will cause plate current to flow in the triode that is cut off, will start a regenerative switching action.

You can best understand the action of the phantastron when it is triggered by dividing it into three successive phases. The first phase is an extremely rapid change which ends at a first balance point. The second phase is a slow linear change which ends at a second balance point. The third phase is the recovery of the circuit to pretrigger condition.

The trigger pulse must have the proper polarity and amplitude to bring grid 3 above cut-off. Since grid 3 is .7 volts beyond cut-off, one volt would do it. A negative pulse of 10 volts amplitude is shown. When this is applied to grid 1, the grid 1 voltage will drop to 30 volts. Cathode follower action will cause the cathode voltage to drop to 30 volts right with the grid. At this point, examine the relation between the cathode and grid 3. Their voltage difference is 27.3 − 30 or −2.7 volts. Grid 3 is not negative enough to prevent electron flow to the plate and plate current will flow as soon as the grid 3-to-cathode voltage drops below 12 volts. Regenerative action starts with this plate current flow. The plate current through R_L causes a drop of plate voltage. Grid 1 is capacitively coupled to the plate, so the grid becomes more negative. This drops the cathode voltage still more, which in turn further reduces the voltage difference between grid 3 and cathode. The current to the plate increases, the plate voltage drops, and the grid is driven farther in the negative direction.

Note that the plate current increase is *not* an overall increase in current with a negative going grid. The negative going grid actually decreases the total current slightly just as it should. But the resultant decrease in cathode voltage lowers the voltage between grid 3 and the cathode, which allows current to flow to the plate at the expense of the screen current. Grid 3 controls the *division* of current between the plate and screen grid. Grid 1 controls the total current. So grid 1 decreases the total current slightly, while grid 3 allows the plate current to increase and causes the screen current to decrease quite rapidly. The 6SA7 characteristic curve shows the relation. In plotting these curves a DC voltage was applied to grid 1 of the circuit in the illustration and varied from minus 20 volts to plus 40 volts. The dotted curve shows the resultant screen current while the other curve shows the plate current. Note that plate current increases by a small amount while screen current decreases a large amount as G_1 goes negative. The sum of the two currents is the cathode current. It is decreasing, because the small current increase is more than overcome by the large decrease.

The regenerative effect just explained is similar to the cutting off of the first multivibrator tube by the second tube. The cutting off process, however, does not go that far in the phantastron, for if the current to the plate were stopped, the plate voltage would have to go up, and this action would be the opposite to the desired feed back. The increase of current to the plate ends at the first balance point. This balance occurs when the negative going grid 1 voltage decreases the overall current so much that the plate current cannot continue to increase. Grid 3 is unable to divert enough electrons to the plate to maintain the increase of plate current, so the plate current stops increasing and becomes steady. This occurs at about −2 volts grid 1 voltage, so the plate has dropped 42 volts before the balance point is reached. Each of these voltages is shown in the curves, under the trigger pulse. The grid voltage has dropped to minus 2 volts. Since this is a 42-volt change, which is caused by the plate, the plate curve shows a 42 volt drop to 258 volts. The cathode voltage drops due to the decrease in plate current until with −2 volts on the grid, the current is .4 ma, caused by the plate, the plate curve shows a at grids 2 and 4 has gone up because the screen current is decreased, reducing the drop across the voltage divider and raising its voltage 84.5 volts.

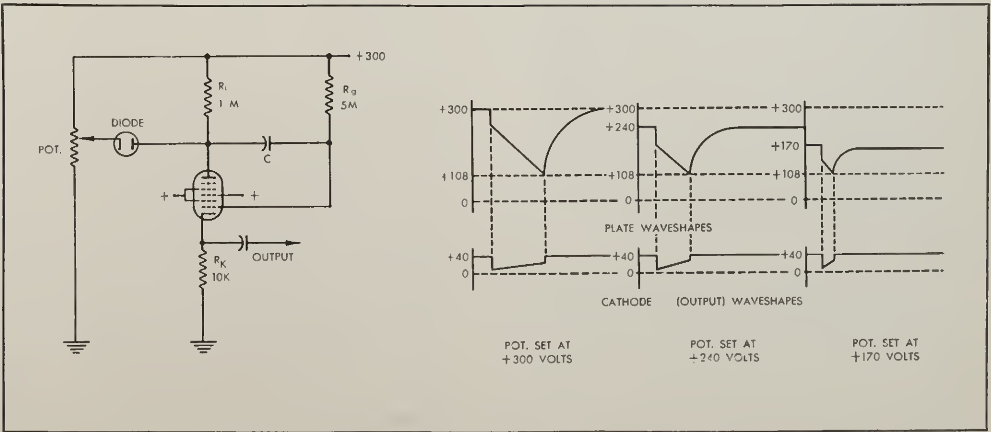
Now consider the action resulting from the starting of the second phase. The condenser from plate to grid at this time starts to discharge to the new lower plate voltage. At this time the grid voltage is −2 and the condenser discharges from 300 to 260. In discharging, the grid end of the grid resistor becomes more positive. This increases the plate current, starting another drop in plate voltage. Grid 3 is not involved directly in this stage as it is straight triode action between the plate, grid 1, and the cathode that affect the circuit operation. The plate voltage drop has an effect opposite to that of the discharging condenser and thereby is degenerative in action. The condenser discharges, the grid goes *positive*, plate current increases, plate voltage drops, tending to make the grid go *negative*. But the plate cannot drop enough to exceed the positive change at the grid because it is the positive grid voltage change that causes the plate voltage drop. So the plate voltage change only counteracts *part* of the effect of the condenser discharge, slowing the discharge considerably. This negative feedback keeps the condenser discharge in the most linear part of the exponential curve. In fact, mathematically it can be shown that the time constant is lengthened by an amount equal to the amplification factor of this triode. Looking at it still another way, the exponential curvature at the grid is counteracted by opposite curvature at the plate due to the inverting properties of the triode tube.

The discharge continues at a linear rate, as shown by the plate voltage curve illustrated on page 6-77. The increase in current raises the cathode voltage to 15.5 volts, and the grid 1 voltage is raised to +12 volts before the second balance point is reached.

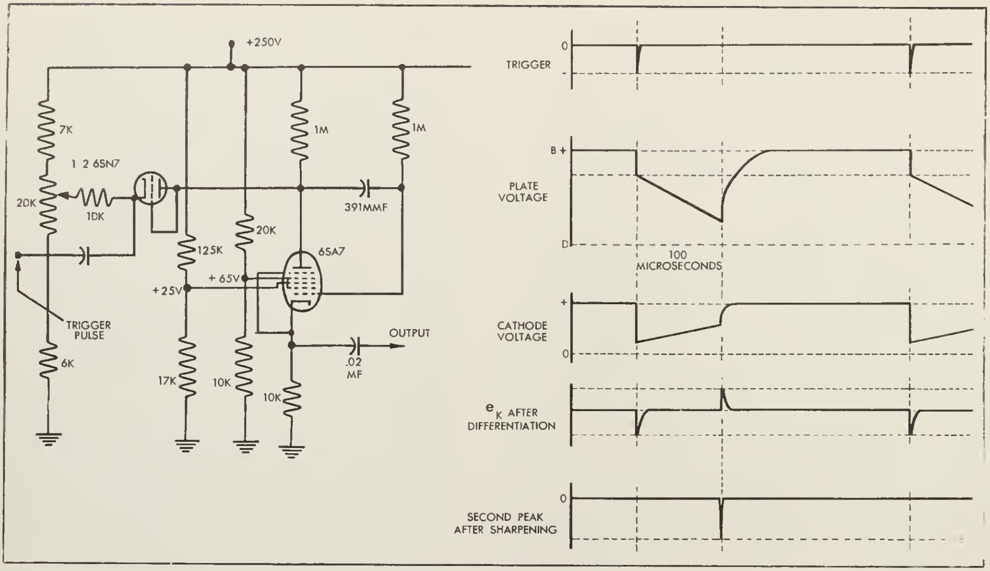
Remember that during all this time, the screen grids (2 & 4) have been continually drawing current. As the plate voltage continues to fall, there is a minimum point where 68 volts on the screens will collect more electrons from the space current than a higher voltage on the plate. This occurs at about 108 volts on the plate. This point is also the peak of the current characteristic in the illustration. To the right of this peak, the plate current decreases and the screen current increases. The screen grids take more current than the plate can attract, so the plate current stops increasing and levels off. This leveling off initiates the rapid switchover to recover the circuit to pretrigger condition.

When the plate current levels off, the voltage stops decreasing at the plate. With the counteracting effect of the plate drop removed, the condenser discharge raises the grid 1 voltage at a very rapid rate. The positive going grid 1 increases the current from the cathode, which raises the cathode voltage. This has the same effect as making grid 3 more negative; so grid 3 reduces the current to the plate. Actually, it is dividing the increasing current in favor of the screen grids. The screen current increases tremendously while the plate current is actually decreased. The plate voltage starts to go in the positive direction. This is coupled to the grid by the 500 mmf condenser to make it go in a positive direction. The cathode voltage is raised some more, which brings grid 3 closer to cut-off, reducing the plate current more and raising the plate voltage. This regenerative action continues until grid 3 is beyond cut-off—which stops plate current completely. During this regenerative action, the grid voltage has increased as fast as the plate voltage, so no change in condenser voltage occurs. But when the grid reaches +40 volts, the plate current reaches zero and the grid no longer affects the plate voltage. Thus, as the grid voltage jumps from 12 to 40, the plate voltage curve shows a 28-volt jump from 108 to 136 volts. When the grid voltage reaches a constant value, the plate rises only as fast as the condenser can recharge along a slow exponential curve up to +300 volts. The cathode voltage rises immediately with the grid voltage while the screen grid voltage drops at recovery time because of the increased screen current.

Now consider the circuit from the viewpoint that it is ready for the next trigger pulse. The duration of operation, and consequently, the duration of the negative pulse on the cathode, is fixed in the circuit shown. This pulse width is determined essentially by the RC in the grid 1 circuit, the plate voltage, and the gain of the triodes involved. The pulse width can be calculated, but the calculations are a bit lengthy to be included here.



Changing Pulse Width in Phantastron Circuit



Typical Radar Phantastron Circuit

CHANGING PULSE WIDTH. The duration of the pulse is generally made variable by applying a DC voltage to the plate from a potentiometer. The circuit is as shown on page 6-79. A potentiometer can select any voltage from zero to +300 volts. A diode is inserted in series to disconnect the phantastron plate from the low-resistance potentiometer during the operation of the circuit. Before the trigger pulse, the plate voltage and voltage at the arm of the potentiometer arm are the same because the load resistor is so large and the current from the potentiometer so small that there is little voltage drop between the arm and 6SA7 plate. Most of the current therefore passes through the diode. When the circuit operates, the plate voltage drops immediately, the diode plate becomes less positive than its cathode, and diode current stops. So the potentiometer-diode circuit has no further effect on the operation of the phantastron.

To see how the potentiometer affects the pulse width, study the effect of varying the position of the potentiometer contact. With the plate voltage at +300 volts, the situation is as previously described. But when this plate voltage is reduced to +240 volts by means of the potentiometer, the pulse width is decreased. The plate voltage drops the same 42 volts at the time of the trigger pulse since the grid drops 42 volts to the first balance point. Then the slow linear voltage decrease occurs. The slope of this decrease is the *same* regardless of plate voltage. This terminates at the same 108 volts as the 300 volts waveshape. So the amount of linear decrease is less, which decreases the time required to reach 108 volts. If the potentiometer is set so the plate voltage is 170 volts before the pulse, the plate will again drop 42 volts, and follow the *same* slope to 108 volts. Again the amplitude of the slope is decreased so the slope time is decreased.

Since the initial drop, the slope, and the final value do not change, the decrease in pulse length is linear with the decrease in plate voltage. This relation holds over the range from the full plate voltage (300 volts in this case) to a minimum of 108+42 or 150 volts. This is the minimum voltage point because the initial 42-volt drop would bring it down to 108 volts, or zero pulse width. Since the pulse width for the example used is 1395 microseconds, the pulse width is variable from zero to 1395 microseconds.

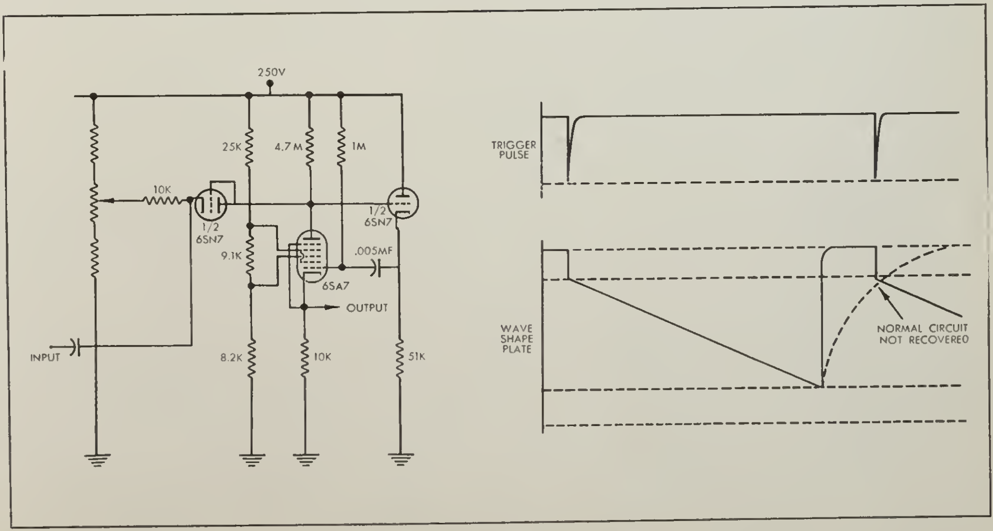
TYPICAL USE. The phantastron circuit shown above is taken from a typical radar set. It is similar to the previous circuit except that here the values and voltages are slightly different, the voltage dividers for grids, 2, 3, and 4 are shown, and the disconnector diode is actually a triode with its grid tied to the plate. This last is done in the set to make use of the other half of a 6SN7 that is employed in another circuit. The potentiometer which sets the plate voltage has fixed resistors in each end of it to limit the pulse width to a certain maximum and minimum width which is adequate for the purpose for which it is designed.

The trigger pulse is applied to the cathode of the disconnector diode. The pulse momentarily changes the plate current in the diode, which causes the negative pulse to appear at the plate of the phantastron tube. The plate-to-grid coupling condenser couples the pulse voltage to the grid where it starts the circuit through a cycle of operation. The output is taken from the cathode of the 6SA7, where a negative, approximately-square waveshape is produced. Actually, the waveshape is not as ideally square as shown on page 6-77, because the trailing edge becomes somewhat exponential in following the condenser charge.

In use, the square wave is usually differentiated and the trailing edge pulse is put through a pulse sharpening circuit to remove the effect of the exponential trailing edge. The resulting sharp pulse is delayed from the original trigger pulse by a few hundred microseconds.

This circuit is satisfactory for short delays, but when the delay becomes nearly as long as the time between pulses, the long exponential recovery of the plate circuit overlaps the next pulse. With the circuit not yet fully recovered, the second pulse may not trigger the circuit. Therefore the circuit is modified for long delay periods as shown below. The change is the addition of a triode between the plate and grid of the phantastron tube. This triode is called a *cathode follower*. Before the trigger pulse arrives, the grid of the cathode follower is at the plate potential, it is drawing current, and the cathode is within a fraction of a volt of the grid voltage. Therefore the .005 mf coupling condenser has the plate voltage on one side and the grid voltage on the other—which is not different from the circuit on page 6-77. When the trigger pulse drops the plate voltage, the grid voltage of the cathode follower, and consequently its cathode voltage goes down. The .005 mf condenser must discharge, and in doing so, makes grid 1 go in the negative direction. Thus by cathode follower action, the grid drops right along with the plate as though capacitive coupling existed between them.

When the end of the pulse arrives, and the plate voltage rises back to 250 volts, it does not have to charge the condenser mentioned above. The plate can raise the cathode follower grid immediately to +250 volts. Then the condenser charges through a path from the 6SA7 cathode



Phantastron with Cathode Follower

to grid one, through the condenser, from cathode to plate of the cathode follower, to +250 volts. The grid of the cathode follower is not involved, other than through interelectrode capacity, which is very small. Only a slight exponential curvature of the plate waveshape is shown in the illustration. This curve is due to stray wiring capacity in the circuit. The dotted curve shows the exponential rise which would interfere with the next pulse if no cathode follower were used. In all other respects, this long delay circuit is the same as the short delay circuits.

The Blocking Oscillator

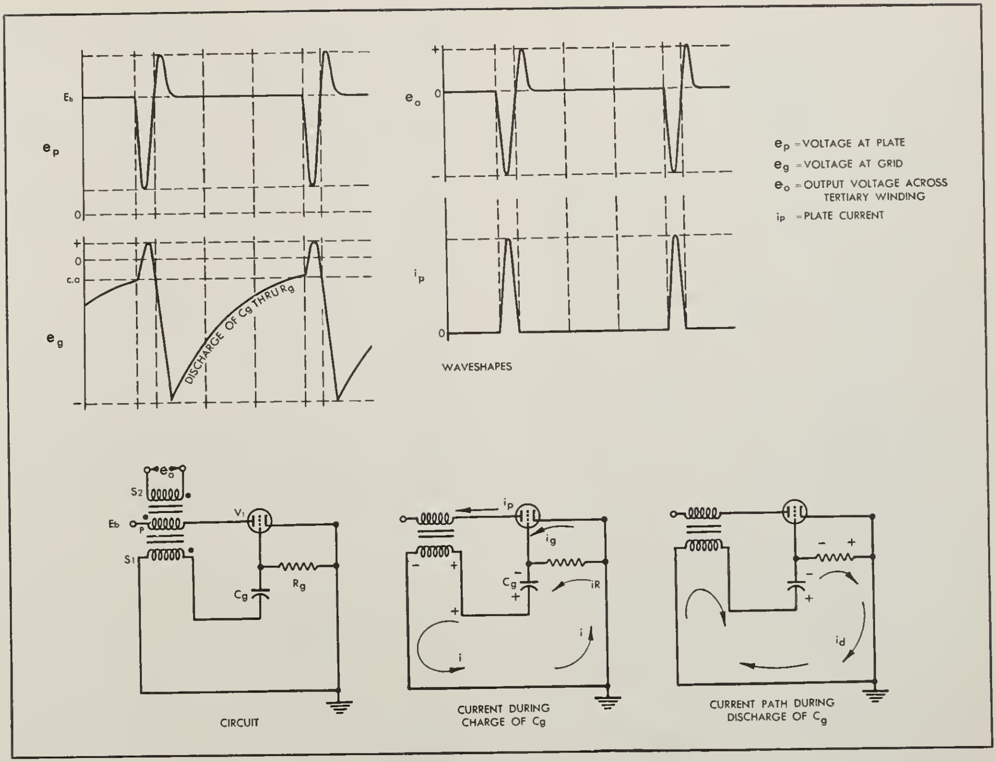
The blocking oscillator is another type of relaxation oscillator. It can do many of the things that a multivibrator does. Thus it can be free running, synchronized at each cycle or at a submultiple frequency, or it can be driven in the manner of a one-shot multivibrator.

A blocking oscillator is any oscillator which cuts itself off after one or more cycles on account of the accumulation of a negative charge on the grid capacitor. Thus, in an oscillator in which the grid swings positive with respect to the cathode, electrons are attracted to the grid and accumulate on the plate of the grid capacitor nearest the grid. Since these electrons cannot return to the cathode through the tube, they must return through the grid-to-cathode resistor. If this resistor is sufficiently large, electrons may accumulate on the capacitor faster than the resistor permits them to return to the cathode. In this case a negative charge is built up at the grid which may bias the tube beyond cut-off. After the tube is cut off, it provides no additional electrons to the grid capacitor. However, the capacitor continues to discharge through the resistor, and a point is reached eventually where the tube again conducts. Thus the process repeats and the tube becomes an intermittent oscillator. The rate of the recurrence of operating conditions is determined by the RC time constant of the grid circuit.

There are two general types of blocking oscillators—the *single swing* type in which the tube is cut off at the completion of one cycle, and the *self-pulsing* type, in which each cycle of oscillation causes the grid to become progressively more negative until the tube is biased out of operation. In radar the single-swing type oscillator usually operates within the audio-frequency range, while the self-pulsing type produces pulses of RF energy.

The single-swing oscillator circuit shown consists of a transformer-coupled oscillator, with a capacitor in series with the grid of the triode V_1. For the purpose of understanding its operation, assume that the grid capacitor C_g has been negatively charged by a preceding cycle. The tube, therefore, is biased well below cut-off. As the charge on the capacitor leaks off, the biasing voltage is reduced to the point where the tube begins to conduct. As plate current starts to flow, a magnetic field is set up around the plate winding P of the transformer. The dots at each winding indicate similar polarities. If, for example, a current flows through a winding so that the dot end is positive, the field set up in the core induces voltages in the other windings that makes the dot end positive in these windings at the same time. The field builds from zero to a maximum in direct proportion to the plate current, and therefore induces increase is slowed some more. This action is impressed upon the grid of the tube through the grid capacitor C_g with a polarity that drives the grid more and more positive as the field in the plate winding is building up. When the grid is driven positive with respect to its cathode, it draws current, and electrons accumulate on the plate of the grid capacitor nearest the grid. As the grid becomes more positive, it eventually takes enough electrons from the plate current to lower the rate of plate current increase. This reduces the induced voltage, the grid stops becoming more positive, and the plate current increase is slowed some more. This action is cumulative and ends with the plate current being settled at a constant value, and the magnetic field being stationary with respect to the secondary.

At this time there is no induced voltage in the grid winding and, because there is no charging potential applied, the capacitor begins to discharge. This discharge causes the potential on the grid to become less positive, thereby causing less plate current to flow in the plate winding. The field around the plate coil starts to collapse. This collapsing field, in turn, induces a voltage in the grid winding in the reverse direction, causing the grid to become more and more negative. This process continues until the grid is driven beyond cut-off, thus completing a cycle of operation. Oscillation does not start again immediately, however, because the flow of grid-current when the grid was positive has built up enough charge on the grid capacitor to hold



Single-Swing Blocking Oscillator

the tube cut-off until some of the charge leaks off through the grid resistor.

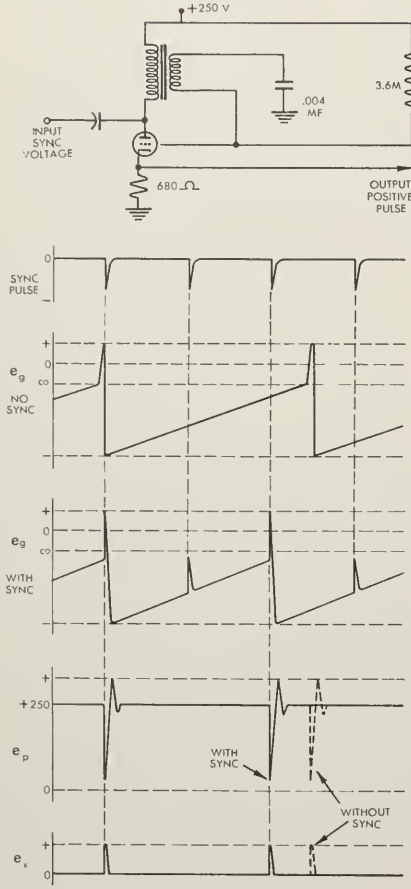
The time consumed by the rise and decay of plate current depends upon the inductance and resistance of the transformer. The time between pulses depends primarily upon the value of resistance since the grid capacitance is fixed because of pulse width requirements.

Note the current paths during charge and discharge. The condenser charges very quickly through the low resistance of the conducting grid-cathode circuit but discharges very slowly through the large grid leak resistance. The rounded waveshapes show that the voltage and current changes occur more slowly than in a multivibrator. This is due to the inductance in the transformer. The inductance also is the cause of the plate voltage rising above E_B when the tube is cut off. The current continues to flow in the same direction as it charges the capacity in the transformer windings, and causes a voltage drop which adds to E_B. The inductance and distributed capacity form a resonant circuit which tends to oscillate due to shock excitation. If the Q of this LC circuit is high, it may oscillate for several cycles. In the case illustrated, damping is at the critical value so that the oscillation occurs for about a half cycle.

The output (e_o) from this circuit is taken from a third winding on the transformer. However, the output can also be taken from the plate, where a large negative pulse is produced, or from a small resistor in the cathode circuit, where a small but sharp positive pulse is available. The waveshape of the cathode pulse is the same as the plate current waveshape shown in the illustration.

The Synchronized Blocking Oscillator

The synchronized blocking oscillator on the next page, is taken from a radar set. This oscillator is employed as the timing oscillator in a small piece of auxiliary equipment. Normally, the set



Synchronized Blocking Oscillator

produces its own PRF independently when the oscillator is free running, as you can see in the waveshapes labeled "e_g, no sync". The natural frequency for its operation is about 300 CPS. However, a synchronizing pulse can be introduced at the plate without making changes in its circuit that will produce a PRF as high as 1000 CPS. The synchronization pulses shown occur at 700 CPS. Their amplitude is such that one pulse makes the circuit operate, the next pulse does not bring the grid above cut-off, but the third pulse does make the tube conduct.

Since the oscillator is synchronized on alternate pulses, its PRF becomes 700/2, or 350 CPS. If the main radar set has a PRF of 1000 CPS and the amplitude slightly less than that shown, the blocking oscillator can be synchronized on every third cycle, and it will operate at $333\frac{1}{3}$ CPS. By synchronizing this auxiliary set with the big set, indications received by the small set can be displayed on the main indicator, relieving the operator of observing two radar indicators at the same time.

The amplitude of the synchronizing pulse has a great effect on the frequency of this circuit. If the amplitude of the synchronization pulse shown is increased 50%, the first pulse that occurs after the circuit begins operation will bring the grid above cut-off and the circuit will be synchronized on every pulse. As stated before, a low amplitude will cause triggering on every third pulse since the first two will not have sufficient amplitude to make the tube conduct.

In each case, it is assumed that the natural period of the circuit is longer than the time between the pulses which cause the circuit to operate. If the period of the blocking oscillator is shorter, it will produce long and short cycles because the circuit will be triggered on some pulses but will conduct on its own accord at times.

Maximum frequency stability is achieved in this circuit by connecting the grid load resistor to the plate voltage supply. This has the same effect as returning the grids to B+ in multivibrators. The useful output from this circuit is taken from a small (680 ohm) resistor in the cathode circuit. This resistor has a negligible effect on the operation of the circuit but develops a voltage pulse due to the plate current flowing through it.

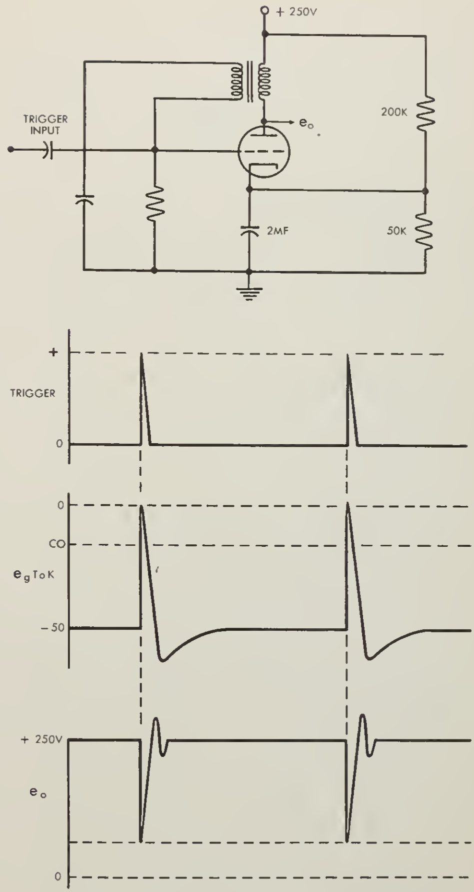
The Driven Blocking Oscillator

Some radar circuits require a blocking oscillator which operates only when triggered. Such a circuit might be called a *blocked oscillator* since a permanent DC voltage prevents the circuit from oscillating of its own accord. Only a positive voltage on the grid or a negative voltage on the cathode will cause the tube to conduct. If the voltage is a pulse, the circuit will go through a cycle of operation, cut itself off, and the DC voltage will keep it cut off. Cut-off is maintained by a voltage divider from the plate supply voltage. With the values shown, the cathode voltage will be plus 50 volts with respect to ground. The grid is grounded, and is therefore 50 volts negative with respect to the cathode. The grid waveshape shows voltages with respect to the cathode since these are the important voltages in understanding the circuit operation. Cut-off for most tubes used in this type circuit occurs at about −18 volts, so the trigger pulse must be more than 50−18, or 32 volts in amplitude to start a cycle of operation. In the grid waveshape the normal discharge of the grid condenser reduces the grid voltage after a cycle of operation, but the discharge is to minus 50 volts, so the circuit does not again conduct without another trigger pulse.

The positive trigger pulse must be applied to the grid only. It is not practical to apply a trigger pulse to the cathode because the filter condenser is usually found there. But a negative pulse may be applied to the plate, where the transformer will invert the pulse to a positive one on the grid. The output may be taken from the plate, where a large negative pulse is available, or from a third transformer winding where a pulse similar to the plate pulse can be obtained with either polarity.

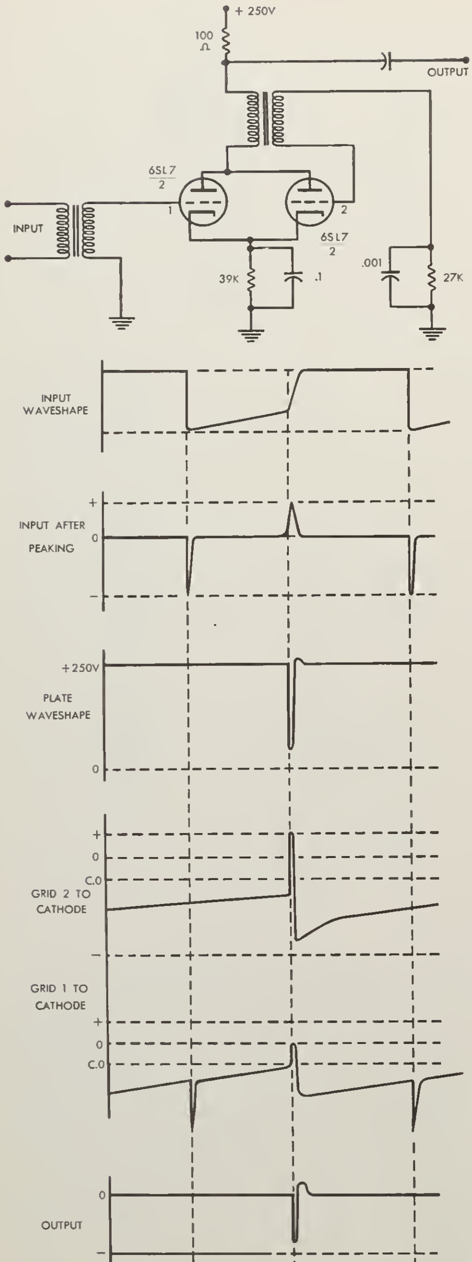
USE AS PULSE SHARPENING CIRCUIT. The driven or blocked oscillator is useful as a pulse sharpening circuit. One set uses the circuit on the next page to improve the pulse shape from the phantastron. The phantastron produces the square wave labeled input waveshape. This square wave is amplified and becomes the input to the differentiating transformer shown in the circuit. The waveshape at the secondary of the transformer is a peaked wave, as shown in the second waveshape, but because of the inherent imperfections in the trailing edge of the phantastron pulse, the peaked wave due to the trailing edge is not a sharp one. For precise triggering of the circuits following the phantastron, a sharper pulse is required. Therefore the peaked wave is applied to a blocked oscillator.

In place of the voltage divider from +250 volts used in the earlier circuits, a long time constant in the cathode circuit provides the cut-off voltage for the preceding circuit. After several cycles of operation, the grid current due to positive grid voltages charges the cathode condenser. It will discharge slowly between cycles and be recharged with each cycle. The RC in the grid circuit is rather short and cannot hold the grid below cut-off for the several hundred microseconds between pulses.



Blocked Oscillator

The negative pulse at the leading edge of the square pulse will not affect the already cut-off tube. The positive pulse at the trailing edge will cause the left tube to conduct. The current through the transformer and associated magnetic field will induce a positive voltage at the grid of the right-hand tube. This starts a regenerative action in the blocking oscillator associated with the right-hand tube and it goes through a normal cycle of operation. Meanwhile the pulse ends at the grid of the left-hand tube, and with no other voltage on the grid, the tube is cut off again.



Blocked Oscillator as a Pulse Sharpening Circuit

The left-hand tube is an amplifier and isolation circuit which causes the blocking oscillator to start very soon after the trigger pulse arrives, then disconnect the triggering circuit from the blocking oscillator so it can go through its normal cycle unaffected by the triggering circuit.

The output is taken from a small resistor above the transformer in the plate lead. A sharp negative pulse of low amplitude is produced across the resistor.

COUNTING CIRCUITS

A counting circuit, also known as a frequency divider, is one which receives uniform pulses, representing units to be counted, and produces a voltage proportional to their frequency. By slight modifications the counting circuit is used in conjunction with a blocking oscillator to produce a trigger pulse which is a submultiple of the frequency of the pulses applied.

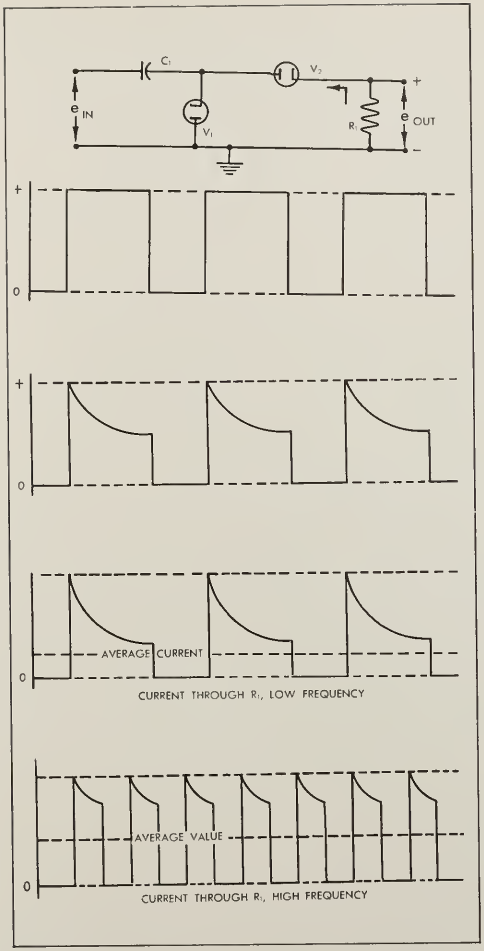
The pulses applied to the counting circuit must have the same amplitude and time duration if accurate frequency division is to be made. Counting circuits, therefore, are ordinarily preceded by shaping and limiting circuits to insure this uniformity of amplitude and width. Under these circumstances the pulse-repetition frequency constitutes the only variable, and frequency variations may be measured.

Positive Counting

Positive pulses, which may vary only in their recurrence frequency, are applied to the input of the positive counter. The charge on the coupling capacitor C_1 cannot change instantaneously as the positive leading edge is applied, so the plate of V_2 becomes positive and the diode conducts. A charging current flows through R_1 during the pulse time and a small charge is developed on C_1. At the end of the pulse the drop in voltage places the diode side of the capacitor at a negative potential equal to the charge accumulated on C_1.

Tube V_2 cannot conduct, as its plate is negative with respect to its cathode. However, V_1 conducts, discharging the small charge from the capacitor, which would otherwise build up during each succeeding positive pulse, eventually rendering the circuit insensitive to the applied pulses.

CIRCUIT. Since a certain amount of current flows through R_1 each time a pulse is applied, an average current flows which increases as the pulse recurrence frequency increases and decreases as

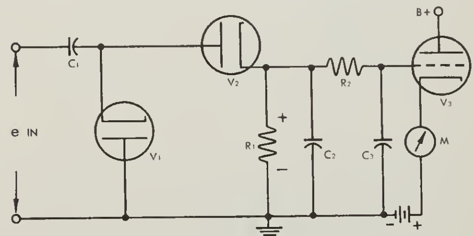


Positive Counter

this frequency decreases. The IR drop developed across R_1 can be used to control a succeeding stage as is illustrated in the circuit controlled by positive counter. The filter in grid circuit of V_3 aids in obtaining smooth operation by removing too rapid changes in voltage developed across R_1. The voltage at the grid of V_3 varies with changes in the pulse frequency and produces variations in the plate current of V_3. A milliammeter is placed in series with the plate circuit so that changes in the average plate current are indicated as a measure of variations in the recurrence frequency of the input pulses.

Negative Counting

By reversing the connections to the diodes V_1 and V_2 in the positive counting circuit the circuit will respond to negative pulses and become a negative counter circuit. The diode V_2 conducts during the time the negative pulse is applied and an electron current flows through R_1 as indicated by the arrow. At the end of the negative pulse, V_1 conducts sufficiently to remove the charge which developed on C_1 during the negative pulse while V_2 was conducting. The current through R_1 increases with an increase in the pulse frequency as before. How-

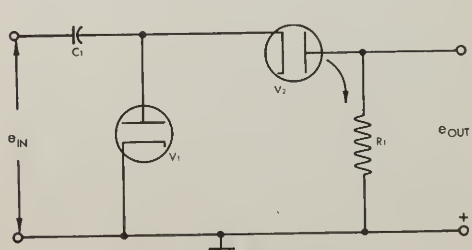


Circuit Controlled by Positive Counter

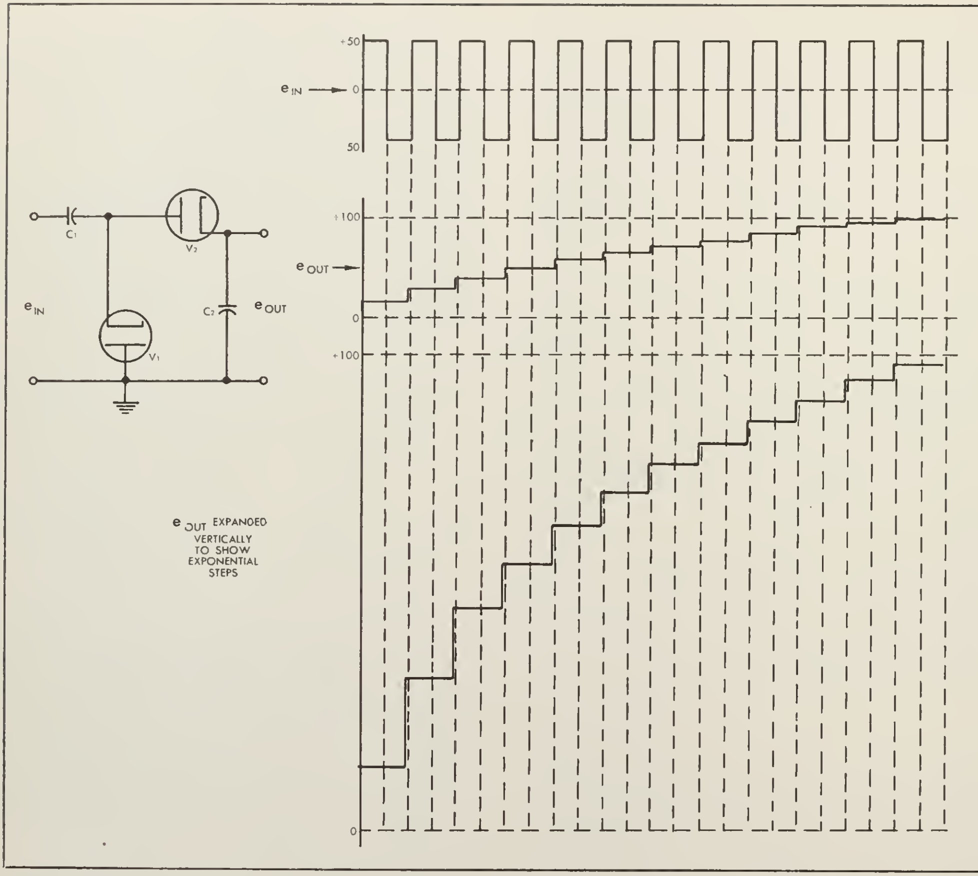
ever, if the voltage developed across R_1 is applied to the same control tube shown previously, the increase in current causes the grid of V_3 to become more negative. This decreases the plate current through V_3 and the meter. Thus an increase in the frequency of the negative pulse causes a drop in the average plate current measured by the meter. This is opposite to the effect in the positive counter.

Step-by-Step Counting

The step-by-step counting circuit is similar to those already discussed except that a capacitor which is large compared to C_1 replaces the resistor R_1 used in the positive counting circuit. The charge on this capacitor, C_2 (in the circuit)



Negative Counting Circuit



Step-by-Step Counting

increases slightly during the time of each positive pulse and produces a step voltage across the output. These steps decrease in size exponentially as the voltage across C_2 approaches the final value, the rate being dependent upon the output impedance of the driving circuit. As long as there is no path through which C_2 can discharge, the voltage across it continues to increase with each successive pulse until it is equal to the amplitude of the applied signals. At this point the cathode of V_2 is held at a positive voltage equal to that on the plate during the pulse time and the tube fails to conduct.

USE. Use is made of the step charging circuit to trigger a blocked oscillator after a certain number of steps. A simple circuit of this type is shown on the next page. The step-by-step counter determines the voltage on the blocked oscillator grid. The cathode has a high fixed bias on it, and a certain number of steps are required to raise the grid near enough the bias level to make the tube conduct. When the tube conducts, the blocking oscillator circuit goes through a normal cycle. The usual grid current during this cycle discharges the condenser C_2. At the beginning of the next input cycle, C_2 begins charging again in steps, starting from zero.

This circuit is not difficult to analyze quantitatively. The following is the calculation of the exact frequency division ratio in this circuit. In the step charging circuit, the time constant of the charging circuit is short since only the diode plate resistance is in series with the condensers. Therefore, the capacity is fully charged during each cycle of input waveshape. But the condensers are in series, so the full charging voltage will be divided between the two. The voltage distribution in condensers is in inverse ratio to the capacity of the respective series-connected condensers and is mathematically expressed,

$$\frac{C_1}{C_2} = \frac{E_2}{E_1}$$

In this equation,

C_1 = *Capacity of first series condenser.*

C_2 = *Capacity of second series condenser.*

E_2 = *Charge on* C_2 *if applied voltage is* $E_1 + E_2$

E_1 = *Charge on* C_1 *if applied voltage is* $E_1 + E_2$

The voltage in question is E_2. Let the applied voltage be E_a. The equation for the E_2 voltage becomes:

$$E_2 = E_a \frac{C_1}{C_1 + C_2}$$

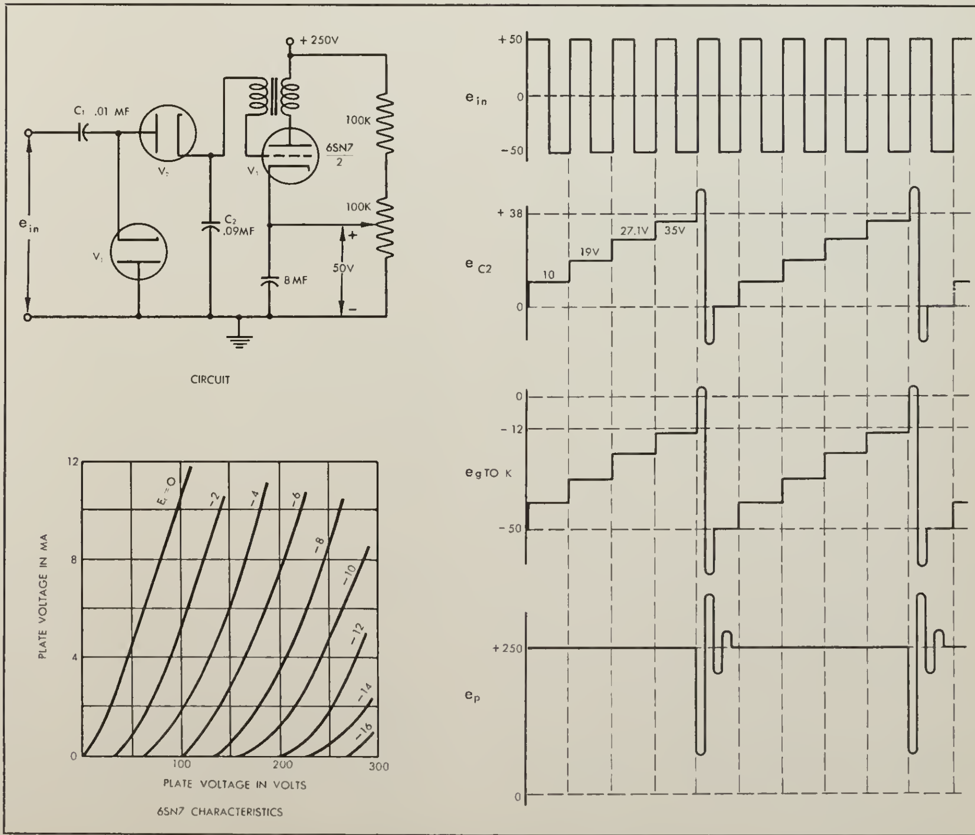
Using the values indicated in the illustration, E_2 is,

$$E_2 = 100 \times \frac{.01}{.01 + .09}$$

$$= 100 \times \frac{.01}{.1}$$

$$E_2 = 10 \text{ volts}$$

The charge on condenser C_2 with 100 volts applied will be 10 volts, while the condenser C_1 will charge to 90 volts. It follows that the first step is 10 volts, which is 10% of the applied voltage. After the circuit has been in operation for more than one complete cycle of the output,



Triggering Blocked Oscillator with Step Charging Circuit

the first step of each cycle is 10 volts when e_{in} is 100 volts peak-to-peak, regardless of the voltage reference level of e_{in}. The plate of V_2 is held at zero by V_1 during each negative half-cycle of e_{in}. Therefore the plate voltage of V_2 changes from zero to plus 100 volts at the beginning of each input cycle and C_2 charges to 10% of the total change during the first input cycle.

The first input pulse ends, C_1 is discharged, but C_2 holds its 10 volt charge. The second pulse occurs. The 10 volt charge on C_2 will oppose the 100 volts of the second pulse, and the total applied voltage, for the condensers to charge to, will be 90 volts. E_a will be 90 volts, and from the previous equation, C_2 will again be charged by 10% or 9 volts. This is in addition to the initial 10 volts, so at the end of the second pulse, the C_2 voltage is 10+9 or 19 volts. The third pulse will be 100 volts again, but 19 volts will oppose it. Therefore the E_a will be 100 − 19 or 81 volts and C_2 will charge an additional 8.1 volts. The new E_2 becomes 19+8.1 or 27.1 volts. Successive input pulses will raise C_2 by 10% of the remaining voltage toward plus 100 until the blocked oscillator works.

In the oscillator, the bias is set at plus 50 volts. The plate supply voltage is 250 volts, so the plate to cathode voltage is 200 volts. Using the plate-to-cathode voltage, you can determine the grid voltage for cut-off. From the curves in the illustration, this value is near − 12 volts, as indicated on the characteristic curves. With the cathode at plus 50 volts, the grid must be within less than 12 volts of this value, that is, above 50 − 12, or 38 volts to allow conduction. So the steps continue until 38 volts is exceeded. Since the fourth step is 35 volts and the fifth step 41 volts, the 38 volt level is crossed between the fourth and fifth step.

The blocked oscillator goes through one cycle of operation at the end of which the "bucket condenser" is discharged. Since every fifth input pulse results in an output pulse, this circuit is a five to one frequency divider.

To cause the circuit to divide by three, or four, or some other value, the bias is set at some new value. For example, a bias of 28 volts would cause the crossover with cut-off voltage to occur at the end of the third pulse, making the circuit a 3 to 1 count divider. The sizes of the steps can be varied by changing the ratio of the two capacitors. For example, if C_2 is changed to .04 mf, it will take 20% of the charge, and the 38-volt level will be crossed at the end of the third pulse, causing division at a 3 to 1 ratio. In practice, this frequency division can be carried to the extent that with six successive frequency dividers, a 100,000 cps sine wave can be reduced to 50 cps. The final circuit will be triggered on each two thousandth cycle of the original voltage.

CHAPTER 7

Amplifiers and Oscillators

Perhaps the two most important circuits in radio and radar equipment are amplifiers and oscillators. They are essentially alike, each depending upon the amplifying property of a vacuum tube with a grid. Depending on how its circuit is connected, a grid-type tube can operate either as an amplifier or as an oscillator.

This chapter deals with the types of amplifiers and oscillators commonly installed in radio and radar equipment. It enumerates their uses, describes their operation, analyzes circuits in which they are used, and, in general, gives you the basic information required for an understanding of amplifiers and oscillators.

AMPLIFIERS

An amplifier is a device, usually containing a vacuum tube, whose output is an enlarged reproduction of the essential characteristics of the input wave. It may be a single-stage amplifier, that is one consisting of a single vacuum tube and its associated circuits, or it may be a number of stages joined together by coupling circuits.

BASIC AMPLIFIER

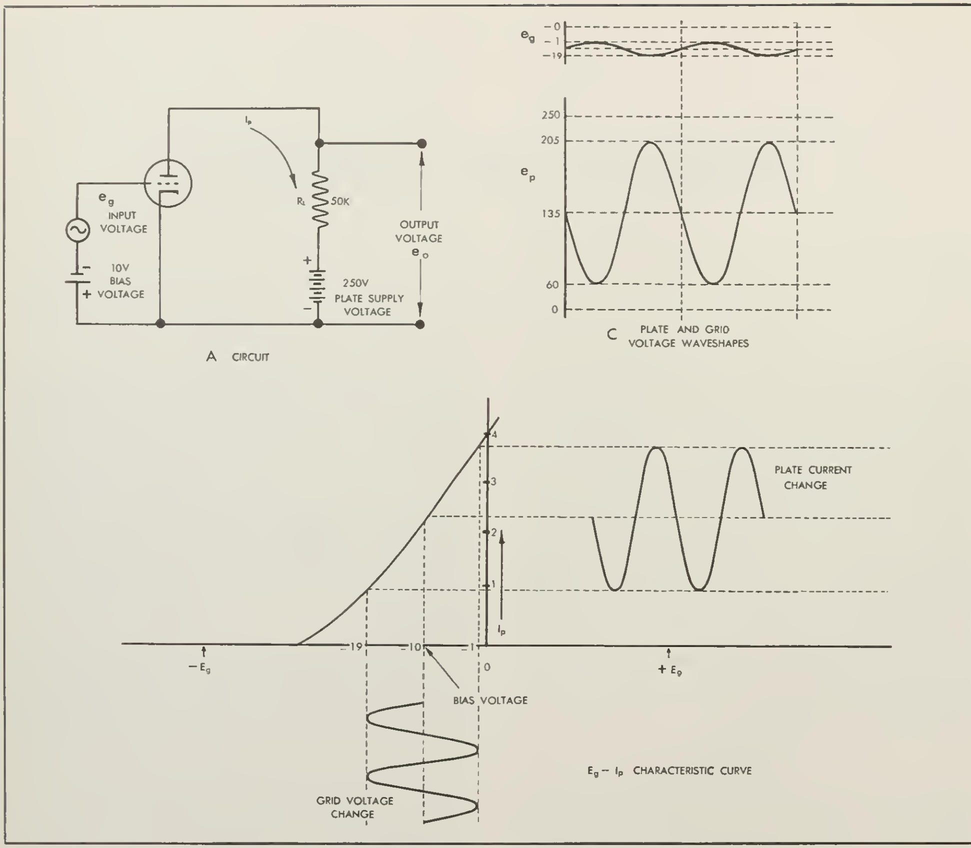
Basically, an amplifier consists of a vacuum tube with a grid and a number of associated circuits, the principal ones of which are the grid-cathode circuit, the grid-bias voltage supply circuit, and the plate circuit.

In the single-stage amplifier circuit shown on the next page, the signal voltage is applied in series with the grid-cathode circuit, and the grid-bias voltage supply. The changes in the grid-to-cathode voltage due to the changes in the signal voltage cause the plate current to change. The changes in the plate current flow through R_L which is in series with the plate-cathode circuit and the plate voltage supply. The resistor R_L is called the plate *load* resistor, and the voltage developed across it varies with the plate current. Since the load resistor is fairly large—approximately five times the internal (plate) resistance of the tube—the voltage developed across R_L is greater in magnitude than the signal voltage applied to the grid. Hence, by the action of the amplifier, a small grid voltage becomes a larger voltage in the plate circuit, and thus is said to be *amplified.*

The grid of the amplifier is biased so that it is always negative with respect to the cathode regardless of the alternating cycle of the input voltage wave. As long as the grid is negative with respect to the cathode, it does not draw any electrons. When no electrons are attracted by the grid, no current flows in the grid circuit. Under this condition, the grid circuit consumes no power. However, if the grid becomes positive, current flows, and the grid circuit does consume power. In the circuit illustrated, the grid is operated at negative potential with respect to the cathode, and there is no power lost in the grid circuit.

In dealing with amplifier circuits, knowledge that an amplifier produces a large voltage in its output when a small voltage is applied to its grid is not sufficient. It is important also to know just how much the voltage is increased by amplification. For determining this increase quantitatively, use characteristic curves like those discussed in Chapter 4.

The characteristic curves which you can use for determining voltage increases in amplifiers are the grid-plate current curve and the plate-voltage plate-current curve. Although the following discussion is based on the grid-plate current curve, in actual problems it is more prac-



Basic Amplifier Circuit and Waveshapes

tical to use the plate-voltage plate-current curve, since it does not involve the mathematical calculations which follow.

Notice the grid-voltage plate-current curve at B below the amplifier circuit. This curve gives the plate current for any value of grid voltage in the useful range for tube operation. To see how this curve is used, assume, for example, that the input voltage to the grid of the amplifier is a sine wave with a peak-to-peak amplitude of 18 volts (about 6.3 volts rms), and that the bias applied to the grid is negative 10 volts. This means that the average grid voltage will be minus 10 volts over a cycle of operation, or, in other words, that the operating point of the stage is minus 10 volts. To determine the plate current for any value of grid voltage, follow a vertical projection from a value of grid voltage up the Eg-Ip curve, and then move across horizontally to the plate current change. The point of intersection is the value of plate current for a specific value of grid voltage. Notice that in the curve the grid voltage varies from −1 to −19 volts, and the resulting plate current varies from 3.8 ma to 0.9 ma. The average plate current is 2.3 ma. At no-signal current condition—that is, when no signal is applied to the grid—the value of plate current is likewise 2.3 ma.

The voltage change across the load resistor (R_L) represents the useful output of the amplifier. From the values of current indicated in the curve, you can calculate the voltage output by multiplying the resistance by the plate current at specific values of grid voltage.

Example

1. *Grid at* -1 *volt*
   $E_o = I_p R_L$
   $= 3.8 \text{ ma} \times 50K$
   $= .0038 \times 50000$
   $= 190$ *volts*
2. *Grid at* -10 *volts*
   $E_o = 2.3 \text{ ma} \times 50K$
   $= .0023 \times 50000$
   $= 115$ *volts*
3. *Grid at* -19 *volts*
   $E_o = 0.9 \text{ ma} \times 50K$
   $= .0009 \times 50000$
   $= 45$ *volts*

The preceding calculations have disregarded one factor—the plate supply voltage. Any voltage drop across the power supply load resistor will be negative at the top of the resistor. This voltage will therefore subtract from the 250 volts plate supply voltage and change the actual output (E_o) of the amplifier.

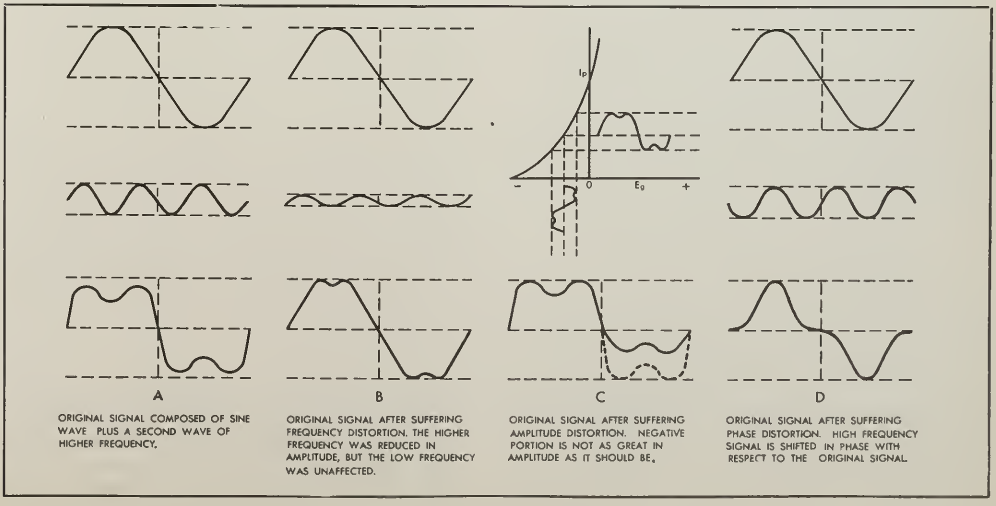
The amplitude of the IR_L drop is the same as that of the output voltage. Nevertheless to find the exact value of E_o, add the 250 volts to the voltage drop across the load resistor as in the following:

$$-190 + 250 = 60 \text{ volts}$$
$$-115 + 250 = 135 \text{ volts}$$
$$-45 + 250 = 205 \text{ volts}$$

At C notice the curves showing the grid and plate voltage curves for the values indicated. These curves show correct amplitude and time relationships. The input voltage (e_g) varies from -1 to -19—that is, the voltage variation of the input signal is 18 volts. The plate current voltage curve represents the voltage variation at the top of resistor R_L. Its range is from 60 to 205 volts, or, in other words, the total output variation is 205 minus 60, or 145 volts. From this figure it is obvious that the output voltage is considerably greater in amplitude (variation) than the input voltage. The ratio between the two voltages is 8.1 to 1. This ratio is called the *gain* of an amplifier stage. The power output of the stage described is less than 2/10 watts. Thus, you can conclude that the circuit in question produces a fairly large voltage change, but produces very little power. Such a stage can be called a voltage amplifier, as it is concerned chiefly with voltage amplification and not power.

DISTORTION IN AMPLIFIERS

Whenever the plate-current waveshape in an amplifier is not identical to the voltage wave impressed on the grid, the output is said to be *distorted*. There are three types of distortion in amplifiers—frequency distortion, amplitude (also called nonlinear) distortion, and phase-shift distortion. Any one of these may be present in amplifiers, either separately or in combination. You can see the effects of each type in the illustration below.



Types of Distortion

At A notice the drawing showing a typical pair of original signals that might appear on the grid of the amplifier. The pair includes a sine wave of one frequency and a second sine wave at a frequency three times that of the first. These two waves are combined in the grid circuit and then amplified by the tube. If both frequencies are not amplified the same amount, the wave form appearing in the plate circuit will not resemble the wave form of the signal on the grid—that is, the output wave will be distorted. This type of distortion is called frequency distortion. Notice at B that the higher frequency signal in the distorted output is only about $\frac{1}{8}$ the amplitude of the low frequency signal, although it started out with an amplitude about $\frac{1}{4}$ as large. The low frequency signal was virtually unaffected by the tube. Obviously, the circuit has discriminated against one frequency. Thus frequency distortion has occurred.

Amplitude distortion affects the magnitude or amplitude of the output wave. This type of distortion is caused by the plate current changing a large amount for a grid voltage change near zero grid voltages, but only a small amount for the same grid voltage change near cutoff. The effect of amplitude distortion is that a part of the output wave does not have as great an amplitude as the other part. In the diagram at C the negative portion of the wave is not as large as it should be. The dotted line shows how the negative part would appear if amplitude distortion were not present.

Phase-shift distortion occurs when reactive components are present in the circuit. Under these conditions, there is a phase shift of one or more of the signals comprising the input signals. As a result, the output signal is not a true replica of the original signal. The diagram at D shows the high-frequency component of the signal shifted 180°, producing a distorted output.

CLASSIFICATION OF AMPLIFIERS

Amplifiers are named or classified in several ways. According to the results they achieve, there are two basic types—voltage amplifiers and power amplifiers. According to the conditions under which they operate, amplifiers, especially power amplifiers, are classified as Class A, Class B, and Class C amplifiers. The class at which an amplifier operates depends on the amount of bias voltage applied to the grid of tube and the portion of AC signal-voltage cycle during which plate current flows. Depending upon the frequency range over which they operate, amplifiers may be further classified as direct current amplifiers, audio frequency amplifiers, intermediate frequency amplifiers, radio frequency amplifiers, and video amplifiers.

Voltage Amplifiers

Voltage amplifiers are primarily intended for amplifying voltage. They are designed to develop the greatest amplified voltage possible across the load in the plate circuit of the amplifier. To accomplish this objective, it is necessary that the load resistance be as high as possible so that it will offer a large opposition to the plate-current change. This results in a large voltage being developed in the output circuit.

Power Amplifiers

Power amplifiers are designed to deliver large amounts of power to the load in the plate circuit without regard to voltage. Since power equals voltage times current, a power amplifier must have a large output voltage across the load in addition to a relatively large current flow. In power amplifiers, the impedance (resistance) of the load is smaller than that in voltage amplifiers.

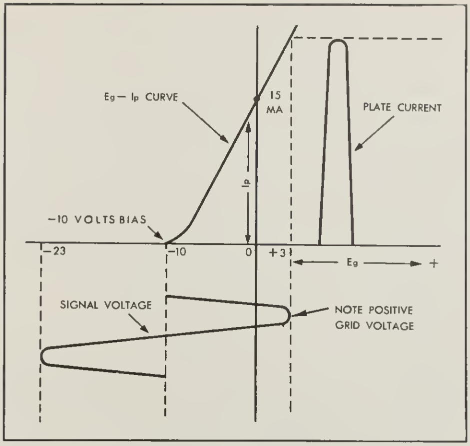
Class A Amplifiers

An amplifier is considered as Class A when the grid bias and alternating grid voltages have such value that plate current flows continuously throughout the cycle of the applied voltage and never reaches zero. Class A amplifiers are biased to about half of cutoff value. The basic amplifier shown on page 7-2 is a Class A amplifier.

Class A amplifiers are characterized by low efficiency and an output having a large ratio of power amplification. By efficiency is meant the ratio of the power output to power input. For practical purposes the efficiency of a Class A amplifier ranges between 20% and 25%. Theoretically, it has a maximum efficiency of about 50%. Class A amplifiers are used as audio- and radio-frequency amplifiers in radio, radar, and sound systems.

Class B Amplifiers

Class B amplifiers are biased to approximately cutoff. Plate current therefore flows only during the positive half-cycle of the applied signal voltage. The efficiency is higher and the current consumption is less in a Class B amplifier than in one operating Class A. Power loss



Class B Amplification

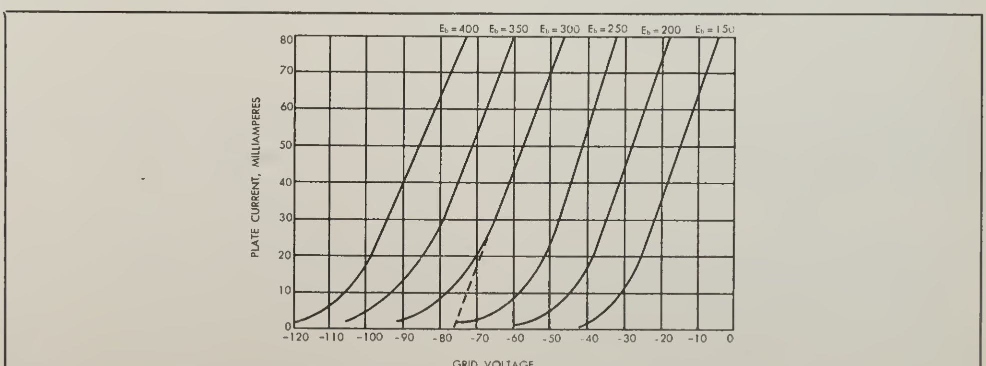
in Class B amplifiers is low for two reasons. First, plate current does not flow when there is no signal applied to the grid, and thus there can be only very little power wasted during the non-operating periods. Secondly, plate current flows only during the positive half of the input cycle. This means that the average current flow will be only 32% of the peak current in the stage.

The above illustration shows the relation between the grid voltage and plate current in a tube operated Class B. Note that plate current flows only during the positive half of the signal voltage. Grid current flows only during the time when the grid is driven positive.

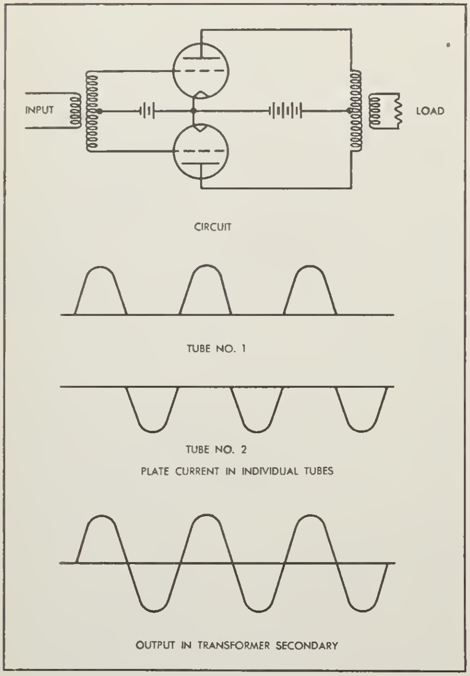
Class B amplifiers are used mostly as power amplifiers. As power amplifiers, their power output is proportional to the square of the grid-excitation voltage. The best bias for Class B operation is that which corresponds roughly to the cutoff bias that would be obtained if the main part of the characteristic curves shown in the chart below were projected as straight lines. Notice the dotted straight line extended from the straight line part of the characteristic curve labeled $E_b = 300$. The point at which this line strikes the grid-voltage volts line, approximately −75 volts, gives the cut-off bias for 300 volts plate operation. Curves of this type provide convenient means of determining grid bias for Class B operation for different plate voltage conditions.

Distortion occurs in Class B amplifiers much under the same general conditions as in Class A amplifiers. Frequency distortion and phase-shift distortion are essentially alike in both. Amplitude distortion in the output of a Class B amplifier operating with the proper load resistance depends upon the departure of the characteristic curve from straight lines and upon the operating point. Refer to the Class B amplification curve shown above. Obviously, the half-cycle output is far from the distortionless reproduction of the full cycle input. One-half of the cycle is missing. This missing cycle is supplied by one of two methods—by an additional Class B amplifier, or by the flywheel action of a resonant circuit. Distortion of the positive peaks of grid voltage may occur due to flow of grid current if the grid becomes positive.

The use of two tubes to supply both halves of the input cycle in the output constitutes a push-pull Class B amplifier. One tube operates



Determining Correct Bias Voltage for Class B Amplifier

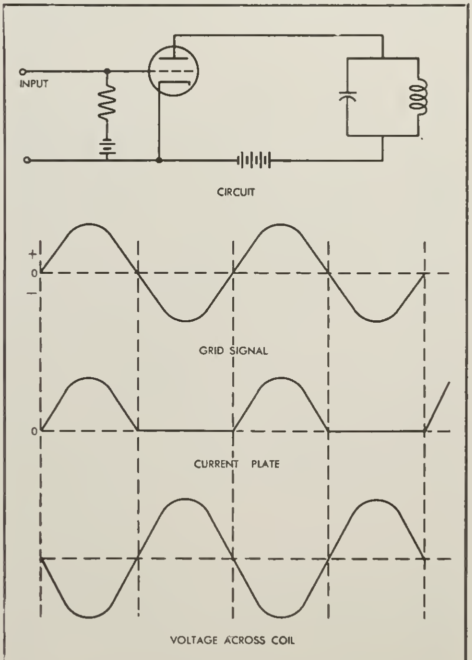


Using Two Class B Tubes to Produce Full Sine Waves

during the first half-cycle of the AC signal voltage and the other tube during the other half-cycle. Since plate current flows during one-half-cycle in one tube and during the next half-cycle in the other, the plate current wave forms can be combined in the load circuit. The load circuit of the push-pull amplifier is the center-tapped primary of the output transformer. During one cycle, one tube generates a voltage across the transformer winding. During the next half-cycle, the other tube generates a voltage of opposite polarity across the winding. Since the plate currents of the two tubes flow in opposite directions through their respective halves of the transformer primary winding, the voltages across the primary windings of each tube combine in the secondary to produce a reasonably undistorted output of the input AC signal voltage.

A single-tube, class B amplifier can be used successfully in RF amplifier stages having a parallel-tuned circuit as the plate load. The parallel-tuned circuit is sometimes called a *tank circuit,* because it has the ability to store power. When it is used as the plate load of a single-ended, Class B amplifier stage, the capacitor in the parallel-tuned circuit is charged by the output voltage produced by the flow of plate current through the load on the positive half-cycles. Although no current flows through the tube on the negative half-cycles of the applied signal voltage, the capacitor discharges into the inductor during this period, and thus supplies the missing half-cycle in the output voltage. This so-called flywheel effect of the tank circuit occurs only when the resonant frequency of the parallel-tuned circuit is equal to the frequency of the applied signal voltage.

The maximum amount of useful power available from a single power tube is limited by the amplitude distortion due to the introduction of undesired harmonic frequencies in the amplifier circuit. A harmonic frequency is a multiple of any given frequency. For example, the second harmonic of 1000 cps is 2000 cps, the third harmonic is 3000 cps, and so on. Second and higher harmonics are generated by a vacuum tube when its grid voltage-plate current changes are nonlinear.

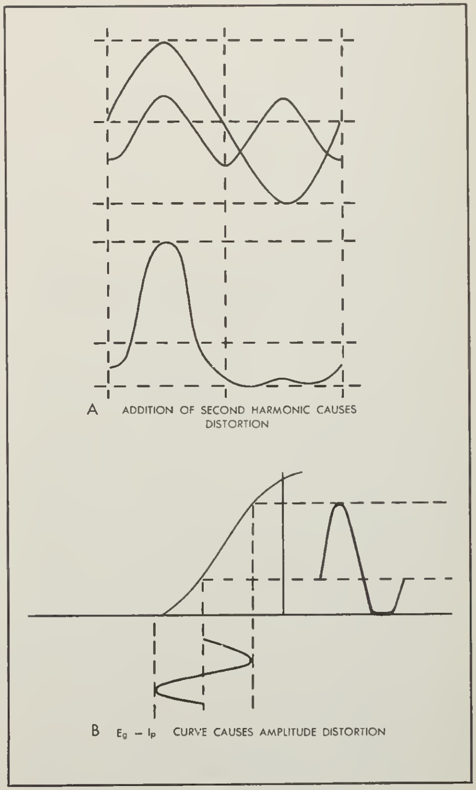


Using Resonant Circuit to Produce Full Sine Wave in Class B Amplifier

The adjacent illustration will help you in understanding the action of harmonics in vacuum tubes. Consider the top diagram (A) first. It shows two frequencies, a sine wave and its second harmonic. Just below the two frequencies is the waveshape produced by adding the two graphically. Notice that the sum is a distorted waveshape of the sine wave in which there is a large positive half-cycle and a small negative cycle. The bottom diagram (B) shows similar results in which a harmonic generated in a vacuum tube is added to the signal frequency (fundamental) and causes the amplitude of the output to be distorted. This action results from operating the tube on the curved portion of the E_g-I_p curve. Here too the output wave has a large positive half-cycle and a small negative half-cycle, a condition evidencing distortion. The output waveshape contains both the original fundamental signal waveshape and the harmonic which was generated in the tube and added to the fundamental by tube action.

Elimination of the even numbered harmonics, second, fourth and so on, is possible when a push-pull circuit arrangement is employed. Only the odd harmonics (particularly the third) will be left to limit the power output. However, the effect of the odd harmonics is minimized by connecting to each tube a load resistance which is more nearly equal to the dynamic plate resistance of the tube. The result of this is that the amount of undistorted power will approach the maximum amount of power obtainable, that is, the power if there were no distortion. Actually, therefore two tubes connected in push-pull will give considerably more than twice the undistorted power which a single tube can deliver.

When two tubes are connected in push-pull, as mentioned before, the plate-current pulse of each individual tube flows through only one half the transformer primary. This makes the combined output of the two tubes equivalent to an AC current having a crest value Imax/2 flowing through the entire transformer primary. If R_L is the equivalent load resistance between the primary terminals of the output transformer, the AC voltage drop between the plate and cathode of a single tube is one-half the voltage drop of the current Imax/2 across the resistance R_L or R_LImax/4. The minimum instantaneous plate voltage can be obtained as follows:



Introduction of Harmonics

$$Emin = E_b - R_L Imax/4$$
$$4Emin = 4E_b - R_L Imax$$
$$R_L Imax = 4E_b - 4Emin.$$

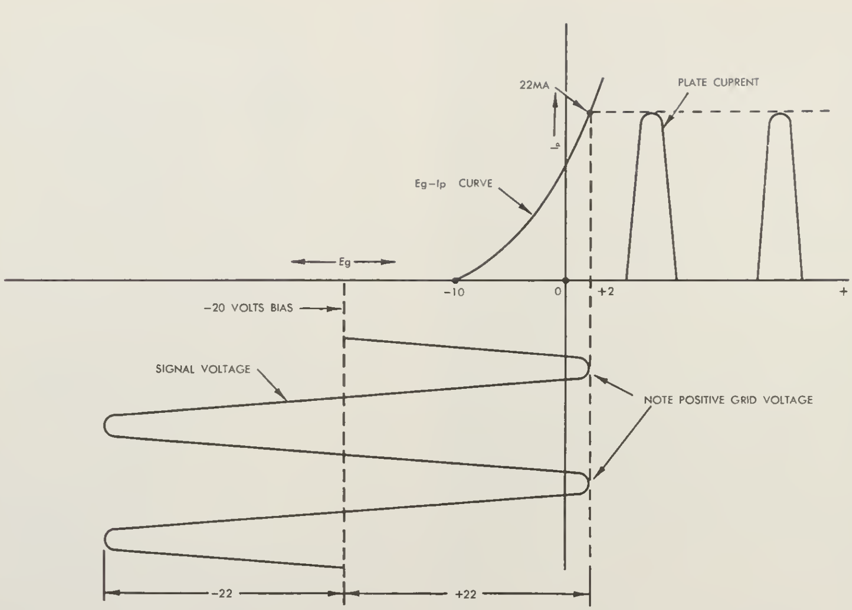
The correct load resistance from plate to plate, then is equal to:

$$R_L = \frac{4\ (E_b - Emin)}{Imax}$$

The power output is equal to the square of the load current multiplied by the load resistance. Mathematically, this relationship is expressed as,

$$Power\ output = \frac{R_L I^2 max}{8}$$

The maximum efficiency possible with class B power amplifiers is theoretically 78.5%. In most practical applications, however, the efficiency attained is about 60% to 65%. Class B amplifiers are principally used in transmitters as voltage amplifiers for amplitude-modulated waves.



Class C Amplification

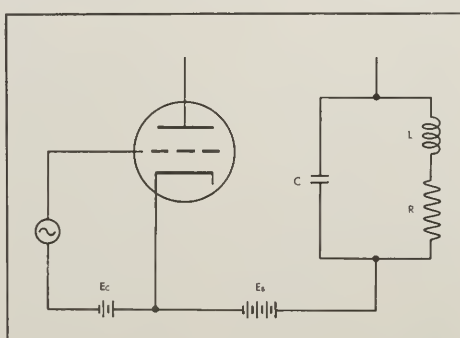
Class C Amplifier

A Class C amplifier is an amplifier in which the grid bias is appreciably greater than cutoff value. When no alternating voltage is applied to the grid the plate current is zero. When an alternating voltage is applied to the grid, plate current flows for appreciably less than ½-cycle.

Except for grid bias, Class C amplifiers operate in the same way as Class B amplifiers. Since current flows for a small part of a cycle, the distortion in Class C amplifiers is very great. A Class C amplifier is also characterized by the fact that it develops its output at a relatively low ratio of power amplification. Still another characteristic is that its grid usually swings sufficiently positive to allow saturation point current to flow through the tube. As a result the plate output waves are not free from harmonics and suitable means must be provided to remove them from the output.

One means of doing this is demonstrated by the illustration at the right. The tuned tank circuit shown offers an impedance to the operating frequency that is quite high and that has a power factor nearly equal to unity, but to the harmonic frequencies, it presents a low impedance, causing them to be attenuated or causing low amplification by the tube. Due to the high impedance offered to the signal frequency, it is thus amplified to a much greater extent than are the harmonic frequencies.

The high efficiency of the Class C amplifier is due to the fact that when plate current is permitted to flow, the instantaneous potential of the



Preventing Distortion by Using Resonant Circuit as Plate Load plate is low compared to the plate supply voltage. In this way energy is supplied to the plate circuit only when most of the plate supply voltage is used up as a voltage drop across the tuned circuit. Therefore, most of the energy is delivered to the tuned circuit instead of being wasted at the plate. Principally because of this fact, the efficiency of the Class C amplifier is in the neighborhood of 60% to 80%.

Class C amplifiers are used as RF amplifiers and modulators in transmitters. They are very useful in high-frequency equipment where it is necessary to deliver appreciable power.

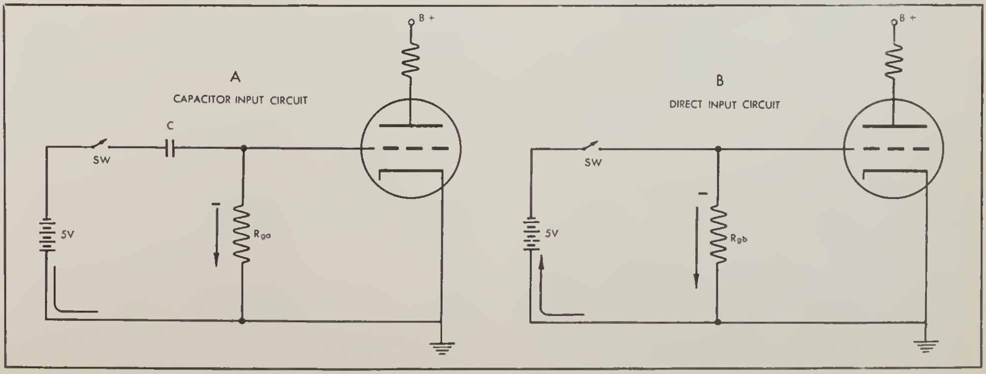
Direct Current Amplifiers

Direct current amplifiers are used to amplify very-low-frequency or DC voltages. A simple DC amplifier consists of a single tube with a grid resistor across the input terminals and with the load in the plate circuit. The load may be some sort of mechanical device, such as a relay or a meter, or the output voltage may be used to control the gain of an amplifier.

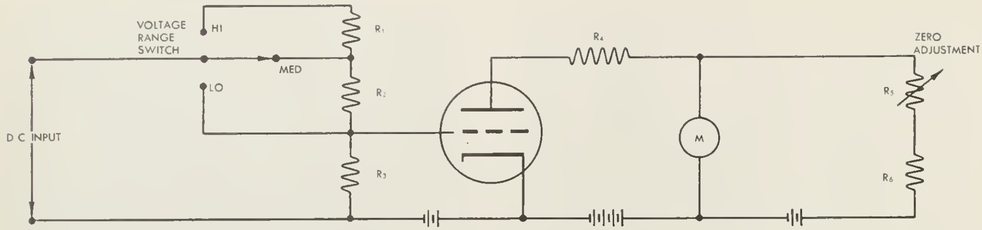
The DC voltage that is to be amplified must be applied directly to the grid of the amplifier tube. For this reason, only direct coupling can be used in the amplifier input circuit. Coupling is shown in the illustration below, where a comparison is shown between a capacitor-input circuit and a direct-input circuit. In both cases direct current voltage is applied to a direct current amplifier.

When the switch in the capacitor-input circuit is closed, the direct current battery voltage charges capacitor C. The electrons moving in the direction indicated cause a voltage equal to the battery voltage to appear momentarily across resistor R. This voltage in turn appears on the grid of the tube. However, as the capacitor continues to charge, and up to the point where its charge is equal to the battery voltage, the voltage across resistor R decreases until it reaches ground potential. During this time, the grid voltage changes from ground to negative 5 volts and then back to ground. These variations appear in the output of the amplifier as a changing voltage. However, they are only momentary because there are no further changes in the circuit values. If the switch remains closed, the applied DC voltage is constant, and just as soon as the capacitor is fully charged, the output of the amplifier will return to its original level. Therefore, this circuit is incapable, as you have seen, of amplifying a direct current voltage.

In the direct input circuit, however, when you close the switch, the battery voltage is applied directly to the grid. Unlike the capacitor-input circuit, therefore, the voltage on the grid remains constant as long as the switch remains closed. Prior to the closing of the switch, a fixed value of current flows in the plate circuit. This results in a fixed voltage drop across the load resistor. When the DC voltage is applied to the input terminals, this voltage is not blocked by a capacitor but is impressed directly upon the grid of the tube. It makes the grid more negative than before and permits less plate current to flow. Accordingly, there is a smaller voltage drop across the load resistor. Since this voltage change is greater than the voltage applied to the grid, the input voltage has been amplified. Therefore, this circuit can serve as a DC amplifier.



Direct and RC Coupling for DC Amplifier

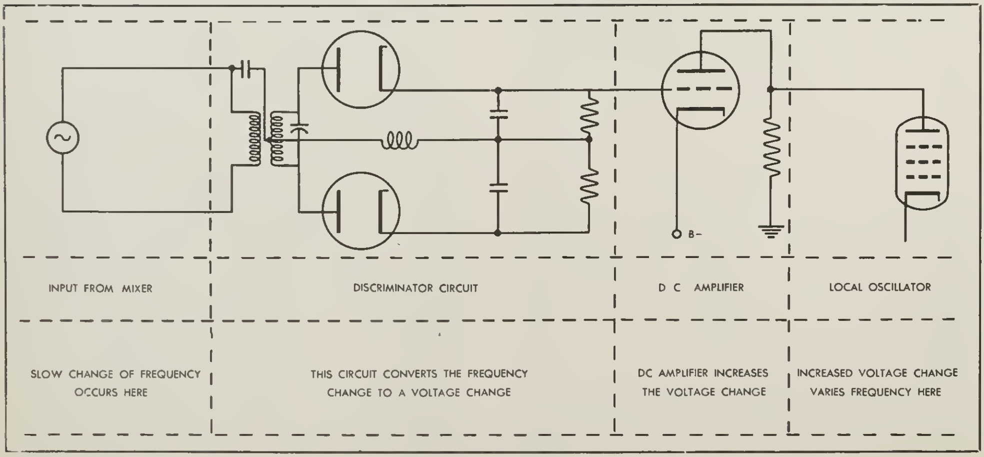


Vacuum Tube Voltmeter

One of the most important uses of a DC amplifier is as a DC vacuum tube voltmeter. Notice the circuit of a vacuum tube voltmeter shown above on this page. The voltage to be measured is applied to the voltage divider, made up of R_1, R_2, and R_3. The ratio of voltage division can be varied by a voltage range switch in such a way that several ranges of voltage can be measured. Resistor R_4 is used to prevent damage to the meter if too high a voltage is applied. In the plate circuit, the additional battery and the variable resistor are used to balance the normal plate current of the circuit. The variable resistor can be adjusted so that meter M reads zero when no signal is applied. Whenever a DC voltage is applied to the input, the tube amplifies it and causes a current to flow through the meter. Since the meter reading is proportional to the voltage applied, you can read the amount of voltage on the calibrated scale of the meter.

In addition to their use in vacuum-tube voltmeters, DC amplifiers are also used for amplifying the output of the discriminator in automatic frequency control circuits. One of the problems in microwave receivers is frequency control—that is, keeping the intermediate frequency constant in spite of minor variations in the transmitted frequency or the local oscillator frequency. To accomplish this, special circuits are used to adjust the frequency of the local oscillator in such a manner that the correct IF will be produced. These circuits, called automatic frequency control circuits, are discussed later, but here the purpose is to show the function of the DC amplifier in the circuit.

To understand the part the DC amplifier plays, first consider the frequency discriminator shown in the illustration below. Its purpose is



Application of DC Amplifier

to convert any frequency deviation from the intermediate frequency of the receiver to a DC voltage that has a magnitude proportional to the amount of deviation from the intermediate frequency and a polarity that depends on whether the deviation is above or below the intermediate frequency.

The DC amplifier tube is connected between ground and a negative potential, making the plate positive with respect to the cathode and grid. The two resistors in the cathode circuit of the balanced diodes act as a load for the discriminator output. This output controls the DC amplifier, causing it to conduct more or less, depending on the effect of the intermediate frequency R.M.S. voltages on the plates of the balanced diodes. The changes in the output of the DC amplifier in turn cause variations in the repeller plate voltage of the local oscillator and change the local oscillator frequency in the direction required to keep the intermediate frequency constant.

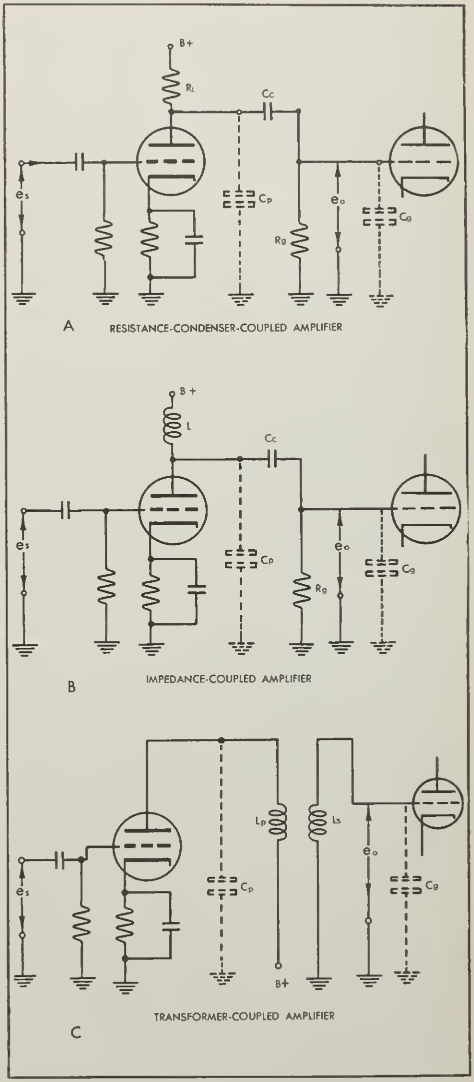
Audio Amplifiers

Amplifiers designed for frequencies within the audio spectrum (approximately 15 to 15,000 cycles) are called audio frequency amplifiers.

To obtain sufficient gain, it is usually necessary to use one or more stages of audio amplification—that is, to use the output of one tube to control the grid circuit of a second, then to use the amplified output of the second tube to control the grid of the following tube, and so on. When two or more amplifiers are connected in this manner, they are said to be connected in *cascade*.

When one stage of audio amplification (a tube and its circuit connections) is coupled to the next, the plate of the amplifier tube cannot be connected directly to the grid of the following stage as with DC amplifiers, but must be coupled through a special circuit. The coupling circuit transfers the varying voltage between the stages, and at the same time supplies the DC potentials and currents necessary for the operation of the tubes. The main types of audio stage coupling are resistance-capacity, impedance, and transformer coupling.

The illustration to the right shows simple circuits of triode audio amplifiers, each employing one of these three types of couplings. Since the method most widely used for coupling audio stages is the resistance-capacity method, it is the only one discussed in this chapter. For information on the other types, refer to Chapter II.

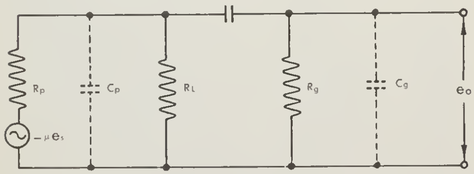


Types of Coupling

Referring to the circuit of the resistance-condenser, or simply the resistance-coupled amplifier, as it is often called, notice that cathode bias is used. This is the most commonly used type of bias that you will encounter in resistance coupled circuits. The condenser C_c, which is called the *coupling condenser*, is for the purpose of providing an AC current path to the grid of the next stage. The resistor R_L is called the *coupling resistor*. Its resistance is high so that as much voltage as possible can be transferred to the grid of the following tube.

The coupling between the two stages illustrated takes place in the following manner. When a signal is applied to the grid of the first stage, voltage variations are produced in the plate circuit of this tube. These variations are impressed upon the grid of the second tube through the coupling condenser and the grid resistor R_g of the second tube.

The coupling condenser serves two functions. As stated before, it provides a low impedance path for the AC to the grid of the following tube. In addition, it keeps the plate voltage of the first tube from the grid of the second tube.



Equivalent Circuit for RC Coupled Amplifier

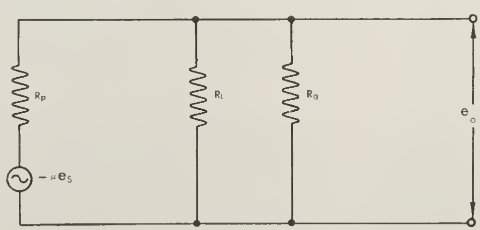
EQUIVALENT CIRCUITS. The variations produced in the plate current of a vacuum tube when a signal voltage is applied to the control grid are exactly the same as would be produced by a generator developing a voltage of $-\mu e_s$ in a circuit consisting of a tube plate resistance in series with the load impedance. The effect on the plate current of applying a signal voltage e_s to the grid is therefore exactly as if the plate cathode circuit were a generator developing a voltage and having an internal resistance equal to the plate resistance of the tube. Thus, the resistance coupled amplifier shown in the preceding illustration can be represented by the equivalent circuit shown above. The minus sign on the equivalent generator voltage is used only to indicate that the voltage is of opposite polarity to the signal voltage. This indicates the fact that there is a 180° phase shift between the grid voltages and plate voltages in a vacuum tube.

Note that the cathode bypass condenser and biasing resistor are omitted in the equivalent circuit. This omission is permissible, first, because the condenser is of such size that it places the cathode at ground potential for AC, and, secondly, because the biasing resistor is considered a part of the plate load. All the circuit constants shown in the equivalent circuit are those which affect any ordinary triode amplifier. In fact, in most practical cases, depending upon the frequency, some of the constants shown may be omitted. Thus, for example, in the mid-frequency band (200 to 10,000 cycles) the electrode capacities C_p and C_g may be omitted in equivalent circuits because their impedances are so large that they act like open circuits. Similarly, the coupling condenser may be left out because of the negligible voltage drop across it.

CALCULATING GAIN. Two important facts to remember relative to audio amplifiers are first that the output of an audio amplifier is always less than the μe_s voltage because of the internal resistance (R_p) of the tube, and, secondly, that the gain is the ratio of the output voltage to the input voltage, which is expressed mathematically,

$$A = \frac{e_o \text{ (output voltage)}}{e_s \text{ (input voltage)}}$$

The following paragraphs discuss and derive gain formulas for triode audio circuits operated at mid-frequencies, circuits in which the grid resistance R_g is much greater than the load resistance R_L ($R_g \gg R_L$), circuits operated at high frequencies, and circuits operated at low frequencies. In each case the equivalent circuit is represented by a constant-voltage generator circuit.



Equivalent Circuit at Mid-Frequencies

GAIN AT MID-FREQUENCIES.

Total resistance, $R_t = \dfrac{R_L R_g}{R_L + R_g} + R_P$

Plate current,

$$i_p = -\frac{\mu e_s}{R_t} = -\frac{\mu e_s}{\dfrac{R_L R_g}{R_L + R_g} + R_P}$$

$$= -\frac{\mu e_s (R_L + R_g)}{R_L R_g + R_P (R_L + R_g)}$$

The output voltage equals the product of i_p and the effective resistance of R_L and R_g in parallel, therefore,

$$e_o = -i_p\frac{R_LR_g}{R_L+R_g} = -\frac{\mu e_s(R_L+R_g)}{R_LR_g+R_P(R_L+R_g)} \times \frac{R_LR_g}{R_L+R_g}$$

$$= -\frac{\mu e_sR_LR_g}{R_LR_g+R_P(R_L+R_g)}$$

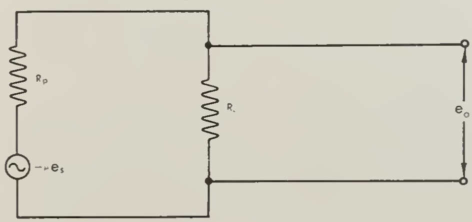
Gain (A) equals $\frac{e_o}{e_s}$, *thus*

$$A = \frac{e_o}{e_s} = \frac{\mu e_sR_LR_g}{R_LR_g+R_P(R_L+R_g)} \div e_s$$

$$A = -\frac{\mu R_LR_g}{R_LR_g+R_P(R_L+R_g)}$$

(The minus sign indicates the 180° phase shift between the input and output voltages.)

GAIN WHEN $R_g \gg R_L$. The following calculations assume that R_g is at least ten times as large as R_L and the frequency is in the mid-frequency band. The calculations are considerably simpler than those for the preceding formula since the parallel combination of R_L and R_g is approximated as equal to R_L. Note that equivalent circuit below is a modification of the circuit for mid-frequencies.



Equivalent Circuit when
$R_g \gg R_L$

Total resistance, $R_t = R_P + R_L$

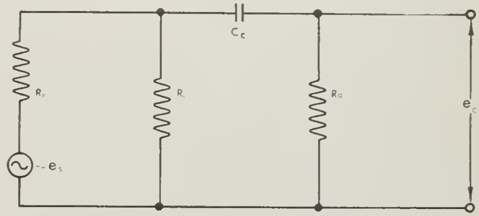
$$i_p = -\frac{\mu e_s}{R_P+R_L}$$

$$e_o = i_pR_L = -\frac{\mu e_sR_L}{R_P+R_L}$$

$$\text{Gain } A = \frac{e_o}{e_s} = -\frac{\mu e_sR_L}{R_P+R_l} \div e_s$$

$$\text{Therefore } A = -\frac{\mu R_L}{R_P+R_L}$$

In connection with the gain equation, $A = -\frac{\mu R_L}{R_P+R_L}$, just derived, it is important to remember that since any vacuum tube has internal resistance (plate resistance), only a portion of the equivalent generator voltage is available across the load resistor. It is therefore very often useful to consider that the plate resistance and the load resistor make up a voltage divider across which the voltage generated with the tube is applied. The ratio $\frac{R_L}{R_p+R_L}$ expresses the proportion of the voltage across the load resistor. You can see from this equation that for maximum output, R_L must be many times as large as R_p.



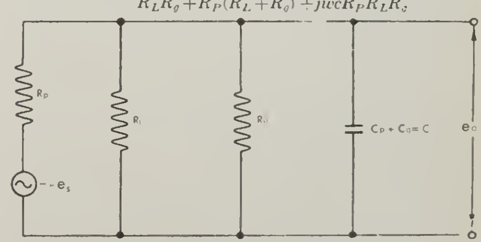
Equivalent Circuit at Low Frequencies

GAIN AT LOW FREQUENCIES (15 TO 200 CPS). Notice the coupling condenser in the equivalent circuit above. At low frequencies, the reactance x_c of this condenser becomes appreciable and must be taken into consideration in the gain formula. At low frequencies, therefore, the gain formula reads,

$$A = \frac{-\mu R_LR_g}{R_LR_g+R_P(R_L+R_g)+R_P+R_L)X_c}$$

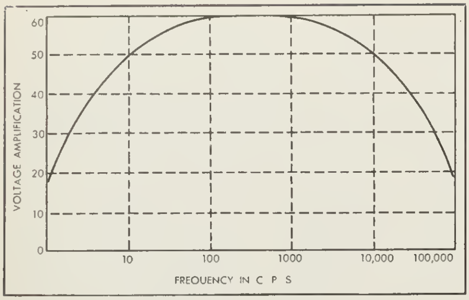
GAIN AT HIGH FREQUENCIES (10,000 cps AND ABOVE). At these frequencies, the plate-to-cathode capacity, the input capacity (grid-to-cathode), and the stray wiring capacity effectively shunt the grid lead resistance, R_g. This means that the reactances of these capacitances is so low that the voltage drop across them is reduced appreciably. The effect of this reduction is reduced output voltage from the amplifier. The equation for gain at high frequencies is,

$$A = -\frac{\mu R_LR_g}{R_LR_g+R_P(R_L+R_g)+j\omega cR_PR_LR_g}$$



Equivalent Circuit at High Frequencies

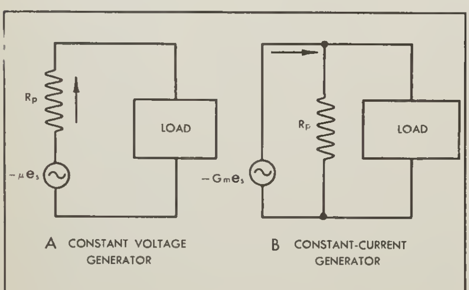
FREQUENCY RESPONSE. The most important property of a resistance-coupled amplifier is the manner in which its amplification varies with frequency. The following frequency response curve represents the response of a characteristic



Response of Audio Amplifier

amplifier over a wide range of frequencies. The low response at low frequencies is due to the large reactance that the coupling capacitor offers at these frequencies. The low response at high frequencies is due to the tube and stray wire capacities in the amplifier circuit.

PENTODE EQUIVALENT CIRCUITS. The discussion thus far has dealt with triodes in which the equivalent circuit was represented by a constant-voltage generator circuit. With some modifications of the equivalent circuit, the principles discussed are applicable also to pentode amplifiers. The major change involves using a constant-current generator rather than the constant-voltage circuit. In pentodes, it is assumed that the plate variations are identical to those produced by a generator developing a current $-G_m e_s$ acting in a circuit consisting of the plate resistance in parallel with the load resistance. Below notice the illustration comparing the constant-voltage circuit and the constant-current circuit. The calculations, previously discussed, are applicable, with some modifications, to pentodes when the constant current type equivalent circuit is used.



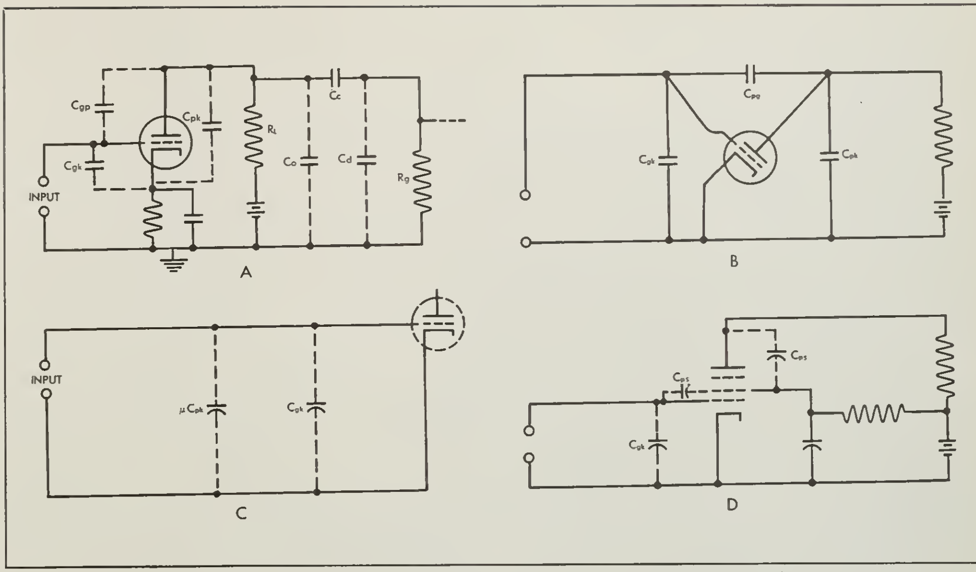
Comporison of Equivalent Circuits

Video Amplifier

A video amplifier is designed for uniformly amplifying a wide range of frequencies, including the audio range, up to several megacycles per second. Typical video amplifiers are resistance-capacitance circuits. They differ from audio amplifiers only in the values of the circuit components.

The resistance-coupled amplifier shown in the next illustration presents several problems which must be solved before the R-C coupled amplifier can be employed as a video amplifier. The high frequency response is limited by the output capacitance C_o, the distributed capacitance C_d, and the input capacitance of the following stage C_i acting in parallel to shunt the load resistance R_L. The low-frequency response, on the other hand, is limited by the time constant R_gC_c which must be long as compared to the period of the lowest frequency to be amplified. Tubes with very low interelectrode capacitance must therefore be used and care must be taken in wiring the circuit to keep leads short and properly spaced in order to reduce distributed capacitance in the wiring. All these problems are reduced by employing pentode tubes, by using small plate load resistors, and frequency compensating circuits.

INPUT CAPACITY. Normally, triodes are not adaptable for use in video amplifier circuits, principally because of their high interelectrode capacity. When an alternating current is applied between the grid and cathode of a triode amplifier tube, an alternating current flows in the small condensers (C_{gk}, C_{pk}, and C_{gp} shown at B) just as AC would flow in any type of condenser. The current flow in the capacity between the plate and cathode (C_{pk}) and in the capacity between the grid and the cathode (C_{gk}) is small, since the size of these capacities is small. Therefore, current flow in them does not seriously affect circuit operation. On the other hand, the instantaneous voltage between the grid and plate is considerably larger than the signal voltage when the tube is amplifying, and consequently the current flowing in the grid-to-plate capacity is considerably larger than it would be if there were no amplification taking place in the tube. From the grid end of the tube, this increased current is equivalent to an increase in the output capacity of the tube, and the effective capacity might be several times greater than that which you could expect from the interelectrode capacities alone. Note in the illustration at C that the input



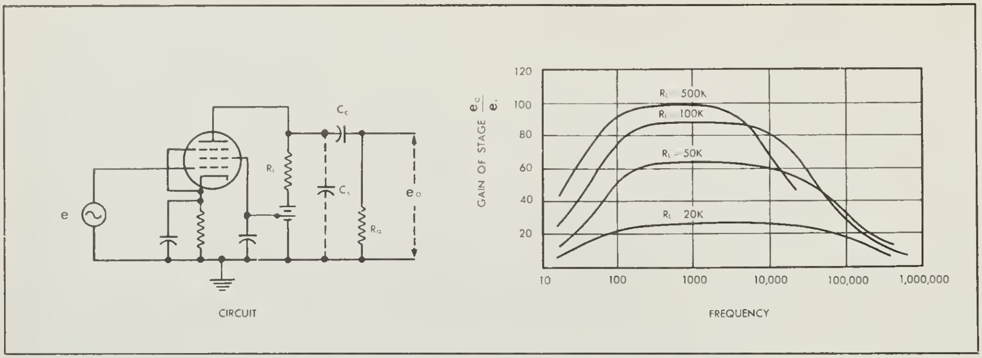
Effective Interelectrode Capacity in Triode Amplifier with Resistive Load

capacity consisting chiefly of C_{gk} is increased by the additional capacity μC_{pk} which is the capacity reflected back into the grid circuit by action of the varying and amplified plate current charging the grid-to-plate capacity. The higher the amplification factor, the greater is the reflected capacity, and the higher the input capacity, the greater the attentuation of high frequencies by the amplifier.

The pentode circuit shown at D shows that the high frequency attenuating effect of high input capacity is practically eliminated in the pentode tube. The screen grid in the pentode tube forms an electrostatic shield between the plate and grid. Instead of the varying plate current charging the capacity between the plate and grid and effectively reflecting a capacity greater by the μ of the tube into the grid circuit, the plate-to-screen capacity takes the charge and, by doing so, leaves the signal voltage virtually unaffected. The capacity between the control grid and the screen grid carries only a very small charging current since the screen grid voltage is maintained at a constant value. Both the C_{gs} capacity and also the C_{gk}, the grid-to-cathode, are small. At high frequencies, therefore, any current flow in these capacities produces no noticeable attenuation. In addition to low input capacity, pentodes offer also the advantage of higher amplification than a triode.

Plate load resistors. The second circuit feature of video amplifiers is the use of small load resistors. The size of the load resistor in resistance-coupled amplifiers is important in that it largely determines the frequency response and gain of the amplifier. In this connection, on the next page refer to the chart showing the effect of various sizes of load resistors on frequency response in a pentode amplifier stage. Notice that a large load resistor causes a high output voltage at the middle frequencies and a steep drop in voltage at high frequencies. In the case of small resistors, the proportionate drop in voltage output (gain) at high frequencies is much smaller than the drop at mid-frequencies.

The smaller proportionate drop at high frequencies in the second case mentioned results chiefly from the employment of a small plate-load resistor. The limit to amplification at the high frequency end in an amplifier is controlled by capacitances which shunt the load resistance of the tube. These capacitances, which are called shunt capacitances, consist of the wiring, the socket, and the inherent capacities of the tube. They act like a low impedance in parallel with the load resistor and effectively reduce the load impedance of the tube. For example, when the reactance of the shunt capacity is equal to the reactance of the tube, the output voltage is

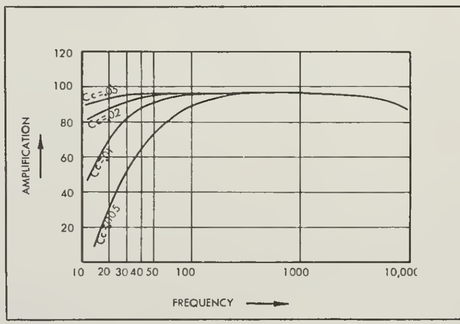


Effect of Size of R_L on Frequency Characteristic of Amplification Stage

reduced by half (two equal parallel impedances have a total impedance of half the individual impedances). In general, the higher the frequency, the lower the output impedance of the tube, and the lower the voltage output, or gain.

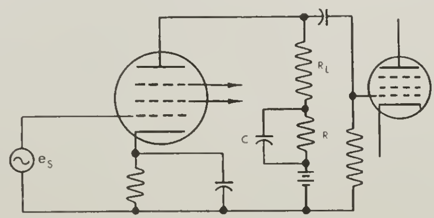
Obviously, this might lead you to conclude that it is not desirable to use low plate load resistors, since at high frequencies the shunt capacity reduces the output impedance enough without further decreasing it by small resistors. However, the video amplifier is primarily concerned with uniformly amplifying a wide band of frequencies and only secondarily with gain. To obtain a wide and uniform frequency response, there must be a sacrifice in gain. To be sure, a small load resistor reduces the gain at high frequencies; but, at the same time, it improves frequency response. At the middle frequencies, the reactance of the shunting capacities are not effective in reducing the output impedance of the tube, and the frequency response is fairly flat there.

LOW FREQUENCY COMPENSATION. As previously mentioned, video amplifiers must amplify frequencies at the low end of the band as well as at the high end. According to the frequency-amplification chart, amplification of low frequencies shows improvement when the coupling condenser C_c is large. A large condenser has less reactance to low frequencies than a small condenser and low reactance means a smaller voltage drop. Since the coupling condenser and the grid resistor form a voltage divider in the input circuit, more voltage appears across the grid resistor when the condenser has a small reactance. Therefore greater amplication takes place in the tube. There is a limit, however, to



Increasing Coupling Condenser between Stages Improves Low Frequency Amplification

the increase in amplification possible by increasing the size of the coupling condenser, particularly in radar. A large C_c discharges too slowly for the reproduction of the steep trailing edges required of square wave pulses.



Low Frequency Compensation Circuit

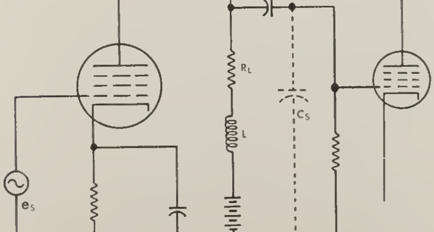
A second and somewhat more satisfactory means for improving low frequency amplification is by using low frequency compensating circuits. A low frequency compensating circuit improves response by virtue of the components

C and R in series with the load resistor. Since C is comparatively large, it offers practically zero reactance to middle and high frequencies and, therefore, does not affect these frequencies. At low frequencies, however, its reactance is high, and it in parallel with R produces a reactance which when added to R_L increases the total load impedance. Earlier you saw how the output voltage of an amplifier may be increased by increasing the load resistance. Therefore, since the compensating circuit produces a larger load impedance, it follows that the gain of the stage is higher.

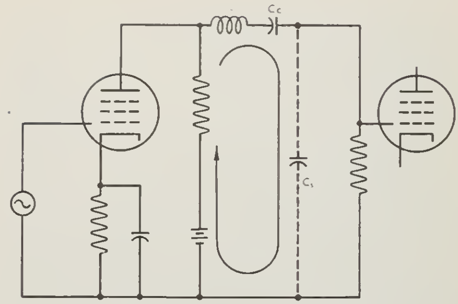
A third method of improving low frequency amplification is a combination of the two just mentioned. It consists of adjusting the coupling condenser size and adding a capacitive reactance in the plate load in such a manner as to obtain a correct proportion between the two. In other words, the voltage drop across C_c is compensated for by the compensating circuit in the plate.

Remember that in frequency amplification, amplification of low frequencies is not as important as amplification of high frequencies. When a radar pulse is very narrow, frequencies below 10,000 cps are of little importance. On the other hand, not all signals are narrow pulses and, since a video signal contains large amplitude low frequency components, the low frequency response should be fairly good.

High frequency compensation. The use of small load resistors in plate circuits of amplifiers does not completely solve the problem of improving high frequency response. Therefore, carefully designed high frequency compensating circuits are often used. Such a circuit consists of a low reactance inductor placed in series with the load resistor. At low frequencies, the in-



Shunt Peaking Coil Provides High Frequency Compensotion



Series Peaking Coil Provides High Frequency Compensation

ductor (L) offers a very low reactance, being practically a short circuit. At high frequencies, however, its reactance is high, and, since larger voltages are developed, it follows that the gain of the amplifier is increased. Since the inductor sets up a resonant circuit with the stray capacities in the circuit, it is so designed that the circuit is resonant at a frequency where the gain is low. This design feature gives an added boost to the signal.

Another and similar compensating circuit is the series peaking coil circuit. In this circuit, the coil (L) is connected in series with the coupling condenser (C_c). (Remember that the cause of reduced high frequency voltage is the decrease in reactance of the stray capacity in the circuit and the grid-to-cathode capacity in the next tube.) At high frequencies, the inductance L resonates with the stray capacity and the grid-to-cathode capacity represented as (C_s), causing an increased current through C_s, and hence a greater voltage across C_s which results in a higher gain.

Improving frequency response by negative feedback. Frequency response can also be improved by negative feedback. This involves feeding a small part of the output voltage back into the input in opposite phase or polarity and proportionately reducing gain. Thus, when the output of an amplifier decreases at high frequencies, less voltage is fed to its input, and when the output increases more voltage is fed back to the input. When the feedback is large, the gain is proportionately decreased and when the feedback is small the gain is proportionately increased. In this way, at high frequencies, when the gain of an amplifier is decreased, less out of phase voltage is fed back to the input and the gain increased.

Analysis of Video Amplifiers

Actually, the preceding discussion was an analysis of video amplifiers in the terms of frequencies found in pulse waveshapes. The following is an analysis of video amplifiers in terms of transient voltages.

The illustration below shows a typical uncompensated circuit which serves as the basis for the analysis. The input voltage is labeled E_i. For easier comparison with other waveshapes, it is inverted in the diagram. The inversion is represented by the negative sign in front of E_i. When the steep leading edge of the input square pulse is applied to the first tube, the plate voltage does not rise as sharply as the input, since it cannot rise any faster than the stray capacity C_s is able to charge. This makes the plate voltage rise along an exponential curve and causes the waveshape on the following grid to be rounded in the same manner as shown at e_{g2}. To get a sharp leading edge, the stray capacity must be small.

The time constant of this curve is the stray capacity times R_t, (the total resistance in parallel with the stray capacity). Mathematically, then, $T.C. = R_t C_s$. R_t is the effective resistance of the load resistor, tube plate resistance, and grid resistance, all in parallel. The relationship between these elements to the frequency (f_2) at which power drops to one-half (and voltage to 70%) is expressed by the equation,

$$T.C. = R_t C_s = \frac{1}{2\pi f_2}$$

During the flat top of the input pulse, the large size of the coupling condenser C_c is important so that it will charge slowly and keep the second grid voltage from falling rapidly. The time constant of this curve is expressed by the equation,

$$T.C. = R_c C_c$$

where R_c is the total of R_g in series with R_l and R_p in parallel.

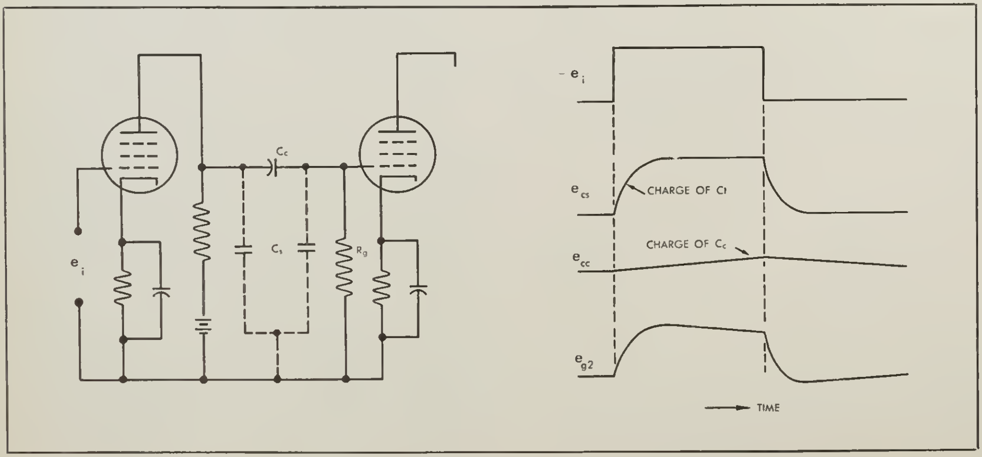
The relation between this time constant and the low frequency is expressed by the equation,

$$T.C. = R_c C_c = \frac{1}{2\pi f_1}$$

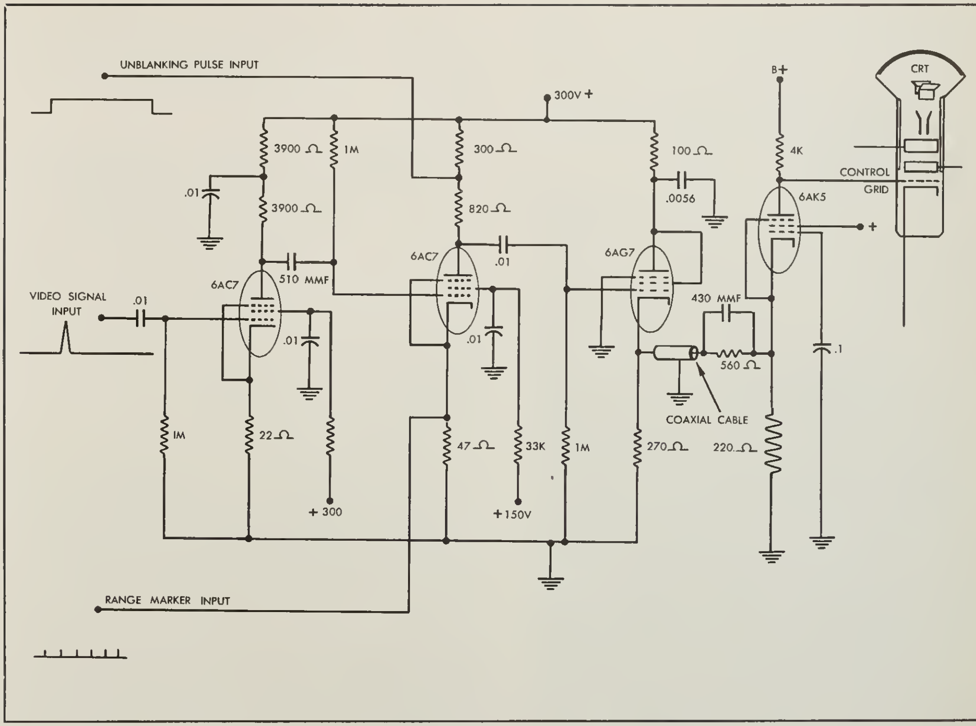
where f_1 is the low frequency at which the power drops to a half and the voltage to 70%.

When the end of the pulse arrives, the plate voltage drops only as fast as C_s discharges through the resistors in the circuit. Therefore this time constant should be short. (The waveshapes in the illustration are exaggerated for purposes of discussion.)

In summary, remember that the square corners of the pulse are due to the high frequency components in the square wave and that the small shunt capacities, such as stray wiring capacity and grid-to-cathode capacity, are responsible for the squareness of the leading and trailing edges of the pulse. The flat top of the pulse is due to the right amplitude of the low frequencies. The coupling condenser and its relation to the resistance in the circuit directly effect the flat top of a pulse.



Uncompensated Circuit



Representative Video Amplifier

AN/APQ-13 Video Amplifier Circuit

On first glance, the AN/APQ-13 video amplifier circuit shown above probably looks like any other amplifier to you. However, closer inspection reveals some unusual features.

The first stage is a very high μ amplifier. The grid receives the video signal directly from the detector. There are two resistors in the plate circuit—each 3900 ohms. The resistor nearest the plate is the plate load resistor. For optimum gain in this amplifier, the load resistor should be a hundred times larger, but in the interest of high frequency response, gain is sacrificed. The second 3900-ohm resistor in combination with the .01 condenser is not, as it appears, a low frequency compensating circuit, but a filter or de-coupling circuit for preventing any of the following stages from feeding out of phase energy to this stage and causing it to oscillate. Frequently, amplifiers depending upon a common plate voltage supply cause variations in this plate voltage supply. A varying plate voltage results in considerable coupling between stages and any energy coupled from one stage to the stages ahead causes these stages to break into oscillation, resulting in considerable frequency instability. Therefore, whenever a number of amplifiers are supplied plate voltage from a common supply, the plate circuits incorporate de-coupling circuits.

The second stage is more like a mixer stage than an amplifier. It is operated with a high plate current which is made possible by placing a high positive voltage on the grid through the one megohm resistor. Its grid is always a fraction of a volt positive with respect to the cathode. The mixer characteristics of this stage are due to the fact that there are three signals applied to it—the sweep voltage pulses which unblank the indicator screen when it is receiving echoes from targets, the positive pulses of the detector which were inverted to negative pulses by the first stage, and the range marker pulse. All three of these are combined into a composite signal and applied to the grid of the third stage through the coupling condenser (.01). This coupling condenser is large in order to retain the flatness of the long, unblanking pulses. Since only short pulses are coupled between the first and second stages, the coupling condenser between them is small.

The third stage is a cathode follower (discussed in more detail later). Briefly its purpose here is to provide power amplification. Even though triode connected, it does not attenuate high frequencies. Its input impedance is high for matching the impedance of the preceding stage: its output impedance is low for matching the coaxial line to the indicator chassis. The coaxial line carries the signal to the indicator which might be from two feet to fifty feet away. The shunt capacity of the coaxial line causes attenuation of high frequencies in the signal fed to the indicator. Therefore, a frequency compensating circuit is employed. It consists of the 430 mmf condenser and the 560 resistor to the left of the 6AK5 tube. The resistor reduces all frequencies equally, and the capacitor, which acts like a low reactance path for high frequencies causes less attenuation of the high frequencies. The high frequencies, therefore, pass through the circuit practically unchanged, but the low are reduced in gain. This effect is opposite to the action of the coaxial cable on the frequencies. Therefore, all frequencies enter the last stage with their original amplitude relationship.

The last stage terminates the coaxial cable in its characteristic impedance and amplifies the signal before feeding it to the cathode ray tube. The distinguishing features of this stage are cathode injection, grounded control grid, plate current drops when cathode is driven positive, positive plate pulses of fair amplitude, and good frequency response.

The output from the last stage in the video amplifier is a series of positive pulses which are applied to the grid of the cathode ray tube (CRT). Each pulse causes a spot of light to appear on the CRT screen so that the range and relative bearing of the target is indicated by the position of the bright spot on the cathode ray tube. Any signals which come between the end of one sweep and the beginning of the following sweep do not appear on the screen, since normal bias keeps the screen dark. The other signals appear on the screen as follows: The unblanking pulse makes the spot bright enough to be visible, the range marks make bright spots along the timebase (the timebase is provided by another circuit) and the echo pulse brightens the spot.

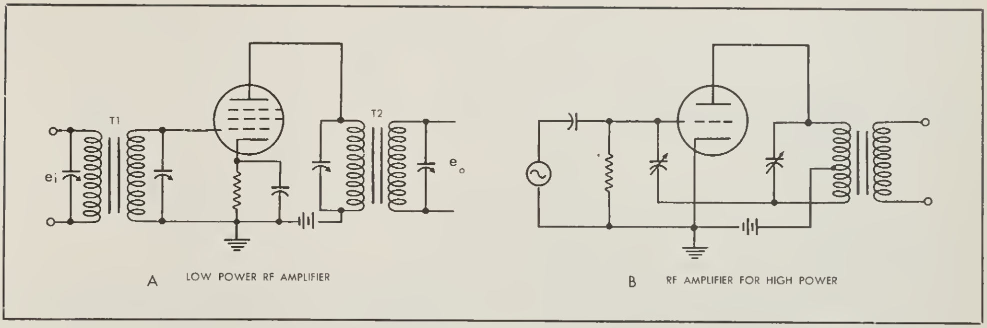
As previously mentioned, video amplifiers are designed so that their frequency response is flat. Flat frequency response is obtained in the amplifier by negative feedback. In this connection, note that all four stages employ un-bypassed cathode resistors. The feedback produced by the un-bypassed resistors in the first and third stages corrects any non-linearity in the Eg-Ip characteristic of these tubes. The second and last stages must be un-bypassed because signals are applied to their cathodes.

IF and RF Amplifiers

IF amplifiers and RF amplifiers are essentially alike. They differ only in their use and frequency. IF amplifiers are designed for amplifying a single frequency, usually higher than 100 kc, while RF amplifiers amplify a band of frequencies on each side of the frequency for which they are designed.

Because of the inefficiencies of conventional vacuum tubes at high frequencies, RF amplifiers are not used for 1000 mc frequencies. For example, the 3-cm and the 10-cm radar sets neither use RF amplifiers in front of the mixer stage in the receiver, nor for amplifying the output of the oscillator in the transmitter. In contrast, however, some low frequency radar sets do use RF amplifiers, and all sets, regardless of frequency, use IF amplifiers.

Operation of RF amplifiers. For low-power operation such as required in receivers, pentode RF amplifiers, because of their low grid-to-plate capacity, provide the highest gain with the least tendency to break into self-oscillation. The RF amplifier shown at A on the next page is actually an IF amplifier, since it is designed to amplify only one frequency. Single frequency amplification occurs because of the parallel resonant circuits at the input and output. These parallel circuits discriminate against all other frequencies by developing maximum voltage at their resonate frequency, and minimum voltage at other frequencies. This ability of an amplifier to select one frequency and reject others is called *selectivity*. Selectivity in an amplifier is greatest when high Q tuned circuits are used. (A high Q tuned circuit is produced by using powdered iron cores in the transformers.)



Typical RF Amplifiers

The RF amplifier circuit at B is one which is commonly used in transmitters. As contrasted with the circuit at A, which uses a pentode tube, this circuit uses a class-A-operated triode. Another difference is that the tuned circuit is in the plate circuit only. Note the coupling circuit from the plate tank to the grid. This is a *neutralization* circuit for feeding out of phase voltage from the output back into the grid circuit. Its function is to prevent the triode from oscillating.

IF amplifiers are used in superheterodyne receivers. In these receivers the incoming signal is converted into a new and lower frequency and amplified by specially-designed, high-gain IF amplifiers. Great care must be given in the choice of the intermediate frequency. For example, the lower the intermediate frequency, the higher the gain and selectivity. Although communications receivers may use intermediate frequencies as low as 75 kc or 275 kc, the tendency is for them to use much higher frequencies because of poor image rejection at the lower frequencies. Unlike communications receivers which have a relatively narrow band-pass, radar receivers are designed with band-passes as wide as 10 to 20 mc. Obviously, with such wide band passes, the intermediate frequency must be greater than 10 mc in order to keep the outer side band above zero mc. The usual practice is to use intermediate frequencies of 30 mc and 60 mc in radar receivers. Although high gain circuits are hard to design at these frequencies, the band pass characteristics nevertheless are good. In fact, the additional band pass devices which must be built into the set further reduce gain. In a radar receiver, therefore, it is common to find six, eight, or more stages of intermediate frequency amplification.

Previously, you saw that the video amplifier must be specially designed. Similarly, RF and IF amplifiers require special design. A pulse-modulated RF signal contains the RF frequency (or IF frequency) plus a side band frequency for each of the frequencies in the pulse. All these frequencies add up to a very wide band of frequencies. To accommodate them, the amplifier band must be fairly wide. The band width of the amplifier may be increased by lowering the Q, stagger tuning, and overcoupling.

IF amplifiers are classified as *single-tuned* or *double-tuned* depending upon whether each interstage coupling circuit includes one or two tuned circuits. Double-tuned coupling produces greater gain for a given band width, but is not used often because of the greater simplicity in alignment and in manufacture of single-tuned circuits.

The illustrations on the next page show the circuit diagrams and equivalent circuits for both types of amplifiers. The following are the gain calculations for the single-tuned intermediate frequency amplifier:

GAIN OF ONE STAGE. From equivalent circuit

$$C,\ e_o = iz = G_m e_s z$$

$$Then,\ A = \frac{e_o}{e_s} = \frac{G_m e_s z}{e_s} = G_m z$$

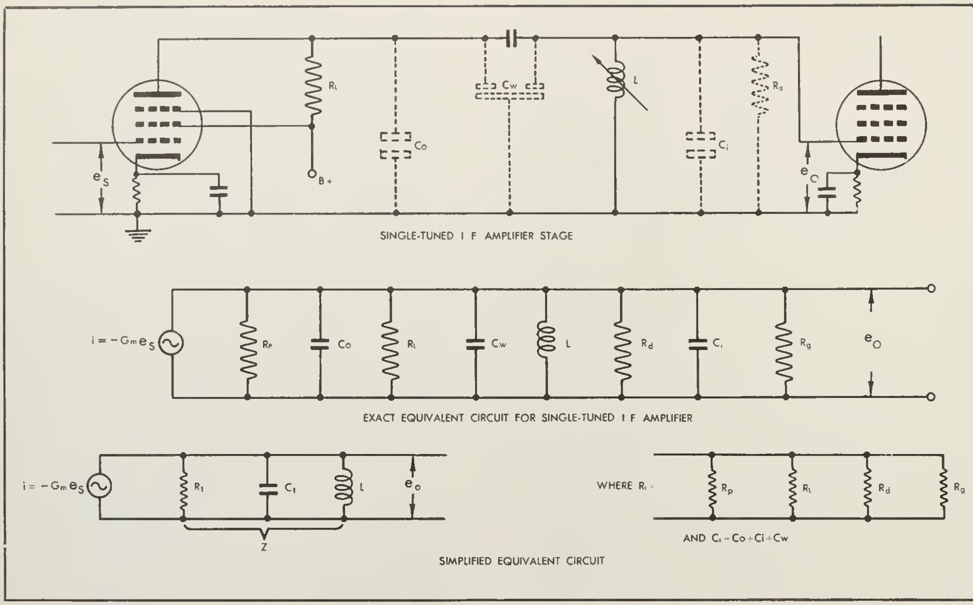
GAIN AT WHICH

$$2\pi f L = \frac{1}{2\pi f C}$$

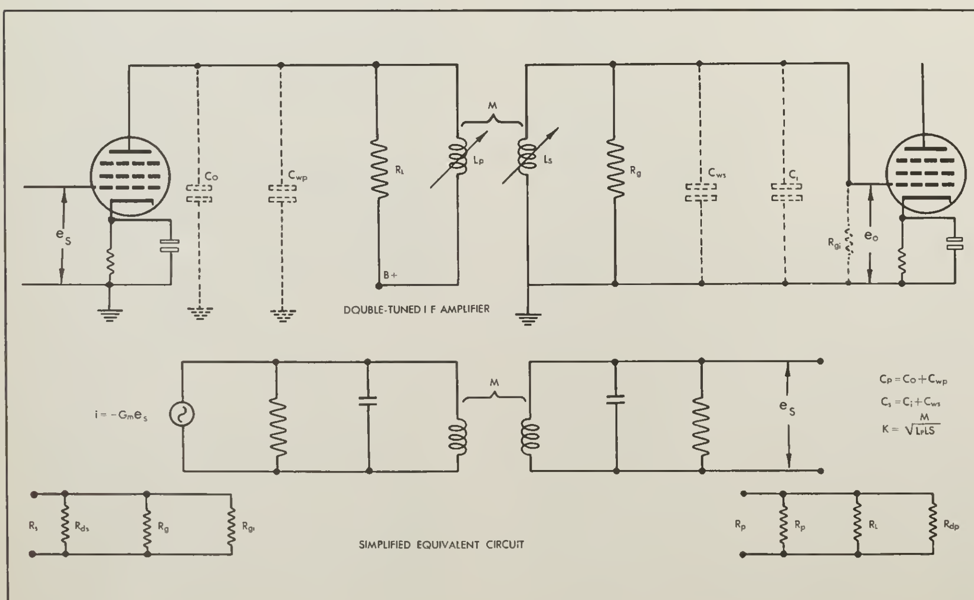
At this frequency, the parallel combination of C_t and L is an infinite impedance in parallel with R_t. Therefore, $A = -G_m R_t$

(Adjusting L in this circuit to the point where maximum gain occurs, denotes the presence of the intermediate frequency.)

Referring to the illustration of the double-tuned IF amplifier, notice that both the primary and secondary circuit are tuned to the inter-



Single Tuned IF Amplifier and Equivalent Circuits



Double Tuned IF Amplifier

mediate frequency. The formulas for calculating the intermediate frequency are as follows:

$$\text{Primary; } f = \frac{1}{2\pi\sqrt{L_pC_p}}$$

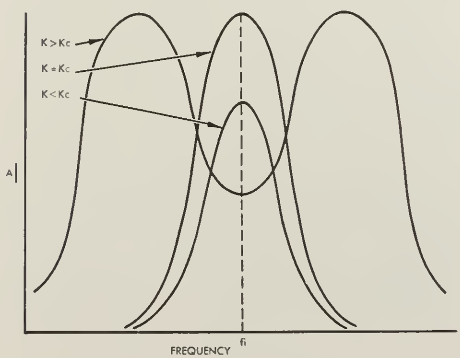
$$\text{Secondary, } f = \frac{1}{2\pi\sqrt{L_sC_s}}$$

Since the primary and secondary tuned circuits must be tuned to the same frequency,

$$IF = \frac{1}{2\pi\sqrt{L_pC_p}} = \frac{1}{2\pi\sqrt{L_sC_s}}$$

Optimum results in double-tuned IF amplifiers dictate the use of coupling that is neither too loose or too close, since it gives a single humped response curve. Such coupling is called critical coupling and is expressed mathematically,

$$K_c = \frac{1}{Q}$$



Effect of Coupling on Frequency Response

Notice the above curves showing the effect of three types of coupling on frequency response. The following are the effects of each type.

Where the coupling is too loose ($K < K_c$), there is decrease in gain.

When the coupling is too close ($K > K_c$), the response curve is double humped and causes each pulse to be followed by an oscillatory transient.

When the coupling is just right ($K = K_c$), there is maximum gain and bandwidth.

(The absolute value of gain is represented by |A|)

RF amplifiers are similar to IF amplifiers except for tubes used at UHF frequencies. An RF amplifier is used when the signal-to-noise ratio of a radar receiver is not high enough.

Several stages of IF amplifiers are used in radar receivers. Each succeeding stage narrows the bandwidth of the receiver. Radar receivers usually make use of a bandwidth from 3 to 6 megacycles. An impairment in fidelity results from use of several stages in cascade, unless the bandwidth is increased. Increasing bandwidth may be accomplished by any of the following methods:

Close coupling the transformers in double-tuned stages.

Loading the primary section of the transformer by making the value of R_t smaller (R_t can be decreased only to a certain limit).

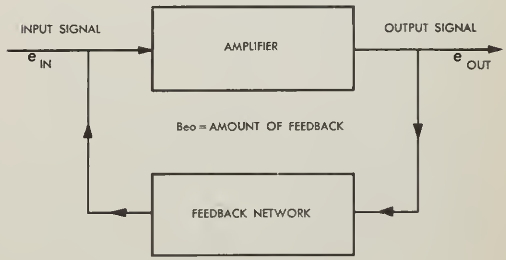
Tuning some IF stages slightly above the intermediate frequency, and some slightly below.

Unbypassing the cathode resistor.

Each of these reduces the gain and the input capacitance of each stage and results in an increase of the over-all bandwidth. Remember, though, that none of these methods can be used without *reducing* the gain of the intermediate frequency amplifier.

FEEDBACK IN AMPLIFIERS

Whenever a portion of the amplified output energy of an amplifier is fed back into the input circuit in such a manner as to reinforce the input voltage, the gain is greatly increased. Such feedback is called *positive* or *regenerative* feedback. If the portion of the amplified output energy is fed back so as to oppose the input voltage, the feedback is called *degenerative, negative,* or *inverse* feedback.



Feedback Amplifier

The above illustration shows a simple block diagram of a feedback amplifier with the voltages as indicated. When there is no phase shift in the input voltage e_i, and when amplification factor is A, the net output voltage equals $e_o - Be_o$ (Be_o representing the fraction of the output fed back to the input). With Be_o fed back to the input, the total input is $e_i + Be_o$. If this input is amplified A times, the equation for the output voltage is,

$$e_o = (e_i + Be_o)A$$

Since the gain is $\frac{e_o}{e_i}$, you can derive the gain formula for a feedback amplifier by solving the equation, $e_o = (e_i + Be_o)A$ as follows:

Multiplying in the right member,

$e_o = Ae_i + BAe_o$

Subtracting BAe_o from both members,

$e_o - BAe_o = Ae_i$

Factoring the left member

$e_o\ (1\text{-}BA) = Ae_i$

Thus,

Gain $= e_o/e_i = A/1\text{-}BA$

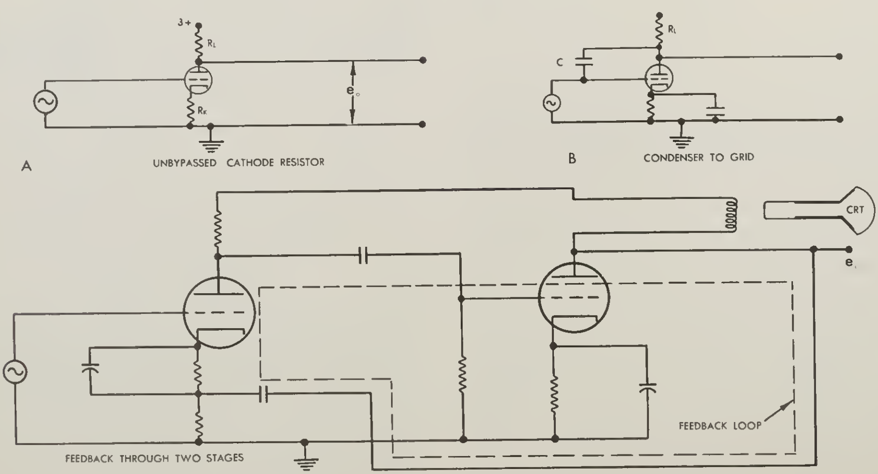
The expression 1-BA is important as it will tell you the nature of the feedback. When the quantity 1-BA is less than 1, the gain of the amplifier is increased, and the feedback is *regenerative*, or *positive*. When the quantity 1-BA is zero, that is, BA = 1, regeneration is large enough to make the stage self-sufficient and the amplifier will act as an oscillator. When the quantity 1-BA is greater than 1, the gain of the amplifier is decreased and the feedback is degenerative, negative, or inverse.

Degenerative feedback decreases gain, but it reduces distortion. Degenerative feedback may be obtained by out of phase feeding a fraction of the plate energy output to the grid circuit input or by using an un-bypassed cathode resistor for self bias.

Methods of Obtaining Feedback

Feedback can be obtained in amplifiers in several ways. One arrangement used to obtain negative feedback is illustrated below in the circuit at A. In this circuit, feedback voltage is developed across the un-bypassed cathode resistor R_k, as a result of the plate current flowing through it. Current flow through an un-bypassed cathode resistor develops a voltage which varies at the same rate as the plate voltage. Since resistor R_k is located between the cathode and the grid of the tube, any voltage developed across it is in series with the input signal of the tube. The phase relation is correct for degeneration to occur. Thus, when a positive signal appears on the grid of the tube, the plate current increases, and the voltage drop across R_k also increases. A voltage increase across the cathode resistor makes the grid more negative relative to the cathode, the reverse of the action by the signal voltage. This arrangement is very useful in cancelling out distortion in the output signal resulting from the E_g-I_p characteristic curve. Since plate voltage changes in a tube do not follow the grid voltage changes exactly there is some distortion in the output waveshape. However, degenerative feedback introduces a portion of the distorted output back into the input in reverse, counter-balancing the conditions causing the original distortion.

Similarly, feedback can be obtained by connecting a condenser from the plate back to the



Methods of Providing Feedback

grid of the tube as shown at B. Here plate voltage changes produced by plate current variations are opposite in polarity to the original grid voltage changes. The condenser introduces a small part of the plate voltage change back into the grid circuit. Thus, the plate change reflects the distortion in the plate current change in reverse, virtually cancelling the distortion originally introduced into the output.

Feedback can be used in amplifiers employing more than one stage. The illustration at C shows a two-stage feedback amplifier. Not only is it useful for counteracting the effects of distortion introduced by the tubes, but it can be used also to cancel out-of-phase relationships caused by certain circuit components. For example, it compensates for any changes in the inductive reactance of the deflection coil in the circuit. When current flows through this coil, its reactance changes, and this causes a change in the plate voltage of the tube. This change in voltage moves around the feedback loop and counteracts the inductance change. This makes the plate voltage normal and insures an undistorted output voltage.

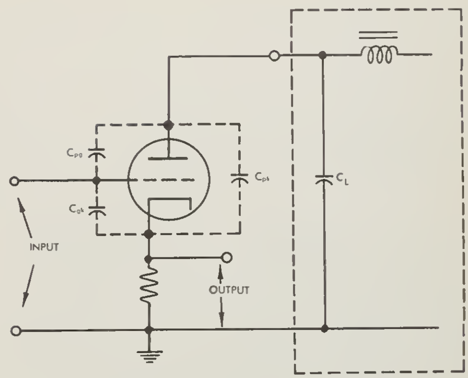
In some circuits, reactance elements introduce positive feedback into the input, and the circuit oscillates. This is desirable in oscillators but not in amplifiers. Therefore, some type of feedback arrangement, similar to that in the two-stage feedback amplifier, must be used to introduce a negative feedback into the input to cancel the effects of the regenerative feedback.

CATHODE FOLLOWER

A cathode follower amplifier is a single-stage degenerative amplifier in which the output is taken from across the cathode resistor. This circuit is essentially an impedance matching device for matching a high-impedance circuit to a low-impedance circuit without discriminating against any AC frequencies. Its voltage output is always less than the input voltage, but it is capable of power amplification. Some of the advantages of cathode followers are low input capacity and distortionless output.

Input Capacity

As previously mentioned, serious losses result from the high input capacity in triode tubes. This is due to the fact that the resistive load causes a leading current in the grid circuit and the plate-to-grid capacity adds to the normal grid-to-cathode capacity to produce a large



Interelectrode Capacities in Cathode Follower Circuit

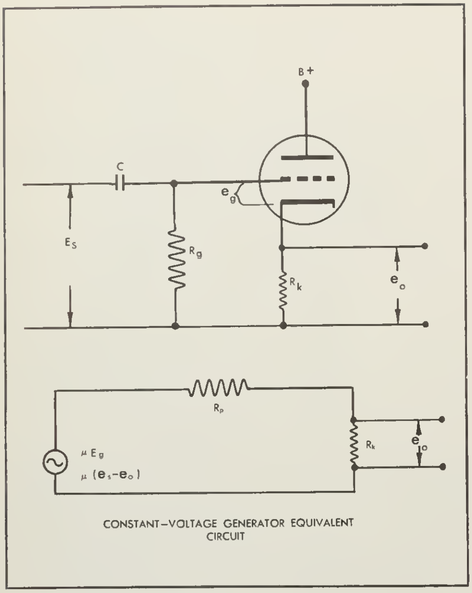
capacity at the input of the triode. Unlike that in the triode, the plate load in the cathode follower is a large capacity (usually the filter condenser in the power supply). Nevertheless, as in the triode, the capacity C_{pg} still must charge when a voltage is applied to it. However, the current in the grid circuit does not lead because the capacitive load in the cathode follower causes the charging current to shift in phase and puts it exactly in phase with the original grid voltage. Since adding in-phase current is the same as lowering the resistance in the grid circuit, the impedance reflected in the grid circuit is purely resistive, and its effect on all frequencies is the same, that is, there is no frequency attenuation. The only reactance left is the actual grid-to-cathode capacity which is too small to be concerned about.

Distortionless Output

Another advantage of the cathode follower is that it introduces very little amplitude distortion into the output, since it is a degenerative circuit in which negative feedback is always produced by an un-bypassed cathode resistor, and since its output is taken from across the cathode resistor and not the plate.

Design Considerations

Three of the most important considerations in the design of cathode followers are gain, input and output impedance, and size of cathode resistor. Since the gain and the size of the cathode resistor are the only variables in a cathode follower, and impedance is in turn related to both, you should be acquainted with the formulas for computing them.



Cathode Follower

Gain formula. Referring to the equivalent circuit in the above diagram, notice the use of the following designations:

e_s = input voltage
e_g = voltage between grid and cathode
e_o = output voltage

From the cathode follower circuit,

$$e_g + e_o = e_s$$

Transposing and substituting,

$$e_g = e_s - e_o = e_s - i_p R_k$$

Since $e_o = i_p R_k$

and $i_p = \frac{\mu e_g}{R_p + R_k} = \frac{\mu e_s - \mu i_p R_k}{R_p + R_k}$

Therefore, $i_p R_p + i_p R_k = \mu e_s - \mu i_p R_k$

and $i_p[R_p + (R_k)\ (1 + \mu)] = \mu e_s$

Therefore, $i_p = \frac{\mu e_s}{R_p + R_k(\mu+1)}$

Then, $e_o = \frac{\mu e_s R_k}{R_p + R_k(\mu+1)}$

Thus, Gain $= e_o/e_s = \frac{\mu R_k}{R_p + R_k(\mu+1)}$

Notice in the equation just derived that the denominator is greater than the numerator. This indicates that the gain of a cathode follower is less than one. Remember, though, that the less than one gain applies only to the AC voltages, and not to DC values. Actually, the cathode to ground DC voltage is greater than the input voltage as long as the grid-to-cathode voltage is negative.

Impedance. The input impedance in a cathode follower is high. Since cathode followers are operated with the grid negative with respect to the cathode, a high amplitude voltage can be applied between the grid and ground without causing grid current flow. This is due to the degenerative action of the cathode resistor and the high input impedance during the positive input signal. Because of its high input impedance, the cathode follower has negligible loading effect on the circuit which drives it.

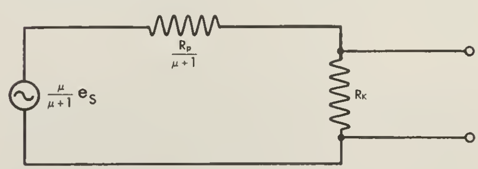
The output impedance is low. The expression for the plate current that flows in the constant-voltage generator equivalent circuit of the cathode follower at the left is,

$$i_p = \frac{\mu e_s}{R_p + R_k\ (\mu+1)}.$$

If both the numerator and the denominator of this expression are divided by $\mu + 1$, the expession will have the form of the plate current of a circuit in which the tube has an amplification factor of $\frac{\mu}{\mu+1}$ and an AC plate resistance of $\frac{r_p}{\mu+1}$. This expression reads,

$$i_p = \frac{\left(\frac{\mu}{\mu+1}\right) e_s}{\left(\frac{R_p}{\mu+1}\right) + R_k}$$

The equivalent circuit is redrawn below to show the new circuit constants.



The output impedance of the cathode follower is the parallel combination of the cathode resistance and the effective AC plate resistance. In general, the output impedance is resistive, and is expressed as,

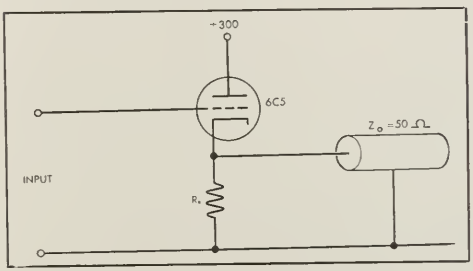
$$Z_o = \frac{\frac{R_p R_k}{\mu+1}}{\frac{R_p}{\mu+1} + R_k} = \frac{\frac{R_p R_k}{\mu+1}}{\frac{R_k\ (\mu+1)\ + R_p}{\mu+1}}$$

$$Z_o = \frac{R_p R_k}{R_p + R_k\ (\mu+1)}$$

Cathode resistor. The size of the cathode resistor is an important design consideration in that it largely determines the output impedance, a factor of special concern in impedance matching. To find the value of the cathode resistor, use the following equation:

$$R_k = \frac{Z_o R_p}{R_p - Z_o(\mu + 1)}$$

where Z_o equals the output impedance, R_p the plate resistance and μ equals the amplification of the tube.



Example

If the 6C5 triode shown above is connected as a cathode follower and has a plate voltage of 300 volts, a plate resistance of 10,000 ohms, and an amplification factor of 20, find the size of the cathode resistor and the gain.

Solution:

To find the size of the cathode resistor, substitute in the formula,

$$R_k = \frac{Z_o R_p}{R_p - Z_o(\mu + 1)}$$
$$= \frac{50 \times 10^4}{10^4 - 50\ (20+1)}$$
$$= \frac{500{,}000}{10{,}000 - 1050}$$
$$= \frac{500{,}000}{8950}$$
$$R_k = 55.8 \text{ ohms}$$

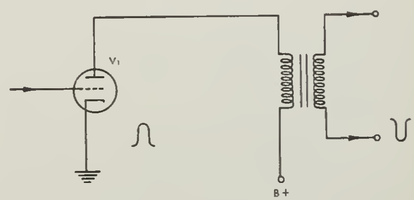
To find the gain, substitute in the formula,

$$Gain = \frac{\mu R_k}{R_p + R_k\ (\mu + 1)}$$
$$= \frac{20 \times 55.8}{10^4 \times 55.8\ (20+1)}$$
$$= \frac{1116}{11171.8}$$
$$Gain = .1 \text{ (approximately)}$$

PHASE INVERTERS

A phase inverter is a circuit which produces an output voltage of opposite polarity to the input voltage without distorting the waveshape. Phase inverters are commonly used in radar equipment. For example, sweep voltages are usually applied to the deflection plates in electrostatic cathode-ray tubes in push-pull, since this type of operation reduces the defocusing effects resulting from applying the sweep voltage to only one of the pair of deflection plates. To obtain the push-pull operation, a phase inverter circuit is used.

Literally, the commonly accepted term, *Phase Inverter* is something of a misnomer, since phase is ordinarily associated with the time and there is no appreciable time difference or phase shift between the output and input circuits of the ordinary phase inverter. Such a circuit is only an *apparent* phase inverter. In reality its a polarity inverter.



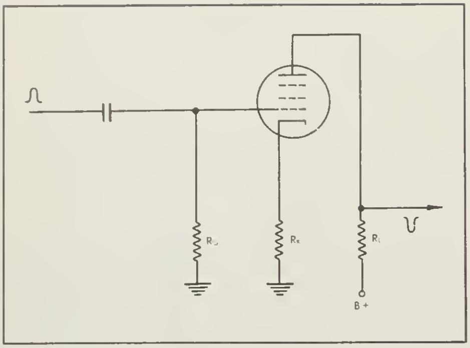
Polarity Inversion

Polarity Inversion by Transformer

A simple method of inverting the polarity of a waveshape is by a transformer. To understand how a transformer inverts polarity, assume that the output wave pulse at the plate of tube V_1 in the above illustration is positive and that tube V_2 requires a negative pulse. This requirement can be met by using the polarity-inverting property of an ordinary transformer, since, in all transformers, a current through the primary induces a voltage in the secondary of opposite polarity to the primary voltage. (Of course if the output or input connections to the transformer are reversed, the output and input voltage will have the same polarity). In transformer inversion, it is clearer to think of the output as a voltage whose polarity is inverted relative to the primary voltage, except perhaps when a sine wave signal is used, where the polarity inversion is referred to as 180° phase shift.

Polarity Inversion Without Amplification

Under some conditions, it is necessary to reverse the polarity of a pulse or a waveform with-

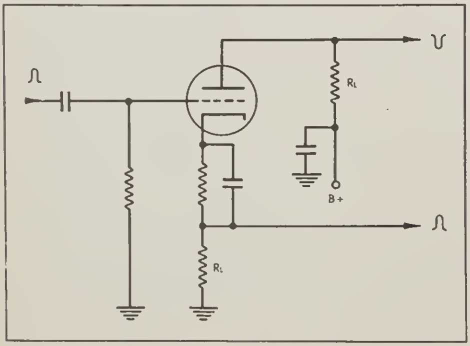


Polarity Inversion Without Amplification

out changing its amplitude. Although a transformer may be used, it is much better to use an ordinary RC coupled amplifier with an un-bypassed cathode resistor as shown above. This circuit inverts phase since any vacuum tube amplifier with a resistive load has an output of opposite polarity to the input. In other words, a positive going signal on the grid produces a negative going signal at the plate. There is no amplification because of the degenerative feedback introduced into the grid by the un-bypassed cathode resistor. This degeneration occurs because the cathode voltage rises as the grid voltage rises, preventing the swing of the voltage between the grid and cathode from reaching the amplitude of the applied grid signal.

Paraphase Amplifier

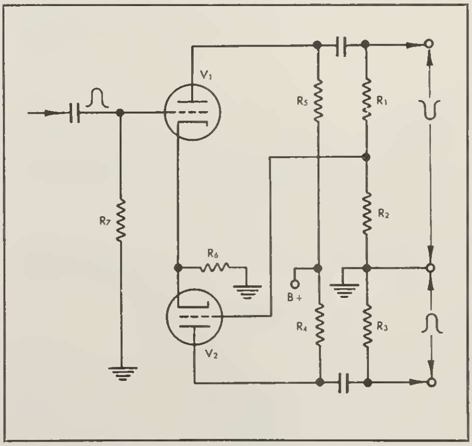
A *paraphase* amplifier is a combination amplifier and phase inverter, which converts a single input into a push-pull output. Paraphase amplifiers are used where waveshapes of equal amplitude and opposite polarity are required for operating circuits in push-pull, as in the electrostatic deflection circuits in cathode-ray tube equipment. There are two types of paraphase amplifiers—the single-tube paraphase amplifier and the two-tube paraphase amplifier.



Single-tube Paraphase Amplifier

Single-Tube Paraphase Amplifier

With a single-tube paraphase amplifier the output is taken from both the cathode and the plate. The cathode resistor R_L and the plate resistor R_L are the load resistors. These resistors are equal and since the same current flows through both, equal voltages appear across them. The voltages across these resistors are opposite in polarity since the output is taken from the positive end of the cathode load resistor and from the negative end of the plate load resistor.

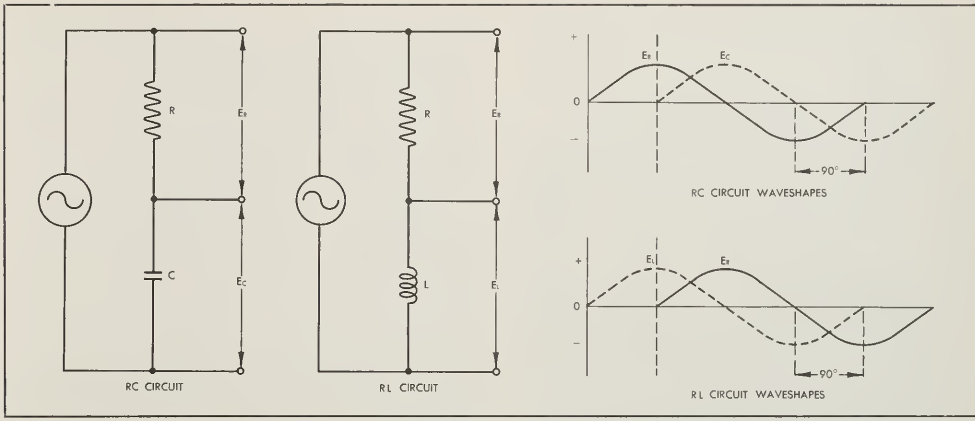


Two-tube Paraphase Amplifier

Two-Tube Paraphase Amplifier

The two-tube paraphase amplifier consists of one tube which acts as a conventional amplifier, and a second which inverts the output of the first tube. The two tubes in combination thus produce two equal output voltages opposite in polarity.

The above illustration shows the circuit diagram of a typical two-tube paraphase amplifier. The first tube V_1 amplifies the input waveform shown at its grid and impresses the amplified output across the voltage divider consisting of R_1 and R_2. The resistor R_2 is of such value that the



RC and RL Phase Splitting

varying voltage across it has the same amplitude as the voltage on the grid of V_1. The voltage across R_2 is impressed on the grid of V_2, the phase inverter tube, where it is amplified. Since the plate load resistors, R_5 for V_1 and R_4 for V_2 are equal, the outputs of the two tubes are equal. The phase inverter inverts the phase of the voltage applied to its grid, making it opposite in phase to the voltage output of V_1. Note in this connection that the waveshape in the output of V_2 is in phase with the grid voltage to V_1. Phase inversion has occurred in V_1 and again in V_2, thus shifting the phase of this voltage back to its original polarity.

PHASE-SPLITTING CIRCUITS

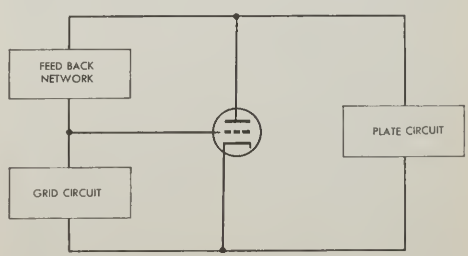
A phase-splitting circuit is one which produces, from the same input, two output waveforms which differ in phase from one another. An example of the use for a phase splitting circuits is in the generation of the circular sweep in the type J scan indicator. The sweep voltage for this scope is generated by two out-of-phase voltages applied to the deflector plates. The two out-of-phase voltages are produced by a phase-splitting circuit.

Two types of circuits used for producing two out-of-phase output waveforms from the same input are the RC phase-splitting circuit and the RL phase-splitting circuit.

Referring to the RC circuit above, notice that R and C which are connected in series represent the load impedance. The same current flows through both R and C. The voltage across R is always in phase with the current in it, since, in a resistive circuit, current and voltage are in phase. In a capacitive circuit, current and voltage are 90° out of phase with current leading. Therefore, the voltage across C is 90° out of phase with the current. Thus, the two voltages—that across R and that across C—are 90° out of phase with each other. An identical phase shift occurs in an LR circuit. The only difference is that the current lags the voltage in L by 90°. The voltages, however, like those across R and C in the RC circuit, are 90° apart.

OSCILLATORS

An oscillator is a device capable of converting direct current into an alternating current at a frequency determined by the values of the constants in the device. Generally, this device is a vacuum tube since a vacuum tube has the ability to amplify which is one requirement of an oscillator. When a portion of the amplified out-



Block Diagram of an Oscillator

put energy of a vacuum tube is fed back into the grid circuit in the correct phase, oscillations occur. Because of this property, an oscillator may be regarded as a self-excited amplifier.

Oscillators in radar sets perform a large number of functions. They are used as UHF generators in transmitters, as local frequency oscillators in receivers, and as master oscillators in timing circuits, pulse-forming circuits, gating circuits, and sweep producing circuits.

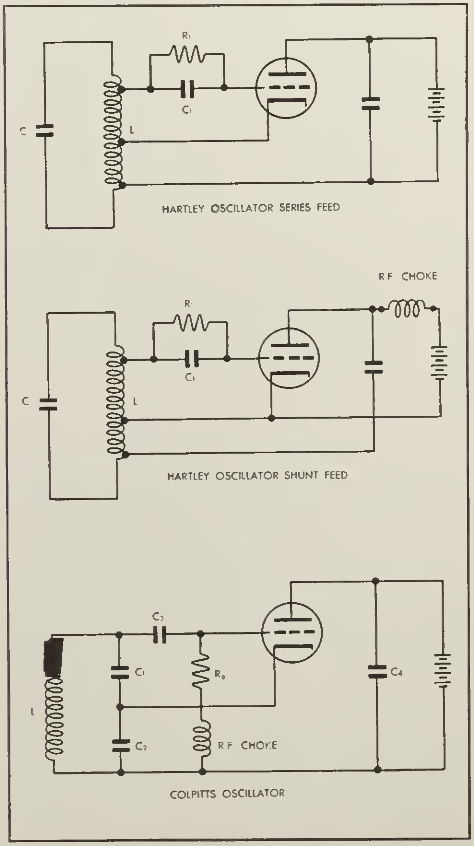
As previously mentioned a vacuum tube is able to oscillate because of its amplifying ability. A vacuum tube oscillates because any small voltage change in the plate or grid circuit can be transferred from one circuit to the other by the process of amplification. Whenever a portion of the amplified output of an oscillator is fed back into the grid circuit in the proper phase, oscillations will start. They will continue as long as in-phase voltage is applied to the grid circuit by the output circuit. Feedback of this kind is called regenerative or positive, since it is in phase with the grid voltage and sustains oscillations.

In a vacuum tube, there is a normal phase shift between the grid and plate. Hence, without an additional shift in phase, the voltage returned to the grid would not have the correct polarity or phase to sustain oscillations in an oscillator. Therefore, a feedback network must be used to shift the feedback voltage another 180° so that it is in phase with the initial grid voltage. The following devices are employed as feedback networks for this purpose:

1. Transformers.
2. RC networks or LC networks.
3. Interelectrode capacitances.
4. Additional vacuum tubes.

A very important characteristic of regenerative feedback is that in supplying the input power to an oscillator, it effectively introduces a negative resistance into the input circuit of the oscillator. Although negative resistance cannot be represented by a physical resistor, it provides a concept which is convenient in discussing oscillators since it represents the source from which energy is obtained to replace the energy losses in the oscillating circuit. All this means that, for an oscillator to keep on oscillating, energy has to be supplied from some source to make up for losses in the feedback circuit. Otherwise, it will stop oscillating after a period because of the accumulated losses. Negative resistance at the input explains how these circuit losses are replaced. A circuit with negative resistance is one which transfers energy from one place to another with an increase in effective energy rather than a decrease. In other words, a negative resistance circuit is a generator of energy and not a user of energy.

There are a number of oscillator types which you should know something about. Two basic types worth studying are the Hartley oscillator and the Colpitts oscillator. Other widely employed oscillators in radar are the tuned-plate-tuned-grid oscillator, the crystal oscillator, the multivibrator (which was discussed in the preceding chapter), the resistance-capacitance coupled oscillator, and the shock-excited oscillator.



Basic Oscillator Circuits

Hartley Oscillator

There are two forms of Hartley oscillators—the series-feed and the shunt-feed. In the series-feed Hartley, direct current flows through part of the tank circuit. In the circuit diagram, note that the tube, a portion of the tank circuit, and the power-supply form a series circuit. In the shunt-feed Hartley, direct current does not flow through any part of the tank circuit at all. The plate supply voltage is in parallel with the tube and tank circuit. RF is kept out of the plate supply by an RF choke, and DC out of the tank circuit by a blocking condenser.

In Hartley oscillators, the frequency of oscillations is determined by the LC constant of the resonant tank circuit. Bias for the tube is provided by the grid leak combination R_1 and C_1. Grid-leak bias or a combination of grid-leak bias and fixed bias is used, since oscillators operate Class C to secure high output and efficiency. If only a high fixed bias were used, plate current would not flow when plate voltage is applied, and thus it would be impossible for oscillations to occur. However, when grid-leak bias is used, the bias is initially zero and plate current flows immediately. After oscillations build up, the grid leak bias becomes proportional to the RF voltage across the grid tank circuit and maintains a value providing a stable oscillatory condition.

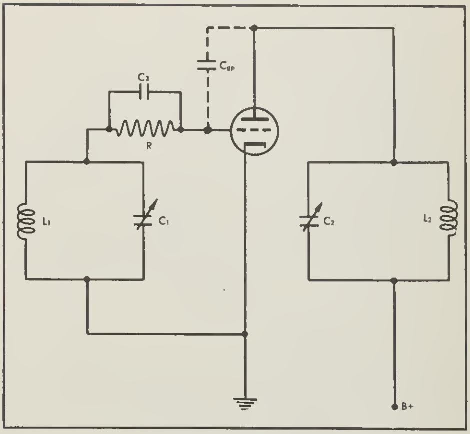
Voltage for sustaining oscillations in the Hartley oscillator is fed from the tank circuit to the grid circuit by mutual inductance between the plate and grid coil. Adjusting the position of the taps connecting the grid, cathode, and plate leads determine the magnitude of the feedback.

Colpitts Oscillator

The colpitts oscillator is essentially the same circuit as the Hartley oscillator except that a pair of capacitances in series is connected across the tank coil. This combination, C_1 and C_2 in the circuit, forms a voltage divider circuit which divides the voltage across the resonant circuit into two parts. The voltage at these ends of the resonant circuit are opposite in polarity with respect to the cathode and in the right phase to sustain oscillation. The total tank capacitance consists of C_1 and C_2. C_3 and R_g comprise the grid leak bias combination.

Tuned-Plate Tuned-Grid Oscillator

The tuned-plate tuned-grid oscillator, sometimes abbreviated as TPTG, employs a tuned circuit in both the grid and plate circuits. This oscillator may be employed at ultra-high frequencies as well as at low frequencies. However, the use of it at low frequencies is not as satisfactory as the oscillators previously described.

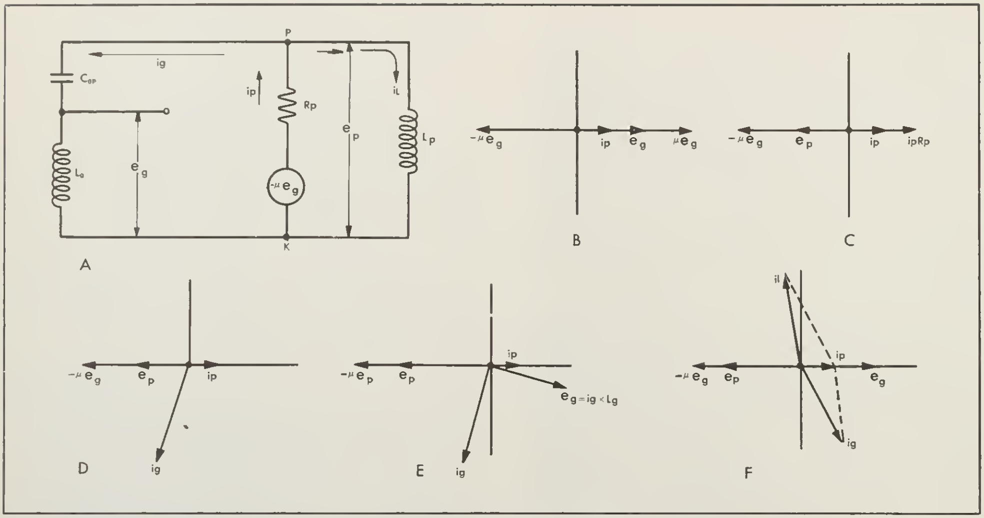


Tuned Plate Tuned Grid Oscillator

In the illustration of the TPTG oscillator, notice that the inductance in the plate tank circuit is not inductively coupled to the inductance in the grid circuit. The feedback necessary for sustaining oscillations occurs through the interelectrode capacitance between the plate and grid of the tube (C_{gp}).

This circuit automatically oscillates at a frequency lower than the natural frequency of both the plate tank (L_2 and C_2) and the grid tank (L_1 and C_2). The tank that is tuned to the lower frequency, whether it is the plate or grid tank, controls the oscillator frequency. The other circuits mainly affect the magnitude of the feedback voltage.

On the next page the illustration containing the equivalent circuit of the TPTG oscillator and vector diagrams explains the operation of this oscillator. Assuming that the oscillator is in operation, a voltage e_g exists at the grid which controls the plate current of the tube so that energy is delivered to the plate-tank circuit at the proper instant to reinforce the existing oscillations. The effect of this grid voltage is represented in the equivalent circuit by a generator which generates a voltage of $-ue_g$. In the vector diagram B the voltage e_g is shown 180° out of phase with the equivalent generator voltage $-ue_g$. The grid voltage directly controls the plate current; there-



Equivalent Circuit of TPTG Oscillator

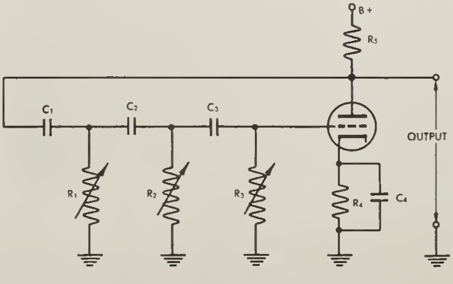
fore, when e_g is a maximum, i_p must be in phase with e_g as shown in the diagram.

The voltage across the tube, P to K, in the circuit diagram A, is the vector difference between $-\mu e_g$ and i_pR_p, the voltage drop across the plate resistance. Thus, the plate voltage, e_p, is shown in the vector diagram C as 180° out of phase with i_pR_p and as the vector difference between $-\mu e_g$ and i_pR_p. This voltage, e_p, is applied across series combination of C_{gp} and L_g. For the oscillator to operate, the reactance of the plate-grid interelectrode capacitance must be greater than the inductance of the grid tank circuit at the operating frequency. Therefore, the current that flows in the grid circuit as a result of the application of e_p across it is a current i_g which leads the voltage by nearly 90° as shown in the vector diagram D. In flowing through the inductance, L_g, this current produces the voltage e_g of the grid tank circuit. This voltage across the inductor must lead the current through the inductor by somewhat less than 90°, since there is some resistance associated with the circuit. On examining vector diagram E, notice that voltage e_g which is fed back is almost in phase with the plate current, so that oscillations can be maintained.

In order for the voltage e_g to be exactly in phase with the plate current, there must be negative resistance present in the TPTG oscillator. When negative resistance is considered, it can be shown that current i_g leads the plate voltage by more than 90°, as shown in vector diagram f, so that the voltage which is fed back is exactly in phase with the plate current.

Resistance-Capacitance Coupled Oscillators

Many oscillators use resistance-capacitance networks for providing regenerative coupling between their output and input circuits and for determining the oscillation frequency. Such oscillators are called resistance-capacitance or simply RC oscillators. Examples of RC oscillators are the phase shift oscillator, the Wien-bridge oscillator, and the free-running multivibrator which was discussed in an earlier chapter.



Phase Shift Oscillator

PHASE SHIFT OSCILLATOR. The phase-shift oscillator circuit consists of a single amplifier tube and a phase-shifting feedback circuit. As previously mentioned, the conventional feedback oscillator requires that the signal fed back from the plate to the grid be shifted 180° in order to sustain oscillations. In phase-shift oscillators, this function is performed by three resistance capacitance sections.

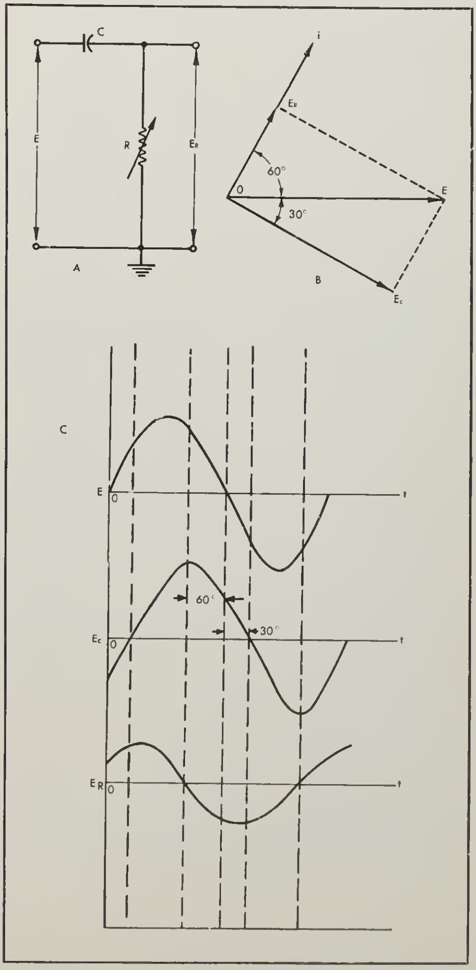
One of these sections is shown at the right at A. When an alternating voltage is applied to this circuit, current having a magnitude determined by the total impedance flows in the circuit. Because of the capacitor C, the impedance is capacitive and the current i leads the impressed voltage E by 60° as shown at B and C in the illustration. The voltage drop E_R which occurs across resistor R is in phase with the current which flows through it. Therefore, the voltage E_R must also lead the impressed voltage, E, by 60°. When the output of this section is applied to a second and similar phase shifter, another shift of 60° of the output voltage occurs, making the output voltage of the second phase shifter 120° ahead of the applied voltage to the first phase shifter.

When the resistance of R_1 (and likewise R_2 and R_3 in phase shifters 2 and 3) is varied, the phase angle of the current flowing in the circuit is also varied. If R were reduced to zero, the current would theoretically lead the applied voltage by 90°, but there are two reasons why adjusting R to zero would be impractical. First, because reducing R to zero would leave no impedance for developing a useful voltage and second because with R equal to zero, the current would lead the voltage slightly less than 90° anyway, because of the inherent circuit resistance. Therefore, resistor R is varied only to the point where the current leads the voltage by 60°.

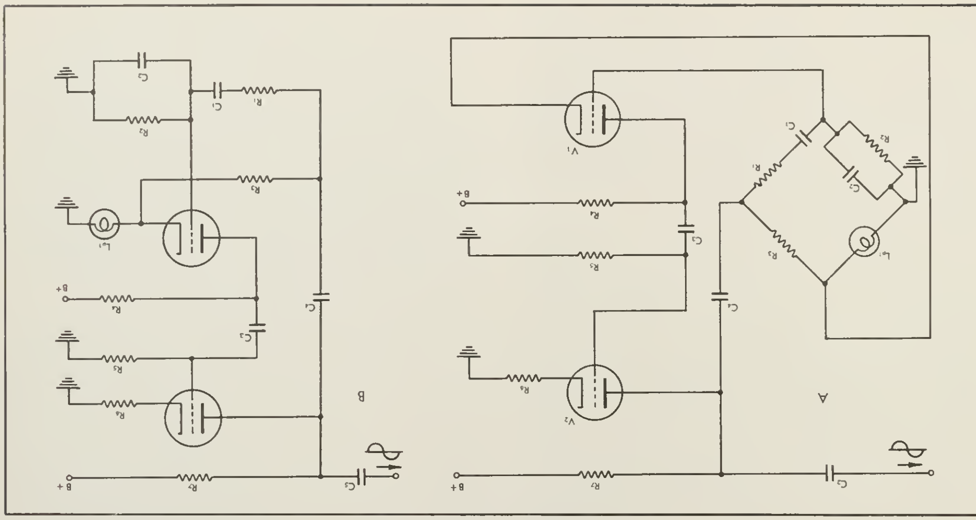
When the output of the second phase shifter is fed to the third, another shift of 60° occurs, making the voltage applied to the grid 180° ahead of the voltage at the input to the first phase shifter. Thus, the voltage at the grid is in the correct phase to sustain oscillations.

Since the three resistors R_1, R_2, and R_3 have to be set to obtain the final phase shift of 180°, the phase-shift oscillator is useful mostly in cases where a fixed frequency is desired. To increase the oscillation frequency, it is necessary to decrease either the resistance or capacitance, and to decrease the frequency, it is necessary to increase the resistance or capacitance.

The oscillations are started in this oscillator by any slight circuit changes such as variations in the plate supply or random noises. When such disturbances occur, the slight change is amplified, inverted 180° at the plate, and inverted another 180° by the RC network from which it is returned in phase to the grid of the tube for reamplification. This process continues until plate-current saturation is reached where the tube cannot amplify further.



RC Phase Shifter



Wien Bridge Oscillator

Wien bridge oscillator. The Wien Bridge oscillator shown at A above is a two-tube RC oscillator which is tuned by a resistance capacitance bridge. One tube V_1 serves as an oscillator and amplifier, and the other V_2 as an inverter. This circuit would oscillate even without the resistance-capacitance bridge because of the 180° phase shift produced by tubes V_1 and V_2. However, this system would oscillate over a very wide range of frequency. The bridge circuit therefore is used to insure the elimination of all feedback voltages except that having the desired frequency.

Referring to the circuit at B which is another way of drawing the same circuit, a degenerative feedback voltage is provided by the resistor R_3 and the lamp L_{p1}. The amplitude of this feedback voltage remains nearly constant for all frequencies since the resistances are practically constant for all frequencies and since there is no phase shift across the voltage divider.

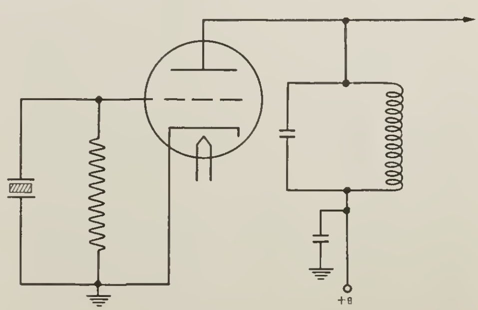
The inverter shifts the output of V_1 180°. Thus, the voltage appearing across R_2 of the bridge is in proper phase to sustain oscillations. This action occurs when $R_1 C_1 = R_2 C_2$. The frequency at which the circuit oscillates is,

$$\frac{1}{2\pi R_1 C_1}$$

The lamp L_{p1} is a thermistor, a device in which resistance varies directly with current. Its function is to compensate automatically for changes in the circuit in order that sufficient feedback will always be applied to V_2. When the current in the circuit increases, the filament of the lamp becomes hotter making its resistance larger. A greater negative feedback voltage is developed across the increased resistance. Thus more degeneration is provided, which reduces the gain of V_1, and thereby holds the output voltage at a nearly constant amplitude.

Crystal Oscillator

A quartz crystal can be used to control the frequency of an oscillator if it is so placed in the circuit that it takes the place of the normal frequency controlling circuit. The arrangement shown in the illustration below is similar to the tuned-plate-tuned-grid oscillator except that a



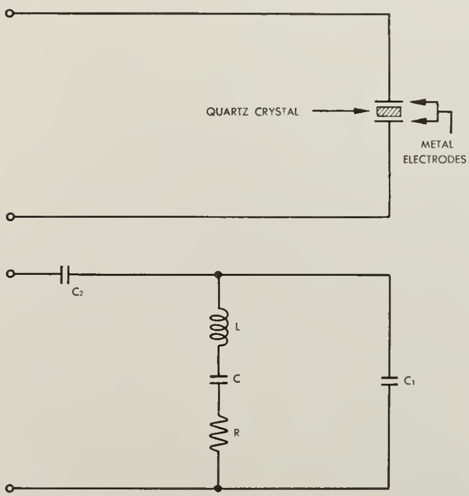
Crystal Oscillator

quartz crystal is used in place of the tuned grid tank circuit.

When an alternating voltage is applied across a crystal in a crystal oscillator circuit, the crystal vibrates. If the frequency of the voltage is approximately equal to the natural vibrating frequency of the crystal, the crystal produces a large amplitude voltage of this frequency.

The current which a crystal develops at its natural or resonant frequency is the same as that produced by a series resonant circuit consisting of resistance, capacitance, and inductance.

Because of the property of displaying series resonance when connected in a circuit, the crystal can be represented by the equivalent circuit shown just below. C_1 represents the capacity



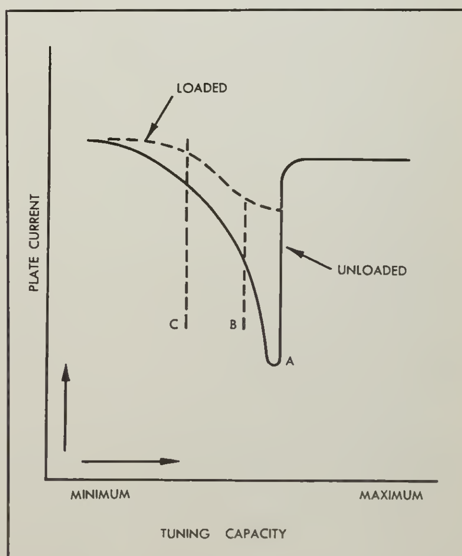
Equivalent Circuit of Crystal

between the crystal electrodes. The crystal acts as the dielectric between the two electrodes. When the crystal is placed in a vacuum tube circuit the capacity of C_1 changes to a value depending upon the crystal holder, the input capacity of the oscillator tube, and the capacity of the connecting wires between the crystal and the tube. The inductance L represents the crystal mass, C the resilience (the ability of the crystal to spring back to its original shape after being distorted by an applied voltage), and R represents frictional losses. The component C_2 represents the series capacity between the crystal and its electrodes.

The frequency at which the reactances of L and C produces series resonance is the natural vibrating frequency of the crystal. When a voltage is applied across the crystal at a slightly higher frequency than its natural frequency, the effective reactance of L and C combined becomes inductive and numerically equal the reactance of C_1. At this frequency, anti-resonance occurs and the crystal becomes the equivalent of an anti-resonant or a parallel resonant circuit. Quartz crystals, employed for controlling the frequency of vacuum tube oscillators, are usually calibrated at their parallel resonant frequencies, since the voltages applied to them are usually higher in frequency than the natural frequency of the crystal. Thus, a crystal acts as series resonant circuit when a voltage having a frequency equal to its natural vibrating frequency is applied to it, and as a parallel resonant circuit when the voltage applied has a frequency slightly higher than the crystal frequency.

C_2 affects the circuit only when the crystal electrodes do not make close contacts with the faces of the crystal. A decrease in its value produces an increase in the resonant frequency.

According to the frequency response curve below, there is a pronounced drop in plate current when a crystal oscillator goes into oscillation. Al-



Variation of Plate Current During Crystal Oscillator Tuning though maximum output occurs then, a crystal oscillator should be operated between the points B and C on the response curve. The reason for this is that the point of maximum output, A on the curve, produces unstable operation. This means that if the crystal is operated at this point, the oscillator becomes very erratic, that is, it may operate intermittently, or stop after a brief oscillation. A further point to consider is that when cathode bias is employed, the plate current, under load, may rise during tuning and exceed the non-oscillating value of current. When this happens, operate the oscillator between the equivalent points B and C on the corresponding rising plate current curve.

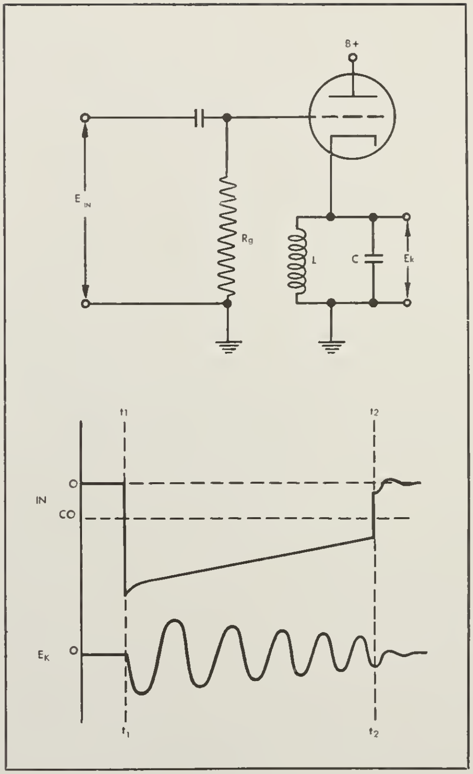
Shock-Excited Oscillators

A shock-excited oscillator is a type of oscillator which employs a vacuum tube as a switch to interrupt the steady flow of plate current through a resonant LC tank circuit as a means of exciting oscillations in the tank. This oscillator is used in operations where periodic oscillations of a certain frequency occurring over short intervals of time are required. Sometimes it is called a ringing oscillator.

One use of the shock-excited oscillator in radar is in range measurement operations. At the time the transmitter generates a pulse, transient oscillations are generated by a shock-excited type oscillator. These oscillations continue for the duration of the range sweep. Before the transmitter sends out another pulse, the transient oscillation set up by the shock-excited oscillator is damped out. In this way the range sweep is divided into a series of known intervals, each determined by the oscillatory period of the shock-excited oscillator.

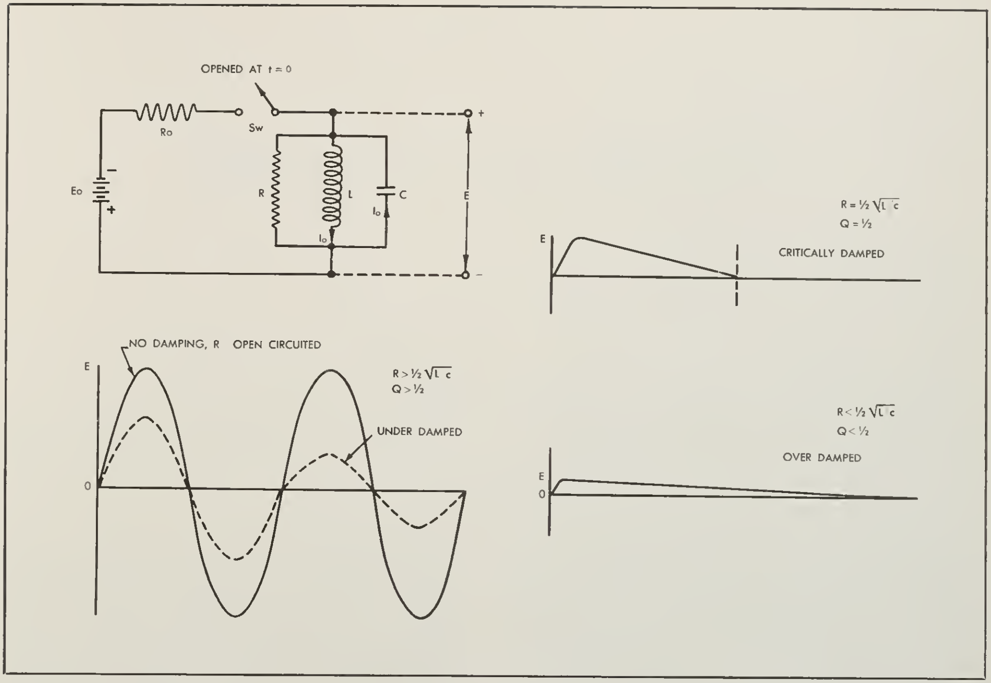
Another use of the shock-excited oscillator is in the peaker circuit. In this use, the sine wave output of the oscillator is used to overdrive an amplifier stage having a low Q resonant tank. The amplifier produces very sharp narrow peaks at a rate controlled by the voltage applied to its grid by the oscillator. The peaked output may be used as trigger or modulator pulses. The peaker circuit is discussed later.

Notice the circuit of a shock-excited oscillator and the input and output waveforms of the voltages applied and produced. Before each transmitter pulse, the tube conducts and a steady current flows through it and inductance L. At the instant t_1 the transmitter sends out a pulse, a negative gate pulse applied to its grid and the tube is cut off. However, the current through the inductor cannot stop instantly and flows into the capacitor, charging it in a direction making it negative, and starting the oscillations. Losses in the LC circuit are made as small as possible, with the result that the amplitude of oscillation is nearly constant throughout the duration of the input gate pulse. At the end of the gate pulse, time t_2, the tube again conducts and a second oscillation is started. However, the conducting tube is equivalent to a resistance shunted across the LC circuit and causes the oscillations to be damped quickly.



Shock Excited Oscillator

Before discussing the peaker circuit it is well, first, to investigate transients in the RLC circuit. Transients in RLC circuits are not all of the same shape. Their form depends upon the Q of the circuit. In the illustration on the next page the initial steady current Io is assumed to be



R-L-C Transients

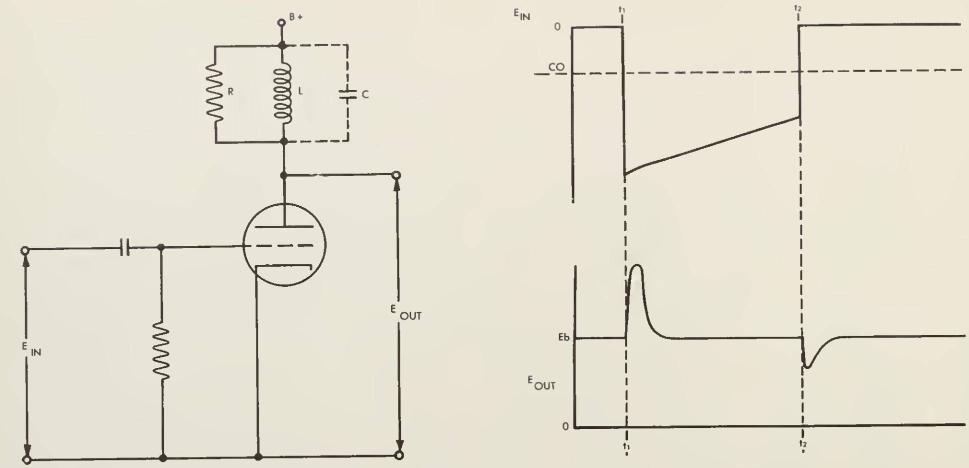
built up in the inductance L. The plotted variations of e are those following the opening of the switch. The curves are all drawn for the same values of L and C but for different values of Q corresponding to different values of R.

If R is open circuited, so that Q is infinite, the transient is a sinusoidal wave. Because the inductor current cannot change instantly when the switch is opened, the current Io flows from L into C and causes the voltage to increase because of the charge placed on C. The voltage across the capacitor causes the current to decrease. The voltage reaches its maximum value when the current is zero, because a further decrease of current means a reversal of direction and a discharge of the capacitor. At t = o, energy is stored in the inductance because of the magnetic field set up by the current I_o. When the inductor current decreases to zero, and the capacitor voltage is at its maximum, this energy is all transferred to the electric field of the capacitor. As the capacitor voltage decreases after reaching its peak, the energy is returned to the inductance, the return of energy being complete when E = o. The reversed current in L then charges C in the direction to make E negative, and so the process continues, energy passing from one element to the other and back, and E oscillating above and below the zero axis. Since the circuit is assumed to be loss-less, the amount of energy transferred does not change, and the oscillation continues at a constant amplitude.

If a high resistance R is used in the circuit, oscillating variations of E are still obtained but the amplitude of oscillation decreases from one cycle to the next, as shown by the dotted curve. Here Q is greater than one-half.

If the value of R is reduced, the rate of delay of the oscillations is increased. When R is reduced to $\frac{1}{2}\sqrt{LC}$ so that $Q = \frac{1}{2}$, the damping is so rapid that no negative swing of E is obtainable, and the circuit is said to be critically damped.

If the value of R is further reduced so that Q is less than one-half, the over damped transient is obtained.



R-L-C Peaker

The RLC peaker circuit illustrated above is actually another form of shock-excited oscillator. The capacitance is usually a lumped inductor and circuit capacitance. A resistance is connected across the LC circuit to produce nearly critical damping when the tube is cut off. Instead of an oscillation, therefore, a single sharp positive peak is produced at time t_1, and a negative peak at time t_2. The negative peak is smaller than the positive one because of the additional damping provided by the conducting tube after time t_2. The circuit is connected to the plate rather than the cathode, in order to obtain an output pulse of greater amplitude and steeper leading edge than the input pulse.

CHAPTER 8

Modulation and Detection

In radio transmission, the steady RF voltage wave generated by a transmitter alone cannot produce an intelligible signal at a receiver. It merely carries the intelligence to the receiver. In other words, it is the *carrier*. At the transmitter, the intelligence is impressed on the carrier by a process called *modulation*. When the modulated carrier is picked up by the receiver, the reverse of modulation occurs. At the receiver, the intelligence is removed from the carrier. This process is called *detection*. This chapter gives you the basic principles of modulation and detection and explains the operation of a large number of detector circuits commonly employed in radar.

MODULATION

Various methods are used for impressing the signal voltage (intelligence) on the RF voltage (the carrier) generated by the transmitter. Sometimes, this process (modulation) is accomplished by varying the *amplitude* of the carrier in accordance with the signal voltage. Sometimes, the *frequency* of the carrier is varied in accordance with the signal voltage, and sometimes it is the *phase* of the carrier that is varied. Thus, there are three common types of modulation—amplitude, frequency, and phase modulation.

Amplitude Modulation

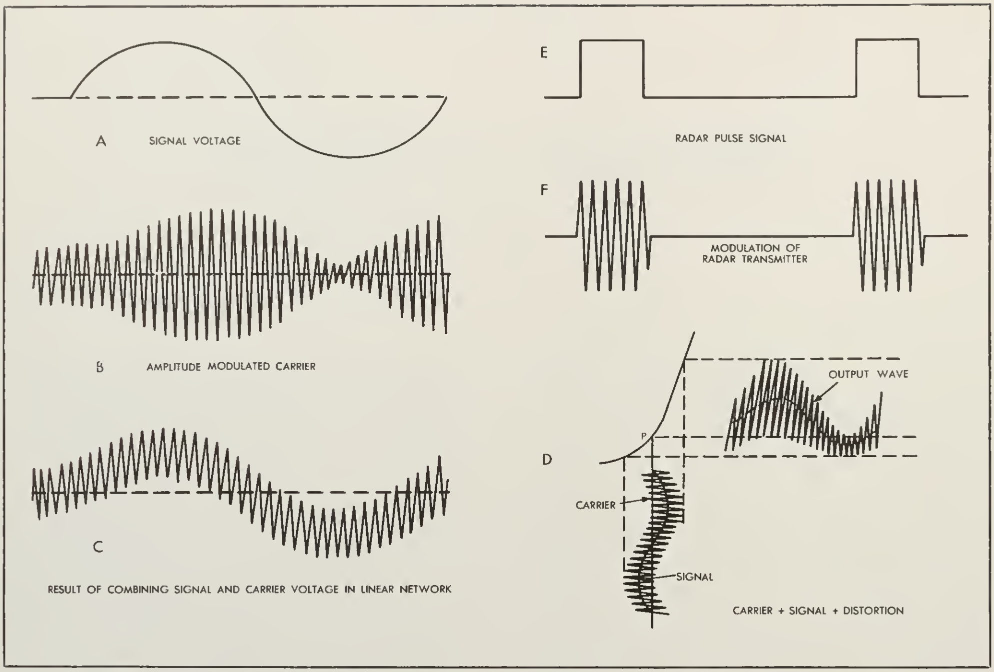
Amplitude modulation is the process by which the magnitude of a carrier wave is varied according to the frequency of the signal voltage. According to the typical amplitude modulated carrier illustrated at B on the next page, you can see that the amplitude of an amplitude modulated wave is high during the positive peaks of the signal voltage, but is near zero during the negative peaks. The shape of this wave is the result of introducing the carrier and signal voltage onto a nonlinear resistance. A nonlinear resistance is one in which the current does *not* vary in direct proportion to the voltage.

Whenever two signals are impressed on a *linear* reistance (one in which the current does vary in direct proportion to the voltage), the output waveshape contains the same frequencies as the input and both the negative and positive half cycles are like those shown in the wave shape at C. In this case, the RF amplifiers in the transmitter amplify the carrier, but discriminate sharply against the signal frequency. This causes only the carrier to be transmitted. Obviously, such an arrangement would defeat the purpose of modulation, as the signal arriving at the receiver would not contain a modulation component at all.

Impressing the carrier and signal frequencies on a nonlinear network, however, produces the output waveshape shown at D. In this waveshape the nonlinear characteristic is indicated by the bend of the curve at point P.

The E_pI_p curve of a nonlinear network is curved, while that of a linear network is straight. The output of a wave impressed on a nonlinear network contains both the carrier and the signal, plus a third component—distortion. As you can see in diagram D, the negative cycles are altered differently than the positive cycles. A sine wave altered in this manner is said to be a combination of the fundamental frequencies and additional frequencies caused by distortion. The distortion is introduced by the nonlinear resistance.

When a carrier frequency and a signal frequency are impressed on a nonlinear resistance, the output wave contains the following frequencies:



Amplitude Modulation

1. The carrier frequency.
2. The *upper side frequency* which is equal to the carrier *plus* the signal frequency.
3. The *lower side frequency* which is equal to the carrier *minus* the signal frequency.
4. Direct current.

When an output wave containing these frequencies is present in a transmitter, the tuned circuits respond to the carrier the upper side frequency and to the lower side frequency, but not to the signal. The result of this frequency response is a composite waveshape, like the amplitude modulated carrier shown at B. The direct current signal indicates the presence of modulation by varying the plate current meter in the modulator stage of the transmitter.

When a carrier frequency of 500 kc is modulated by a 1000 cps signal, the frequencies in the modulated stage become 1000 cps, 500 kc, 501 kc, 499 kc, and zero kc (DC). Of these, only the 500 kc, 501 kc, and 499 kc components are within the range of the transmitter circuit, and can be acted upon by the tuned circuits. Therefore, they are the ones that are transmitted.

An important point to remember in modulation is that the transmitted signal, which the receiver picks up, does *not* contain the signal frequency at all. Because of this, simple low-pass filters, which are commonly used to separate frequencies, are not capable of separating the signal from the carrier. Instead, receivers employ a detector stage which *generates*, or reproduces, the signal frequency when the carrier and side frequencies are impressed upon it. Later you will see that this detector is also a nonlinear electrical circuit.

For radar systems, the same principles hold true. A typical signal frequency such as shown at E, is a 1-microsecond square wave pulse which recurs at about 1000-microsecond intervals. This wave is composed of many harmonics (sine wave) frequencies. When a carrier is modulated by these pulses, wave shapes like those at F result which contain the carrier and a wide band of side frequencies. The side frequencies are located on each side of the carrier and are called *side bands.*

Frequency Modulation

Frequency modulation is the process which impresses the signal on the carrier by varying the frequency of the carrier in accordance with the frequency of the signal. In frequency modulation, the frequency of the carrier when it is not being modulated is called the *center frequency*. When the carrier is modulated by a positive signal voltage, its frequency becomes higher in proportion to the amount of positive signal voltage, and when it is modulated by a negative voltage, its frequency becomes lower. Therefore, when the sine wave shown at A is applied to a carrier, its frequency changes from the normal center frequency to *high*, to *normal*, to *low*, and to *normal* again in accordance with the voltage of the sine wave. The maximum frequency change from center, which depends upon the amplitude of the signal, is called the *deviation*. (Note that the amplitude of the modulated carrier is constant.)

The principle of frequency modulation was developed as early as 1920, but only recently has it been put to practical use. Radar equipment uses a special detector circuit, called a *discriminator* for detecting frequency modulated radio signals. This circuit gets its name from the fact that it discriminates between different frequencies. The discriminator is taken up later in this chapter.

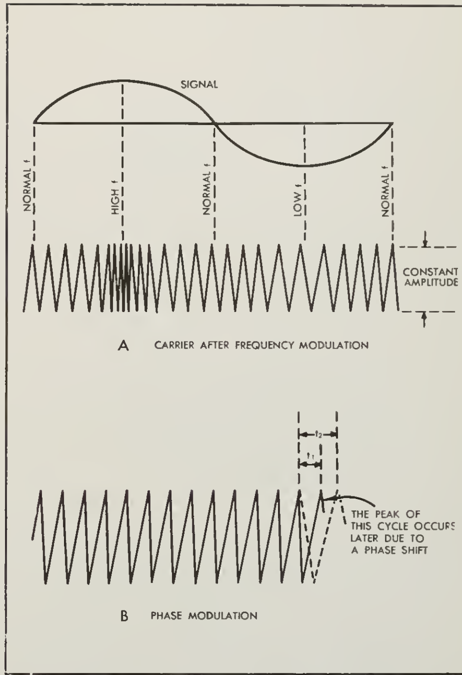
Phase Modulation

Phase modulation, which is a slight modification of frequency modulation, is characterized by changes in the phase of the AC carrier voltage instead of in its frequency. This means that the phase of a cycle shifts with respect to each preceding cycle and that the last cycle is ahead or behind its normal phase. This causes the peak to occur earlier or later than normally, as you can see in the illustration, at B which shows the normal period as t_1 and the longer period as t_2. The shorter period corresponds to the higher frequency. A frequency modulation detector responds to phase changes in the same manner as it responds to frequency changes.

DETECTION

Previously, it was stated that an amplitude-modulated waveshape contains the radio frequencies, but not the signal frequencies, and that it is the function of the detector in the receiver to reproduce the signal frequencies from the variations of the carrier and its sidebands.

Detection of amplitude-modulated signals requires a nonlinear electrical network. An ideal nonlinear curve for this is one that affects the positive half-cycles of the modulated wave differently than the negative half-cycles, and that so distorts a wave with an applied voltage of zero average value that the average resultant current varies as the signal frequency. The curve shown on the next page at A is called an ideal curve because it is linear on each side of the operating point P and does not introduce harmonic frequencies.

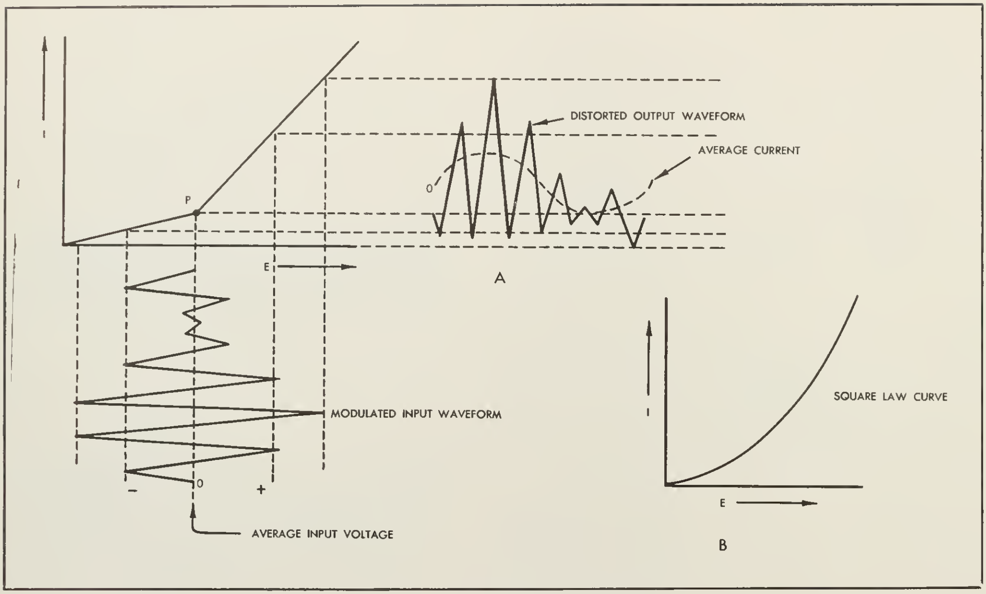


Frequency and Phase Modulation

When the input to an ideal nonlinear curve is a carrier and its sidebands, the output contains the following frequencies:

1. The carrier frequency.
2. The upper sideband.
3. The lower sideband.
4. A frequency equal to the carrier minus the lower sideband (or, a frequency equal to the upper sideband minus the carrier), which is the original *signal frequency*.
5. A frequency of zero or a DC voltage.

The detector reproduces the signal frequency by producing a distortion of a desirable kind



Nonlinear Device used as Detector

in its output circuit. When the output voltage of the detector is impressed upon a low-pass filter, which suppresses the radio frequencies, only the original or signal frequency is left.

In some practical detector circuits, the nearest approach to the ideal curve is the square-law curve shown above at B. The output of a device using this curve contains, in addition to all the frequencies which were listed, the harmonics of each of the input frequencies. The harmonics occur because voltage inputs which have a large amplitude are distorted differently than voltage inputs which have a low amplitude. The harmonics of radio frequencies can be filtered out, but harmonics of the signal frequencies, even though they produce an undesirable distortion, have to be tolerated. However, the square-law curve offers an advantage in that the output amplitude varies as the square of the input. In contrast, a linear curve produces an output which varies directly with the input. Therefore the square-law device produces a greater output for a given input.

Diode Detector

One of the simplest and most effective types of detectors and one with nearly an ideal nonlinear resistance characteristic is the diode detector. Notice the E_p I_p curve in the illustration at B on the next page. This is the type of curve on which the diode detector at A operates. The curved part of the curve is the region of low plate current, and indicates that for small signals the output of the detector will follow the square law. For input signals with large amplitudes, however, the output is essentially linear in the positive direction from the operating point. This type of detector is classed as a power detector since it handles large input amplitudes without much distortion.

On examining the diode detector circuit, note that the modulated carrier is introduced into the tuned circuit made up of LC_1. This circuit is designed so that the receiver has high selectivity. The waveshape of the input to the diode plate is shown at C. As a diode conducts only during positive half-cycles, it removes all the negative half-cycles and gives the result shown at D. The average output is shown at E. Although the average input voltage is zero, the average output voltage across R always varies above zero and has an average voltage of ½ x .637 x peak voltage for any positive half-cycle.

The low-pass filter, made up of the condenser C2 and the resistor R removes the RF (carrier) frequency, which, so far as the receiver is concerned, serves no useful purpose. The condenser C_2 charges rapidly to the peak voltage through the small resistance of the conducting diode, but discharges slowly through the high resistance of R. The sizes of R and C_2 normally form a rather short time-constant at the signal frequency and a very long time-constant at the RF frequencies. The resultant output with C_2 in the circuit is a varying voltage which follows the peak variation of the modulated carrier (see F). The DC component produced by the detector circuit is still in the waveshape but may be removed by the condenser C_3 producing the AC voltage waveshape at G. In communications receivers, the DC component is often retained for providing automatic volume (gain) control.

Advantages. The following are the advantages of diode detectors:

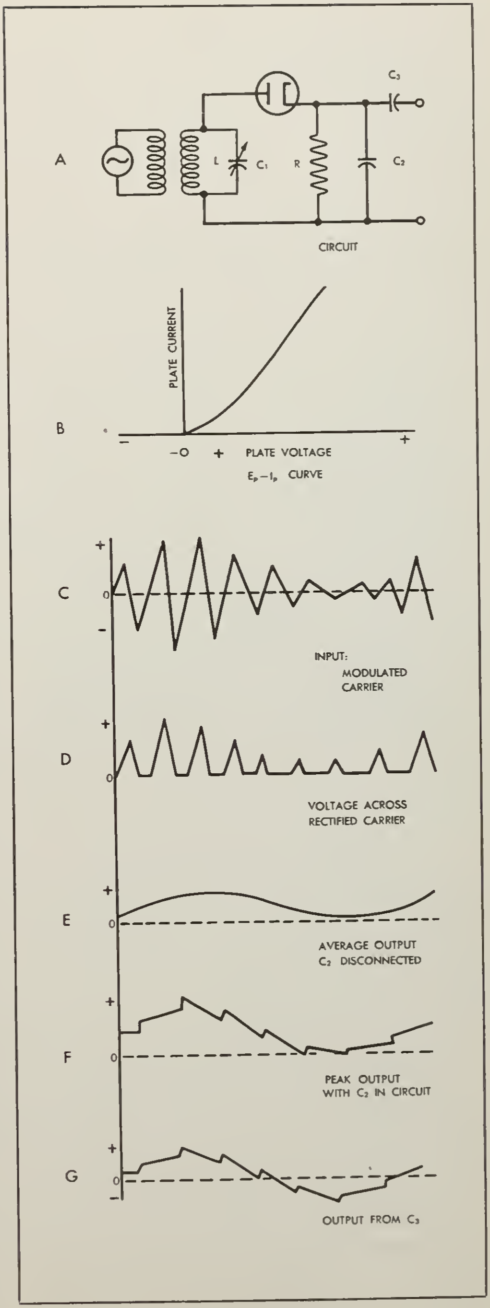
1. Ability to handle relatively high-power signals. There is no practical limit to the amplitude of the input signal.
2. Low distortion. Distortion decreases as the amplitude increases.
3. High efficiency. When properly designed, 90% efficiency is obtainable.
4. Diode detectors develop a readily usable DC voltage for the automatic gain control circuits.

Disadvantages. The following are the disadvantages of the diode detector:

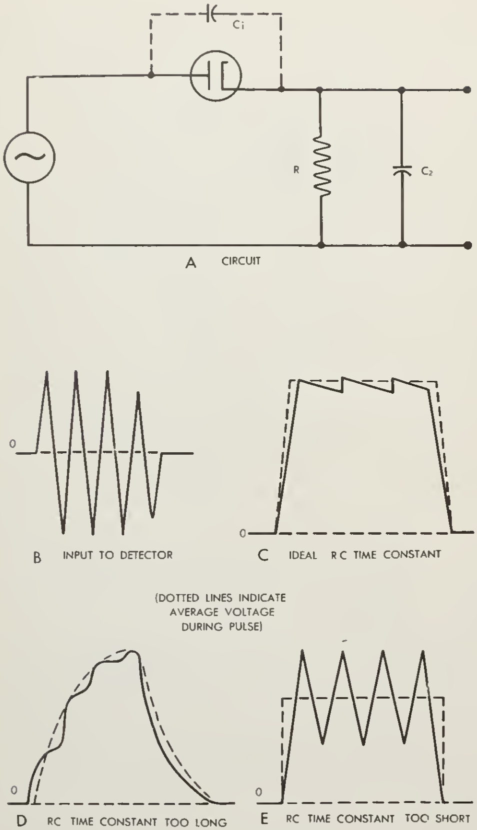
1. Power is absorbed from the tuned circuit by the circuit. This reduces the Q and selectivity of the tuned input circuit.
2. No amplification occurs in a diode detector circuit.

Component parts. Careful selection of component parts is necessary for obtaining optimum efficiency in diode detector circuits. One very important fact to consider is the value of the time constant RC_2, particularly in the case of pulse modulation. When a carrier modulated by a square pulse such as shown at B on page 8-6 is applied to an ideal diode detector, the detector produces a waveshape such as shown at C. Notice that for clarity the amplitude of the wave at C is exaggerated in comparison to the high-frequency carrier shown at B.

If the time constant of RC_2 is too long, several cycles are required to charge C, and the leading edge is sloped as shown at D, on the next page. After the pulse passes by, the condenser charges slowly and the trailing edge is exponential rather than square as desired. If the time constant is too short, both the leading and trailing



Diode Detector



Diode Detector Component Considerations

edges are steep, but the condenser discharges considerably between cycles, and thus reduces the average amplitude of the pulse and leaves a sizable component of the carrier frequency in the output as at E.

For these reasons, the selection of the time constant is a compromise between an output that is large but distorted, and an output that is undistorted but small and only partly filtered. To achieve this compromise, the load resistor R must be large so that the total input voltage will divide across R and the internal resistance of the diode when the tube is conducting. A large value of load resistance insures that the greatest proportion of this voltage will be in the output where it is desired. On the other hand, the load resistance must not be so high that the condenser C_2 becomes small enough to approximate the size of C_i, the internal capacity of the tube. When this occurs, the condenser C_2 will try to discharge through C_i during the non-conducting periods and reduce the amplitude of the detector output.

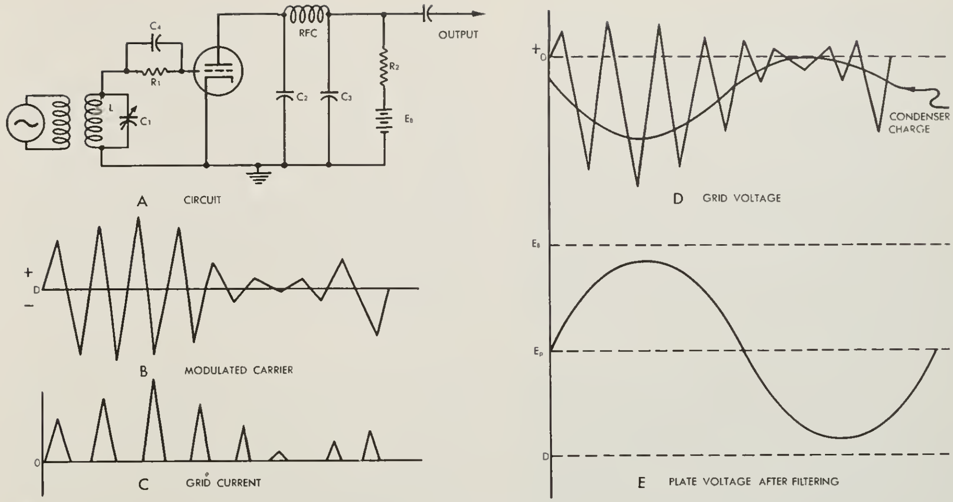
Grid-Leak Detector

The grid-leak power detector is similar to the diode detector with the advantage that it serves as a stage of amplification. Grid-leak detectors use no fixed bias arrangement in the grid circuit, and depend for their operation on the flow of grid current whenever the grid is positive.

In the typical grid-leak circuit illustrated at A on the next page, the grid, cathode, C_4, R_1 and the tuned circuit form a diode detector circuit. If the condenser C_4 and the resistor R_1 are correctly proportioned, the condenser charges during the period of current flow and discharges slowly through R_1 during periods of no current flow. Since the condenser maintains an average charge, it places an average negative voltage on the grid with respect to the cathode. (This voltage is represented by the heavy line in the grid voltage curve at D.) The RF voltage at the grid varies around this average value. If the condenser should discharge slightly, the next positive cycle of the modulated voltage will go higher in the positive direction, recharging the condenser. As the modulated waveform decreases in amplitude (that is, as the positive peaks become smaller), the condenser does not fully recharge with each cycle, and the average charge (and grid bias) reaches a lower value.

All this means that the average grid voltage varies in accordance with the variation of the amplitude of the modulated wave and, in doing so, reproduces the signal voltage. Consequently, the plate current and the plate voltage shown at E also vary at the signal rate. The low-pass filter in the plate circuit prevents the plate voltage from changing at an RF rate and thus removes the RF frequencies from the output. The normal gain of the triode stage produces an amplified waveshape in the plate circuit. The time constant of C_4R_1 is governed by the same factors as the filter used in the diode detector.

Advantages and Disadvantages. The voltage applied to a grid-leak detector must not be so high that it causes the average grid voltage to exceed the plate current cutoff voltage for the tube. This characteristic limits the power handling capacity of the grid-leak detector. Like the diode detector, the grid-leak detector pro-



Grid Leak Detector

duces little distortion when the received signal is strong. Although it produces only a limited amount of DC voltage for AVC, the grid current required for this purpose reduces the Q of the preceding stage. Due to the amplification properties of a grid-leak circuit, its output for a given input is much greater than that of the diode detector.

Plate Detector

The plate detector shown below at B, sometimes called a bias detector or anode detector, is similar to a class A amplifier stage except that it is operated at very near cutoff bias. For ideal operation, this bias may be provided from a fixed DC potential, but for practical purposes, a well bypassed cathode resistor will bias the tube near

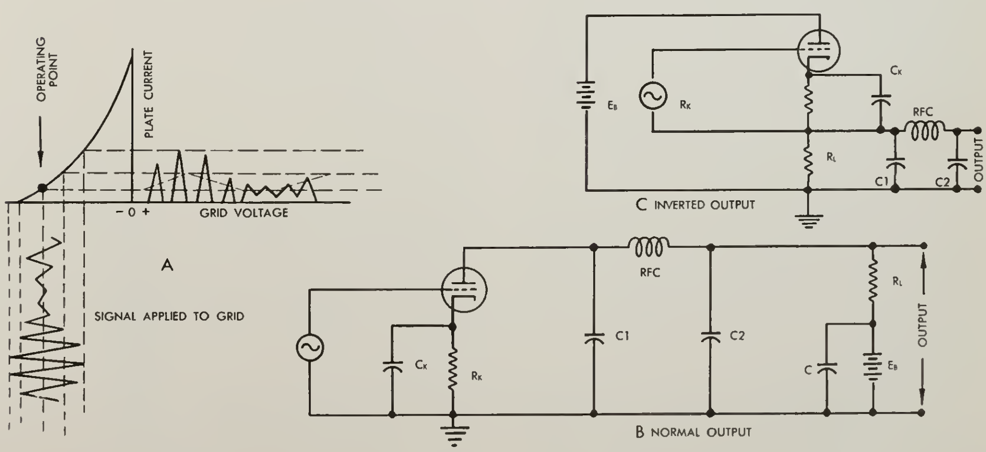


Plate Detector

enough to cutoff. For this reason, R_k, the cathode resistor, shown in the circuit at B, is rather large, and the bypass condenser C_K is large enough to bypass both carrier and modulation frequencies. Due to the extremely high operating bias, the tube characteristics are very nonlinear as you can see by examining the E_g-I_p curve at A. Plate current flows only during positive halves of the input signal and a small portion of the negative halves. The grid does not draw current as long as the signal does not drive the grid into the positive region of the E_g-I_p curve. The average plate current, shown at A by dotted lines, varies as the modulation varies.

The removal of the RF component from the plate circuit is accomplished by the low-pass filter, formed by the choke RFC and condenser C_1 and C_2. The variation remaining after removal constitutes the original modulation. Because of the amplification factor of the tube, the voltage representing the modulation which appears across the load resistor is greater than the input (grid) voltage. All other AC voltages are effectively placed at ground potential by condenser C_1.

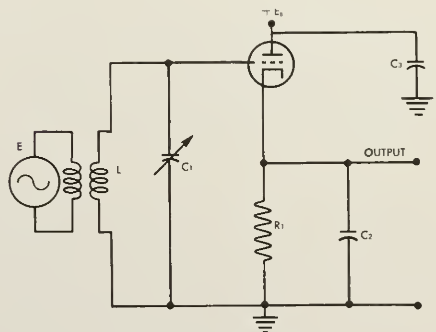
For pulse type modulation, the plate detector operates in the following manner. When a signal is applied to the detector, a pulse appears as a positive pulse at the grid and as a negative pulse in the plate circuit. In some applications positive output pulses are desired. This is done by changing the circuit so that R_L is located between the cathode-to-grid connection instead of in the plate circuit. By this arrangement, the output has the correct polarity and may be taken across R_L. The other ground connections are removed, as you can see, by examining the inverted output circuit for pulse type modulation in the illustration at C.

ADVANTAGES AND DISADVANTAGES. The advantage of the plate detector is that, since the grid does not draw current, the Q of the input circuit remains the same and amplification occurs in the circuit. Its disadvantages are higher distortion and lower power handling ability than the diode detector.

Infinite-Impedance Detector

The infinite-impedance detector is a modified type of plate detector. It has the same ability as the diode detector to handle large input amplitudes with little distortion and without drawing current from the circuit driving it. Drawing current by any type of detector causes reduction in the Q of the tuned circuit and a decrease in the selectivity.

In the illustration showing the infinite impedance detector, E is the modulated carrier voltage. LC_1 is a tuned circuit which contributes to the selectivity of the receiver. The plate of the triode is connected directly to the power supply. Condenser C_3 is a virtual short circuit to ground for AC voltages. The load resistance R_1 is in series with the cathode. The condenser C_2, in conjunction with R_1, forms a long RC to the RF frequency but a short RC to signal frequencies.



Infinite Impedance Detector

Under no-signal conditions, the plate current through R_1 provides a bias which keeps the grid very near cutoff. Under signal conditions, when an RF voltage is applied, the positive peak of the modulated voltage charges the condenser C_2 to a higher voltage. This condenser holds its charge between peaks because of the long-time constant of R_1C_2. No grid current flows during the positive peaks because the cathode bias changes at the same time that grid voltage changes, and the cathode-grid voltage difference is never great enough to exceed the cutoff value. When the crest of the modulation peak has passed, the amplitude of the peak decreases and the condenser discharges at the same rate. Therefore, the condenser charge follows the modulation peaks in the same manner as in the diode detector.

The infinite-impedance detector employs a very high input impedance for two reasons. First, the grid does not draw current. Second, because of the capacity in the plate circuit, the impedance reflected into the grid circuit is purely resistive.

Another characteristic of this detector is that like the cathode follower, negative feedback exists due to the common input and output impedance at signal frequencies. This negative feedback is responsible for reducing the distortion due to characteristics of the vacuum tube.

Distortion is also minimized by the employment of high input voltages. This input voltage can be as large as half the DC voltage on the plate without overdriving the tube. The ability of the tube to handle large signals is due largely to the fact that the bias increases as the signal input increases.

Crystal Detector

A small piece of silicon, galena, iron pyrite or carborundum is capable of providing the nonlinear characteristic required for detection. When a crystal of one of these elements is in contact with a sharp pointed wire, the arrangement will offer a different resistance to current flow from the crystal to the wire than it does from the wire to the crystal. Such a device is called a crystal rectifier or a crystal detector. (Do not confuse this type of crystal with piezoelectric crystal used to control the frequency of oscillators.)

For use in radar sets, the crystal is silicon and the wire is tungsten as shown below at D. Since some spots on a crystal are more sensitive than others, the point of the tungsten wire is placed in contact with the most sensitive part, and the contact is made permanent by using a sealing compound and mounting the assembly in a watertight, standard sized cartridge.

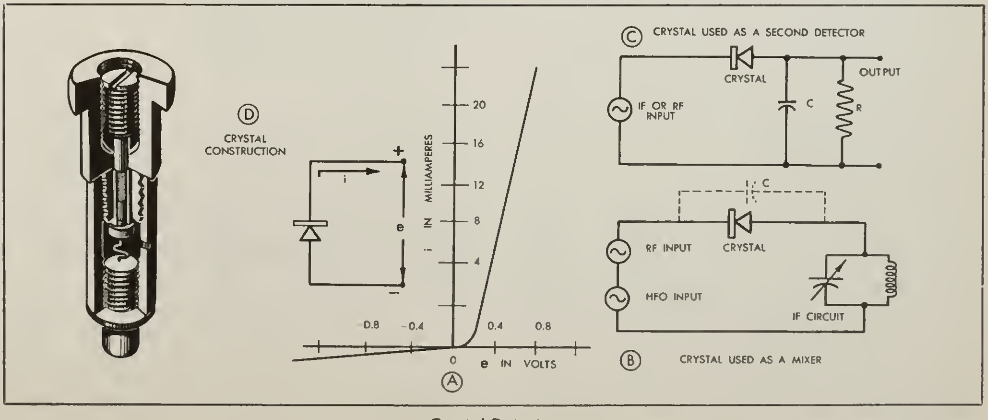
USE OF CRYSTALS. Crystals are most frequently used in the mixer stage in the superheterodyne receivers of radar sets. Seldom are they used as detectors to demodulate the modulated carrier. Their use as mixers requires the same nonlinear characteristics as are required in a detector circuit.

Mixer crystals are designated according to use and frequency. The crystal suitable for 10 cm waves comes in types 1N21, 1N21A, or 1N21B. A crystal for 3 cm waves comes in types 1N23, 1N23A, or 1N23B. The three types for each frequency are characterized by poor, medium, and high frequency sensitivity in the order given. The crystal type 1N27 is the type used for detector circuits.

As shown in the crystal mixer circuit at B, the radio-frequency and the local oscillator signals are applied to the resonant circuit through the crystal. The nonlinear characteristics of the crystal cause the voltage across the resonant circuit to contain the sum and difference frequencies, and in addition the original frequencies. The coil condenser circuit resonates to the difference frequencies. The other frequencies are rejected and the difference frequency is applied to the IF amplifiers.

The second detector circuit at C using the crystal operates similarly in all respects to the diode detector.

DC VOLTAGE-CURRENT CHARACTERISTICS. Notice in the DC voltage-current characteristic curve of a crystal at A that the current flow in the crystal, due to negative voltages, is very small.



Crystal Detector

When the crystal is used as a rectifier at 3000 mc, the capacity between the point and the crystal is about 1mmf, making the capacitive reactance about 50 ohms. With one negative volt applied, the curve shows that the resistance of the crystal is about 300 ohms. At 3000 mc, the 50 ohms capacitive reactance is in parallel with the 300 ohm DC resistance, and the actual back impedance is about 43 ohms.

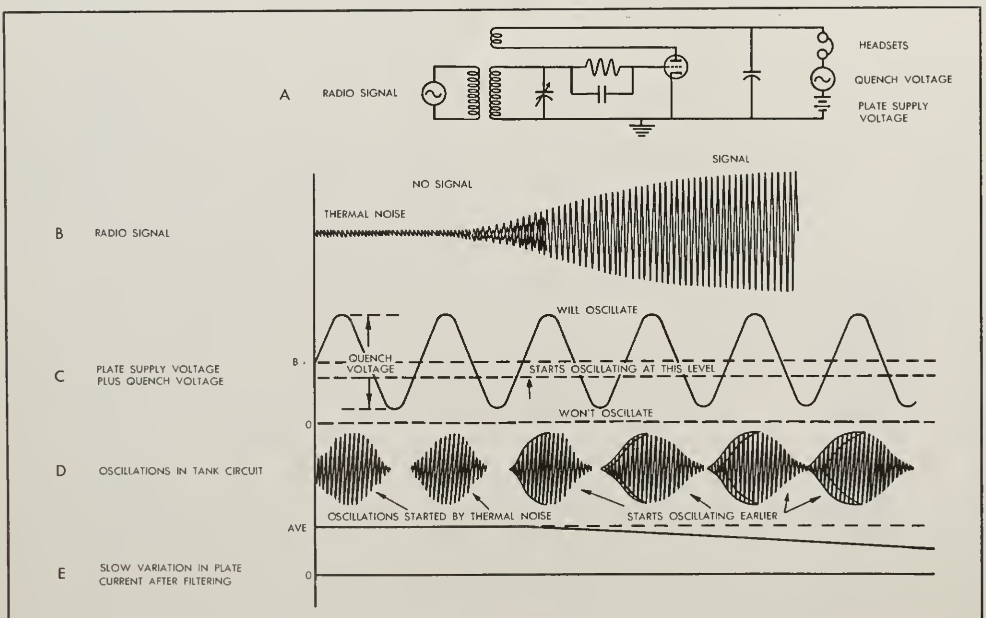
CRYSTAL SENSITIVITY. According to the curve, a normal crystal has a resistance ratio of 6 to 1 when measured either way. Measuring the resistance each way with an ohmmeter is a rough test of the crystal sensitivity. A somewhat better test for checking the crystal sensitivity is the following one. When you know that the back (high) resistance of the crystal at the time is good, but later find it has dropped one-half this value, the sensitivity in the crystal is too low for use. The only solution is to replace it with a known good crystal.

Neither of these tests are conclusive because of the change in back impedance at high frequencies. Since the factory test equipment is not available in the field, you may test the crystal at high frequencies by comparing the signal-to-noise ratio of the receiver with a questionable crystal installed against this ratio with a good crystal installed.

ADVANTAGES AND DISADVANTAGES. Even though the contact point causes reduced sensitivity, the crystal is superior to vacuum tube mixers at 3000 and 10,000 mc since it introduces less noise into the receiver system. This means a better signal-to-noise ratio. In addition, there are no transit time difficulties involved in crystals. Disadvantages of crystals are that they damage easily, even while installed in the radar set, and have a gain of less than one. Damage in the radar set usually results from inadequate protection from the transmitter pulse. As a second detector, the crystal has the advantage of being smaller in size than vacuum tube detector circuits and requiring no heater voltage.

Superregenerative Detector

The superregenerative detector is an extremely sensitive detector. It has a gain comparable to a complete superheterodyne receiver. On examining the circuit of the superregenerative detector below, you can see that it consists of an RF oscillator and a tuned circuit which is resonant to the frequency being detected. The superregenerative



Superregenerative Detector

detector also includes a special circuit which allows the detector to break into oscillation, then quench or stop the oscillations at regular intervals. This action is accomplished by variations in the detector plate voltage. When the plate voltage is above a certain value, the circuit oscillates; below this value, it ceases oscillations.

The voltage supplied by the special circuit is called the quench voltage. It varies the plate voltage above and below the value just mentioned at a rate above the audible range as shown at C on page 8-10. As the plate voltage gets near the value called the critical value, small irregular current changes in the cathode emission cause the plate voltage to rise above the critical point, and the circuit oscillates until the plate voltage drops below the critical value. When the plate voltage comes back up again, the process is repeated. However, when a radio signal is applied to the tuned circuit, the signal adds enough voltage to the grid circuit to cause oscillations at a somewhat lower value of plate voltage. This causes oscillation to start sooner and to last until the plate voltage drops back to the critical value.

Under signal conditions, the duration of the oscillation cycle is longer than under no signal conditions. Within limits this time varies directly with the signal strength. Short periods of oscillation build up small charges on the grid leak condenser, while long periods of oscillation cause grid current to flow longer and consequently build up greater charges. The charge in turn determines the average plate current, and it in turn varies with the signal strength. If the signal is amplitude modulated, the variation in plate current will roughly resemble the modulation voltage.

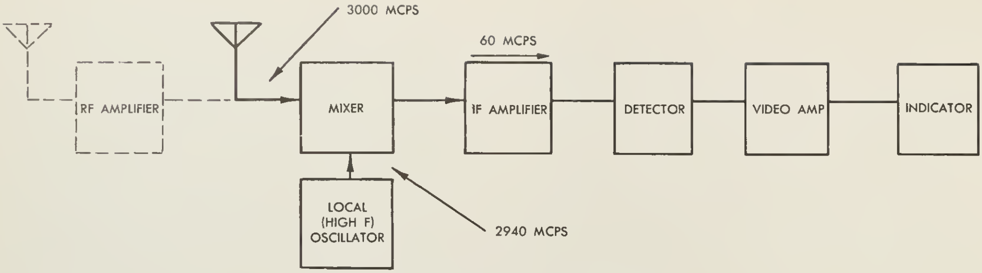
Since the incoming radio signal only initiates the oscillation while the circuit itself determines the amplitude, the circuit realizes a tremendous gain when properly adjusted. In fact, any signals stronger than the thermal noise level are audible. Since this is the limit of amplification in a superheterodyne receiver, the sensitivity of a superregenerative detector is near that of the superheterodyne receiver.

The superregenerative detector has the advantage of high gain and light-weight components. It finds its greatest use at very high frequencies where distortions and noise are compensated by the high gain, which is not obtainable with conventional circuits. Another advantage is the fact that the output voltage is about the same for weak or strong signals. Weak signals bring it to full oscillation equally as well as strong signals. (For this reason, any noise voltages stronger than the signals are reducible to the level of the signal. Suppressing ignition noises and other types of interference further reduces the annoyance to the listener.) Its disadvantages are excessive distortion, poor selectivity, and a characteristic hiss when no signal is being received. The poor selectivity is due to the fact that the detector responds to signals off frequency as well as to those on frequency, even though the amplitude of the off-frequency signal is less. The hiss, when no signal is being received, is due to the noise produced by the random thermal voltages in the tube which start oscillations at irregular times. The hiss disappears when a signal is present, since the signal has a definite amplitude and since the oscillations start at a definite time. The distortion in the detector results from grid current clipping the positive peaks and from accentuation of the negative modulation peaks.

Frequency Conversion

Many stages of amplification are required in a radar receiver in order to make a very weak echo voltage at the antenna give a good indication on the radar indicator screen. In the simplest type receiver, there are several RF stages followed by a detector and by a video amplifier. In spite of this, the gain (amplification) and selectivity of an RF amplifier decreases as the operating frequency increases. Therefore, it is desirable to convert the high radio frequency to a lower frequency, and then to amplify the low frequency. Furthermore, it is easier to design RF amplifiers with a higher gain when the frequency is constant. Therefore, the modern receiver employs a high gain and highly selective amplifier in the IF stage designed to operate at one frequency to which all received frequencies are converted.

Frequency conversion is accomplished by mixing the RF frequency with the local oscillator frequency in a superheteordyne receiver. This mixing produces a frequency equal to the difference of the two frequencies. If the two frequencies are nearly the same, the difference-frequency will be low, but not as low as the video frequencies. This frequency lies between the RF and video frequencies, and is called the *intermediate frequency.* If the RF frequency is below 1000 mcps, it is usually amplified by an RF amplifier before being fed to the mixer. If

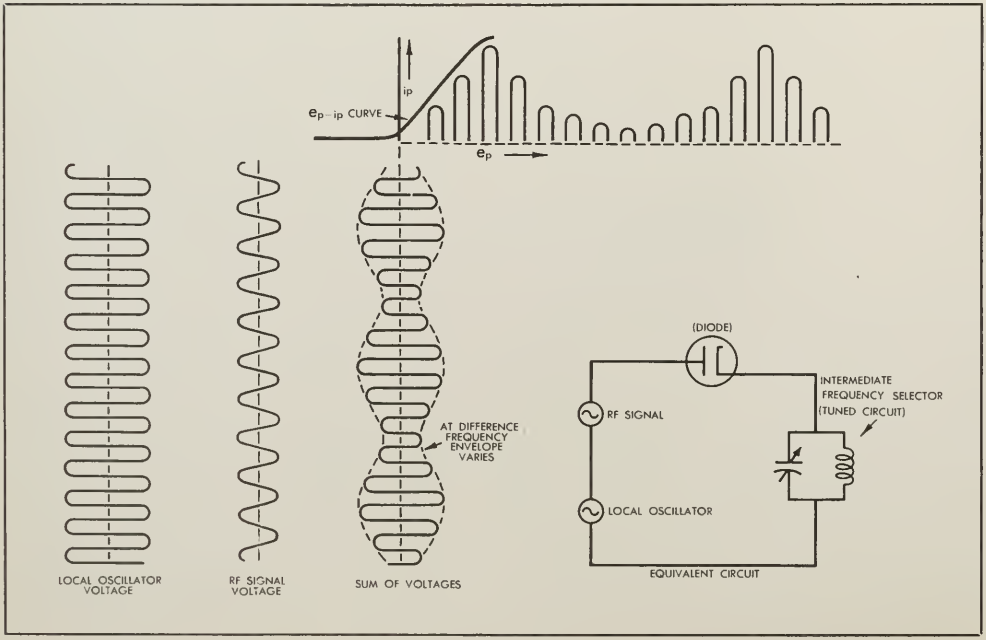


Block Diagram of Superheterodyne Receiver

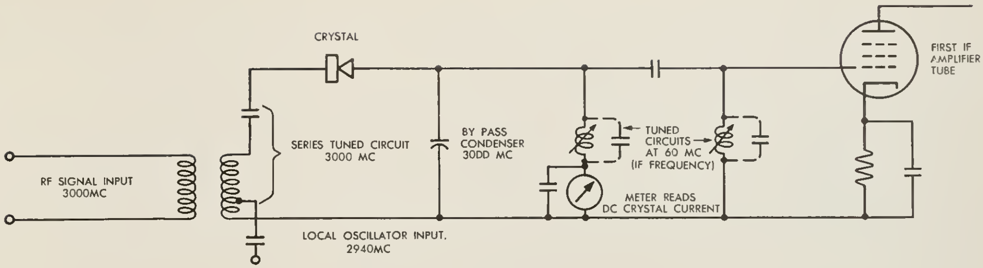
higher, the RF signal is introduced directly into the mixer. After mixing, the IF frequency is amplified by an IF amplifier and then is detected in the same manner as any other signal.

Production of a difference-frequency in the mixer stage requires the use of a nonlinear electrical circuit for the generation of the new frequency. The mixer, as this circuit is called, produces the new frequency in the same way as a modulator or a detector. For this reason, this stage is often called the *first detector*.

When two slightly different frequencies are added as shown below, the waveshape representing the combining of the local oscillator frequency and the RF signal frequency contains only these two frequencies. After distortion by the mixer the waveshape contains not only these two frequencies, but a frequency equal to the sum of the two frequencies and another equal to the difference between the two frequencies. Since it is the difference frequency that is desired, the tuned circuits in the IF amplifier select this frequency and amplify it. If the RF signal is modulated, the difference-frequency will also be modulated in the same manner.



Mixer Action



Crystal Mixer

Any of the detector circuits previously discussed will work in this stage. Grid-leak and plate detectors are usually used in the mixer stages of communications receivers in order to utilize the gain of the triode or pentode used. In low-frequency radar sets, the diode is sometimes used. In high frequency sets, however, the transit time in the tube and the thermal noise it introduces makes the vacuum tube inferior to the crystal.

Most 10-cm and all 3-cm sets use the silicon crystal for the nonlinear element in the mixer stage. In the simplified crystal mixer circuit shown above, currents from both the RF signal and local oscillator signal flow through the crystal. The difference-frequency is selected by the tuned circuits, which are connected to the grid of the first amplifier tube. Any nonlinear device produces a direct current which may be read on a meter. The adjustment of the oscillator for maximum output is determined by observing the meter reading, which varies directly with the amplitude of the oscillator signal.

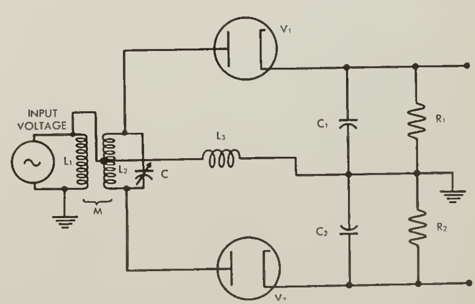
Discriminator Circuit

Even though the design and construction of radar equipment is quite exacting, the frequencies of the transmitter occasionally change. This in turn causes considerable trouble at the receiver. The receiver is no longer tuned exactly to the RF signal frequency, and signals which supply weak returns may no longer be visible on the indicator screen. Any change of the transmitter frequency, therefore, requires either continual readjustment of the receiver tuning by the operator or the employment of an automatic tuning device. Constant frequency checking by the operator not only is time consuming, but certainly is not in the interest of optimum efficiency. Therefore, most radar receivers employ an automatic frequency control circuit in which changes of frequency are detected by a discriminator circuit.

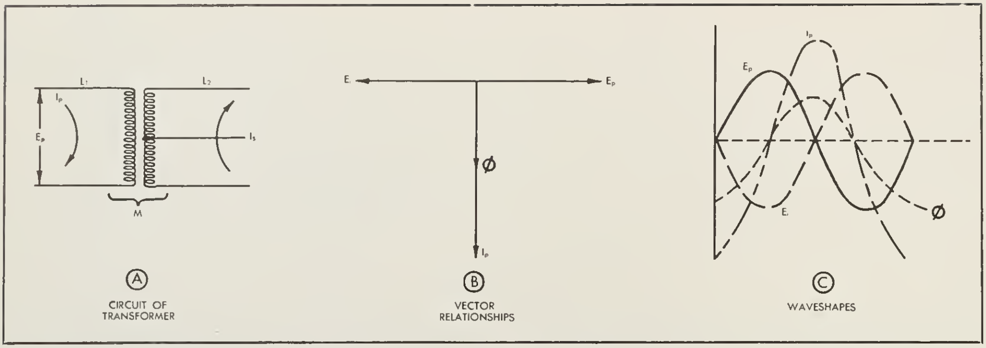
The detection of a change of frequency by the discriminator is accomplished by a transformer, which, due to special connections, distorts the frequency change into a voltage or varying amplitude change. After distortion, the voltage of varying amplitude is detected in the manner employed by any amplitude modulation detector. The illustration below shows the circuit of a typical discriminator circuit.

The two coils L_1 and L_2 are the primary and the secondary, respectively, of the IF transformers. The secondary coil is tuned to the correct transmitter frequency by condenser C. Tube V_1 and V_2 are diode detector tubes. The filters R_1C_1 and R_2C_2 serve the same purpose as the filters in the diode detector, that is, the removal of the RF component from the circuit.

The presence of L_3 and its connection to the primary at the top and to the secondary at the center presents an unusual feature in the discriminator circuit. L_3 is an RF choke which has a high reactance to the RF frequency. Since it is connected across the primary, the primary volt-



Discriminator Circuit



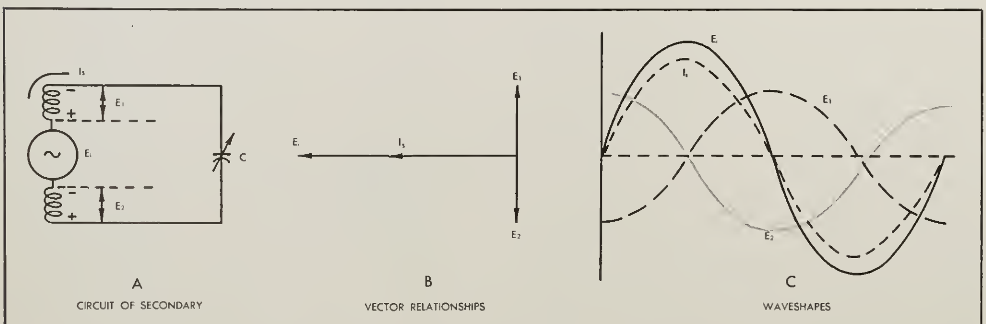
Transformer Vector Relationships and Waveshapes

age appears across it at all times. The connection to each diode from this choke causes the primary voltage (E_p) to appear at the plates of the diode with the same phase shift in each case. This voltage causes currents to flow in the opposite directions in the resistors R_1 and R_2, resulting in zero output. Therefore, the discriminator is not affected when the amplitude of the applied voltage changes.

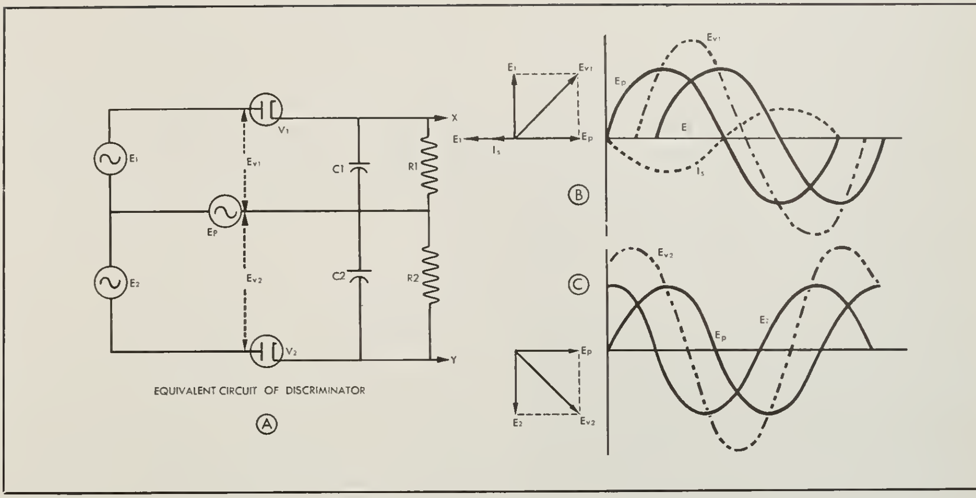
The phase relationship of the voltage across the RF choke to the voltage induced in the secondary is the key to the operation of the discriminator circuit. For this reason you will find it is desirable to analyze its operation in detail. To make the analysis refer to the illustration above showing the discriminator transformer at A and the vector relationships of the current and voltages in the transformer and their waveshapes.

At the desired frequency the transformer vector relationships are as follows: The primary voltage E_p is the reference vector. Since the mutual inductance (coupling) is small, the primary is inductive. The primary current I_p lags E_p by 90°. The magnetic field which affects the secondary (called the *flux* and shown by the symbol ϕ) is in phase with the primary current. Due to normal transformer action, the voltage induced in the secondary is 90° behind the flux. This induced voltage, labeled E_i, is simulated by a generator in the equivalent circuit of the secondary shown at A below. The generator is inserted at the center of the transformer secondary where the voltage across the secondary is divided between the two tubes. A current I_s flows around the loop composed of the secondary windings and the condenser C. This loop is a series-resonant circuit insofar as the generator sees it. Since the circuit is at resonance, the current I_s is in phase with the induced voltage.

Each half of the secondary has considerable inductive reactance at the RF frequency. Therefore, there is a voltage drop across this reactance due to the current I_s. These voltages are shown



Secondary Voltages at Resonance



Resultant Plate Voltages at Resonance

at A in the secondary circuit as E_1 and E_2. Since voltage leads current by 90° in an inductor, E_1 and E_2 are 90° out of phase with I_s. The voltage E_2 actually leads I_s as it should, but the voltage E_1 is 180° out of phase; that is, its polarity is reversed, due to the manner in which it is connected to the tube. Current flows in the same direction through both coils, but the vector voltages are both measured with respect to the center. This causes opposite polarities to exist.

Referring above at A, to the equivalent circuit of the discriminator in which the voltages are replaced by generators, note that the plate voltage on V_1 is the sum of the voltages E_p and E_1, and the plate voltage on V_2 is the sum of E_p and E_2. The vector diagram at B shows E_p and E_1 added vectorially to produce a vector sum E_{V1}. E_{V1} is the actual plate voltage on V_1. The two sine wave voltates E_p and E_1 are also shown. They add to produce a third sine wave called Ev_1 When plate current flows, C charges to near the peak voltage of E_{V1}, producing a constant DC voltage almost equal to E_{V1} across resistor R_1. In the same manner, E_p and E_2 add as E_p and E_1 did previously resulting in E_{V2}, the plate voltage on the other diode. Consequently, C_2 charges to E_{V2} and a constant voltage almost equal to E_{V2} appears across R_2 as shown in diagram C. The vector and sine waves are shown at C.

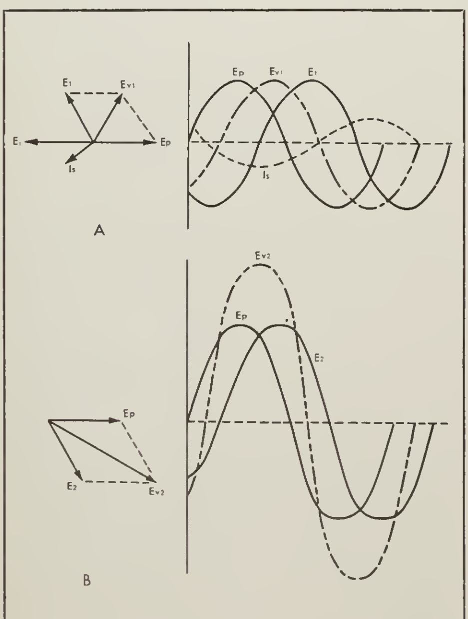
Observe that the voltage across the filters R_1C_1 and R_2C_2 are DC voltages. Obviously, phase relationships cannot exist there. In this case only amplitude relationships are important. Since E_p is common to both and E_1 equals E_2, the vector sums of E_{V1} and E_{V2} are equal. The sine-wave illustrations show that the resultant of their sine-waves, although of different phase, have equal amplitudes. The DC voltages across each filter are alike, but have opposite polarities. Therefore, the sum of these voltages will be zero between X and Y. This fact indicates that the output at resonance is zero.

OUTPUT AT FREQUENCIES BELOW RESONANCE. When the frequency is lower than the center or resonant frequency the output is no longer zero. Using the primary voltage as a reference, the voltage induced in the secondary is still 180° out of phase, as shown by the vector diagrams at A on page 8-16. Since inductive reactance decreases as frequency decreases, while capacitive reactance increases, the secondary circuit of the transformer at A on bottom of page 8-14 will be capacitive at frequencies below resonance. The current in a capacitive circuit leads the applied voltage. Here, the applied voltage is E_i, and I_s will lead. An arbitrary amount of lead is used in the vector diagram at A. The voltage drop E_1 and E_2 will still be 90° out-of-phase with the current since this is not a function of frequency. Since under these conditions voltage E_1 is more than 90° away from E_P when added vectorially, their sum will be less than at resonance even though the magnitude of the components E_p and E_1 has not changed. At the same time, E_2 is less than 90° away from E_p—or they are more nearly in phase—so the magnitude of the resultant E_{V2} is greater than at resonance. Again, since the resistor voltages depend upon the length of the resultant vectors, there will be a high voltage across R_2 and a lower voltage across R_1. Therefore, the output voltage will be equal to the R_2 voltage minus the R_1 voltage and the output point X will be negative with respect to Y which may be grounded.

The waveshape shows the same thing. I_s is shifted in phase; voltage E_1 is further out-of-phase with E_p, and E_2 is more nearly in phase with E_p. The sum of the more out-of-phase components is small; but the resultant waveshape of the two nearly in-phase components is large.

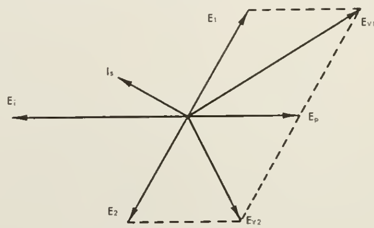
Studying the vector diagrams, you can see that the greater the deviation from resonance, the greater the DC output voltage.

Output at frequencies above resonance. When the input frequency is greater than the center frequency, the circuit L_2C is inductive, and the current I_s lags the induced voltage.



Voltages at Frequencies Below Resonance

The illustration below shows the vector relationship of voltages at frequencies above resonance. Note that at frequencies above resonance, E_1 is more nearly in-phase, and E_2 is more out-of-phase than at the center frequency. Further notice that E_{V1} has a greater magnitude than E_{V2}. Therefore, the DC voltage across R_1 is greater than the R_2 voltage, and consequently the output voltage is positive.



Voltages at Frequencies Above Resonance

In summary, the following facts are worth remembering: A frequency below resonance causes a positive output at X with respect to Y. At resonance, the output is zero, and above resonance the output is negative. These polarities may be reversed by measuring the voltage at Y with respect to X.

Use. The discriminator circuit is employed as a part of the automatic tuning circuit in a receiver. Part of the IF signal is diverted to the input transformer of the discriminator. The output of the discriminator is connected to the local oscillator of the receiver in such a manner that the DC voltage affects the frequency of the oscillator. If the IF frequency is correct, the receiver is operating at peak efficiency, the discriminator has a zero output, and the oscillator frequency is not affected. If the IF frequency changes, the receiver is said to be detuned, and the IF section operates at lower gain. However, the discriminator develops a DC voltage from this incorrect frequency and this DC voltage changes the frequency of the oscillator. By proper circuit design, this change can be made to occur in a direction which will correct the IF frequency.

The oscillator, of course, beats with the incoming RF signal to form the IF signal. Therefore, a change of oscillator frequency, or a change of RF frequency, changes the IF frequency. The discriminator and its associated circuit corrects either change. This system is known as *automatic frequency control*, and is usually abbreviated as the AFC.

CHAPTER 9

Radio Frequency Lines

A radio frequency line is a conductor or a group of conductors used singly or in multiple to transmit electrical energy from a source to a load. Radio frequency lines are necessary because surrounding objects modify radiation patterns and reduce the effectiveness of any kind of antenna.

The purpose of this chapter is to acquaint you with the principles of radio frequency lines. It discusses their theory, uses, and construction. It also explains the types of losses to which radio frequency lines are subject, methods of making measurements on them, and how the principles of radio frequency lines are used in constructing artificial lines.

DEFINITION

A radio frequency line is a special form of transmission line. A transmission line is a device, usually a pair of electrical conductors, which transfers electrical energy from a power source to a distant electrical circuit requiring power. The using device is called the *load*.

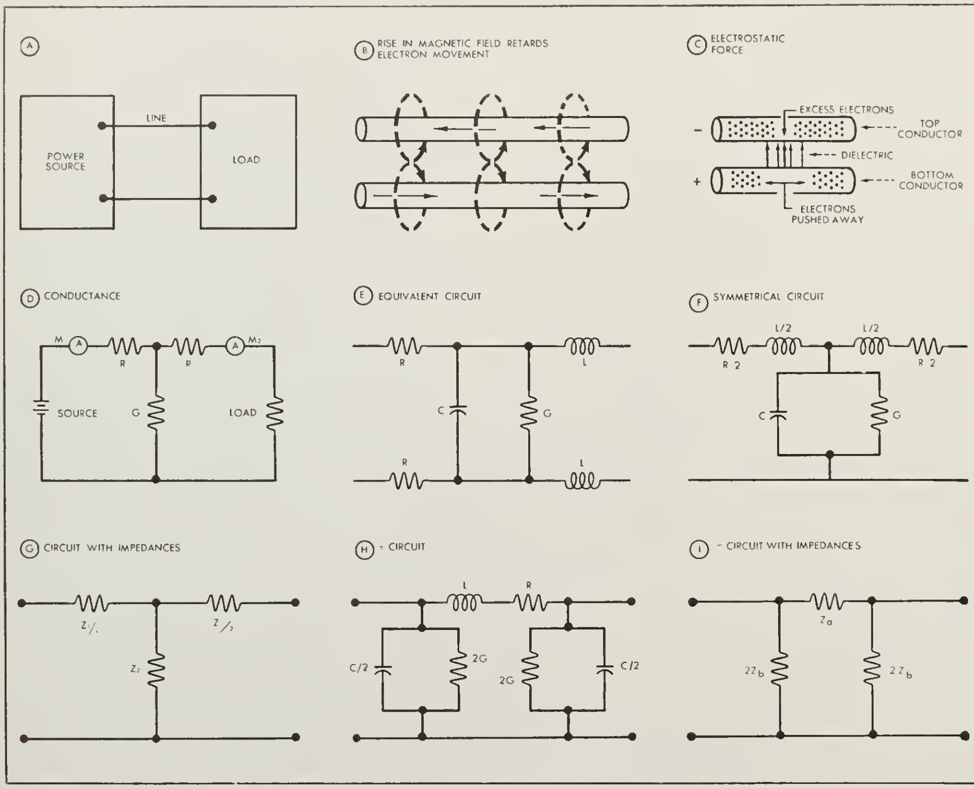
In transferring DC power, the sole requirement of a transmission line is that it offers low DC resistance. In transferring AC power at frequencies as low as 60 or 400 cycles per second (commercial and airplane power frequencies), the reactance of a transmission line is slightly noticeable, but it is negligible compared to its resistance. As frequency increases, however, the effects of capacity and inductance increase, until at radio frequencies (above 50 kc) the reactive components retard the transfer of energy to such an extent that a complete cycle may be generated at the source before the voltage at the start of the cycle reaches the load. Whenever this occurs, the line is said to be *electrically long*. At the power frequency of 400 cps, this phenomenon occurs over a distance of 400 miles. Since the longest power line possible in an airplane is about 200 feet, such a line is considered very short electrically and it is possible to disregard the reactive effects. At the radar frequency of 10,000 mc, however, this electrical distance is only 2 inches and at these frequencies you must consider the length of the conductors from one vacuum tube to another as electrically long. Whenever the line is electrically long, it is considered an *RF line*, and as such there are special considerations in design and adjustment.

FUNCTION

In radar equipment, RF lines serve to transfer power from the transmitter to the antenna and from the antenna to the receiver. Since the RF line displays reactive characteristics, short sections of RF lines are used as tuned oscillator circuits, as inductors or capacitors, and as delay networks. The capacity in an RF line can be charged, and upon discharging, produces a square wave rather than an exponential one. The RF line is also used to produce a square radar transmitter pulse.

RF LINE THEORY

By resolving an RF line into a simple equivalent circuit and analyzing this circuit, you will be able to understand the characteristics of the RF line. This equivalent circuit is made by lumping all the resistances into a single large resistor, all the inductances into a single large inductor, and so on. Such a circuit is an RF line with lumped constants. To evaluate the effect of the RF line shown at A on the next page, it is necessary to put only the constants of a small section of a line into the equivalent circuit and to multiply the effect of this circuit by the number of these sections.



Equivalent Circuit of RF Transmission Line

The first constant to consider is the DC resistance of the conductors. In the equivalent circuit, this constant is a series connected element. It is represented by R in each line of the equivalent circuit shown at E above.

The second constant is inductance. Previously, you learned that whenever electrons move in a wire, a magnetic field is produced about the wire and, because of self-induction, retards the acceleration of the electrons as it builds up. In an RF line there is a similar self-inductive action. This action is represented in the equivalent circuit by inductors in each line, as shown at B and E.

Another constant to consider is capacitance. To understand the effect of capacitance in RF lines, recall that the term voltage difference means that an excessive number of electrons exist at a certain point in an electrical circuit. In an RF line, voltage difference means that there are more electrons on one conductor than on the other. The point of excess electrons exerts a force through the dielectric material between the two conductors and pushes the electrons away from this point as indicated at C. This action is identical to the action that takes place between the two plates of a capacitor, and indicates that there is a capacity between the two conductors. The parallel capacitor C in the equivalent circuit at E represents the capacity of the RF line.

The fourth constant in an RF line is the leakage which takes place from one wire to the other through the insulation between the wires. There is no such thing as a perfect insulator. Any insulator, therefore, will act as high resistance rather than an open circuit. The insulation conductance between wires is represented as G and is shown at D. In this case note that with G out of the equivalent circuit, the current on the meter (M_1) at the source will read the same as the meter (M_2) at the load. When you replace the shunt conductance (G) in the circuit, more current will flow at the source. Thus, you see that without changing the load resistance or the source voltage, you can increase the current.

This completes the equivalent circuit and it appears as shown at E. However, it will not perform exactly like the actual line itself. An actual line is uniformly constructed so that energy going either way will encounter the same effects. This is not the case in the equivalent circuit, for if you look into the line shown at E from the left end, you see two R's and one C connected series, and if you look into the right end, you see two L's and one G in series. This is not the same circuit.

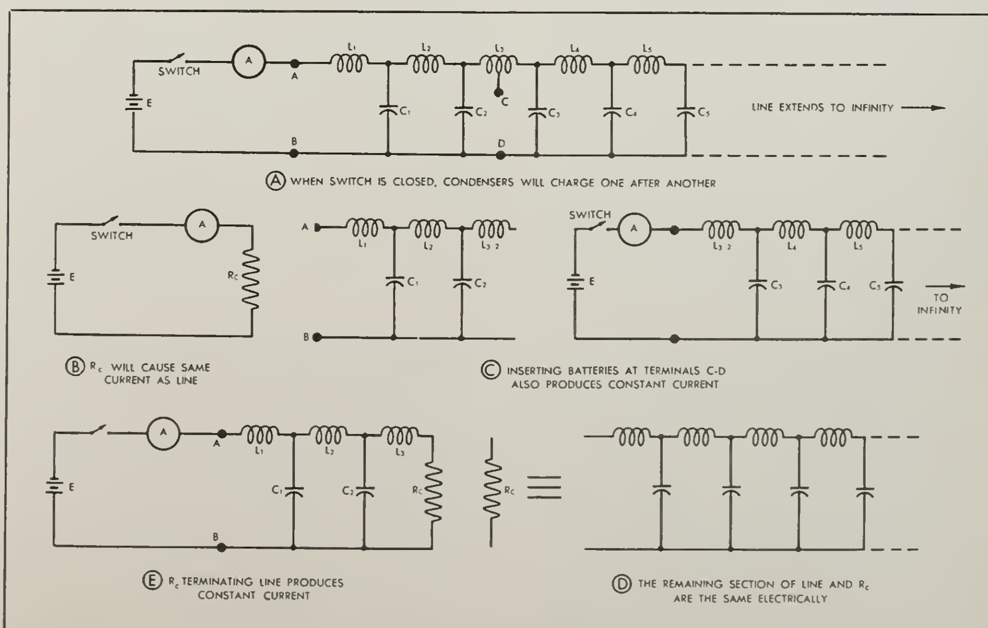
To balance and simplify the circuit, it is changed as you can see at F so that all the series resistance is put into the top line, where the electrical effect is the same. The series resistance is then divided into halves and each half is located as shown at F. The same is done with the inductance. The net result is that the circuit looks the same from either end.

You can further reduce the circuit to the simple impedance network as shown at G. In this case all the series impedances (represented by Z_1) are symmetrically spaced about the parallel impedance (represented by Z_2). When the circuit is reduced to this form, it is possible to develop certain general equations in terms of Z_1 and Z_2. These equations, which are taken up later, apply to any circuit regardless of the resistance, the inductance, or capacity in Z_1 and Z_2.

Another diagram (H) shows another form into which the RF line constants can be lumped. In this circuit the shunt constants are symmetrically placed about the series constants. This is known as a Pi section and is shown in terms of impedances as shown at I.

The Characteristic Impedance of a Lossless Line

In an RF line there is an inherent impedance which is constant regardless of the length or the applied voltage. In the equivalent circuit shown below at A, the RF line is assumed to be without losses. A point to remember here is that inductance or capacity will store energy, then return it to a circuit, while resistance will dissipate energy in the form of heat. Since no resistors



Developing Characteristic Impedance

appear in the circuit at A, no energy will be lost. A line can be constructed with extremely small losses, which approximates the ideal situation shown. Such a line is assumed to be infinitely long. Although this, too, is a theoretical situation, cross-country telephone lines approach the condition described.

Notice what happens when you apply a DC voltage to this circuit. At the time you close the switch, the full voltage E will appear across the line terminals A-B. This end is referred to as the *sending end*. The end to which energy travels is called the *receiving end*. The first capacitor C_1 starts to charge. However, its charge is not instantaneous, for the current increase to the value required is impeded by the inductance of L_1. After a little time, however, C_1 will charge, and when the voltage across it nearly reaches the applied voltage, C_2 will start to charge through L_2, again taking some time. Meanwhile, the ammeter A in the circuit will indicate the current flow. A current will flow through the inductors to charge the next capacitor. Since the line is infinitely long, there will always be more capacitors to be charged and current will not stop flowing until they are all charged. Thus, current will flow indefinitely in the line.

Another circuit which will display identical characteristics is shown at B at bottom of page 9-3. The resistor R in this circuit will cause a certain current to flow indefinitely when voltage E is applied. (To determine the value of R, use the Ohm's law formula, $R = E/I$.) Returning to the RF line, since a constant current I is indicated by the meter and the voltage E is known, the RF line displays a resistance equal to the value E/I. This resistance is the characteristic resistance of the RF line, and may be called R_c. Mathematically, it is equal to,

$$R_c = \frac{E}{I}$$

If you break the line at the center of L_3 and connect the battery to terminals C-D as shown at C, again the same current would flow, for the same constants are displayed and the receiving section is still infinitely long.

If a resistance equal to the characteristic resistance of the section removed is placed across receiving end terminals of the short section and the battery again connected across terminals A-B as shown at E, the same current will flow again. Therefore, whether the line is infinitely long or fixed in length and terminated in a resistance equal to its characteristic resistance, the effect at the battery end is the same. The display of constant resistance that you see at the sending end is equally important, for it is this resistance that must be matched by the generator for maximum power transfer. When AC voltages are applied to the line, the characteristics of the line might be resistive with a reactive component—either capacitive or inductive. Therefore, the characteristic of a line generally is described as the characteristic impedance, for which the symbol is Z_0.

Voltage Changes Moving Along a Line

Previously, it was stated that a voltage suddenly applied will appear at a point some distance from the sending end at some later time. This is true regardless of whether the voltage change is a jump from zero to some value, a drop from some value to zero, or for that matter, any other change in voltage. Whatever the change, the voltage will be conducted down the line at a constant rate.

If you know the inductance and capacity of an RF line, you can easily compute the time required by any waveform to traverse the length of the line. The circuit at A on the next page is the equivalent circuit of a line which is reduced to T sections and is assumed to be lossless. The inductance per section is represented by L; C is the capacitance per section. In this connection recall a few fundamental relationships, such as $Q = IT$, which means that the quantity of electricity in coulombs is equal to the current multiplied by the time the current flows, and $Q = CE$, which means that the quantity of electricity in a capacitor is equal to the capacity times the voltage across it. In the case of the transmission line shown at A, the (Q) of electricity leaving the battery is also shown by the equation,

$$Q = IT \qquad (1)$$

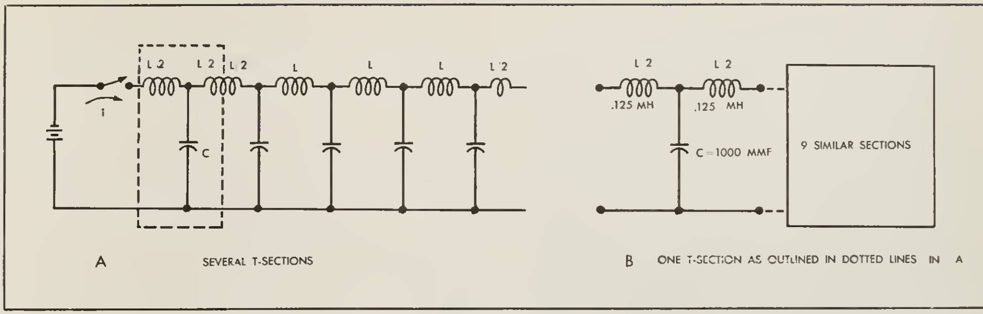
where I is the current and T the time the current flowed.

The electricity which leaves the battery goes into the line, where it is deposited in the capacitors as a charge (Q). This charge is shown by the equation,

$$Q = CE \qquad (2)$$

where C is the capacity per section and E the voltage applied.

The quantity of electricity which left the battery is equal to the quantity in the line; that is,



Circuit for Calculating Time for Wove to Travel Down the Line.

Q=Q. Therefore, by equating (1) and (2),

$$IT = CE \qquad (3)$$

As each condenser charges to CE, the voltage at the end of the inductor must change by E with respect to the other end. This occurs because the first condenser charges to E voltage, while the second condenser is still at zero voltage. This makes the voltage E appear across the inductor. In an inductor these circuit components are related mathematically as,

$$E = L\frac{dI}{dT} \qquad (4)$$

The voltage across the inductor is directly proportional to the inductance and the change in current, but increases as the time becomes shorter. This is expressed mathematically by the equation,

$$EdT = LdI \qquad (5)$$

which is the result of multiplying the preceding equation by dT.

Since current and time start from zero, the change in time (dT) and the change in current (dI) are equal to the final time T and final current I. For this case, equation (5) becomes,

$$ET = LI \qquad (6)$$

If voltage E is applied for time T across inductor L, the final current I will flow. The following equations contain all three terms (T, L, and C).

$$IT = CE \qquad (3)$$

$$ET = LI \qquad (6)$$

Another way of using them is to get an equation for T in terms of L and C. This is done in the following manner: Multiply the left and right members of each equation as follows

$$(IT)\ (ET) = (CE)\ (LI) \qquad (7)$$

$$EIT^2 = LCEI$$

Cancelling the EI terms, $T^2 = LC$ (8)

and $T = \sqrt{LC}$ (9)

Equation (9) is the equation for the time required for a voltage change to travel along a section composed of L and C.

Since the characteristic impedance is resistive and equal to R=E/I, it is equal to the voltage-to-current ratio. As E and I appear in the preceding equations, these equations can be solved for the E/I ratio. This is done as follows:

Divide equation (6) by (3): $\frac{ET}{IT} = \frac{LI}{CE}$ (10)

Multiply both sides by $\frac{E}{I}$: $\frac{E^2T}{I^2T} = \frac{L}{C}$ (11)

Cancel the T's: $\frac{E^2}{I^2} = \frac{L}{C}$ (12)

Take square root: $\frac{E}{I} = \sqrt{\frac{L}{C}}$ (13)

Since E/I is the Z_0 *of the line,* $Z_0 = \sqrt{\frac{L}{C}}$ (14)

Example

Problem. Assume that the line illustrated above at A is 1000 feet long and a hundred foot section is measured and found to have an inductance of ¼ millihenry and a capacity of 1000 mmf as shown at B. Find the characteristic impedance of the line and the time required by a voltage to travel the length of one section.

Solution:

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{.25 \times 10^{-3}}{10^{-9}}} = \sqrt{.25 \times 10^{6}} = .5 \times 10^{3}$$

$Z_0 = 500$ *ohms*

The time required by a voltage to travel the length of one section is equal to,

$$T = \sqrt{LC} = \sqrt{.25 \times 10^{-3} \times 10^{-9}} = \sqrt{.25 \times 10^{-12}}$$
$$= .5 \times 10^{-6}\ sec.$$

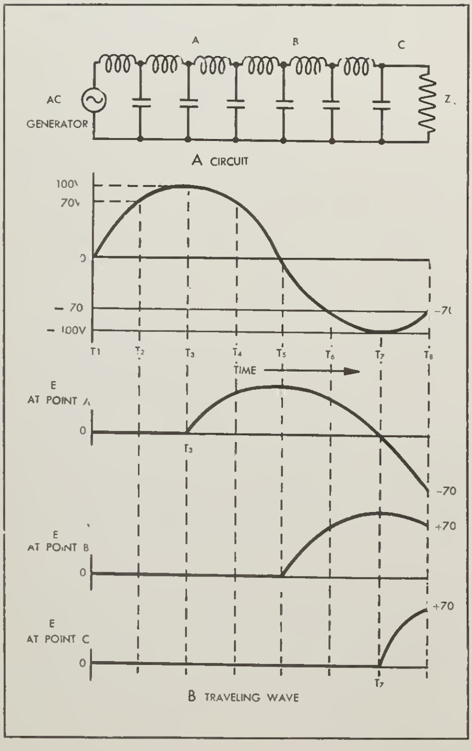
$T = .5$ *microseconds*

The time required for voltage to travel over ten sections (1000 feet) is equal to .5 × 10 = 5 microseconds.

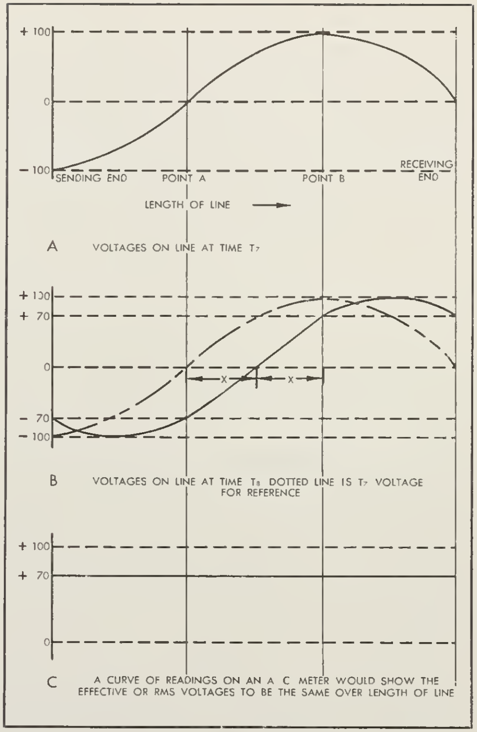
This example was based on a DC change. An RF line displays similar characteristics when an AC voltage is applied to its sending end terminals. Assume that an AC voltage is applied to the line simulated by the circuit shown below at A. The generator voltage starts from zero, and produces the voltage indicated. As soon as a small voltage change is produced, it starts its journey down the line, and the generator in the meantime continues to produce new voltages along a sine curve. At time T_2, the generator voltage is 70 volts. The voltages still move along the line. Finally, at time T_3, the first small change arrives at point A, and the voltage at that point starts increasing. At time T_5 the same voltage arrives at point B on the line and finally at time T_7 the first small change arrives at the receiving end of the line. Meanwhile all the changes in the sine wave pass each point in turn. The time required for the changes to travel the length of the line is the same time required for a DC voltage to travel the distance.

At time T_7, the voltage at the various points on the line are the following:

At generator:	*−100v*
At point A:	*0v*
At point B:	*+100v*
At point C:	*0v*



Sinusoidal Voltage Travels Down Line



Instantaneous Voltages are Different Along Length of Line Voltages for moving Sine Wave

If you plot these voltages along the length of the line, the resulting curve is like the curve shown above at A. Note that such a curve of instantaneous voltages resembles a sine wave.

At time T_8 in the illustration at the left, the voltages line up in the following manner:

At generator:	*Drop from −100 to −70v*
At point A:	*Rise from zero to −70v*
At point B:	*Drop from 100 to 70v*
At point C:	*Rise from zero to 70v*

A plot of these voltages produces the solid curve shown in diagram B above. For reference, the curve from time T_7 is drawn as a dotted line. The solid curve has exactly the same shape as the dotted curve, but it has moved bodily down the line by the distance X. Another plot at T_9 would show a new curve similar to the one at T_8 but moved to the right by the distance X.

Summary

By analyzing the points just discussed, you will find that the action associated with voltage changes along an RF line are as follows:

1. All parts of the sine wave produced by the generator travel down the line in the order produced.

2. At any point, you will obtain a sine wave if you plot all the instantaneous voltages passing the point during 8 of the time intervals previously shown. The oscilloscope will plot all values instantaneously.

3. The oscilloscope pictures (instantaneous voltages) will be the same in all cases except that a phase difference will exist in each picture taken down the line. The phase will change continually with respect to the generator until over a certain length of line the change will be 360°.

4. Since all parts of a sine wave pass every point along the line, a plot of the readings of an AC meter (which reads the magnitude of the voltage over a period of time) will show that the voltage is constant at all points as shown at C in the last illustration.

5. Since the line is terminated with Z_0, the energy arriving at the end is absorbed by the resistance.

The distance traveled by the start of one cycle, during the time required to generate the entire cycle, is one wavelength. In the next illustration the voltage arrives at the receiving end at the same time a similar voltage is generated in the next cycle. This line is exactly one electrical wavelength long, and contains a full wave of voltage at one time.

The electrical length of a line is a function of the frequency of applied voltage and the velocity of energy travel. The time required for travel per section is equal to,

$$T = \sqrt{LC} \qquad (9)$$

Velocity is the distance per unit of time and distance D is the length of one section, thus,

$$V_g = \frac{D}{T} \qquad (15)$$

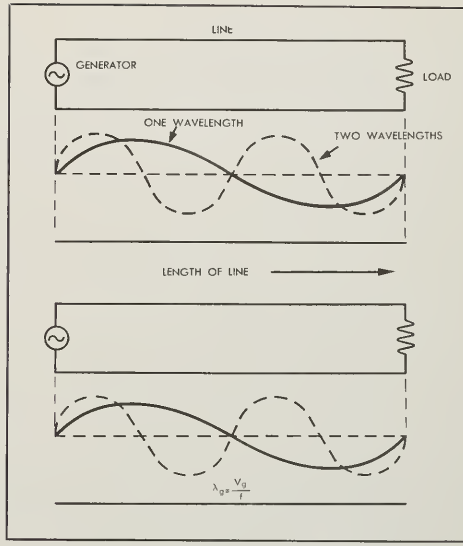
The subscript of g indicates velocity of waves on a line as distinguished from velocity of waves in space. In terms of L and C, equation (15) may be expressed,

$$V_g = \frac{D}{\sqrt{LC}} \qquad (16)$$

If, for example, $\sqrt{LC}$, or T, is .5 microsecond, and the section 100 feet long, the velocity of the wavefront, V_g, is found as follows:

$$V_g = \frac{D}{\sqrt{LC}} = \frac{100}{\sqrt{.25 \times 10^{-3} \times 10^{-9}}} = \frac{100}{.5 \times 10^{-6}}$$

$$V_g = \frac{100 \times 10^6}{5}$$



Meaning of Wave Length

$V_g = 200 \times 10^6$ or 200 million feet per second or approximately 39,000 miles per second.

The wavelength (of the group) of voltages on the line is found by the equation,

$$\lambda g = \frac{V_g}{f} \qquad (17)$$

where f is the frequency in cycles per second λg is wavelength in feet if V_g is velocity in feet per second.

Example

Problem. Assuming that the generator frequency in the preceding example is 200 kc, find λg.

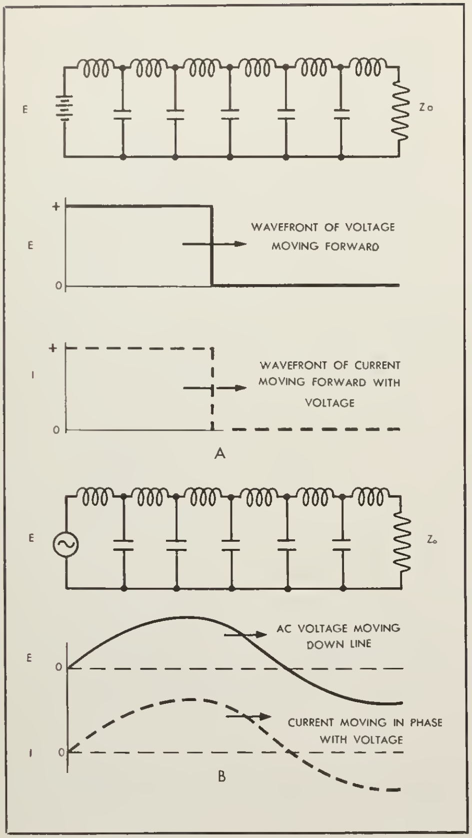
Solution:

$$\lambda_g = \frac{V_g}{f} = \frac{200 \times 10^6}{200 \times 10^3} = \frac{2 \times 10^8}{2 \times 10^5} = 10^3$$

$$\lambda_g = 1000 \text{ feet}$$

One wavelength will occupy one thousand feet of the line. Notice that if you double the frequency, the wavelength is only one-half of the former value. A 1000-foot line will have two full cycles of voltage on it at any one time. This is shown by dotted line in the illustration above.

There has been little said about the current changes in the RF lines thus far. The RF lines described were all assumed to display a resistance to the current, but no reactance to it, and the termination to have the characteristic impedance of the line. Under these conditions, the



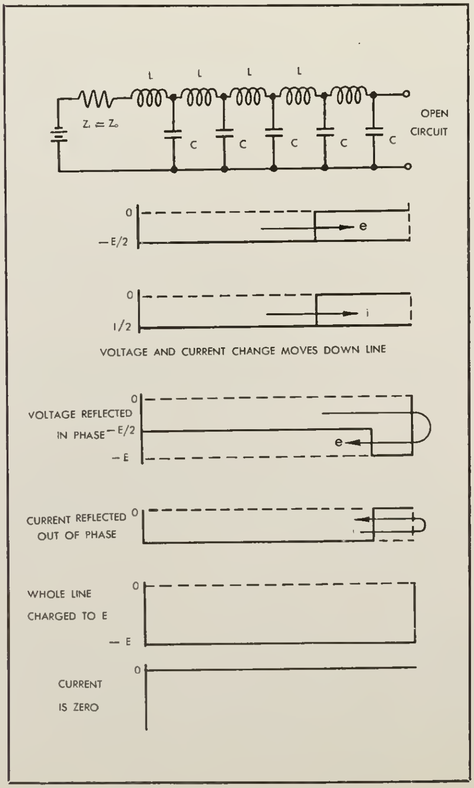
Current Changes Move in Step with Voltage whether Change is a Sudden One or a Sine Wave

current is always directly proportional to the voltage, and whenever a voltage appears across a point on the line, a current equal to E/Z_0 will immediately flow at that point as shown above. As a DC voltage change moves along the line, a current change moves with it. As an AC voltage moves along the line, a current exactly in phase with it is present at all points.

REFLECTION

All the RF lines previously analyzed were assumed to be terminated in a manner that the terminating resistor absorbed all the energy. However, RF lines are not always terminated in this manner. In some cases you will find the receiving end open-circuited or shorted.

Take the case of the line which is open-circuited at the receiving end and which has DC voltage applied at the sending end. As the switch in the equivalent line with an open-circuited receiving end shown below is closed at the sending end, the negative voltage moves down the line, charging each condenser through the preceding inductor. When the voltage change reaches the last capacitor, and it is charged, there is no longer any place for the voltage change to go. The capacitor at each end of the inductor is charged to 1/2E (since $Z_i = Z_o$, one-half the voltage drop is across Z_i and one-half across Z_o). As there is no voltage across the inductor to maintain the magnetic field, the field collapses and causes current to flow into the last capacitor. Furthermore, as the direction of current has not changed, the condenser charges in the same direction.



Reflection from Open Ended Line

Since the energy in the magnetic field described equals the energy in the capacitors, the transfer to the condenser is just enough to double the voltage across the last capacitor. The current in the inductor drops to zero. Since there is no place for current to flow from the next to the last inductor, its field collapses, doubling the voltage across the next to the last capacitor. This change of voltage (moving backward) is the same as though the voltage on arriving at the end found no place to go, started back in the same polarity. Such action is called *reflection*. A reflection of voltage occurs always in the same polarity. The voltage change moves back to the source, charging each condenser in turn until the last condenser is charged to the source voltage and the action stops.

In the meantime, the field collapsed around the coil and the current there became zero. As each condenser is charged, the current in each inductor drops to zero, effectively reflecting the changed current in opposite polarity. The opposite polarity cancels the current at each point and it drops to zero at that point. When the last condenser is charged, the current from the source stops flowing.

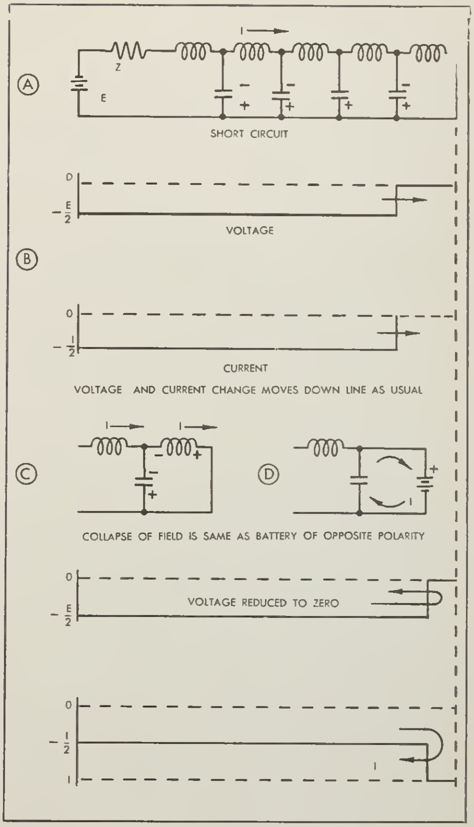
Notice that the line current flowed through the resistor called Z_1. This is the internal resistance of the source. It must equal the Z_0 of the line for maximum power transfer to the line. Because of this resistance, the voltage change delivered to the line is one-half of the source voltage. On the return trip, the additional half charges each condenser to the source voltage.

The following summarizes the principal facts in the reflection of voltages in open end lines:

1. Voltage is reflected in phase and without change in amplitude or shape from an *open* end.
2. Current is reflected in opposite phase and without change of amplitude or shape from an *open* end.

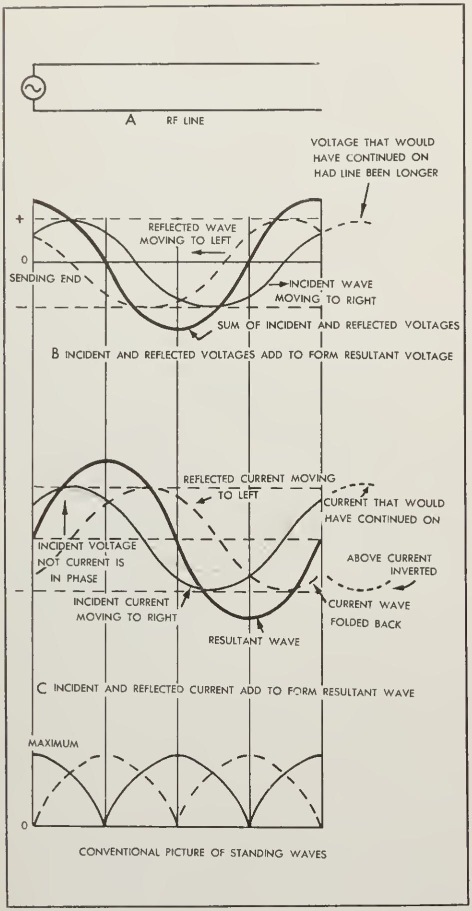
Reflection from Short Circuit

A short circuited line has a different effect on voltage change than an open circuited line. In a short-circuited line the voltage change from zero to E 2 arrives at the last inductor as usual, but in this case the current in the inductor has no condenser to charge. As you can see at C in the illustration, the current in the final inductor causes the indicated polarity to exist. When the field collapses, the inductor becomes a source and the polarity reverses. In the instantaneous equivalent circuit at D showing this action, the inductor acts like a battery and forces current through the capacitor in the opposite direction, causing it to discharge. Since the energy in the magnetic field is the same as that in the condenser, the condenser discharges exactly to zero. The second inductor current consequently has no voltage to maintain it; therefore, it will discharge the inductor which is next to the last one.



Reflection from Short Circuited Line

The reduction of voltage to zero by the inductor voltage becomes effectively a new voltage which appears at the end of the line. It is equal to E 2, but has opposite polarity. The collapsing field adds a current, which, in turn, adds to the current created by the source. In effect, the current change appears at the end of the line in the same direction as the original current.



Formation of Standing Waves

An RF line with a short circuit across it is characterized by:

1. Voltage changes which are reflected in opposite polarity but in same amplitude.

2. Current changes which are reflected in the same polarity and the same amplitude.

Reflection of AC Voltages

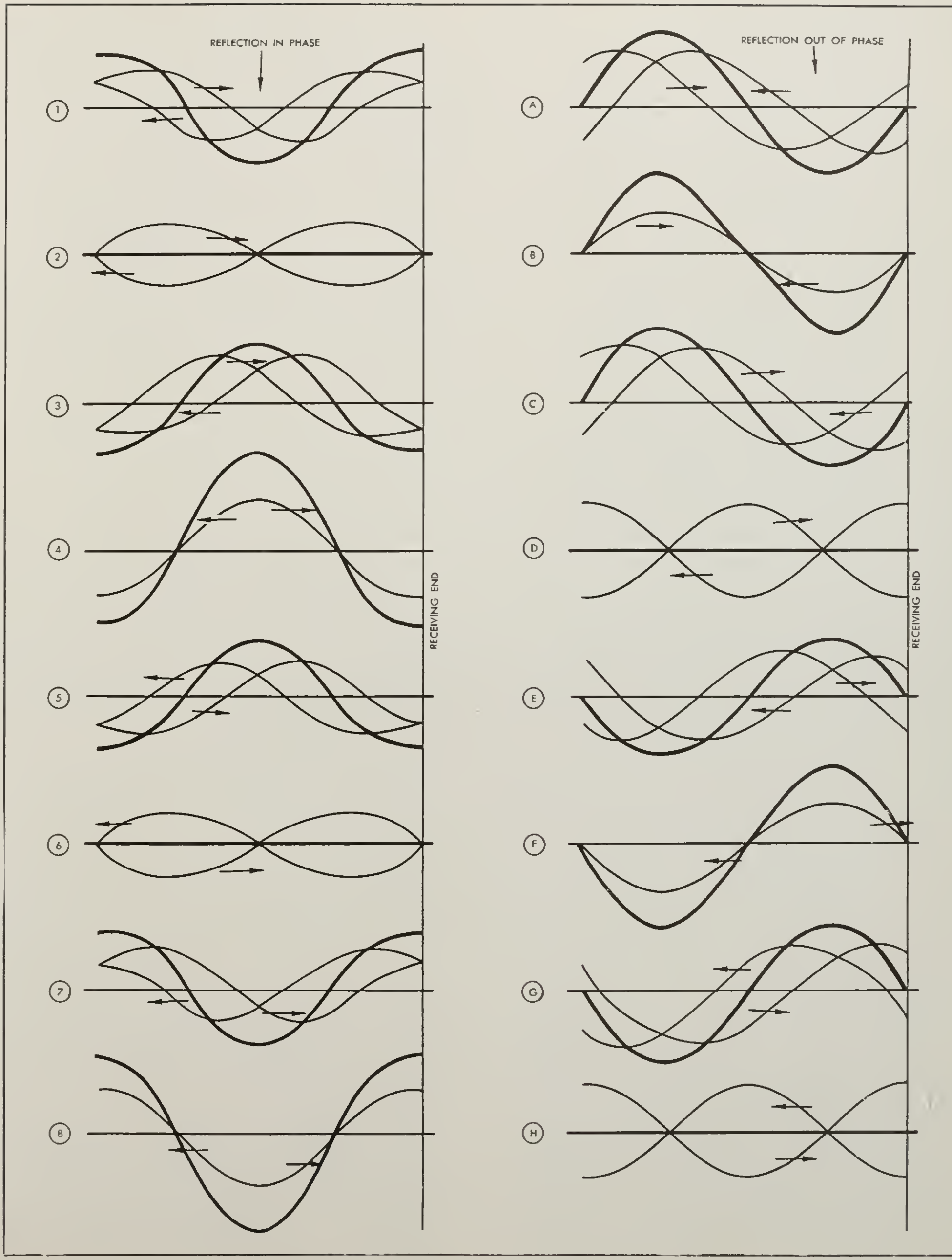
In most cases where RF lines are used, the voltages which are applied to the sending line are AC voltages. When AC voltages are applied, the action at the receiving end of the line is exactly the same as for DC. In the open ended line shown above at A the generated AC voltage is distributed along the line as shown at B in such a way that as each voltage arrives at the end, it is reflected back in the same polarity and amplitude. When AC is used, this reflection is in phase. Each of the reflected voltages travels back along the line until they reach the generator. If the generator impedance is the same as the line impedance, energy arriving at it is absorbed. There are now two voltages on the line. However, most indicating instruments are unable to indicate these voltages separately. Instead, they read their vector sum.

An oscilloscope is usually used to study the instantaneous voltages in RF lines. It is used because meters read average values over a period of time. However, you may first add the two voltages algebraically on paper as at B to see how they will appear on the oscilloscope.

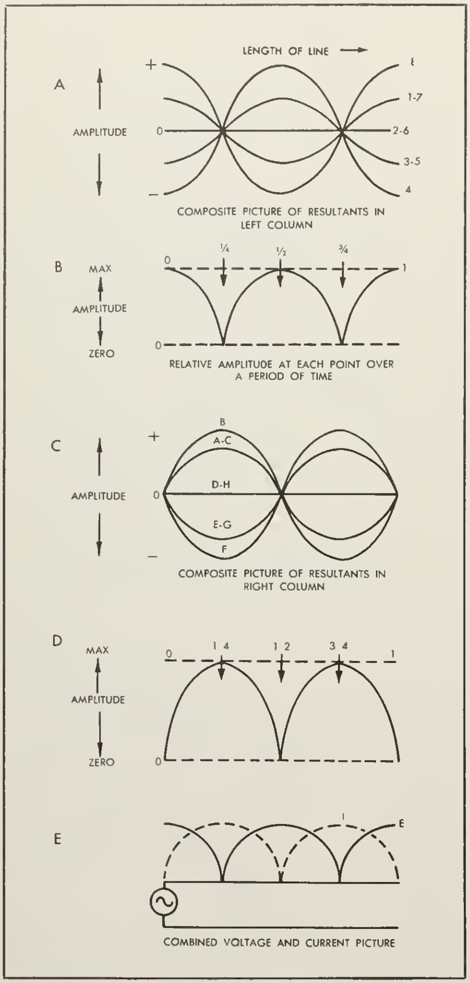
Since there are two waves of voltage moving on the line, it is desirable to distinguish between the two. This is done by giving them names. The voltages moving toward the receiving end are called *incident voltages*, and the whole waveshape is called *the incident wave.* The wave moving back to the sending end after reflection is called the *reflected wave*.

Another step in investigating the open ended RF line is to see how the current waves act. Note that the incident current wave is the solid line in the diagram at C. The voltage at the moment is represented by the dotted line. The current is in phase with the voltage while traveling toward the receiving end. Upon arrival at this end, it is reflected in opposite polarity, that is, it is shifted 180 degrees in phase. However, its amplitude remains the same. The reflected wave of current is shown in dashed lines at C. Since a current indicating device would read the sum of the two currents, you can add them to determine what the current indicator will show. The heavy-line curve represents the sum of the two currents. Notice that current is zero at the end of the line. This is reasonable since there can be no current flow through an open circuit. The voltage curve shows that the voltage is maximum at the end of the line, a condition which can easily occur across an open circuit.

Another factor to keep in mind is that when you use an AC meter to determine the current or voltage on an RF line, it will indicate only the magnitude of the current, but not the polarity. If you plot all the readings along the length of the line, you will obtain a curve like the one



Instantaneous Values of Incident and Reflected Waves



Composite Results of Instantaneous Waves

shown in D. Notice that all the readings are positive due to current rectification in the meter circuit. The curves in diagram D on page 9-10 represent the conventional method of showing current and voltage on an RF line.

The voltages and currents shown at B and C on the same page are those at a single instant. During the generation of a complete cycle by the generator, there are a large number of these pictures generated. The instantaneous incident and reflected waveshapes are shown in the illustration on page 9-11.

The pictures to the left in the illustration on page 9-11 show reflection in phase. This is the case with voltage when the line is open circuited at the end. If you check each picture individually, you will see that the heavy curve is the algebraic sum of the other two curves. In going from picture to picture, you can see that one curve moves to the right and the other moves to the left. The curve moving to the right is the incident wave of voltage; the one to the left is the reflected wave of voltage. The composite picture at A in the adjacent illustration shows all resultant curves over a complete cycle. Notice that the voltage varies between zero and maximum in both directions at the center and at both ends as well. But the voltage is always zero at intermediate points; in this case, one fourth of the distance from each end. The resultant waveshape, which is not traveling like the other waves, is referred to as a *standing wave* of voltage.

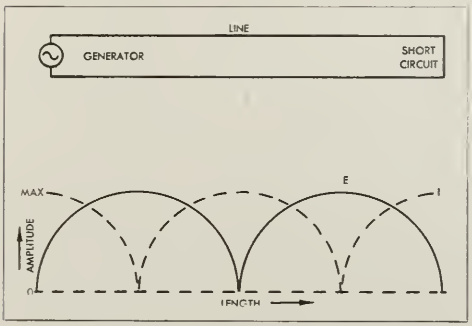
The voltage at the center and the ends varies at a sinusoidal rate between the limits shown, and is always at zero at the ¼ and ¾ points. A continuous series of pictures such as these are difficult to see with conventional test equipment, for this equipment reads the effective or average voltage over a number of cycles. The curve of amplitudes over the length of line for several cycles is shown at B in the adjacent illustration. A meter would read zero at the points shown and show a maximum voltage at the center no matter how many cycles went by.

The current waveshapes on the open ended line are shown in the right hand column in the illustration on page 9-11. Since the reflection of current is out of phase at an open end, the resultant waveshapes differ from those on the left side. The two out of phase components always cancel one another at the end, and thus the resultant is always zero at that point. (See the composite picture of resultants at C on this page.) Again the voltage is zero at certain points at all times, but maximum at other points. Therefore, it is a standing wave. As shown at D, meters would show the amplitude along the length. In this case, it is zero at the end and center, but maximum at the ¼ and ¾ points.

The entire picture of the open circuited line conditions is usually as shown at E. The standing wave of voltage and current appear together. Observe that one is maximum when the other is minimum. The current and voltage standing waves are ¼ cycle, or 90 degrees out of phase with one another.

Reflection from a Short Circuited RF Line

Before, it was stated that when an RF line is terminated in a short circuit, reflection is complete but the effect on voltage and current differs from the effect in an open circuited line. Voltage is reflected in opposite phase, while current is reflected in phase. Again refer to the series of pictures shown on page 9-11. The left column represents current since it shows reflection in phase. The right column of pictures shows the voltage changes on the shorted line. The final effect of this is the diagram just below. It shows the voltage in a solid line, in which the voltage is zero at the end and center, and maximum at the ¼ and ¾ points.



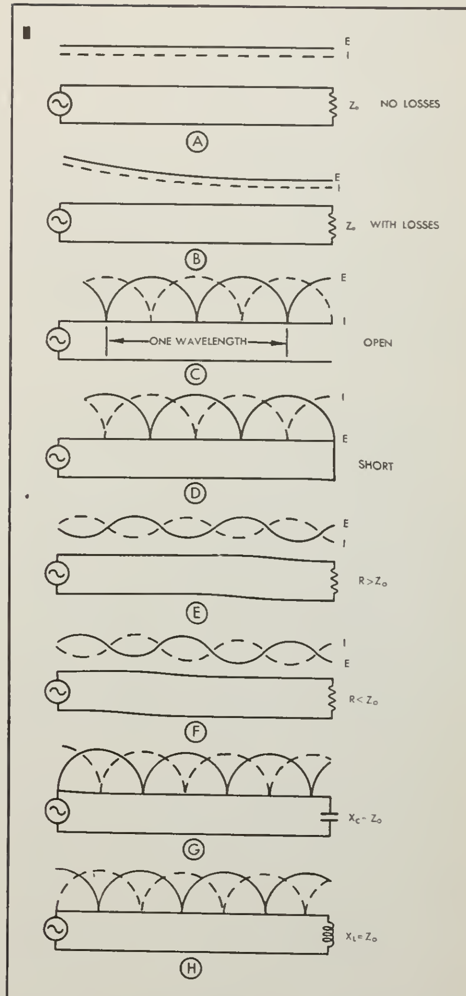
Standing Waves on Shorted Line

COMPARISON OF STANDING WAVES WITH VARIOUS TYPES OF TERMINATORS

There is a large variety of terminations for RF lines. Each type of termination has a characteristic effect on the standing waves on the line. From the nature of the standing waves, you can determine the type of termination which produces the waves.

Termination in Z_o

Termination in Z_o (characteristic impedance) will cause a constant reading on an AC meter when it is moved along the length. As illustrated at the right at A, the curve for such a condition, provided there are no losses in the line, will be a straight line. If there are losses in the line, the amplitude of the voltage and current will diminish as they move down the line. The losses are due to DC resistance in the line itself. In radar equipment, as the DC losses are often small enough to be disregarded, the assumption of no losses is not too theoretical.



Effects of Various Terminations on Standing Waves

Termination in Open Circuit

In an open circuited RF line at C above, the voltage is maximum at the end, but the current is minimum. The distance between the two zero points is ½ wavelength, and the distance between every other zero point is one wavelength. It can be said that the voltage is zero at a distance of ¼ wave from the end of the line. (This is true at any frequency.) A voltage peak occurs at the end of the line, and also ½ wavelength from the end.

Termination in Short Circuit

On the line terminated in a short circuit at D, the voltage is zero at the end and maximum at ¼ wavelength from the end.

Termination is a Resistance Larger than the Characteristic Impedance

Whenever the termination is anything but Z_o, there are reflections. For example, if the terminating element contains resistance, it will absorb some energy, but if the resistive element does not equal the Z_o of the line as shown at E, some (not all) of the energy will be reflected. The amount of voltage reflected may be found by the equation,

$$E_R = E_i \frac{R_L - Z_o}{R_L + Z_o}$$

Where E_R is reflected voltage, E_i is incident voltage, R_L is terminating resistor and Z_o the characteristic impedance of line.

If you try different values in the preceding equation, you will find that the reflected voltage is always less than the incident voltage except when R_L is zero or infinitely high, or when E_R and E_i are equal. A smaller value of reflected voltage causes a standing wave of smaller variation. A large resistance, when compared to Z_o makes the end of the line somwhat like an open circuit. Thus, the standing wave of voltage will be greatest at the end of the line, while the current will be at a minimum (but not zero).

Termination with R_L Smaller Than Z_o

When R_L is less than Z_o as at F, the line is terminated approximately as a short circuit. In this case the voltage is minimum at the end of the line (this fact may be verified by the preceding equation), and the value of E_R becomes negative, subtracting from E_i at the end.

You probably have noticed that the variation of standing waves will show how near the RF line is to being terminated in Z_o. A wide variation in voltage along the length means a termination far from Z_o. A small variation means termination near Z_o. Thus, the ratio of the maximum to minimum standing wave voltage is a measure of the perfection of the termination of a line. It is called *standing wave ratio* and is always expressed in whole numbers. For example, a ratio of 1:1 describes a line terminated in its characteristic impedance (Z_o).

Terminating the Line in Capacity

When a line is terminated in capacity, the capacitor does not absorb energy, but returns it all to the circuit. This means 100% reflection. The current and voltage relationships are somewhat more involved than in previous types of termination. For purposes of understanding, assume that the capacitive reactance is equal to the Z_o of the line. The current and voltage will be in phase upon arriving at the end, but the current in flowing through the capacitor and Z_o connected in series with it, is shifted 45° in phase relationship. The reflected current is thus shifted 45° and the reflected voltage shifted 45° in the opposite direction. Thus, they arrive in phase and leave 90° out of phase, due to the 45° shift introduced by the capacitor. This results in the standing wave configuration shown at G, in which the standing wave of voltage is minimum at a distance of exactly one-eighth wavelength from the end. If the capacitive reactance is greater than Z_o (smaller capacity), the termination looks more like an open circuit, the voltage goes up and the voltage minimum moves away from the end. If the capacitive reactance is smaller voltage, the minimum moves toward the end.

Termination in Inductance

When the line is terminated in inductance, both the current and voltage are shifted the other way in phase upon arriving at the end of the line. When X_L is equal to Z_o, the shift is exactly 45° and the resulting standing waves are as shown at H. The current minimum is located ⅛ of a wavelength from the end of the line. When the inductive reactance (and inductance) is increased, the open circuit situation is approached and all standing waves appear closer to the end. When inductive reactance is decreased, the standing waves move away from the end of the line.

SENDING END IMPEDANCE OF LINES OF VARIOUS LENGTHS AND TERMINATIONS

The impedance at the generator sending end varies widely from the characteristic impedance when a line is terminated in something other than Z_o, as was shown by the previous illustration in which a line was not terminated in its characteristic impedance. An example of a line of random length terminated in a short circuit is shown at A on the next page. In it, the voltage is low and the current is high at the end of the line. This is the correct voltage to current ratio for low resistance. Notice at a point ½ of a wavelength from the end of the line that not only the voltage has shifted a half cycle or a half wavelength, but that the current also has shifted one-half cycle. The magnitude of current and voltage at this point is the same as their magnitude at the shorted end. This means that the voltage to current ratio is the same as that of a short circuit.

In observing the situation at a point one quarter of a wavelength from the end of the line, you can see that there the voltage is maximum and the current is zero. Since the voltage-current ratio is infinitely high, the impedance is likewise infinitely high. The impedance at this point is the same as that for an open circuit. If the line is cut at this point, the generator connected to it would work into a virtual open circuit.

Eighth-Wave Section

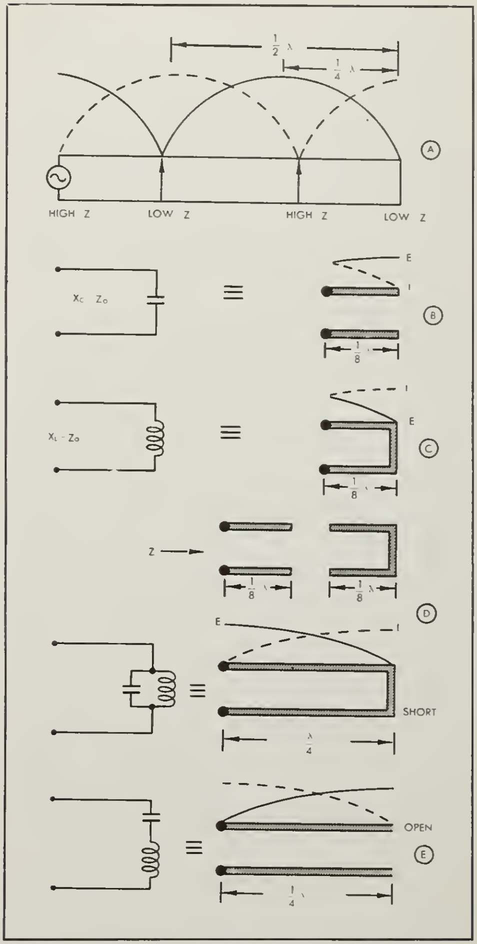
When a section of RF line ⅛ wavelength long, which is open at the end, is connected to a generator, the section displays capacity. For example, notice the open end ⅛ wave section illustrated at B to the right. Since it is an open circuit, there is very little current on the line or very little magnetic field about it. Yet an electrostatic field of any size can develop about it. The predominant characteristic of the short section is capacity. Looking at it another way, this means that if the current is shifted 90° in phase over ¼ of a wavelength, it will likewise shift 45° over ⅛ of a wavelength. As this section is ⅛ of a wavelength long, the shift is 45° in which the current leads the voltage. This action is similar to that caused by a condenser which causes a phase shift of 45°. When a condenser is used, the capacitive reactance must equal the resistance of the line. In the case of the RF line in which $X_c = X_o$, the ⅛ wave section displays exactly the same characteristics as a line having a capacitor in which $X_c = Z_o$.

When the section is shorted at the end, current then predominates, the voltage is mostly shorted out, and the input appears inductive. From the viewpoint of standing waves, a shift of 45° has occurred since leaving the end with the current now lagging the voltage. This can only occur in an inductor when the inductive reactance is equal to Z_o and thus the input to the shorted section is $X_L = Z_o$, as you can see at C.

Notice that the standing wave configuration always starts from the receiving end. This is due to the fact that reflections occur there. Standing wave patterns start at the receiving end, regardless of the length of the line.

Quarter-Wave Section

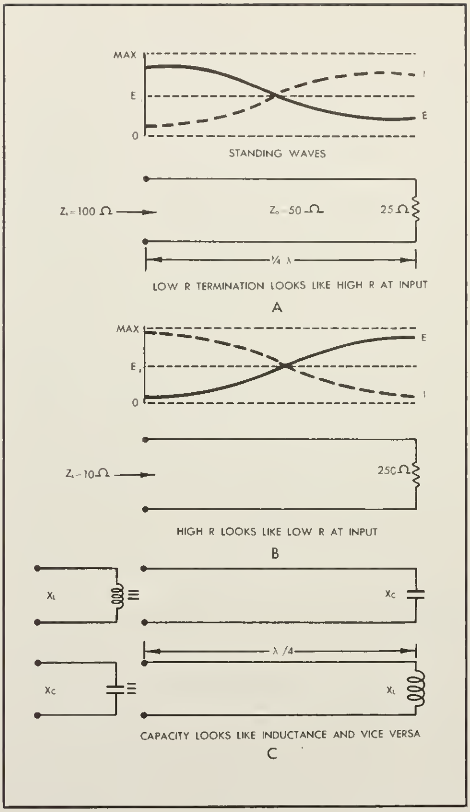
In the section ¼ wave long, shorted at the end, shown at D, notice that the voltage is high and the current is zero at the sending end terminals.



Sending End Impedance of Various Lengths and Terminations

This means that the section has a very high input impedance and that it compares with a resistor of several megohms. Or, from another viewpoint, it means that the quarter wave section is the sum of an ⅛ wave section and an ⅛ wave shorted section. When you connect an inductor and capacitor with the same reactance in parallel, there results a parallel resonant circuit. This parallel resonant circuit is characterized by a high input impedance. It is also capacitively reactive to frequencies higher than resonance and inductive to frequencies lower than resonance. The ¼ wave line has the same characteristics. At higher frequencies (shorter wavelengths) a ¼ wave section of a given physical length will be longer than a quarter wave, that is, when the open ended section is longer than the ⅛ wave shorted section—it displays capacity. At lower frequencies, when it is physically short for a quarter wavelength and it (the capacitive section) is not long enough to balance the shorted section electrically, it is inductive. There are many uses of the quarter wave section. High resistance and resonant circuits in radar equipment commonly use the quarter wave section.

When the quarter wave section is open at the end, the voltage at the end is high and the current is low, as you can see at E on page 9-15. At the sending end due to standing wave phenomena, the voltage-current ratio is completely inverted. At the sending end, the current is maximum and the voltage is zero. In fact, the quarter wave section will always invert any impedance at its receiving end. In this case, the low voltage with high current indicates low impedance. Since the section is reactive to frequencies at which it is not a quarter wavelength, it is electrically similar to a parallel resonant circuit which has zero impedance (theoretically) at resonance. The open circuit at the end of this line has been inverted to a short circuit of zero impedance.



Impedance Inversion with Quarter Wave Section

Impedance inversion. There are other terminations for the quarter wave sections. When a quarter-wave section is terminated in a resistance greater than Z_0, the section will invert it to resemble a resistance less than Z_0. Mathematically, the inversion is expressed by the equation,

$$\frac{Z_R}{Z_o}=\frac{Z_o}{Z_s} \qquad (19)$$

where Z_R is the impedance at receiving end, Z_s the impedance at sending end, and Z_0 the characteristic impedance of line.

For example, the impedance Z_s at the sending end of a 50-ohm line, terminated with a 25 ohm resistor, is equal to,

$$Z_s=\frac{Z_o^2}{Z_R}=\frac{50^2}{25}=\frac{2500}{25}=100 \text{ ohms}$$

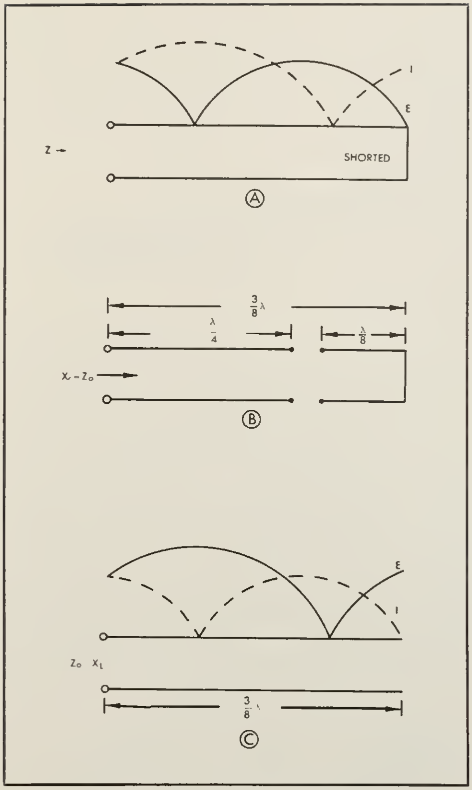
From equation (18) shown on page 9-14, you can calculate the reflected voltage from this termination to be ⅓ of the incident voltage with a negative sign. In other words, the standing wave of voltage at the end is equal to ⅔ of the incident voltage. With this information you can determine the standing wave pattern to be as shown at A above. At the sending end, the voltage is higher, but the current ratio at the sending end terminals is 100 (ohms).

If the terminating resistance is greater than Z_0, the input impedance will be lower than Z_0. For example, if the same line were terminated with 250 ohms, its input impedance would be equal to,

$$Z_s=\frac{Z_o^2}{Z_R}=\frac{50^2}{250}=\frac{2500}{250}=10 \text{ ohms}$$

When the 250 ohms is inverted, it will look like 10 ohms, as you can see in the diagram at B above.

Effectively the quarter wave section is an impedance matching transformer at *one frequency*. This inverting property also holds for reactance. If the section is terminated in capacitive reactance, the input will be inductively reactive. If inductive reactance is at the end, the input will display all the characteristics of a capacitor, as you can see in diagram C.



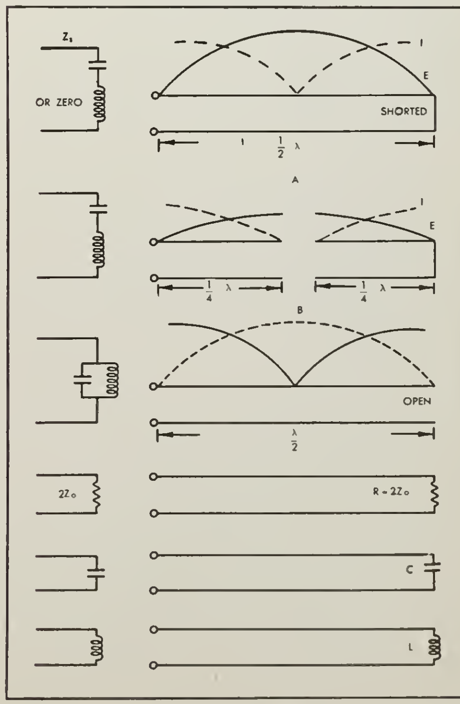
Three-Eighth-Wave Length Section

Three-Eighth-Wave Sections

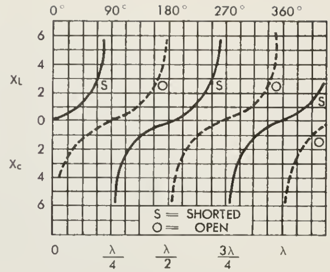
When the RF line is exactly 3/8 wave long, you can determine the input impedance as easily as you can with a quarter wave section. For example, consider the shorted section shown above at A. This can be reduced, for simplification, to two sections as shown at B. Here, a 1/8 wave section is connected to a 1/4 wave section. The 1/8 wave shorted section displays inductive reactance to the 1/4 wave section. Since 1/4 wave section inverts it to look like capacity, the 3/8 wave section is capacitive. The capacitive reactance is equal to Z_0. If the end is open as at C, it can be shown in the same manner to be inductively reactive at the input terminals. Reference to the standing waves will verify this reasoning, for at the sending end you see the same situation as at the end of a line terminated in reactance —that is, the standing waves are between minimum and maximum.

Half-Wave Section

When the section is a half wave long, the situation resembles that of two quarter wave sections connected together. The first 1/4 wave section inverts the terminal impedance and the second inverts it again as shown at B below, making the input impedance the same as the terminating impedance as shown at A. The input impedance to a shorted half-wave section is zero impedance because the voltage is zero and the current maximum. In addition, the properties of the quarter wave resonant circuit also are inverted. For example, in diagram B, the shorted 1/4 wave section is like a parallel resonant circuit. It is inverted, and looks like a series resonant circuit at the input to the half wave section. In a like manner, the 1/2 wave section, open at the end, has a high impedance at the input terminals and displays reactance to frequencies at which it is not a half wave, so it is like a parallel resonant circuit. When the half wave section is terminated in resistance (other than Z_0), in capacity, or in inductance, it will always repeat this impedance at the sending end.



Half-Wave Section



Graph for Approximating Reactance

Lengths of line which are not exact multiples of those discussed will display reactance if terminated in short circuit or open circuit, but by varying the length and termination, they can be made to display resistance. If the $\frac{1}{8}$ wave section is shortened, it becomes less than an eighth wave length, and it will still be inductively reactive, but by an amount less than Z_0. If it is longer than an $\frac{1}{8}$ wavelength, it will display an inductive reactance greater than Z_0. The above graph will aid you to approximate the reactance of odd lengths of line.

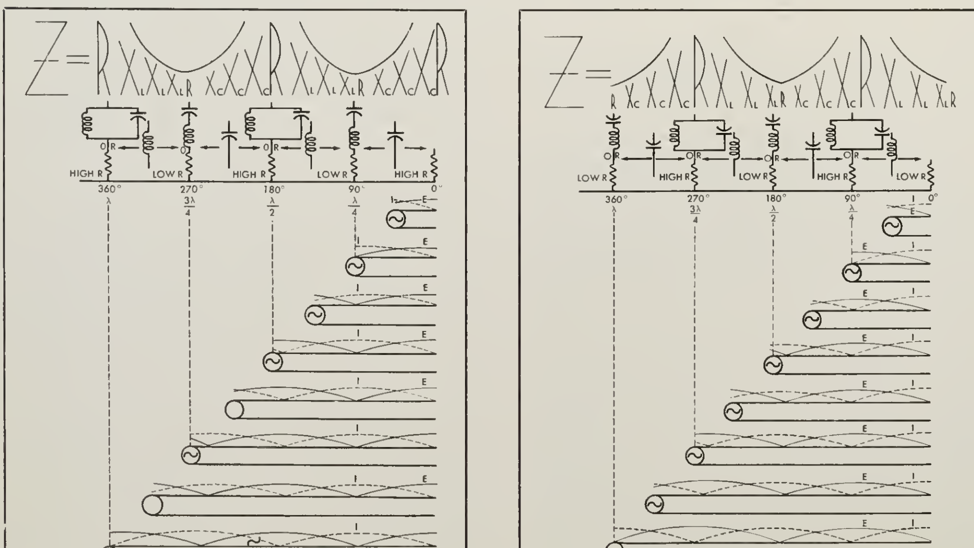
Summary

The following summarizes the characteristics of the sending end impedances of lines of various lengths and terminations:

1. $\frac{1}{8}$ wave sections display a reactance equal to Z_0 when shorted or open.
2. $\frac{1}{4}$ wave sections invert the impedance at the receiving end.
3. $\frac{1}{4}$ wave sections act as resonant circuits.
4. $\frac{1}{2}$ wave sections repeat the impedance at the receiving end.
5. $\frac{1}{2}$ wave sections act as resonant circuits.

Impedance Chart for Open and Shorted Lines

The following charts show the impedance for a number of lengths of lines.



Impedance Charts

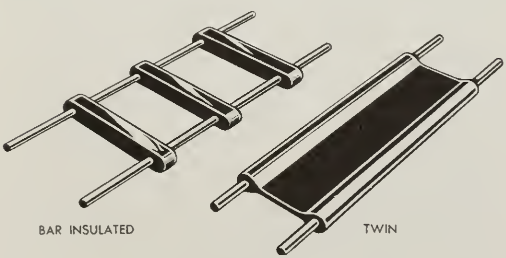
TYPES OF RF LINES

The principal types of RF lines are the twisted two-wire line, the spaced two-wire line, the shielded two-wire line, the flexible coaxial cable, and the rigid coaxial cable.

Twisted Two-Wire Line

Twisted Two-Wire Line

The twisted two-wire line consists of a pair of insulated and twisted conductors. The twisting holds the lines together mechanically, and aids in balancing each conductor to the influence of nearby magnetic and electrostatic fields. Twisted wires usually have a characteristic impedance of 70 to 100 ohms. Although the twisted line is flexible, occupies little space and is simple to manufacture, it is not usable at radar frequencies because of capacity between the wires, lack of shielding, and changes in characteristics due to moisture on the line.



Spaced Two-Wire Line

Spaced Two-Wire Line

The spaced two-wire line consists of two parallel conductors which are maintained at a fixed distance apart by insulating spacers or spreaders. There are two types of this line—the old type, the spreader bar, and the new type, the twin-lead line. The spreader bar type uses ceramic or polystyrene bars for spacing wires. The spacing may be up to 6 inches or more. The twin-lead line is the result of improved insulating materials and manufacturing techniques. It consists of two wires which are molded into a low-loss plastic called polyethylene. Impedances ranging from 75 to 300 are available in the twin lead line. Lines in general, however, may be constructed with any impedance up to 700 ohms by varying the wire diameter and spacing. The twin-lead line has a number of advantages over the spreader bar two-wire line. Wide spacing between the leads in the twin-lead type gives lower shunt capacity between the leads. Insulation losses are lower in it, since the wires are separated by other than air as are the wires in the spreader bar line.

The twin-lead two-wire line is not perfect. It has two principal disadvantages. In the first place, the line acts like an antenna by which it causes energy to be lost. This is due to the fact that magnetic fields extend some distance around from the conducting wires. When these fields collapse at a change in frequency cycle, some of the lines of force do not retract when the change is rapid, and consequently a few lines of force are projected outward in the form of radiation, just like radiated energy on an antenna. This energy is therefore lost. Secondly, nearby objects set up capacity with the wires and disturb current flow through the line. This obstructs the orderly transfer of energy by the line.

THE CHARACTERISTIC IMPEDANCE OF THE TWO-WIRE LINE. The relationship between the characteristic impedance and the dimensions of a two-wire line is expressed mathematically by the equation,

$$Z_o = 276 \log_{10} \frac{2D}{d} \qquad (20)$$

Where D is the distance between wires (center to center), and d the diameter of one wire.

Example

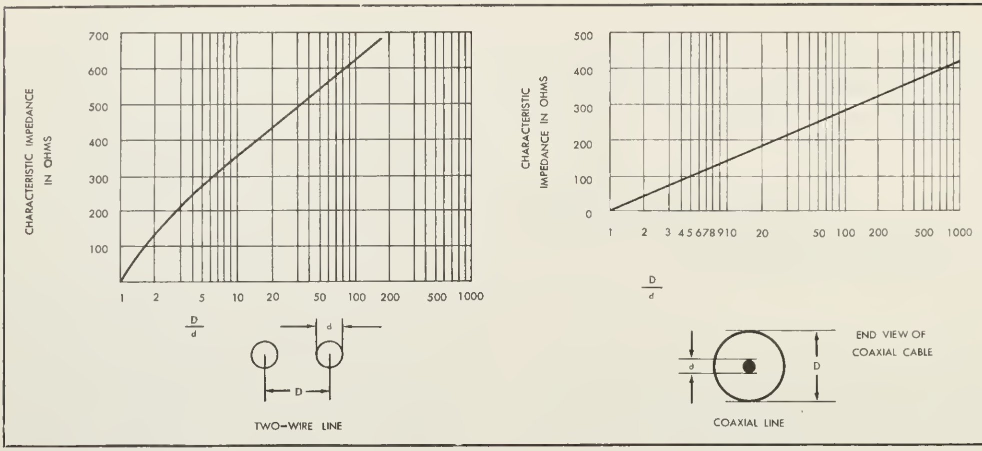
Problem. For a two-wire line constructed of No. 12 wire (.08 inch in diameter) spaced 6 inches apart, find the Z_o.

Solution:

$$\begin{aligned} Z_o &= 276 \log_{10} \frac{2D}{d} \\ &= 276 \log_{10} \frac{2 \times 6}{.08} \\ &= 276 \log_{10} 150 \\ &= 276 \times 2.1761 \end{aligned}$$

The characteristic impedance (Z_o) then equals approximately 600 ohms.

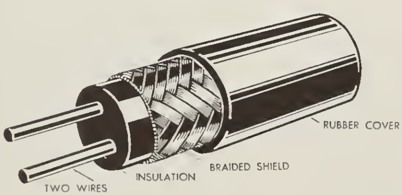
The graphs at the top of the next page are for plotting characteristic impedances. One is for a two-wire line. The other is for coaxial lines, which are explained later. In the graph for two-wire lines the preceding example is plotted in dotted lines.



Characteristic Impedance Graphs for Two-Wire and Coaxial Lines

Shielded Two-Wire Line

A shielded two-wire line is similar to the two-wire spaced line except that it is shielded by a copper braid. Notice the practical form of this line in the illustration just below. The copper braid gives it flexibility. Flexibility could be obtained by using unbraided tubing. However, the use of braid permits uniform spacing of each conductor during manufacture. This results in



Shielded Two-Wire Line

each wire being perfectly balanced capacitively to the surrounding conductor. As long as the balance is maintained, certain detrimental effects, such as high capacity to ground when the shield is grounded, are very slight. This line does not radiate energy because of the shield and thereby is not affected by nearby magnetic fields.

Coaxial Line

The coaxial line is the most universally used type of line for RF transmission. It consists of two conductors. One of the conductors is hollow; the other conductor is located inside the hollow conductor. The two conductors are made concentric in cross section to maintain uniform characteristics. Early (prewar) coaxial lines were rigid and were made of copper tubing in which there was a wire held in the center by ceramic or polystyrene washers or beads. Attempts to build flexible cables led to the use of outer conductors of copper braid spaced by a continuous row of beads or rubber insulators. However, rubber insulation caused excessive losses at high frequencies. In addition, the bead arrangement permitted air to enter the line causing moisture to collect. This resulted in high leakage currents and arc over when high voltages were applied. Wartime research solved this problem by using polyethylene plastic, a hard solid material, safely flexible over a temperature range of −40 to +80° C. Polyethylene is unaffected by such fluids as acids, alkalis, aviation gasoline, oil hydraulic brake fluid or sea water. Furthermore, there is no known solvent for polyethylene at ordinary temperatures. Its extremely low losses at high frequencies is important for radar. At 200 mc, the losses may be only one-hundredth of those of rubber under certain conditions. When carefully manufactured, a coaxial cable insulated with polyethylene will be uniform within .005 inches and will have no air between the conductors. Thus, due to its uniform characteristics, it will operate indefinitely with low leakage and with little danger of arc-over.

Where installations require still lower losses, such as 3000 mc airborne radar installations, an air-insulated rigid coaxial cable meets these requirements. In this cable, the center conductor is supported with metallic insulators, which are quarter wave sections of coaxial line. Since there is air in the interior, the interior of the line is usually pressurized to keep moisture out of it.

Rigid Coaxial Cable with Metallic Insulators

Coaxial lines have a number of advantages. First, the shielding is perfect for both magnetic and electrostatic fields. If you refer to the two end views of the coaxial cable in the illustration below, you can see that the electrostatic field is terminated at the outer conductor in a manner that none of the field is outside of the line. However, the magnetic field from the inner conductor does extend beyond the outer conductor, but as an equal amount of current flows in the outer conductor, setting up the field shown, which is outside of the outer conductor, it cancels the other field. Inside the cable these two fields add. Hence, there is neither an electrostatic nor a magnetic field outside of the cable. Secondly, since a coaxial line does not radiate, it does not pick up any energy. Therefore, it can be installed anywhere without being influenced by other strong fields. Thirdly, its capacity is lower than that of a shielded pair line.

Among the disadvantages of coaxial lines are that they cost more than two-wire lines, are more difficult to install than flexible cables, and it is more difficult to measure the fields in a coaxial cable than on an open-wire line.

CHARACTERISTIC IMPEDANCE OF COAXIAL LINES FROM DIMENSIONS. When you know the dimensions of a coaxial line, you can determine its characteristic impedance by the following equation,

$$Z_o = 138 \log_{10} \frac{D}{d} \qquad (21)$$

where D is the inside diameter of the outer conductor and d is the diameter of the inner conductor.

Example

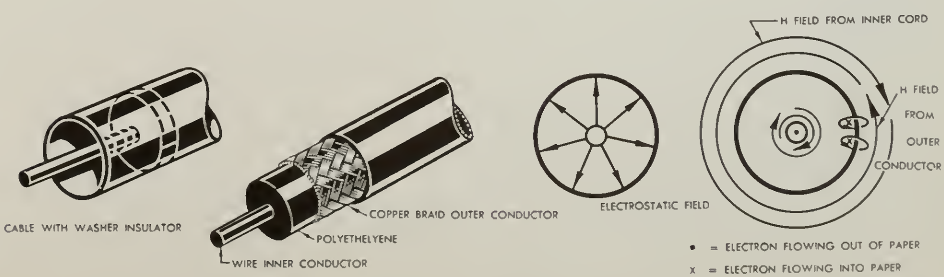
Problem. If the inner conductor is a ¼-inch rod, and the outer conductor is a .875 inch tubing (inside diameter), find Z_o.

Solution:

$$Z_o = 138 \log_{10} \frac{D}{d}$$
$$= 138 \log_{10} \frac{.875}{.25}$$
$$= 138 \log_{10} 3.5$$
$$= 138 \times .5428$$

The characteristic impedance (Z_o) equals 75 ohms.

You can conveniently plot this equation in the impedance vs dimensions graph shown on page 9-20. This graph shows this example drawn in dotted lines. Notice that 75 ohms is a popular value for characteristic impedance since the Q of a coaxial line is highest when its dimensions provide for 75 ohms resistance.



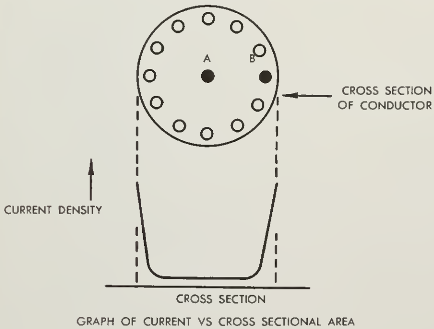
Coaxial Cables

LOSSES IN RF LINE

Although the theory of RF lines just given dealt with lines assumed to be loss-less, actually there are some losses in all lines. Line-losses may be of three types—copper losses, dielectric losses and radiation or induction losses.

Copper Losses

One type of copper loss is I^2R loss. In RF lines the resistance of the conductors is never equal to zero. Any time current flows through one of these conductors, there is some energy dissipated in the form of heat. This loss of energy is an I^2R or power loss, since power is equal to $P=I^2R$. With copper braid, in which the resistance is higher than the resistance in a solid tube, this power loss is accordingly higher.



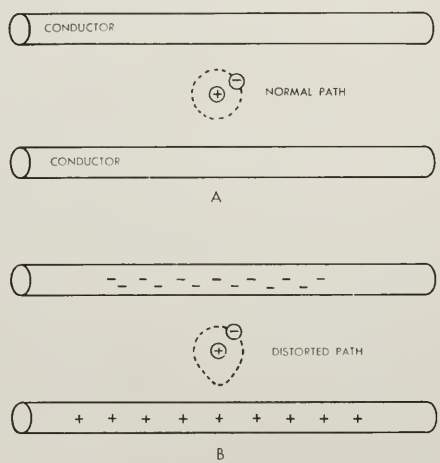
Skin Effect Losses in RF Lines

Another copper loss results from skin effect. It is explained as follows: When a DC current flows through a conductor, electrons move uniformly through the cross-section of the conductor. On the other hand, when an AC current flows, this is not the case. Expanding and contracting magnetic fields about each electron encircle other electrons and retard their movement. This is the phenomenon of self-induction. If you refer to the above diagram, you can notice that the electrons moving in area A expand and effect the movement of all the electrons near the surface at B of the conductor. At the same time, all the electrons moving at the surface have an equal field. These fields expand and retard the electrons at A. While a single electron exerts only a small effect on the outer electrons, the outer electrons in combination exert a tremendous effect on the center electron—so much so, that at frequencies of 100 mc and above, electron movement (current flow) is so small at the center that the center of the conductor can be removed without noticeable effect on current flow. Further, since the resistance of a conductor varies inversely with the cross-section, the cross-section (because current flows only in the outer portion of the conductor) is less and the actual resistance is greater. In conclusion, skin effect is the tendency of alternating currents to flow near the surface of the conductor, thus being restricted to a small part of the total cross-sectional area. This effect increases the resistance and becomes more marked as the frequency rises.

Because of the flow of current near the surface of a conductor, the conductivity of an RF line can be increased by plating it with silver. Most of the current will flow in the silver layer. The tubing serves mostly for mechanical support.

Dielectric Losses

Dielectric losses are due to the heating of the dielectric material (insulation) between conductors. This heating of the dielectric material takes power from the source. The heat itself is dissipated into the surrounding medium. The illustration at A below shows the mechanics of dielectric losses. When there is no potential difference between two conductors, the atoms in the dielectric material between them are normal, that is, the orbits of the electrons are circular.



Origin of Dielectric Losses

But, when a potential difference exists, as shown at B, the condition of the electron orbits changes. The excessive negative charge on one conductor repels electrons in the dielectric toward the positive conductor and thereby distorts their orbits. This change in the path of electrons requires work or power, which in turn is furnished from the power source for the RF line, thus introducing a power loss. The structure of the atoms, in some dielectric materials, rubber for example, is harder to distort than the structure of other materials. On the other hand, the atoms of some materials distort easily. Polyethylene is this type of material. It is used as a dielectric because little power is consumed when its electron orbits are distorted.

Radiation and Induction Losses

Radiation and induction losses are similar in that they are both due to the fields surrounding conductors. When the field about conductors cuts a nearby metallic object, there is a current induced in the object. This induced current results in a power being dissipated by the object. The power thus lost is supplied by transformer action from the source for the RF line. Radiation losses are due to the fact that some lines of force about a conductor do not return to it when the frequency cycle changes. These lines of force are projected into space as radiation, and since they do not return, the energy they use must be supplied by the power source, thus introducing additional power loss.

Importance of Losses

Although in cross country telephone lines resistance is a major problem from the standpoint of losses, it (resistance) is of little concern in RF systems, since RF lines are short and the DC resistance is quite small. However, the other types of losses previously discussed are of more concern. Most copper losses in RF lines are due to skin effect. These, however, can be minimized by using large diameter conductors with a low resistance surface material. Dielectric losses, on the other hand, are of considerable concern. To obtain minimum dielectric loss, it is necessary to use a *dry* air dielectric. When the line is to be flexible, it is necessary to use a low loss dielectric, such as polyethylene. Furtheremore, since dielectric losses increase with the length of the line, systems requiring extremely low loss use a line as short as possible. In two-wire lines, radiation losses can be decreased either by reducing the spacing between wires or by shielding. Both methods increase the capacity between wires, which is usually undesirable. For the least losses from this cause, coaxial cables are recommended because of their natural shielding. This type of loss increases as the square of the frequency. Thus, two-wire lines are not usable at 3000 mc and up. Shielded and coaxial cables are used for high frequency radar systems.

MEASUREMENTS IN RF LINES

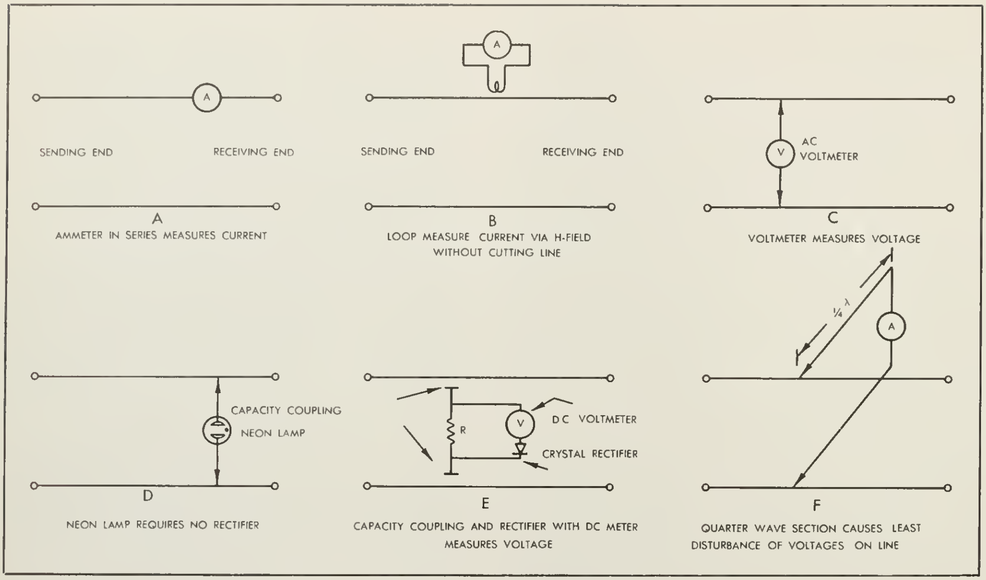
To be certain that an RF line is transferring energy with the least losses possible or that it is actually a certain number of quarter waves long, or that any other situation exists, it is necessary to make some electrical measurements on the line.

There are several simple methods for determining the magnitude of voltage or current at any point of an RF line. In using these methods, remember that the magnetic field around a line varies directly with current, and that the electrostatic field about the line varies directly with the voltage.

One method for observing current at any point on a line is either to break the line and insert an ammeter, or to place a loop in the magnetic field. The diagram at A at top of page 9-24 shows an AC ammeter connected in series with the line. Here the meter reads the standing wave of current at the point indicated. Sometimes it is not practical to cut a line. In this case, you can move a magnetic loop which is connected to the meter along the line. Here the meter reads maximum at the point where the greatest current is induced in the loop—that is, at the current loop—and it reads minimum at the nodes or point of current minimum.

To read voltage at any point, connect an AC voltmeter across the line at the point you are checking as shown at C.

Another simple method is to connect a neon lamp across the line. When the lamp is connected across the line, the voltage in it will cause the lamp to glow. The greater the voltage, the brighter the lamp will glow. An electrostatic field will make a neon lamp glow, even when it is not directly connected to the line, provided that the field is strong enough. In high powered installations, you need only to place the neon lamp in the vicinity of the line. A more accurate device is the sensitive DC meter with a rectifier. A crystal or diode may be used as the rectifier, as shown in diagram E. The resistor shown is



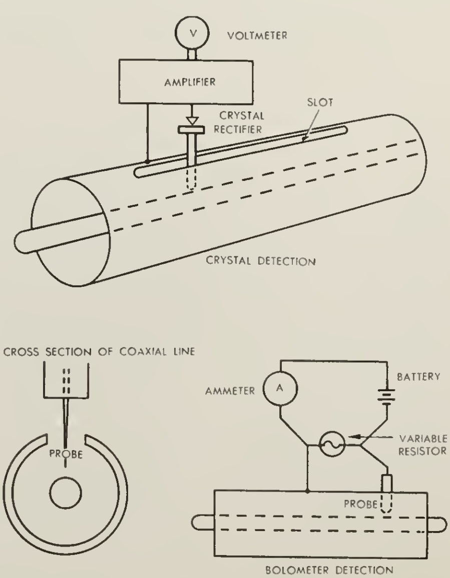
Making Measurements on RF Lines

capacitively coupled to both wires, and the voltage drop across the resistor measured by the meter.

All the preceding methods of measuring have the common disadvantage that ordinary measuring devices take a certain amount of energy from the line being measured. Such use of energy represents a lowered impedance at the point of measurement. Any change in impedance at any point along a line causes reflections. Thus, ordinary measuring instruments load (take energy) a line, and upset its normal chracteristics.

A high impedance measuring device suitable for making measurements in a two-wire line is the quarter wave section shown above in the illustration at F. This device, which consists of an AC ammeter, forms a short circuit across the section. As the ¼ wave section is shorted at the ammeter end, it presents an extremely high impedance to the line and very little current is needed to energize it.

When coaxial lines are used, as is the case with radar equipment, the magnetic and electrostatic fields are not accessible to loops and contacts. In measuring the field in this type of line, it is necessary to use a slotted line section as in the adjacent illustration. In the slot in the line a probe is inserted at a small depth, not far enough to touch the center conductor. The probe is a slender rod which acts as an antenna and which is excited by the electrostatic field which is



Crystal and Bolometer Detection

parallel to it. It is unaffected by the magnetic field. The slot does not reduce the effectiveness of the coaxial line because the currents flow parallel to the slot. With this arrangement the coupling is very slight and therefore very little energy is extracted by the probe.

The energy at the probe is AC at the RF frequency and is detected with some form of detector. A vacuum tube detector can be used, but usually a crystal detector or a bolometer is used. In the crystal detector a crystal rectifier changes the RF to DC, the magnitude of which varies with the amplitude of the AC voltage. The DC can be connected to a DC meter. Since the DC is usually quite small, it is first amplified by a vacuum tube amplifier before it is fed to the meter.

A bolometer is a piece of very fine wire, in which the resistance varies with temperature. As it is used in the diagram on page 9-24, the RF energy heats the wire. The heat changes the resistance of the wire to the DC current flowing from the battery. The meter then changes its reading. This change also can be amplified by a vacuum tube amplifier. The resistance element in the bolometer is usually a standard 1 100 ampere fuse, which is commercially known as a *Littlefuse*.

THE LECHER LINE

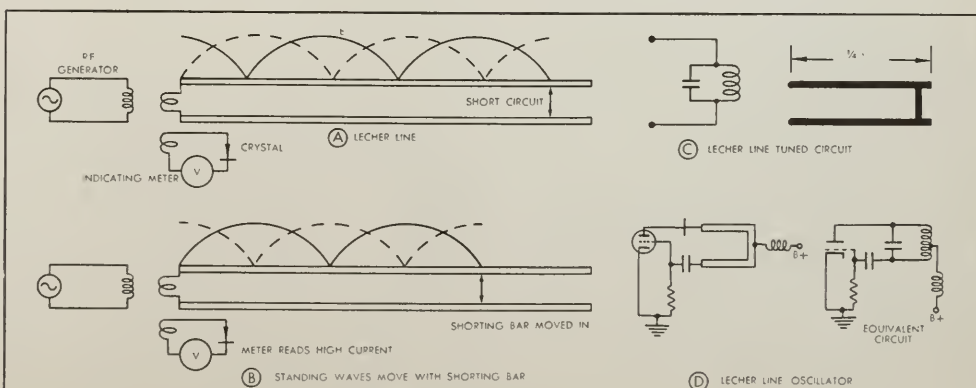
A Lecher line is the term applied to a two-wire RF line which is ¼ to 2 or 3 wavelengths long. Such a device is used for either investigating standing wave phenomena, measuring wavelengths, or utilizing the principles of an RF line. In the Lecher line below at A, two wires run parallel for about one wavelength. With this line, you can by experimentation determine several facts. For example, when the frequency of a transmitter is unknown, you can use a Lecher line to measure the wavelength of the transmitted frequency. Further, since the velocity at which energy travels on a line of this type is about 97.5% of the velocity in free space (300 million meters per second), you can compute the frequency by the formula,

$$f = \frac{300 \times 10^6 \times .975}{2\,w} \qquad (22)$$

where w is the distance between two minimum or maximum standing wave points, and f the frequency in cycles per second.

Generally, a sliding shorting bar is a part of a Lecher line. As the sliding bar moves toward the generator, the standing waves move with it (shorting bar). Results are read by a pick-up coil with an indicating meter which is placed near the excitation coil of the Lecher line.

The use of the shorting bar and the pick-up coil to determine current minimums and maximums is illustrated as follows. The standing current wave shown at A shows that the standing wave of current is minimum at the location of the excitation coil. Thus, very little current flows through it, and the small magnetic field resulting causes the meter to be excited very little, indicating a current minimum. In diagram B, which shows the shorting bar moved, the current is maximum at the excitation coil, and the meter pick-up coil, which is excited by the strong magnetic field about the excitation coil, indicating a high current and the presence of a current maximum.



Use of Lecher Lines

A Lecher line ¼ wavelength long displays the characteristics of a parallel resonant circuit. For this reason, a ¼-wave Lecher line often is used as a tuned circuit in an ultra high frequency oscillator. At 200 mc, a quarter wave is only about 14 inches, which is small enough for practical use as is shown at C on page 9-25. (Note the typical Lecher line oscillator at D.) A Lecher line oscillator is impractical at lower frequencies. For example, at a frequency of one mc, a quarter wave line is 250 feet long. Obviously, this is too long for the ordinary transmitter operating at these frequencies.

USE OF RF LINES

A radar set may be compared to the original use of RF line as a device for transferring energy from a generator to a distant load. In radar sets, the generator is the transmitter and the means of transferring energy from the transmitter is the antenna.

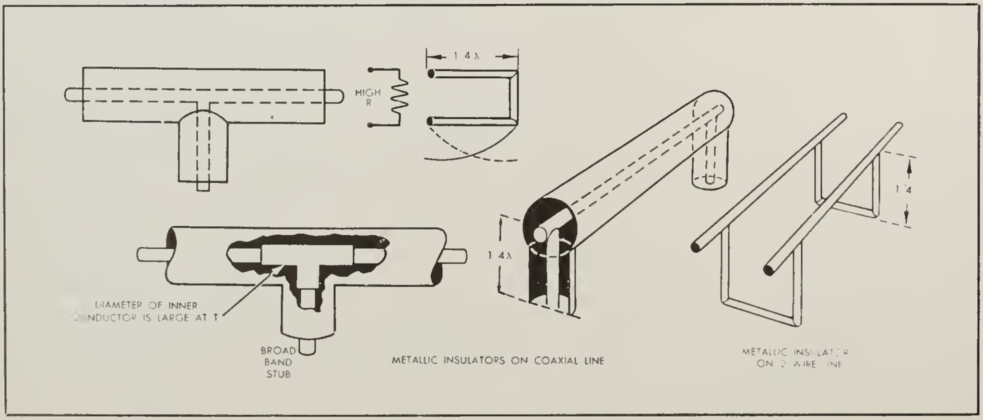
In radar, standing wave ratios must be as near unity as possible for satisfactory operation. However, even with a near unity ratio, standing waves cause several deficiencies. First, they reduce power handling capacity. The maximum power that a line can handle is limited by the maximum voltage that can be impressed across the line. The maximum voltage in turn is limited by the break-down voltage of the insulation. Standing waves increase the voltage at some points without additional power. Second, the efficiency of the line is lower with standing waves because of the high circulating currents at the current loops. The I^2R losses introduced by this current cannot be tolerated. Third, standing waves change the input impedance of the line. The frequency and power output of the usual radar transmitter varies with load impedance, so the line can change both the transmitter frequency and power output. In practice it is usually possible to get the standing wave ratio down to less than 1.5:1, a ratio which is considered satisfactory.

When energy is transferred from the antenna to the receiver, it is also desirable to get a low standing wave ratio, since any losses that occur subtract from the energy radiated by the antenna system. A low standing wave ratio also prevents excessive attenuation of the received echoes which are usually too weak to stand much loss. For this reason, the amplitude of a radar "echo" signal is kept very low at the antenna in order to minimize any undesirable attenuation.

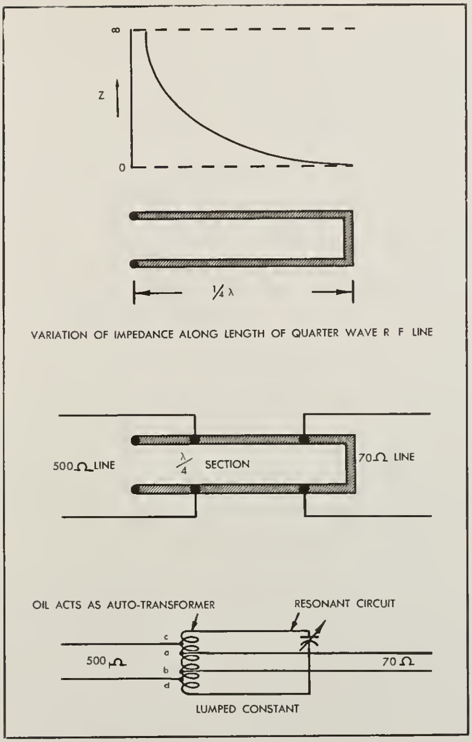
RF lines also are used in radar equipment to transfer intermediate frequencies, video signals and pulse voltages. Usually these lines are short electrically. Furthermore, when they are matched properly, the losses in them are small.

METALLIC INSULATORS

In any of the previous uses of RF lines described, standing waves should not exist because of the unusual phenomena which results. However, there are many uses of RF lines other than the transfer of energy in which standing waves are desirable. The metallic insulator is an example of such a wave. Since radar sets operate at a single frequency, it is possible to support the center conductor or the coaxial line by a ¼



RF Line Used as Metallic Insulator



RF Line used as Impedance Matching Device

wave section of RF line. The inverting properties of this ¼ wave section make it possible to short the receiving end of the section. This makes the input ends look like an infinitely high resistance. However, because of slight copper losses, the infinite resistance is not quite attained, but the input resistance nevertheless is very high. The metallic insulator has no dielectric losses and is not damaged by arc-overs. Ordinary insulating materials are carbonized by an arc. The carbonized path acts as a low resistance conductor at any voltage. Air insulation in quarter wave sections does not carbonize. Thus, with it the line is not shorted after the arc.

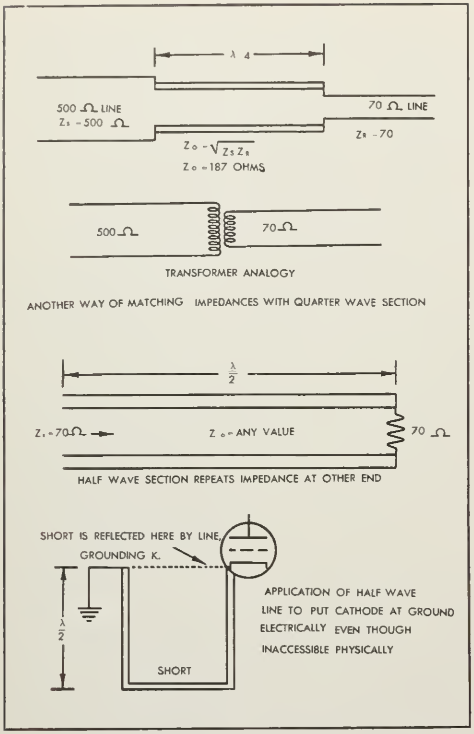
Metallic insulators are practical only at the higher frequencies, such as 3000 mc, where the quarter-wave "stub" is about 2 inches long. At lower frequencies the physical length of the insulator would be too great for practical use. A disadvantage of this insulator is that it is a quarter-wave at only one frequency. This limits its use to constant-frequency equipment.

This is the function of the broad band stub. In pulse modulation, many side band frequencies are present in a radar wave. At frequencies slightly different from the radar frequency a stub is normally reactive. This causes reflections in the main line. These reflections can be minimized by using a broad (frequency) band stub, like the one shown. Note that the diameter of the center conductor is enlarged at the junction.

IMPEDANCE MATCHING DEVICE

The impedance of a shorted RF lines varies widely over its length. Notice to the left the curve showing the variation of impedance along the length of a quarter wave RF line. When this section of the line is shorted, the impedance varies from zero value (at location of short) to infinity at a distance of a quarter wave from the short. The curve shown is a hyperbola. When this section of the RF line is excited, it is possible to connect to it with perfect match somewhere along its length devices of any impedance. For example, you can match a 500-ohm line and a 70-ohm line to this section so that there will be no standing waves on either line. The illustration on this page shows a 500-ohm and a 70-ohm line connected to the quarter wave section. Energy from the 500-ohm line sets up standing wave on the section. When the section is excited, the voltage to current ratio is correct at the point indicated. This means that the 500-ohm line is terminated in its Z_0. At the shorted end of the section, the voltage to current ratio is less. Therefore, at some point near this end of the section the value of $\frac{E}{I}$ must equal 70 ohms. At this point you can properly match the 70-ohm line.

The action involved in matching a 500-ohm line to a 70-ohm line by a quarter-wave section is comparable to the action involved in a tuned auto-transformer. Notice the diagram in the illustration labeled, "Lumped Constant Analogy." The coil in this diagram acts like an auto-transformer. When a 70-ohm line is connected to points a and b on it the reflected impedance at c to d is equal to 500 ohms. Like transformer action, the direction of energy transfer may be reversed without changing connections. The 70-ohm line can feed the quarter-wave section—a coil in this case—and the 500-ohm line will present the proper impedance to it. Any other value of impedance can be found for either line by moving the connections along the quarter-wave section.



RF Lines as Transformer

The above illustration shows still another way to get transformer action by using a quarter-wave section. As a quarter-wave section will invert any value of impedance, by selecting a section of a certain impedance, you can invert the 70 ohm impedance of the line used in the preceding example to look like 500 ohms at the other end. The equation for this relationship is,

$$\frac{Z_s}{Z_o} = \frac{Z_o}{Z_r}$$

Using this equation and solving for Z_o,

$$Z_o = \sqrt{Z_s Z_r} \qquad (23)$$

shows that the impedance of the quarter wave in this example is equal to,

$$Z_o = \sqrt{Z_s Z_r} = \sqrt{500 \times 70} = \sqrt{35000} = 187 \text{ ohms}$$

When you choose conductors of the proper dimensions and correctly space them, you can construct a 187 ohm line. As you can see in this arrangement, there are standing waves on the quarter-wave section, but none on either line connected to it.

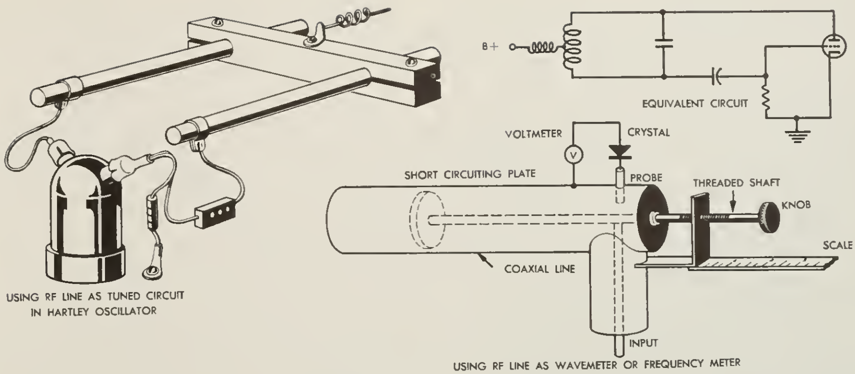
In the section which is a half-wave long, the output impedance is repeated at the input. For example, a 70-ohm load is inverted over a quarter wavelength to some other value, depending on the characteristic impedance of the line. It is further re-inverted to 70 ohms over the remaining quarter wave section. The entire action is like that of a one-to-one transformer. This type of inversion is frequently used in radar. For example, a certain impedance might be required between two tube elements where the lengths of the connecting leads is long at the operating frequency and a connection of the impedance outside of the tube does not provide the desired results. By using a half-wave line in which the impedance is located at one end and using the tube leads for part of the half-wave, it is possible to produce a virtual impedance right at the tube elements themselves. The illustration on this page shows a half-wave line used to put the cathode at ground electrically even though the ground is not accessible physically.

Just because the preceding discussion of the principles of impedance matching lines dealt only with those which had transmission lines connected to them, do not conclude that impedance matching lines are used solely with transmission lines. Quite to the contrary, there are several types of devices which can be connected to an impedance matching line. For example, the input or load to an impedance matching line could be, in addition to a transmission line, an RF device such as a transmitter, an antenna, a receiver input network, or a vacuum tube plate or grid circuit.

TUNED CIRCUITS

As was explained previously, an RF line having a certain length displays the properties of a resonant circuit. There are many cases in high-frequency equipment where this property is used. For example, quarter-wave and half-wave sections are commonly used as tuned circuits in radar equipment. The Hartley oscillator shown in the next illustration uses a quarter-wave Lecher line for the tuned circuit. Compare this circuit with its equivalent circuit and note how the radar circuits differ from the conventional representation used in the equivalent circuit.

Another use of a section of an RF line is as a wave meter. Notice the illustration showing the RF wave meter. The meter consists principally of a coaxial line with a movable short circuiting plate. When this plate is at a distance any mul-

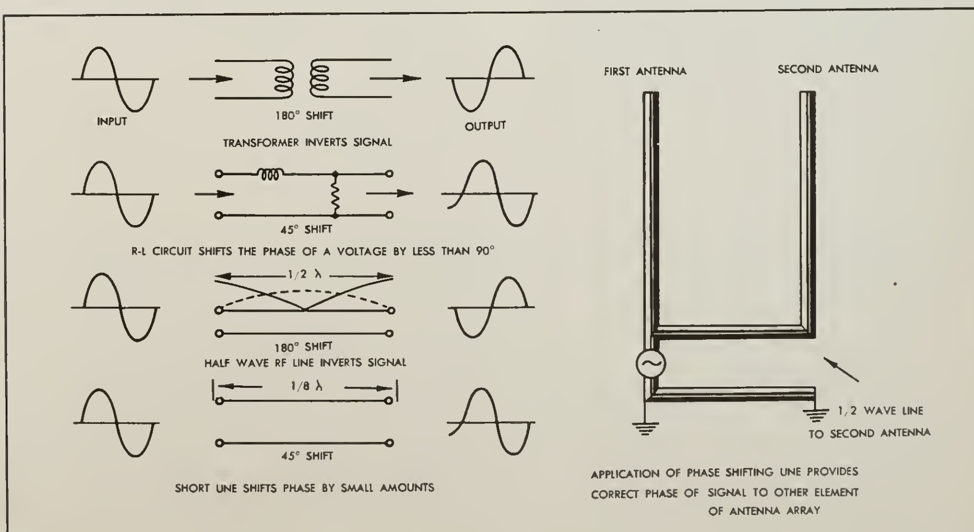


Using RF Lines as Tuned Circuits

tiple of a half wave-length from the probe in the meter, the voltage at the probe drops to zero. The amount of voltage at the probe can be read on a rectifier type voltmeter. The distance represented by any two consecutive zero voltage readings on this meter represents one wave-length. The knob at the right of the meter provides a means for moving the position of the short-circuiting plate. The distance traveled by this plate can be read from the metric scale which is calibrated in frequencies.

PHASE INVERTING OR SHIFTING NETWORKS

Sometimes it is desirable in electronic operations to shift the phase of current or voltage a certain number of degrees, or in some cases to invert the phase completely. Previously it was explained that the phase of voltage changes 360° in one wave-length of an RF line. Thus it follows that the phase relationship of a sine wave at a distance of a half-wave-length from the sending end will be 180° out of phase with the sine wave at the sending end. This is the same



Shifting Phase with RF Line Sections

action that occurs in a transformer. An LR circuit as shown in the illustration at bottom of page 9-29, however, will shift phase less than 90°. Lines made up of small parts of a wavelength shift phase only in small amounts.

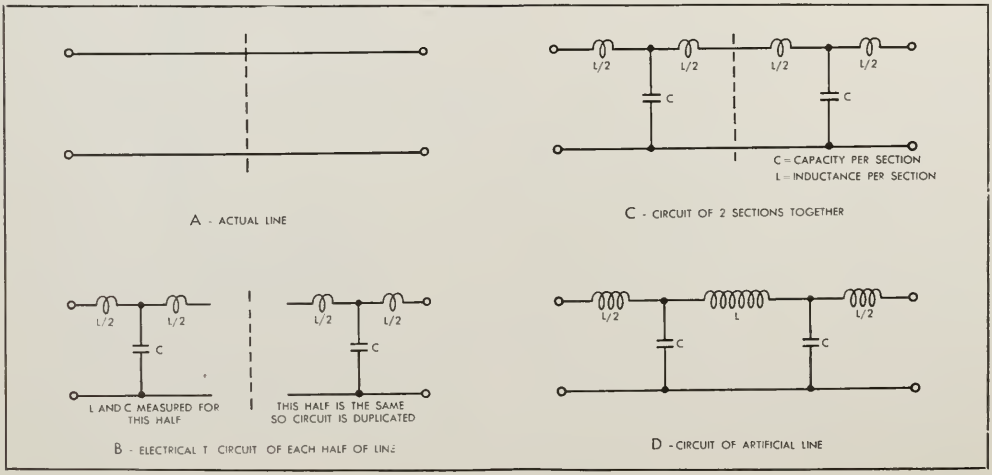
An example of a use of phase shifting is antenna excitation. Whenever two antennas are excited out of phase, a certain antenna pattern is produced. One easy way to obtain a 180° phase difference for exciting the antennas is with a half-wave RF line. The same illustration shows the use of a phase shifting line to provide the correct phase difference in the elements of the antenna array. The antenna on the left is excited by the RF generator. The generator also applies the same signal to a half-wave line. Due to the phase shifting characteristics of the RF line, the antenna at the other end is fed a signal which differs 180° from that fed to the first antenna.

Other antenna directional characteristics can be obtained by varying the length of the RF line. This variation of line length causes a change in phase, which in turn gives different antenna radiation characteristics.

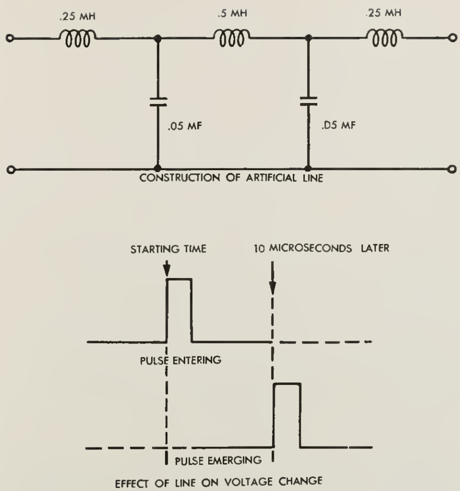
ARTIFICIAL LINES

Although the delay characteristics and charging action of RF lines make them useful in some radar applications, they cannot be used in others, chiefly because they would have to be long. An actual RF line would have to be considerably long, perhaps several hundred feet, for some uses in radar. Even if a line of such length could be coiled up, it would occupy more space than there is available in an airplane. Thus, so far as airborne equipment is concerned, actual transmission lines are out of the question. For this reason artificial lines are used instead of actual lines. Artificial lines possess all the characteristics of actual lines, but at the same time do not have the bulk and physical length of an actual line. They are constructed by first determining the capacity and inductance of the desired line and then lumping these variables into equivalent inductors and capacitors. The result is an electrically equivalent, but physically different line. The illustration on this page shows an actual two-wire line. It also shows the equivalent circuit for each section as it is when the section is split into two parts and the characteristics of it measured. Since the other section has identical characteristics, the two can be joined together. The final circuit shown at D, lumps the inductive and capacitive characteristics of each section and simulates the artificial line. As you can see, the artificial line performs just like an actual line, but does not occupy nearly as much space.

In any line, time is required for any voltage change to travel the length of a line. This time characteristic makes it possible to slow down the transfer of a voltage change in its travel from one circuit to another.



Design of the Artificial RF Line



Artificial Line Used to Delay a Pulse Voltage 10 Microseconds

The artificial line shown above provides a means of delaying a pulse voltage 10 microseconds. This line is constructed with an inductance of ½ millihenry and a capacity of 0.05 mf in each of the two sections. To compute the time for a voltage change to move through one section, use equation (9) on page 9-5. The equation for time delay for the entire line is the following:

$$T_d = N\sqrt{LC} \qquad (24)$$

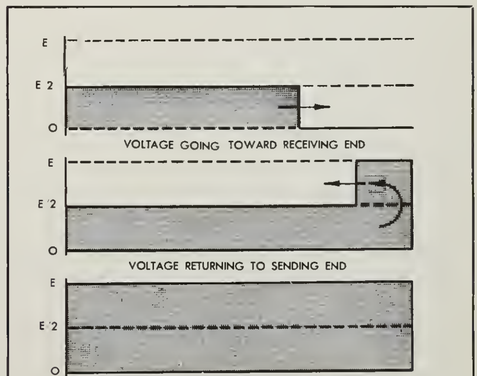
where T_d is the time delay in seconds for the entire line, N, the number of sections, and L and C, the inductance and capacity per section in henries and farads respectively.

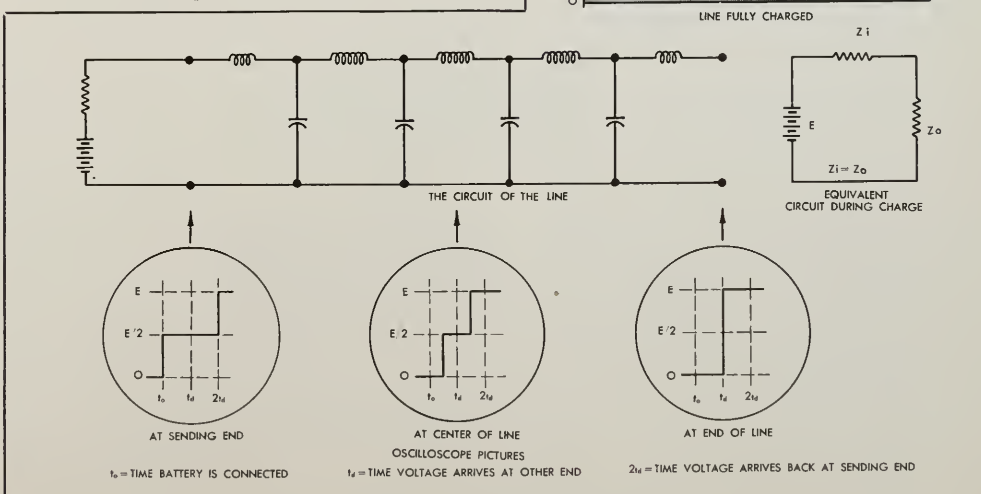
The following gives the calculations for finding the time delay for the entire artificial line shown at the left.

$$T_d = N\sqrt{LC} = 2\times\sqrt{.5\times10^{-3}\times.05\times10^{-6}}$$
$$= 2\times\sqrt{5\times10^{-4}\times5\times10^{-8}}$$
$$= 2\times\sqrt{25\times10^{-12}}$$
$$= 2\times5\times10^{-6}$$
$$= 10^{-5} \text{ seconds}$$

Time delay equals, $T_d = 10$ *microseconds*

One particular radar set uses the time delay characteristics of an RF line in its operation.



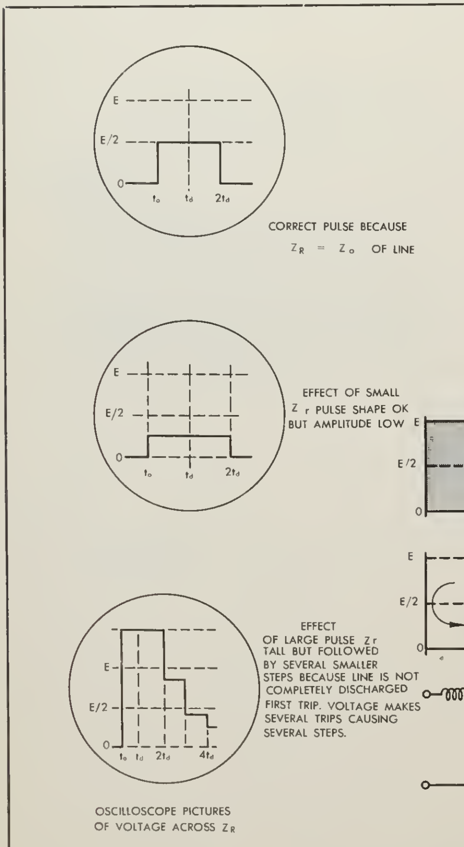


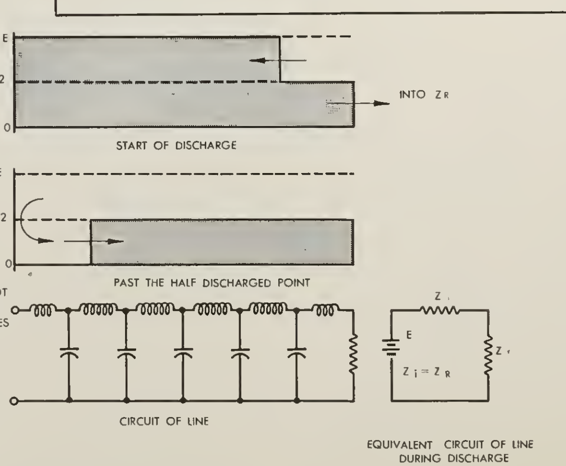
Mechanism of Charging Artificial Line

In this set, the time base is generated in such a way that it does not start until after the first radar echo returns to the receiver. In order for these signals to be displayed at their correct range mark, they are sent through an artificial RF delay line, which delays the signals just enough so that the very first returning signal does not arrive at the Cathode Ray Tube (CRT) until the time base has begun to form. Although a large number of video frequencies are present in the signal going through the line, all of them are delayed an equal amount.

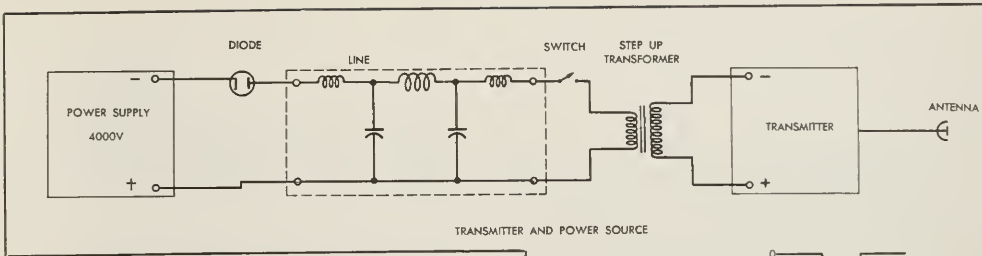
Another phenomenon of RF lines is the manner in which the lines charge and discharge when a DC voltage is applied to them. When DC is applied to a line in which the source impedance is matched to the line impedance, half the source voltage appears across the line impedance at the time the battery is connected to the line. This produces a change in voltage across the line. This voltage change travels down the line, charging it as it goes. When it reaches the open end, it is reflected and starts back along the line. All condensers charge to half the battery voltage when the voltage is going down, and to full battery voltage when the voltage is coming back. When all condensers have been charged to a value equal to the battery voltage, current stops flowing from the battery. If you view these voltage changes with an oscilloscope, their wave-shapes will appear like those shown in the illustration at the bottom of page 9-31. Note that a voltage which is shorter in time nearer the receiving end than at the sending end is produced. This voltage is called a *step* voltage.

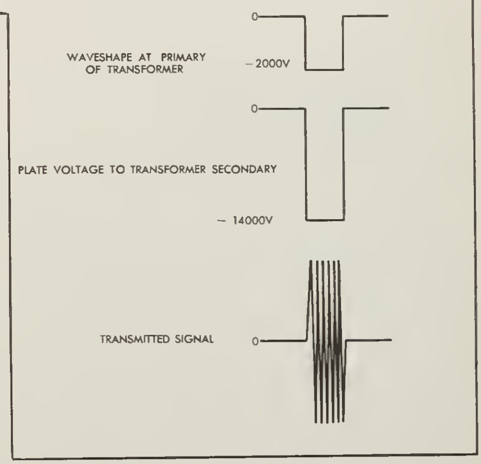
When this line is discharged into a resistance equal to its characteristic impedance, the result is a square wave which has a constant amplitude and a duration equal to twice the time for the pulse to travel the length of the line. The amplitude is equal to half the charge voltage because the line discharges through its own impedance and the load impedance (Z_R), which are connected in series. This cuts the load voltage by half and produces a drop of voltage which immediately travels down the line and back to the starting place as shown in the illustration below. As the end of the line is open, the reflection there is in phase. This produces a voltage across the load impedance (Z_R)





Discharging Artificial Line Through Z_R





Simple Circuit Showing Pulse Line used in Radar Transmitter Circuit

equal to half the charge voltage and which lasts for twice the time of travel on the line.

Usually, an artificial line is constructed to generate this square wave. When the L and C for each section is known, the time duration of the square pulse, t_p, is equal to

$$t_p = 2n\sqrt{LC}$$

The following are the calculations for the time duration of the pulse generated in the artificial line at top of page 9-31:

$L = 5$ *MH per section*

$C = .05$ *MF per section*

$N = 2$ *sections*

$$t_p = 2N\sqrt{LC} = 2 \times 2 \times \sqrt{.5 \times 10^{-3} \times 5 \times 10^{-8}}$$
$$= 4 \times \sqrt{25 \times 10^{-12}}$$
$$= 4 \times 5 \times 10^{-6}$$
$$= 20 \times 10^{-6}$$

The time duration equals,
$t_p = 20$ *microseconds*

When an artificial line is not terminated in its characteristic impedance (Z_o), several trips (reflections) are required for complete discharge. This means that discharge cannot be completed in a single round trip.

In the illustration on page 9-32 are a number of oscilloscope pictures which show the discharge action of an artificial line across a small Z_r and a large Z_r. When Z_r is small, most of the energy is dissipated in the line. In this case the shape of the pulse produced is satisfactory, but its amplitude is low. When Z_r is large, the pulse is tall. However, it is followed by several smaller steps—called tails. These tails exist because the line did not completely discharge during the first trip and more trips are necessary for complete discharge. One of the pictures shows the discharge action in which Z_r equals the Z_o of the line.

To calculate the Z_o of the line, use the formula

$$Z_o = \sqrt{\frac{L}{C}}$$

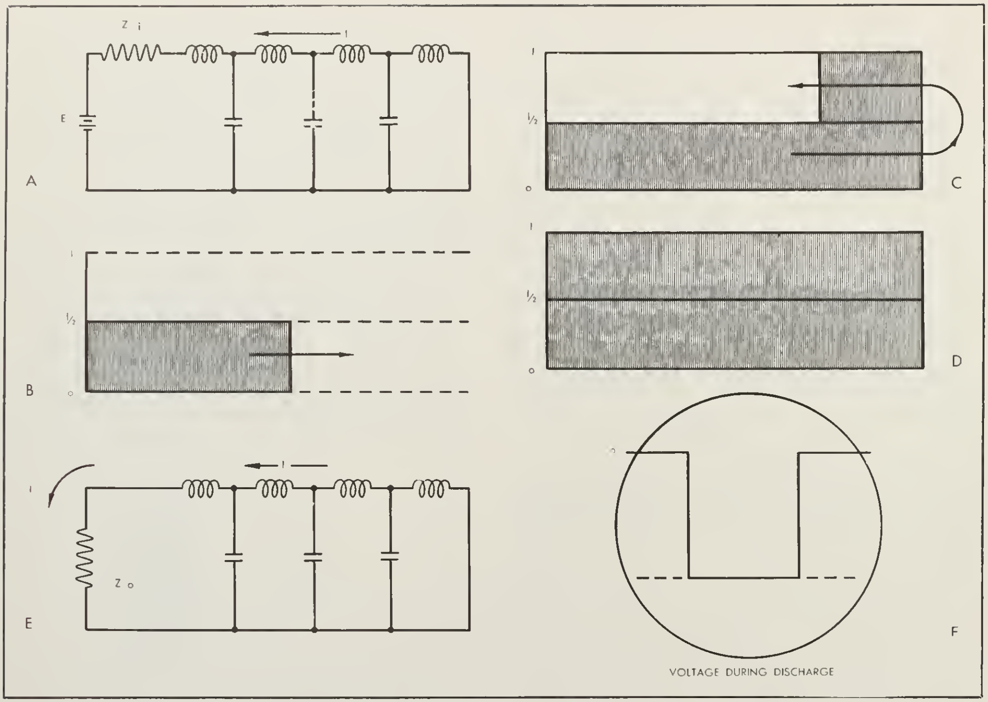
The following are the calculations for finding the Z_o of the artificial delay line shown at the top of page 9-31:

$$Z_o = \sqrt{\frac{L}{C}} = \sqrt{\frac{.5 \times 10^{-3}}{5 \times 10^{-8}}}$$
$$= \sqrt{\frac{5 \times 10^{-4}}{5 \times 10^{-8}}}$$

The characteristic impedance is equal to

$$Z_o = \sqrt{10^4} = 10^2 = 100 \text{ ohms}$$

An artificial pulse line is capable of producing a perfectly measured square wave pulse which has an amplitude equal to several thousand volts. As a matter of comparison, a vacuum tube multivibrator and amplifier circuit which could provide a similar output would weigh a hundred pounds and occupy a few cubic feet of space. On the other hand, a pulse forming line of equal capacity would consist of a few inductors and capacitors and occupy an area only about 4 inches square. Notice the above illustration showing a simplified radar transmitter which uses a pulse line. In it a conventional power supply



Pulse Formation with Shorted Line

charges the pulse line through a diode tube to 4000 volts. The diode prevents the line from discharging through the power supply. When you close the switch, a one-microsecond 2000 volt pulse is generated. The transformer, a step-up type, increases this 2000 volts several times. This voltage serves as the plate voltage for the transmitter. The transmitter oscillates during the time that plate voltage is applied to it for a period of one microsecond. Thus, a radio signal of one microsecond duration (or a radar pulse) is radiated by the antenna.

Another use of an artificial line is for forming pulses when the receiving end is short circuited. The operation of the line in this case depends on energy stored in the magnetic fields around the inductors in the line. Notice the above illustration labeled pulse formation with a shorted line. When the battery voltage E is applied to the line as shown in A, it moves down the line and charges each condenser. At the short circuit, the voltage turns around and returns up the line. On its return trip, it discharges all the condensers. Meanwhile, the current likewise moves down the line and is reflected back at the short. See B and C above. On its return trip up the line its value is double the current when it went down. After one round trip, the voltage across the line at all points is equal to zero while the current at all inductors is high as shown in D above. When a load is suddenly connected to the line (circuit shown in E) the current in the inductors will flow through the load. This current drops to one-half, goes to the shorted end and back, maintaining a current flow for the period of one round trip on the line. At F is shown the voltage output pulse.

CHAPTER 10

Waveguides and Cavity Resonators

The high frequencies employed by radar make possible the use of two unique, but very practical devices—waveguides and cavity resonators. A waveguide is a hollow pipe for transferring high frequency energy. A cavity resonator is a hollow metallic cavity in which electromagnetic oscillation can exist when the cavity is properly excited. The purpose of this chapter is to acquaint you with these devices. It discusses their theory, operation and uses.

WAVEGUIDES

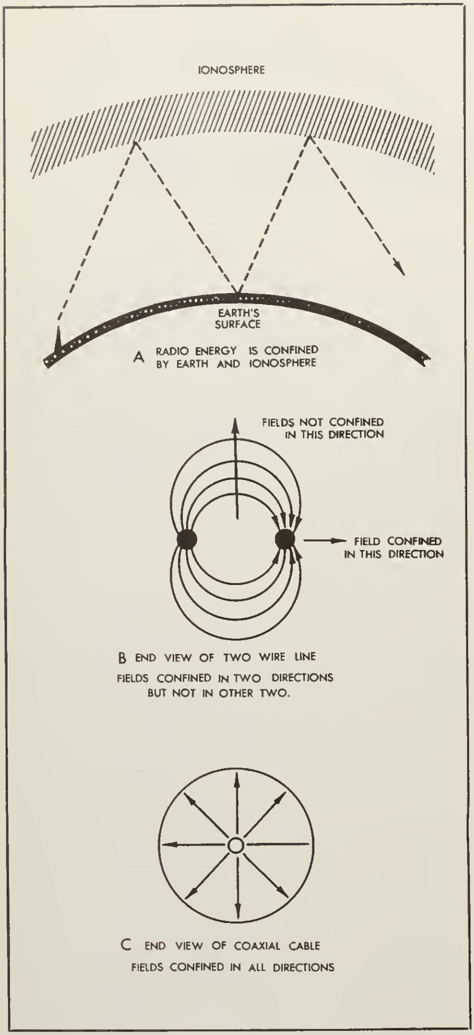
GUIDED AND UNGUIDED ELECTROMAGNETIC WAVES

In general, there are two methods for transferring electrical energy—one is by current flow in wires; the other is by movement of electromagnetic fields in space. Electrical energy can be transferred as current flow in a number of types of transmission lines, for example, two wire lines and coaxial lines; in space it moves as electromagnetic fields whenever a radio antenna radiates energy.

Electromagnetic fields which move in space are confined largely to the area between the earth and the ionosphere as you can see at A on the next page. (The ionosphere is the thicklayer or cloud of free ions and electrons which exists at a height of about 60 miles above the earth.) The change in the dielectric constant of this part of the atmosphere is sufficient to reflect electromagnetic fields which strike it except xtremely high frequency radiation that strike it almost perpendicularly. An electromagnetic field that does not start out in a direction parallel to the earth and the ionosphere, follows a zig-zag path in between these two areas and may or may not be reflected back to the earth.

Although the transfer of energy by electromagnetic fields and by currents in wires may seem to be unrelated phenomena, actually the present trend on the part of electronic scientists is to look even on two-wire lines as elements which guide electromagnetic fields from one place to another. The currents in the wires are merely considered incidental to the action and the result of the moving fields.

Strictly speaking, a two-wire line is a poor guide for transferring electromagnetic fields, because it does not confine the fields in a direction perpendicular to the plane which contains the wires as shown at B. This results in some energy escaping in the form of radiation. Electromagnetic fields may be completely confined in this direction when one conductor is extended around the other to form a coaxial cable as shown at C. In a coaxial cable, energy transfer also is said to take place by electromagnetic fields, rather than by current flow. However, this method is not too efficient at high radio frequencies, since skin effect limits the current carrying area of a conductor to a thin layer at its surface. Another disadvantage is that the ability of the fields to form is limited by the amount of current flow associated with it. When the resistance in a conductor is increased, current flow in it is reduced depending on the amount of increase. This reduces the magnitude of the fields. On inspecting the cross-section of the coaxial cable, you can see that the surface area of the inner conductor is much less than the surface area of the outer conductor. This causes the inner conductor to retard the current considerably more than the outer conductor and results in a reduction in the efficiency of energy transfer. If you could remove the center conductor and retain the fields, energy might be transferred with less loss.



Guiding Waves

Electromagnetic fields can transfer energy in a line which does not have a center conductor provided the configuration of the fields is changed to compensate for the missing conductor. The area remaining, which is virtually a hollow pipe, is called a *waveguide*. A waveguide does not necessarily have to be circular in cross-section. Practical waveguides, for example, are sometimes square, rectangular, or eliptical in cross-section.

Metallic walls are not necessary to guide electromagnetic fields in a waveguide, for the fields will be reflected whenever they encounter any kind of a substance which has a different dielectric constant than the substance in which they are traveling. For example, fields can be made to travel through a ceramic rod with a little loss of energy. When they encounter the air at the surface of the rod, they are reflected back into the rod.

Waveguides vs. RF Lines

As you previously learned, the three types of losses in the RF lines are copper losses, dielectric losses, and radiation losses. Briefly these losses are described as follows: Copper loss is an I^2R loss. It becomes appreciable whenever skin effect reduces the conducting area of the lines. Dielectric losses, as you recall, are losses due to the heating of the insulation between conductors. Radiation losses are losses due to energy escaping from RF lines in the form of radiation.

ADVANTAGES OF WAVEGUIDES. Considering waveguides from the point of view of these losses, they have the following advantages:

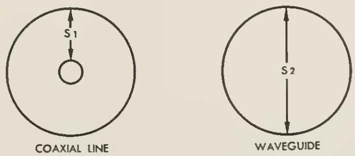
1. Copper losses are small in waveguides. Since a two-wire line consists of a pair of conductors which are small, the surface area of each is likewise small. In the case of a coaxial cable, although the surface area of the outer conductor is large, the inner conductor is small, and it produces considerable copper losses. On the other hand, a waveguide, as it does not have a center conductor, has a large surface area. Therefore, whenever current flows, the copper losses in it are less than those in other types of lines.

2. Dielectric losses are small in waveguides. In two conductor lines, some form of insulation is used between the conductors. The fields which move around this insulator cause heat, and the heat in turn takes power from the line. A waveguide has no center conductor to support. Furthermore, there is only air in the hollow pipes. Since the dielectric loss of air is negligible, it follows that the dielectric losses in it are small.

3. Radiation losses are less in a waveguide than in a two-wire line. In a waveguide, fields are contained wholly within the guide itself just as in a coaxial line. Therefore, only a negligible amount of energy is radiated.

4. The power handling capacity of a waveguide is greater than that of a coaxial line having an equal size. Power is a function of E^2/Z_0, where

E is the maximum voltage in the traveling wave and Z_0 is the characteristic impedance of the line. E is limited by the distance between the conductors. In the coaxial line illustrated just below, this distance is S_1. In the waveguide, the distance which is S_2 is much greater than S_1. Therefore, the waveguide is able to handle greater power before the voltage exceeds the breakdown potential of the insulation.



Comparison of Spacing in Coaxial Line and Waveguide

5. A waveguide is simpler to construct than a coaxial line. This is due to the fact that in the waveguide the center conductor is eliminated completely.

6. The physical stamina of a waveguide is greater than that of a coaxial line. This is due to the fact that unlike a coaxial line, it has no center conductor or insulators which can be displaced or broken.

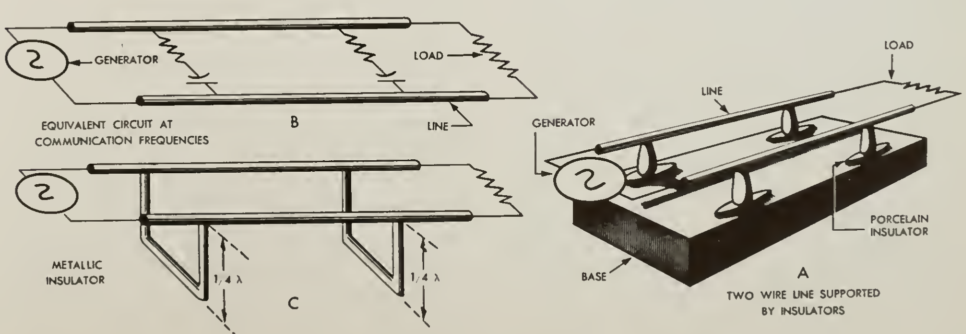
DISADVANTAGES. In view of these advantages, you may wonder why waveguides are not used exclusively for transferring energy. There are, however, two disadvantages which make waveguides impractical to use at any but extremely high frequencies. In the first place, the cross-sectional dimensions of a waveguide must be in the order of a half-wave-length for it to contain electromagnetic fields properly. A waveguide used at one megacycle, for example, would be about 700 feet wide. At lower radar frequencies, 200 mc for example, this waveguide would have to be about 4 feet wide, while at higher radar frequencies, such as 10,000 mc, it need be only one inch wide. Therefore, dimensions which waveguides require make them impractical at any frequency lower than about 3000 mc. In the second place, if the dimension of the guide is a half wavelength or less, energy will not be propagated through the waveguide. The reason for this is that for any given waveguide there is a cutoff frequency, below which it does not function as a power transfer device. This also limits the frequency range of any system using waveguides.

WAVEGUIDE THEORY

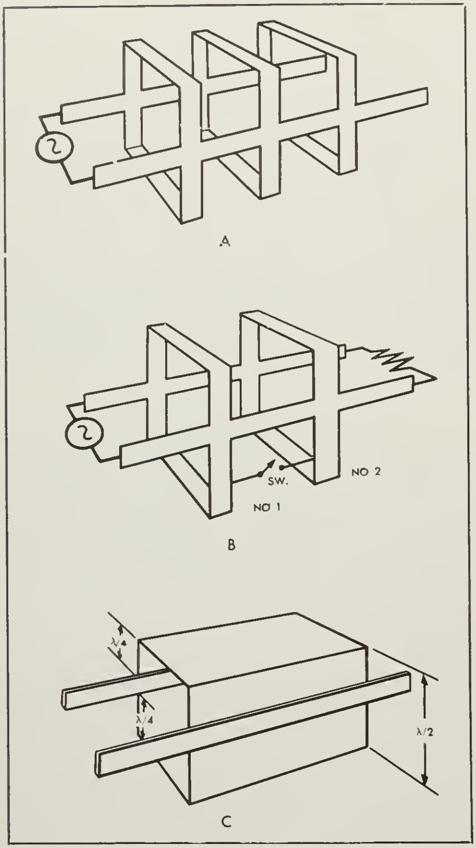
An exact mathematical analysis of the way in which fields exist in a waveguide is quite complicated. However, it is possible to obtain an understanding of many of the properties of waveguide propagation by using the following simple analogy, which shows both how the fields are able to exist in a waveguide and how you can handle them.

Analogy of Waveguide Action from a Two-wire RF Line

To understand the action of a waveguide, assume that a waveguide has the form of a two-wire line. In this condition there must be some means of supporting the two wires. Furthermore, the support must be a non-conductor, so that no power will be lost by radiation leakage. An efficient way for both insulating and supporting the two-wire line is shown in the illustration at the bottom of this page. This line is spaced, insulated, and supported by porcelain stand-off insulators. At communication frequencies, the



Insulating the Two-Wire Line



Development of Waveguide by Adding Quarter-Wave Sections

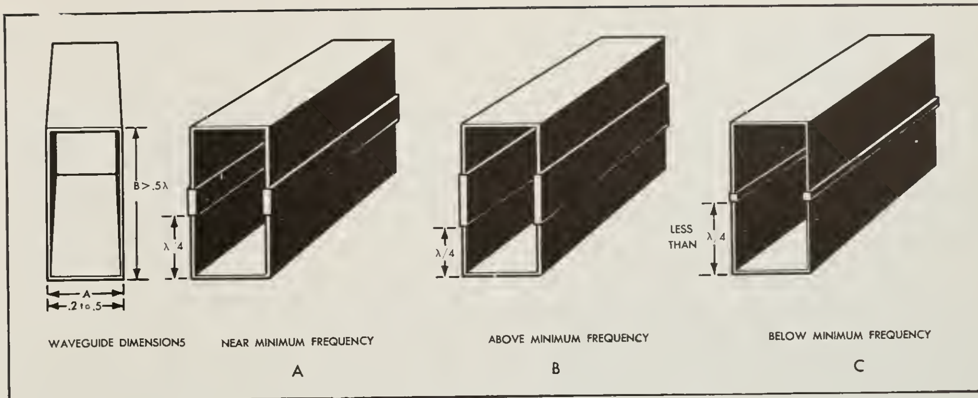
absorption of power by the dielectric material (insulators) causes them to look like a low resistance and capacity. The equivalent electrical circuit at higher frequencies is shown in diagram B in the illustration on page 10-3. For frequencies of 3000 mc and up, a better insulator than non-conducting procelain insulators must be used. A superior high frequency insulator for this purpose is a quarter wave section of RF line called a *metallic* insulator which was discussed in the preceding chapter. Such an insulator is shown at C. As there are no dielectric losses in a quarter wave section of an RF line, the impedance at the open end (the junction of the two wire line) is very high.

A metallic insulator can be placed anywhere along a two wire line. Diagram A above shows several on each side of a two-wire line. A point to note in this line is that the supports are a quarter wave at only one frequency. This limits the high efficiency of the two-wire line to only one frequency.

The use of several insulators results in improved conductivity of a two wire line when the sections are connected together. This connection is made between the two adjacent insulators through a switch, as you can see at B in this illustration. When the switch is open, both quarter-wave sections are excited by the main line. In this condition there will be standing waves on the quarter wave sections. When the switch is connected to the same place on each section, the relative phase relationship of the voltages at the connection will be the same for each section. In this condition the No. 1 section will be excited first by the generator. When the switch is closed, the No. 2 section will be partly excited by the No. 1 section through the switch connection. In this condition, less energy from the main line will be required to excite the No. 2 section. The parallel paths shown cause less resistance to exist along a given length of line, and energy is transferred with less copper loss.

When more and more sections are added to the line until each section makes contact with the next, the result is a rectangular box in which the line is at the center, as shown in C. The line itself is actually part of the wall of the box. The rectangular box thus formed is a waveguide.

EFFECT OF DIFFERENT FREQUENCIES ON A WAVEGUIDE. Previously it was stated that a quarter wave section is limited in operation to a certain frequency. However, when a solid wall of insulators is added, the section will operate at other frequencies. The waveguide shown at A at the top of the next page is the one just discussed. When the frequency being transferred by the guide is made higher, the quarter wave section must be shorter. These shorter wavelengths are easily accommodated if you assume this two wire line is made up of a wide bar or strip in each wall of the guide, as shown at B. The shorter distance remaining is the shorter quarter wave section. Thus, the wide bar shown is theoretically well insulated at any frequency higher than the one which creates the near minimum frequency at A. In reality there is a practical upper frequency limit at which this analogy is applicable. For example, when the bar is a half-wavelength across, the



Effect of Different Frequencies in the Waveguide

waveguide will be 4 quarter wavelengths across and may act as though there were two bars instead of one, as shown later in the chapter.

The next consideration is a frequency which is lower than the original frequency. Lowering the frequency in a given waveguide will lengthen the sections and narrow the bar. Beyond some lower frequency, this bar does not exist, because the quarter wave sections meet one another. At a still lower frequency, the sections become less than a quarter wave as in C above. A section less than a quarter wave is inductive. So the impedance across the place where the conducting bars belong is not a high resistance, but an inductance. The inductive reactance will dissipate the energy in the line very swiftly through high currents which flow back and forth in one place, as though the inductance were taking energy during one half cycle and then returning it to the line during the next half cycle.

It follows thus that in a waveguide there is a low frequency limit or cutoff frequency, below which the waveguide cannot transfer energy. The width of the waveguide at the cutoff frequency is equal to one half wavelength, since at this frequency the two quarter wave sections touch and add to one another. Mathematically, the cutoff wavelength is expressed by the equation,

$$\lambda c = 2B$$

where λc is the cutoff wavelength, and B is the long dimension of the rectangular cross-section.

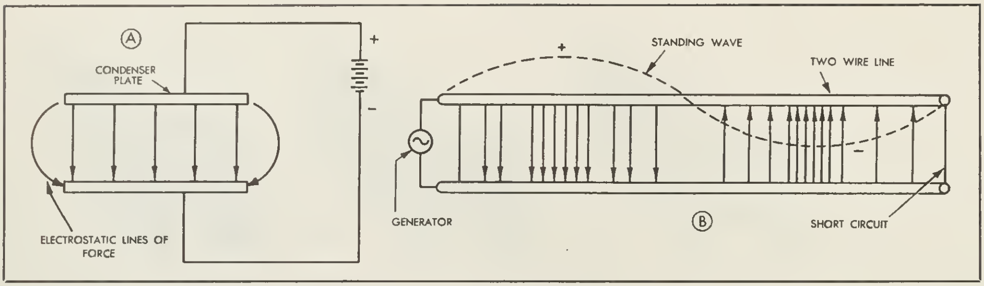
As cutoff occurs when width of a waveguide is below a half wavelength, most waveguides are made .7 wavelengths in the wide dimension to give a margin between the actual size and the size for cutoff. The other dimension is the distance between conductors, and similarly, as in the two-wire lines, is governed by the voltage breakdown potential of the dielectric, which is usually air. Width of .2 to .5 wavelengths are common.

ELECTROMAGNETIC FIELDS IN A WAVEGUIDE

A good working knowledge of the fields present in a waveguide is necessary if you want to use them intelligently. Energy in a waveguide is transferred by the electromagnetic fields, while currents and voltages merely aid in forming these fields. You should know such things as when a good current path is required or where the voltages will be high. As energy is normally introduced into and removed from the waveguide by the fields, you should know where the fields will exist. In any waveguide two fields—the electromagnetic and the electrostatic field—are always present.

The Electric Field

The existence of an electrostatic (electric field) indicates that there is a difference in the number of electrons between two points. An electric field consists of a stress in the dielectric field and is represented by arrows in diagrams. The simplest form of electrostatic field is the field which forms between the two parallel plates of a condenser as is shown at A at the top of page 10-6. When the top plate of the condenser is made positive by a battery, electrons move from the top plate and deposit themselves on the bottom plate. This immediately sets up a

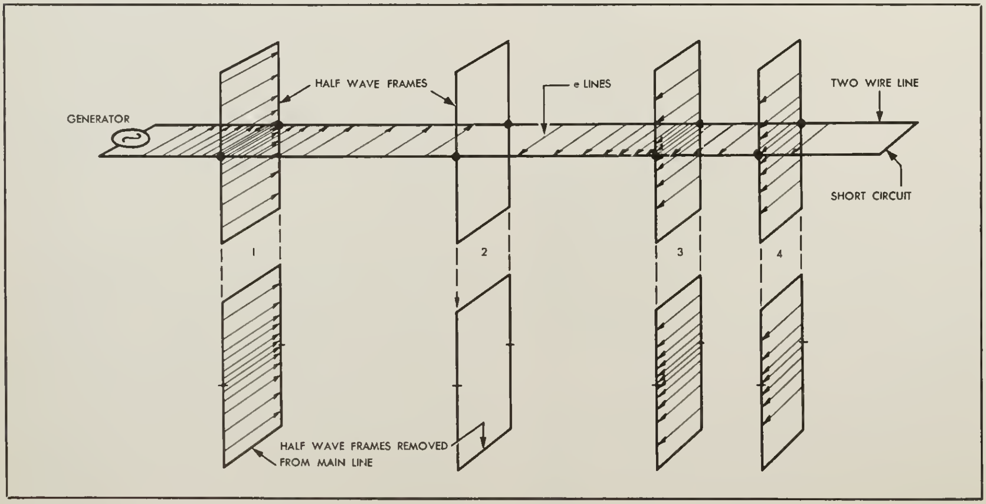


Electric Field between Condenser Plates and a Full-Wave Section of Two-Wire Line

stress in the dielectric between the plates. This stress is represented by arrows, the direction of which point from the more positive voltage point to the more negative voltage point. The amount of stress sometimes is indicated by the length of the arrow. A long arrow represents more stress than a short arrow. In representing electrostatic fields, the *number* of arrows indicates the strength of the field. In the case of the condenser, note that the arrows are evenly spaced across the area between the two plates. As the voltage across the plates is the same at all points, the electrostatic lines between the plates are evenly distributed. This set of lines forms an electrostatic field. This field is usually called the *electric field* and is abbreviated the *e-field*. The lines of stress are called the *e-lines*.

Notice above at B the two-wire transmission line which has an instantaneous standing wave of voltage applied to it. This line is equal to one wavelength. At the same time, part of it is positive, while another part is negative. The instantaneous electrostatic field (e-field) is the same at the negative and positive points, but the arrows representing each field point in opposite directions. The voltage along the line varies sinusoidally. Therefore, the density of the e-lines varies sinusoidally.

An easy way to show the development of the e-field in a waveguide is from a two wire line which has quarter wave insulators. The illustration below shows the two-wire line previously discussed with several double quarter wave insulators, or half wave frames used as insulators.



Magnitude of Fields on Half-Wave Frames Vary with Strength of Field on Main Line

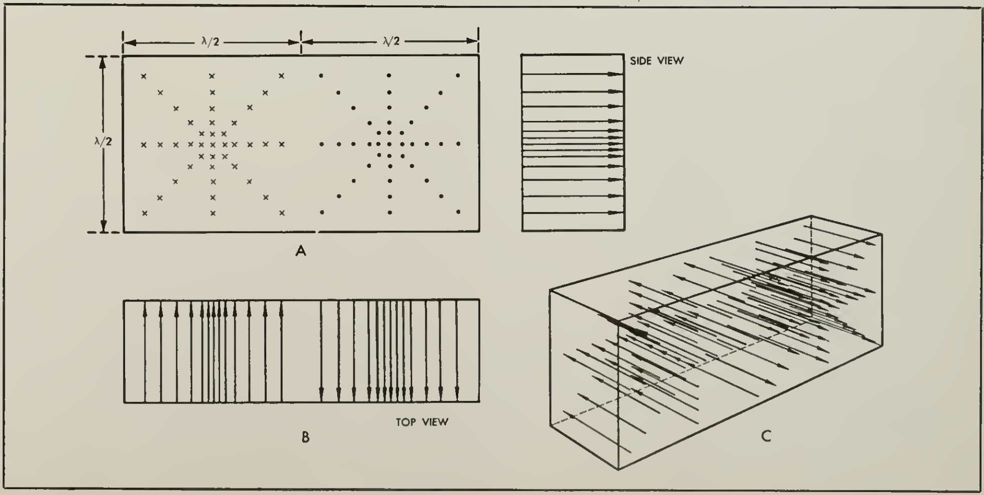
The e-field on the main line is the same as that in the transmission line illustrated at the top of page 10-6. The half-wave frames located at points of high voltage (strong e-field) will have a strong e-field across them. The half-wave frames located at a point voltage minimum will have no e-field on them. Frame No. 1 in the illustration at the bottom of page 10-6 is an example of an insulator which has a strong e-field across it. Each frame is shown separately below the main line for a clearer view. Frame No. 2 is at a zero voltage point, so it will have no field on it. Frame No. 3 also has a strong field, but its polarity is reversed. Frame No. 4 has a weaker field on it due to being at a lower voltage point on the main line. The picture shown is a build-up to the three dimensional aspect of the full e-field in a waveguide.

The illustration below shows the e-field in an actual waveguide. This is the field which results when an infinite number of quarter wave sections are connected to the line to form a rectangular box. The e-field is strong at one quarter and three quarter distances from the shorted end, but becomes weaker at the sine rate toward the upper and lower walls and toward the ends and center. Again the phenomenon of wavelength is present, as shown at A below. You should realize, of course, that this is an instantaneous picture taken at the time the standing wave of voltage is at its peak. At other times, the voltage and E-field varies from zero to the peak value, reversing direction every alternation of the applied voltage.

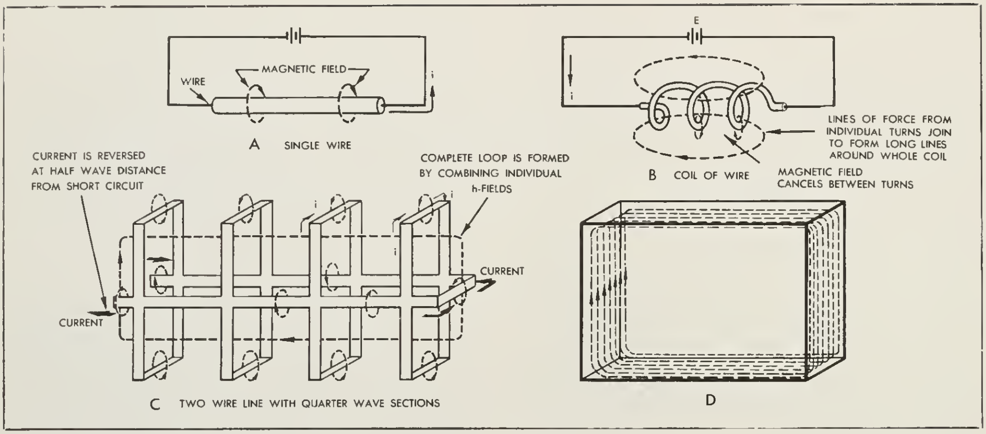
Certain boundary conditions must exist in order for propagation to occur in a waveguide. The principal one is that there must be no electric field tangent to the walls of the guide. This is satisfied by the e-field diminishing to zero at the top and at the bottom of the guide by natural RF line action, while at other places, the field is perpendicular to the walls.

The Magnetic Field

The second field which must always exist in a waveguide is the magnetic field. The magnetic lines of force which make up the magnetic field are caused by the *movement* of electrons in the conducting material. All the tiny magnetic forces exerted by the individual moving electrons add together and form a large force around the conductor. The presence of the force is shown by closed loops around the single wire in A of the diagram at the top of page 10-8. The h-line must be a continuous closed loop in order to exist. The line forming the loop is a magnetic line of force or an *h-line.* All the lines associated with current are collectively called a magnetic field or an *h-field.* The strength of the h-field varies directly with the current. Each h-line has a certain direction. You can determine this direction by the left-hand rule. The strength of the h-field is indicated by the number of h-lines in a given area.



E-Field in Actual Waveguide

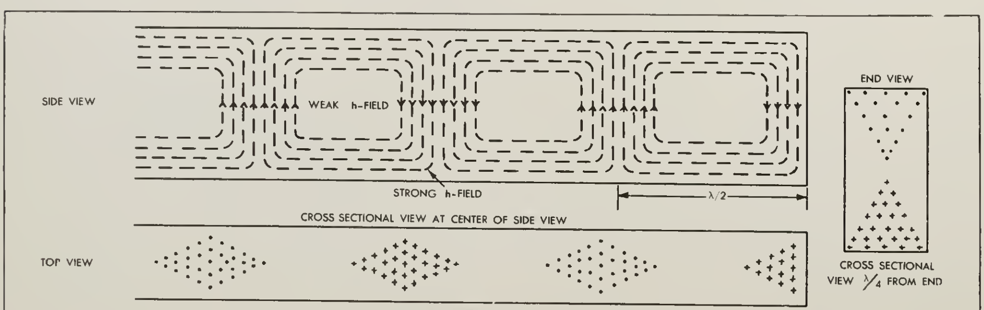


Development of Magnetic or H-Field in the Waveguide

Although h-lines encircle a single straight wire, they behave differently when the wire is formed into a coil as is shown at B above. In a coil the individual h-lines tend to form around each turn of wire, but in doing so take opposite directions between adjacent turns. This causes cancellations and results in zero field strength between the turns. But inside and outside the coil, the directions are the same for each h-field. Therefore, the fields here join and form a continuous h-line around the entire coil as shown.

Similar action also takes place in a waveguide. In diagram C above a two-wire line with quarter wave section is shown. Currents flow in the main line and in the quarter wave section. The current direction produces the individual h-lines around each conductor as shown. When a large number of sections exist, the field cancels between the sections but their directions are the same inside and outside the waveguide. At half wave intervals on the main line, current will flow in opposite directions. This produces h-lines loops having opposite direction. At C current at the left end is opposite to the current at the right end. The individual loops on the main line are opposite in direction. All around the framework, they join such that the long loop shown at D is formed. Outside of the waveguide, the individual loops cannot join to form a continuous loop. Thus there is no magnetic field outside of a waveguide.

Below in the illustration showing a conventional presentation of the magnetic field in a waveguide three half wavelengths, note that the field is strongest at the edges of the waveguide. This is where the current is the highest. The current is lowest at the center of each set of loops because there the standing wave of current is



Magnetic Field in Waveguide Three Half Wavelengths Long zero at all times. In the illustration the picture shown represents an instantaneous condition. During the peak of the other half cycle of AC input, all field directions are reversed. An instantaneous picture of this condition would show the fields reversed at half-wave intervals, since the current in the long line is reversed over half-wave distances.

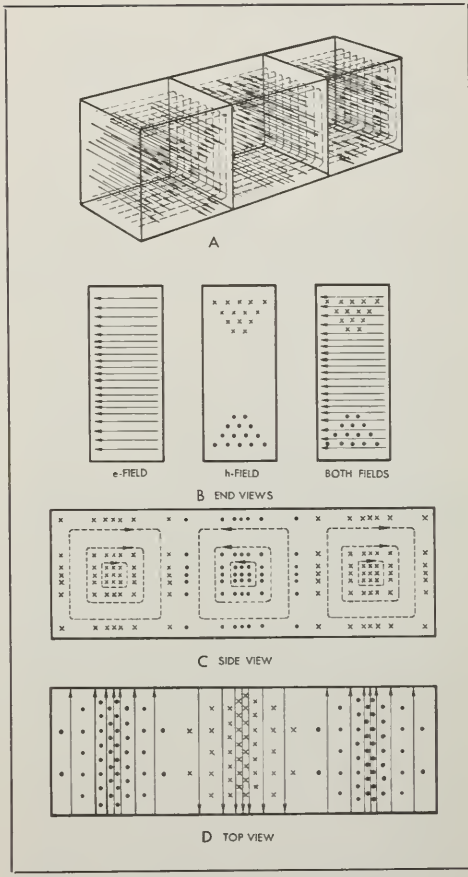
A second boundary condition necessary for electromagnetic fields to transfer power is satisfied by the configuration of the magnetic field. This condition requires that at the surface of the waveguide there be no perpendicular component of the magnetic field. Since all the h-lines are parallel to the surface, this condition is satisfied.

Electric and magnetic fields exist simultaneously in the waveguide. In fact, the h-field causes a current which in turn causes a voltage difference. This causes an e-field and it in turn causes a current which causes an h-field and so on. One field is dependent on the other as energy is continually transferred from one field to the other.

At the right the conventional picture of both fields in the waveguide at A. Since this picture is rather complicated, the presence and direction of the field is usually indicated in more simple diagrams, such as those shown in B, C, and D. In these diagrams the number of e-lines in a given area indicates the strength of the electrostatic field, while the number of h-lines in any given cross-section indicates the strength of the magnetic field in that area.

The field configuration shown in this illustration represents only one of the many ways in which fields are able to exist in a waveguide. Such a field configuration is called a *mode* of operation. In the case of the rectangular waveguide illustrated, the configuration is known as the *dominant* mode, since it is the easiest one to produce. Other higher modes—that is, different field configurations—may occur accidentally or may be caused deliberately in a waveguide.

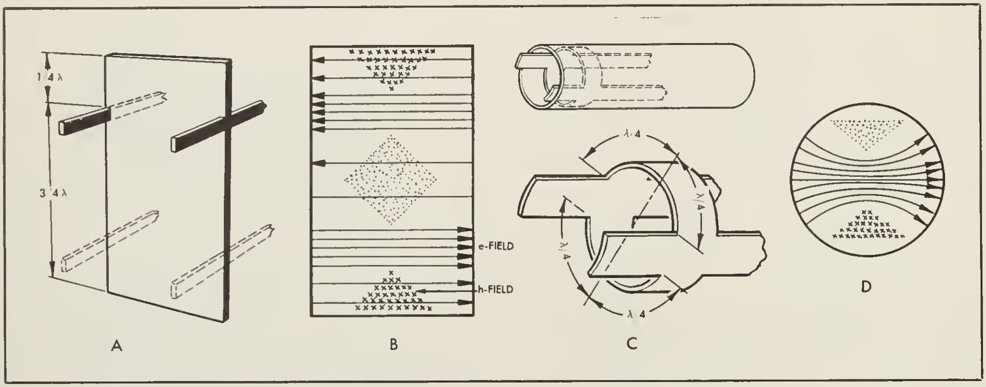
An example of another field configuration is developed in the illustration at A on page 10-10. If the size of this waveguide is doubled over that of the waveguide shown in the previous illustrations, the cross-section will be a full wave rather than a half wave. The two-wire conductor can be assumed to be a quarter wave down from the top (or a quarter wave up from the bottom). The remaining distance to the bottom is $\frac{3}{4}$ wave.



Conventional Picture of Both Fields in Waveguide

A $\frac{3}{4}$-wave section has the same high impedance input as the quarter wave section. Thus the two-wire line is properly insulated and will transfer energy. The field configuration will show a full wave across the wide dimension, as you can see at B in this illustration.

This field configuration can be applied to a circular waveguide. The two conductors shown at C are assumed to be part of the waveguide wall. The remaining part of the wall forms the quarter wave sections. The quarter wave section insulates the two conductors. This makes it possible to transfer energy with minimum losses. The resulting field configuration shown at D is the dominant mode for a waveguide with a circular cross section.



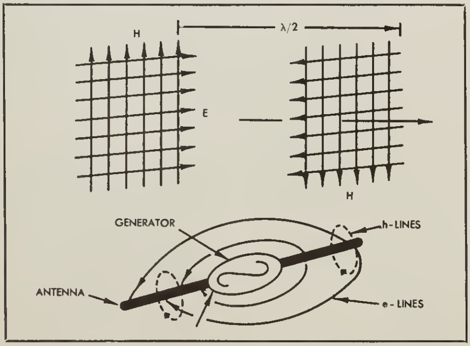
Another Field Configuration in the Rectangular Waveguide

ANALOGY OF WAVEGUIDE ACTION BY ELECTRIC WAVES

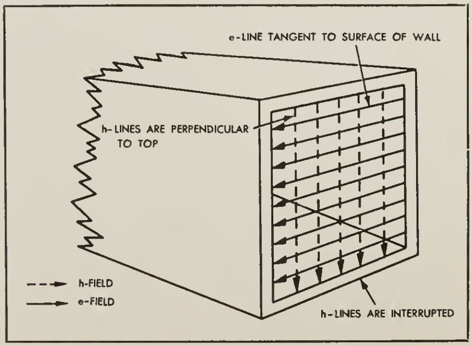
A somewhat different analogy involving waveguide action deals with the field rather than with current and voltages. This analogy is somewhat more exact than the previous explanation which dealt with voltages and current in two-wire lines, for power flow in this case is assumed to be in the fields rather than in the conductors.

In a waveguide the fields are the same as those radiated into space by an antenna. Below notice the illustration showing a small portion of a field which is radiated into space from an antenna. In it the electrostatic line of force or e-lines are parallel to the antenna, and the electromagnetic line of force or h-lines are perpendicular to the antenna. They move away from the antenna at the speed of light. At each half cycle the polarity is reversed. Therefore, at half-wave intervals, the fields are in opposite direction (or polarity). Although only a small part of the total field is shown, actually the e-lines and h-lines form huge closed loops after they leave the antenna.

As previously mentioned, the energy which moves through a waveguide and the energy which is radiated by an antenna are both the same form of electromagnetic radiation. Nevertheless, the field configuration shown in the diagram just below cannot exist in a waveguide because it does not satisfy the required boundary conditions. First, there cannot be any e-lines tangent to the surface of the walls. Since the e-lines are evenly distributed across the area, some will be across the top and bottom wall. This causes the voltage to short out and this in turn causes the e-lines to vanish. Other e-lines which are pushed up to the wall by the repulsion between e-lines likewise short out. This shorting out is accumulative and eventually removes the entire e-field.

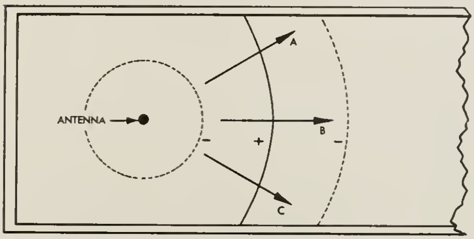


Small Portion of Field Radiated into Space by an Antenna



Fields in a Waveguide Must Satisfy Boundary Conditions to be Radiated

The second boundary condition which must be satisfied is that there must be no component of the magnetic field perpendicular to the wall. Note again in the illustration that the h-lines are parallel to the side walls which is correct—but are perpendicular to the bottom, which cannot be; so h-lines of this type cannot exist in the waveguide either. Furthermore, an h-line cannot exist unless it is a closed loop.



How Radiation Fields are Made to Fit a Hollow Pipe

When a small antenna is placed in the waveguide and excited at an RF frequency, both positive and negative half-cycles are radiated as shown just above. The wavefront produced is like an expanding circle. The part which travels in the direction of arrow B goes straight down the waveguide and is quickly attenuated, as previously described. However, the part of the wavefront which travels in the direction of arrow A is reflected from the wall. The wall is a short circuit and causes the wavefront to be reflected in reverse phase. Meanwhile, the wavefront which travels in direction C is reflected from the other wall and proceeds in opposite phase. Thus, the radiation fields are contained in the waveguide.

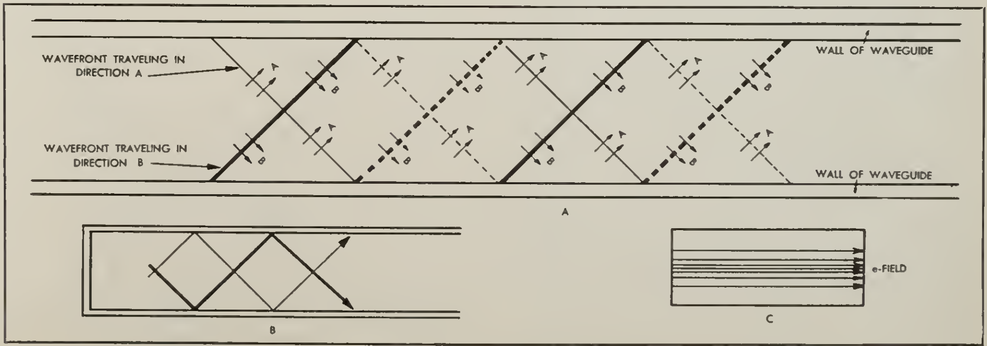
Path of Wavefronts in a Waveguide

In the side view of the waveguide at A below the light solid and broken lines represent the wavefront going in direction A. The heavy lines and dashes represent the wavefront going in direction B. Note that all parts of the wavefront A are traveling upward at an angle across the guide. Wavefront B is traveling at the same angle but downward.

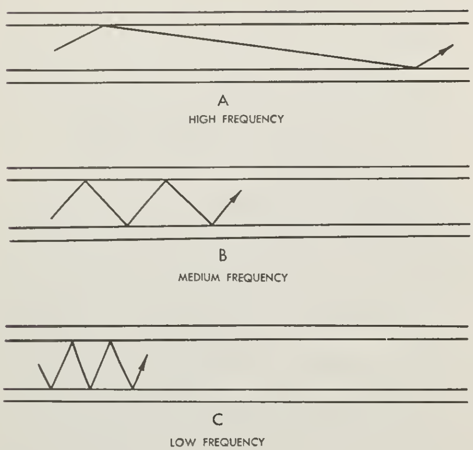
When the wave travels in this fashion in a waveguide, propagation is possible. In understanding what happens, note that the positive wavefront (represented by solid lines) occurs simultaneously throughout the center of the guide. These fronts add and cause a maximum voltage to occur at the center. (The e-field is shown maximum at the center in diagram C.) The negative wavefront adds in the same manner as the positive wavefront. When the negative wavefront meets the positive wavefront at the walls, the two wavefronts cancel each other, making the total voltage equal to zero. This verifies the e-field condition shown at C. With the e-field zero at the edges, it is possible for the e-field to exist in the waveguide.

Crossing Angle

The angle at which a wavefront crosses a waveguide is a function of the wavelength and the cross-sectional dimension of the waveguide. At some intermediate frequency the reflection is as shown at B on the next page. But as the frequency increases, the angle of incidence becomes less and the signal travels farther before it reaches the other side (see A). At lower frequencies, the wavefront crosses the guide at more nearly right angles to the walls. At some frequency, the angle will be 90°. At this point, the wave travels back and forth across the guide



Paths of Wavefronts in Waveguide



Angle at which Fields Cross Waveguide Varies with Frequency

until the energy is dissipated by the resistance of the walls of the guide. At this frequency, the distance from side to side is one-half wavelength for the waveguide. At the cutoff frequency, the attenuation is a linear function of length and is very high.

The velocity of propagation of a wave along a two-wire line is less than its velocity in air. The same is true in a waveguide. Movement of a wave along a two-wire line is slower than its movement in air because of the retarding effect of the DC resistance, the conductors, and conductance of the insulation. In the waveguide, the lower velocity is due to the way the field travels. As shown in the above illustration at C, the path of a wavefront at a relatively low frequency is along the zig-zag arrow at the velocity of light. But due to the long path, the wavefront actually travels very slowly along the waveguide. In the same illustration at A, the frequency is higher, and the wavefront or the group of waves actually travel a given distance in less time than those at C.

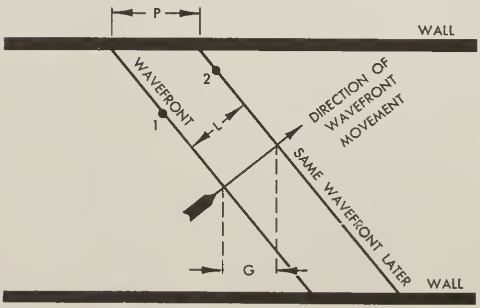
The axial velocity of a wavefront or a group of waves is called the *group velocity.* The relationship of the group velocity to diagonal velocity causes an unusual phenomenon. The velocity of propagation appears to be greater than the speed of light. As you can see in the illustration to the right during a given time, a wavefront will move from point one to point two, or a distance L at the velocity of light (V_L). Due to this diagonal movement (direction of the arrow), during this time the wavefront has actually moved down the guide only the distance G, which is necessarily a lower velocity. This is called group velocity (V_g). But if an instrument were used to detect the two positions at the wall, they would be the distance P apart. This is greater than the distance L or G. The movement of the contact point between the wave and the wall is at a greater velocity. Since the phase of the RF has changed over the distance P, this velocity is called the phase Velocity (V_p), The mathematical relationship between the three velocities is stated by the equation

$$V_L = \sqrt{V_p V_g}$$

where

V_L = *velocity of light* $= 3 \times 10^8$ *meters/second*
V_p = *Phase Velocity*
V_g = *Group Velocity*

This equation indicates that it is possible for the phase velocity to be greater than the velocity of light. As the frequency decreases, the angle of crossing is more of a right angle. In this condition the phase velocity increases. For measuring standing waves in a waveguide, it is the phase velocity which determines the distance between voltage maximum and minimum. For this reason, the wavelength measured in the guide will actually be greater than the wavelength in free space. From a practical standpoint, the different velocities are related in the following manner: If the RF frequency being propagated is sine wave modulated, the modulation envelope will move forward through the waveguide at the group velocity, while the individual cycles of RF energy will move forward through the modulation envelope at the



Relation of Phase, Group, and Wavefront Velocity phase velocity. If the modulation is a square wave, as in radar transmissions, again the square wave will travel at group velocity, while the RF waveshape will move forward within the envelope. Since the standing wave measuring equipment is affected by each RF cycle, the wavelength will be governed by the rapid movement of the changes in RF voltage. Since intelligence is conveyed by the modulation, the transfer of intelligence through the waveguide will be slower than the speed of light, as is the case in other types of RF lines.

Because of the way the fields are assumed to move across the waveguide, it is possible to establish a number of trigonometric relationships between certain factors. As shown below the angle that the wavefront makes with the wall, (angle θ) is related to the wavelength and dimension of the guide and is equal to,

$$\cos \theta = \frac{\lambda}{2B}$$

where λ is the wavelength in free space of the signal in the guide, and B is the inside wide dimension of the guide. The group velocity (V_g) is related to the velocity of light (V_L) as follows:

$$\frac{V_g}{V_L} = \sin \theta = \sqrt{1 - \left(\frac{\lambda}{2B}\right)^2}$$

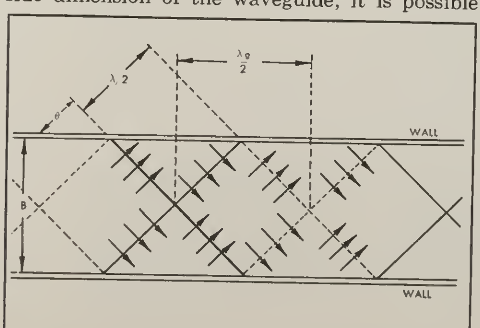
Further, since it is possible to measure the wavelength in the guide (λg), the wavelength in space is equal to,

$$\frac{\lambda g}{\lambda} = \frac{1}{\sin \theta} = \frac{1}{\sqrt{1 - \left(\frac{\lambda}{2B}\right)^2}}$$

Solving this for λ, the equation becomes equal to,

$$\lambda = \frac{2B\lambda g}{\sqrt{\lambda g^2 + 4B^2}}$$

After measuring the wavelength and the inside dimension of the waveguide, it is possible



Trigonometric Relations Exist between Factors Indicated

to calculate most other quantities associated with the waveguide.

Numbering System of the Modes

The normal configuration of the electromagnetic field within a waveguide is called the *dominant* mode of operation. The mode developed for the rectangular waveguide, as was explained before, is the dominant mode of operation. The dominant mode for the circular waveguide was also shown in a previous illustration. A wide variety of higher modes are possible in either type of waveguide. The higher modes in the rectangular waveguide are seldom used in radar, but some of the higher modes in the circular waveguide are useful.

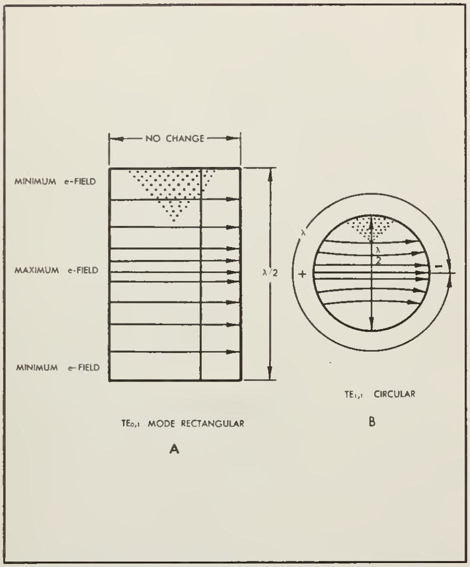
For ease in identifying modes, any field configuration can be classified as either a transverse electric mode or a transverse magnetic mode. These modes are abbreviated TE or TM respectively.

In a transverse electric mode, all parts of the electric field are perpendicular to the length of the guide and no e-line is parallel to the direction of propagation. The TE mode is sometimes called the H-mode.

In a transverse magnetic mode, the plane of the h-field is perpendicular to the length of the waveguide. No h-line is parallel to direction of propagation. This mode is sometimes called an E-mode.

It is interesting to note from these definitions that the wavefront in free space or in a coaxial line is a TEM mode, since both fields are perpendicular to the direction of propagation. This mode cannot exist in a waveguide.

In addition to the letters TE or TM, subscript numbers are used to complete the description of the field pattern. In describing field configurations in rectangular guides, the first small number indicates the number of half-wave patterns of the transverse lines which exist along the short dimension of the guide through the center of the cross-section. The second small number indicates the number of transverse half-wave patterns that exist along the long dimension of the guide through the center of the cross-section. For circular waveguides the first number indicates the number of full waves of the transverse field encountered around the circumference of the guide. The second number indicates the number of half-wave patterns that exist across the diameter.



How to Count Wavelengths for Numbering Modes

Counting Wavelengths for Measuring Modes

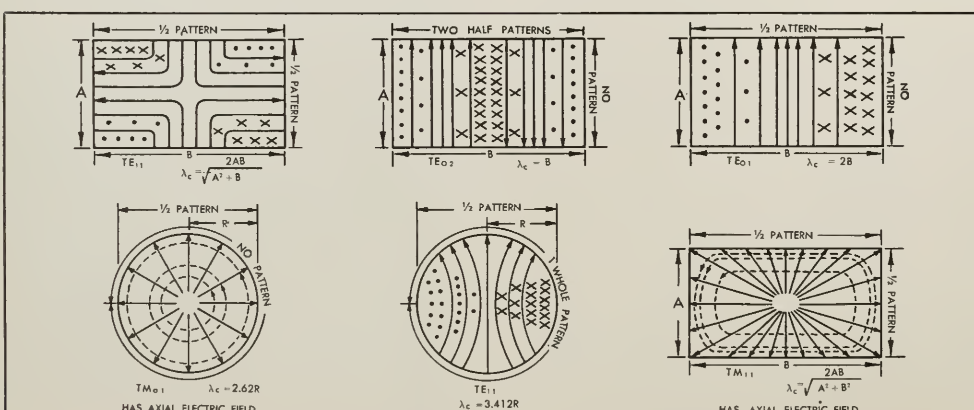
In the rectangular mode illustrated above at A, note that all the electric lines are perpendicular to the direction of movement. This makes it a TE mode. In the direction across the narrow dimension of the guide parallel to the e-line, the intensity change is zero. Across the guide along the wide dimension, the e-field varies from zero at the top through maximum at the center to zero on the bottom. Since this is one-half wave, the second subscript is one. Thus, the complete description of this mode is TE_{01}.

In the circular waveguide at B, the e-field is transverse and the letters which describe it are TE. Moving around the circumference starting at the top, the fields goes from zero, through maximum positive (tail of arrows), through zero, through maximum negative (head of arrows), to zero. This is one full wave, so the number is one. Going through the diameter, the start is from zero at the top wall, through maximum in the center to zero at the bottom, one-half wave. The second subscript is one. The complete designation for the circular mode becomes $TE_{1,1}$.

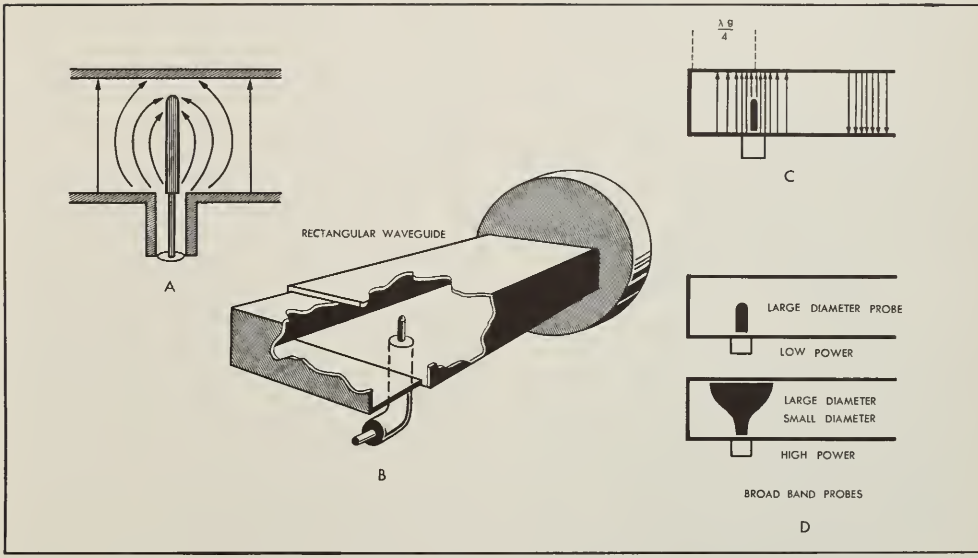
Several circular and rectangular modes are possible. On each diagram illustrated below you can verify the numbering system. Note that the magnetic and electric fields are maximum in intensity in the same area. This indicates that the current and voltage are in phase. This is the condition which exists when there are no reflections to cause standing waves. In previous examples in which fields were developed, the fields were out of phase because of a short circuit at the end of the two-wire line.

INTRODUCING FIELDS INTO A WAVEGUIDE

A waveguide, as was explained before, is a single conductor. Therefore, it does not have the two connections which ordinary RF lines have, and it is necessary to use special devices to put energy into a waveguide at one end and to remove it from the other. In a waveguide, as with



Various Modes in Waveguides



Exciting the Waveguide with Electric Field

many other electrical networks, reciprocity exists in any excitation system—that is, energy may be transferred either to the waveguide or from the waveguide with the same efficiency.

Waveguides may be excited by three principal methods, namely, electric fields, magnetic fields and electromagnetic fields.

Exciting with Electric Fields

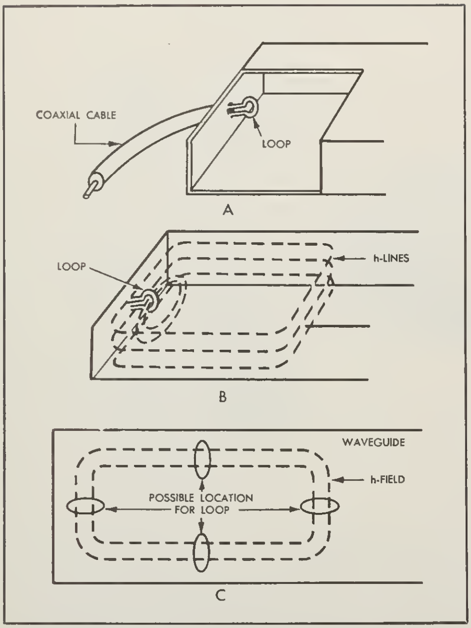
When a small probe or antenna is placed in a waveguide and fed with an RF signal, current will flow in the probe and set up an electrostatic field such as shown above at A. This causes the e-lines to detach themselves from the probe and to form in the waveguide. When the probe is located in the right place, a field having considerable intensity will be set up. The best place to locate the probe is in the center, parallel to the narrow dimension and one quarter wavelength away from the shorted end of the guide, as shown at C. Note here that the field is strongest at the quarter wave point. This is the point of maximum coupling between the probe and the field. Of course, the probe will work equally well in the center of any unidirectional field. For example, a ¾-wave distance from the shorted end will also be a good spot to place the probe.

Usually, the probe is fed with a coaxial cable. In comparison with the waveguide, this cable is extremely short. This insures that the greatest benefit will be derived from the waveguide. Impedance matching between the coaxial cable and the waveguide is accomplished by varying the distance from the probe to the end of the waveguide (by moving the shorted end) and by varying the length of the probe (see above). A mismatch will cause unwanted reflections in the waveguide.

The degree of excitation can be reduced by reducing the length of the probe, moving it out of the center of the e-field, or shielding it. Where it is necessary to vary the degree of excitation frequently, the probe is made retractable and the end of the waveguide fitted with a movable plunger. In airborne radar systems, the position of the probe and the end piece is often predetermined by the factory and fixed permanently.

In pulse-modulated radar systems there are wide side bands on each side of the carrier. In order that a probe feeding system does not discriminate too sharply against frequencies which differ from the carrier frequencies, *wide-band* probes are often used. This probe is large in diameter and is conical or door knob in shape. A conical probe is capable of handling high powered signals.

The same kind of probe is used when a probe is used to take energy out of the guide and deliver it to the coaxial cable.



Excitation with Magnetic Fields

Excitation with a Magnetic Field

Another way of exciting a waveguide is by setting up a magnetic (h) field in the waveguide. This can be accomplished by a small loop which carries a high current and placing the loop in the waveguide. This is what happens. A magnetic field builds up around the loop. The field expands and fits the guide. If the frequency of the current is correct, energy will be transferred from the loop to the waveguide. A loop for transferring energy into a guide is shown at A and B in the above illustration. Notice that the loop is fed by a coaxial cable. The location of the loop for optimum coupling to the guide is at the place where the magnetic field which is to be set up is of greatest strength. There are a series of places where this is true. Several are shown at C.

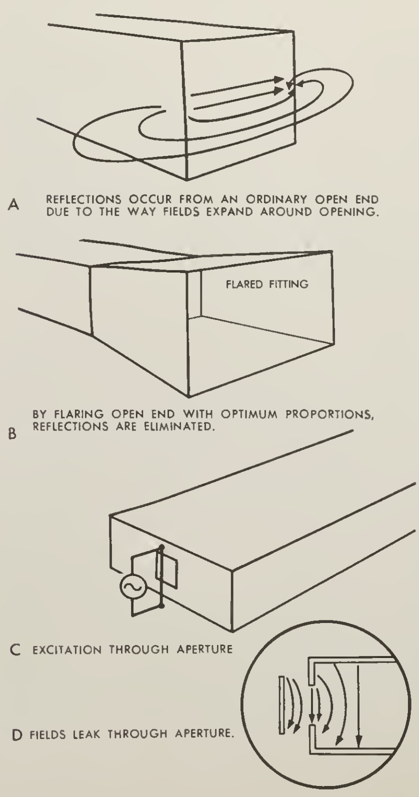
When less coupling is desired, you can rotate or move the loop until it encircles a smaller number of lines of force.

When an excitation loop is used in radar equipment, its proper location is often predetermined and fixed either during construction or final tuning at the factory. In test or laboratory equipment, the loop is often made adjustable.

When a loop is introduced in a guide in which an h-field is present, a current will be induced in the loop itself. When this condition exists, the loop will take energy out of the waveguide as well as put energy into it.

Excitation with Electromagnetic Fields

After learning what fields are like, you might think that a good way to either excite the waveguide or to let energy out of it is simply to leave the end open. However, this is not the case, for when energy leaves a guide, fields form around the end of the guide and cause an impedance mismatch as shown below at A. In other words, reflections and standing waves would result if the end were left open. Thus, simply leaving the end open is not an efficient way of letting energy out of the waveguide.



Excitation with Electromagnetic Fields

In order for energy to move smoothly in or out of a guide, the opening of the guide may be flared like a funnel as shown at B. This makes the guide similar to a V-type antenna. The *funnel* in effect eliminates reflection by matching the impedance of free space to the impedance of the waveguide. When the mouth of the funnel is exposed to electromagnetic fields, they enter the funnel where they are gradually shaped to fit the waveguide. The funnel is directional in characteristic. It sends or receives the greatest amount of energy from in front of the opening.

Another method for either putting energy into or removing it from waveguides is through slots or openings. This method is sometimes used when very loose coupling is desired. In this method energy enters the guide through a small aperture, as you can see at C. Any device which will generate an e-field may be placed near the aperture and the e-field will expand into the waveguide. A single wire is shown at D. On it e-lines are set up parallel to the wire due to the voltage difference between different parts of the wire. The e-lines, in expanding, will exist first across the aperture, then across the interior of the waveguide. If the frequency is correct and the size of the aperture properly proportioned, energy will be transferred to the waveguide with a minimum of reflections.

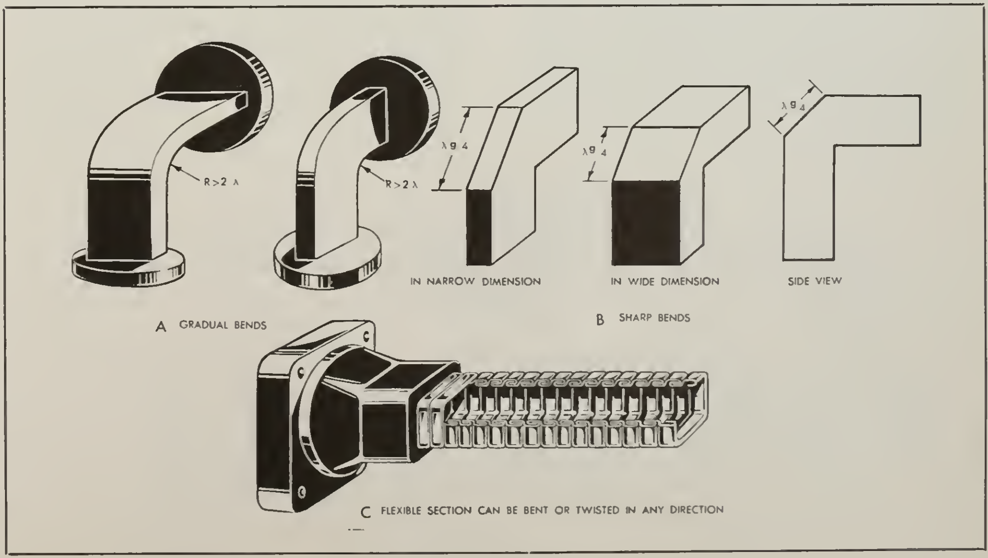
BENDS, TWISTS, JOINTS AND TERMINATIONS

In order for energy to move from one end of a waveguide to the other without reflections, the size, shape, and dielectric material of the waveguide must be constant throughout its entire length. Any abrupt change in its size or shape results in reflections. Therefore, if no reflections are desired, any change in the direction or the size of the waveguide must be gradual. When it is necessary that the change in direction or size be abrupt, then special devices, such as bends, twists, joints, or terminations, must be used.

Bends

Waveguides may be bent in several ways to avoid reflections. One is to make the bend gradual. It must have a radius of bend greater than two wavelengths in order to minimize any reflection. Some bends may be 90° bends. Other bends may be greater or less than 90°, depending upon the requirements of the system. Still another type of bend is the sharp bend.

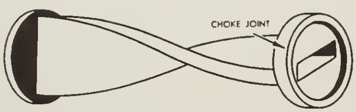
A bend can be made in either the narrow or wide dimension of a guide without changing the mode of operation. In a sharp 90° bend, normally reflections will occur. To avoid this, the guide is bent twice at 45°-one quarter wave apart. The combination of the direct reflection at one bend



Types of Bends

and the inverted reflection from the other bend will cancel and leave the fields as though no reflection had occurred.

To permit using any special bend which an installation might require, sections of a waveguide are often made flexible. These sections can be bent or twisted in any desired direction. They consist of a spiral wound ribbon of brass. In cross section the winding is exactly the same size as a waveguide. The entire assembly is like a spiral spring in that it can be bent or twisted into any desired shape. As skin effect keeps the current at the inner surface of the waveguide, the inside surfaces of the flexible section are chromium plated. This provides for maximum current conductivity. The outside of the section is covered with rubber. This gives the section flexibility and at the same time makes it both air- and watertight.



Twisted Section of Waveguide Rotates the Field with Minimum Reflections

Rotating the Field

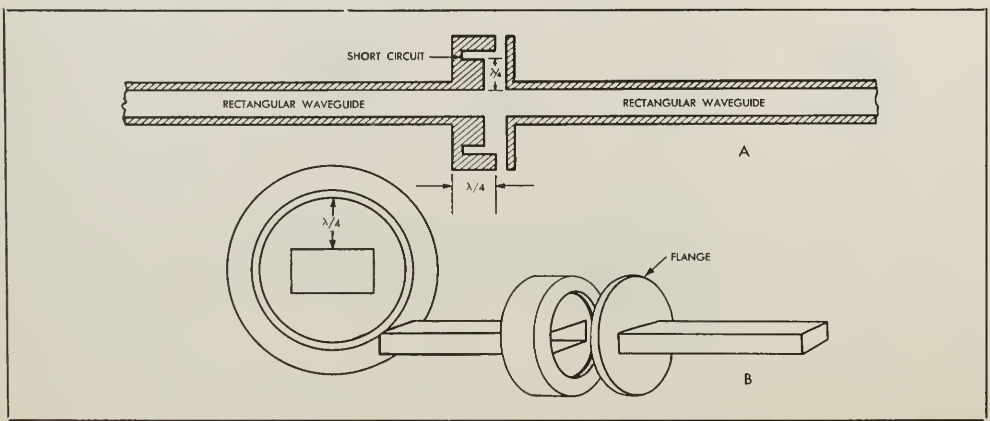
Sometimes it is desired to rotate the electromagnetic fields so that they are in the proper direction for matching. This may be accomplished by twisting the waveguide as shown above. The twist should be gradual and extend over two wavelengths or more to prevent excessive reflections. Flexible sections also are used to rotate fields.

Joints

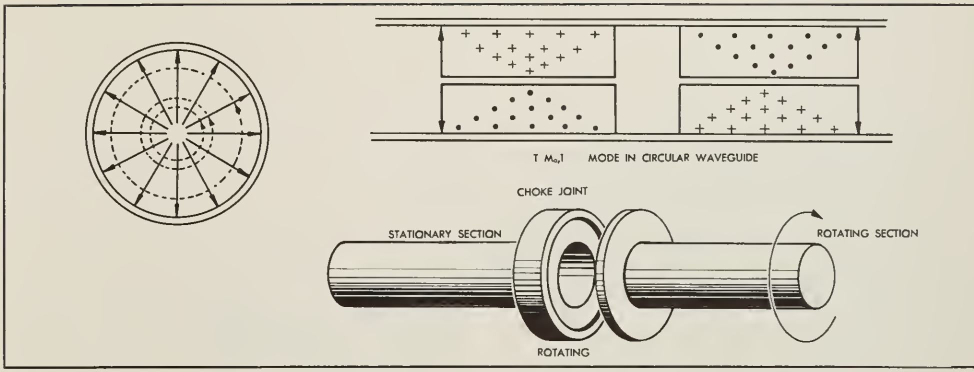
Since it is impossible to mold an entire waveguide system in a radar set into one piece, it is necessary to construct it in sections and then to connect the sections together by joints. There are three main types of joints: These are the permanent, the semi-permanent, and the rotating joints.

On the surface it would appear that joining two waveguide sections together would only require that the sections be the same size and fit tightly at the joint. However, irregularities at the joints set up standing waves and allow energy to escape. One kind of joint which affords a good connection between the parts of a waveguide and which has very little effect on the fields is the permanent type. This joint is made at the factory. When it is used, the waveguide sections are machined within a few thousandths of an inch and then welded together. The result is a hermetically sealed and mirror smooth joint.

Where it is necessary that sections be taken apart for normal maintenance and repair, it is impractical to use a permanent joint. To permit portions of the waveguide to be taken apart, they are commonly connected together with semi-permanent joints. The most common type of semi-permanent joint is the choke joint. A cross-sectional view of a choke joint is shown in the illustration below at A. It consists of two flanges which are connected to the waveguide at the center. The right-hand flange is flat, and the one at the left is slotted a quarter wave deep at a



Choke Joints Keeps RF Fields Inside Waveguide

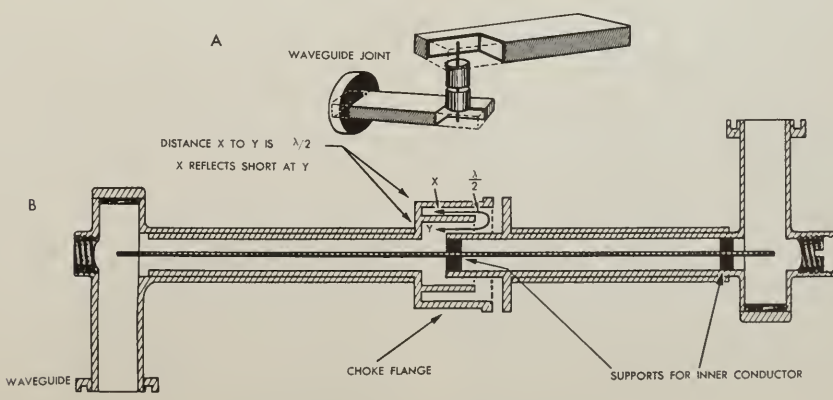


Rotating Joint and TM_{01} Mode in Circular Waveguide

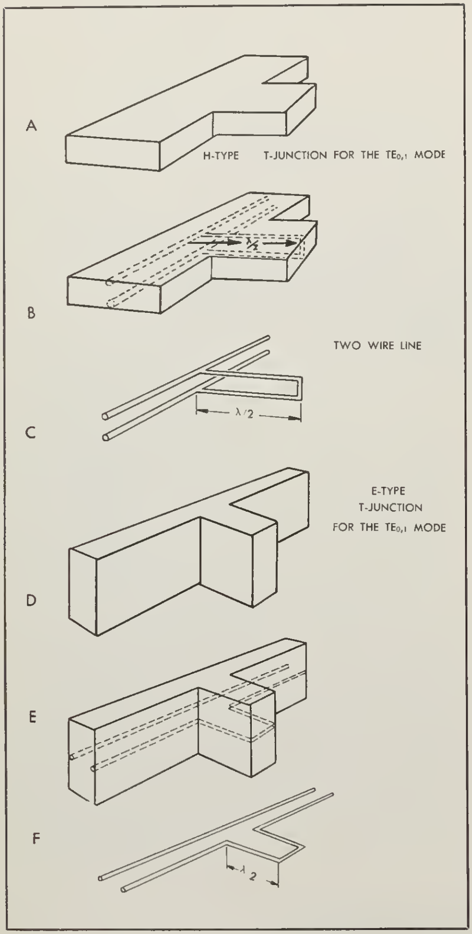
distance a quarter wave from the point where the walls of the guide are joined. The quarter wave slot is shorted at the end. The two quarter waves together become a half wave and reflect a short circuit at the place where the walls are joined together. Electrically, this creates a short circuit at the junction of the two waveguides. The two sections actually can be separated as much as a tenth of a wavelength without excessive loss of energy at the joint. This separation allows room to seal the interior of the waveguide with a rubber gasket for pressurization. The quarter wave distance from the walls to the slot is modified slightly to compensate for the slight reactance introduced by the short space and the open circuit from the slot to the periphery of the flange.

The name choke joint is said to come from the similarity between the action of this joint on RF fields and the action of an RF choke in a power supply lead. An RF choke keeps RF in the circuit where it belongs. Similarly, the choke joint keeps the electromagnetic fields in the waveguide where they belong. The loss introduced by the well designed choke is less than .03 db, while an unsoldered permanent joint, well machined, has a loss of .05 db or more.

Rotating joints are usually required in an airborne radar system where the transmitter is stationary and the antenna is rotatable. A simple method for rotating part of a waveguide system is by using a mode of operation that is symmetrical about the axis as shown above. This requirement is met by using a circular waveguide and



Rotating Joint with Rectangular Waveguide



T-Junctions

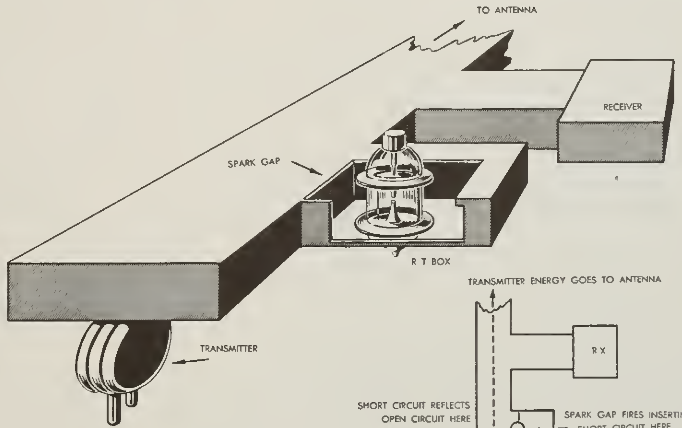
a mode such at $TM_{0,1}$. In this method a choke joint may be used to separate the sections mechanically and to join them electrically. With this method the waveguide rotates but not the field. The continuous field produced by the fixed field minimizes reflections. Although this method is satisfactory, current airborne equipment uses rectangular waveguides. This requires a different system for connecting the sections of the waveguide. For mechanical reasons the rotating joint must be circular and involves the use of a coaxial cable as shown at the bottom of page 10-19. This cable provides axial symmetry of the fields and the circular cross-section for rotation. In the rotating joint, a probe forms the end of the center conductor of the coaxial cable. The probe takes energy from one waveguide. Then it is conducted through the coaxial cable and delivered through another probe to the other waveguide. The probe in the other waveguide is the other end of the center conductor. The center conductor remains stationary with respect to one waveguide and rotates with respect to the other. To make the rotating electrical connection, the outer conductor can either be fitted with sliding contacts or with the tubing by a half wave slot. The slot which is shorted at the end reflects a short at the junction of the two outer conductors. In this method, no mechanical contact is required between the two sections of the outer conductor. The inner conductor of the coaxial cable is supported by insulating washers.

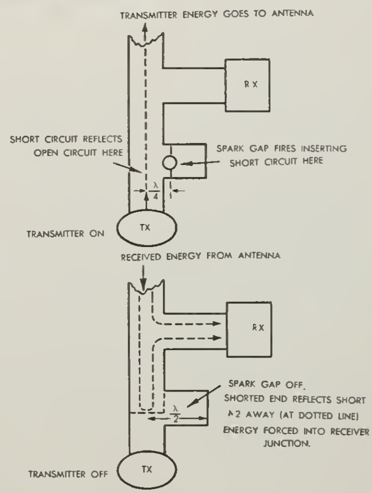
T-Junctions

Sometimes it is desirable to connect a section of waveguide into the side of another waveguide. This type of connection forms a T-junction. It may be connected either in the narrow side as shown at A or in the wide side of the waveguide, as shown in diagram D. When the T-junction is in the plane of an h-field of a $TE_{0,1}$ mode, it is called an *H-type* junction, and when the junction is in the plane of the e-lines, it is called an *E-type* junction.

The H-type junction effectively is a parallel connection with the main line. For example, when the end of the T-joint shown at B to the left is short circuited at a distance of a half wave from the *center* of the waveguide, the result is the equivalent parallel circuit shown at C. Note that C shows a half-wave section which is connected to a wire line. This section will reflect a short circuit at the line and will not allow any energy to pass. Similarly, in the waveguide itself a short circuit is reflected to the center, where the e-lines are supposed to be. Since an e-line cannot exist at a short circuit, no energy will pass that point. If the shorted end of the T-section were only a quarter wave from the center of the waveguide, an open section would be reflected there and the passage of energy would be unaffected.

The E-type joint effectively is a series connection with one side of the main line, as you can see at E. Its two-wire counterpart is shown at F





T-Junction as Transmit-Receive Switch

In this case, if the section added to the waveguide is a half wave long, it will act as a short at the junction but will allow the energy to pass. The length of the added section will have to be a quarter wave in order to open the circuit at the junction and stop the ordinary flow of energy past it.

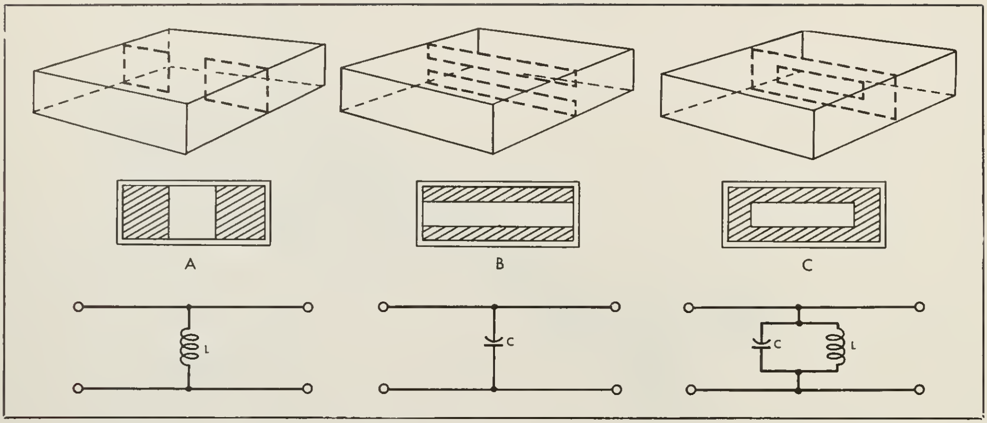
One important use of the T-joint is with the automatic Receive-Transmit switch (also called the R/T Box). In the partial Transmit-Receive system illustrated above an H-type T-junction is connected in the main waveguide. A spark gap is located one quarter wave from the center of the main guide. The junction is shorted at the distance of a half wave from the center of the main waveguide. When the powerful transmitter is turned on, a spark jumps the spark gap. This causes the waveguide branch to short at that point and is inverted to an open circuit at the center. Therefore, transmitted energy can pass unhindered.

Energy received from the radar echo enters in the waveguide from the other end after the transmitter is turned off. This energy is not great enough to cause a spark to jump across the spark gap. This time the shorted end of the branch reflects a short at the center of the waveguide and reflects the received energy back to another T-junction, which is located at the input to the receiver. Thus no energy is absorbed by the inoperative transmitter. There is a similar arrangement called a T/R Box placed in the waveguide leading to the receiver. It offers high impedance to the transmitted pulse but low impedance to the received echo.

In practice the actual length of the sections in T-junctions is not an exact quarter or a half wave length because the e-fields and h-fields are not perfect around the T-junction. A distortion around the fringe called *fringe effect* requires some variations from exact wavelengths.

Matching Devices

Some devices which are used in radar introduce inductance or capacitance. Sometimes these reactances are deliberately introduced. Other times when they are present and not desired, then can be tuned out with small fins or plates in a waveguide.



Reactive Plates in Waveguide

The illustration above shows a number of reactive plates which are deliberately used to introduce capacity or inductance in a waveguide. When these plates are employed as shown at A above they set up oscillations in the higher modes. Since a waveguide is too small for higher modes at the same frequency, these frequencies are not propagated, but remain in the vicinity of the plates. If the edges of the plates are vertical with respect to the plane of the h-field the modes produced are the TM type. The effect of this on power flow is that of inductance across the two-wire line. This causes reflections and a shift in the standing wave pattern. The wider the space between the plates, the greater the inductive reactance.

When the partitions are arranged perpendicularly to the e-field as in B above, a local e-field and the higher modes of oscillation are set up between the edges of the plates. These oscillations cannot be propagated but do change the dominant mode to a TE mode and introduce capacitive reactance. As with the TM mode, the wider the opening, the greater the reactance.

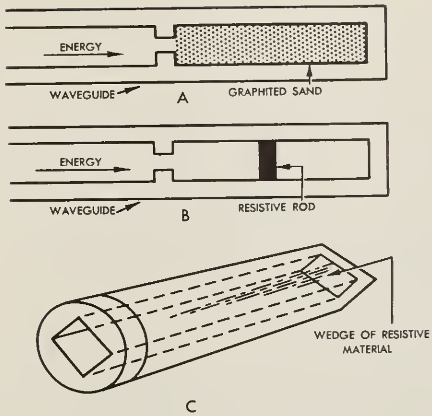
From these facts it would seem that by combining both types of plates and leaving a small opening in a large guide as in C would produce a resonant circuit. This is approximately true, providing the dimensions are correct. At resonance a resonant circuit acts like high resistance. In this condition, a small opening would introduce a high shunt resistance and the guide would in effect have connected across it a resonant circuit, since at resonance a resonant circuit acts like a high resistance.

Terminating a Waveguide

Since a waveguide is a single conductor, it is not as easy to define its characteristic impedance (Z_0) as it is for a coaxial line. Nevertheless, you can think of the characteristic impedance of a waveguide as being approximately equal to the ratio of the strength of the electric field to the strength of the magnetic field for energy traveling in one direction. This ratio is equivalent to the voltage to current ratio in coaxial lines on which there are no standing waves.

The lowest characteristic impedance of a circular waveguide is about 350 ohms. In a rectangular waveguide, it may be any value, depending on the dimensions of the waveguide and the frequency of the electrical energy. In this guide, it is directly proportional to the narrow dimension when the other dimension and the frequency are fixed and may vary from approximately zero to 465 ohms.

On a waveguide there is no place to connect a fixed resistor to terminate it in its characteristic impedance as there is on a coaxial cable. However, there are a number of special arrangements which accomplish the same thing. One consists of filling the end of the waveguide with graphited-sand as shown at A at top of page 10-23. As the fields enter the sand, currents flow in it. These currents create heat, which is instrumental in dissipating energy. None of the energy thus dissipated as heat is reflected back into the guide. Another arrangement B



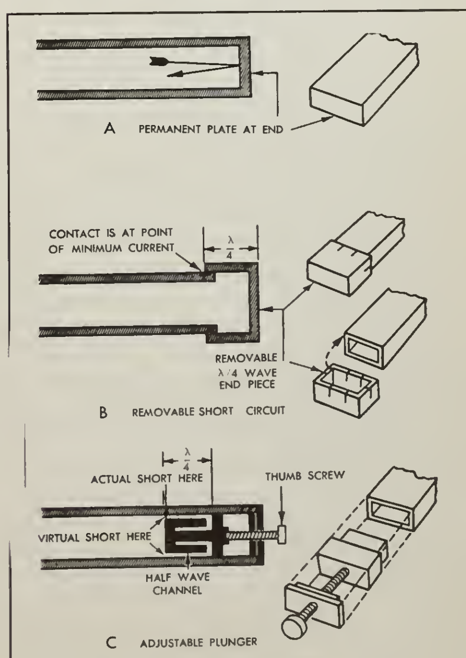
Termination for Minimum Reflections

uses a high resistance rod, which is placed at the center of the e-field. The e-field (voltage) causes current to flow through the rod. The high resistance of the rod dissipates the energy as an I^2R loss. Still another method for terminating a waveguide is to use a wedge of high resistance material C. The plane of the wedge is placed perpendicular to the magnetic lines of force. When the h-lines cut the wedge, a voltage is induced in it. The current produced by the induced voltage on flowing through the high resistance of the wedge produces an I_2R loss. This loss is dissipated in the form of heat. This permits very little energy to reach the closed end to be reflected.

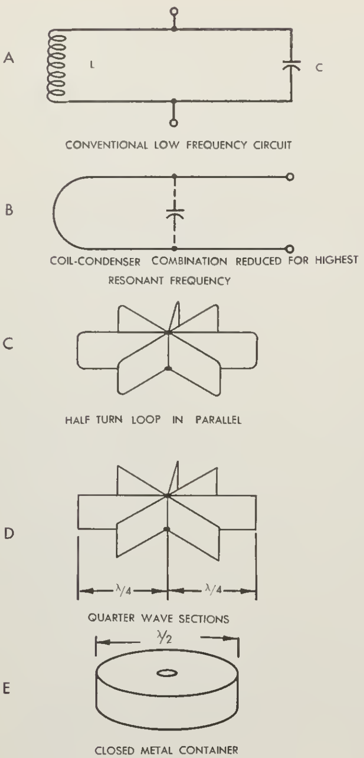
Each of the preceding terminations is designed to match the impedance of the guide in order to insure a minimum of reflection. On the other hand, there are many instances where it is desirable for all the energy to be reflected from the end of the waveguide. The best way to accomplish this is to permanently weld a metal plate at the end of the waveguide as shown at A to the right.

When it is necessary that the end be removable, a removable end plate is attached to the end of the guide. For this method to be satisfactory, the contact between the guide and the end plate must be exceptionally good in order that the h-field will not be attenuated when current flows. Perfect contact is not required when the connection is made at a point of minimum current. This point is located at a quarter wave from the end. When connection is made at this point, a cup is used instead of the end plate B. This cup is a quarter wave long and large enough to fit over the end of the guide. The voltage between opposite sides of the cup opening is high, but as the reflected h-field cancels the incident h-field, the resulting current is very small and reflection is at a minimum.

When the end must be adjustable, the contact must be nearly perfect. However, it is impossible to get a perfect contact. The best arrangement, one which is similar to the choke joint previously explained, consists principally of an adjustable plunger which fits into the guide as shown at C. The walls of the waveguide and the plunger form a half-wave channel. The half-wave channel is closed at the end and reflects a short circuit across the other end, where a perfect connection is supposed to exist between the wall and the plunger. The actual physical contact is made at a quarter wave distance from the short circuit, where the current is minimum due to the standing waves. This makes it possible for the plunger to slide loosely in the guide at a point where the contact resistance to current flow is very low.



Termination for Maximum Reflection



Development of Cavity from λ/4 Sections

CAVITY RESONATORS

In ordinary radio work the conventional low frequency resonant circuit consists of a coil and condenser which are connected either in series or in parallel as shown above at A. To increase the resonant frequency, it is necessary to decrease either the capacity or the inductance or both. However, a frequency is reached where the inductance is a half turn coil and where the capacity consists only of the stray capacity in the coil. At extremely high frequencies this resonant circuit would consist of a coil about an inch long and a quarter inch across.

In this circuit the current handling capacity and breakdown voltage for the spacing would be low.

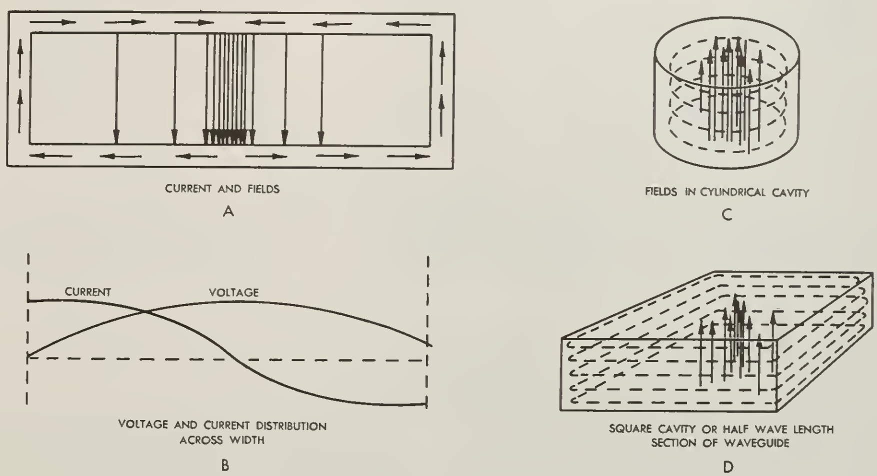
The current carrying ability of a resonant circuit may be increased by adding half turn loops in parallel. This does not change the resonant frequency appreciably because it adds capacity in parallel—which lowers the frequency, and inductance in parallel—which increases the frequency. As the effects of each cancel, the frequency remains about the same.

In the diagram at C several half turn loops are added in parallel. In D are shown in parallel several quarter wave Lecher Lines, which you recall are resonant when they are near a quarter wavelength. When more and more loops are added in parallel, the assembly eventually becomes a closed resonant box as shown at E which is a quarter wave in radius or, in other words, a half wave in diameter. This box is called a *resonant cavity*.

A resonant cavity displays the same resonant characteristics as a tuned circuit composed of a coil and condenser. In it there are a large number of current paths. This means that the resistance of the box to current flow is very low and that the Q of the resonant circuit is very high. While it is difficult to attain a Q of several hundred in a coil of wire, it is fairly easy to construct a resonant cavity with a Q of many thousand. Although a cavity is as efficient at low frequencies as at high frequencies, the large size required at low frequencies prohibits its use at those frequencies. For example, at one megacycle a resonant cavity would be a cylinder about 500 feet in diameter. When the frequency is in the vicinity of 10,000 megacycles, the diameter of the cavity is only 0.6 inch. This makes the cavity smaller than a conventional tuned circuit. Therefore, equipment which operates at a frequency of 3000 mc or above usually employs resonant cavities as resonant circuits.

The Fields in a Cavity

A resonant cavity may be compared to a waveguide, since its operation is best described in terms of the fields rather than in terms of the currents and voltages present. As in waveguides, the different field configurations in cavities are called *modes*. In the next illustration showing the dominant mode of the cylindrical cavity, note that in A the voltage is represented by e-lines between the top and the bottom of the cavity. The current, due to skin effect, flows in a thin layer on the surface of the cavity. The strength of the current is indicated by graduated arrows. The magnetic field is strong where the current is high. The strongest h-field is at the vertical walls of the cylinder and diminishes toward the center where the current is zero. This is due to the standing waves on the quarter



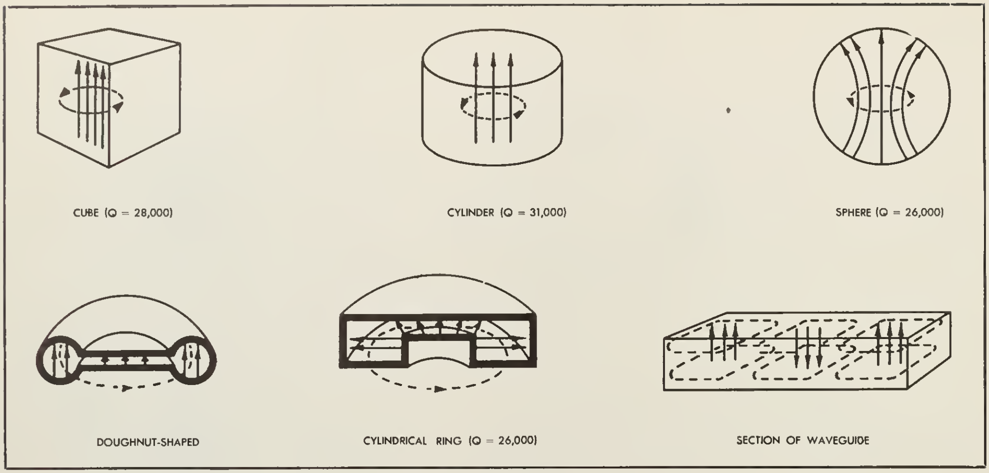
Voltages, Currents and Associated Fields for Simple Mode in a Cavity Resonator

wave section. The e-field is maximum at the center and decreases to zero at the edge, where the vertical wall is a short circuit to the voltage. The curve of e-field and h-field density are shown above at B, C and D show two types of cavities with their fields represented.

The modes in a cavity are identified by the same numbering system that is used with waveguides, except that a third subscript is used to indicate the number of patterns of the transverse field along the axis of the cavity (perpendicular to the transverse field). For example, the cylindrival cavity shown at C is a form of circular waveguide. The axis is the center of the circle. The transverse field is the magnetic field. Therefore, it is marked TM. Around the circumference there is a constant magnetic field. (The h-lines are parallel to the circumference.) Therefore, the first subscript is zero. The distance across the diameter is one half wave. Thus the second subscript is one. Through the center along the axis the h-field strength is a constant zero. This makes the third subscript zero. Therefore, the complete description of the mode is $TM_{0,1,0}$.

When a section of waveguide which is one-half wave long is closed on both ends in the form of a rectangular cavity, as shown at D, standing waves are set up and resonance occurs. The simple mode in this cavity is the same as the dominant mode of a rectangular waveguide, that is, it is $TE_{0,1}$. The third subscript of the mode, which is determined by the plane of the e-field, is 1. Thus the complete description of the simple mode in the rectangular cavity at D is $TE_{0,1,1}$.

Cavities may have various physical shapes, for any chamber enclosed in conducting walls resonates at several frequencies and produces a number of modes. Note the illustration on the next page showing examples of several types of cavities. The Q of each cavity is indicated. Of those shown the cylinder type cavity is useful in wavemeters or in frequency measuring devices. The cylindrical ring type is used in super high frequency oscillators as the frequency determining element. The section of waveguide which is shown diagrammatically is used in some radar systems as a mixing chamber for combining signals from two sources.

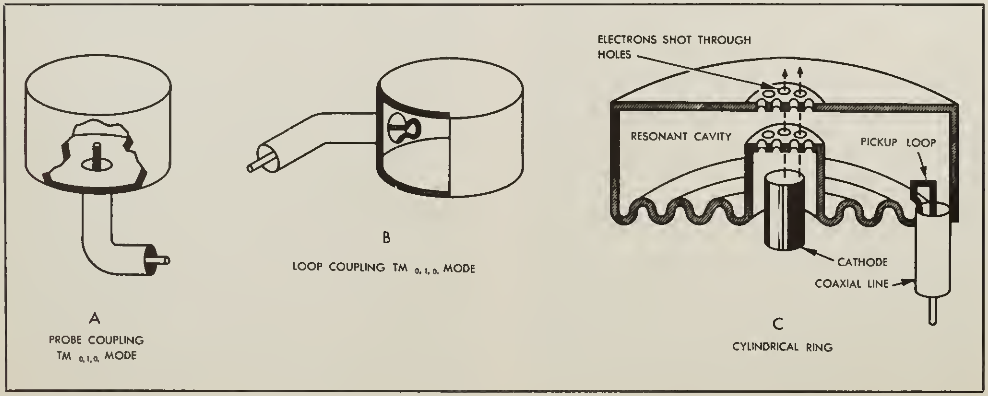


Several Types of Cavities

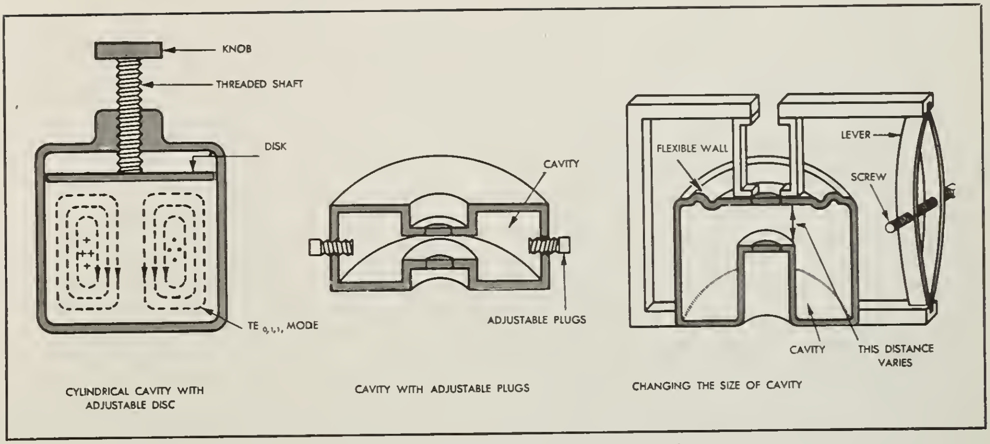
Exciting the Cavity

There are three principal ways in which energy can be inserted into and removed from a cavity resonator. The first is by inserting a probe into the cavity. The current which flows in the probe sets up e-lines parallel to it, and they in turn start oscillation. Another method uses a magnetic loop. The loop is placed in the region where the magnetic field will be located. The currents in the loop start an h-field in the cavity. Either of these methods can be used either to remove energy from or to put energy into the cavity. A third method uses a cylindrical ring type cavity. In this method the energy is placed into the cavity by clouds of electrons, which are virtually shot through the holes in the center of a perforated plate. As each cloud goes through, it creates a disturbance in the space inside the cavity until a field is set up. In terms of current, it may be said that the approaching cloud of electrons makes the perforated plate positive by repelling the electrons away from it. This current starts an h-field. Energy may be removed from the cavity by placing a loop at the outside edge.

Other methods of feeding the cavity are discussed later in the analysis of systems using cavities.



Methods of Exciting the Cavity



Method of Changing the Frequency of a Cavity

Varying the Resonant Frequency of the Cavity

Three methods for setting the resonant frequency in a cavity are shown above. One method uses a cylindrical cavity with an adjustable disc. When a $TE_{0,1,1}$ mode is used, the size of the cylinder may be changed along the axis to change the resonant frequency. The smaller the volume of the cavity, the higher the resonant frequency. The movement of the disc may be calibrated in terms of frequency. Usually in the case of high frequency equipment a micrometer scale is used to indicate the position of the disc, and a calibration chart is used to determine the frequency.

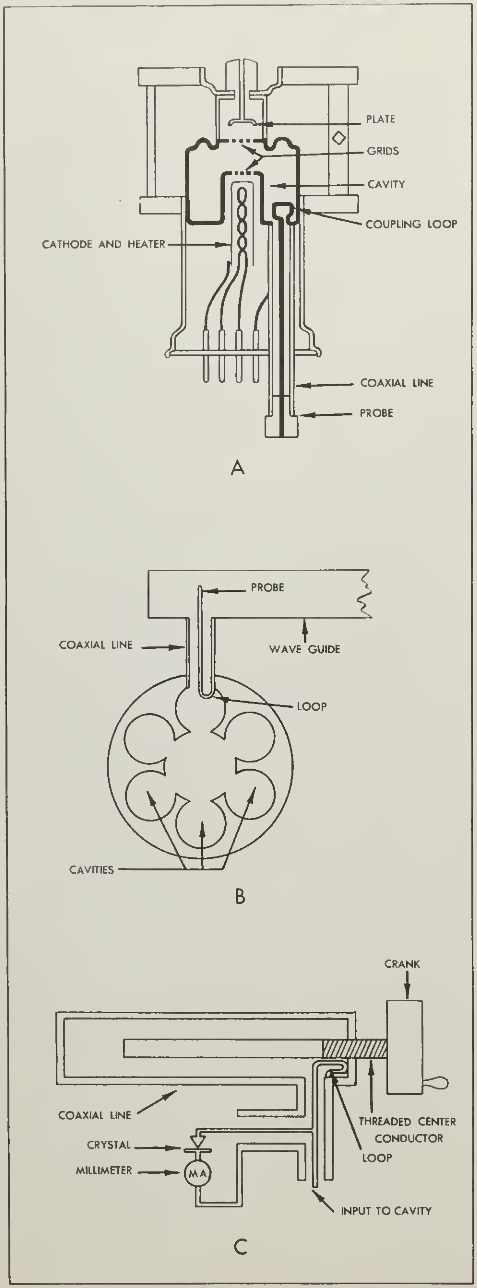
A second method for varying the frequency employs threaded plugs which are inserted in the cavity. The plug reduces the strength of the magnetic field in the cavity in a manner similar to reducing the inductance of the tuned circuit. The deeper the plug extends into the cavity, the higher the frequency.

In a third method, the interior of the cavity, which is part of the interior of a vacuum tube, is sealed and evacuated. In this method a frequency change occurs whenever the top and bottom of the cavity are moved away from (or toward) each other. This is accomplished by turning a screw at the end of a lever. As the screw is turned, the two pieces of metal to which the lever is attached are either pulled together or allowed to spring apart. This moves the top of the cavity up and down with respect to the bottom. As the distance varies, the volume and the capacity between the top and bottom of the cavity are changed. As the change in capacity is the chief result of the change in the distance between the plates, the resonant frequency is inversely proportional to the distance from the top to the bottom.

Another way of changing the frequency is by changing the method of exciting the circuit. This can be done by tuning the exciting loop either capacitively or inductively. This can occur either accidentially due to improperly tuned circuits or deliberately as a means of tuning the cavity.

Uses of Cavities

There are many ways in which resonant cavities are used in radar. One important use is as a tuned circuit in 10,000 mc oscillator circuits. For example, the klystron shown at A on page 10-28 which is usually the local oscillator in a radar receiver, is tuned by the resonant cavity at C. In the klystron, the cathode emits electrons which pass through the grids toward the plate. When they pass through this area, they disturb the field and excite the cavity. The resonant frequency is changed by the spacing of the grids. Another use of resonant cavities is the series of cavities arranged around a circle in the magnetron oscillator as shown at B. These cavities, which are coupled one to another through capacity at the openings, are excited by moving electrons. The energy from the electrons is passed around the ring from all cavities to the one with a loop. This one serves to transfer the energy to the waveguide. From the waveguide, the energy goes to the antenna where it is radiated into space.



Uses of Cavities

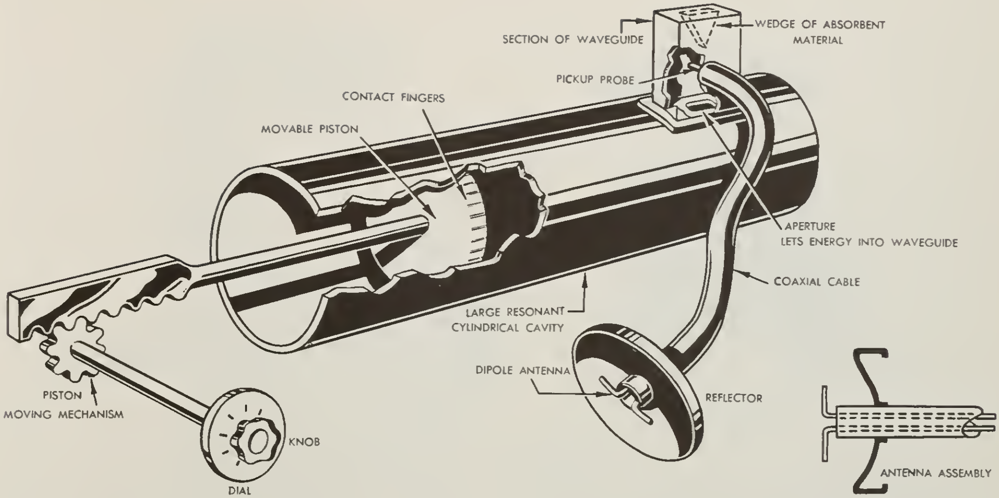
Since the frequency of the transmitter is fixed, there are no frequency adjusting devices on a waveguide. However, the different speeds at which the electrons travel and the variable impedances in the waveguide cause either a change in frequency, or oscillations in a mode which differs from the one intended.

Cavities also may be used as wavemeters, devices for measuring frequency. One type of wavemeter consists of a cylindrical cavity with an adjustable disc. Another type operates essentially like the first one but employs a coaxial cavity and operates in a different mode as shown at C to the left. In the second type, the signal is introduced into the cavity by a loop which is located in one end of the cavity. The signal is strongest when the loop is at a high current point in the standing wave. The standing wave of current is maximum at odd multiples of quarter wave distances from the open end of the guide. When the length of the threaded center conductor is varied by the crank, the distance from the open end to the loop can be made equal to an odd multiple of quarter wave. This causes the current maximum to occur at the loop and this in turn causes the input impedance at the loop to be zero. When the distance from the open end of the guide to the short (closed end) is a quarter wave or any odd multiple of a quarter wave, the current introduced into the guide will be at the location of parallel resonance.

The threaded shaft shown in the illustration at C can be calibrated either in wavelength or in frequency. The input impedance can be indicated by any suitable indicator, such as a crystal rectifier and a DC milliammeter. In operation, when the input impedance is zero, the signal is shorted out. In this case, the meter will read zero. To check the frequency for a zero impedance reading, turn the crank until a dip occurs on the meter and read the frequency on the threaded shaft.

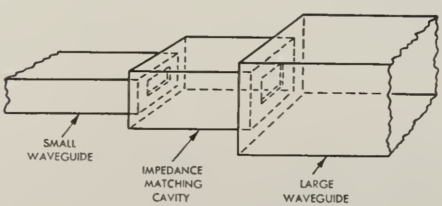
A cavity can be used as a mixing chamber. When two or more signals are put into a cavity, it is possible to remove the combination of their signal voltages. An example of this use is discussed later in this chapter.

Another use of a cavity is as an impedance matching device such as shown by the illustration of a small waveguide connected to a large waveguide on the next page. Normally, when you connect the two waveguides together, there will be reflections in the waveguides. However, a



3 CM Ringing Circuit

small section of waveguide or an intermediate size can be matched without reflections, provided that plates are used at the junction to cancel the reactive effects which result from the different sizes of the waveguides. Thus, although the cavity itself has standing waves in it, there are none on either waveguide.

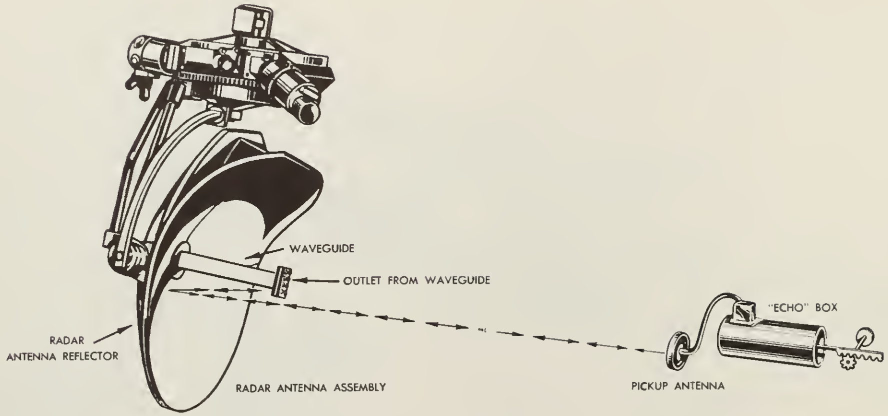


Cavity used for Matching an Impedance Device to One of Different Impedance

Another use of a cavity is as a *ringing circuit* in a resonant circuit. The ringing circuit gets its name from the fact that it oscillates for many cycles after being started by an external circuit. In this use the cavity is resonant and after started oscillating will continue for a few microseconds after the source of voltage has been removed. This action is analogous to a bell which will ring for many seconds after being struck. In the 3cm (10,000 mc) ringing circuit shown above, the cavity itself is a cylinder which is three inches in diameter and 10 inches long. A movable piston in the cylinder determines the volume of the cavity. As the position of the piston is calibrated in megacycles, it serves as a frequency meter. The cavity is energized by a signal which enters through a small aperture from the small waveguide. The energy is put into the short waveguide by a probe, which in turn is fed by a coaxial cable. A good method to put the signal into the coaxial cable is to connect a small antenna to the end of the cable.

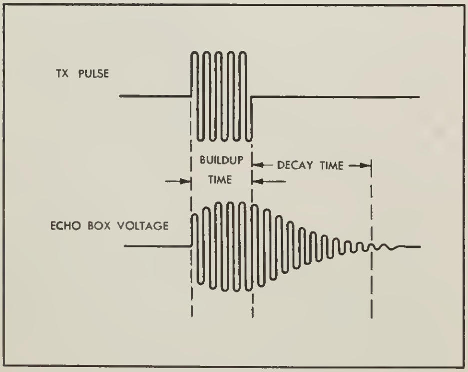
The resonant cavity shown above is used as an echo box as shown on page 10-30. Echo boxes are resonant cavities which have a very high Q. They are used with microwave radar sets to provide an artificial or "phantom" target, which may be used to tune the receiver to the transmitter when no real targets are available.

This is how the echo box works. When the radar transmitter is turned on, energy is fed into the echo box cavity. There it sets up violent oscillations, and the cavity itself acts like a transmitter for a period of 20 microseconds after the pulse generated by the radar transmitter ends. (When you view the oscillations on an oscilloscope, their waveshapes appear as shown in



Cavity as an Echo Box

the illustration below.) The energy which the echo box radiates is picked up by the small pickup antenna located in front of the radar antenna assembly and fed into the radar receiver. The receiver displays it on the indicator screen. Since this signal simulates a radar echo, the cavity which transmitted it is appropriately called an echo box. It is thus a means of supplying a signal for tuning the receiver when there is no real target available.



Oscillations in Echo Box

A COMPLETE WAVEGUIDE RF SYSTEM

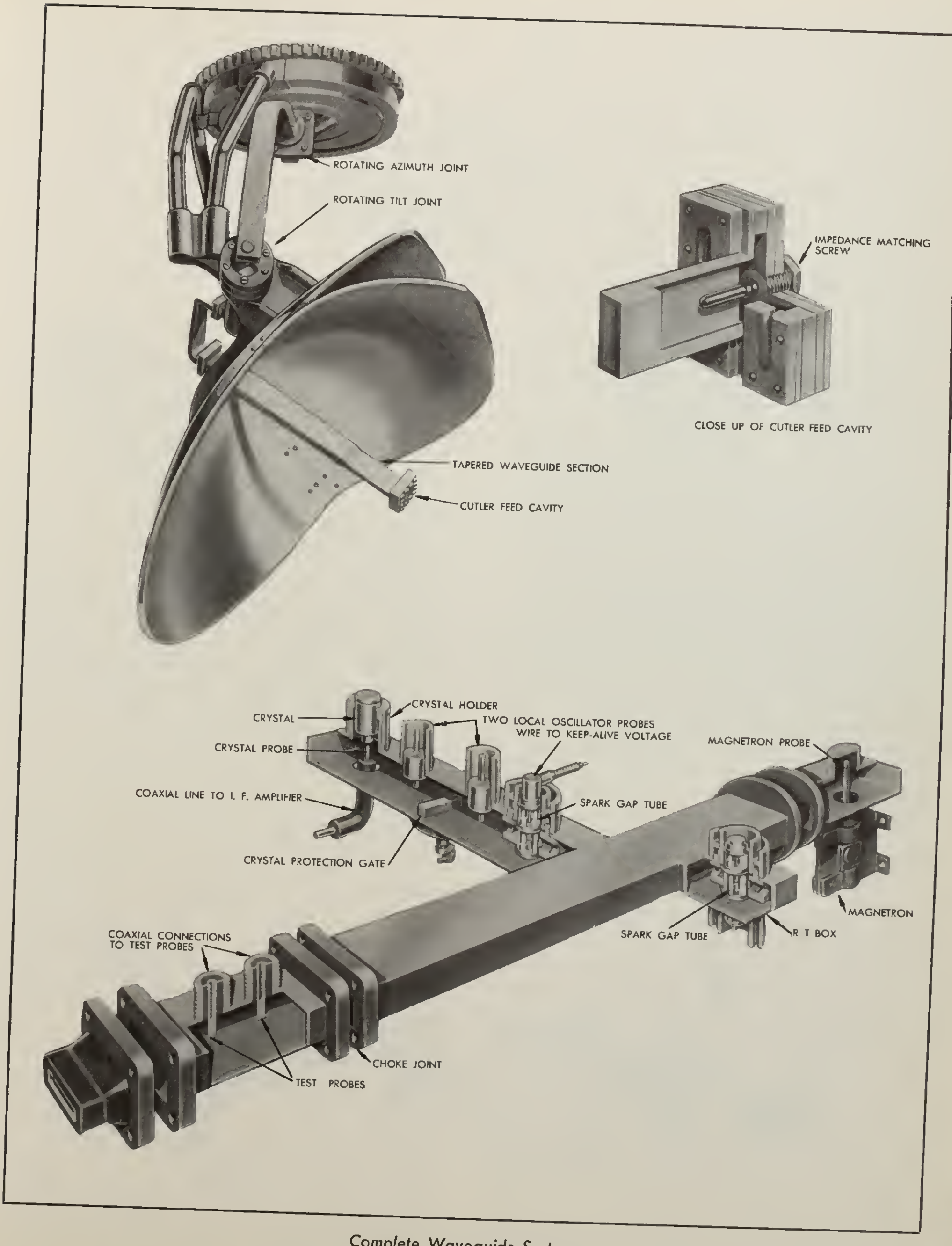
In a complete waveguide RF system the signal source is the magnetron oscillator. The antenna which transmits the output of the magnetron is located as closely as possible to the radar transmitter. A principal requirement in an efficient radar system is that energy must be transferred from the transmitter to the antenna with minimum losses, for any losses that occur will shorten the maximum range and thereby decrease the efficiency of the entire radar set. Since the transmitting antenna is pointed at the *target* at the time the transmitter pulse leaves, and since it is conveniently located, it is preferable to use it as the receiving antenna, too. Similarly, it is possible to use the same waveguide system for transferring the received energy to the radar receiver.

When the same elements are used for both transmitting and receiving, it is necessary to use a high speed automatic switch to prevent the powerful transmitter pulse from going into the receiver circuit and causing serious damage. The T/R and R/T Boxes discussed earlier perform this function in many radar sets.

The highly directional antenna is made rotatable in order to cover all areas. To make rotation possible, signals are coupled to the antenna through vertical and horizontal rotating joints.

Since airborne equipment is flown at high altitudes, temperature changes will cause moisture condensation inside the waveguide. As water causes extremely high losses, the interior of the waveguide must be maintained at a higher pressure than the outside to drive the moisture out and keep it out. Furthermore, all joints have to be airtight to maintain this pressure.

The illustration shows on the next page a waveguide system which uses the components just discussed. The source of the RF energy is the

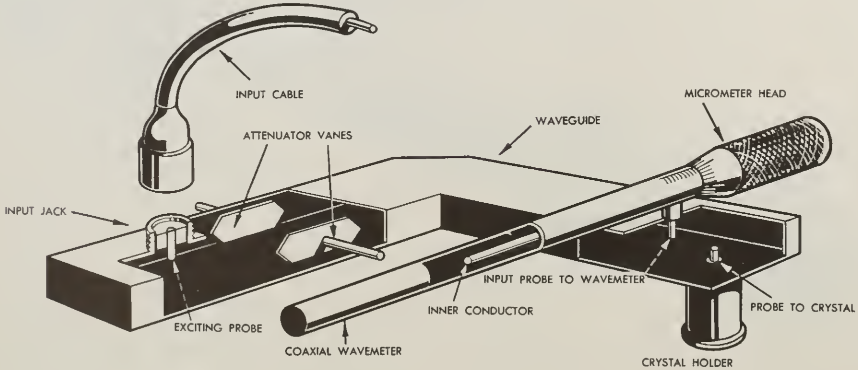


Complete Waveguide System

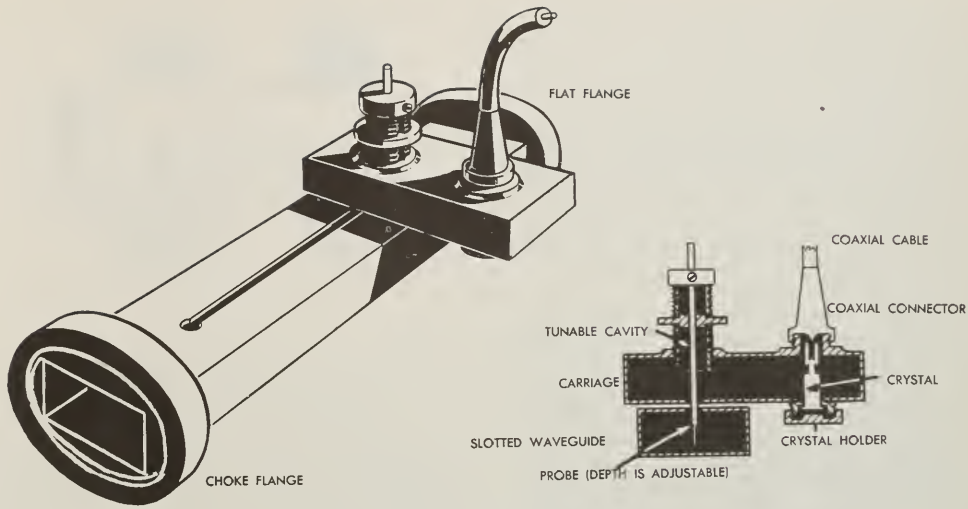
magnetron tube which is shown at the right in the lower part of the illustration. A probe couples it to the waveguide, which operates in the $TE_{0,2}$ mode. A very short section of waveguide is fixed permanently to the magnetron output. When the magnetron is installed, its waveguide section is connected to the end of the main waveguide. In operating conditions the transmitter energy travels down the main waveguide where it enters a section which has two T-junctions—one called the R T box, and the other the T R box and mixing chamber. In each T-junction, there is a spark gap at a distance of a quarter wave from the center of the main guide. Both these gaps arc-over when the transmitter is on. Thus, no energy goes by either spark gap tube into the T-junction waveguides. The energy flow is continuous down the main guide to the antenna. Each section is connected to the next with a choke joint. In going to the antenna the signal passes through a special section which contains a pair of probes. Each probe is a quarter wave from the other. Test equipment can be inserted at each probe to measure the signal strength at the probe. The ratio of the two signal strengths is the standing wave ratio. In going through the stationary transmitter unit to the moving antenna, the signal goes through a flexible section of waveguide which insulates the transmitter unit from the mechanical vibration of the antenna.

The antenna is designed to rotate horizontally and to tilt vertically. Thus, the signal is transferred to a rotating coaxial joint twice before being radiated and each time it is immediately transferred back to a waveguide again. The antenna itself is a resonant cavity on which there are a pair of slots one half-wave apart. As you can see in the enlargement, the waveguide is tapered in the narrow dimension to fit between the slots. The effect of the cavity is as though the waveguide were split in half and each side bent through 180°. Tapering the waveguide prevents the reflections that would occur if the size were changed abruptly. Tapering in the narrow dimension does not affect the operation of the waveguide, since this dimension is not critical. The actual impedance match is accomplished by adjusting the matching screw. This adjustment is made at the factory and soldered in place.

The energy comes out of the back of the cavity and goes to the large reflector. The reflector sends the energy out forward in a narrow beam. The energy which returns from the radar target is reflected by the reflector into the slots. It passes through the slots into the waveguide and travels toward the magnetron, where it stops at the R/T box. The length of the T-junction is a half-wave from the center of the waveguide. So the closed end of the section reflects a short circuit at the outer center of the



Accurate 3CM Frequency Meter using a Waveguide



Slotted Waveguide Section Permits Access to Fields in Waveguide System

main waveguide. This reflects the signal back along the guide to the mixing chamber. Of course, neither spark gap is fired because the signal is too weak to strike an arc. The signal enters the mixing chamber, where it strikes the end wall and reflections are again set up. The chamber becomes a cavity resonator. An additional signal is introduced from another oscillator—the heterodyne oscillator for the receiver. The signal is injected with a probe at the center of the E-field in the cavity. Both signals cause current flow through a pick-up probe at a quarter wave from the end of the cavity. This probe is connected to a crystal mixer. The second oscillator is set at a different frequency and produces the correct IF frequency when the returning signal is from a radar beacon transmitter.

As the crystal is easily damaged by high powered signals, the T/R box spark gap protection must be certain. To insure that it will arc-over at once when the transmitter signal appears in the mixing chamber or when other nearby radar transmitters accidentally send in a weaker—but still damaging—signal, a *keep alive* voltage is provided. This is a high DC voltage which causes the spark gap to be partially ionized. Then the additional voltage of any signal that is stronger than a radar echo will cause complete ionization or an arc between the electrodes. When the set is turned off, no keep alive voltage is present, and the crystal is vulnerable to signals from other nearby radar transmitters. So a spring operated gate closes the waveguide when the set is turned off. No signal can get to the crystal from the antenna. When the set is turned on, an electrically operated relay opens the gate.

The entire system is assembled with rubber gaskets at the joints, or otherwise sealed to permit pressurizing the interior of the waveguide. The small windows or slots in the antenna cavity are covered with mica. It lets RF energy through, but keeps moisture out.

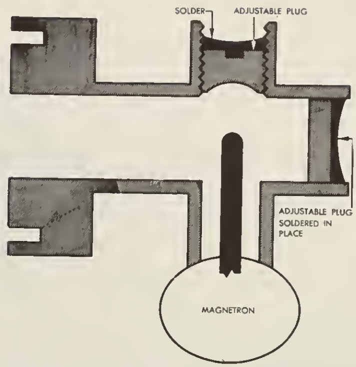
TEST EQUIPMENT

A resonant cavity can be used to measure the frequency of the signal in a waveguide. When approximate measurements are to be made, either the ring box or echo box can be used. When more accurate measurement is desired, a frequency meter of the type shown on the preceding page is used. In this meter a coaxial cable brings the signal from a test jack in the waveguide of the radar set to the input jack of the wavemeter waveguide. A probe sends the signal down the guide through attenuator vanes. When these vanes are at the walls of guide, they have no effect, but when they are moved together, they reduce the wide dimension until it is less than cutoff wave length for the frequency involved. The signal bounces back and forth between the vanes, with very little of it getting out of the waveguide proper. It then goes to a coaxial cavity. This cavity is similar to that previously illustrated on page 10-28, except that it has a micrometer head to insure maximum accuracy of reading. The coaxial line is resonant when it is 15 quarter wavelengths long in this test set. Resonance is indicated by the rectified output from the crystal, which is used to swing a meter, or if pulsed signal is used, the amplitude of the pulse can be observed on a test oscilloscope.

You can measure the standing waves on any waveguide system by inserting a slotted section into the waveguide as shown on page 10-33 and then using a travel-probe to measure the electric field strength. A special assembly fits on the slotted section onto the waveguide. This assembly picks up energy by a probe and transmits it through a short waveguide to a crystal, where it is rectified and then carried by a coaxial cable to an indicator, which is similar to that described earlier in the slotted coaxial line.

ADJUSTMENTS

In a waveguide system in an airborne radar set, although there are many adjustments, most of them are made at the factory. For example, almost all the probes which a waveguide system uses cause an impedance mismatch. This mismatch must be tuned out. The plugs of the type shown above are provided to get rid of the mismatch, but adjustment is made at the factory wherever possible and the adjustment is fixed by soldering.



Mismatch and Reflections Introduced by Magnetron Probe are tuned out by plugs

On the other hand, an adjustment which the radar mechanic does adjust deals with the amount of signal introduced by the local oscillator. This can be varied by making the probe in the mixing chamber longer or shorter. This probe, which is permanently fixed to the oscillator, moves the entire oscillator up and down with respect to the waveguide.

CHAPTER 11

Oscillators at Radar Frequencies

A radar transmitter consists of an oscillator and its associated pulsing circuits. The theory of pulsing circuits has been discussed to some extent in previous chapters. One purpose of this chapter is to acquaint you with oscillators and oscillatory circuits employed in radar transmitters; another is to give you additional information about pulsing circuits in transmitters, and to acquaint you with local oscillators used in radar receivers. Specifically, it discusses problems involved in generating high frequency signals, explains the theory and operation of a number of oscillators employed in radar equipment, and describes various methods for pulsing oscillators.

ADVANTAGES OF HIGH FREQUENCIES

Radar oscillators are operated at high frequencies for three reasons—high frequencies produce the best echoes, they make it possible to detect smaller targets, and permit the use of small antennas.

As mentioned previously radar transmitters send out short pulses at regular intervals. Usually the duration of these pulses is a microsecond or so to prevent the echo from any particular pulse from being returned by nearby targets before that pulse ends at the transmitter. Furthermore it is proved by experiments that satisfactory echoes require that the transmitted pulse consists of 200 or more oscillations per second. Obviously then the only way for the pulse to contain the necessary number of oscillations within the specified time limit is to operate the transmitter oscillator at ultra-high or microwave frequencies.

The size of the target that radar can detect depends mostly on the frequency. In general the higher the frequency, the smaller the target that the radar set can detect. Furthermore small targets produce better echoes than large targets. A target gives best echoes when it is about a half wave or some multiple of the radar frequency being transmitted. If the target is much smaller than a half wave, the intensity of pulse that it reflects is low. However as the operating frequency is increased the half wave becomes smaller and makes it possible to detect smaller targets.

Perhaps the main reason for using higher frequencies is that it is possible to construct directive antennas in smaller space than when lower frequencies are used. Small directive antennas are especially an important consideration in airborne equipment where space and weight are major problems.

FREQUENCY LIMITATIONS OF USUAL OSCILLATORS

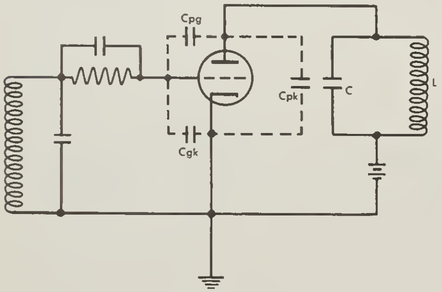
The principal factors which limit the frequency of the ordinary type oscillator are the construction of the tube and the external circuits which connect to the tube.

Tube Construction

The tube construction factors which limit the operating frequency of the usual or conventional oscillator are the interelectrode capacitances in the tube, the inductances of the leads, and the transit time.

INTERELECTRODE CAPACITANCES. At ordinary radio frequencies the interelectrode capacitances in a vacuum tube form reactances which are large enough not to cause any serious trouble. However, as frequencies increase, the reactance of these capacitances become small enough to materially affect the performance of a circuit. A 1 mmf condenser, for example, has a reactance of 1590 ohms at 100 mc. If this condenser is the plate to grid capacitance and the RF voltage between these electrodes is 500 volts, there will be an interelectrode capacitance current flow of $\frac{500}{1590}$ or .315 amperes, an amount of current which will disturb circuit operation. On the other hand, at 1mc the reactance of this condenser becomes approximately 159,000 ohms and the current flow is only $\frac{500}{159000}$ or 3.15 milliamperes, an amount which will not seriously affect circuit performance.

A good point to remember is that the higher the frequency or the larger the interelectrode capacitance, the higher the current flow through this capacitance. In most UHF oscillators interelectrode capacitance currents are much greater than the power currents supplied by the tube. These higher currents cause power losses in the resistance in the oscillatory circuit.



Effect of Interelectrode Capacitance on Frequency of TPTG Oscillator

Since interelectrode capacitances are effectively in parallel with the tuned circuit, they affect the frequency at which the tuned circuit resonates. As you can see in the circuit above, the plate-to-cathode capacitance is in parallel with the series combination of the plate-to-grid capacitance and the grid-to-cathode capacitance. All these capacitances together form a part of the total capacity of the tuned circuit. You can find their total capacitance by the formula,

$$C_t = C + C_{pk} + \frac{C_{gk}\ C_{gp}}{C_{gk} + C_{gp}}$$

Not only does interelectrode capacitance limit the frequency by establishing a minimum capacity below which it is impossible to go, but it also varies with the applied voltages and with the loading of the oscillator. This causes frequency instability, particularly when the interelectrode capacitance forms a large part of the tuning capacitance.

Inductance of leads. Another frequency limiting factor within a tube is the inductance of the leads to each tube element. While these inductances do not necessarily impair the efficiency of the oscillator, they may represent a major portion of the inductance of the tuned circuit and limit the frequency by setting a minimum limit on the inductance. Furthermore, since the cathode lead is common to both plate and grid circuits, feedback takes place through it and produces an additional loss of efficiency.

Transit time. A third limitation imposed by tube construction is *transit time.* Transit time is the time required for electrons to travel from cathode to plate. At radio frequencies transit time is negligible, since it occupies only a comparatively small portion of an oscillatory period. But as the frequency becomes higher, transit time occupies an appreciable portion of this period and produces undesirable effects in tube operation. The effect of transit time is of especial concern in connection with the input impedance of the tube. Part of the current that flows in the grid circuit is the current which charges the grid to plate capacitance, C_{gp}. The voltage that produces this current is the vector sum of the input voltage (grid to cathode) and the output voltage across the plate load. At lower frequencies with a resistive load, these two voltages are 180° out of phase and add algebraically to determine the charging current. This current is 90° out of phase with the input voltage. However, at higher frequencies, where transit time is an important factor, the plate current begins to lag the input voltage. This causes the plate voltage to be less than 180° out of phase with the input voltage, and the voltage across the condenser to lag the input voltage slightly As a result, the charging current will no longer be 90° out of phase but will have an in-phase component. This means that power is consumed in the grid circuit. This consumption of power is effectively the same as adding a high resistance in the input impedance. This resistance decreases as the frequency increases.

Overcoming tube limitations. There are several ways to reduce the effect of interelectrode capacitances in vacuum tubes. None are completely satisfactory. One is to move the electrodes farther apart. However, this is hardly desirable for it increases the transit time. Another method is to reduce the size of the tube and electrodes. This is satisfactory except for one ill effect. It decreases the power handling ability of the tube directly with the square of the factor by which the electrodes are reduced. Another method is to separate the leads and to bring them out of the envelope at the nearest point. This results in a slight decrease in the capacitances.

Similarly there are several ways to reduce the inductances of the leads. As just mentioned, bringing out electrode leads through the envelope at the nearest point produces a slight decrease in the electrode capacitance in a tube. This also decreases the inductances of the leads. Another method is to make double connections to the electrode. This makes two parallel inductances which cut the lead inductance in half. Another method is to arrange the leads as extensions of external transmission lines.

For reducing transit time, closer spacing of electrodes and higher plate voltage are employed. Closer spacing of electrodes, however, causes higher interelectrode capacitances. Therefore, it can be used only when the electrodes are made smaller. For this reason radar oscillators operate at very high plate voltage. The fact that a radar transmitter does not operate continuously is an advantage in that high enough voltages may be used to reduce transit time without exceeding the maximum power rating of the tube. Operating with high voltage does, however, necessitate some precautionary measures, such as separating the seals of the leads to avoid the presence of excessive voltage gradients in or on the surface of the glass envelope and avoiding sharp projections on the high voltage leads or electrodes to prevent arcing.

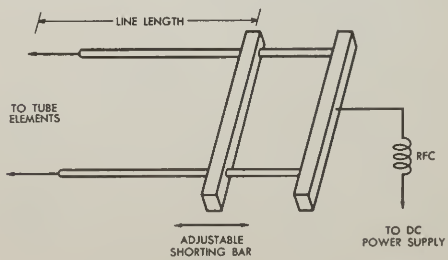
Limitations Due to External Circuits

Increases in frequency cause a quite rapid increase in the power losses in the external oscillatory circuit. These power losses are due principally to skin effect and radiation.

SKIN EFFECT. Skin effect causes a considerable increase in the resistance in a vacuum tube circuit. This results in a lower Q and increased I^2R losses. To prevent skin effect losses conductors are made large in size and tubular in shape since current flows only in the surface. In addition they are plated with silver, since it has a higher conductivity than copper, the material out of which most conductors are made.

RADIATION. Radiation, the other cause of power loss, is due to incomplete cancellation of electromagnetic fields in the region surrounding the circuit. When the frequency is low enough so that the spacing between two parallel conductors equals only a very small fraction of a half wavelength, there is almost complete cancellation of fields in all directions. At higher frequencies, however, such an identical spacing would represent a larger fraction of a half wavelength. This means less cancellation. In extreme cases, such as where the spacing is a half wavelength, the fields add in the direction of the plane containing the two conductors. This causes the tuned circuit to radiate energy like an antenna. Therefore, as the frequency increases it is necessary to reduce the spacing between the parallel elements. However, there is a limit on how far you can go in reducing the spacing. Too close spacing, for example, causes arcing which materially increases the RF resistance of the tank circuit. Another means of eliminating radiation is using concentric lines instead of open wire lines. This eliminates radiation entirely, since the outer conductor acts like a shield which prevents the electromagnetic field from expanding into space.

LIMIT OF INDUCTANCE AND CAPACITANCE. At UHF frequencies it is necessary that the inductances and the capacitances in the oscillatory circuit be very small. The limit of capacitance is the sum of the interelectrode capacitances and the distributed capacitance of the leads. The limit of inductance is the lead inductance plus the inductance necessary to connect the tube electrodes externally. UHF oscillators approach both these limits, the only capacitance in the tank circuit being a small trimmer condenser for fine tuning adjustments, and the inductance being a short circuited transmission line less than one quarter wavelength in length.



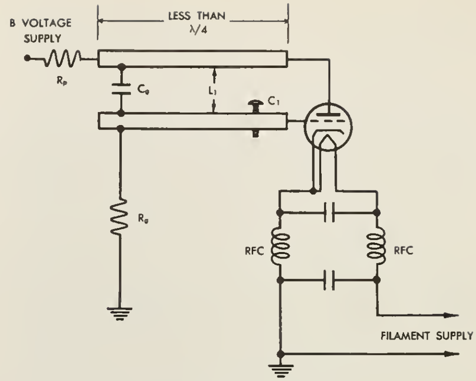
Typical Parallel Wire Resonant Circuit

In the preceding typical parallel line tank circuit, the expression for the input reactance (looking at the circuit from the tube) of the tank circuit is

$$Z = {}_jR_c \tan B_s$$

where R_c is the characteristic resistance of the line and B_s the electrical length.

As you can see in the expression, if B_s is less than 90° ($\lambda/4$), then tan B_s is positive and the reactance is inductive. For smaller values of inductance the length of the line is decreased. Further, this expression is true only when the resistance losses are low enough to be neglected. Frequently the tank circuit is located near a flat surface and parallel to it. The flat surface acts as a shield and, while it does not affect the electrical length of the line, it does lower the characteristic resistance of the line.



Ultra Audion Oscillator

TRIODE OSCILLATORS AS TRANSMITTERS

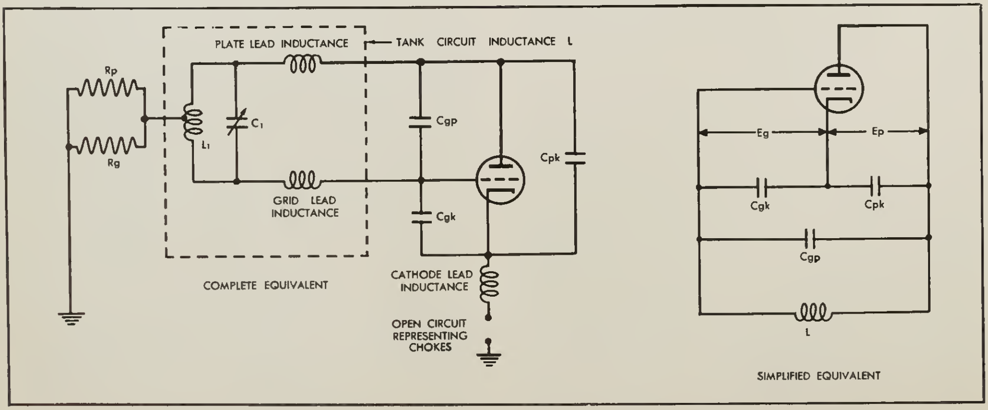
A variety of oscillators are used in the so-called long-wave types of radar equipment. While the discussion deals primarily with their use in the transmitting circuits of this equipment, it is also applicable to local oscillators in receivers. The main difference between the two kinds of oscillators is that the transmitter oscillator is pulsed and produces considerable power during each pulse, while the local oscillator operates on CW at a lower power output than the transmitter oscillator.

Ultra Audion Oscillator

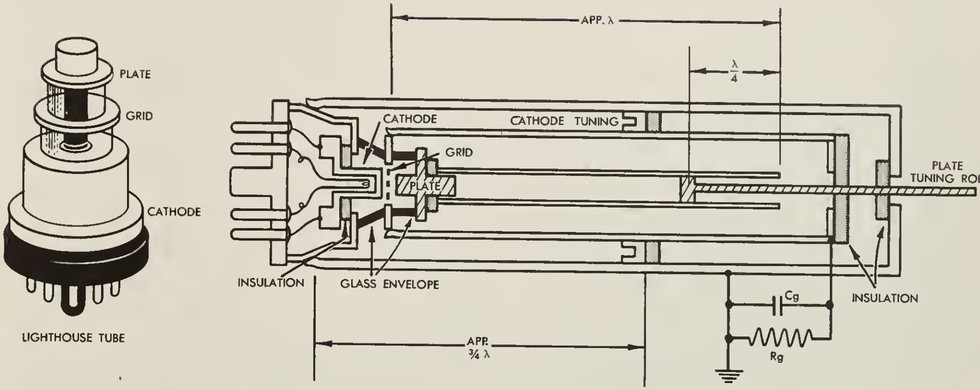
The typical ultra audion oscillator above is used as a local oscillator in some radar receivers and as a transmitter in certain IFF (Identification, Friend or Foe) sets. The circuit shown is incapable of generating enough power for it to be useful for high-power long-range radar sets.

The DC circuits consist of the components R_p, C_g, R_g and the small DC resistances in the tank circuit and leads. The components, C_g and R_g form the grid leak bias circuit. The charge built up on C_g biases the tube to class C operation.

To understand the operation of the ultra audion oscillator, consider its equivalent circuits shown below. So far as RF is concerned, as you can see in the equivalent circuit, the chokes in the cathode are open circuits. The tank circuit inductance L consists of the plate and grid lead inductances connected in series with the parallel combination of L_1 and C_1.



Equivalent Circuits of Ultra Audion Oscillator for RF Currents



Lighthouse Tube Oscillator

The simplified equivalent shows that this tank circuit inductance forms a resonant circuit with the combined interelectrode capacitance, C, which is given by the equation,

$$C = C_{gp} + \frac{C_{gk} \times C_{pk}}{C_{gk} + C_{pk}}$$

The expression for the frequency of this resonant circuit is,

$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

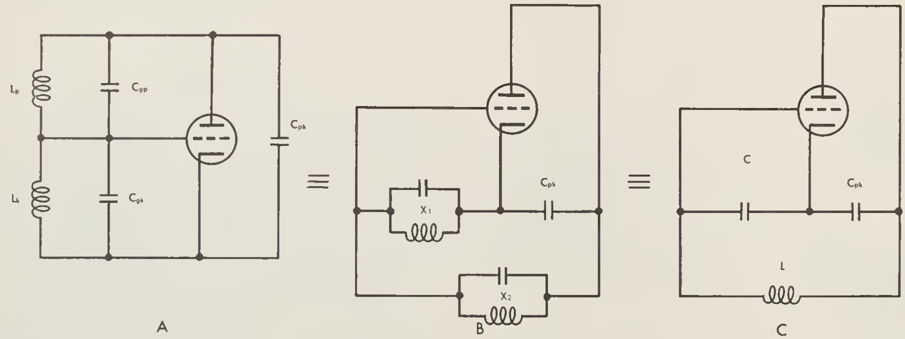
Basically the circuit of the ultra audion oscillator, as you can see in the simplified equivalent is that of a Colpitts oscillator. The capacitances C_{gk} and C_{pk} form a voltage divider circuit. The voltage E_g developed across C_{gk} is in the right phase and high enough to sustain oscillations. The minimum amplification required to sustain oscillations is given by the formula $A = C_{gk}/C_{pk}$. If the initial gain is greater than this value, oscillations build up to such an amplitude that the gain is reduced to the value given by this formula.

Lighthouse Tube Oscillator

The lighthouse tube oscillator like the ultra-audio oscillator is a single tube oscillator. It is used in a few low-powered radar sets such as short-range gun-laying equipment and in some types of transportable beacon equipment. In a typical application the lighthouse tube, which gets its name from its peculiar construction; operates at a frequency of 2500 mc. Its plate, grid and cathode connections are arranged so that it may be mounted directly in the tuning assembly, as you can see in the above illustration. The component called the tuner consists of three coaxial cylindrical conductors, of which the inner cylinder makes contact with the plate, the next with the grid, and the outer conductor with the shell of the tube. The outer conductor provides an RF connection to the cathode through capacitance. The space between the cathode and grid conductors forms a coaxial cathode line which in turn is shorted by means of a plunger. This plunger does not touch the grid conductor since a connection would form a DC connection from grid to cathode. The capacitance between the plunger and the grid conductor is great enough to form a virtual short circuit for RF. The DC grid-to-cathode path consisting of R_g and C_g, form the grid leak bias circuit.

In the plate circuit, the plate line consists of the grid and plate conductors. It is open circuited at the end away from the tube. The plate voltage is applied to the plate lead through the plate tuning rod at the point where the plate line is shorted. This point is a ¼ wavelength from the end. This quarter-wave section which is shorted at the point where B+ is applied presents a high impedance to RF from the open-ended plate line and therefore prevents RF from flowing into the power supply.

Equivalent circuits. The following equivalent circuits show the evolution of the lighthouse tube oscillator. In the first equivalent, the components, L_k and L_p, represent the inductances of the shorted cathode line and the open circuited plate line respectively. This is a true representation of the actual circuit provided



Evolution of Equivalent Circuit of Lighthouse Tube Oscillator

that the cathode line is less than ¾ wavelength long and provided that the plate line is less than a full wavelength long. The ¾ wavelength and full wavelength were chosen for this circuit rather than the ¼ and ½ wavelength due to the fact that at the frequency of operation the latter fractions of a wavelength are inconveniently short. However, keep in mind that if the cathode line is slightly less than ¼ wavelength, it acts like an inductance and that if the plate line is slightly less than ½ wavelength, it will act like an inductance. There is no lead inductance shown in the circuit since the lines connect directly to the electrodes. The interelectrode capacitances complete the equivalent circuit.

In the equivalent circuit B, the parallel combination of L_p and C_{gp} are represented by X_2 and the parallel combination of L_k and C_{gk} by X_1. Obviously the reactances X_2 and X_1 along with C_{pk}, which you can consider unknown reactances, must form a resonant circuit. The reactance X_1 forms a voltage divider circuit along with C_{pk}. Since the voltage across X_1 is the voltage fed back to the grid, it must be 180° out of phase with the plate voltage which can happen only if X_1 is a capacitive reactance. This condition leads to the third equivalent circuit C in which the reactance X_1 is represented by a condenser.

Now an oscillatory circuit requires at least one inductance; therefore X_2 must be an inductive reactance. Hence in the final or the third equivalent circuit C the reactance X_2 is shown as the inductance L. This circuit is representative of the oscillator circuit only at the resonant frequency. On examining this circuit, notice that basically it is that of a Colpitts oscillator. Now, if X_1 is capacitive, C_{gk} must conduct more heavily than L_k; therefore the oscillator frequency must be above the resonant frequency of L_k and C_{gk}. Likewise for X_2 to be inductive the oscillator frequency must be below the resonant frequency of L_p and C_{gp}. Thus you see that the oscillator frequency is between the resonant frequencies of the plate circuit and the cathode circuit.

FEEDBACK. In this oscillator, the amount of feedback depends upon the size of C in relation to C_{pk}. This is, in turn, a function of the tuning of the cathode line; hence, the cathode tuning serves principally for controlling the amount of feedback. If feedback is too small, oscillations will be weak and may cease completely. If feedback is too great, power consumption in the grid circuit will become too large and will result in a decrease in output of the oscillator. The value of C, hence of the cathode tuning, has some effect on the frequency of oscillations but, since C and C_{pk} are in series and C is usually much larger than C_{pk}, C must be changed considerably to effect the frequency appreciably.

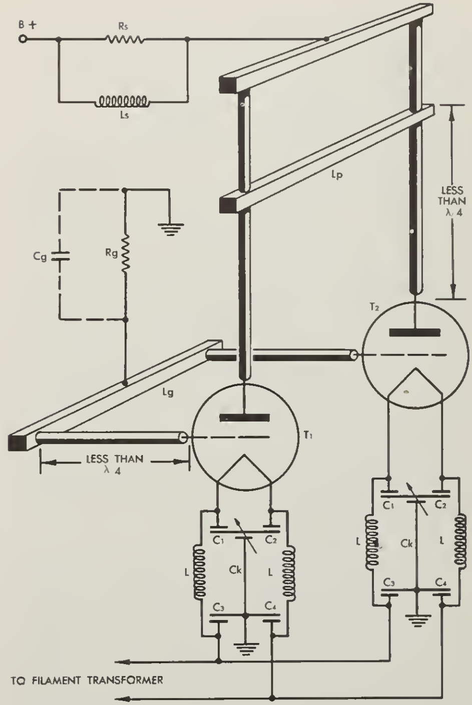
PLATE TUNING. Since the tuning of the plate line determines the resonant frequency of the plate circuit, which in turn determines the amount of inductance represented by the plate circuit at the oscillator frequency, the plate tuning is the chief control of the oscillator frequency. It has a secondary effect of feedback since C is determined by the relative frequencies of the cathode circuit and of the oscillator. You can see from the foregoing discussion that the plate line is tuned to give the correct frequency and the cathode line to adjust feedback for maximum output from the oscillator but, that since there is some interaction, it will be necessary to readjust each line for maximum accuracy.

Push Pull Oscillators

To obtain higher power outputs, two tubes in push pull circuits or four or more tubes in a ring oscillator are sometimes used. In push pull circuits, the tuned circuits are the parallel line (Lecher line) type. This arrangement forms a balanced-to-ground system.

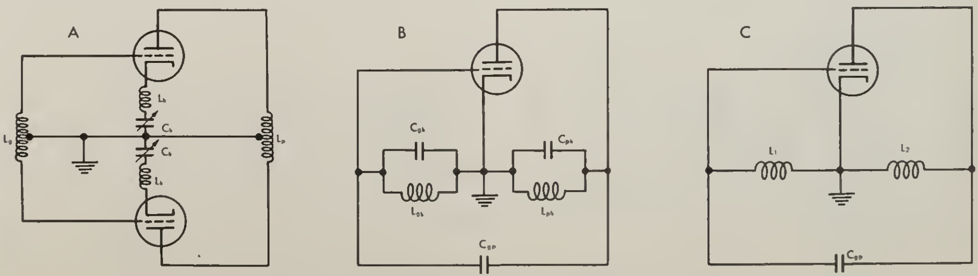
The oscillator circuit discussed first is the push pull tuned-plate tuned-grid oscillator which was used in an early radar set operating in the UHF range. Only the oscillatory action is described here. Its pulsing circuits are taken up later in this chapter.

TUNED-PLATE TUNED-GRID OSCILLATOR. In the tuned-plate tuned-grid oscillator, both the plate and grid lines are slightly under λ/4 in length. They are indicated in the circuit of the tuned-plate tuned-grid oscillator as inductances L_p and L_g. The length of the plate line in the circuit may be varied by the shorting bar. While the grid line is of fixed length in this particular arrangement, a little greater flexibility would result if it too were adjustable in length. The inductances L and capacitors C_1, C_2, C_3, and C_4 in the cathode circuits form filters to prevent RF from reaching the filament transformer. The grid leak bias set up includes R_g and C_g. The component C_g is shown as a dotted portion since it is actually the distributed capacity of the grid line to ground. B+ voltage is applied to the plate lines through R_s and L_s. L_s serves the double purpose of preventing RF from getting into the power supply and suppressing parasitic oscillations.



Tuned-Plate Tuned-Grid Oscillator using Lecher Lines as Tuned Circuits

Equivalent Circuits. In equivalent circuit A, below, both tubes of the oscillator are shown. L_p is the inductance, which represents both the plate line and the plate lead inductances. L_g represents the inductance of the grid line and grid leads. L_k is the cathode lead inductance. With proper adjustment of C_k, L_k forms a series resonant circuit that effectively grounds the cathode within the tube.



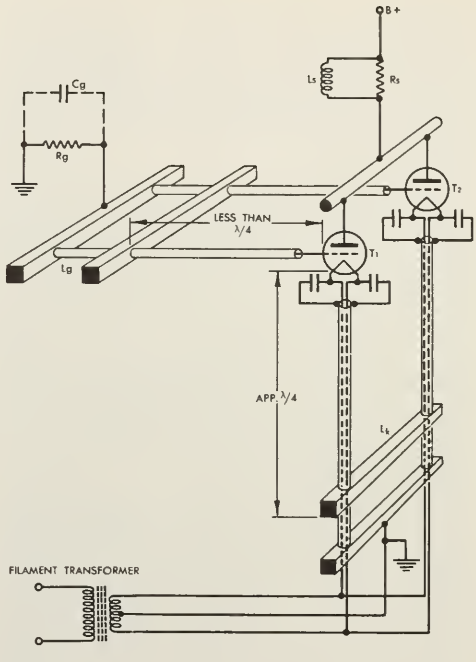
Equivalent Circuits of TPTG Oscillator

The second equivalent circuit B shows only one tube since both circuits are identical. L_{pk} is one half L_p and along with C_{pk} forms a circuit that resonates at a frequency above that at which the entire circuit oscillates.

In the third equivalent circuit C the parallel combination of L_{pk} and C_{pk} is represented as an inductance L_2. Likewise L_{gk}, which is one half L_g and along with C_{gk} forms a circuit which resonates at a higher frequency than that of the oscillator, is represented as an inductive at the oscillator frequency. This is true because the tank circuit must have at least one inductive element to oscillate and the two reactances must be alike for the feedback to be in the correct phase. In this equivalent circuit, the oscillator is the Hartley type. Since the oscillator frequency must be lower than either resonant frequency (plate or grid circuits), the grid or plate circuit, whichever is lower in resonant frequency, controls the oscillator frequency and the other controls the amount of feedback.

TUNED - GRID TUNED - CATHODE OSCILLATOR. Another push pull oscillator circuit which is used in a number of types of UHF radar equipment is the tuned-grid tuned-cathode circuit. In this oscillator, the plates are effectively short-circuited by tubes which have very little plate lead inductance and which are connected by a section of line. When the lead inductance is large enough to become objectionable, then the plates are connected by means of a half wave section of line to insure a short circuit. The DC circuits are the same as in the tuned-plate tuned-grid oscillator.

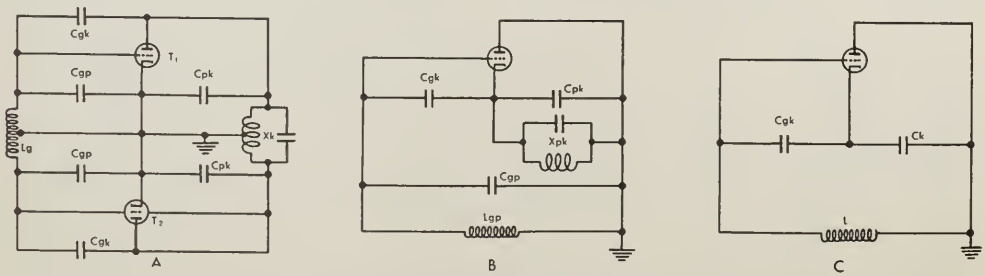
Equivalent Circuits. In the first equivalent circuit A below both tubes are shown. The grid line is represented as inductance L_g, the cathode line as reactance X_k, and the interelectrode capacitances as indicated.



Tuned-Grid Tuned-Cathode Oscillator

Circuit B shows only one tube. In this circuit L_{gp} represents one half of L_g and X_{pk} represents one half of X_k. By following the line of reasoning similar to that used in the TPTG circuit, you see that the parallel combination of X_{pk} and C_{pk} must be capacitive at the oscillator frequency. Thus it is shown as condenser C_k in the third equivalent circuit C.

The oscillator must have an inductive element; hence L_{gp} and C_{gp} in parallel act like an inductance L as shown in C. For L_{gp} and



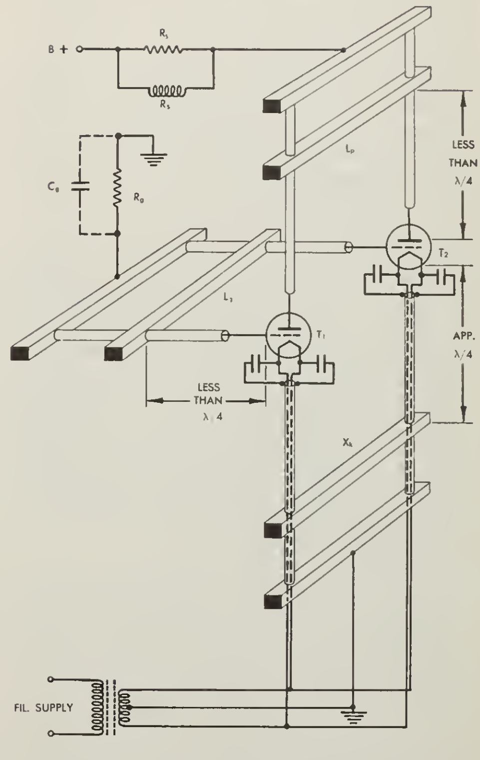
Equivalent Circuits of Tuned-Grid Tuned-Cathode Oscillator

C_{gp} to be inductive, the grid line must always be less than $\lambda/4$ in length. Since variations in Lo have the greatest effect on the frequency of oscillations, *the grid line is the primary frequency tuning device.* The cathode line (X_{pk}) controls only a portion of the total capacitance of the circuit and therefore does not greatly affect the oscillator frequency. However, since X_{pk} is in parallel with C_{pk}, forms a voltage divider with C_{gk}, it has a *considerable effect of feedback.* X_{pk} may be inductive, purely resistive, or capacitive depending on whether the cathode line length is less than, equal to, or greater than a quarter wavelength at the frequency of oscillation. If X_{pk} is a high resistance, the feedback is determined entirely by the voltage divider composed of C_{gk} and C_{pk}. If X_{pk} is capacitive (greater than $\lambda/4$), then the two capacitances in parallel combine to increase C_k and results in an increase in feedback. Tuning the cathode line to slightly less than $\lambda/4$ makes it inductive and decreases C_k, which results in a reduction of feedback. Oscillations will cease if X_{pk} and C_{pk} in parallel can no longer be represented as a condenser C_k. As can be seen by the third equivalent circuit, this oscillator is basically a Colpitts oscillator.

Advantages. The tuned-grid tuned-cathode type oscillator has three advantages. First, the cathode line offers a convenient means of connecting the load to the oscillator because it is at ground potential with reference to DC. Second, since only DC voltages appear at the plate and only RF voltages appear in the cathode circuit, the two are not present in the same circuit and the high peak voltages present in a tuned plate circuit are avoided. Third, the tuning of the oscillator is quite simple, the grid line being the frequency control and the cathode line the feedback control. There is some interaction between the tuning controls so that readjustment of each may improve both accuracy and output.

Tuned-Plate, Tuned-Grid, Tuned-Cathode Oscillator Circuit

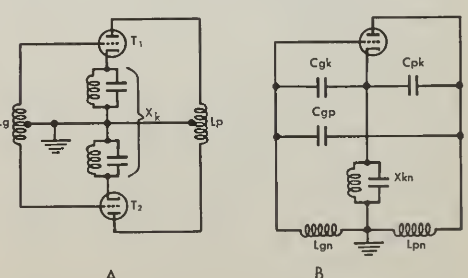
A number of radar oscillators in the UHF range employ tuned lines in all three leads of the triode, as in the oscillator circuit above. So far as RF is concerned, all three tuned lines are grounded at their midpoints. The DC circuits and filament circuits are identical to those described previously. In the circuit shown, the plate and grid lines, which are less than a quarter wavelength long, are labelled L_p and L_g respectively. The cathode line, which may be



Tuned-Plate, Tuned-Grid, Tuned-Cathode Oscillator Circuit

equal to, less than or greater than a quarter wavelength, is labelled X_k.

Equivalent circuit A below shows both tubes, the inductances L_p and L_g center tapped to ground, and the unknown reactance X_k center tapped to ground. The interelectrode capaci-



Equivalent Circuits of T-P T-G T-K Oscillator tances are not shown in A. Since the circuits of the two tubes are identical, the second equivalent circuit shows only one tube. The inductances L_{pn} and L_{gn} and the reactance X_{kn} represent one half of L_p, L_g, and X_k. The interelectrode capacitances are added. Examination of circuit B shows that if the cathode line is exactly a quarter wavelength so that X_{kn} becomes a high resistance, then the circuit is that of a conventional Colpitts oscillator. If the ratio of L_{gn} to L_{pn} is the same as the ratio of the reactances of C_{gk} and C_{pk}, there is no difference in potential between the cathode and ground and X_{kn} is in effect out of the circuit. This condition also reduces the circuit to that of a Colpitts oscillator. In either case, the frequency of oscillations is determined by the sum of the inductances of the plate and grid lines, and the amount of feedback by the ratio of C_{gk} and C_{pk}. Ordinarily this ratio does not give the amount of feedback necessary to cause oscillations; for best efficiency, therefore, the cathode line is detuned from the pure resistive quarter wavelength so that X_{kn} is capacitive or inductive.

Making X_{kn} capacitative or inductive while keeping the sum of L_{gn} and L_{pn} constant does not change the frequency of oscillation but does increase feedback. The exact amount of feedback depends on the ratio between L_{gn} and L_{pn}. Thus, for example, consider the case where L_{pn} is reduced to zero and examine the nature of X_{kn}. In this case X_{kn} is in parallel with C_{pk} as shown in equivalent circuit B of the TGTC oscillator on page 11-8. Thus, in controlling feedback, the effect of X_{kn} changes is the same as in the tuned-grid tuned-cathode oscillator—that is, the cathode line must be increased in length to increase feedback or be decreased in length to decrease feedback.

On the other hand, if L_{gn} is decreased to zero and L_{pn} is increased to equal the value of their former sum, then X_{kn} is in parallel with C_{gk}. Its effect on feedback then is just the opposite to that described. For example, if the cathode line is greater in length than a quarter wave, it becomes capacitive and the parallel combination C_{gk} and X_{kn} becomes a larger capacitance. This reduces the feedback. Reducing the length of the cathode line below a quarter wavelength makes it inductive, and the parallel combination becomes a smaller capacitance. This results in an increase in feedback. This demonstrates that *feedback is a function of both the ratio of L_{gn} and L_{pn} and of the tuning of the cathode line.* Since the frequency of oscillations is controlled by the sum of L_{pn} and L_{gn} and feedback both by their ratio and the cathode line length, there are numerous settings of the three variable line lengths for any given frequency of oscillations and for proper feedback. This is readily understood if you consider that you have three variable elements with which to control two variables, and that there is a certain amount of interaction among the three controls.

Pulsing of Triode Oscillators

There are two methods of pulsing UHF oscillators. One method uses separate modulator tubes; the other employs a self-pulsing oscillator.

THE SELF-PULSING OSCILLATOR. Most oscillators use grid leak bias or a combination of grid leak and fixed bias. An advantage of grid leak bias is that the amplitude of oscillations is limited and stabilized. Although grid leak bias was taken up in Chapter 4, it is discussed briefly again to introduce the subject of self-pulsing. In this connection notice the graphs of plate current and grid voltage versus time on page 11-11. These waveshapes are the same for all oscillators which use grid leak bias. At zero time, plate voltage is applied but the grids are grounded so that a steady plate current flows. At time t_1 the ground is removed, and any random change in plate current will cause a change in voltage in the plate tank circuit, a portion of which is fed back to the grid circuit. Since this feedback is in the right phase and amplitude, the oscillations will grow in amplitude. Now each time the grid goes positive, grid current flows and charges the grid leak condenser more negatively so that on each successive oscillation the average grid voltage (the bias) goes more negative. This continues until a state of balance is reached where just enough grid current is drawn to replace the charge which leaks off through the grid leak resistor during the remainder of the cycle. In addition as the average grid voltage becomes more negative the plate current flows for a shorter portion of each cycle until the flow is just sufficient to replace the energy lost in the tank circuit and to supply feedback to the grid circuit.

With perfect circuit elements and smooth electron emission, the circuit would oscillate indefinitely at this amplitude. But minute irregularities occur in the movement of electrons through resistors and conductors; besides the emission from a cathode is always in non-uniform bunches. The non-uniformity is slight, but enough to affect the equilibrium in the oscil-

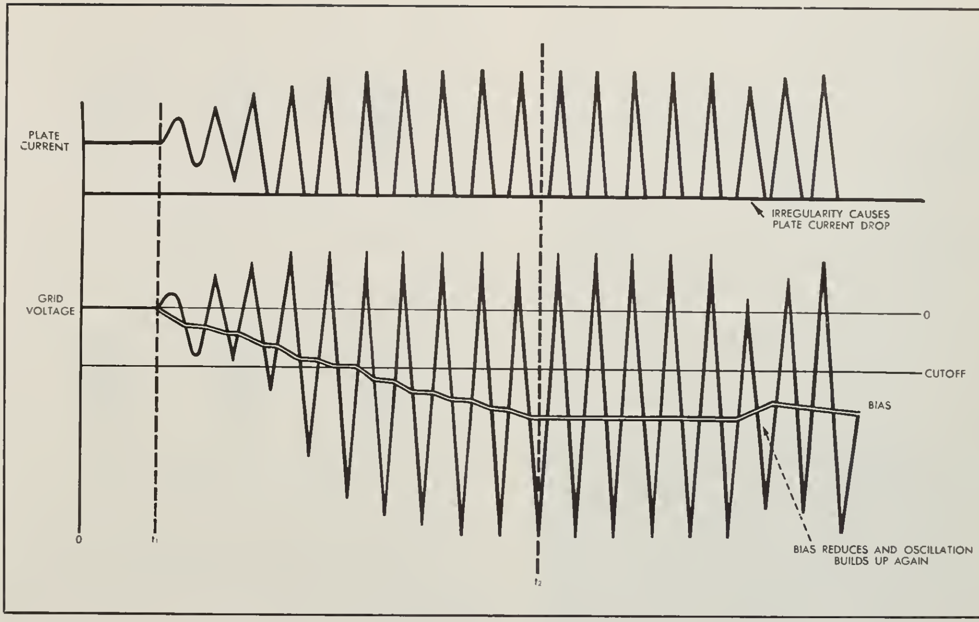
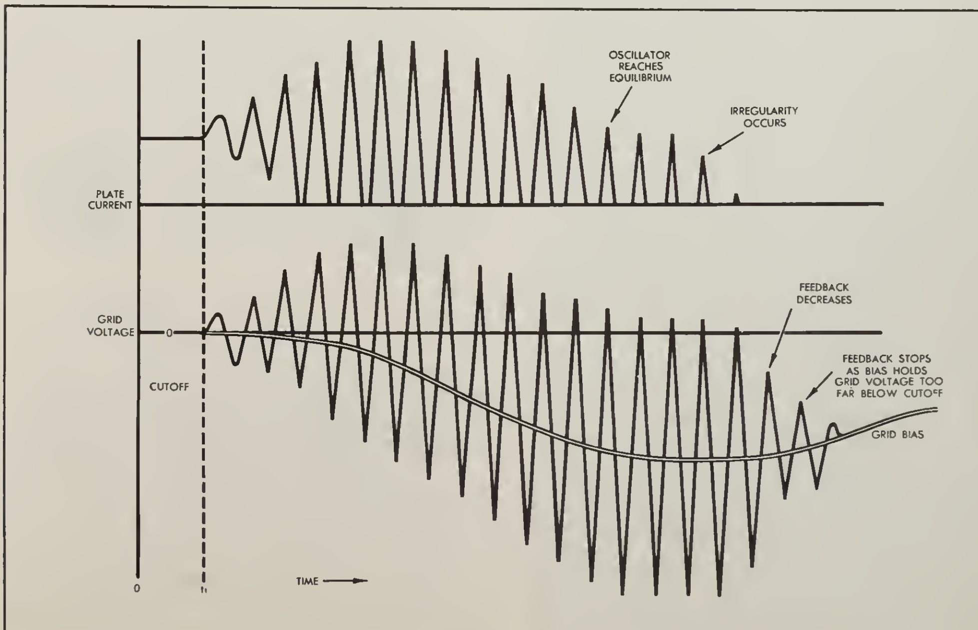


Plate Current and Grid Voltage vs Time showing Build up of Oscillations

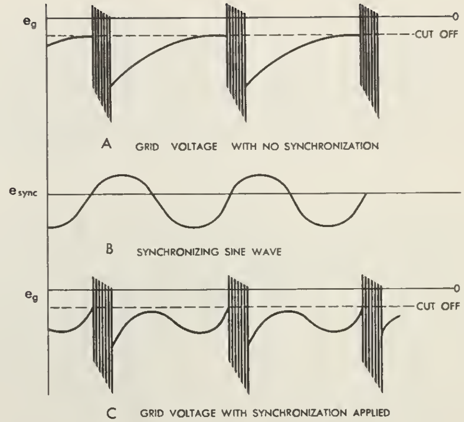


Effect of Long RC

lator circuit. Any slight decrease in current decreases the feedback, and if the time constant of the RC circuit is not too long, the bias is reduced slightly by discharge of the condenser. The decrease in bias allows a little more plate current, which re-establishes the equilibrium at the old level.

In a circuit with a long RC, however, the circuit starts oscillating, then settles down to a balanced condition as previously described, but any slight decrease in plate current now will not be compensated for by a lowered bias as shown in the bottom graph on page 11-11. The condenser discharges too slowly. The resulting continued high bias prevents the feedback from being restored to the proper level and after a few cycles, the oscillation dies out competely. The condenser slowly dischärges and when its voltage gets below the cut-off voltage, the current again flows and the process repeats. For continuous operation then, the RC of the self biasing circuit must be short enough to allow automatic adjustment of the bias for changing conditions in the circuit. To make the circuit oscillate briefly, then stop, use a long RC. You can see that the time the tube remains cut off depends upon both the grid condenser and grid resistance. By properly choosing these values, you can have the oscillator operate for one or two microseconds and then cease operation for 500 to 1000 microseconds as you desire. You can get an idea of how the variations of grid voltage would look from the above diagram. It is not possible to show the actual time relations since the "off" period is several hundred times as long as the "on" period.

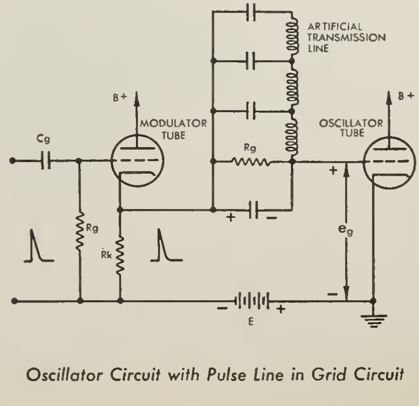
In a radar set which uses a self-pulsing oscillator, synchronizing pulses are taken from the oscillator for starting the sweep on the CRT and for any other circuits which need to be synchronized. In most cases, the exact pulse recurrence frequency is unimportant. There are a few sets, however, which use the 400 cycle per second power supply to regulate the PRF. This is done by applying a small synchronizing voltage to the grid as shown above at B and C. Without synchronization, the grid rises exponentially and crosses the cut off line at a small angle which means that the actual time that oscillations begin may vary somewhat. Notice at C that the point at which the grid voltage crosses the cut off bias line is much more definite. Other ways of controlling the PRF more exactly are to feed in a positive going pulse



Grid Waveshape With and Without Synchronization

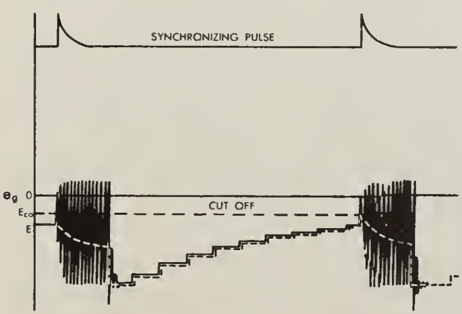
when the grid voltage approaches cut off or to return the grid to a positive voltage rather than to ground.

In order to make the duration of the pulse more definite, some self-pulsed oscillators use an artificial transmission line instead of the grid leak condenser. This has the added advantage of making the operating conditions more nearly constant throughout the pulse. There are two places where the pulse line can be located. One is to put it in the grid circuit as shown below; the other in the cathode circuit. (The construction and operation of artificial transmission lines was explained in Chapter 9.)



Oscillator Circuit with Pulse Line in Grid Circuit

A pulse line behaves like a long transmission line. Your chief interest in it is that it acts like a resistive circuit until the charging pulse has traveled the length of it and returned, at which time it is fully charged. This causes the voltage to change in two steps, the first from the bias E to the voltage determined by the product of the grid current times the characteristic resistance of the line and the second step of equal magnitude. The PRF is controlled by the synchronizing pulses but the discharge of the pulse line through the grid resistor must be such that the grid voltage approaches the cut off point by the time the next synchronizing pulse comes in. The discharge of the line is in steps, as you can see below in the diagram of the grid waveshape of a line-controlled self-pulsed oscillator. The magnitude of the discharge voltage is determined by the relative values of resistance of R_g and the characteristic resistance of the line. R_g is usually much larger, hence the steps are small. These steps follow the general path of the exponential discharge curve with time constant R_gC_o where C_o is the capacitance of the pulse line.



Grid Waveshape of a Line Controlled Self Pulsed Oscillator

SEPARATE MODULATORS. As mentioned earlier, some triode UHF oscillators are not self-pulsed, but employ separate modulators. These modulators are very much like those used with magnetron oscillators, which are discussed later in the chapter. Here merely note that the oscillator may be plate modulated, grid modulated, or cathode modulated. In the above circuit for plate modulation, short positive rectangular pulses are applied to the grid of the modulator tube resulting in a negative-going pulse at the plate. The transformer is connected so that plate

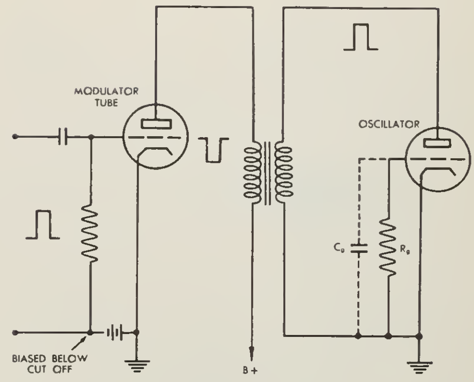
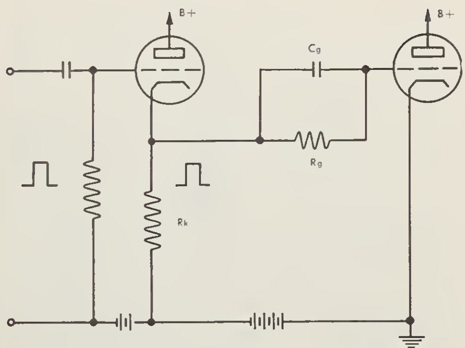


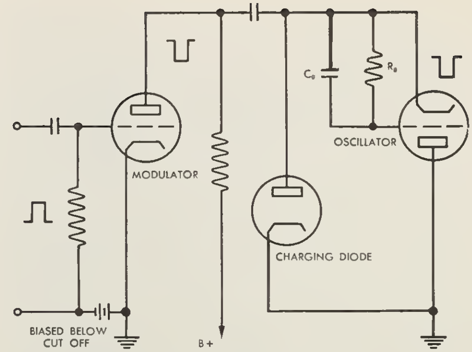
Plate Modulated Oscillator

voltage is applied to the oscillator tube or tubes in short rectangular pulses. RF components for the oscillator are not shown in the diagram. The transformer is specially constructed so that it passes the rectangular pulse without undue distortion. Since the purpose of the transformer is polarity inversion, the question arises in regard to why not apply negative going pulses to the grid of the modulator tube to cut it off and get a positive going pulse at the plate for use in modulating the oscillator. This would require that the modulator tube, or tubes, conduct heavily during all the cycle except during the pulse and would result in excessive waste of power. The grid capacitance C_g of the oscillator is quite small, usually consisting of only the capacitance to ground of the grid tuned circuit. If made too large, C_g, might cause the oscillator to be self-pulsed and cut itself off before the modulator pulse is over. Hence C_g and R_g have values to cause the oscillator to operate CW if plate voltage were applied all the time. The modulator pulse then is most effective in controlling the output pulse. Oscillations must build up rapidly so that the leading edge of the output pulse will be as steep as possible. The plate voltage of the oscillator must remain constant throughout the pulse.

A cathode-pulsed (modulated) oscillator does away with the necessity of inverting the pulse, hence no transformer is used. The charging diode serves to recharge the coupling condenser between pulses. The condenser must have quite a large capacitance so that it will not discharge appreciably during the pulse and thereby lower the plate-to-cathode voltage.



Cathode Modulated Oscillator

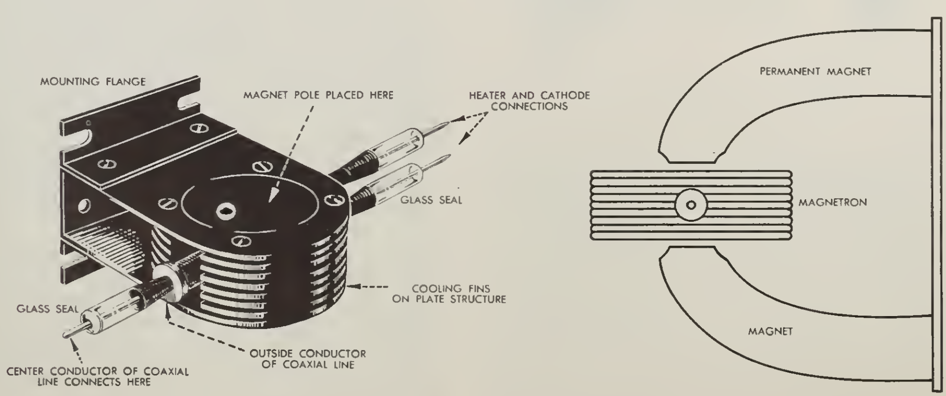


Grid Modulated Oscillator

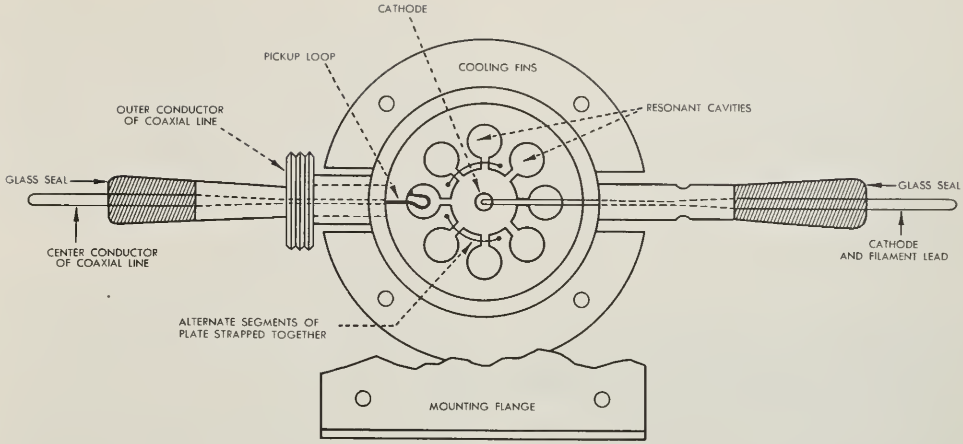
Grid modulation requires less power from the modulator for operation than either of the other two types of modulation. In grid modulation a positive going rectangular pulse is applied to the grid and raises the grid voltage well above the cut off point and oscillations start. Oscillations end whenever the grid voltage drops to the normal bias. Usually the arrangement for this purpose is the one in which oscillations are started by a positive trigger pulse and ended by a self-pulsing grid bias voltage. C_g in the grid-pulsed oscillator shown is a small capacitance which is in addition to the grid to ground capacitance. Its purpose is to couple the pulse from the cathode of the modulator tube to the grid of the oscillator to insure a rapid rise of voltage at the grid and cause oscillations to build up quickly.

THE MAGNETRON OSCILLATOR AS TRANSMITTER

Practically all radar transmitters which operate above 600 mc use magnetrons for transmitting tubes. The main advantage of the magnetron is that transit time, which is a determining factor in the upper limit of frequency of operation in conventional tubes, is not a disturbing influence. There are some special type triodes employed in unusual transit time oscillator circuits that will oscillate in the frequency range useful for radar but unlike the magnetron their power output is too small for them to be useful in most cases. In addition, there are also the velocity-modulated tubes, such as the Klystron and the Shepherd Pierce tube, which operate in the desired frequency range with better power output than the triodes just mentioned, but they are not readily adaptable for pulsed operation.



Typical 10cm Magnetron



Cut-Away View of 10cm Magnetron

The magnetron is a diode in which the magnetic field between the cathode and the plate s perpendicular to the electric field. The tuned circuits included in the tube are in the form of cylindrical cavities. As shown in the typical 10 cm magnetron on page 11-14, the cathode and filament structure is at the center of the tube and is supported by filament leads, which are large and stiff enough to keep the cathode and filament structure fixed in position under ordinary circumstances.

Exercise care in handling the magnetron tube. Undue jarring of the tube or bending the filament leads might change the position of the cathode causing it to short to the plate, or change its spacing with the plate. In either case the operation of the tube would be materially affected. The plate structure as shown above is a solid block of copper in which the resonant cavities are cylindrical holes. A narrow slot opens each cavity into the central portion of the tube and divides the inner plate structure into as many segments as there are cavities. Alternate segments are strapped together to put the cavities in parallel so far as the output is concerned. This makes it possible to take the output from a pickup loop placed inside any one of the cavities. Since the outer conductor of the output coaxial line is connected to the shell of the magnetron, which is a part of the plate structure, there cannot be a high positive DC voltage applied to the plate. The same result is realized whenever negative voltage is applied to the cathode. The magnets, which are made of a special alloy called Alnico, are unusually strong. Protect them from sharp blows of metallic substances or any hard material, since such blows will cause them to lose magnetism.

Operation of the Magnetron Oscillator

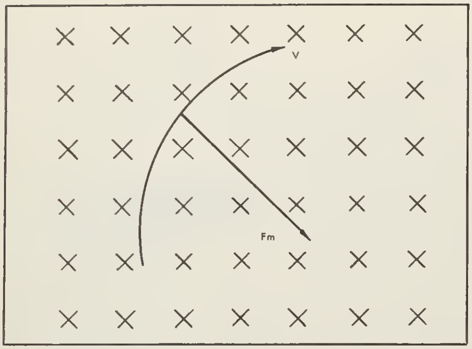
The theory of operation of the magnetron is based on the motion of electrons in combined electric and magnetic fields. The following laws govern this motion.

In electric field alone. The law governing the motion of an electron in an electric field states that the force exerted by an electric field on an electron is proportional to the strength of the field and the direction of the force is opposite to that of the field. In other words, electrons tend to move from low potential toward high potential. Mathematically this law is stated by the vector equation,

$$F_e = 1.6 \times 10^{-12} E \qquad (1)$$

where E is the electric field intensity in volts per centimeter, and F_e, the force in dynes.

In magnetic field alone. The law of motion of an electron in a magnetic field states that the force exerted on an electron in a magnetic field is at right angles to both the field and the path of the electron; the direction of the force being such that the electron trajectories are clockwise, when viewed in the direction of the magnetic field as in the following diagram.



Direction Fm of the force acting on an electron moving at a velocity V at right angles to a magnetic field directed into the page

Mathematically the law governing the magnitude of a force in dynes is expressed,

$$F_m = 1.6 \times 10^{-20} V_p B \qquad (2)$$

where V_p is the component of the velocity in centimeters per second of the electron perpendicular to the direction of the magnetic field, and B the magnetic flux density in gausses.

In regions of uniform magnetic field the electrons move in circles of radius according to the equation,

$$R = 5.68 \times 10^{-8} \frac{V_p}{B} \text{ centimeters} \qquad (3)$$

with a period of

$$T = \frac{35.7 \times 10^{-8}}{B} \text{ seconds} \qquad (4)$$

IN MUTUALLY PERPENDICULAR ELECTRIC AND MAGNETIC FIELDS. The path that an electron takes when traveling normally with respect to the magnetic field depends upon the initial velocity of the electron. Generally, however, the paths of electrons in a plane normal to the magnetic fields are cycloids (figures generated by points on the radius of a wheel rolling on a smooth plane). Different paths are obtained by different points along the radius of the wheel. For example, the center point of a wheel generates a straight line. A point on the circumference generates a typical cycloid as shown below. Points lying beyond the radius generate cycloids also, but these are closed loops.

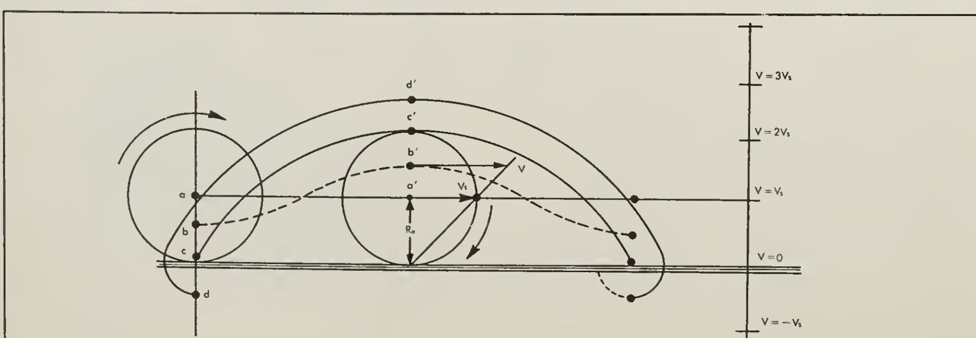
Paths Generated By Electrons

All paths generated by the rolling wheel may also be generated by electrons traveling at different velocities. This is understandable if you consider that the uniform magnetic field is directed inward perpendicular to this page and that the electric field is directed downward from the top to the bottom of the page. If an electron is projected horizontally from left to right, there will be two forces acting on it—a magnetic force downward and an electric force upward. If these forces are equal, the path of the electron E will be a straight line. Equating equation (1) and (2) gives the velocity required to produce a straight line path. The result is expressed by the equation,

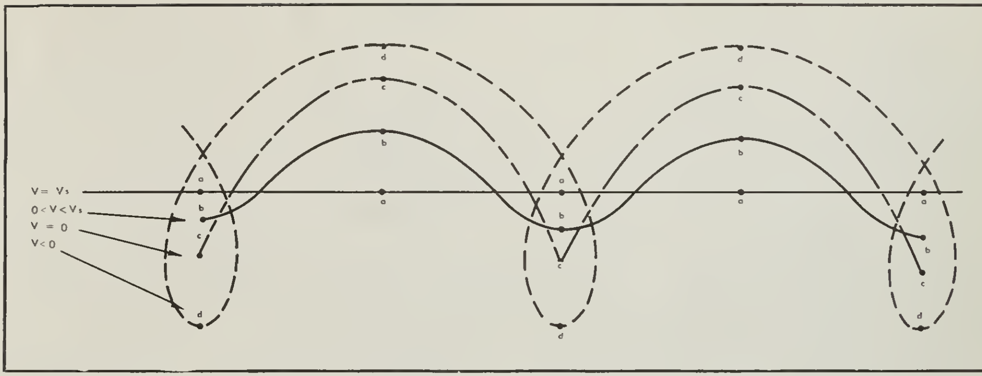
$$V_p = V_s = 10^8 \frac{E}{B} \qquad (5)$$

where the velocity to produce a straight path is V_s.

For all other velocities, the electron path will be curved as shown on page 11-17. All electrons traveling with velocities between zero and V_s take a path similar to that shown in curve b. The qualitative explanation of this path is the following: Because the initial velocity of the electron is less than V_s, there is a smaller magnetic force acting on it. Hence, the elec-



Showing how cycloids may be generated by points on a rolling wheel. When a point is moving horizontally, its velocity is proportional to its distance above the surface.



Four possible orbits for electrons moving in a plane at right angles to a uniform magnetic field which is perpendicular to a uniform electric field

tron curves upward at first. But as it moves, it gains kinetic energy from the electric field with a resultant increase in velocity. The magnetic force thus increases and after a time the electron path is bent downward where the electron begins to lose velocity, and finally returns to its original state.

Curve c above shows the path which an electron takes when it starts with a zero velocity. At the start the magnetic field exerts no force on the electron as no magnetic lines of force are being cut. The electric field, however, moves the electron directly upward to a point where the magnetic field acts on it and deflects the electron to the right. As the electron moves upward, it gains velocity, the magnetic force increases, and eventually becomes large enough to bend the trajectory back until the velocity of the electron again becomes zero and has the same potential in the electric field as that at which it started.

Finally, an electron traveling from right to left. (mathematically, a negative velocity) is at first pushed upward by the magnetic and electric fields as you can see in curve d. Its initial radius of curvature is small. Soon it goes straight upward gaining velocity rapidly. As the magnetic force increases steadily, the electron is pushed to the right and finally downward. The path is bent back and the electron finally returns to its initial condition.

The time for each cycle and the horizontal distance covered by each electron in one cycle is the same for each of the cases given.

All the paths above are cycloids. It follows that electron a corresponds to the center of a wheel rolling on a smooth plane. Electrons b, c and d correspond to a point on the radius, to a point on the circumference, and to a point outside the wheel, respectively. The period of the wheel is identical to the period of an electron in the magnetic field alone, according to equation 4 on page 11-16.

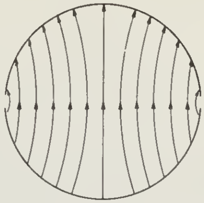
Source of Energy in Magnetron Oscillators

In an electric field alone, a free electron which always tends to move parallel to the field, gains a large amount of kinetic energy. In combined electric and magnetic fields, the electron, which on the average moves at right angles to the field, does not change its average kinetic energy. The magnetic field always exerts a force which tends to change the direction of motion of the electron.

In a magnetron oscillator there are two components in the electric field—a steady component furnished by the DC source and an alternating component furnished by the oscillations of the tank circuit. In a properly designed magnetron, the electrons absorb energy from the DC component of the field and deliver it to the alternating component. The function of the magnetic field in many respects is the same as that of the positive feed back to the grid of the triode oscillator.

The Split-Anode Transit-Time Magnetron

The two segmented-plate cylindrical anode magnetron is a practical oscillator from which the multi-anode magnetron using resonant cavities as tank circuits was developed. In it the electric field is the result of the difference of potential between the two plate segments. In addition to this field, there is a steady field (not



Electric Field between Plate Segments

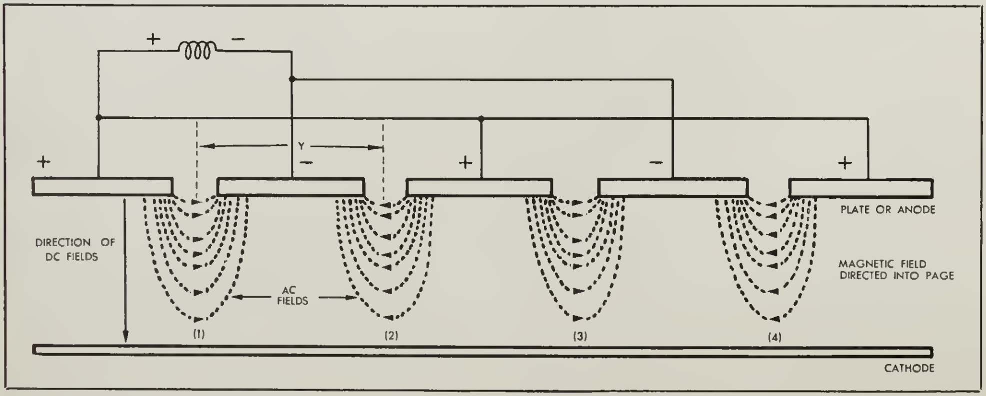
shown), which is due to the potential difference between the cathode and the two anodes. The illustration below shows the motion of an electron that delivers energy to the source on alternating potential between the magnetron plates. Note that the electron trajectory is approximately cycloidal.

Path of an Electron Delivering Energy to the Alternating Field

As you can see, the electric field between plate segments may either increase or decrease the angular velocity of an electron, depending upon the angular position of the electron. If an alternating voltage having a period equal to the time taken for an electron to move once around the filament is impressed between the anode segments, the direction of the field reverses twice in each revolution of the electron about the filament. As a result, the electrons that pass the gaps at the instants at which the field is maximum will experience continuous angular acceleration or deceleration. Electrons that are accelerated gain energy from the source of alternating voltage. The electrons that absorb energy are speeded up and the component of force toward the filament also increases. As a result, these electrons enter the cathode after one excursion. Electrons that deliver energy, on the other hand, are decelerated. This decrease in angular velocity reduces the average force toward the cathode, and causes electrons to drift toward the plate as shown to the left. Experiments have proved that electrons which give up energy make on an average between 80 and 100 excursions. Since the electrons which deliver energy make so many more radial oscillations before striking the plate than the ones that absorb energy, the source of alternating voltage gains energy. Sustained oscillations will take place whenever the oscillatory circuit has sufficiently low dissipation, such as a Lecher line or a resonant cavity.

The Multi-Anode Transit-Time Magnetron

PLANE FORM. Since the exact theory of the cylindrical multi-anode magnetron is more complicated than that of the plane form, the plane form is discussed first. The ideas presented for the plane form may be carried over in the discussion of the more practical cylindrical multi-anode magnetron.



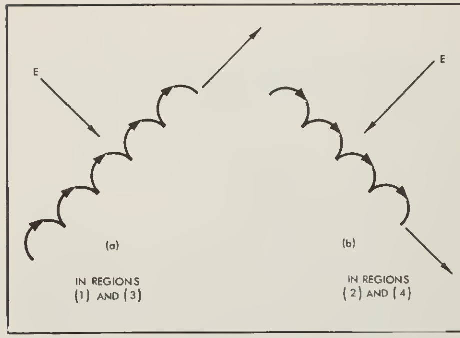
AC Electric Field in Plane Magnetron with Multi-Anodes with Maximum Potential between Adjacent Anode Segments

To understand the operation of the plane form magnetron consider a magnetron oscillator which is composed of a continuous plane cathode and a segmented anode of which the alternate sections are connected to the opposite sides of its tank circuit. In the illustration on page 11-18 showing this magnetron, note that the alternating electric field is sketched in the space between the cathode and anode at the instant when alternate anode segments are at their maximum positive and negative values.

Important considerations in the operation of this magnetron are which electrons tend to sustain oscillations, which ones will absorb energy, and their probable paths. In this connection consider an electron in a uniform electric field directed as shown in the region (1) in the preceding magnetron illustration. The direction of the resultant field is obtained by adding the AC and DC fields at point (1) vectorially. (The DC field, which is not shown, is directed downward.) The approximate direction of the electric field and the path of the electron are shown in the above illustration.

As shown in this illustration, the electron makes cycloid paths. Since the general progression of a cycloid is always at right angles to the resultant electric field, these cycloids progress upward and to the right, with the magnetic field directed into the page. On the other hand, an electron in a uniform field whose direction is the same as the resultant shown in region (2) in the magnetron shown on page 11-18 would progress downward and to the right as shown above at B. In the diagram showing this, the direction of the uniform electric field is also indicated. Since the electron of diagram A above moves on the average, from cathode to anode, it must absorb energy from the source of steady field. This electron, however, tends to deliver energy to the alternating field since it moves in a direction which is opposed by the field of the tank circuit. Because the average velocity V of an electron does not change as it progresses toward the anode, it is a very efficient means for converting energy from the DC source into energy of oscillation for the tank circuit. By the same reasoning, the electron in region (2) tends to absorb energy from the alternating field and to transfer it to the steady field.

Whenever oscillation is to be maintained, the electron in diagram A must be made to continue its path, while the other electrons—the ones which absorb energy—must be removed from the field quickly. An electron of the type in diagram B, starting from the cathode with a small or zero velocity, will strike the cathode before completing its first loop. Had the fields of regions (1) and (2) been static, the electron of diagram A would tend to turn downward as it approached region (2). Since the fields are alternating, this will not necessarily be the case. In fact, if the left-to-right velocity of the electron of diagram B is such that it arrives in region (2) in exactly one-half cycle, it will encounter there exactly the same kind of field that it encountered in region (1) one-half cycle earlier. Thus, if this electron continues to progress distance Y shown in the magnetron on page 11-18 from left to right during each half-cycle, it will go all the way to the anode, following a path similar to that of an electron in diagram A. This condition is necessary for oscillation, since the electron of diagram A contributes energy to the tank circuit for a much longer time than the electron of diagram B takes energy away from the tank circuit.



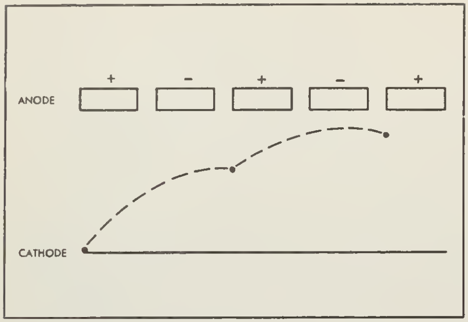
Electron Paths

The following analyzes the factors affecting the frequency of oscillation in a plane magnetron. The average left-to-right velocity Va of the electron depends only on the average downward component of the electric field. The DC field is the average electric field. Since an electron should progress horizontally at a distance of Y centimeters in half a period, T/2, it follows that

$$(Va)\frac{T}{2} = Y \qquad (6)$$

and $$f = \frac{1}{T} = \frac{Va}{2Y} = \frac{(10^8 E)}{B}\frac{1}{2Y} \qquad (7)$$

where $Va = \frac{10^8 E}{B}$ *from equation (5) and E is the DC component of the electric field.*



Showing that if the time for one loop equals the period of the tank circuit, each loop will have a length equal to two anode segments

These conditions clearly indicate that the plane magnetron is a transit-time oscillator, for an electron is allowed a certain definite time to travel from cathode to anode. Thus, in addition to the usual condition of resonance, $f = \frac{1}{2\pi\sqrt{LC}}$, there is equation (7) to be fulfilled also. When this is done you may rewrite equation (7) as

$$f/(E/B) = const. \qquad (8)$$

The question arises whether you could obtain greater efficiency if in addition to adjusting the ratio (E/B) to the frequency as required by equation (8) or (7), you also made the time for one cycloidal loop equal to the period of the tank circuit. When both of these conditions are filled, the lengths of the cycloidal loops become equal to 2Y or the length of two anode segments as illustrated above. It turns out, however, that the efficiency is not increased by doing this.

This is why the efficiency is not increased. Since an electron returns to zero kinetic energy at each cusp, the efficiency of conversion from energy to tank circuit energy is 100 percent from cusp to cusp. The only DC energy which is not converted into tank-circuit energy by electrons which reach the anode is that which appears as the kinetic energy which arrives at the anode. As this kinetic energy is all acquired after the last cusp, the closer the last cusp is to the anode, the less will be the energy lost relative to the total available DC energy. Therefore, the smaller the cycloids, the smaller will be the fraction of energy lost as kinetic energy at the plate. Thus, for these reasons efficiency is not increased by making the looping frequency equal to the tank-circuit frequency.

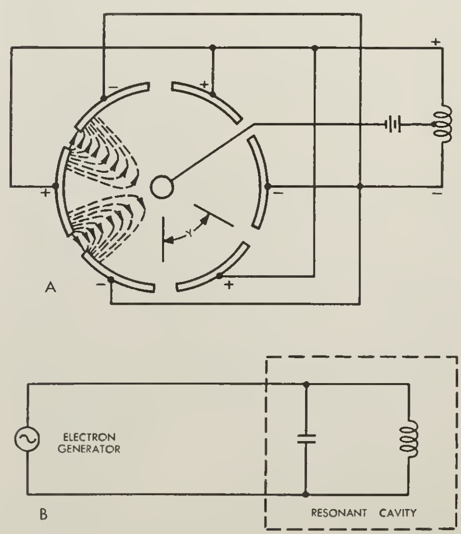
To increase the efficiency in a magnetron, it is necessary to make the cycloids smaller. As can be proved from the properties of a rolling wheel, the size of the orbits depends on the radius of the wheel R, and the equation,

$$R = 5.68\frac{E}{B^2} = 5.68\frac{(E)}{B}\frac{(1)}{B} \qquad (9)$$

Thus, it follows that for smaller cycloids the value of B must be increased for a given value of (E/B). As previously stated the ratio (E/B) is fixed for a given frequency, as this is one of the conditions for oscillation as you recall was indicated in equation (8).

In practice it is possible for the value of B to be 1.5 to 3 times the value which makes the looping frequency equal to the tank-circuit frequency. Not only does increasing B increase the efficiency, but it also increases the power input since E must be increased in direct proportion to B in order to keep E/B constant.

CYLINDRICAL FORM. The cylindrical form magnetron is analogous to the plane magnetron rolled up into a circle. The number of segments (anodes) which may be used in it varies widely. As many as 32 have been tried on certain occasions. Here, however, for simplicity, the discussion considers only a magnetron with six segments. The illustration below shows the circuit of a six anode magnetron and the approximate AC field between the segments and its equivalent circuit.



(A) Simplified tank circuit with six anode magnetron, showing the alternating electric field
(B) The equivalent AC circuit in which the generator corresponds to the electrons between anode segments

As mentioned previously, the exact theory involved in the cylindrical magnetron is more complicated than the theory of the plane magnetron. The discussion for that reason does not go into this theory. However, you should at least qualitatively understand the performance of the cylindrical magnetron. For this purpose, assume that the DC field is uniform and that its magnitude is Eb b-a, where Eb is the DC supply voltage, b the anode radius in centimeters and a the cathode radius. Since the AC field is much narrower near the cathode than near the anode in the cylindrical case, as shown at A, the linear distance (the distance which the electron must progress around the cathode per anode segment) is less near the cathode than near the anode. If the theory of the plane magnetron is to be applicable to the cylindrical magnetron, it will be necessary to choose an effective or average value of Y. This value is one-sixth of the circumference of the circle halfway between the cathode and anode. These assumptions about Y and the electric field lead to the same resonance condition as before—namely, equation (8) on page 11-20. In addition, as with the plane magnetron, it is inefficient to fulfill the resonance condition for B alone.

The following table shows quantitatively the actual efficiencies of typical individual electrons and verifies the fact that more energy is delivered to the tank circuit by the electrons that strike the anode than is absorbed by the ones that strike the cathode. The four electrons which are used in the table are the ones that leave the cathode directly opposite an anode segment at times separated by a quarter of a cycle, as you can see in the second column of the table. The data in columns three and four are the results of calculations too complex to be included here. The figures in column five are based on the equation (10) which states that for each electron the energy from AC source is equal to the kinetic energy plus the energy to the tank circuit and that the available energy per electron is 10,000 electron-volts.

The efficiency of the cylindrical magnetron can be calculated from the results of the table. For example, if the four electrons in the table are considered, the efficiency of the magnetron would be the following:

$$\textit{Efficiency} = \frac{13{,}000}{30{,}000} = 43 \textit{ per cent.}$$

This rate of efficiency is fairly representative of 10-cm magnetrons employing plate voltages

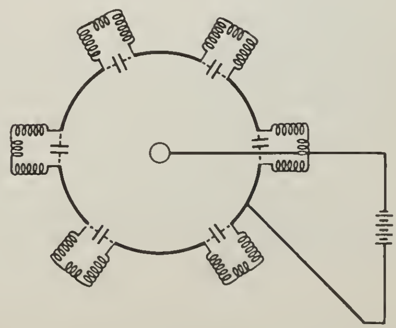
Electron No.	Opposite Anode Potential when Electron leaves	Energy Striking Cathode	Energy Striking Anode	Energy Contributed to AC	Energy Contributed by DC
1	+	+1,500	....	−1,500	0
2	0	......	+4,000	+6,000	+10,000
3	—	......	+2,900	+7,100	+10,000
4	0	......	+8,500	+1,500	+10,000
TOTALS		+1,500	+15,400	+13,000	+30,000

(Anode voltage =10 KV. All energies in electron volts)

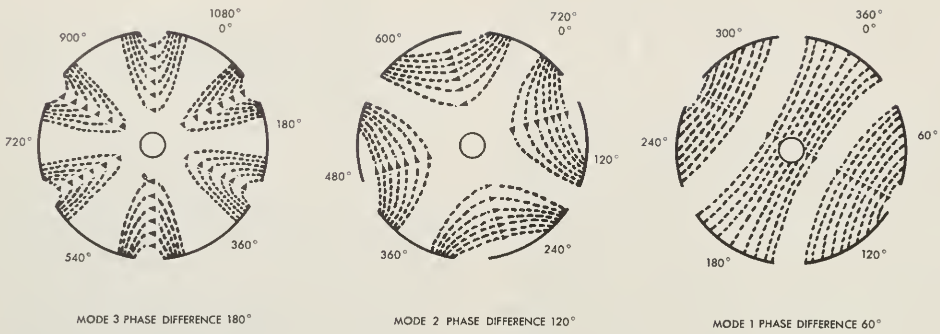
Typical Electron Energies in a Possible Magnetron

shown in the table. Practical magnetrons under these conditions have efficiencies which run between 30-40 per cent. Higher efficiencies, between 50-60 per cent, are obtained by using greater plate voltage, magnetic field and power.

Modes of oscillation of the magnetron. As shown, the simplified circuit of the cylindrical magnetron at A on page 11-20 may be split up to form six resonant circuits with magnetic coupling between them. While there is usually considerable coupling between inductances that are close together, in actual magnetrons which employ cavities as resonant circuits there is no coupling between adjacent cavities—that is, coupling in a manner characteristic of lumped circuit constants. On the other hand, there can be coupling resulting from radiation of energy from one cavity to another. As you recall from the theory of coupled circuits, when two circuits tuned to the same frequency are closely coupled, they will resonate not at that frequency but at two different fre-



Forming Six Magnetically Coupled Resonant Circuits



Configuration of AC field between Adjacent Anode Segments in the Various Modes of Oscillation

quencies. These six identical resonant circuits, all closely coupled, resonate, not at two, but at three different frequencies which are separated by relatively few cycles from each other. A point of interest is whether all these frequencies are associated with the same configuration of the AC field in the space between the cathode and anode. As it is they are not, but instead, for each frequency there is one configuration or *mode* of oscillation. For example, for three frequencies, there are three configurations as above.

In connection with the excitation of mode 3, which has been discussed, the average velocity, Va, which is equal to 10^8 E/B, must be so adjusted that an electron passes two segments per cycle. In mode 2 the ratio of (E/B) must be higher since the electron must pass three segments per cycle. It follows thus that this mode requires very high voltage and is difficult to excite strongly.

Another consideration is that the radar magnetron should oscillate only in one mode under operating conditions, since the frequencies of these three modes are generally different, and since the radar receiver can be tuned to only one frequency at a time. If the modes are widely separated in frequency, it will be easier to excite only the one desired. Since mode three requires the least voltage, it is the mode generally used. Sometimes special circuits are used to *strap* magnetrons to a single mode.

Summary

The following summarizes the preceding discussion of magnetrons:

1. By using a segmented anode, you can make the electrons work against an alternating electric field which is crosswise to the steady field. In this way it is possible to get the electrons to go all the way to the anode without much increase in kinetic energy. This results in high power and high efficiency.

2. Oscillations which are obtained when the average crosswise velocity Va are such that the electron passes two plates per cycle for mode 3. The equation for resonance for these oscillations is,

$$f = 10^8 \frac{E}{2BY}$$

In using this equation replace factor 2 by 3 and 6 respectively for modes two and one.

3. The frequency of looping which depends on B alone is not usually matched to the other frequencies, because doing this causes too low a power input and too low an efficiency.

4. Quantitative calculations show that there are more electrons which give energy to the tank circuit than electrons which remove energy from the tank circuit. Further, it shows that on the average each of the former electrons gives more energy to the tank circuit than each of the latter takes away.

5. Magnetrons in use at present have a wavelength range from about 50 cm down to less than one cm.

6. Efficiencies of magnetrons at moderate power in service at present have values from 30 to 40 percent, and from 50 to 60 per cent at powers greater than 200 kw. (peak).

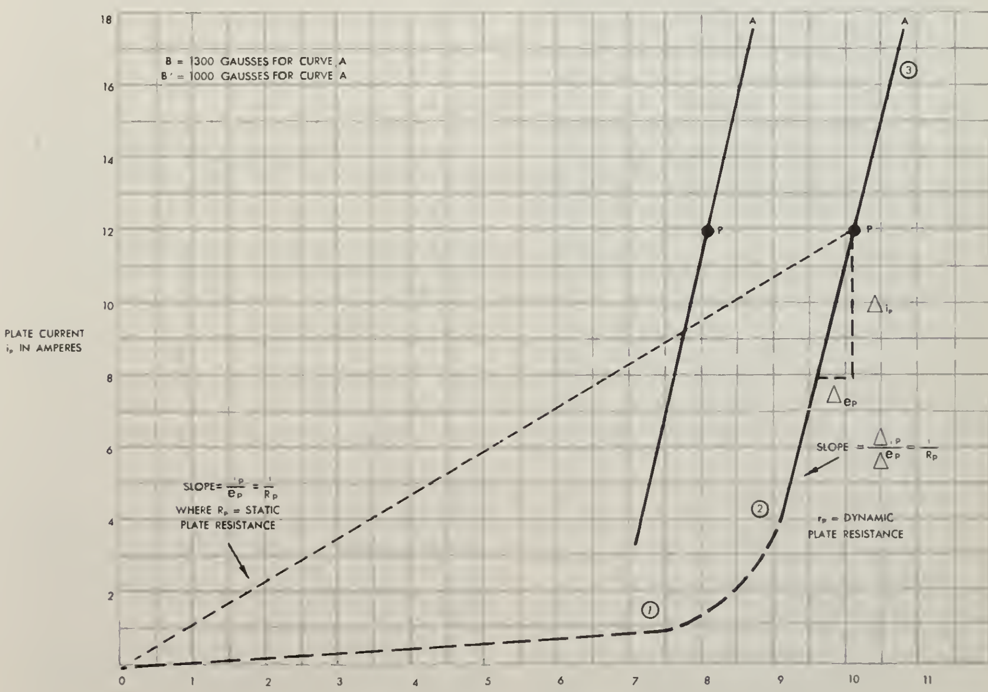
7. The actual tuned circuit will resonate at several different frequencies each of which corresponds to a distinct phase difference between anode segments.

Plate Current-Plate Voltage Curves of a Magnetron

A useful curve for studying magnetrons is the e_p-i_p characteristic curve. The curve below is a typical curve for a 10 cm magnetron operating at about 1300 gausses. Ordinarily, these curves are plotted by holding the magnetron field constant.

These curves are important in that you can get much information by properly interpreting them. Keep in mind the following factors about these curves. The resistance between the cathode and anode is non-linear. The AC and DC resistances are different from each other at any one voltage and vary with the voltage. For example, at point P the static resistance r_p is the ratio of e_p/i_p at the point, while the ratio of $\Delta e_p/\Delta i_p$ is the dynamic resistance. Since the magnetron operates on the straight line portion of the curve, the reciprocal of the slope of this straight portion, ② to ③, is the dynamic resistance r_p.

Note that from the origin to point ① on curve A that the magnetron current is very small. There a negligible coherent RF energy is produced. Also a considerable amount of random RF noise is produced. Somewhere between 6 and 8 KV appreciable oscillations start in a service tube. In the region from point ① to point ②, the operation of the magnetron varies greatly, not only from one magnetron to another, but also from one time to another in the same magnetron. Oscillations of more than one wavelength can be detected in the output, indicating that the magnetron is oscillating in more than one mode. The value of (E/B),



Variation of Plate Current with Plate Voltage for Constant Flux Density of 1000 Gauss (Curve A') and for 1300 Gauss (Curve A) which excites oscillations of higher modes, is probably a submultiple of that which fulfills the resonance condition for these modes, and at the same time is not very far from the resonance condition for the 180° mode.

In the straight-line portion, between points ② and ③, the magnetron oscillates in the stable mode 3. In this region the frequency varies only a few megacycles over this range. In the strapped 10 cm service magnetrons the straight portion continues to the highest current obtainable. Theoretically, if the range be made great enough, mode 2 might be expected to appear. However, this mode is rarely, if ever, encountered in strapped tubes. Unstrapped magnetrons, however, have an additional multi-moding region near or about point 3.

Even though the percentage variation in frequency is small over the straight line portion, the actual change may be greater than the receiver bandwidth. For this reason it is important to maintain the pulse amplitude constant.

The power input is $e_p i_p$. Since the straight line portion is fairly steep, the input power rises rapidly with increase in voltage.

The magnetron has a very low efficiency in the region 0-①-②. In the single-mode region ②-③, the efficiency is maximum somewhere in the middle of the straight line portion and decreases slowly toward point ③. In this region you may safely assume that the efficiency is constant and equal to 30-40% over the entire region. It may seem surprising in view of the resonance condition specified earlier that a single-mode oscillation is excited over such a wide range of voltage (region ②-③). This is explained by the fact that the point of resonance is not a sharp one, and the point of resonance on the characteristic curve is not the only point of which the oscillation is excited, but as the point where the *efficiency* of operation is a maximum.

The power output increases approximately as the power input in region ②-③ since the efficiency is practically constant within this region.

Curve A′ shows the effect of reducing the magnetic flux density to 1000 gausses. To a close approximation, the new operating point P′ will occur at a voltage 25% less than that of P. The current will be the same. As is evident, the dynamic resistance will be essentially unchanged. If B were increased 25%, the curve would be shifted to the right of curve A by about the same amount. The general properties of i_p-e_p curves are not altered except that both the input power and efficiency increases in the linear portion as B is increased. The increase in power input is the result of the higher voltage necessary to keep (E B) constant. The increase in efficiency is accounted for by the fact that the kinetic energy with which the electrons strike the anode remains roughly constant (from equation (5)) while the total energy available increases with the voltage. Hence, a larger fraction of the energy goes into the oscillation, as indicated by equation (10). Efficiencies of 50% to 60% are obtained by increasing B to 2500 gausses and plate voltage to 25 KV.

The following table summarizes the characteristics of the magnetron:

Region	Mode	Power Input	Power Output	Efficiency
0-①	Noise	Very Small	Zero	Zero
①-②	More than one	Small	Very Small	Very Small
②-③	Mode 3	Increases rapidly	Increases rapidly	High and fairly constant (40%)
Beyond ③ Unstrapped	Double-mode	Large	Moderate	Low
Beyond ③ Strapped	Continues same mode	Increasing	Large	Moderate and decreasing

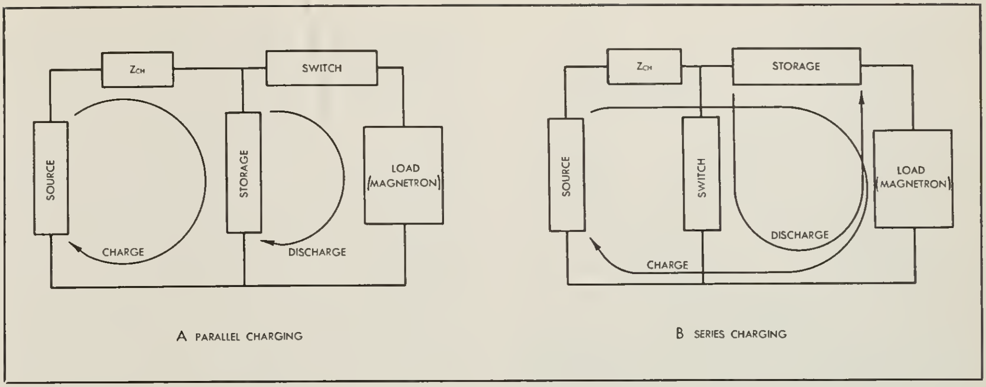
Summary of Magnetron Characteristics

Modulating the Magnetron Transmitter

The modulator for the magnetron serves a double purpose—it determines the waveform of the output pulse, and it stores energy between pulses and releases it through the magnetron during the pulse.

There are two circuits which make it possible for the modulator to store and release energy. These are the parallel charging circuit and the series charging circuit.

As you can see, in the parallel charging circuit the source of energy is connected at all times to the storage element through the charging im-



Parallel and Series Circuits for Storage and Release of Energy

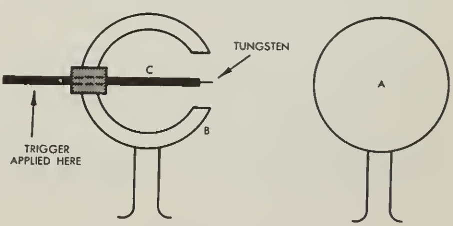
pedance Z_{ch}. During the output pulse the closed switch applies the voltage from the storage element to the load.

A disadvantage of this circuit is that it is not possible to prevent grounding all three of the circuit elements—the source, switch, and load.

Because of the disadvantages of the parallel charging circuit, the series charging circuit is most generally employed. In this circuit the storage element is charged through the charging impedance and the load. It is discharged through the load when the switch is closed. On examining this circuit, you can see that closing the switch also completes a circuit containing the source and Z_{ch}. This has little consequence, however, since the switch is closed for only a microsecond or so at a time and Z_{ch} is large enough to prevent any damage.

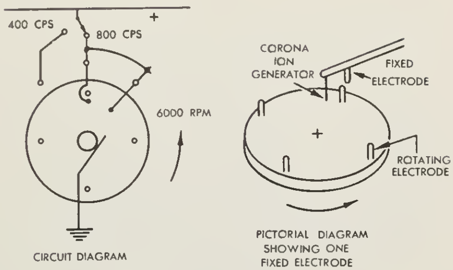
It is well to find out the components of some of the modulator circuit elements which are shown as blocks in the above circuits. In the modulator, the source from which the storage element is charged may be either a DC voltage or an AC voltage. The statement was made earlier in this chapter that the magnetron requires a voltage of 10 to 15 kilovolts or higher for proper operation. In some types of equipment a transformer is used between the storage element and the magnetron to step up the voltage while in many others a direct connection is made between the storage element and the magnetron. Actual charging circuits are discussed in a later section dealing with line pulsing modulators. In these circuits the charging impedance may be a high resistance or a high inductance or both. Its purpose is to control the charging time for the storage element and to prevent short circuiting of the source during the pulse.

Switches for controlling the charging time for storage may be of two general types—a soft switch or a hard tube. A soft switch is one in which conduction takes place through a gas. Soft switches include gas thyratron, mercury thyratron, fixed trigger gap, and rotary spark gap switches. A hard tube switch is a high vacuum tube switch. Earlier radar equipment employed only hard tube and thyratron switches, while more recent sets largely employ spark gap switches.



Fixed Trigger Gap

In the fixed trigger gap switch break-down occurs at the desired instant when a trigger pulse is applied to electrode C at that instant. This in turn causes break-down from C to B resulting in the formation of a sufficient amount of ions which cause break-down between the high voltage electrodes A and B. One difficulty with this switch is that the trigger wire C corrodes and is consumed in too short a time to make it general in use.

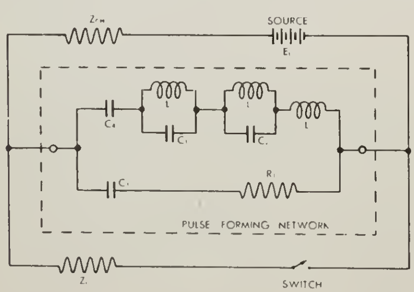


Rotary Spark Gap

In the rotary type of spark gap shown above, there are four electrodes on the rotating wheel, which in turn is grounded. The electrodes pass between two pairs of fixed electrodes when the switch is in the 800 cps position and between one pair when it is in the 400 cps position. Each time one of the moving electrodes passes a pair of fixed electrodes, there is a breakdown of voltage and the switch closes for a brief period of time. In operation there is some irregularity in the intervals between break-down which may cause the period to vary as much as 50 microseconds but in a self synchronous radar set it presents no difficulty. In fact, there is some advantage to the irregularity since it causes pulses from other radar sets to move on the scope and also lessens the likelihood of their being confused by echoes occuring on the second trace after the transmitted pulse.

As each type of switch mentioned has certain disadvantages. None is perfect for the job. Therefore, you may expect to find any or all of them in use in various sets.

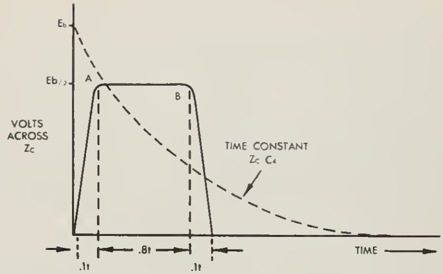
The storage element can be either a condenser or a pulse forming network. A condenser is most often used with the hard tube modulator in which the sole purpose is to store energy. In this type of modulator circuit, it is necessary that the time constant be such that the condenser does not discharge appreciably during the pulse for as you can see in the graph on page 11-23 the magnetron current and in turn the magnetron output, drops very rapidly when the voltage decreases. The pulse forming network in the storage element serves the double purpose of storing energy and of shaping the output waveform. In operation it is quite similar to the artificial transmission line, that is, it furnishes a steady output voltage for a certain length of time at which time its voltage drops to zero. It is different, however, in that instead of being made up of a number of identical sections, each section is different. It has the advantage of forming more nearly rectangular pulses than an artificial transmission line with a similar number of sections.



Pulse Forming Network as Storage Element

Although the theory of operation of the pulse line shown to the left during discharge is somewhat complex, it may be summed up as follows: C_4 stores most of the energy for the pulse while the components C_1, C_2, L_1, L_2, and L_3, which are in series with it, mainly shape the pulse. During the first part of the pulse (when the charge on C_4 is nearly equal to the supply voltage E_b) the series components offer a high impedance and cause the voltage across Z_c (the load) to jump to $E_b/2$. As the pulse continues, C_4 discharges but the impedance of the series components drops also. This causes the voltage across Z_c to remain at $E_b/2$. Later when the charge on C_4 is less than $E_b/2$ the inductances maintain the level of current flow. This continues until the condenser is discharged and the magnetic fields are collapsed. At this time the voltage drops rather rapidly to zero. The components C_3 and R_1 make the leading edge of the pulse steeper. There being no inductance in this branch of the network, the current can rise immediately to its full value. The pulse which appears at the load is essentially a rectangular pulse of the desired duration and equal in amplitude to $E_b/2$. The values of the components determine the length of the pulse and the characteristic impedance of the line. The characteristic impedance of the line must be matched to that of the load. This accounts for the fact that even though the line is charged to E_b, the voltage across Z_c is only $E_b/2$.

Study the illustration to the right to get a comparison of the output pulse and the discharge curve of C_4. The time of rise of the voltage is about one-tenth of the pulse duration and the time of fall a like amount of time. This type of pulse network is usually sealed in a metal box filled with oil which acts as an insulator. For that reason they are sometimes called potted networks. It is impossible to replace components in the line. Some lines using only a few elements are not potted. These do not generally give as nearly a rectangular output pulse, due to the smaller number of sections.

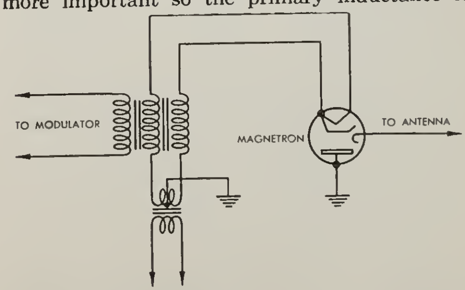


Comparison of Pulse from Pulse Line with Exponential Discharge of Condenser

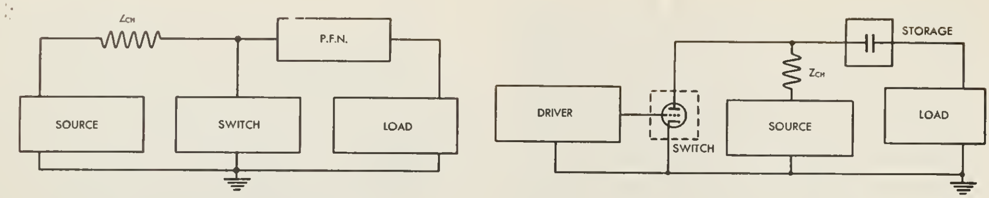
In the circuits discussed thus far the load can be either a magnetron and its associated circuits or the primary of a pulse transformer which has its secondary connected to the magnetron. In addition a pulse transformer can be used in the amplifier stages preceeding the hard tube modulator. When this is done all tubes are cut off between pulses. The interstage transformer is a low-voltage and low-turns-ratio device. Its only purpose is polarity inversion. The use of a high voltage step up transformer to furnish voltage for the magnetron makes possible the use of lower voltages in the pulse forming network and thus simplifies the insulating problem. Another advantage in using a transformer with two secondaries as shown below is that it avoids having high voltage in the magnetron filament transformer. A third need for the high voltage pulse transformer is matching impedances, particularly where the radar antenna is remote from the set itself. Since the use of long transmission lines between the magnetron and the antenna make the magnetron's operation somewhat erratic, it is desirable to have the magnetron near the antenna. This necessitates the use of long lines to carry the modulator pulse to the magnetron. In order to be able to use 50 ohm line for this purpose and still get a match of impedance with the approximately 1000 ohms of the magnetron, it is necessary to use a 4½: 1 step up transformer at the magnetron end.

The pulse transformer must be specially designed because there are very high frequency components present in a rectangular pulse of such short duration. The core is usually silicon steel and has laminations in the order of 0.003 inches in thickness. In general a good pulse transformer must have low leakage inductance, low interwinding capacitance, and high primary inductance.

Low leakage inductance is necessary in transformers to preserve the steep leading edge of the pulse. This is achieved by very close coupling between primary and secondary and by using fewer turns on each. Low interwinding capacitance causes high frequency oscillations between it and the leakage inductance and makes it easier to damp them out. Since low capacitance between windings can be achieved by separating the windings farther, you see that a compromise is necessary. In most cases the spacing is rather close and the number of turns low, the primary and secondary consisting of a single layer each, wound one on top of the other on the same leg of the core. To damp oscillations a resistor is shunted across the primary or secondary or both. The high primary inductance is desirable to induce a high voltage into the secondary and to insure that there is little change in the secondary voltage if the primary current remains constant. High primary inductance requires many turns, which is in conflict with the requirement for low leakage inductance. Usually the latter is considered more important so the primary inductance is



A transformer with two secondaries makes possible the grounding of the filament transformer secondary



LINE PULSING MODULATOR

DRIVER HARD TUBE MODULATOR

Two Types of Modulators

not as high as would be desirable. For this reason the secondary voltage does drop some during the pulse. The maximum allowable drop is about 5%.

Types of Modulators

There are two types of modulators—the line pulsing modulator and the driver—hard tube modulator. The line pulsing modulator stores energy and forms pulses in the same circuit element. This element is usually the pulse forming network. The driver-hard tube modulator forms the pulse in the driver and stores the energy in the modulator circuit.

Before discussing the line pulsing modulator, it is well to consider briefly and to note the advantages and limitations of the various types of charging circuits. To do this, study the illustration on the next page showing the various types of charging circuits.

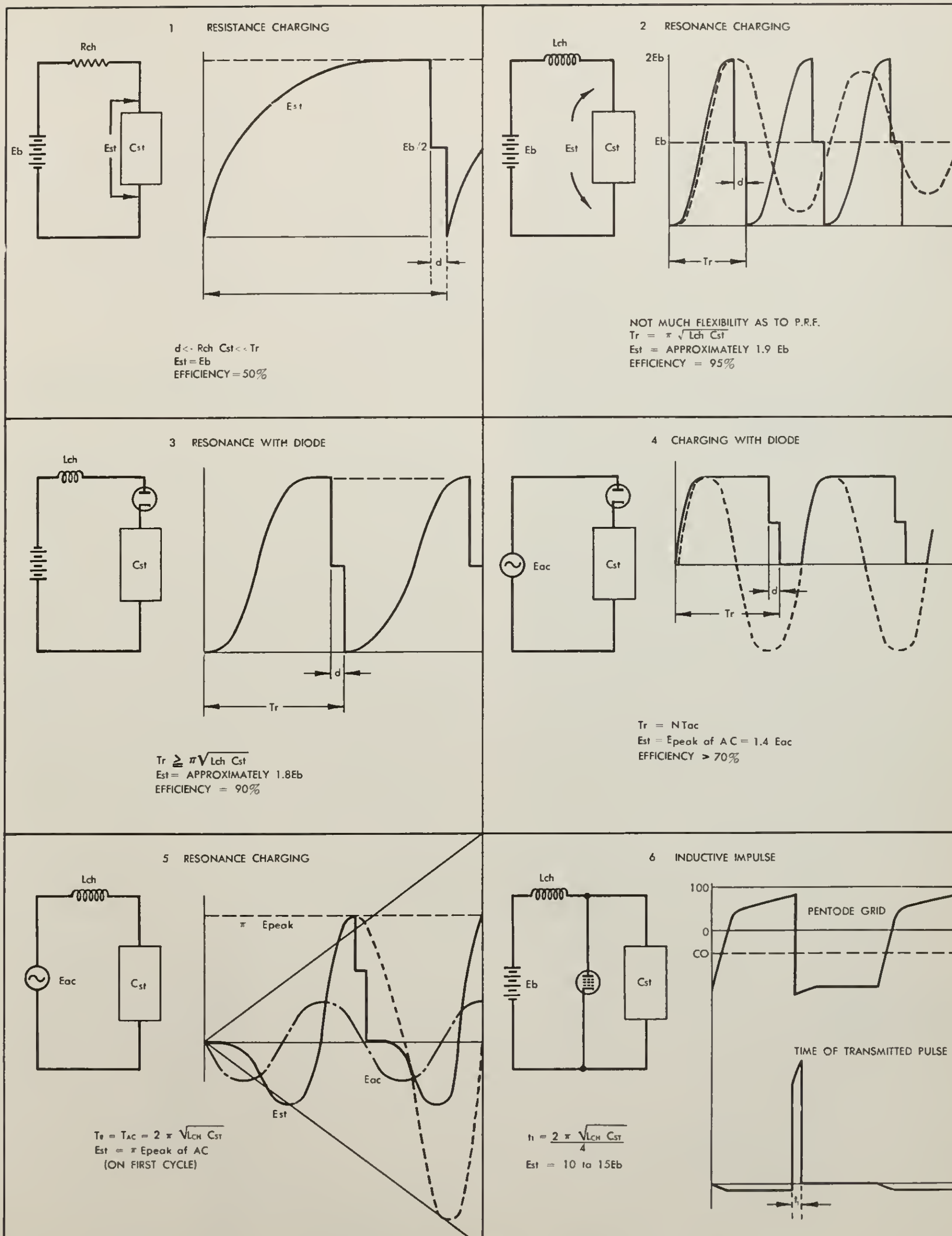
In the DC resistance charging circuit the pulse line is represented as C_{st}. It is charged through R_{ch} to the value of the DC voltage. During the pulse one half the DC voltage is applied to the magnetron (assuming the line impedance is matched to that of the magnetron) for a period of time determined by the line components. This is only 50% efficient. Therefore, it is necessary that the DC voltage be twice the value needed for the magnetron. The time constant $R_{ch}C_{st}$ must be large in comparison to the pulse length *d* but small in comparison to the period of the repetition frequency. This insures that the line will fully charge between pulses but will not discharge appreciably through the source and E_{ch} during the pulse.

In the DC resonance charging circuit resistor R_{ch} is replaced by an inductance L_{ch}. As C_{st} charges the current through L_{ch} builds up a magnetic field. This field causes current to continue after C_{st} is charged to E_b. The charge will reach a peak of approximately 1.9 E_b after which it starts decreasing. If the modulator is triggered at the instant of maximum charge, the voltage applied to the magnetron is about 95% of E_b. This necessitates that the resonant frequency of L_{ch} and C_{st} be one-half the PRF of the radar. The firing time is very critical if you are to take advantage of the maximum voltage.

The difficulty encountered in the preceding circuit is overcome by adding a diode in series with the charging element. Circuit 3 prevents the line from discharging after reaching its peak. This means that the time of firing can occur any time after the peak is reached. Therefore, the frequency at which L_{ch} and C_{st} resonate is equal to or greater than one-half the PRF. The disadvantage of making the firing time less critical is that the voltage to which the pulse forming line charges is decreased by the drop across the diode to about 1.8 E_b. This makes the efficiency of this circuit about 90%.

Circuit 4 has the pulse line charging from an AC source. In this circuit, the charging line is prevented from discharging by a diode in series with it. Further, in this circuit, it is necessary that every trigger pulse come at a time when the diode plate is negative; otherwise, the diode and the switch will short circuit the AC source. Therefore, $T_r = N\ T_{ac}$. The voltage to which the line charges is almost the peak value of the AC wave so that the efficiency is about 70%.

In circuit 5, the diode is replaced by the inductance L_{ch}. This inductance, along with C_{st}, forms a resonant circuit which is resonant at the frequency of the AC supply voltage. In this circuit the switch may be closed at any peak of oscillations. Usually, though, it is closed at the first peak. At this time the voltage is



Types of Charging Circuits

approximately π times the peak value of the AC voltage. In addition $T_r = T_{ac} = 2\pi\sqrt{L_{ch}C_{st}}$. If the switch is closed at the second peak, the voltage will be nearly twice as high. This type of charging gives voltages higher than the source voltage.

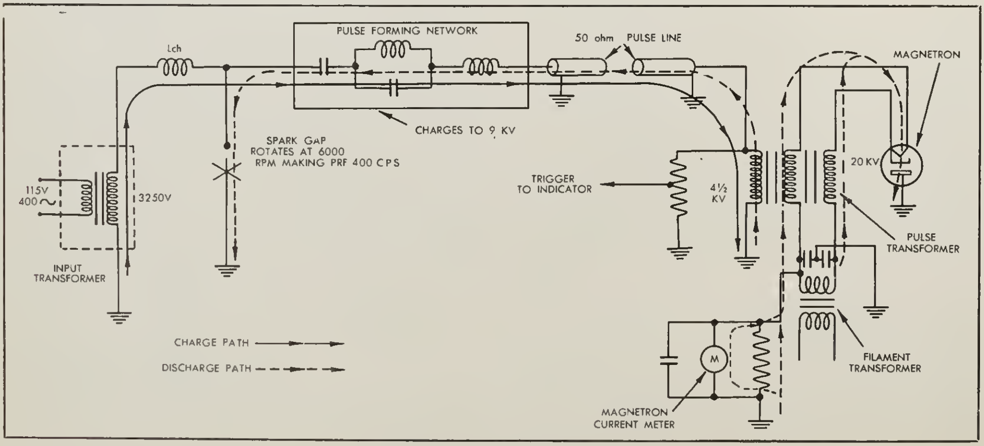
The inductive impulse charging circuits provides voltages 10 to 15 times as great as the supply voltage. It is used where the available voltage is of the order of 1 KV and operates as follows: The pentode grid is raised above cut off and into the positive voltage region so that the current through the inductance builds up a magnetic field around it. At a time t_1 (the time prior to the beginning of the transmitter pulse) the pentode is suddenly cut off. The magnetic field in collapsing sends current through the pulse line and charges it to a very high voltage. To take full advantage of this high voltage the transmitter must be keyed at the instant of maximum voltage. The correct timing for $t_1 = \frac{1}{4}(2\pi\sqrt{L_{ch}C_{st}})$. The value of t_1 is usually 10 to 15 microseconds. This type charging circuit is used only in conjunction with a master timer.

Line pulsing modulator. The typical line pulsing modulator circuit, shown below, employs an AC resonance type charging circuit. It also employs a 115 V AC input which is stepped up to 3250 V by an input transformer. A resonance charging system charges the line to approximately 2.8 times the 3250 volts or about 9 KV. The switch is a rotary spark gap which fixes the PRF at 400 cps. The pulse forming line is connected through a 50 ohm pulse line to the primary of a pulse transformer. The voltage across the primary is approximately $4\frac{1}{2}$ KV. This is stepped up to about 20 KV. by the transformer.

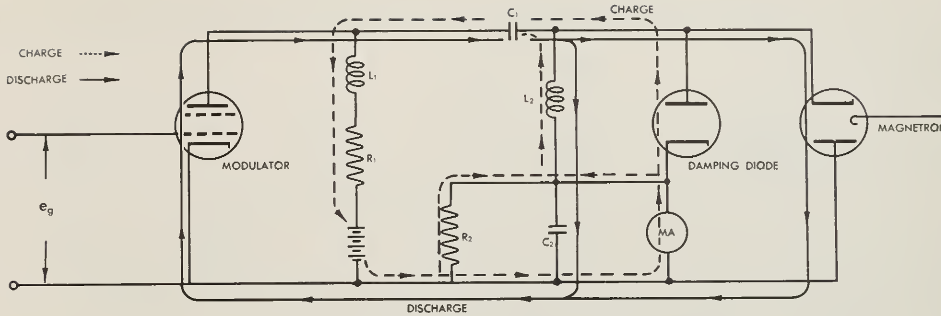
The driver-hard tube modulator. The driver-hard tube modulator, as the name indicates, consists of two parts—the driver and the hard tube modulator. The hard tube modulator is the part which stores energy between pulses to be released when the hard tube is made conducting during the pulse. The driver is the part that shapes the pulse applied to the grid of the modulator.

In the earlier days of radar the hard tube modulator was the most commonly used circuit because at that time there was no good switching arrangement available for the line pulsing modulator.

A simplified circuit of a hard tube modulator employing DC resonance with diode charging is shown on the next page. This is how it operates. Between pulses from the driver the modulator tube is non-conducting due to a very high negative voltage on its grid. The storage condenser charges through the path indicated. This path includes R_1, L and R_2 and the milliammeter in parallel. As R_2 is about 1000 ohms, most of the current flows through the meter. The only purpose of R_2 is to switch the meter out of the circuit. This permits it to measure other currents without leaving the charge path open. As the meter reads the charging current, it will indicate the mag-



Typical Line Pulsing Modulator



Hard Tube Modulator with Diode Charging

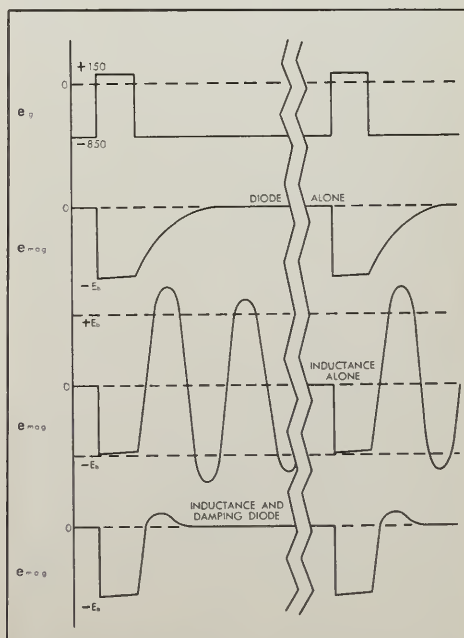
netron current since the charge that is lost during the pulse is replaced between pulses.

When the driver raises the grid potential of the modulator, it starts conducting and the condenser discharges through the path containing the magnetron and the modulator tube. The charging diode prevents any discharge through that branch. A negligible amount of current will flow through the source, R_1 and L. When the driver lowers the modulator grid voltage below cut off, the pulse ends.

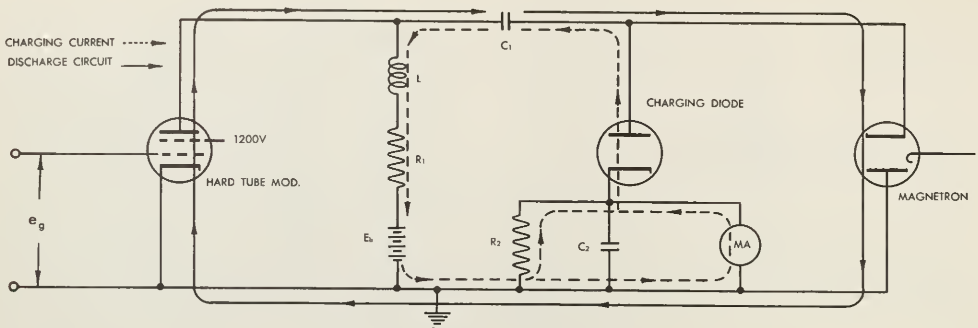
The amount of voltage which the modulator applies to the magnetron is determined by a voltage divider circuit consisting of the static plate resistances of the magnetron and the modulator. These resistances have values of approximately 1000 ohms and 100 ohms respectively.

The voltage which is applied to the magnetron is roughly 90% of the condenser voltage. It is not a true rectangular pulse, however, because of the distributed capacitances in the leads to the two condenser plates and those in the magnetron. The total distributed capacitance, which equals 50 to 120 micromicrofarads, must be charged as the voltage rises (in a negative sense) and discharged as the voltage decreases. The time constant for charging the distributed capacitance, which is determined mainly by the static plate resistance of the modulator tube, is in the order of 0.01 microsecond, a time which is small enough to ignore.

Discharge of the distributed capacitance takes place through R_1, L, and the magnetron itself. The calculation of the discharge time constant is rather complicated but you can estimate it by ignoring the effect of the inductance and figuring the parallel combination of R_1 and plate resistance of the magnetron as the voltage decreases. For example, an average discharge time constant is about 0.5 microseconds as R_1 and r_p of the magnetron are relatively high. This discharge time causes the voltage to trail off rather slowly as shown below and in turn produces considerable noise, and erratic oscillations for one or two microseconds as the voltage passes through the region below 8 KV. All this makes it impossible to detect a nearby target.



Voltage Waveshapes in Hard Tube Modulator Circuit



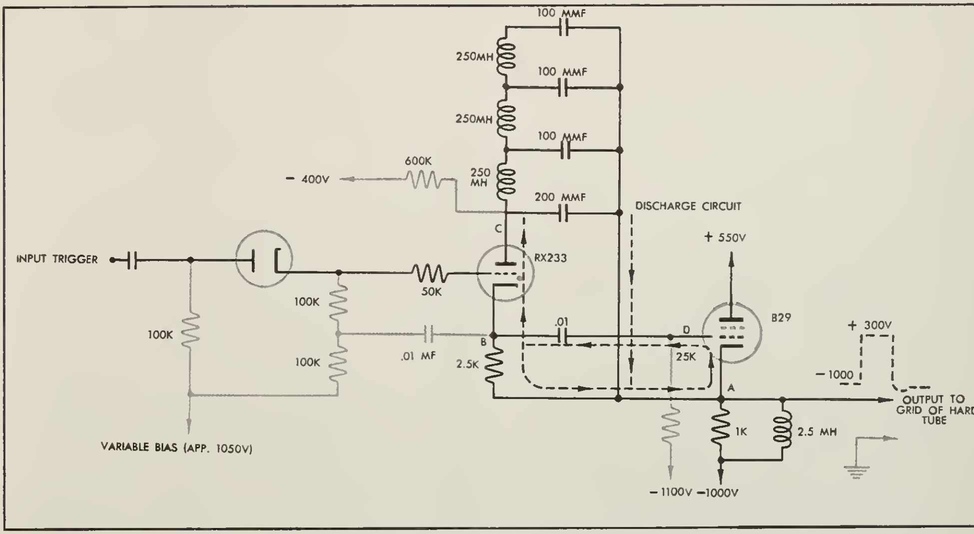
Hard Tube Modulator with Tail Clipping

The difficulty is overcome by using a tail-clipping circuit. This circuit, which consists of an inductance and a diode connected in parallel, replaces the charging diode. In the modulator tail-clipping circuit above the inductance (L_2) is small, being 3 to 5 mh. In the charge and discharge paths, most of the charging current flows through L_2. This means that you could leave the diode out of the circuit as far as charging is concerned except for a very important function. The discharge path includes L_2 in parallel with the magnetron. At the beginning of the pulse little current flows through the inductance, but, nevertheless, it builds up because the inductance is small so that at the end of the pulse the inductance discharges through the distributed capacitance. This causes the voltage to return more rapidly to zero. Another point is that L_2 and the distributed capacitance can be thought of as a resonant circuit. With a small value of L_2 the resonant frequency is such that the period is only two or three microseconds. The oscillations of this tuned circuit would cause the voltage to swing positive and negative if it were not for the damping diode, which conducts when its plate swings positive and ends the oscillations. For the waveshapes with and without the damping diode, see the bottom illustration on page 11-31. The fact that the current through the inductance has to build up during the pulse means that there will be more change in magnetron voltage than with diode charging alone. The larger the tail clipping inductance, the less the magnetron voltage will change. Already you have seen that in order to return the magnetron voltage to zero quickly at the end of the pulse, you need a small inductance. It is evident that the inductance cannot be both large during the pulse and small at the end of the pulse, so the value that is used is a compromise between the two.

Types of driver stages. The two principal types of driver stages which are employed for shaping the pulse applied to the hard-tube modulator are the boot-strap driver and the blocking oscillator.

The boot-strap circuit was widely used in earlier radar sets because it did not require a pulse transformer, a device which was not perfected at that time. The boot-strap driver circuit forms the pulse by discharging an artificial transmission line through a gas tube. Since the voltages of the pulse required to trigger the hard-tube modulator exceed the voltages that can be present in a gas tube, an amplifier is added between the gas tube and the modulator. To permit "off" tube operation of all three stages, cathode coupling is used. This coupling is unique in that it accomplishes amplification of the pulse without polarity inversion.

In the basic circuit diagram of the boot-strap driver in the Raytheon service modulator on the next page, the circuits in heavy lines are those along which pulse formation and amplification progress. One side of the pulse forming line is connected to −400 v. and the other to −1000 v. This line charges to 600 v. when it is allowed sufficient time. In this modulator, the pulses are only 500 microseconds apart and the time constant of the charge circuit is about 300 microseconds. The pulse forming line only has time to charge to approximately 500 volts. The discharge circuit which the line uses when the gas tube is made to conduct by the trigger pulse is through



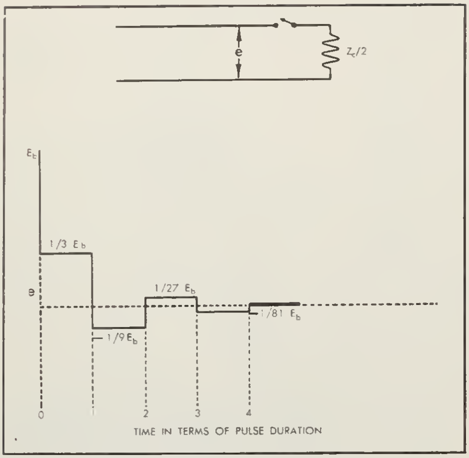
The Boot-strap Driver for Hard Tube Modulator

the parallel combination of the cathode to grid resistance of the 829 tube and the 2.5 K cathode resistor of the RX233 in series with the RX233 tube. The resistance of the RX233 is negligible and of the parallel combination is approximately $Z_c/2$ where Z_c is the characteristic impedance of the line. (In some circuits the parallel combination has a value of Z_c.) With the line discharging through $Z_c/2$, one-third of the voltage to which the line is charged appears across the 2.5 K and grid to cathode resistance of the 829 for a period of time determined by the components of the pulse line. This 167 volt pulse being applied between grid and cathode of the 829 causes it to conduct heavily since its bias is 100 volts (cathode is at −1000 v. and grid at −1100 v.). When the 829 is conducting heavily the drop across the tube is only about 250 volts so the cathode rises to about +300 volts. This represents a change of about 1300 volts since in the non-conducting period the cathode is at −1000 volts. The purpose of the 2.5 mh. inductance in parallel with the 1 K cathode resistor of the 829 tube is to produce the tail at the end of the pulse. The tail results from the collapsing of the magnetic field about the pulse. When the field collapses, current goes through the 1 K resistor in the downward direction and causes point A to be more negative than −1000 v. for a brief time. This action insures a rapid cut off of the hard tube. The discharge of a transmission line through an impedance of $Z_c/2$ is oscillatory as shown on page 11-34. On examining this diagram you can see that at the end of the time for a pulse, the voltage across the line swings to $-E_b/9$. (In this circuit this equals −500/9 or −55 volts.) Thus the gas tube and the 829 are cut off very quickly. The purpose of the diode in the input circuit is to isolate the trigger circuit from any transients.

The following list sums up the action at points A, B, C, and D before and during each pulse.

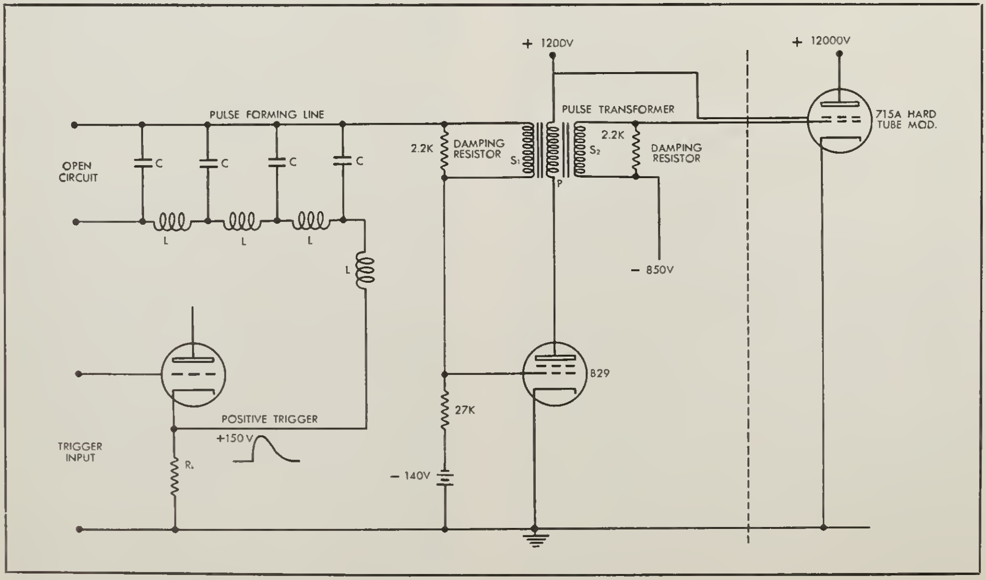
	A	*B*	*C*	*D*
Before	*−1000 v.*	*−1000 v.*	*−500 v.*	*−1100 v.*
During	*+300 v.*	*+467 v. app.*	*+467 v.*	*app. +367 v.*

In a line-controlled blocking-oscillator driver circuit, the pulse forming line controls the duration and the waveshape of each pulse. Normally the oscillator is cut off by the 140 volts bias. During the interval between pulses the line is charged to 140 volts since it is connected to the bias supply through the 27 K resistor, the transformer secondary, S_1, and R_K (the cathode resistor of the trigger tube). The trigger pulse, which is about 150 volts in amplitude, is sufficient to raise the grid of the blocking oscillator well above cut-off. It is coupled to the grid through the capacitance of the pulse forming line.

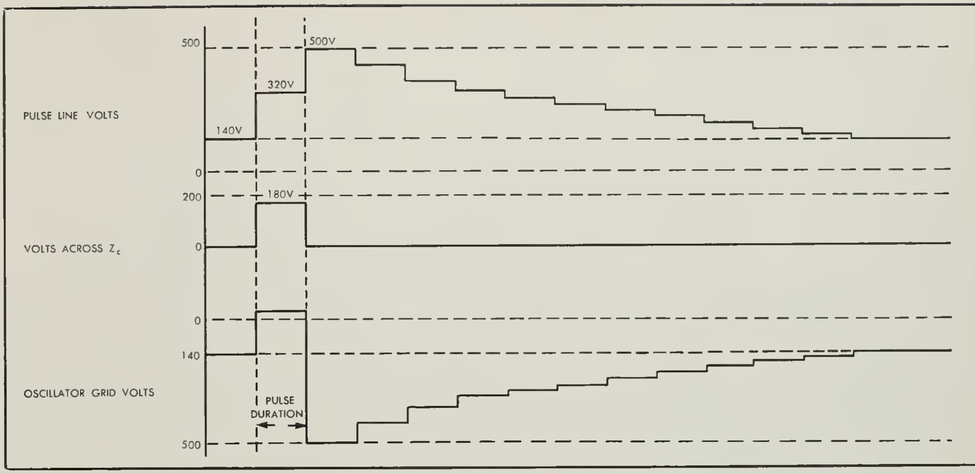


Discharge of Pulse Line into Zc/2

As in an ordinary blocking oscillator, the plate voltage drops when the tube begins to conduct. This drop is coupled back to the grid in opposite phase due to the transformer action and causes the grid to go more positive. This action is rapid enough that it quickly drives the tube to saturation. In this condition the voltage across the primary P of the transformer is about 1000 volts and that across the secondary S_1 is about 500 volts. Since the pulse line is charged to 140 volts and since it is in series with the secondary voltage of 500 volts, it will charge to the higher voltage through the grid to cathode resistance of the oscillator, the secondary S_1, and R_K. Further, the impedance of the circuit is matched to the characteristic impedance of the line. Thus the voltage across the line goes from 140 to 500 volts in two steps as shown on the next page. The length of the step at 320 volts is the time it takes the voltage impulse to travel the length of the line, be reflected and return to the sending end. At that time the line is fully charged and current ceases to flow. The current flow through the grid to cathode resistance is practically constant during this time and keeps the grid to cathode potential steady insuring little change in the primary voltage. At the time the line becomes charged, current flow ceases through the charging circuit and causes a drop in the oscillator grid voltage. This decreases the plate current and causes a rise in plate voltage. Through the transformer this drop appears as a drop in voltage at the grid and still further reduces conduction. The action is cumulative,



Line Controlled Blocking Oscillator Driver



Blocking Oscillator Voltages

eventually resulting in the tube being cut off quickly. In this condition the line is charged to 500 volts and as it has only 140 volts applied, it starts discharging. It discharges through the 27 K resistor, S_1 and R_K. As this path has a much higher impedance than the Z_c of the line, the discharge appears as a series of steps which follows the general exponential curve. The output voltage which is applied to the grid of the hard tube modulator is taken from another secondary winding S_2 and is a 1000 volt pulse. Damping resistors shunt both secondaries.

VELOCITY-MODULATED TUBES AS LOCAL OSCILLATORS

In radar, most receivers use 30 or 60 mc IF frequencies. In receiver operation, a highly important factor is the stability of the frequency of the local oscillator which generates the frequency that differs from the transmitted frequency by the IF. For example, in a radar receiver which receives a frequency of 3000 mc a frequency shift in the local oscillator as much as 0.1 per cent would be a 3 mc frequency shift. This is equal to the bandwidth in most receivers and would cause a considerable gain loss in the receiver.

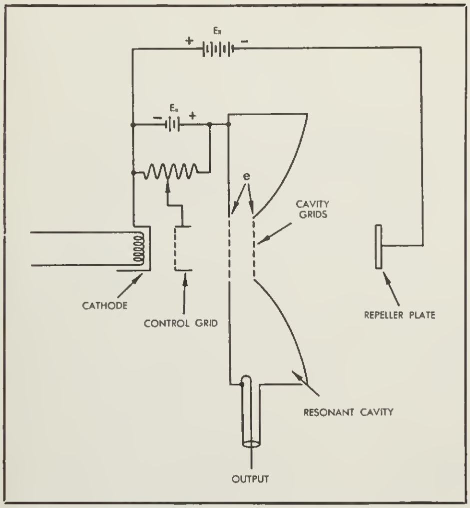
Still another consideration in radar is a receiver which uses a crystal mixer. In this receiver the power required of the local oscillator is small, being only 20 to 50 mw in the 3000 mc region. Due to the very loose coupling, only about one milliwatt actually reaches the crystal.

A third requirement of a local oscillator is that it must be tunable over a range of several megacycles. This is to compensate for changes in the transmitted frequencies and in its own frequency. In cases where the RF units are located in a remote position, it is desirable to mount the converter section of the receiver remotely. This insures that the IF frequency rather than the transmitted frequency is cabled back to the receiver, and makes it necessary that the local oscillator be tuned from a position some distance away, preferably by varying the voltage applied to it.

Because the reflex-velocity-modulated tube meets these three requirements better than any other type tube now in production, it is used almost exclusively for local oscillators in microwave radar receivers. The following deals with its operation. In it the cavity resonators are treated as parallel resonant circuits, as was explained in the theory of the resonant cavity in the preceding chapter.

Theory of Operation

In the circuit of the reflex-velocity-modulated tube, on the next page, note the arrangement of electrodes and the voltages involved for operation. Electrons are emitted by an indirectly heated cathode. These electrons are attracted by the



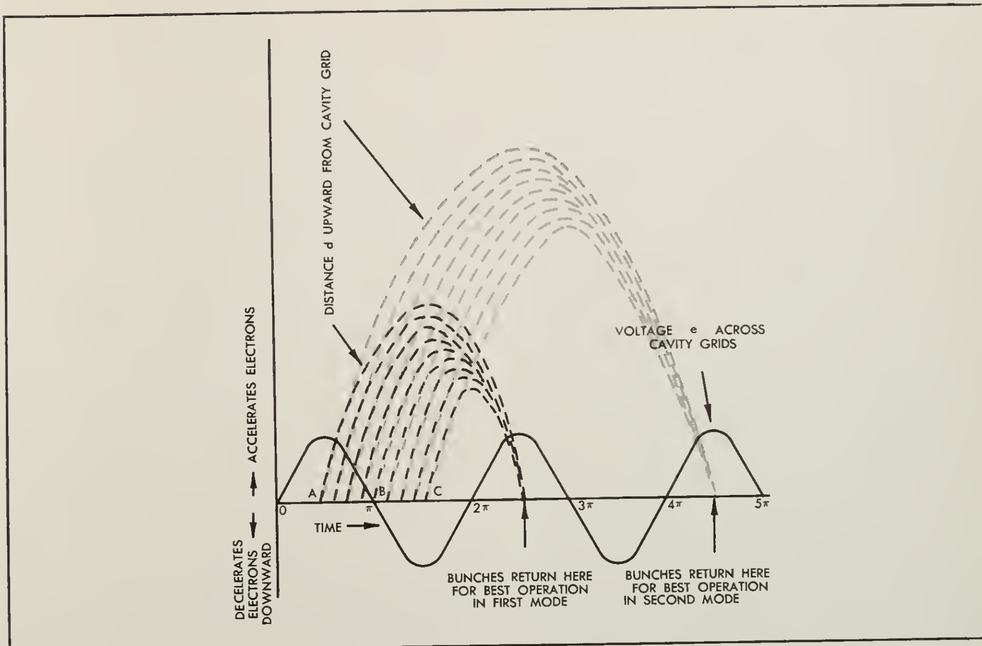
Circuit of Reflex-Velocity Modulated Tube Oscillator

cavity grids which are more positive than the cathode by the voltage E_a. The control grid is located between the cathode and the cavity grids and has as its purpose the control of electron flow. It also has a *positive* potential which is usually about two or three hundred volts. The electrons emitted from the cathode travel toward the cavity grids at a velocity determined mainly by E_a. Most of the electrons pass through the control grid, the cavity grids, and continue on toward the repeller plate. After passing the cavity grids they come to a region where the electrical field opposes their motion since the repeller plate is negative with respect to the cathode by the voltage E_r. This voltage is variable and equals about −100 volts. This in turn makes the voltage from repeller to cavity grids three or four hundred volts and slows down the electrons causing them to come to a stop, after which they reverse direction and pass back through the grids. They are collected either by the control grid, the shell or the cathode of the tube.

An important consideration is the cause of oscillations in resonant cavities. In most oscillators, oscillations start from some irregularity in current flow, such as a transient that results from voltage being suddenly applied to the tube or from shot effect. With this in mind, assume that the oscillations in the resonant cavities are already taking place. From this assumption examine the source of the energy needed to sustain these oscillations. With the tank circuit oscillating, a high frequency voltage, e, appears between the two cavity grids. This makes the electric field between these grids reverse twice each complete cycle of operation. In this condition as the electrons approach these grids, the electron stream is uniform. The time that is required for the electrons to pass through the small distance between the grids is small compared to the period of oscillations. Electrons which enter the space between the grids when e is zero will encounter no AC electrical field and they pass on through at the same velocity. The electrons which enter the space when e makes the left grid negative with respect to the right will encounter a field which tends to accelerate them. The amount which they are accelerated is a function of e. Electrons entering the space when e is reversed in polarity are decelerated. The change in velocity due to acceleration and deceleration is small in comparison with the original velocity. Those electrons which are accelerated most will travel farther toward the repeller plate before being turned back, while those that are decelerated most will be turned back before approaching very close to the repeller. Thus it is conceivable that with the proper magnitudes of e, E_a and E_R, electrons returning to the cavity grids will arrive in bunches.

The diagram on the next page shows the position of electrons in the tube at various times during their transit. The zero distance position is midway between the cavity grids. Electron A, which arrives when e is positive, is accelerated and travels farther before being turned back; electron B is unaffected; electron C is decelerated and turns back after a shorter excursion. Hence, in the diagram these electrons and the ones passing through at intermediate times are shown as arriving back at the grids at the same instant of time. This is the ideal situation, but it is not difficult to see that electrons will return to these grids in a stream which varies in intensity at the frequency of the oscillations. It is due to this fact that this tube is called a velocity-modulated tube. It is called a *reflex* velocity-modulated tube because the electrons reverse direction and travel through the interelectrode space twice.

On the return trip the electric fields set up by the voltage e again act upon the electrons. Since they are now traveling in the opposite direction, they will be decelerated if they return



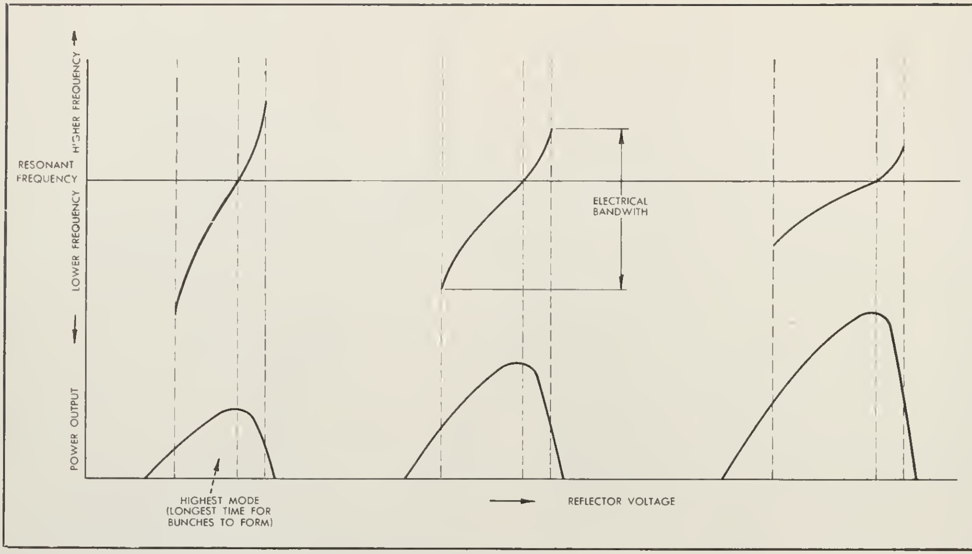
Bunching Action of Reflex Velocity Modulated Tube

when e is positive and accelerated if they arrive when e is negative. You are already familiar with the fact that an electron which is accelerated by an electric field has its kinetic energy increased and that this additional energy is taken from the electric field. On the other hand, an electron which is decelerated gives up energy to the electric field. Now, if the bunches of electrons can be made to arrive back at the grid when e is positive they will give up energy to the alternating field. For maximum transfer of energy the bunches must arrive when e is maximum positive. The question arises as to where did this energy originate. You have seen that if the electron stream from the cathode is uniform, some electrons are accelerated and some are decelerated, on the outbound trip, by the electric field of e. On the average, as many electrons absorb energy from the field as give up energy to it. Hence, very little net energy is taken from the oscillating circuit during the bunching process. The average kinetic energy of the electron is that imparted to it by the DC voltage E_a. Thus, you see how energy is taken from the DC electric field and transferred to the AC field to sustain the oscillations.

Modes of Operation

It is not necessary that the bunches of electrons return to the grids on the first positive swing of e after they leave them. The illustration directly above indicates that if the bunches arrive on the second positive swing, the net result is still the same. You can see that the time in transit for the average electron B is $\frac{3}{4}$ cycle, $1\frac{3}{4}$ cycle, $2\frac{3}{4}$ cycle, etc. In actual practice there are three or four "modes" in which it is possible for the reflex-velocity-modulated tube to oscillate. The black lines in the illustration indicate paths of electrons operating in the first mode while the red lines indicate paths of electrons operating in the second mode. A third mode is possible when the average transit time is $2\frac{3}{4}$ cycles, etc.

The following shows you how the transit time is controlled to produce oscillations in the different modes. If you give a little thought to this, you can see that since the original velocity of an electron depends on the DC voltage E_a, and since the distance that the electron travels before turning back and the speed with which it returns depend upon the difference between E_a and E_r, it is possible to adjust the two volt-



Power Output and Frequency versus Reflector Voltage in Different Voltage Modes for a Reflex Velocity Modulated Tube

ages E_a and E_r for any of the modes. The voltage E_a is usually fixed in magnitude since varying it produces greater initial velocity, which in turn causes a farther excursion and a greater return velocity. Since it is not feasible to make E_a variable, E_r is variable. For operation in the first mode, the round trip must be completed in the shortest time. This is accomplished by making the repeller plate most negative. For greater time in the interelectrode space, the repeller is made less negative.

In the above diagram showing power output and frequency of oscillations as functions of the repeller voltage for three modes of operation, notice that the frequency at the point of maximum output is same for all three modes and is the resonant frequency of the cavity. In addition, note that the power output for the various modes at the resonant frequency are not the same and that the output is least in the highest mode. This can be explained by examining the factors which limit the amplitude of oscillations and which in turn limit the power output.

Power and amplitude limitations are due to overbunching as well as the usual losses in the oscillatory circuit. Overbunching occurs in the following way. As oscillations build up and e becomes greater, the amount of acceleration and deceleration increases. This causes bunching to occur within a shorter period of time, that is, in a time before the electrons reach the grids on the return trip. This tends to reduce the magnitude of oscillations. In the higher modes of oscillations where the bunches are formed more slowly the electrons are more susceptible to overbunching. The magnitude of e which results in overbunching is therefore lower and oscillations are limited by this action to a lower amplitude than in the lower modes of operation.

As shown in the above diagram, the frequency of oscillations in a reflex-velocity-modulated tube is variable to a limited degree in any of the modes of operation by means of varying the repeller voltage. When the repeller voltage is varied, it causes a bunch to return a little sooner or a little later than normal. Off resonance, the amplitude of oscillations decreases by an amount which depends on the Q of the cavity. In this tube the tuning range is small in comparison to the frequency of oscillations and varies somewhat from one mode to another, being greatest in the highest mode. This can be explained from the fact that in the highest mode, bunching and debunching take place at a slower rate and that greater variation from the ideal time of return is possible without debunching, causing the amplitude of oscillations to drop below this usable output level. Another way to look at it is that in the highest mode the interval between leaving the grids and returning is greater, and the change in period represented by a given change in frequency is a smaller portion of the interval. To illustrate, in the third mode the interval before return must be about 2¾ cycles. A small change in the period of e would therefore be only 3/11 as great a portion of the interval as it would if operation were in the first mode where the ideal time interval is ¾ cycle.

The band of frequencies which can be obtained by varying the repeller voltage lie between the half power points shown in the diagram on page 11-38. This range of frequencies is known as the *electrical bandwidth*. The output curves of the bandwidth are unsymmetrical about the maximum output points. This results from the fact that if E_r is increased, not only does the bunching voltage e decrease and cause bunches to form at a later time, but the repeller voltage causes a quicker return and the effects of the two actions add to cause poor bunching at the time the electrons return, resulting in a rapid drop in output on the high side of the hump. At lower voltages, however, even though the bunching voltage e decreases and causes slower bunching, the decreased repeller voltage causes a later return to the grids so that the two effects are counteracting and a greater change in repeller voltage is possible before the output drops below the usable level.

The choice of the point and mode of operation is a compromise among several factors. To begin with, there are three or four modes which have the necessary power output. On the whole, it would seem then that the correct choice would be the highest mode, for it gives the largest tuning range. The highest mode, however, is too sensitive to a change in voltage to be very well regulated. A change of one volt may cause a change of 0.5 mc in the 3000 mc oscillator. Since the humps are unsymmetrical, the point of operation is usually chosen a little below the point of maximum output. This makes possible the tuning above the operating frequency by a greater amount than if the maximum point were used.

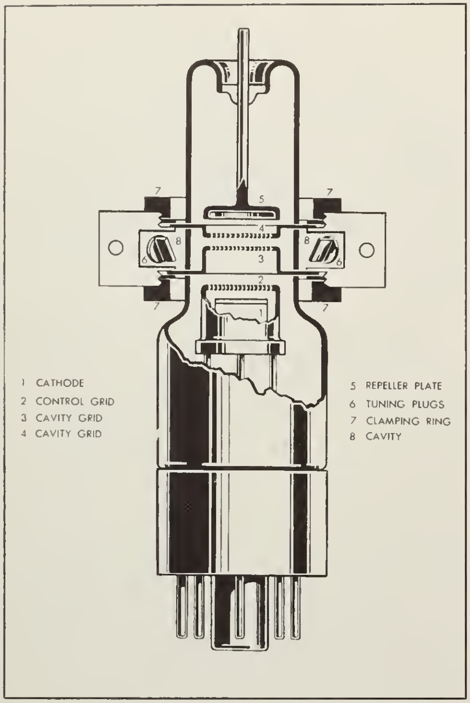
In practice, the reflex velocity-modulated tube is usually used in conjunction with an automatic frequency control circuit, a circuit which controls the repeller voltage in such a way as to maintain the correct intermediate frequency. The details of the discriminator circuit and its operation were taken up in Chapter 8. Keep in mind that the frequency of oscillations is primarily determined by the dimensions of the cavity and that the repeller voltage is effective in making small changes in the frequency. Hence, in most reflex velocity-modulated tubes there is a course frequency adjustment which varies the cavity size in some way and the repeller voltage is the fine frequency adjustment.

Tubes

The following table gives some of the operating characteristics of reflex velocity-modulated tubes. The data in this table is given only to give an idea of the order of magnitude of tube quantities since there is wide variation between different tubes and different conditions of operation.

No.	Name	Mfr.	Freq. (Mcps)	Acc. Voltage	Beam Cur. (Ma)	Repeller Voltage	Control Grid Voltage	Power Output (mw)	Elec. Tuning (Mcps)
K417	Klystron	Sperry	3000	300-600	5-30	+50 to −500	−5 to +50	150	5
707A	McNally	W. E. and Raytheon	3000	250-325	25-35	0 to −250	same as acc.	75	30
726A	Shepherd-Pierce	W. E.	3000	300	22	−20 to −300	same as acc.	100	20
723A	Shepherd-Pierce	W. E.	9400	300	18-25	−20 to −300	same as acc. (internal connection)	20	45

Typical Reflex Velocity-Modulated Tubes



707A (McNally) Tube

The K417 reflex Klystron is one of the earlier types which was used for 10 cm. operation. One feature of this tube was that in its early application it did not have provision for controlling the frequency through a change in the repeller voltage since both coarse and fine frequency controls changed the cavity grid spacing.

Another 10 cm. tube is the 707A (McNally) tube shown to the left. In it the cavities are external to the tube and are not evacuated. This makes them susceptible to changes in temperature and results in changes in frequency. To get good frequency stability in it, it is necessary to control the cavity temperature. The coarse frequency control consists of plugs which when screwed into or out of the cavity change its size. Fine frequency is controlled by the variable repeller plate voltage control.

An all metal tube which is available for both 10 cm. and 3 cm. operation is the Shepherd-Pierce tube. In it the cavities are located inside the tube. Mechanical coarse tuning is accomplished by means of struts on the side of the tube. The struts are adjusted by a screw, which in turn varies the size of the cavity. The repeller voltage control also serves as the fine frequency control. The 10 cm. and 3 cm. type Shepherd-pierce tube differ in the shape of the cavity and in the method of coupling the output.

CHAPTER 12

Radar Antennas

This chapter deals with the basic principles of antennas. It discusses antennas in general, the principles of electromagnetic radiation and its application to radar antennas, various antenna arrays, and typical airborne radar antenna systems.

FUNCTION OF THE ANTENNA

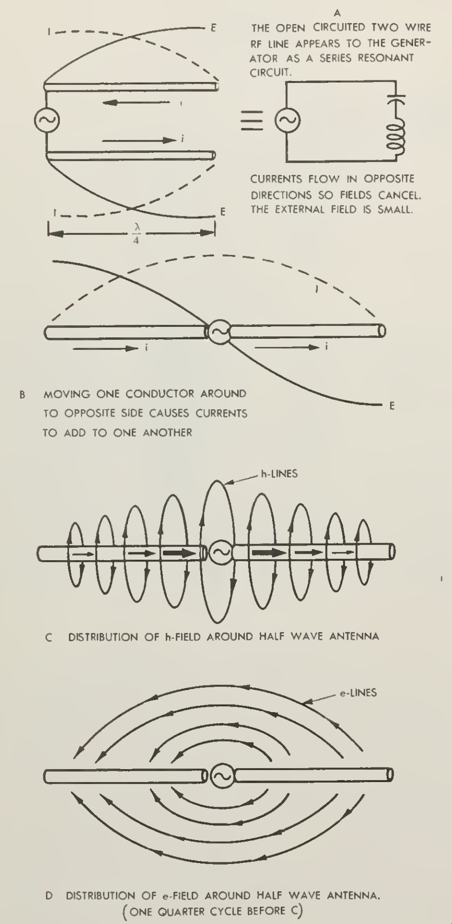
An antenna is an electronic device that is used either for radiating electromagnetic energy into space or for collecting electromagnetic energy from space. In the radar transmitter, the magnetron generates the high frequency signal, but an antenna is needed to change this signal into electromagnetic fields which are suitable for propagation into space. The radar receiver will amplify any signal that appears at its input terminals, but an antenna is required to intercept the electromagnetic fields that are in space and to change these fields into a voltage which the receiver can interpret.

ANTENNA RECIPROCITY

Fortunately, separate antennas seldom are required for transmitting and receiving radio energy, for any antenna transfers energy from space to its input terminals with the same efficiency with which it transfers energy from the output terminals into space, assuming, of course, that the frequency is the same. This property of interchangeability of the same antenna for transmitting and receiving operations is known as antenna *reciprocity*. Antenna reciprocity is possible chiefly because antenna characteristics are essentially the same regardless of whether an antenna is sending or receiving electromagnetic energy. Because of antenna reciprocity, most radar sets installed in aircraft employ the same antenna both for receiving and transmitting. An automatic switch in the radio frequency line first connects the single antenna to the transmitter, then to the receiver, depending upon the sequence of operation. Because of reciprocity of radar antennas, this chapter treats antennas from the viewpoint of the transmitting antenna with the understanding that the same principles apply equally well when the antennas are used for receiving electromagnetic energy.

DIRECTIONAL PROPERTIES

Usually, the most important characteristic of a radar antenna is its directional property or simply its directivity. Directivity means that an antenna radiates more energy in one direction than in another. For that matter, all antennas are directional, some slightly; others, almost entirely. In radar operations, some antennas are required to send all energy in one direction in order that as much as possible of the electromagnetic energy generated by the transmitter will strike an object in a given direction. In other systems, it is desirable for the energy to be radiated equally well in all directions from the source. An example of an antenna system in radar which radiates energy in a given direction is the airborne navigation and bombing set. In this set, there is only a limited amount of power available at the transmitter. In order to achieve maximum benefit from this minimum power, all of it is sent in the same direction. Since the antenna in this set is also used for reception, it likewise receives electromagnetic energy only from one direction. Because of design features, it is possible to tell the direction of an object at which this directional type an-



Development of Half-Wave Antenna from Open-End Quarter-Wave RF Line

tenna is sending energy or the direction of the object from which the antenna is receiving energy. Furthermore, the physical position of the antenna is indicative of the direction of the object. An example of a non-directional radar antenna is the antenna installed in the radar beacon. This antenna must receive energy equally well in all directions in order that a radar equipped airplane can ascertain its position regardless of its direction from the beacon antenna.

ANTENNA EFFICIENCY

An efficient antenna is one which wastes very little energy. The higher the antenna efficiency, the less the energy loss. The greatest cause of decreased antenna efficiency during the conversion of RF energy into electromagnetic energy is the loss of energy in the form of heat. Heat losses are discussed later.

Actually antenna efficiency may be high; in a single antenna element the efficiency may be as great as 90%. An entire antenna system, properly designed, can concentrate all its radiated energy in a given direction with virtually none being lost in other directions. A directional receiving antenna is more sensitive to signals in a given direction than a similar non-directional antenna is to signals in all directions. This greater sensitivity, which is another way of stating there is less energy lost by the antenna, represents greater receiving ability, and accordingly greater efficiency.

ELECTROMAGNETIC ENERGY IN SPACE

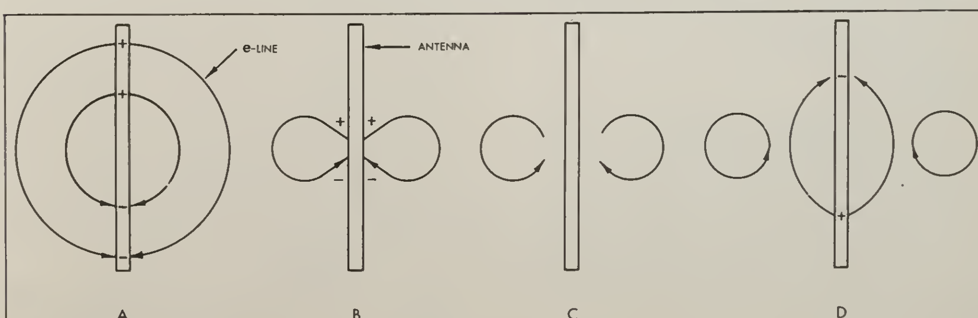
Because the half wave dipole type antenna is the fundamental element in an antenna system, it can be used as a starting point for discussing the radiation of electromagnetic energy into space. Electrically, you can think of the dipole antenna as an open circuited and shorted quarter wave length RF line which is excited by a generator and which appears to the generator as a resonant circuit. At the left in the top diagram A, there is a dipole which consists of a piece of wire which is cut in half and attached to the terminals of a high frequency alternating current generator. The frequency of the generator is such that each half of the wire is one-quarter of the wave length of the generator output. Due to the resonant characteristics of this arrangement, high circulating currents will flow in the antenna when it is excited by the generator. This current flow sets up two strong fields about the antenna—one the magnetic field (H) and the other the electric field (E). However, due to the U-shape of the antenna, and to the fact that the currents in each wire move in opposite directions, the H-fields and E-fields are in opposition and cancel, making the overall field strength of the antenna assembly very low. Suppose, however, the two wires are moved as far apart as possible, as shown at B. In this case, the current flow in each wire is in the same direction, and the resulting E-fields and H-fields no longer oppose, but now add and consequently create strong fields about the antenna, as you can see in diagrams C and D. Later you shall see that one of these fields is responsible for electromagnetic radiation.

The distribution of current (I) and voltage (E) about an antenna is shown by curves shown in the illustration. The dotted line represents current distribution, and the full line, voltage distribution. The current curve at B shows that most of the current flows at the center and none of it at the ends of the antenna. The voltage, on the other hand, is maximum at the ends and minimum at the center. The magnetic field is greatest where the current is greatest, as you can see in the diagram at C, and the electric field is strongest at the outer ends and weakest at the center. Note that, as in the case of a resonant circuit where voltage is maximum, that similarly about the antenna the E-field is maximum at the time the H-field is minimum. Both of these fields vary at a sinusoidal rate with a time difference of one-fourth cycle or a difference of 90° between them.

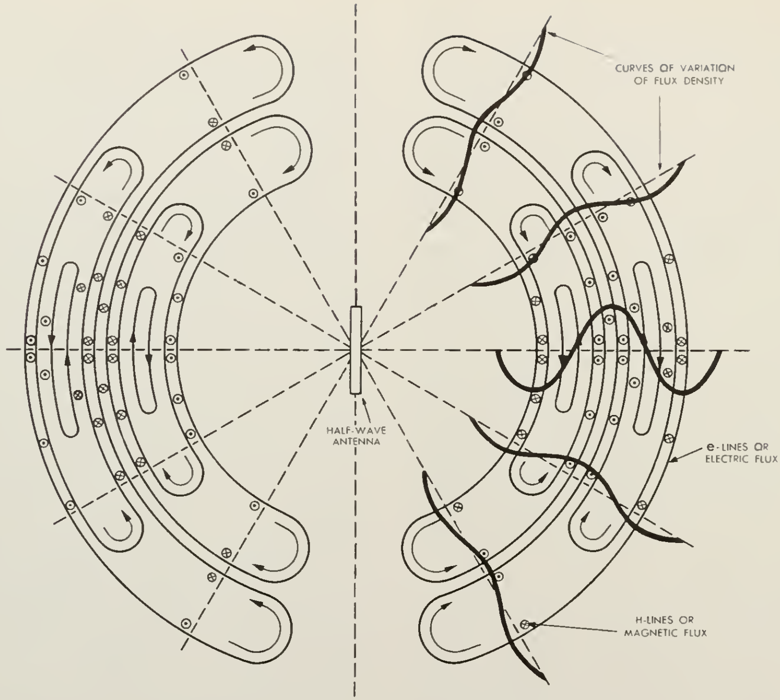
The fields associated with the energy stored in the resonant circuit (antenna) are called the *induction* fields. These fields decrease with the square of the distance from the antenna. They are only local in effect and play no part in the transmission of electromagnetic energy. They represent only the stored energy in the antenna and are responsible only for the resonant effects which the antenna reflects to the generator. The fields set up in the transfer of energy through space are known as the *radiation* fields. Although these fields decrease with distance from the antenna, this decrease is much less rapid than the decrease of the induction fields. This is because it is linear and is not according to the square rule. Therefore, the radiation fields reach great distance from the antenna. It is the radiation fields that are responsible for electromagnetic radiation.

Radiation of Fields

The exact cause and the mechanism for the phenomenon of radiation is not known to the electronic scientist. However, you can get a pretty good idea of what takes place by studying the illustration below showing a simple picture of an E-field detaching itself from an antenna. At A in this illustration, the E-field is maximum. Notice that the outer e-lines are stretched away from the inner ones. This is due to the repelling force between the lines of force. As the voltage decreases toward zero, the radiation field decreases and the e-lines retract back into the conductor. If the voltage decreases slowly, the entire field will collapse back into the conductor. On the other hand, if the decrease of voltage is rapid, the outer parts of the field cannot move in very fast and when the voltage is almost zero, as shown at B, a relatively large E-field will still exist around the conductor. At C an exact zero voltage condition is shown. The E-field in this case does not reach zero, and it is left with no voltage to support it. When the next E-field develops around, it will be pushed away by the preceding E-field in the manner shown. The action of one field pushing away the preceding field is analogous to the snapping of a whip, in that a part of the E-field is snapped off the antenna with each cycle. The snapped off field is projected into space and moves away from the antenna at a constant speed of 186,000 miles per second. A similar action projects H-fields into space.



E-Field Detaching Itself from an Antenna



Fields in Space Around a Half-Wave Antenna

Another factor to remember about radiation is that the ease with which it may occur, varies with frequency. At lower frequencies, such as 60 cps power frequency, for example, voltage on the antenna changes so slowly that the component of energy radiated is so extremely small that it is of no practical value. At higher frequencies, such as 50,000 cps and up, the radiated energy is great enough to meet communications requirements.

Fields in Space

The above illustration shows the manner in which radiation fields are propagated from the antenna. The E-field and the H-field are shown as separate sets of flux lines about the antenna. The electric flux lines are the closed loops on each side of the antenna. The magnetic flux lines are closed circular loops which have their axis around the antenna, or in other words, the antenna is the center of each loop. They are represented as dots and crosses. The sine waves which are labeled the curves of radial variation of flux density indicate the relative field strength at various distances and angles from the antenna. Since the field configuration is not a standing wave, but a traveling wave phenomenon, the magnetic and electric fields are in phase and thus the sine waves apply to the flux density of either field.

In the direction perpendicular to the antenna, both the H-field and E-field are strong, for this is the direction where both fields originally formed. In the direction parallel to the antenna, or off the ends, no H-field forms at all and only a very small E-field. The flux density, therefore, is small in these two directions. In other words, due to the directional properties of the half-wave antenna, most of the radiated energy travels in the direction perpendicular to the antenna, but very little energy in the direction along its axis.

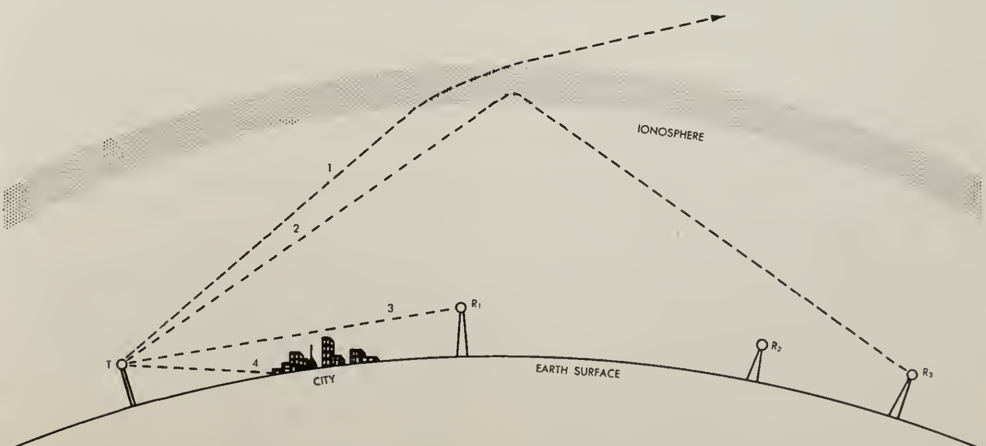
As previously stated in the discussion of wave guides, no moving E-field can exist without an H-field, and no H-field can be propagated without an associated E-field. Similarly, with the propagation of electromagnetic energy into space, no moving electrostatic forces can exist without magnetic stresses existing in space and no magnetic force in motion can exist without an electrostatic stress. Each creates the other and one cannot exist without the other. The direction of motion is the travel of the fields away from the antenna. The current associated with the magnetic field does not flow because in space there is no conductor to carry current, but the field exists, nevertheless. To visualize a field existing without current flow, think of a moving magnetic field cutting a glass rod. A voltage (electrostatic stress) is induced in the rod, but there is little actual electron movement because the rod is a good insulator. Magnetic lines move in space, and set up electric stresses in space in a similar manner.

Propagation into Space

Any part of the electromagnetic field set up by an antenna travels away from the antenna in a straight line. There are many parts to this field and many directions in which energy travels. In the illustration below are four of the large number of paths which the energy can take. These and other paths which energy takes affect reception of radiated energy. If there is nothing between the emitting antenna and receiving station (R_1) some energy will travel directly to it via path 3. Receiving station R_2 cannot receive energy because the earth is between the two points and because the energy cannot go through the earth for any distance. Some of the energy will follow path 2 out into space. At a height of some 60 miles above the earth, there is a heavily ionized layer of atmosphere, called the *ionosphere*. This constitutes a change of media through which the energy must travel, and is sufficient to refract a wave. In the case of path 2, the refraction is sufficient to bend the energy wave out of the ionosphere back to the earth. Therefore, receiving station R_2, which is beyond the horizon and not located for receiving a direct ray, receives the reflected ray from the ionosphere.

It is the ionosphere that makes possible round-the-world communications. If the angle of incidence of an energy wave with the ionosphere is too great, the angle refraction may not be sufficient to bend the energy back toward the earth and it may go on through the ionosphere and be lost, as shown by path 1. Paths 3 and 4 follow along the surface of the earth and are called *ground waves*. The waves bent back to the earth by the ionosphere are called *sky waves*.

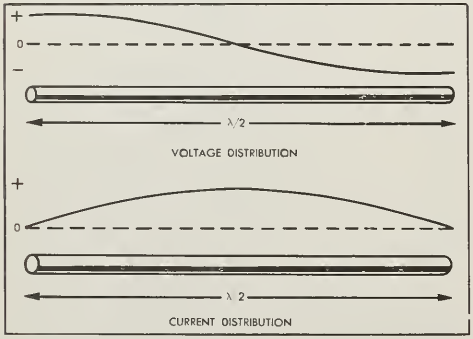
Generally, medium and high frequency waves are reflected readily, whereas ultra high and super high frequencies are more likely to be lost through penetration of the ionosphere. Airborne radar operates in the super high-frequency (3000 mc to 30,000 mc) part of the spectrum. Irregularity of the sky waves makes them unreliable for radar use. Therefore, radar uses only the direct ground wave. Path 4, for example, where the ray goes directly to some object and returns by a direct path is a direct ground wave.



Effectiveness of Energy Traveling in Various Directions from an Antenna

THE HALF-WAVE ANTENNA

As stated previously in the discussion of electromagnetic radiation, a half-wave length conductor is the simplest of the radiating elements. Considerable radiation occurs in this element because of its resonant characteristics and its ability to store large amounts of energy in induction fields. Resonance causes high voltages and high circulating currents and they in turn produce strong fields around the antenna.



Current and Voltage Distribution in Half-Wave Antenna

Current and Voltage Distribution

According to the above illustration showing voltage distribution in a half wave antenna, voltage standing wave is high at the ends of the antenna and low at the center. As previously explained, this is also the case with the quarter wave open circuit two-wire line from which the half-wave antenna is developed.

An examination of the current distribution curve shows that the current standing wave reaches maximum a quarter cycle after the voltage reaches maximum at which time the current is maximum at the center and zero at the ends. At the ends where there is no place for electrons to go, the current is zero. In practical applications the ends of a half-wave antenna must be insulated due to the high voltages there and the center of the antenna must have low resistance in order to minimize the I^2R losses due to the high current there.

RESONANCE AND DIMENSIONS

Electromagnetic waves travel through space at a speed of 300 million meters per second. The length of one cycle in space depends upon frequency and is called the *wave length.* Mathematically, the length of an electromagnetic wave is expressed by the formula,

$$\lambda = \frac{3 \times 10^8}{f}$$

where λ is the wave length in meters, and f is the frequency in cycles per second.

Since an electromagnetic wave travels on an antenna, the antenna, too, has wave length. But, due to the resistance of the wire, the movement of waves along a wire (antenna) is somewhat slower than wave movement in space. Wave length on a wire, therefore, is slightly less than that of a wave traveling in space and is expressed by the equation,

$$l = \frac{3 \times 10^8 \times .94}{f} \text{ meters}$$

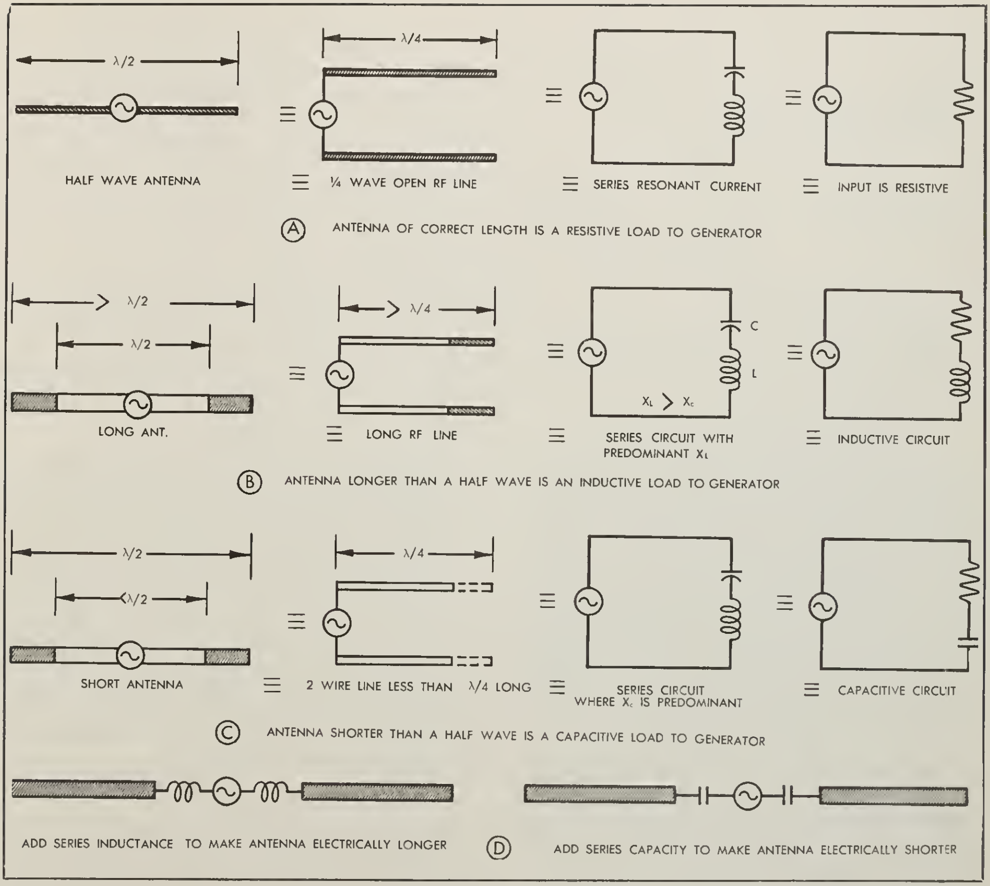
Physically, an antenna is about 6% shorter than a half wave traveling in space. (This accounts for the fact that it is necessary to multiply the wavelength in the antenna formula by .94.) In this manual, assume that the correction in antenna length has been made whenever the length of the antenna is given in wavelength.

Effect of Length on Antenna Impedance

An antenna of the correct length acts like a resonant circuit and presents pure resistance to the excitation circuit. (See A in the illustration on the next page). An antenna having other than the correct length displays both resistance and reactance to the excitation circuit. An antenna slightly longer than a half wave, for example, acts like an inductive circuit. This is understandable if you think of the antenna in terms of a quarter wave RF line. When the antenna is excited at the center, it is equivalent to a ¼ wave RF line, looking at it from the generator end. Any two-wire line, longer than a quarter wave, is like a quarter wave section with a short, open-circuited section added to it. The open section, which is capacitive, is inverted to inductance at the input terminals by the quarter wave. In the same manner a slightly long antenna looks inductive.

According to the equivalent lumped circuit at B, the inductance is not entirely balanced by the capacitance. The remedy for correcting the length is either cutting the conductor shorter, or tuning out the inductive reactance by adding capacity in series.

If the antenna is physically shorter than its resonant length, the input impedance becomes capacitively reactive. The two-wire line shorter



Effect of Length on Antenna Impedance

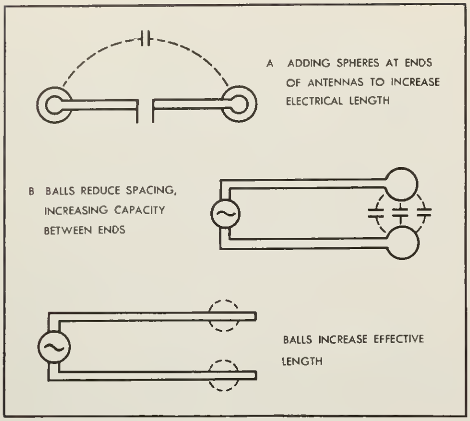
than a quarter wavelength displays capacity at its input terminals. (See diagram C.) The correction in this case is either to add inductance in series with the antenna to bring it back to resonance, or to add physical length to the antenna.

Increasing Antenna Length

A practical method for increasing the length of an antenna which is physically short is shown at A in the illustration at the top of page 12-8. The spheres fitted at the ends add to the capacity from end to end. As shown at B, this is equivalent to adding capacity at the end of the two-wire line. The additional capacity comes from the closer spacing between conductors. Adding capacity to the line gives the same result as adding length. Therefore, the effective length of the antenna in this case is equal to one-quarter wavelength.

With radar equipment, a wide band of frequencies—rather than a single frequency—must be handled by an antenna. This means that an antenna which is designed for the center frequency is short for frequencies in the lower (low frequencies) sidebands and long for frequencies in the upper (high frequency) sidebands. The antenna must be corrected for both long and short conditions at the same time.

Note in the chart at the bottom of the next page showing the variations in the reactance



Increasing Electrical Length of Antenna

of an antenna that each side of an antenna must be less than a quarter wave (in space) in order to present zero reactance. This agrees with the previous statement that the physical length of the half-wave element must be 6% shorter than the half wave in space. Two curves are shown. One is for a large diameter antenna, and the other for a small diameter antenna. Note that the larger diameter displays less reactance at lengths off resonance.

Antenna diameters as great as 1/20 wave length are not uncommon in radar equipment. A large diameter increases the capacity of the antenna. The inductance is decreased for the same resonant frequency. Lower inductance with the same resistance lowers the X_L/R ratio, or Q of the antenna. A lower Q causes the resonance curve to be broader and gives the antenna a more uniform response to a greater band of frequencies.

INPEDANCE AT INPUT TERMINALS

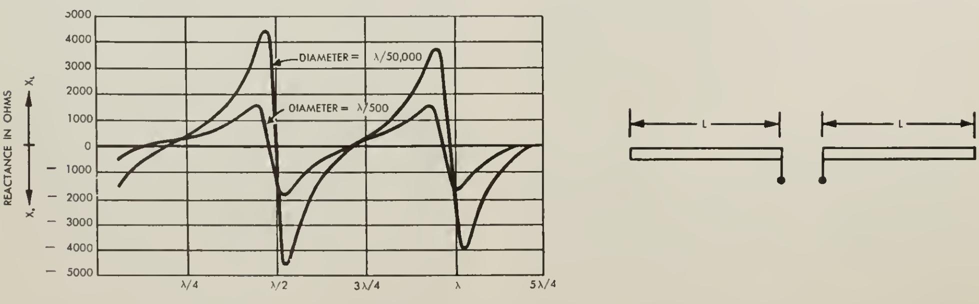
In practice an antenna acts like a resonant or near resonant circuit and, theoretically, like a perfectly tuned circuit. When it looks like a series resonant circuit, its input terminals must display zero impedance, and when it is connected like a parallel resonant circuit, it must display infinite impedance. Further, if it is not of the correct length, it must display reactance as well as resistance.

Practical antennas, like practical resonant circuits, have losses which must be replaced by the source. These losses make the input impedance of the antenna somewhat resistive. This resistance is a combination of two resistances—the resistance of the conductors themselves, which is increased by the skin effect at high frequencies, and the radiation resistance.

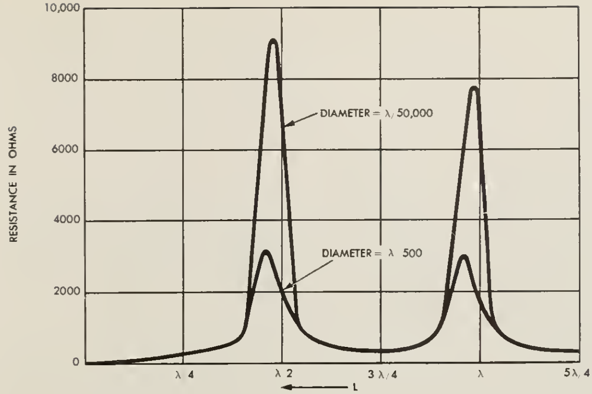
Radiation resistance is a fictitious resistance which dissipates the same amount of power in the form of heat that is actually dissipated as radiated energy. Because of radiation resistance, an antenna allows part of its fields to escape into space. This makes it necessary for additional current to flow into the antenna from the source of power to replace the part of the field that escaped. Therefore, the source must provide the energy lost by radiation. The energy lost is power (P) and its ratio to the square of the current (I) is the radiation resistance, R_r. Mathematically, it is expressed,

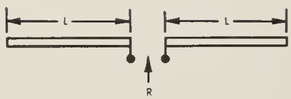
$$R_r = \frac{P}{I^2}$$

Radiation resistance may be defined as the ratio of the power radiated to the square of the current in the antenna. Radiation resistance and conductor resistance constitute the total input resistance to an antenna. It is possible to make the conductor resistance small in comparison



Reactance at Input of a Center-Fed Antenna of Arbitrary Length





Resistance at Input of a Center-Fed Antenna of Arbitrary Length

to the radiation resistance by using large diameter low resistance conductors. It is desirable that a major portion of the input resistance becomes the radiation resistance. In practice radiation ratios as high as 9:1 are obtainable. This means that an antenna can be 90% efficient as a radiator.

Above is the graph showing the input resistance at the center of antennas of various lengths. For the half-wave dipole (L=λ 4), the input is 73 ohms. The input resistance of a full wave antenna is (L=λ 2) is as high as 9000 ohms, depending on the diameter of the conductor.

For the lengths of antennas in which the reactance is zero, the input resistance is the input impedance of the antenna. However, for lengths in which the reactance is not zero, the input impedance is the vector sum of the resistance (graph above) and the reactance at the input of the antenna, graph page 12-8.

Other factors which affect the input impedance of an antenna are nearby conducting objects, such as the earth, and other antennas or the skin of the aircraft. The graphs above and on the preceding page apply only to a center fed antenna which is located in free space and not close to any conductor.

POLARIZATION OF AN ELECTROMAGNETIC WAVE

Electromagnetic fields in space are said to be *polarized* and the direction of the electric field is considered the *direction* of polarization. As the electric field is parallel to the axis of a half-wave dipole, the antenna is in the plane of polarization. When the half-wave dipole is horizontally orientated, the emitted wave is horizontally polarized. A vertically polarized wave is emitted when the antenna is erected vertically.

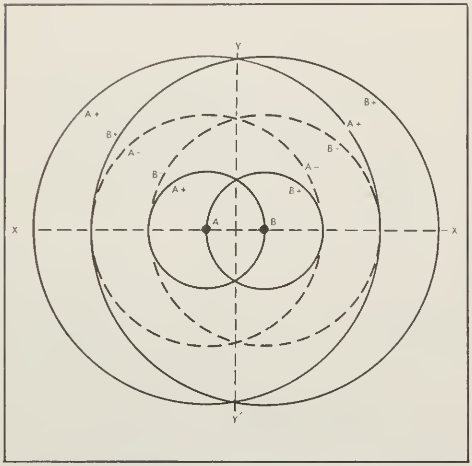
For maximum absorption of energy from the electromagnetic fields, it is necessary that a half-wave dipole be located in the plane of polarization. This places the conductor at right angles to the magnetic lines of force that are moving through the antenna, and parallel to the electric lines.

In general, the polarization of a wave does not change over short distances. Therefore, transmitting and receiving antennas are orientated alike, especially if short distances separate them.

Over long distances, the polarization changes. The change is usually small at low frequencies. At high frequencies, the change is quite rapid.

With radar transmissions, a received signal is a reflected wave from some object. As the polarization of the reflected signal varies with the type of object, no set position of the receiving antenna is correct for all returning signals. Generally, the receiving antenna is polarized in the same direction as the transmitting antenna.

When the transmitting antenna is close to the ground insofar as propagation is concerned, vertically polarized waves cause a greater signal strength along the earth's surface. On the other hand, antennas high above the ground should be horizontally polarized to get the greatest signal strength possible to the earth's surface. For this reason most airborne radar systems emit horizontally polarized waves.



Vector Addition of Electromagnetic Energy from Points A and B

DIRECTIONAL CHARACTERISTICS

If you measured the radiation from a single point in space, you would find that this point radiates equally well in all directions. The strength of the radiated energy varies inversely as the distance. If all the points where the energy was of the same strength were plotted, the points would form a sphere, with the radiating point at the center.

If radiation occurs from more than one point in space, the radiated signal from each point will add vectorially to produce the total radiation strength.

Note two points in space in the above illustration showing vector addition of electromagnetic energy. Each point is radiating a similar field in all directions. Concentric circles are shown around each point. The circles represent positive and negative peaks of the radiation field. The solid lines are positive peaks. For simplicity, the points are placed a half wave apart. The picture is stopped just as each point is ready to generate a negative peak.

Note the peaks as they cross the vertical line YY′. Adding the two fields there doubles the field strength.

Next observe the peaks along the line XX′. Wherever a positive peak from one point appears, a negative peak from the other also appears. This means that cancellation occurs along the line XX′.

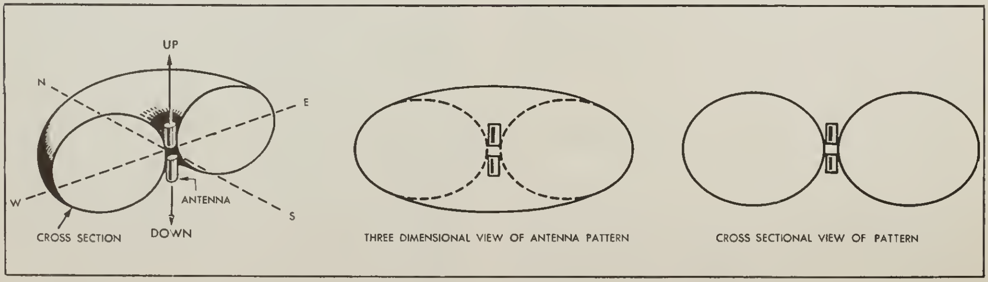
The radiation from a pair of points is directional. Most of the energy is detected along YY′, while little or no energy is detected along XX . The strength at angles between X and Y varies as the cosine of the angle.

When you apply these facts to a half wave antenna, you can look at the conductor as a series of points arranged in a straight line. The radiation from points equally spaced from the center will add in directions perpendicular to a line through them, but will cancel along the line through them. A plot of all points of equal strength would produce a figure which is not a sphere, but which is a doughnut-shaped, three dimensional figure as shown below. Note that it resembles two almost perfect circles which are adjacent to one another.

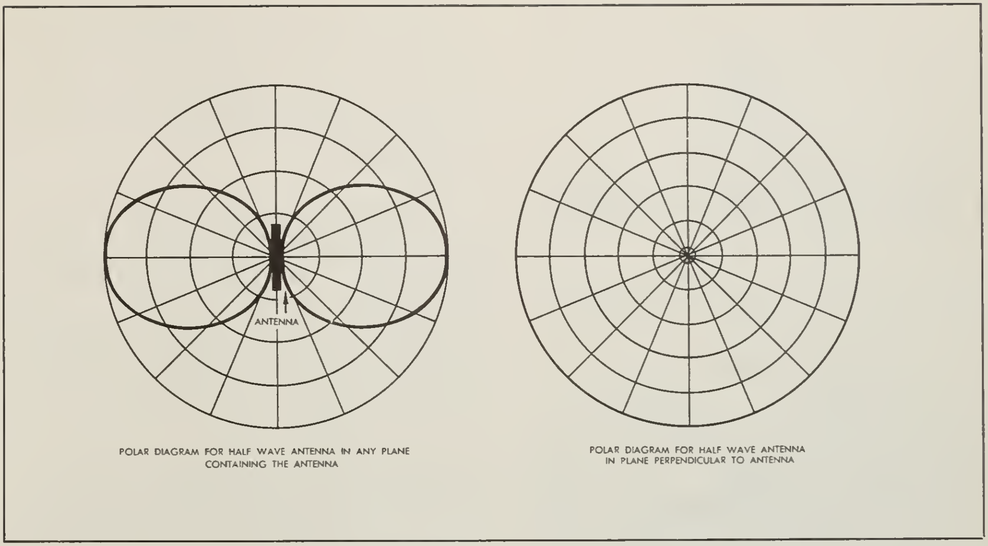
The cross-sectional view of the preceding illustration is called a *polar diagram*. This sort of diagram is standard for delineating antenna performance on paper.

Constructing Polar Diagram

There are two methods of constructing a polar diagram. One method consists of measuring the field strength at the same distance in all directions from the antenna and plotting the angle against field strength on polar coordinate paper.



Cross Sectional View of Vector Addition Figure



Polar Diagrams of Half-Wave Dipole

The other method consists of moving around the antenna in all directions, finding the points where the field strength is the same, and plotting the angle against distance on polar coordinate paper. Either type of measurement will produce identical polar diagrams.

Above note the polar diagrams for the half-wave dipole. The left diagram holds good for any plane containing the antenna. For the plane perpendicular to the half-wave antenna, the antenna forms the single point which was derived previously. The polar diagram for this point becomes a circle with the point at the center as in the right diagram. Thus, the simple half-wave dipole is bidirectional in any plane containing the antenna, but nondirectional in the plane perpendicular to the antenna.

The strength of a radiation field is called *field strength* and is measured in units called *volts per meter*. One unit is defined as the field strength which will induce one volt in a conductor one meter long. As field strengths of a volt per meter are seldom encountered, weaker fields are measured in *microbolts per meter* or *millivolts per meter*.

IMAGE ANTENNAS

The preceding discussion dealt with antennas that are isolated from any conductor. However, in all practical cases, usually there are conductors somewhere near the antenna. The effect of the conductors is often undesirable but can sometimes provide advantages.

The effect of a nearby conductor is illustrated diagrammatically below at A. A real antenna is shown perpendicular to a horizontal perfectly conducting plane. When a field is radiated, energy will be sent out in all directions. There are two

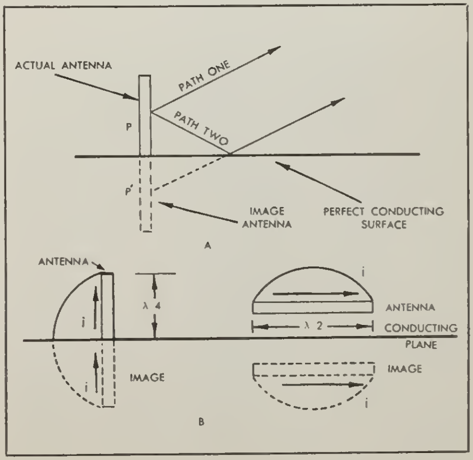
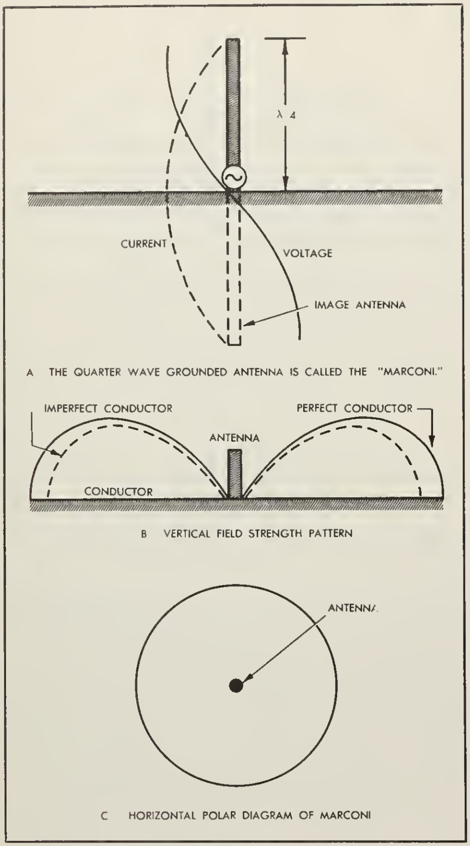


Image Antennas



Marconi Antenna

possible paths for the field to take. Path one starts out directly toward a distant point. The energy in path two starts out toward the conducting plane. Upon striking the perfect conductor, it is reflected in reverse phase. (This is analogous to an incident wave being reflected in reverse phase from the shorted end of an RF line.) As with light waves, the angle of reflection is the same as the angle of incidence. The reflected wave also proceeds toward some point in space. The sum of the two waves make up the total wave at any point in space. So far as the action in the conducting plane is concerned, it can be replaced by another antenna, which is a mirror image of the actual antenna. The reflected wave can be assumed to have originated at point P′ on the image antenna.

In general, current in vertical image antennas flows in the same direction as in actual antennas, while in horizontal antennas, the current flows in opposite directions. Note at B on page 12-11 that the combination of the real and image antennas for vertical quarter wave makes a half-wave dipole.

Marconi Antenna

A quarter wave grounded antenna is a common type of grounded antenna. This type is often called a *Marconi* antenna, as contrasted with the half-wave (ungrounded) dipole, which is called a *Hertz* antenna. Note the standing wave amplitude of current and voltage on the Marconi antenna shown to the left. Note also the similarity to the half wave element when the image is included.

In the Marconi antenna the vertical field strength pattern (polar diagram) shown at B is the same as that of the half-wave element, except that the conducting plane cuts it off at the center. The image is only effective above the plane because no energy penetrates the conducting plane there.

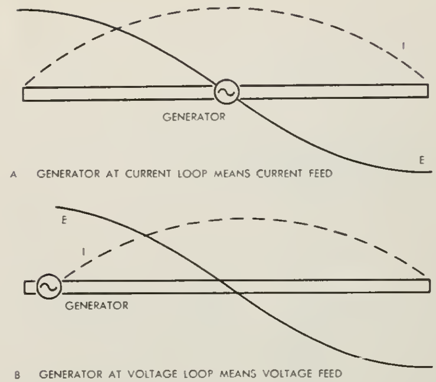
In the horizontal polar diagram at C, the vertical field strength pattern can be rotated with the antenna as the axis to form the horizontal polar diagram of a Marconi. It is non-directional in a plane perpendicular to the length of the antenna.

The input impedance to the Marconi is approximately 37 ohms when the antenna is fed at its base as illustrated at A. In addition a quarter-wave Marconi is resonant and displays zero reactance just like a half-wave antenna.

This discussion has assumed that the conducting plane is a perfect conductor. If it is not a perfect conductor, as is the case in practice, some of the conditions just discussed must be altered. The principal alteration is the reduction in size of the polar diagram. This results in decreased signal strength from the antenna as shown at B. The conducting plane is usually the skin of the aircraft where airborne equipment is concerned and with ground equipment, it is the earth's surface.

Other types of conductors that might be near an antenna are the aircraft tail pieces, the airplane wing, or steel antenna towers. As the radiation field passes any conductor, currents are induced in them. These currents vary at the same radio frequency and make the conductor itself a radiator. In other words, when the conductor is excited it re-radiates, creating a radiation field of its own. This is the way in which radar "echoes" are produced.

When the radiation field of the conductor is strong, it will change the radiation pattern of the original antenna in two ways. First, IR^2 losses in the conductor absorb power from the radiation field. This reduces the field strength in areas beyond the conductor. Second, the re-radiated field combines with the antenna radiation field and produces random phase relationships. If they occur in phase, a strong resultant field will be present in the area. If they are out of phase, the field strength will be low. Usually the presence of conductors is undesirable. Sometimes they are deliberately placed by the antenna to distort the field strength pattern. This is discussed later under arrays.



Current and Voltage Feed

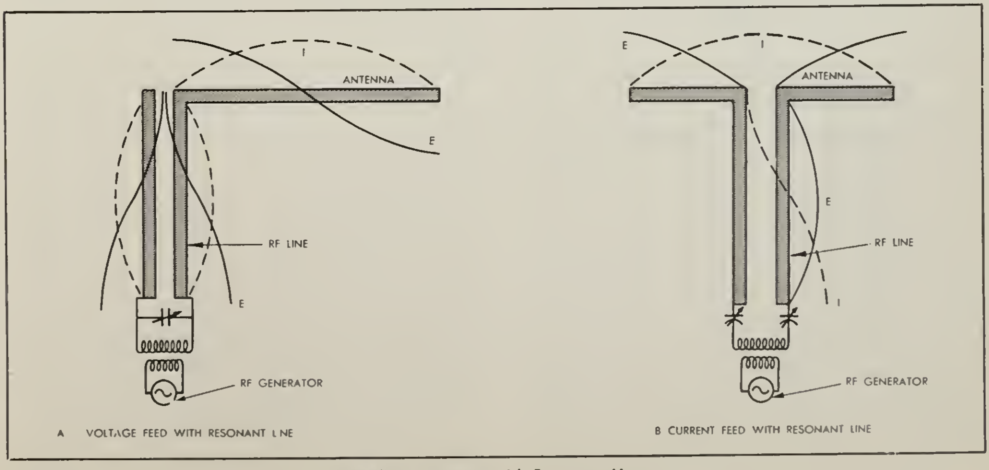
INTRODUCING ENERGY TO THE ANTENNA

Although energy may be fed to an antenna in a variety of ways, most antennas are *voltage fed* or *current fed* as shown above. When the excitation energy is introduced to the antenna at the point of high circulating currents, the antenna is said to be *current fed*. When the energy is introduced at the point of maximum voltage, the antenna is said to be *voltage fed*.

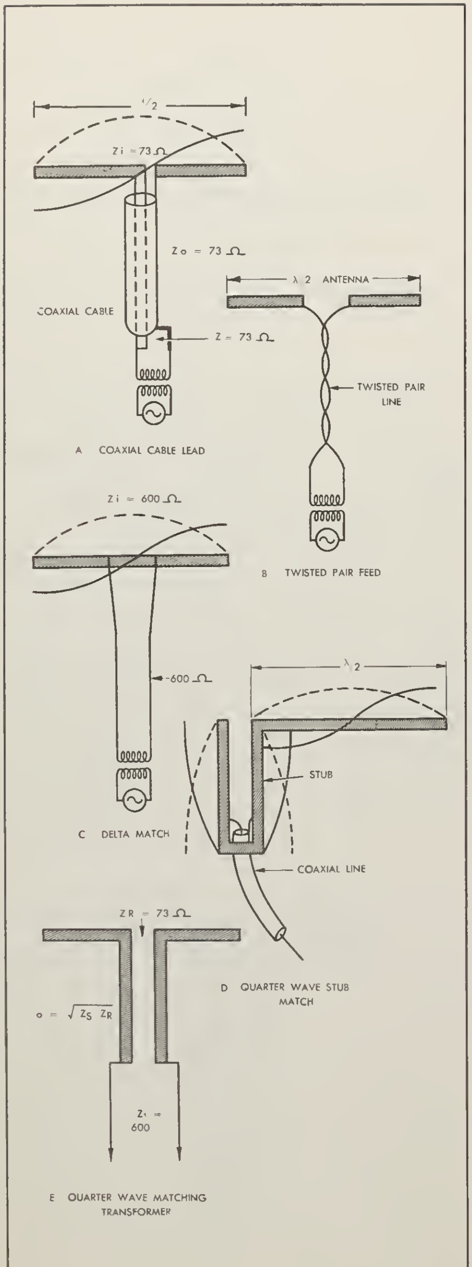
It is seldom possible to connect a generator directly to an antenna. Instead it is necessary to use RF lines to carry the energy from the generator to the antenna. The RF lines which carry the excitation energy might be resonant lines, non-resonant lines, or a combination of both.

Feeding Antennas with Resonant Lines

In a voltage fed half-wave antenna which is excited by a resonant RF line, as shown at A below, the end of the antenna is connected to one side of the line. Voltage changes at that point excite the antenna into oscillation. The impedance at the end of the antenna is high, since there the voltage is high and the current there is low. The RF line shown is a half-wave length two-wire line. The impedance of such a line is usually low in order that the high impedance at which it is terminated will set up standing waves on it. In addition it is al-



Feeding Antennas with Resonant Lines



Feeding Antennas with Non-Resonant Lines

ways a multiple of quarter wavelengths electrically in order that it will act like a resonant circuit. This makes the input to the line also a high impedance. A parallel resonant circuit is used to develop a high voltage across this high input impedance. Small irregularities in line lengths can be compensated for by tuning the parallel resonant circuit at the input.

In the current fed antenna with a resonant line shown at B on the preceding page, the RF line is connected to the center of the antenna. This antenna has a very low impedance at the center and, like the voltage-fed antenna, has standing waves on it. It also is a multiple of quarter wavelengths and is a resonant circuit. Making it exactly a half-wave length causes the impedance at the sending end to be low. A series resonant circuit is used to develop the high current necessary to excite the line. Adjusting the condensers at the input compensates for irregularities in line and in antenna length.

Although these examples of feeding antennas are simple ones, the principles described apply to antennas and to lines of any length—providing both are of resonant length.

The line which is connected to the antenna can be a two-wire or a coaxial line. In radar equipment the coaxial line is preferred.

One advantage of the resonant line is that it makes matching devices unnecessary. In addition it makes it possible to compensate for incorrect antenna lengths by transmitter tuning. Its disadvantages are increased I^2R losses due to high standing waves of current on the resonant feed lines, increased possibility of arc-over between lines due to standing waves of voltage, very critical length, and production of radiation fields by the two-wire line due to the standing waves on it. Radiation fields are undesirable because they distort the normal radiation pattern of the antenna.

Feeding Antennas with Non-Resonant Lines

The illustrations to the left show the excitation of a half-wave antenna by non-resonant lines. As the input to the center of the antenna at A is 73 ohms and as a coaxial cable with the highest Q has a characteristic impedance of 73 ohms, the most common method to feed this antenna is through a coaxial cable connected directly to the center of the antenna. This method of connection produces no standing waves on the line (coaxial cable) when a generator is matched to the line. Coupling to the generator is usually made, as you see at A, through a simple untuned transformer secondary. Another means of efficient transfer of energy to the antenna can be accomplished by a twisted pair line as shown at B. In a twisted pair the impedance is about 70 ohms. It, like the coaxial cable, provides a good impedance match.

When a line does not match the impedance of the antenna, it is necessary to use special impedance matching devices. Any of the impedance matching RF lines discussed in Chapter 9 are adequate for this purpose. An example of a type of impedance matching device is the delta match shown at C. It gets its name from the fact that the wires form the Greek letter Delta at the antenna.

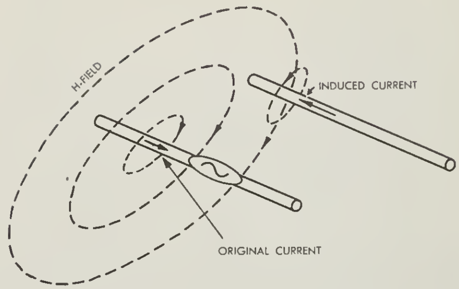
The delta match is used to match a line having more than 73 ohms impedance to the center of an antenna. The line in the example shown is a 600-ohm line. The connections from the line are spread apart at the antenna until the impedance between connections is equal to 600 ohms.

In the delta match as the connections are moved away from the center, the voltage becomes higher and the current becomes lower. The ratio between these, known as the E/I ratio, increases from a very low value at the center to several thousand ohms at the ends. The two connections are set at the point where this ratio is equal to 600 ohms. This is the same as matching various impedances along a shorted quarter-wave section of line.

Another method of matching impedance is the quarter-wave stub match shown at D. In the quarter-wave stub match the high impedance at the end of the antenna matches the open end of the stub. The impedance on the stub varies from zero at the short circuit to several thousand ohms at the open end. This makes it possible to connect a 70 ohm coaxial line a short distance from the shorted end at the 70 ohm point. It is possible to match almost any impedance along the length of the stub.

Still another impedance matching device is the quarter-wave transformer (see E). This device is used for matching the low impedance at the antenna to the line of higher impedance. In the example shown, the characteristic impedance, Z_0, of the quarter wave section is equal to 210 ohms. With this matching device standing waves will exist on the antenna and the quarter wave section but not on the 600 ohm line itself.

Non-resonant lines are characterized by small radiation, low voltage at all points for any given power, and non-critical length. In addition they are preferred for the transfer of power over distances greater than one or two wavelengths.



Exciting Antenna by Radiation Fields from Nearby Antenna

Exciting Antennas by Radiation Fields

The antenna may be excited either by a direct connection or by radiation (induction fields) from nearby antennas. A half-wave antenna driven by a directly connected generator is shown in the illustration above. The flux from the driven antenna expands and cuts the other element and induces in it a current which varies at a frequency nearly equal to the exciting frequency of the driven antennas and which in turn causes additional radiation fields. When the antenna elements are properly designed, the radiation field resulting from the induced current will be nearly as strong as the original fields.

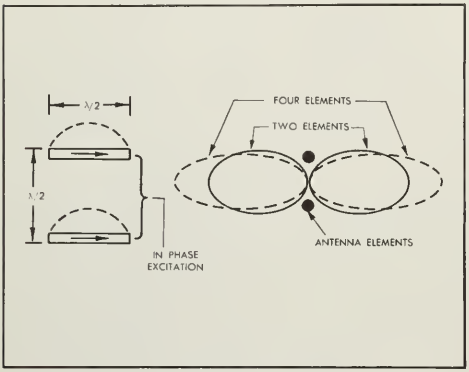
Besides being excited by a nearby antenna, an antenna may also be excited by radiation fields from a distant transmitting antenna. This takes place in any receiving antenna. In a receiving antenna the current which is conducted to the receiver is the result of current induced by radiation fields generated by the transmitting antenna.

ANTENNA ARRAYS

There are uses for antennas where it is desirable that the radiation be absolutely non-directional and other uses where the antenna must be very sharply directional. In the latter case the antenna system usually consists of two or more simple half-wave elements so spaced that the fields from the elements add in some directions and cancel in others. The set of antenna elements is called an *antenna array*. There are two types of antenna arrays—driven arrays and parasitic arrays.

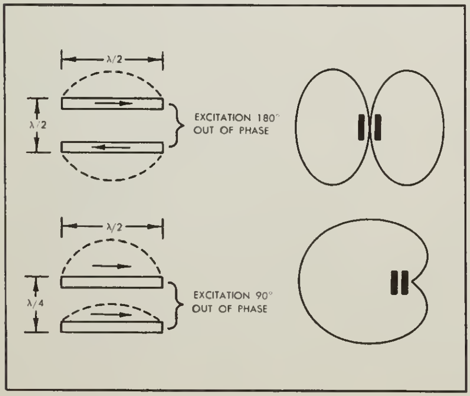
Driven Array

The driven array consists of two or more elements with all elements connected to the generator. They may be divided into three basic types—the broadside array, the end fire array, and the collinear array.



Broadside Array

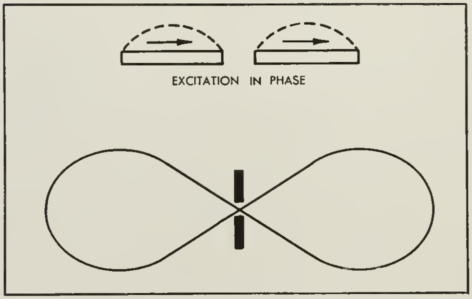
BROADSIDE ARRAY. When two half-wave elements are placed one half wave apart and parallel to each other and excited in phase, most of the radiation takes place in a direction perpendicular to a plane through the elements. The arrangement and radiation pattern of a broadside array is shown in the illustration above. In it, the radiation pattern is shown in solid lines. Increasing the number of elements makes the pattern more directional. This increased directivity is shown by the dotted lines.



End Fire Array

END FIRE ARRAY. When two elements are spaced a certain fraction of a wavelength apart and are excited out of phase by the same fraction of a cycle, the radiation is directional in the plane of the array and perpendicular to the elements. It is also off the end of the array. This is the reason why the array is called *end fire*. If the spacing in an end fire array, for example, is ½ wave, the two elements are excited ½ cycle or 180° out of phase. This causes the bidirectional pattern shown in illustration at the bottom of the page. A uni-directional cardioid pattern can be obtained with ¼ wave spacing and ¼ cycle (90°) phase difference in excitation.

COLLINEAR ARRAY. A collinear array is formed when two half-wave elements are placed end to end and excited in phase. In a collinear array there is no directivity in the plane perpendicular to the antenna, but there is a sharp pattern in any plane containing the antenna. As with any of the other arrays, increasing the number of half wave elements increases the directivity of the pattern.



Collinear Array

Parasitic Arrays

A parasitic array is an antenna system which consists of two or more elements in which only one of the elements is driven. The other element (or elements) is excited by induction and radiation fields which are produced by the driven element. With parasitic arrays it is possible to obtain highly directional patterns.

The action in a parasitic array is analogous to the action in a transformer in which the primary induces a current in the secondary and the current in the secondary produces a magnetic field, which in turn induces current back into the primary. Its action differs from that in a transformer in that where the phase relationship between the primary voltage and secondary current in a transformer is always 90°, the phase relationship between elements in an array varies according to the spacing between the elements. The elements are usually spaced an appreciable part of a wavelength apart. This makes it possible for the phase difference to vary over a range of more than 80°.

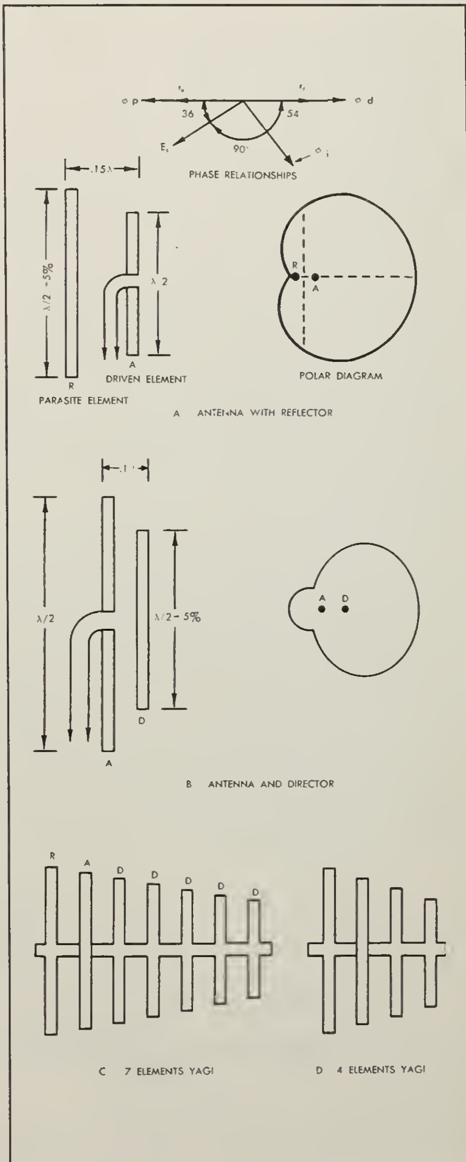
In a two element parasitic array, shown to the right at A, the driven element (labeled A) is cut at the center for connecting a low impedance feedline. The length of the driven element is a half wavelength. This makes it self resonant. The parasitic element called the parasite is located 15% of a wavelength in space from the driven element. It is about 5% longer than the driven element.

Another diagram at A shows the phase relationships in the parasitic array vectorially. Vector i_d, which represents the current in the driven element, is in phase with the h-field. The part of the h-field that cuts the parasitic element lags the field which leaves the driven element by .15 of a cycle. (This is the time elapsed during the travel between elements.) The lag is equivalent to 54°. The flux at the parasite is shown by vector ϕ_i. It lags i_d by 54°. The voltage induced by the field is 90° out of phase with the field. This voltage is represented by the vector E_p. If the parasitic element were resonant, the current in it would be in phase with E_p. But the parasite is longer than the resonant half wave length. A long antenna is inductive and the current in this element will lag the voltage by 36° if it is approximately 5% longer. The radiated field will be in phase with this current. In summary, the field starting out from the parasitic element will be 180° out of phase with the field leaving the driven element.

If the polar diagram for two elements spaced .15 wavelength apart and excited out of phase is plotted, the curve at A results which shows that most of the radiation is on the side of the driven element which is away from the parasitic element, while very little occurs on the side of the parasite.

The parasite acts somewhat like a reflector, since it directs most of the power to the other side of the driven element. Because of this action, the parasite in this parasitic array is also called the *reflector*.

Another point is that the field passing from the reflector cuts the antenna (driven element) and induces a voltage in it. This voltage changes the input current. The input impedance, which is a function of this current is about 50 ohms as compared to 73 ohms for the antenna alone.



Parasitic Array

A parasitic element becomes a director when it is made shorter than the antenna element. In a director most of the energy is sent in a direction from the antenna element to the parasitic element. To see what takes place note the radiation pattern and the arrangement of the array in the diagram at B on the preceding page. The director is usually 5% shorter and about .1 wavelength from the antenna. Sometimes the impedance is reduced to 20 ohms at the driven element in this array.

In radar systems, the parasitic array usually consists of 2 to 7 elements. Notice the element array called a Yagi antenna at C. In it the antenna or driven element is insulated, but the reflector and all directors are welded to a piece of tubing which runs parallel to the direction of propagation. The beam width of this array is 19°. The four element Yagi shown at D is constructed similarly to the one at C. It has a beam width of about 50°.

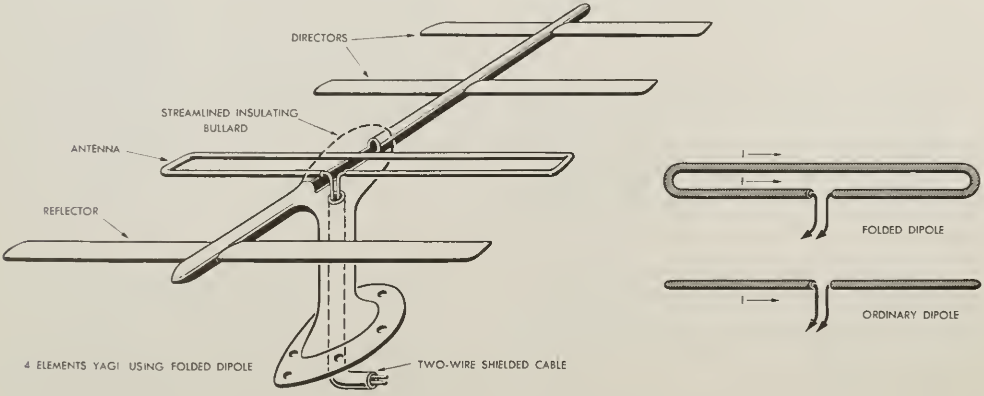
The conventional method for describing the directivity of an antenna array is in terms either of the ratio of the power in the best lobe to the power radiated by a simple half wave antenna, or the ratio of the power in the direction of the best lobe to the power in the opposite direction. The direction of the best lobe is usually called *forward* direction, while that in the opposite direction is called the *backward* direction. In other words, this ratio is the power in the forward field to the power in the backward field. For example, the front to back ratio of the antenna array illustrated at B was about 5:1. In decibels, this represents approximately a 7 db gain. The ratio between forward power and power from a single half wave antenna is about 4 db. The gain of an array using a reflector is somewhat less than the gain in the arrays just illustrated. For this reason, two element arrays used with radar equipment usually employ a driven element and a director.

Comparisons between the more directive types of arrays are made in terms of the beam angle. This angle is the angle between half power points in the main lobe.

In arrays the half power points are the points where the electric field strength is .707 as great as that along the axis of the beam.

In multi-element arrays the input impedance drops as low as 15 ohms. In these arrays special matching devices are needed for matching this low impedance to the higher impedance of most RF lines. One method for making the match is to use the Delta matching system previously described. Another method, the one most often employed in radar equipment, uses the folded dipole as shown below.

A folded dipole is a full wavelength conductor which is folded to form a half wave element. A better description is that it consists of a pair of half-wave elements connected together at the ends. In it the voltage at the ends of each element must be the same. In operation the field from the driven element induces a current in the second element. This current is the *same* as the current in the driven element.



Folded Dipole

An ordinary dipole with a given current I produces a certain field intensity in space. Due to this field, there is also a certain power density per square meter in space. This power density is produced by the input power P. The relationship between the input resistance, the current, and the input power is expressed by the equation,

$$R = \frac{P}{I^2}$$

Another consideration with the folded dipole is that when the same current I exists in each of the two sections, the field strength in space is doubled. This causes the power density per square meter to increase four times. In turn the input power must be four times as great. In this case to balance the equation, it is necessary to multiply R by 4 as follows:

$$\frac{4P}{I^2} = 4R$$

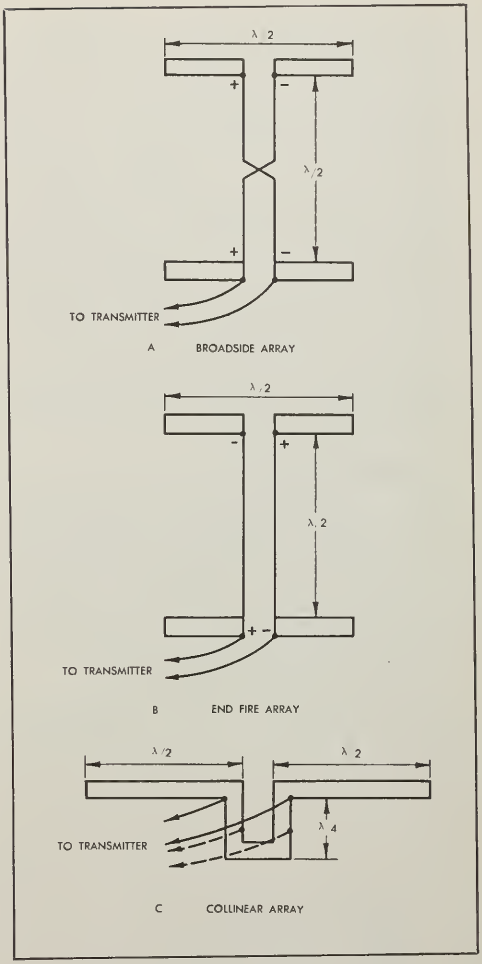
So long as each section of a dipole has the same diameter, the input resistance is four times that of the simple half wave dipole. Increasing the diameter of one section makes the increase in impedance still greater. The input impedance to the driven element of a parasitic array drops to about a fourth of the value of the coaxial cable impedance, but the use of a folded dipole increases the impedance by about four times. In this way a good impedance match is effected.

Interconnection of Elements in Arrays

There are a variety of ways to connect the elements in an antenna array to obtain the required phase of excitation. The most convenient method to change the phase from one element to another element a ½ wave apart is with an RF line. In the broadside array shown above at A, one element is excited directly from the transmitter. This element is connected to the other element by a half-wave length of RF line. The phase of the voltage along the RF line is shifted 180° per each half wavelength. The leads to the second driven element are reversed. This causes another 180° phase shift. The second element is then excited in the same phase as the first.

It is possible to connect any number of elements in this manner. In addition increasing the number of elements makes the array more directive.

In the end-fire array at B the elements are also interconnected with RF lines. But instead of reversing the leads to elements, they are connected directly to take advantage of the phase shift. With half-wave spacing, the interconnecting line will be half wave long and provide a 180° phase shift. If quarter-wave spacing is used, the interconnecting line will be a quarter wave long. This arrangement produces a 90° shift.



Connecting Half-Wave Elements to Obtain Proper Phase Relationship

The usual method for providing the correct phase in a collinear array is shown at C. Since the current direction changes for each half wave length of an antenna, it is not possible to connect the half-wave sections together directly.

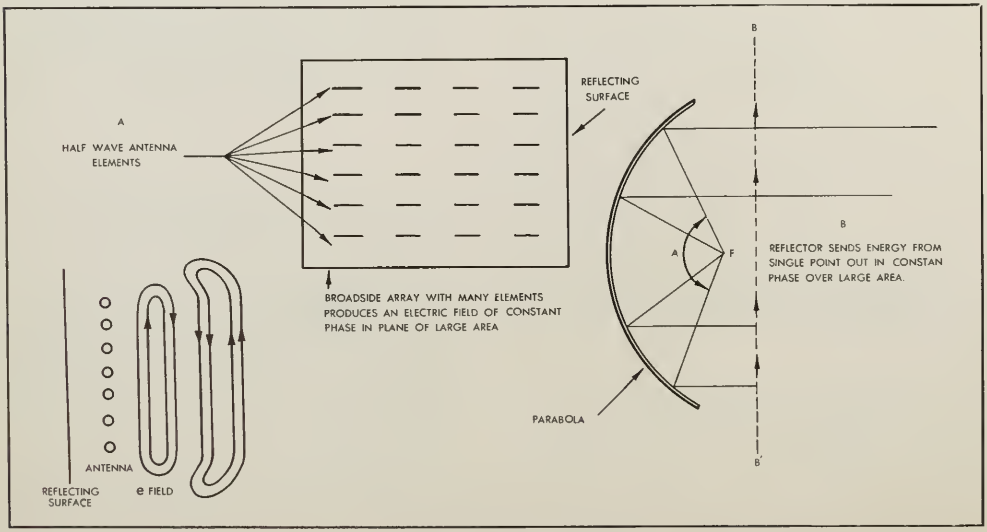
Instead, the half-wave section which carries the current in the wrong direction is folded to form a quarter wave section of RF line. This brings the ends of the sections in which current flows in the same direction together. In other words, in terms of voltage, it is necessary that the two antenna sections have voltages which are opposite polarity at the open end. Further, note how the line to the transmitter is connected at C. The connection is made at a point of very high impedance. This requires the use of a resonant line. If, however, a non-resonant line is to be used, it must be connected as shown by dotted lines. In the diagram the connections near the closed end of the quarter wave section is a point of low impedance. Another way of correctly exciting a collinear array is by a waveguide. This method is described later.

The Parabolic Reflector

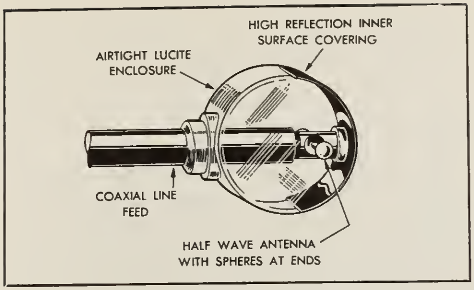
When a multi-element broadside array is excited, the e-field which exists in front of the antenna will be in a single plane as shown at A below rather than in an arc as is the case in a single half-wave element. The larger the dimensions of this plane in terms of wavelengths, the greater is the directivity of the antenna system, and the narrower the radiated beam.

Although a multi-element broadside array gives good results, it is quite complicated in structure. In it every element must be driven, and all spacings and dimensions must be quite exact. A much simpler device for producing an electric field in a single plane is the parabolic reflector. As you can see in diagram B below, the parabola has its focal point at F. If a single antenna is placed at F and caused to radiate a field, the electric field will leave the antenna in all directions at the same rate in the form of an arc as indicated at point A. As each part of the wavefront reaches the reflecting surface, it is shifted 180° in phase and sent outward at an angle of reflection that is equal to the angle of incidence. All parts of the field will arrive at line BB′ at the same time after reflection because all paths from F, to the reflector, to line BB′ are equal in length. Thus you see that with only one antenna and a specially shaped reflector, it is possible to produce a large electric field in a single plane. Or, looking at it another way, all parts of the field travel in parallel paths after reflection from the parabola in a way that the rays are focused like the headlight beam from an automobile.

Like the broadside array, high directivity is not obtained until the diameter of the parabolic reflector is made many wavelengths long. This prohibits the use of the parabolic reflector at low frequencies, but for three- and ten-centimeter radar equipment they are very practical.

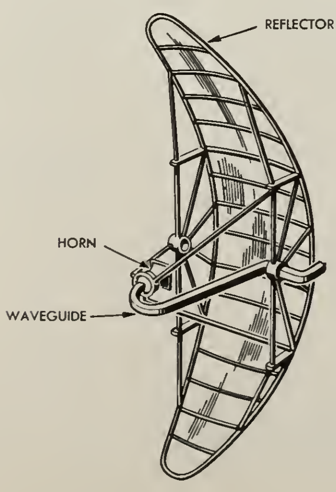


Development of Parabolic Reflector



Dipole for Exciting Parabolic Reflector in 10 CM System

In the illustration above showing the exciting antenna for a parabolic reflector, the half-wave dipole is mounted a quarter wave back from the short on the coaxial line. To sharpen the focal point the antenna is physically less than a half wave long. However, the balls at the end make it electrically a half wave long. This broadens the band of frequencies it will handle. The airtight cylinder in which the antenna is enclosed permits the coaxial line to be pressurized. The inner surface of half the cylinder, that is, the side away from the parabolic reflector, is coated with a reflecting foil. This reflecting surface directs energy from that side of the antenna into the large reflector. Without this reflector half of the antenna radiation would be non-directional.



Orange Peel Parabola with Waveguide Feed

The natural directivity of a dipole causes the pattern from a parabolic reflector to be somewhat sharper in the plane containing the dipole than in the other plane. For this reason, the dipole is erected horizontally for maximum azimuth accuracy in radar systems. If vertical accuracy is of primary importance, the dipole is mounted vertically.

Other methods of preventing the direct forward radiation use a parasitic reflector with the half-wave dipole, and a disc which is placed in front of the dipole.

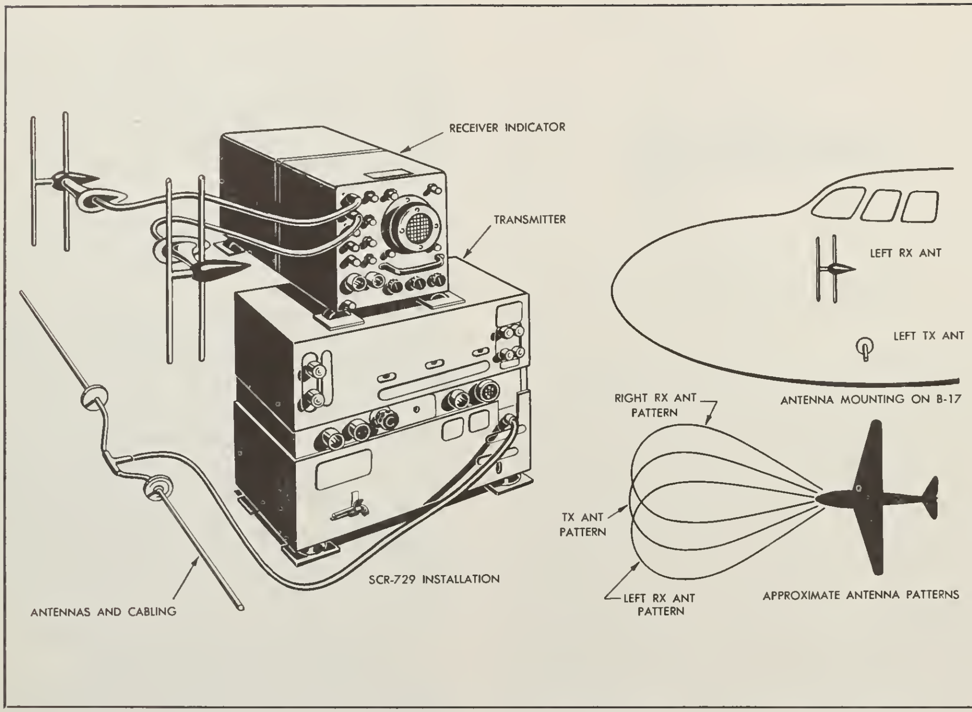
When a wave guide is used in the RF system, it is possible to send the energy into the parabolic reflector with a horn radiator. In this case an orange-peel reflector, or section of a complete circular paraboloid as shown below may be used as the reflector. The feed is by a waveguide, and a horn type radiator. The horn just about covers the shape of the reflector and prevents very little RF energy from escaping at the sides. This arrangement is highly directive in the vertical plane. It is used principally to determine the altitude of airplanes.

TYPICAL AIRBORNE ANTENNA SYSTEMS

THE 175 MC ANTENNA SYSTEM

Before 10-cm and 3-cm radar equipment was developed most radar sets operated at a frequency of about 175 mc. Today most of these low frequency sets are obsolete. However, one which was still in use at the close of the last war is a small airborne set which is used to trigger radar beacons and radar blind landing systems. In brief, this is how it works. The transmitter sends out a 175 mc pulse. When this signal is intercepted by the radar beacon, the beacon immediately transmits another pulse which has a frequency slightly different than the one sent out by the transmitter. This new pulse is picked up by the set's receiving antenna and as this antenna is directional, it indicates the direction from which the beacon pulse came. From the direction indicated by the antenna and the range indicated by the time base, it is possible to determine the position of the beacon from the aircraft.

The illustration on the next page shows the placement of the antennas used with this set on the aircraft. As you can see, the transmitting antenna consists of two quarter-wave grounded antennas, one on each side of the airplane. The skin of the airplane forms the ground,



Typical 175 MC Antenna System

or the reflecting surface. These antennas are equivalent to two half-wave antennas placed end to end and spaced somewhat apart. Both are fed by a coaxial cable from the transmitter. This cable is split by a T-junction into two cables, each of which goes to the antennas. As these cables are equal in length, each antenna can be fed in phase. Further, since no reflectors are used, the antenna system is only directional in the plane perpendicular to its axis.

The receiving antennas consist of a pair of separate two-element arrays. Each array is connected to the receiver by a separate coaxial cable. This cable in turn connects to a half-wave dipole. The dipole removes most of the rear lobe. This makes the forward gain high. (The impedance at the input is not reduced enough in this array to require a folded dipole.) A parasitic director is mounted ahead of the antenna. The mounting, director, and extension holding the director are welded into one piece, while the antenna is insulated from this with a streamlined block of low-loss material.

The receiving antennas are mounted on the side of the aircraft in the manner illustrated. One on each side produces a pair of antenna patterns. The right antenna is mostly directional to the right, while the left antenna has its greatest lobe to the left. This means that a signal source to the left of the aircraft will cause a greater current to be induced in the left antenna than in the right antenna. By rapidly switching the indicator between left and right antennas it is possible to determine by the relative signal strengths the approximate position of the signal source.

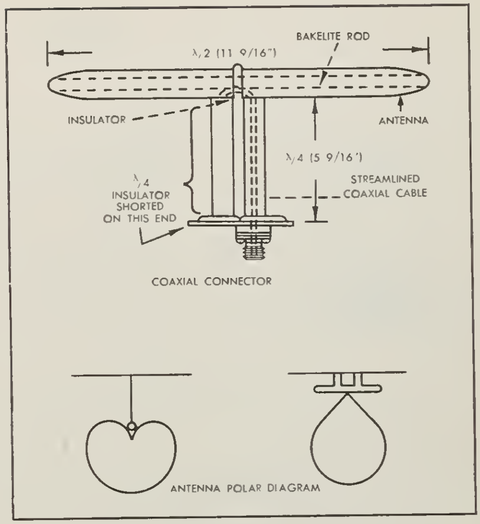
A 400 MC ANTENNA SYSTEM

A typical example of a set which uses a 400 mc antenna system is the 400 mc radar altimeter. In this set both a frequency modulation system and a pulse modulation system are employed. Both systems operate at the same frequency and cover a wide band characteristic. The antenna used is interchangeable between the two modulation systems.

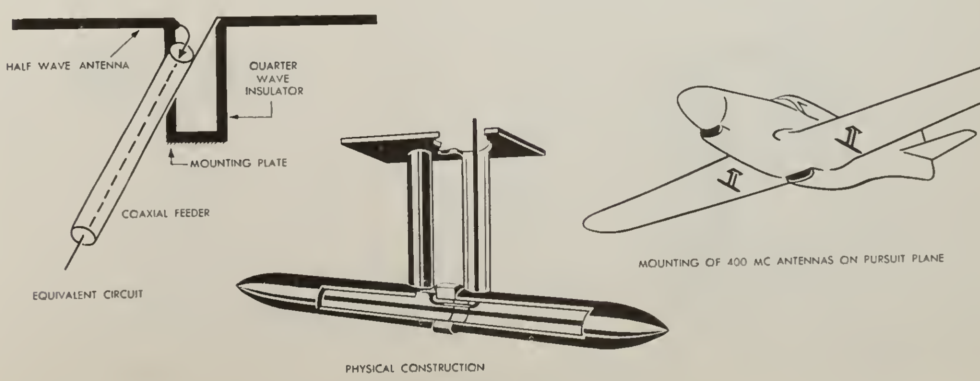
As you can see in the illustration below the antenna is a half-wave dipole constructed of a large diameter tube which is divided at the middle and mounted in the slip stream. As it must be able to withstand air pressures at speeds up to 600 MPH, it is streamlined by rounding at the ends. It is rigidly mounted on a pair of brass tube supports. Each tube is exactly equal to a quarter wavelength and is welded to a plate that mounts on the airplane. Together, the tubes form a quarter-wave section which is shorted at the bottom end and is open at the antenna end. Each half of the antenna connects to a part of the leg of this quarter-wave section. The entire antenna is perfectly insulated by a quarter-wave metallic insulator.

The entire section is connected electrically to the antenna through a coaxial cable. One of the tubes supporting the antenna is the outer conductor for the coaxial cable. A wire through the center of this tube starts from a coaxial connector at the mounting plate and ends at the opposite half of the antenna. Because of the split sections, a bakelite rod inside the antenna tubing is used to give the antenna mechanical strength. The half sections are insulated from one another by a ceramic bushing.

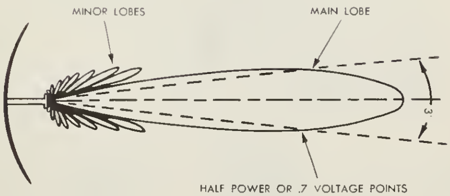
Two antennas are employed with each installation—one for transmitting and the other for receiving. Usually each is mounted under a wing. Usually the part of the aircraft which is located between the antennas serves to attenuate any direct signal radiation between them. The signal desired is the one that goes from the transmitting antenna to the earth and then returns to the receiving antenna. Each antenna is mounted a quarter-wave length from the metal skin of the airplane. The image antenna in the skin is a half wave from the real antenna and excited out of phase. This produces a virtual two-element end-fire array. Since no radiation fields go through the skin, all radiation is directed downward in a somewhat narrowed lobe, as you can see in the side and front view of the radiation pattern shown in the illustration above. The large diameter of the antenna itself makes it broadly resonant and causes its characteristics to be fairly uniform over the 40 mc band width of the equipment.



400 MC Antenna Dimensions and Polar Diagrams



400 MC Antenna

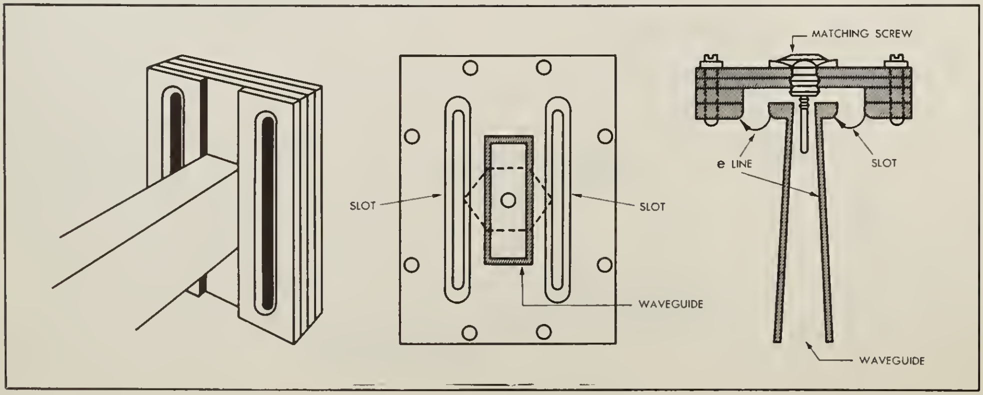


Polar Diagram

A 3-CM ANTENNA SYSTEM USING A PARABOLIC REFLECTOR

A typical example of a set with a 3-cm antenna system using a parabolic reflector is a radar-navigation bombing set. As a navigational aid, this set is required to present a clearly defined picture of the terrain in all directions, both near and far, from the aircraft. This requires high definition in both azimuth and range. Good azimuth definition is provided by a sharply focused parabolic reflector. All directions are covered by rotating the antenna system through 360°. Good range definition is provided by a transmitted signal composed of a super-high frequency carrier and a very short pulse. Presentation of near and far areas simultaneously is accomplished with a special-shaped reflector called a cosecant-squared reflector.

The entire antenna assembly is mounted on a rotating base which is turned by an electric motor. The system uses a 1 x 1/2 inch waveguide for RF plumbing. A rotary waveguide joint, of the type previously shown, brings the signal through the rotating base. When the reflector is tilted vertically, the signal goes through a second rotary joint at the tilt axis. The waveguide is finally terminated in front of the reflector in a Cutler-feed radiator, which in reality is a cavity at the end of the waveguide. It can be considered as two small waveguides formed as a result of dividing the main wave into two halves and bending each half through 180 degrees. The fields emerge from openings or slots that are a half wave apart as shown below. The waveguide is tapered in the non-critical dimension to fit between the slots. The cavity impedance is matched to the waveguide by the matching screw. This adjustment is made at the factory and soldered in position. The radiated field is horizontally polarized. The design of the cavity is such that the energy striking the reflector decreases sinusoidally from the center toward each edge. Upon being reflected, the parabolic shape forms a horizontal radiation pattern which is only 3° wide at the half power points as shown above.

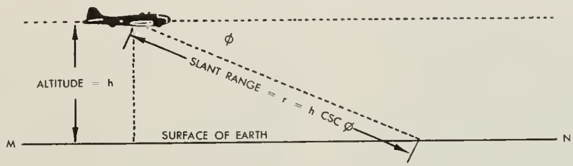


Termination to Cutler-Feed Radiator

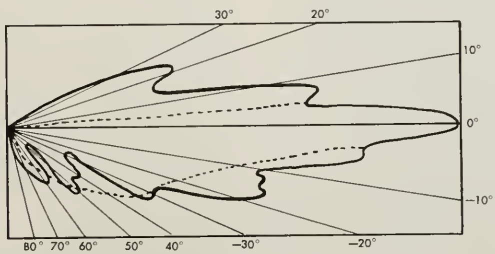
ONE ANTENNA POSITION FOR NEAR OBJECTS

ANOTHER ANTENNA POSITION FOR FAR OBJECTS

A SHARP VERTICAL ANTENNA DOES NOT "SEE" WHOLE FIELD OF INTEREST AT ONCE.



B GEOMETRY OF DEVELOPING COSECANT SQUARED PATTERN



C SPECIAL PATTERN DUE TO COSECANT SQUARED REFLECTOR

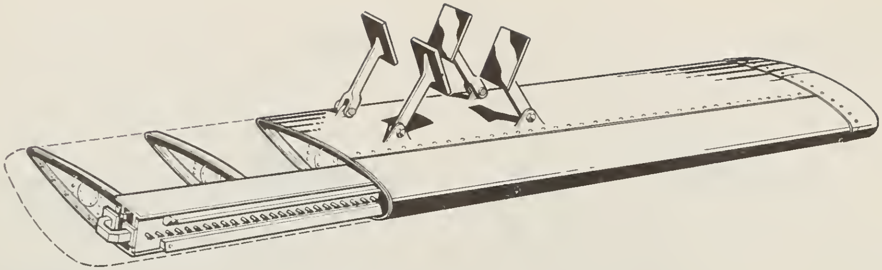
Vertical Coverage of 3 CM Antenna Array

When the parabolic shape is held in the vertical plane, an equally sharp, but undesirable, beam forms like the one illustrated above at A.

As mentioned before, it is necessary to obtain satisfactory indications from objects at different ranges with this set. There are a number of devices on it that make this possible. For example, when the adjustment is set for receiving near objects, the gain of the receiver is too low for the returns from far objects to be visible. This is gotten around by using the specially shaped reflector which spreads the available RF field evenly over the entire usable range.

When the aircraft is airborne, the polar diagram should be a straight line which is parallel to the earth as shown at B. If you designate the distance from the airborne antenna to the point of given electric field intensity as r, then this distance varies as the cosecant of the angle ϕ. This produces a series of points of equal field intensity, all of which lie on the straight line MN. If the electric field intensity varies as the cosecant of the angle, then the *power* density pattern will vary as the *square* of the cosecant of ϕ. This gives rise to the name *cosecant-squared antenna*. The special *skirt* added to the antenna produces the pattern illustrated above at C.

Usually, the antenna assembly is mounted below the fuselage on airplanes where the radiation is unobstructed in all directions. It is protected from the slipstream by a streamlined housing called the *radome*. The radome is constructed of material which causes low attenuation to fields which pass through it. Inferior radome materials will cause reflections from the inside surface. These radiations set up standing waves all the way back to the magnetron and cause it to change frequency. In addition, the reflections change as the antenna rotates.



Cutaway View of 3 CM Collinear Array

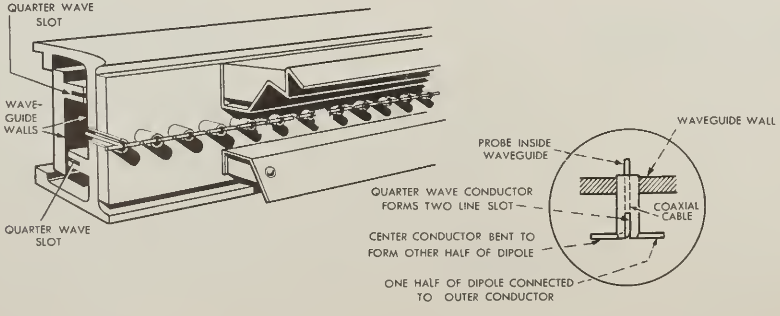
A 3-CM ANTENNA SYSTEM USING A COLLINEAR ARRAY

Another radar set, which has many of the same features as the preceding set,uses a 3 cm antenna system with a collinear array. It obtains superior azimuth definition by virtue of its improved antenna system. This set requires high azimuth and range definition for area bombing. In it high range definition is provided by a high-frequency short-duration pulse. Azimuth definition is provided by an antenna with a pattern only .4° wide.

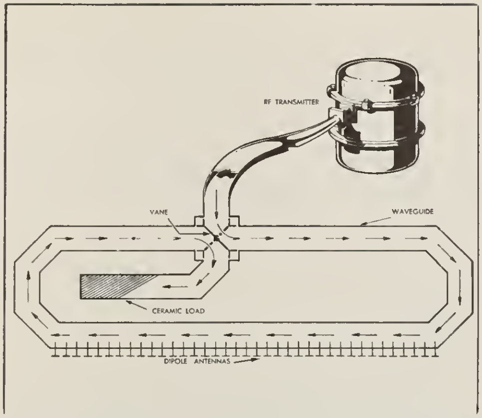
For transmitting energy from the transmitter to the antenna, this set uses 3 cm 1″ x ½″ waveguide plumbing.

Although the three centimeter antenna itself looks like a small *wing*, it is actually 18 feet long. It is practical as equipment on medium-heavy and heavy bombers. High directivity is obtained with this antenna by using a collinear array with 250 half-wave elements placed end to end. The cutaway view above shows a few of these antennas located at the leading edge of the "wing". The RF can go through the plastic covering of the leading edge. The trailing edge is covered with the usual sheet aluminum. The entire interior of the wing is sealed for pressurization and to permit heating for de-icing.

All 250 of the antennas are mounted on the side of a wave guide as shown below. Each is fed by a coaxial line from a probe in the interior of the waveguide. The center conductor comes out the antenna end of the coaxial line and is bent over to form one-half of the dipole. The other half is connected directly to the side of the coaxial line. The outer conductor is split with a slot a quarter wave deep. In order to insulate the inner and outer conductor, a half of the dipole is connected to one side of the quarter wave slot while the center conductor and the other half of the dipole is soldered to the other side of the slot at point where it passes. The probe associated with each antenna extracts energy from the waveguide or delivers energy to it. Energy enters from one end of the



Detail of Antenna and Waveguide Construction



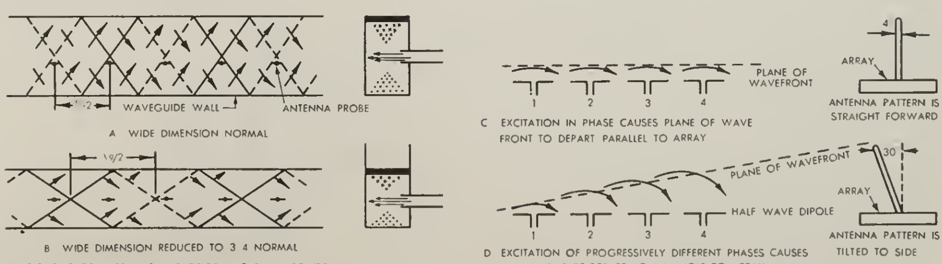
Path of RF Energy in 3 CM Collinear Array

waveguide and proceeds to the other. As each probe is excited, the field becomes weaker. For even excitation of all antennas, the probe depth is very small at the end of the waveguide. Each succeeding probe is inserted a bit deeper into the waveguide. Since energy is sent from either end of the waveguide, the probes are actually short at each end and deep in the center.

It is necessary to excite a collinear array in such a manner that all elements are in phase electrically. This is so that most of the directivity be perpendicular to the antenna. The manner of exciting the antennas is shown in the illustration above and at A below. The power from the transmitter proceeds through a two-way RF switch which diverts the energy to the right or left in the waveguide loop. As is shown, the power flow is to the right, around to the end of the waveguide which has the probes in it. Energy goes through this section, exciting all antennas, then around through a return waveguide, through the RF switch, and into a ceramic block, which absorbs the remaining power with very little reflection.

In passing the probes, the traveling waves are orientated with the probes as shown in the illustration below at A. Here you see the analogy of plane electromagnetic fields being employed. The antennas are spaced so that a half wave of the RF field exists between each antenna. This will excite each probe exactly 180 degrees out of phase. To cause the antennas to be in phase, alternate antenna connections are reversed.

An additional requirement in this system is that the beam direction must be variable so that a wide area may be scanned. As mechanical rotation of the huge "wing" is impractical, its direction is changed electrically. The main waveguide which contains the probes is actually composed of two parts. A stationary part forms one wall and the bottom of the guide, while a movable part forms the opposite wall and the top. The movable part slides vertically, maintaining the distance between walls, but changing the distance between top and bottom. Changing this distance changes the relation between the wavelength of the RF energy and the wide dimension of the rectangular waveguide. A physical gap is present at the junction of the moving and stationary parts, right at a high current point. However, an electrical short circuit exists at the junction by virtue of a quarter wave slot located a quarter wave from the junction. A sort of choke joint is formed by the slot. You will recall from waveguide theory that such a change in the presence of an RF signal of constant frequency will change the wavelength in the waveguide. The change of this dimension is illustrated



Changing the Direction of the Beam

in diagram B. Here the change is from normal to $\frac{3}{4}$ normal. When normal, the antennas are exactly a half wave apart in the waveguide, and each is excited in the same part of the cycle at any time. When the dimension is decreased, the wave is said to travel across the waveguide at a greater angle, which makes the waveguide wavelength longer. In illustration B, the antennas are only .3 wavelength apart. With this relationship, each antenna will be excited in a slightly different phase. Note that the antennas occur at different places in relation to the RF field that is passing at the moment.

The illustration at D shows the effect of this. The curved arrows represent the E-field at the peak of the cycle. Because of the phase difference between antennas, antenna 4 reaches the peak first, then antennas 3, 2, and 1 reach the peak in quick succession. At the instant shown, this part of the electric field is farther away from antenna 4 than from antenna 3. This inclines the plane of the electric field from the combination of antennas, and the greatest field strength will occur in a direction perpendicular to the plane. When all antennas are fed in phase, all electric fields start out simultaneously, causing the plane of the combined E-fields to be parallel to the line of the array.

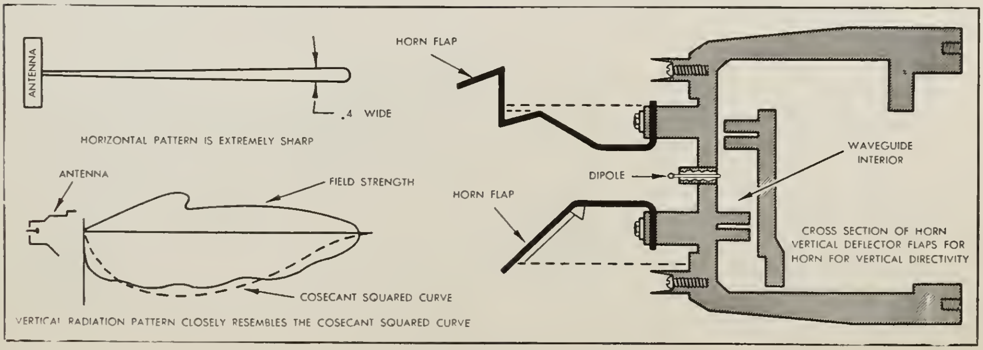
The limit of the phase shift is the point where the waveguide dimension is reduced to cutoff. For optimum transmission the waveguide should be .7 wavelength. This is the dimension for a beam dead ahead. For changing the beam, the dimension is reduced to .525 wavelength. This is the just above cutoff dimension and will divert the beam to an angle of 30° from the perpendicular. So the maximum deviation of the beam from dead ahead is 30°.

The beam is directed to any direction in between these limits by adjusting the size of the waveguide to intermediate values.

The beam is pointed 30° in the other direction by reversing the direction of power flow through the waveguide. If the vane of the RF switch is turned 90°, it will channel the transmitter energy into the other end of the waveguide, through the section with the probes, through the RF switch, to the ceramic load. In passing through in reverse direction when the size of the waveguide is reduced, the antennas are excited in reverse order, tipping the beam the other way.

A powerful 28-volt motor serves to move the waveguide wall up and down. When the beam gets to the dead ahead position, the vane flips over and causes the beam to reverse direction. When the waveguide is moved continually, the beam likewise continuously sweeps to the left and right over a 60° sector. Although a sweep of 60° is far less than the 360° scan possible with the rotating parabola previously described, the sharpness of the beam compensates for this small range in sweep. The indicator employs the p scan (PPI) in which the azimuth is limited to 60° arc.

Another requirement is that the vertical pattern be carefully controlled. A horn type radiator as is shown below, is effected by adding the flap parallel to the waveguide. Notice the cross section view of the flap. As with the previous set, it is necessary in this set that far and near objects be visible at the same time. With this set vertical adjustment of the antenna is not possible. This makes control of the vertical pattern extremely important.



Radiation Pattern

CHAPTER 13

Selsyns and Servomechanisms

Up to now, the manual has analyzed a great variety of circuits employed in radar in general, and in a number of radar sets in particular. The purpose of this chapter is to acquaint you with devices which transmit position between these circuits and remotely located gages, dials, and controls. The chapter discusses the basic theory of two of these devices—the selsyn, a machine which converts mechanical position into electrical position, or electrical position into mechanical position; and the servomechanism, a system which amplifies and transmits mechanical position from one position to another by electrical means.

USES

There are many operations in radar where it is desirable for two shafts to rotate in synchronization. But because the distance between the two shafts is too great, or because one shaft cannot develop enough torque to rotate the other one, it is not always feasible for the shafts to be connected together mechanically. Therefore, shafts which are designed to rotate in synchronization are most generally connected electrically.

Systems employing electricity for rotating two shafts in synchronization are called *remote indication* or *remote control* systems. Where it is necessary that the torque of the shaft which controls the operation (the *control* shaft) be amplified before it is applied to the shaft which is connected to the load (the *load* shaft), the system then is called a *servomechanism* or *servo* system.

As you can see by the name, a remote indication system is designed to produce either an indication at one position of an operation at another position, or to transfer information from one position to another. Sometimes this system is called a *data transmission system*. Many radar antennas, for example, are designed to rotate 360° but are located in places where it is not possible for the operator to see them. Therefore a data transmission system is employed to give the operator the direction in which the antenna is pointed. A pointer attached to a shaft rotates in synchronization with a shaft connected to the antenna, and registers the position of the antenna on a dial at the operator's position. The devices through which the position of the antenna is transmitted to the dial are called *selsyns*. They resemble small electric motors. The selsyn at the antenna is called a selsyn *generator* or transmitter; the one at the operator's position is called the selsyn *motor* or *receiver*.

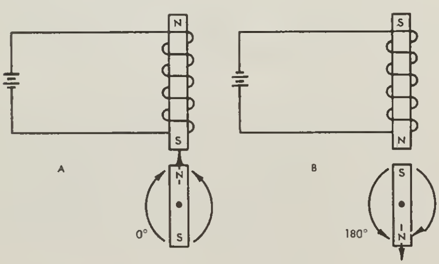
The word selsyn, as you can see, is a contraction of the words, *self synchronous*. It is a trade name which has been so widely used that it is applied to all remote indication devices regardless of who manufactures them. Other names frequently used for these devices are synchro, autosyn, and synchrotie.

In the remote indication system described, it is possible for the operator to rotate the antenna itself from his position. However, this requires more torque than can be transmitted efficiently through the system. To provide the necessary torque a servo system must be used. In the servo system, the operator turns a hand wheel (or some other control). This moves the shaft connected to the hand wheel out of step with the shaft connected to the antenna. The out-of-step condition produces a voltage which is called the error voltage. This voltage is amplified by the torque amplifier and is applied to a motor which in turn rotates the antenna. The motor rotates in a direction which counterbalances the error produced by the out of step relationship between the hand shaft and the antenna shaft, and thus causes the antenna to take the same position as established by the hand control.

POSITIONING A SHAFT WITH DC VOLTAGES

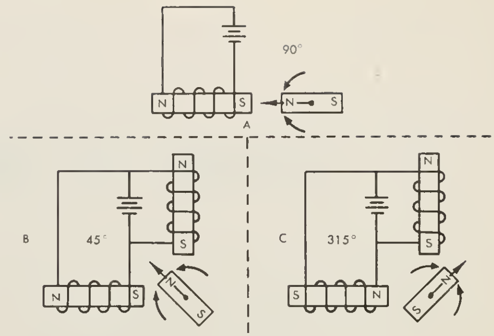
Although the selsyn is an AC device, DC voltages are sometimes used for positioning shafts. Since an understanding of shaft positioning with DC will help you to analyze the operation of the selsyn, it is well to take it up first.

There are a number of ways for positioning a shaft with DC voltages. One is to mount a bar magnet on the shaft and to locate it near a solenoid (coil). When a voltage is applied to the coil, the shaft will rotate to either of the two positions shown in the illustration directly below. When you reverse the polarity of the DC voltage, the magnet will likewise reverse its position. The position of the magnet shown is arbitrarily designed the 0° position for facilitating description of this position. Furthermore, the positive angles represent rotation in the counterclockwise direction. In examining this arrangement closely, you can see that only two positions of rotation are possible. Therefore it would not be very satisfactory as a remote indicating system.



Using One Coil to Position a Magnet

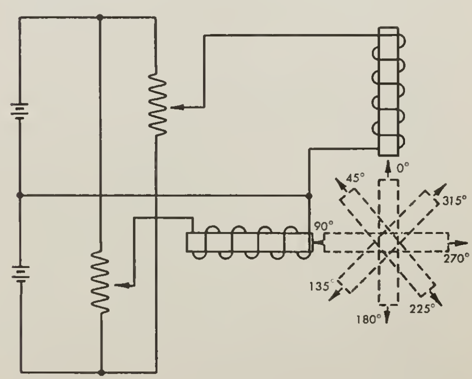
Another system, capable of rotating a shaft to more positions than a system using one coil, employs two coils which are placed at right angles about the magnet. When a single coil is placed as shown at the upper right in diagram A, the magnet will rotate to the 90° position, and when the polarity of the coil is reversed, the magnet will rotate to the opposite or 270° position. If, however, you connect two identical coils at right angles into the circuit (B), the magnet will rotate to the 45° position. If you reverse the polarity of both coils, then the magnet will reverse its position and come to rest at the 225° position. Reversing the



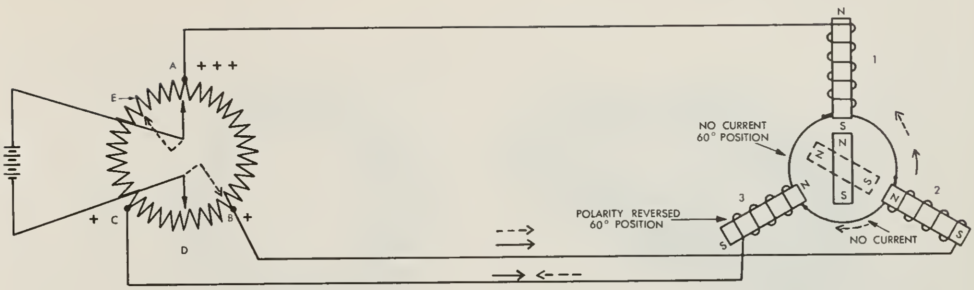
Using Two Coils Placed at Right Angles

polarity of the coil in the horizontal position causes the magnet to assume the 315° position shown at (C). If the polarity of both coils at (C) is reversed, the magnet will take a 135° position. Thus, you can see that by applying the same voltage to one or both of two coils which are placed at right angles, and by reversing the polarity of the coils, the magnet will take any one of eight positions.

A further method of positioning a shaft with DC as shown below involves using a potentiometer for controlling the polarity and magnitude of the voltage applied to each coil. Since the potentiometer makes it possible to vary the magnitude of the voltage, the magnet will take any position around the circle. While this arrangement produces accurate results, it is not satisfactory as a practical remote indicating device since it requires independently controlled potentiometers for each coil.



Two Coils at Right Angles and Potentiometers to Vary Polarity and Magnitude of DC Voltage



Potentiometer and Magnet Remote Indicator

POTENTIOMETER AND MAGNET REMOTE INDICATING SYSTEM

A practical DC remote indicating system which uses the principles described in positioning shafts with DC is the Potentiometer and Magnet Indicating System as shown in the illustration directly above. Sometimes this system is called a DC selsyn.

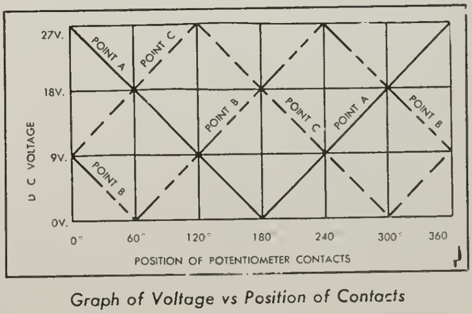
The potentiometer in this remote indicating system is circular. DC voltage is applied to it through contacts which slide along the inner side of the potentiometer. These contacts are insulated from each other to prevent the DC voltage from being shorted out. In the three-coil permanent magnet arrangement at the right of the diagram, the coils may be connected either in delta or in wye. In either connection, there is little difference in operation of the system.

The potentiometer is step-wound to compensate for the shunting effect between the potentiometer coils. This type of winding, however, causes the voltages at points A, B, and C to vary as the potentiometer rotates in a counterclockwise direction, as shown to the right in the graph of the voltages at different positions of the potentiometer contacts. All are with respect to the negative (ground) side of the DC voltage supply.

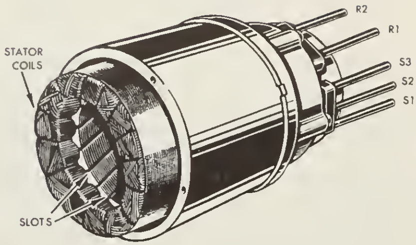
As previously mentioned, there is little difference in the operation of a wye- and a delta-connected system. Thus, if you follow through the theory of operation of the wye-connected type of three-coil-permanent magnet assembly, you will be able to understand the delta connected type also.

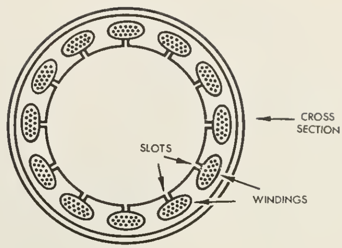
During operation of the wye-connected system when the contacts are at the 0° position (+ to A and − to D), the voltages at A, B, and C respectively are 27v, 9v, and 9v. One end of coil 1 is connected to point A; one end of coil 2 is connected to B; and one end of coil 3 to C. The other ends of each coil are connected together. Current flows from B to A through coils 2 and 1, and from C to A through coils 3 and 1. Since B and C are at the same potential, there is no current flow between them. Coils 1, 2, and 3 become magnetized, with their poles as shown in the diagram above. Coil 1 carries more current flow and therefore exerts a strong attraction for the north pole of the magnet. Coils 2 and 3 carry equal currents and attract the south pole of the magnet with equal force. The magnet will therefore assume the position shown (the 0° position).

When the potentiometer contacts move to the 60° position, then E is at 27v, and B at 0v. As you can see in the graph below, A is at 18v, and C at 18v. This causes current to flow from B to A through coils 2 and 1 and from B to C through coils 2 and 3. No current flows from A to C. In this condition, coil 2 now carries more current than the other coils and strongly attracts the south pole. The current in coil 3 is reversed. Its polarity is also reversed. Thus, coils 1 and 3 attract the north pole with equal force, and the magnet assumes the 60° position corresponding to the potentiometer contacts.



Graph of Voltage vs Position of Contacts





The Stator of a Selsyn

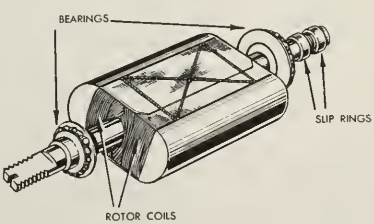
As the potentiometer contacts move from the 0° position to the 60° position, the voltages vary uniformly as follows: A from 27 to 18v; B from 9 to 0v; and C from 9 to 18v. Hence the current through coil 1 decreases, that through coil 2 increases, and that through coil 3 decreases to zero at 30° and then increases in the opposite direction. The rotation of the magnet is uniform and follows the rotation of the contacts.

Since the smooth variation of voltages is upset if the coils draw much current, this system is limited to applications where the current in the coils is not appreciable when compared with the current in the potentiometer. Since the potentiometer current is not useful except for maintaining the voltages, the system wastes considerable power. Its principal use is as an indicating device, which tells the operator the angle of elevation of the antenna. It may be used to indicate the azimuth to the operator, but he usually gets this information from the C.R.T.

PHYSICAL CHARACTERISTICS OF SELSYNS

Before going into the operation of the selsyn generator and motor, it is well to consider first the construction of each. Both resemble small electric motors and are identical except for several details. Each unit contains a fixed element called the *stator* and a movable element called the *rotor*. The stator consists of a number of coils which are placed in slots around the inside of a laminated iron field structure, very much like an ordinary AC motor. These coils are divided into three groups, spaced 120° apart around the inside of the field. Actually the groups overlap somewhat so that the attractive force tending to pull the rotor into position is the same for all positions of the rotor. The 120° spacing of windings sometimes leads a person to believe that three-phase voltages and currents are used, but such is not the case. *Only single phase voltage and current are present.*

In a typical selsyn generator, the rotor consists of a single coil of wire wound on a soft iron core and mounted on a shaft in a way that the axis of the coil is perpendicular to the shaft. The ends of the core are curved so that the air gap between them and the stator is small and uniform. To make friction low, the shaft is mounted on ball bearing mounts. The rotor turns inside the stator and is electrically connected to the rotor through two slip rings on the shaft.



The Rotor of a Selsyn

Because a selsyn motor is similar to an ordinary AC motor, it tends either to oscillate violently or to spin continuously under some conditions. This is particularly likely to happen when the shaft is turned suddenly, as for example, when power is first applied to the system. To prevent this undesirable oscillation, there is a heavy metal flywheel, called an *inertia damper*, mounted on one end of the shaft. This flywheel is mounted so that it turns freely on the shaft for 45° or so, and then runs into a keyed bushing.

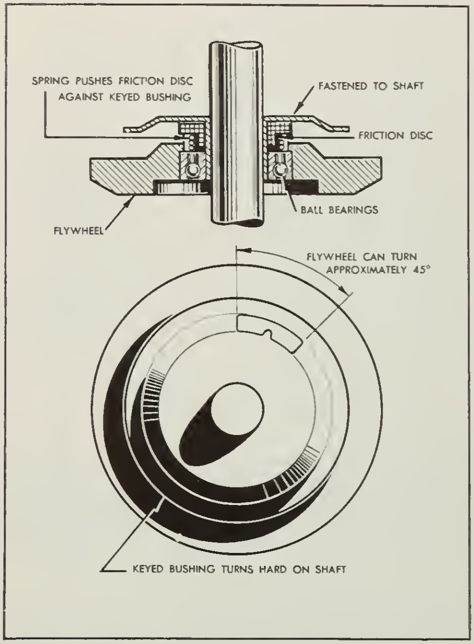
This bushing which is fastened to the shaft through a friction disc also turns on the shaft but with a great deal of friction. For slow changes in position of the shaft, the flywheel follows along without much oscillatory effect. But if the shaft tries to turn suddenly, the flywheel tends to stand still, and the friction disc acts as a brake to slow down the motion of the shaft. This keeps the shaft from reaching a speed fast enough to start oscillating or spinning. If oscillation or spinning does occur, you usually can be certain that there is something wrong with the damper.

In the selsyn, the leads to the rotor and stator are brought out through two insulating strips at the back of the motor or generator. The stator connections are marked S1, S2, and S3; and the rotor connections R1 and R2 as shown in the stator illustration. This system of marking connections is standard among the various manufacturers.

OPERATION OF SELSYN SYSTEM

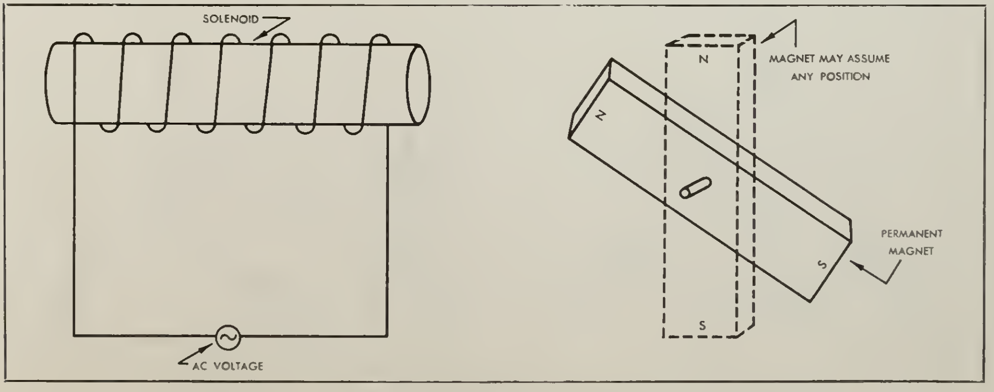
Positioning a Shaft with AC Voltages

When a 60 cps voltage is applied to a solenoid located close to a permanent magnet that is mounted so that it can rotate on an axis as shown below, the torque exerted on the magnet will reverse direction 120 times per second, first in one direction; then in the opposite direction. Since the magnet cannot reverse direction this rapidly, it will respond only to the average torque and since the average torque is zero, the magnet will assume some position independent of action by the coil. Thus since the magnet is not affected by the coil, it is obvious that it cannot be used with an AC coil as it can be with a DC coil.

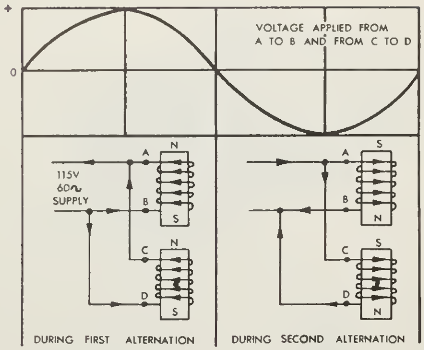


Typical Inertia Damper

But when you substitute another AC coil for the permanent magnet, connected as in the illustration at the top of the next page, the current in both solenoids reverses at the same instant. This causes the same end of the rotating solenoid to be attracted to the nearest pole of the fixed solenoid throughout the entire cycle. As shown, the C-end of the rotating solenoid is attracted to the B-end of the fixed solenoid at



Effect of AC Solenoid on Magnet



Effect of One AC Coil on Another

all times. Reversing the connections on either, but not both coils, reverses the rotating coil and causes B to attract D. This reversal of connections is equivalent to a reversal of phase in that coil.

Symbols for Selsyn Diagrams

It is not very practical to show selsyn circuits in the form of photographs or drawings of the actual apparatus, since they merely show the outer appearance and leave the inside obscure. Therefore, to make it easier to describe the operation of a selsyn motor (or generator), standard diagrams which are referred to as *selsyn diagrams* are used. In forming these diagrams, first, there is a standard way for describing the shaft position. This is necessary to insure that if the rotor shaft is properly connected mechanically in the system, it will turn to the correct position when it is connected electrically to a selsyn generator.

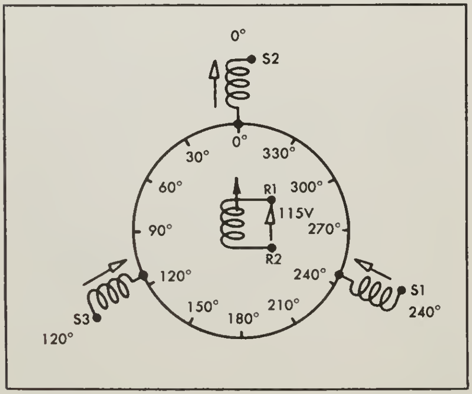
In the diagram below, the position which the rotor takes when it lines up with the stator coil which is connected to S2 is called the *electrical zero* position. Other positions are measured in degrees, assuming that you are looking at the shaft end of the unit and that the shaft turns counterclockwise for an increasing number of degrees.

In studying the selsyn diagram, think of the rotor as turning around an axis in the center of the diagram and that the arrow on the R1 end of the rotor points at the number which indicates the electrical position of the rotor.

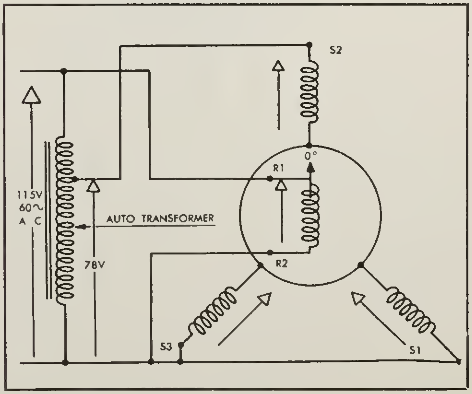
Another point to consider is that the rotor coil in the diagram is assumed to be wound in the same direction, going from R1 to R2, and each of the stator coils is wound from the outside to inside.

The open-headed arrows are important in that they indicate the following *relative phase* polarity: If two arrows point in the same direction, the voltages are in phase; if the arrows point in opposite directions, they are 180° out of phase.

All voltages are either in phase or 180° out of phase with each other. The diagram indicates that the voltages from R2 to R1, from common to S2, from S1 to common, and from S3 to common are in phase, while from S2 to common they are 180° out of phase with either S1 to common or S3 to common.



A Selsyn Diagram

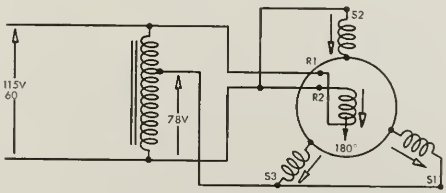


The 0° Position

Voltages Required to Position a Selsyn Motor

In order for a standard selsyn motor to turn to the electrical zero position, 115v AC (R.M.S. value) must be applied to the rotor leads, zero voltage between S1 and S3, 78v between S2 and S1-S3 in such a way that the voltage from R1 to R2 is as shown in the preceding illustration. In this condition, the coils connected to S1 and S3 exert equal attractive forces on the R2 end of the rotor. The coil connected to S2 attracts the R1 end of the rotor with greater force, for it carries the current of both the other coils. Therefore this causes the rotor to assume the 0° position.

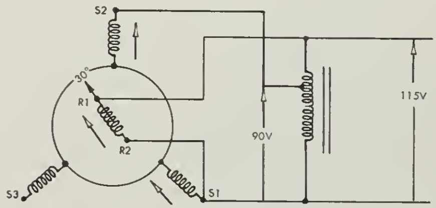
Reversing the leads to the 78v supply causes the voltages across the stator coils to reverse in phase and the rotor coil to reverse in position, causing it to point to the 180° position below.



The 180° Position

Connecting together the other pairs of stator leads and applying the 78v to the third lead either in phase or out of phase with R1 and R2 voltage will cause the rotor to turn to the four positions—60°, 120°, 240°, 300°—as shown directly below.

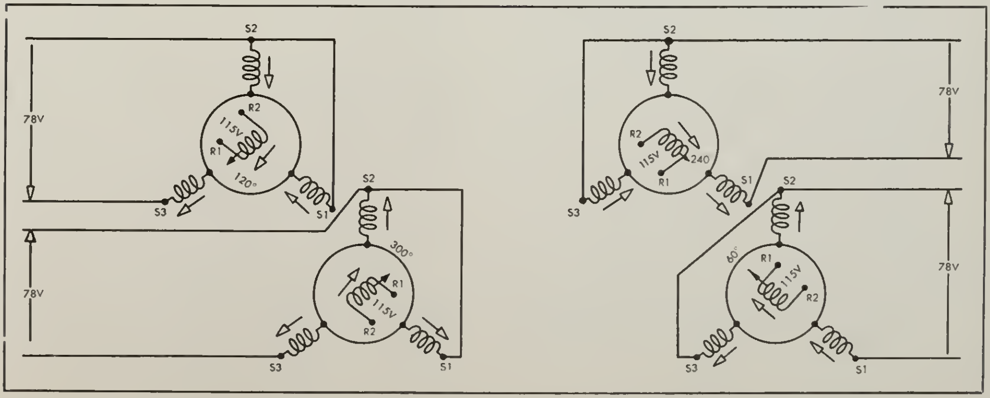
When no voltage is applied to S3 and 90v is applied between the terminals, S1 and S2 (90v are required to produce the same attraction as before), the rotor will come to rest in the 30° position, since the R1 end of the rotor is attracted to the coil connected to S2, and the R2 end to the S1 coil with equal force, as you can see in the illustration directly below. Reversing the phase polarity of the 90v causes the rotor to turn to the 210° position. This leaves S1 or S2 open. Applying the 90v to the remaining two terminals either in phase or 180° out of phase produces four other positions shown on the next page.



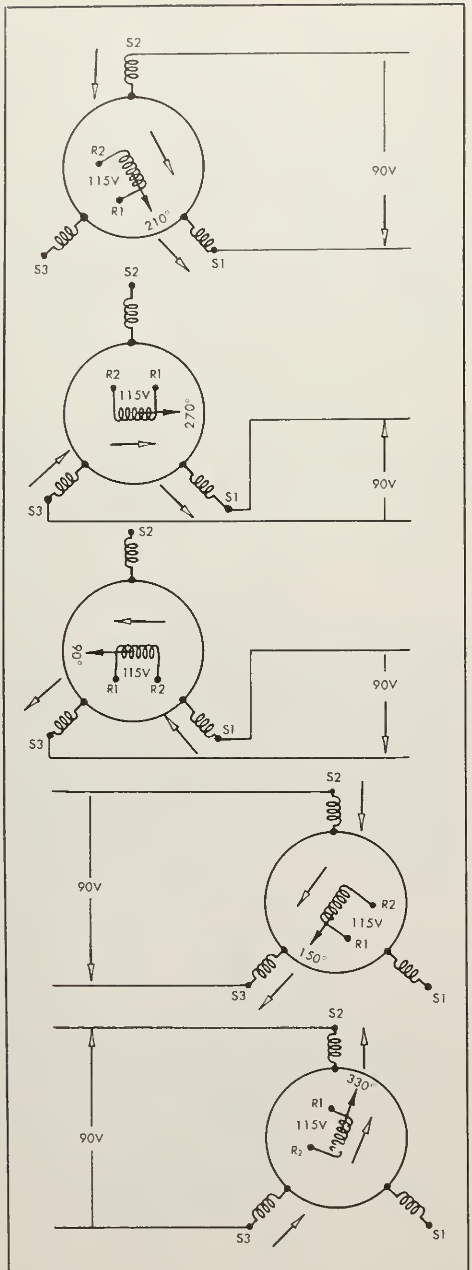
The 30° Position

So far, only the means of obtaining 12 positions, spaced 30° apart, has been described. Next consider methods for obtaining position at intermediate points between these positions.

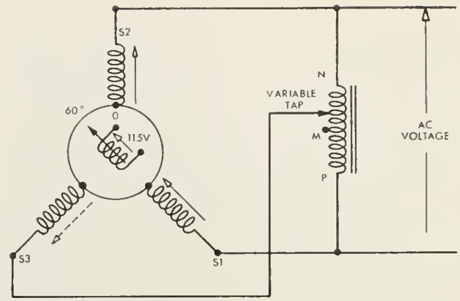
The illustration at the top of page 13-8 shows a set-up which gives positioning at any point from 0° to 60°. In it, fixed AC voltages are applied to S1 and S2, and the voltage at S3 is variable from the voltage at S1 to the voltage at S2 by means of a variable auto transformer.



Four Positions Possible with Two Leads Connected



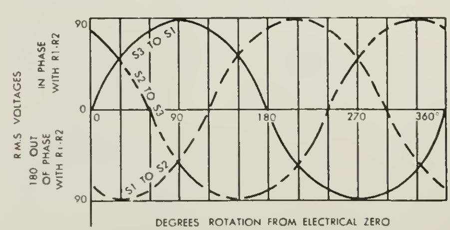
Other Positions with One Lead Open



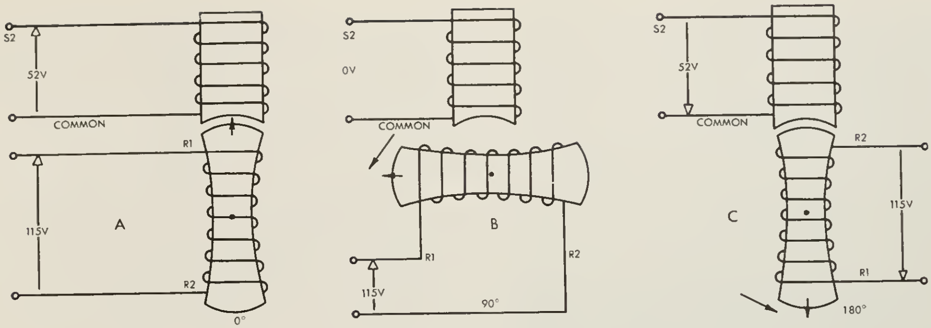
Using Variable Tap to Obtain Positions from 0 to 120

If the voltage at S3 is the same as at S2, the rotor will turn to the 60° position as shown in the illustration at the bottom of page 13-7. If the voltage at S3 is equal to the voltage at S1, the rotor turns to the 0° position as indicated in the diagram on page 13-6. These two conditions occur when the variable tap is at N and P respectively. When the tap is at M, midway between N and P, S3 is at zero potential with respect to the common connection of the coils and exerts no attraction or repulsion. Hence the rotor turns to the 30° position just as it did when S3 was disconnected. At intermediate points between M and N, the phase of the voltage in coil S3 will be as indicated by the dotted phase arrow and the rotor will assume a position between 30° and 60°, depending upon the magnitude of the voltage at S3. Between N and P the rotor turns to some point between 30° and 0° which depends on the magnitude of the voltage at S3, the phase polarity of which is opposite to that indicated by the arrow.

By using stator voltages intermediate to the voltages for the 12 basic positions, it is possible for the rotor to assume any position around the circle. The voltages necessary to produce any angular position may be plotted as in the graph below. The three curves show the voltages be-



Voltage vs Rotor Positions



Transformer Action

tween pairs of stator leads. The voltages above the zero axis are in phase with the voltage from R1 to R2, and those below the axis are opposite in phase. Keep in mind that the curve of the S2 to S1 voltages will be 180° out of phase with the curve shown for the S1 to S2 voltage.

Next suppose you check these voltages against the voltages in some of the cases previously discussed. Notice that when the rotor turns to the 0° position, the voltage from S3 to S1 is 0v, and the other two are 78v. This is the same situation shown in the 0° position illustrated on page 13-6. When the voltage between S1 and S2 is increased to 90v and each of the others is changed to 45v the shaft turns to the 30° position as you can see in the illustration of the 30° position on page 13-7. Note that S3 being open is at a common potential, that is, half way between S1 and S2; hence S3 to S1 and S2 to S3 fall to 45 volts. When S3 to S1 and S1 to S2 go to 78v, the shaft turns to 60°; and so on. At each position of the rotor, there is a certain definite value and phase condition for each of the three voltages.

Due to the method of arranging the rotor and stator windings on a standard selsyn, the curves shown in the graph are in the shape of sine curves. In other words, the curves look like time graphs of sinusoidal voltages. They show something entirely different, however, for time is not involved in any way in the preceding graph. Another thing to remember is that while they have the appearance of the voltages in three phase, they are really single phase. What is actually shown in the curves is the effective values of the three AC voltages plotted against position of the rotor and whether the voltages are in phase with or 180° out of phase with the rotor voltage.

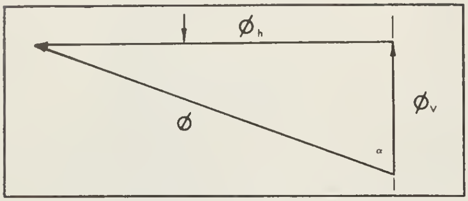
Transformer Action of Selsyn

Another important factor in selsyn operation is transformer action. This deals with the voltages induced in the stator windings when an AC voltage is applied at various positions of the rotor. In studying this action, assume that the stator leads are open at first and then refer to the illustration above showing one stator coil with the rotor in three positions. At (A), the axis of the rotor and stator are aligned and the flux produced in the rotor due to the 115v applied to it induces a voltage in the stator coil. Due to the turns ratio this induced voltage in a standard selsyn is 52 volts. This voltage is in phase with the rotor voltage as indicated by the open headed arrows.

In (B), the rotor is in the 90° position. In this position, the flux set up in the rotor is not effective in inducing a voltage in the stator because the flux that passes through the coil is at right angles to its axis and induces equal and opposite voltages on the opposite sides of each turn. Thus at this position, the net voltage is zero.

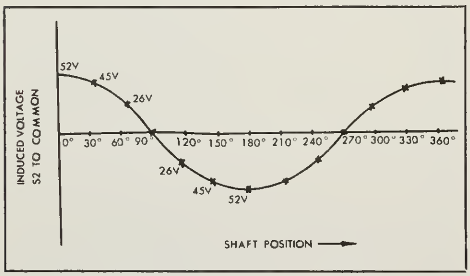
In (C), the rotor is in the 180° position. Here the flux at any instant is 180° out of phase with what it would be in the 0° position. Thus the voltage induced in the stator coil is 180° out of phase with the rotor voltage. The magnitude is the same as for the 0° position, 52 volts.

At intermediate points, the magnitude and relative phase polarity depends upon the flux threading the coil. If you resolve the flux ϕ into the components ϕh and ϕv, as shown in the next illustration, you can determine the relation between the flux and the induced voltage. The horizontal component is ineffective, while the vertical component of flux induces a voltage in the coil. But $\phi v / \phi$ is the cosine of angle a between the axis of the stator and rotor.



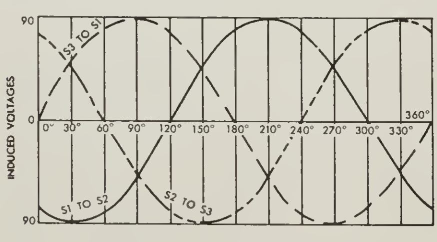
Components of Flux φ

Thus you see that for any angle of rotation the voltage induced in the stator coil S2 is 52 cos α volts. (Below is the graph of these voltages vs shaft position.) Due to the 120° spacing of the stator coils, maximum voltage is induced in S3 at the 120° position and in S1 at the 240° position.



Induced Voltage vs Shaft Position

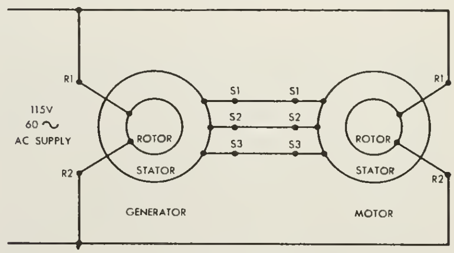
Since the connections are made to the points S1, S2 and S3, the voltages appearing between those points are of special interest. These voltages are the sums or differences between the voltages induced in the individual stator coils, depending on whether the voltages are in phase or 180° out of phase. For example, the voltage from S1 to S2 is the sum of the S1 to common voltage and the common to S2 voltage. Below is shown the magnitude of the voltages between stator connections for any position of the rotor. In comparing the voltage shown in this illustration and those shown at the bottom of page 13-8, note that the voltages induced in the stator coils are the same as the voltages required to position the rotor to the same angle.



Induced Voltage vs Rotor Position

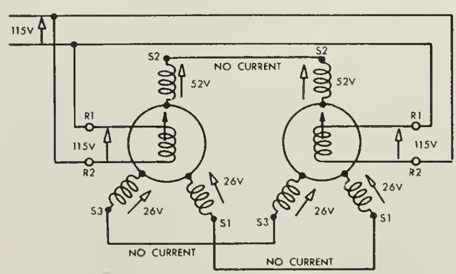
How a Selsyn Motor Follows a Generator

The selsyn system in its simplest form consists of one motor connected to one generator as shown below. With this connection, whenever the shaft of the generator turns, the shaft of the motor turns such that its *electrical position*



Selsyn System Connections

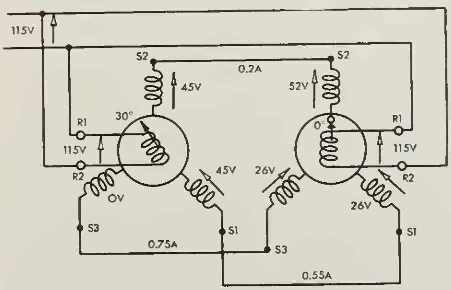
is the same as that of the shaft of the generator. Thus, for example, when the generator shaft turns to electrical zero, the motor shaft turns to 0°, or when the generator shaft turns to 30°, the motor shaft turns to 30°, and so on. In order to see how the motor shaft follows the generator shaft, first examine the internal conditions of each selsyn. To do this, consider both shafts at the same position and both shafts at different positions.



Generator and Motor Both at 0° Position

Both shafts at the same position. Above both shafts are turned to the 0° position. In this position there will be voltages induced in the stator windings of both generator and motor, with the phase polarity and magnitude indicated.

Note that the voltages of the motor are equal to those of the generator, but because of the method of connection, the voltages in one oppose the voltages in the other. This causes no current to flow in the stator coils. Hence, there is no magnetic field set up by them and thus there is no attraction or repulsion to cause either rotor to turn. The rotors will remain in the 0° position indefinitely if no change in outside conditions occurs. The same situation exists if the two rotors are at any other position, except that the magnitudes of the induced voltages vary in accordance with the curves shown in the induced voltage versus rotor position graph on page 13-10.

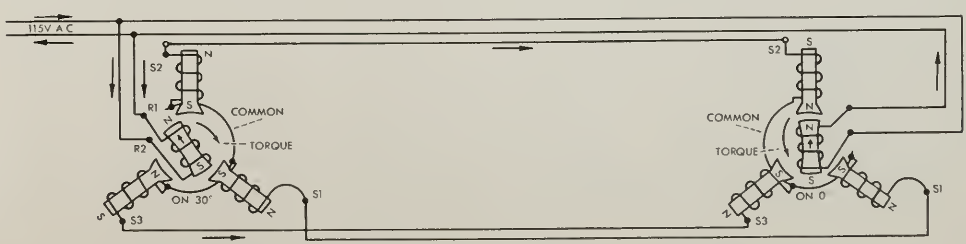


Shafts in Different Positions

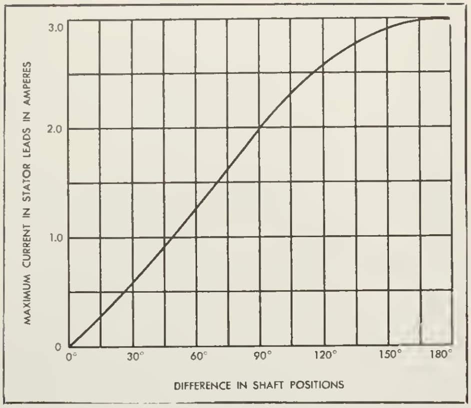
Shafts at different positions. When the two shafts occupy different positions, the voltages induced in the stator windings are not balanced as they are when the shafts occupy the same position. Suppose, for example, that the generator rotor is at the 30° position and that the motor rotor is at the 0° position. By referring to the illustration on page 13-10 showing rotor position and the illustration just above showing the shaft at different positions, you can determine the value of the induced voltages. Note that there is no balance among the voltages and that currents flow in all three stator leads and all stator coils of both generator and motor. The currents are greatest in the circuits where the voltage unbalance is greatest, that is, in the circuit composed of S3 coils and common. The effect of these currents is to produce magnetic fields which tend to turn the motor shaft to the same position as the generator shaft. Thus if the motor shaft is free to turn, it will take the 30° position. As the motor shaft turns, the misalignment decreases and in turn the unbalance of voltages decreases and the currents produced by them decrease. When the shaft reaches the 30° position, the same conditions that prevailed when both shafts occupied the same position now prevail.

A torque also is set up in the generator tending to turn it in the clockwise direction, resulting in both rotors tending to align themselves. In the usual system the generator shaft is attached to some device so that it is not free to rotate.

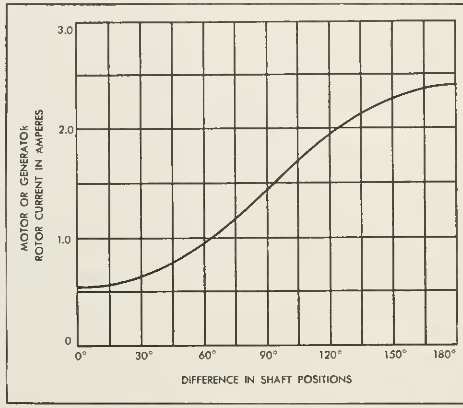
To see more clearly the magnetic effects producing the torques, influencing motor operation, note below the diagram of the conditions in the circuit *at a given instant*. Keep in mind the same conditions do not hold except for that particular instant. Examination of the diagram shows the following action: (1) that the lower end of the S2 coil of the generator attracts the R1 end of the rotor, (2) that the S end of the S1 coil repels the R2 end of the rotor and (3) that the N end of the S3 coil attracts the R2 end of the rotor. All these forces tend to cause rotation in the clockwise direction. In the motor the magnetic poles of the stator coils are the reverse of those of the generator. Consequently, the forces of attraction and repulsion are reversed and the torque is counterclockwise. At one-half cycle (of the input voltage) later, the polarities of the magnets and the directions of current flow will reverse. The forces of attraction and repulsion do not change, however, since the polarities of the stator and rotor magnets reverse at the same instant.



Instantaneous Picture of Forces Set Up when Shafts are in Different Positions



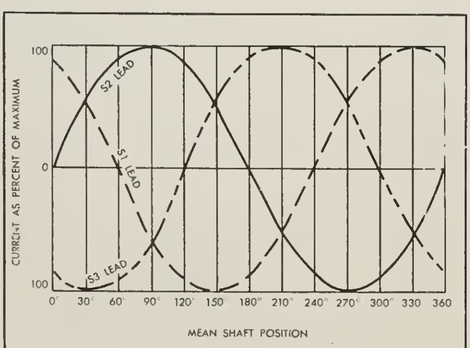
Stator Current vs Difference in Shaft Positions



Rotor Current vs Difference in Shaft Positions

How the Currents Flowing in a Selsyn System Depend upon Shaft Positions

As was shown, current flows in the stator circuits whenever the two shafts occupy different positions. In any stator lead the amount of this current flow depends on the difference between the voltage induced in the two coils to which that lead connects. Suppose that the current in any one stator lead (S2 for example) is measured and that the actual shaft positions are adjusted so that this current is maximum for each difference of the shaft positions. From the above graph showing this maximum possible current versus the difference in shaft positions, you see that currents as high as 3 amperes are possible if the shafts are held in positions differing by 180°. However, in practical operation the shafts are seldom more than a degree apart, since the motor shaft starts turning immediately when the generator shaft starts to rotate. Hence the current in any lead seldom becomes more than one-tenth of an ampere. The current in one stator lead depends not only on the difference of shaft positions, but also on the positions themselves. The curves on this page showing the percent of maximum current in each lead for the various mean shaft positions (average positions between the two shaft positions) completes the representation.



Stator Current vs Mean Shaft Position

To illustrate how these two diagrams may be used in conjunction, take the case where the generator shaft is at 60° and the motor shaft is at 120°, a difference of 60°. By referring to the difference in shaft position diagram, you can see that the maximum possible current in any one lead is 1.25 amperes. Since the mean shaft position is 90°, by referring to the mean shaft position chart you see what percentage of this maximum current flows in each lead. The S2 lead carries 100% or 1.25 amperes, and the S1 and S3 leads carry 50% or 0.625 amperes. The fact that S1 and S3 carry current below the zero axis shows that the phase is opposite from that flowing in the S2 lead. Remember that currents are usually not high since small angular differences are the usual or normal condition.

Since the selsyn generator or motor acts like a transformer, an increase in stator current causes a corresponding increase in rotor current. (See the above diagram of rotor current versus difference in shaft position). Although all three stator currents are zero when the shafts are aligned, the rotor current is still approximately 0.5 ampere.

As in any transformer the primary draws some current with no load on the secondary. This current produces magnetization of the rotor and supplies its losses.

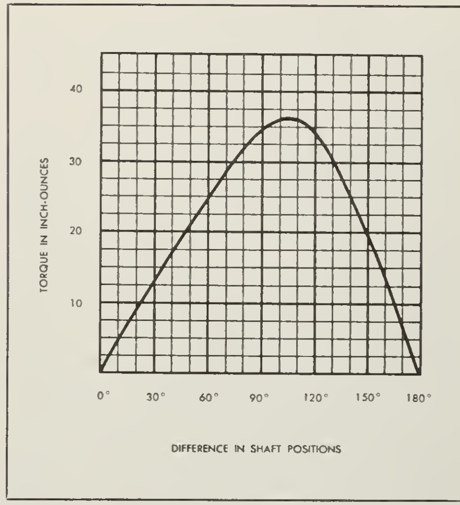
The following statements summarize the factors concerning the relative shaft positions:

1. Reversing the rotor leads on either motor or generator causes a 180° difference in shaft positions.

2. Interchanging the S1 and S3 leads on either the generator or motor causes rotation of the motor shaft in the opposite direction to that of the generator.

3. Shifting the stator leads so that S1, S2, S3 of the generator connect to S2, S3, S1 of the motor produces a fixed difference of the 120° in shaft positions.

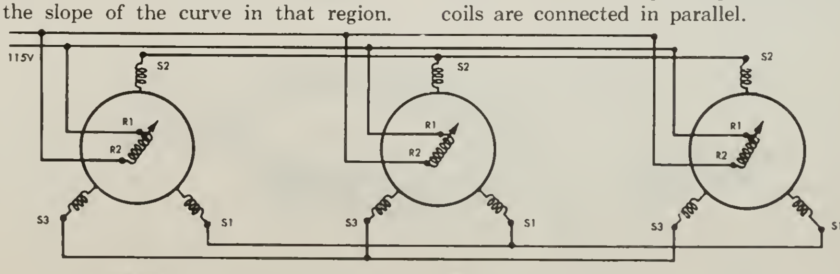
Since both the selsyn motor and the load attached to it have some friction and inertia, some torque is required both during motion and during acceleration. Since no torque is developed when the two rotor positions coincide, perfect accuracy is impossible. Some angular displacement is necessary to overcome friction and inertia. This lag is not appreciable, in most cases being of the order of a fraction of one degree. To minimize the inaccuracy, the friction of the motor must be as low as possible and the torque gradient (torque per degree displacement) must be as high as practicable. The torque gradient depends upon the internal impedance of its stator coils. Hence both the motor and generator must be large enough, for if either is too small, the higher impedance of the windings will reduce the current for small unbalances of voltages and the torque will accordingly be less. Note in the above graph of the torque versus angular displacement relations for a typical selsyn that for small angles the graph is practically a straight line. The term *torque gradient* is simply the slope of the curve in that region.



Torque vs Displacement

Maximum torque is developed at approximately 100°. The graph shows zero torque at 180° displacement. This is true because the powerful forces that are exerted are in a state of balance. This is, therefore, an unstable condition and any small change in either rotor position results in a return to the zero displacement condition.

One generator is sometimes used to transmit data to several motors at various locations, for example, the generator connected to the gyrocompass aboard a plane is used for driving motors at several of the compass repeater stations. If there are more than two motors, a larger generator must be used to maintain accuracy. If any motor becomes jammed, it affects the accuracy of the entire system since it acts as a generator competing with the real generator. As shown just below all the corresponding stator and rotor coils are connected in parallel.



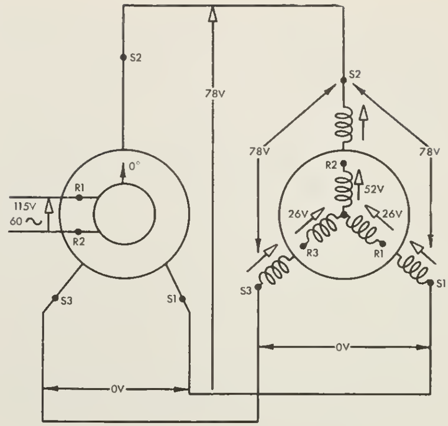
Two Motors Connected to One Generator

THE DIFFERENTIAL SELSYN

In a mechanical system, a differential connects three shafts together in such a way that the amount that one shaft turns is equal to the difference between the amounts that the other two turn. The differential selsyn is named according to the two functions it performs. If the shaft of the differential selsyn serves to indicate the difference in shafts positions of two other selsyns, the other two selsyns are generators and the differential selsyn is a *motor*. If the differential selsyn shaft position is to be subtracted from that of a selsyn generator (or vice versa) and the difference to be indicated by the rotor position of a motor, the differential selsyn is then a *generator*.

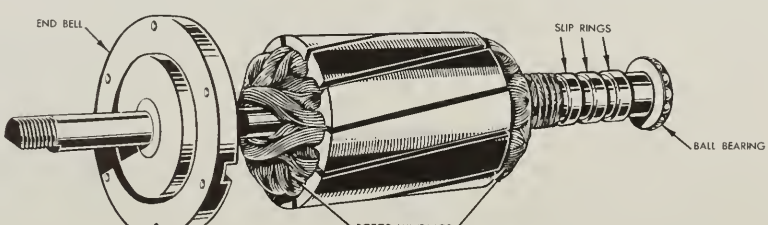
The stator of the differential generator or motor is very similar to the stator of the ordinary selsyn. It consists of three sets of coils wound in slots and spaced 120° apart around the inside of the field structure. The rotor, however, differs considerably from that of an ordinary selsyn. It is cylindrical in shape and has *three* sets of coils wound in slots and equally spaced around the circumference. Connections to the external circuits are made through three brushes riding on three slip rings on the rotor shaft. As in the ordinary selsyn, the motor has an oscillation damper and the generator does not. There is no connection to the 110v AC as in the case of the ordinary selsyn. The three leads to the rotor are designated R1, R2, and R3. Electrical zero is the position in which rotor coil R2 is aligned with stator coil S2, R1 with S1, and R3 with S3.

Electrically, the differential selsyn acts as a 1:1 transformer with the stator coils the primary and the rotor coils the secondary. Due to the air gap, there must be a few more turns on the stator coils than on the rotor; hence, the transformation is not a 1:1 ratio when the rotor coils are connected as the primary.

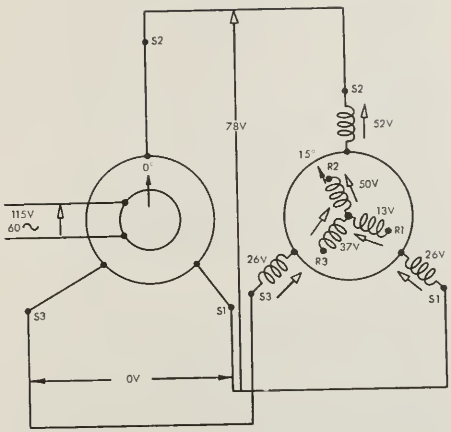


Differential Transformer Action at 0° Position

The selsyn generator just above is connected to a differential selsyn in which the rotor leads are open. Both shafts are in the 0° position. Since the stator coils of the differential selsyn connect to the stator coils of the generator, the voltages are equal in magnitude and are in phase. Since the differential is a 1:1 transformer, the induced voltages are equal in magnitude and are in phase with the stator voltages. When the generator shaft is turned to some other position, the voltages in the stator coils of the differential and, consequently, those induced in the rotor coils will be the same as those in the stator coils of the generator. Thus the rotor voltages of the differential are as shown in the rotor position diagram on page 13-10, as the generator shaft is rotated with the differential shaft remaining in the 0° position. A motor with its stator windings connected to the rotor windings of the differential will have the correct voltages applied causing it



Rotor of Differential Selsyn

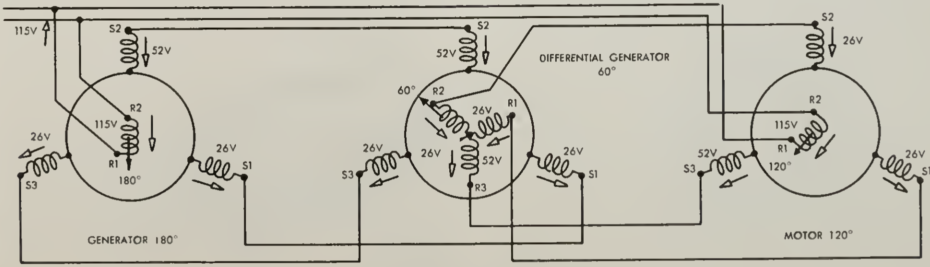


Transformer Action at 15° Position

to assume the same position as the generator rotor. This means, that with the differential rotor in the 0° position the system functions as it would if the differential were out of the circuit and connections were directly from generator to motor.

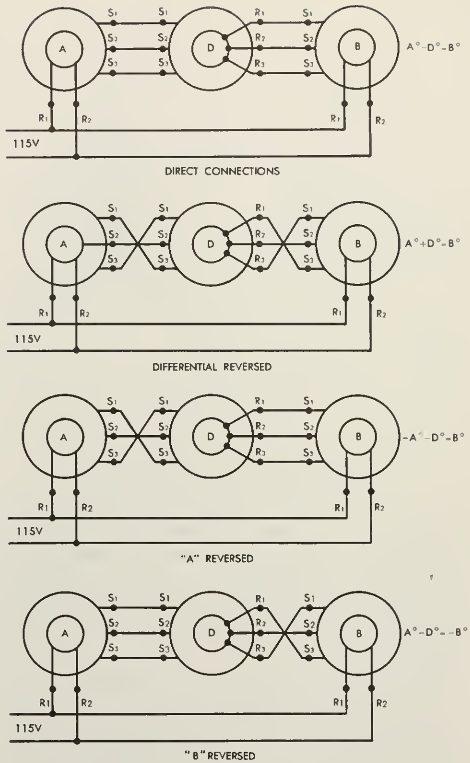
When the generator rotor is in the 0° position, and the rotor of the differential is turned to 15° as shown directly above, the voltages induced in the rotor coils will be as indicated. The voltages between terminals R1-R2, R2-R3 and R3-R1 are respectively 63v, 87v, and 24v and the R2-R3 voltage is in phase with the R1-R2 voltage across the rotor of the generator. The other voltages are 180° out of phase. These voltages correspond to the voltages induced in the stator windings of a generator in which the rotor is at the 345° position. To understand this action, consider the relative angle of the inducing magnetic field axis and the axis of the secondary coils. When a generator rotor is in the 345° position, the field axis is 15° *clockwise* from the axis of the S2 secondary coil, the voltage induced in it equals 52 cos (−15°). Likewise the voltages induced in S1 and S3 equal 52 cos (−255°) and 52 cos (−135°). These voltages set up a field in the motor which equals the vector sum of the three fields of the stator coils, resulting in a field with an axis at the 345° position. The rotor of a motor accordingly moves to the 345° position. Now in the differential selsyn, the composite primary field has reached the 0° position while the secondary (rotor) coils have rotated 15°. This makes the field axis 15° *clockwise* from the axis of the R2 coil. The induced voltages then have the values previously stated and shown in the transformer action illustration to the left. If the motor had its stator coils connected to the terminals R1, R2, R3 of the differential, the voltages applied to it would cause the rotor of the motor to move to the 345° position. Conversely, if the rotor of the differential were turned to the 345° position, the motor would turn to the 15° position. If the generator shaft position is at 0°, the voltages induced in the rotor of the differential selsyn are the same as if the rotor of the differential were at 0° and the rotor of the generator were rotated *in the opposite direction.*

Thus, whenever the shafts of both generator and differential rotate, the resultant field and induced voltages vary with the difference of the angular position of the shafts. A motor connected to the rotor coils of the differential will assume a shaft position equal to the difference of the angular positions of the other two shafts. The mathematical proof, a quite long one, is beyond the scope of this manual. Briefly, it involves setting up equations for the voltages induced in each of the rotor coils of the differential by each of the stator coils and adding the induced voltages



Selsyn System with Differential Generator in pairs. A qualitative explanation of what causes the motor shaft to turn to the indicated position is fairly simple to understand. For example, take the case illustrated at the bottom of page 13-15. The generator rotor is at the 180° position. Hence the voltages induced in S1, S2, S3 are 26v, 52v, 26v, respectively in the phases indicated. All the voltages S1, S2, S3 of the differential are identical. These voltages set up a magnetic field in the differential which has the same position, 180°, for its axis as the field in the generator. If the axis of the rotor of the differential generator were aligned with the 0° position, the induced voltages would be identical to the stator voltages. But since the rotor coil axis is shifted to the 60° position, the axis is only 120° from that of the magnetic field. Therefore the voltages induced in R1, R2, R3 are 52 cos 120°, 52 cos 240°, 52 cos 0° volts respectively. The voltages from R1 to R2, from R2 to R3 and from R3 to R1 are therefore 0v, 78v and 78v with R2-R3 being 180° out of phase with R1-R2 of the generator. When these voltages are applied to the motor, it will cause its rotor to turn to the position such that the voltages induced in its stator coils by the rotor just balance these voltages. This happens when the rotor is at the 120° position, the difference between 180° and 60°.

By reversing pairs of leads either between the generator and the differential or between the differential and the motor, you can make any one of the shafts assume a position equal to the sum or difference of the angular positions of the other shafts. The proper connections for various types of operation are shown in the illustration to the right.



Various Connection for Selsyn System

A differential selsyn cannot be a 1:1 transformer in both directions. Therefore the perfect balance of voltages indicated is not strictly accurate. The differential selsyn is designed such that the voltages induced in the rotor coils are equal to the voltages applied to the stator windings provided the shaft is at 0°. When a differential is connected between a generator and a motor, some current flows in the generator's stator circuit, reducing the output voltages due to the IR drop. These lower than normal voltages are induced in the rotor windings of the differential. They are too small to buck out the motor voltages; and some current flows in the motor stator circuits. The generator currents, usually higher, reduce the accuracy of the system. Therefore, when a very high degree of accuracy is required, the use of a differential is avoided if at all possible.

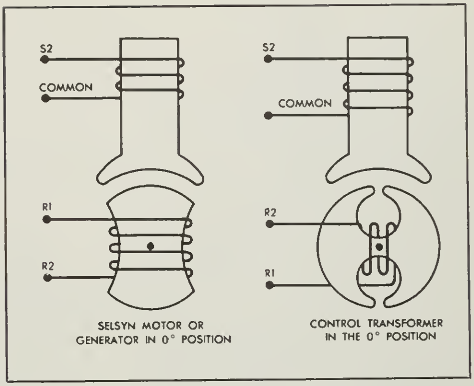
The Control Transformer

The uses of selsyn are to cause two shafts to rotate in synchronization and to produce an error voltage which indicates the difference in position of two shafts. The systems discussed to this point are for synchronization purposes. For the purpose of producing error voltages, there is a type of selsyn known as the control transformer which is used in conjunction with a generator. Both the generator and the transformer have their shafts connected to loads. The voltages induced in the stator coils of the generator are applied to the stator of the control transformer. These voltages induce in the rotor of the control transformer (not connected to the 115v AC) a voltage, the magnitude and polarity of which depends upon the relative positions of the two shafts.

The stator of a control transformer is very similar to that of a generator or motor except that the coils are wound with more turns and finer wire. The purpose of this is to make the impedance high enough to prevent high currents from flowing. Remember that no voltage is induced in the stator coils of the control transformer to buck the applied voltage and keep the current within safe limits. The stator current is determined by the impedance of the stator windings and is practically independent of the rotor's position.

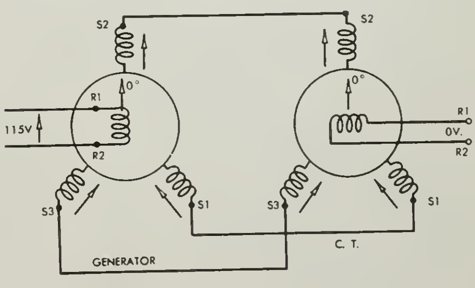
The rotor of a control transformer looks somewhat like the rotor of a differential selsyn. Yet there are some important differences. Instead of three groups of windings spaced 120° apart, all the windings are in series and have their external connections through two slip rings and brushes. The shaft always connects to a load making an oscillation damper unnecessary. The coils contain many turns of fine wire. The turns ratio is such that the maximum output voltage, with normal stator voltages, is 55 volts. There is no appreciable current in the rotor windings. The rotor is not free to turn to any particular position when voltages are applied to the stator windings. The electrical zero position of the rotor is such that no voltage is induced in the rotor windings. This means that the coil axis is at the right angle position as shown in the illustration at the upper right. Note that the terminals are labeled exactly as are those of a motor or generator.

This is how the control transformer operates. In the illustration just to the right, the generator shaft is at 0° and the control transformer shaft is at 0° position. The axis of the magnetic field set up by the stator voltages is in the 0° direction—at right angles to the rotor windings. It induces no voltage in the rotor windings. Now assume that the control transformer rotor is turned to

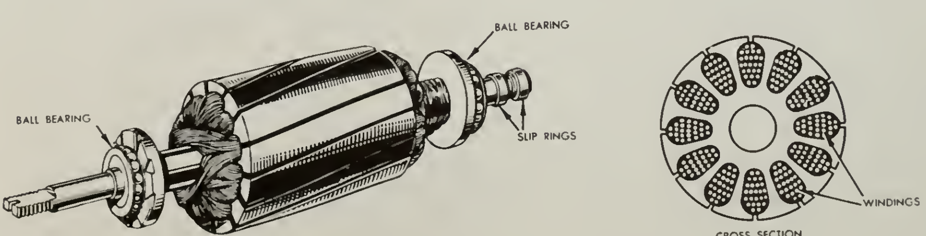


Comparing Electrical Zero Position

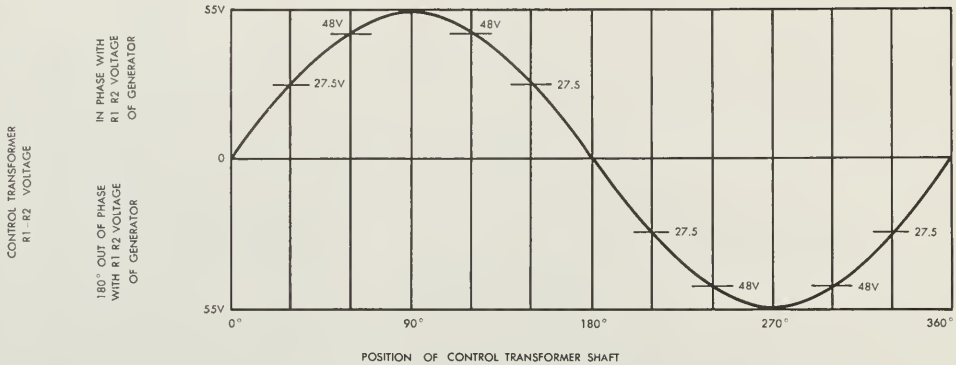
the 90° position. The axis of the coil will then be aligned with the composite magnetic field and there will be a maximum (55v) voltage induced in it. The phase polarity of the R1-R2 voltage will be identical to the R1-R2 voltage of the generator rotor. At other positions you can determine the magnitude of the induced voltage, by resolving the magnetic field into two components one of which is perpendicular to the axis of the coil and induces no voltage, and the other which is aligned with the coil and induces



No Voltage Induced with Rotor at 0° Position



Typical Control Transformer Rotor



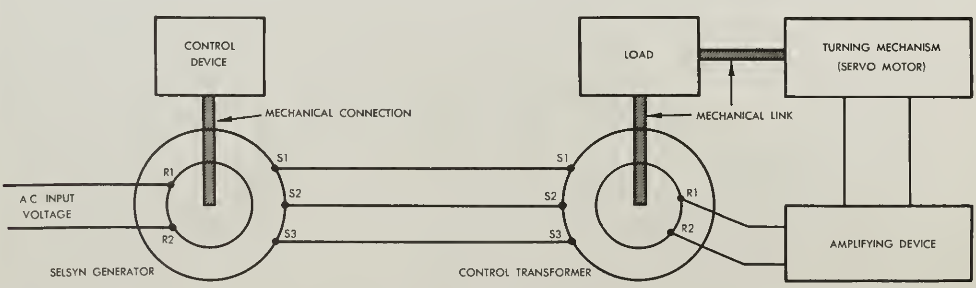
Induced Voltage vs Shaft Position

a voltage. The component of the field which is parallel to the coil axis is proportional to the sine of the angle to which the control transformer rotor is turned. Note in the induced voltage vs shaft position graph directly above that voltages above the zero axis are in phase with the generator rotor voltage, R1-R2, and the voltages below the axis are 180° out of phase. Obviously then the voltages from the 180° position to the 360° position are 180° out of phase.

When the rotor of the generator is turned to some position other than 0°, the magnetic field of the control transformer will follow and assume the same position, and when the rotor of the control transformer is turned the same position, the output will be zero volts, since the axis of the coil is perpendicular to the field. The rotor, starting from this position, on rotating in a counterclockwise direction will cause the voltages induced in the control transformer to assume the values indicated in the above graph, where the abscissa is the difference in angular position of the two rotors. This action indicates that if the control transformer (CT) shaft *leads* the generator shaft by any number of degrees up to 180, the output is *in phase* with the generator rotor voltage but, that if the CT shaft *lags* the generator shaft by any number of degrees up to 180, the output is opposite in phase.

SERVOMECHANISMS

You have already seen how selsyns transmit data from one place to another. When the shaft to be driven at the remote position is connected to an indicating device or some light load, the selsyn motor is capable of developing the necessary torque. But if the load is greater and more torque is necessary, then torque amplification is required. In this case the load is driven by a reversible adjustable speed motor which is called a servo motor. The complete system, including the motor and its control devices, is called a *servomechanism*. The speed and direction of rotation of the motor are controlled by the



Simple Servomechanism

magnitude and phase of an "error voltage" which is produced by any misalignment of the two shafts (the load shaft and the control shaft). Obviously, the direction of rotation must be such as to reduce the misalignment causing the error voltage.

The principles of the servo system in its simplest form are shown in conjunction with the illustration on the preceding page. Assume that the control device is a crank or a handwheel and the load is the antenna. To the handwheel there is connected a selsyn generator, and to the antenna a control transformer. The stator windings of the generator and control transformer are connected electrically as shown. The output of the control transformer is an AC voltage that depends upon the relative shaft positions for magnitude and phase. There is no output when the shafts are in the same position, and accordingly no output from the amplifying device. Hence the servo motor does not turn and the antenna remains stationary. To change the position of antenna, you merely turn the handwheel to the desired position. This produces an error voltage in the control transformer. This voltage is rectified and increased in magnitude by the amplifying device and then applied to the servo motor, causing it to turn the antenna to the desired position.

The selsyn generator and control transformer may be interchanged in position. However, such an interchange necessitates five electrical leads from one position to the other and gives no advantages. If the control device rotates continuously in one direction, the servo motor will rotate in the same direction at the same rate but with a small angle of lag. Remember that an error voltage is necessary before torque is developed. The greater the torque required, the greater will be the angle of lag.

The Servo Motor

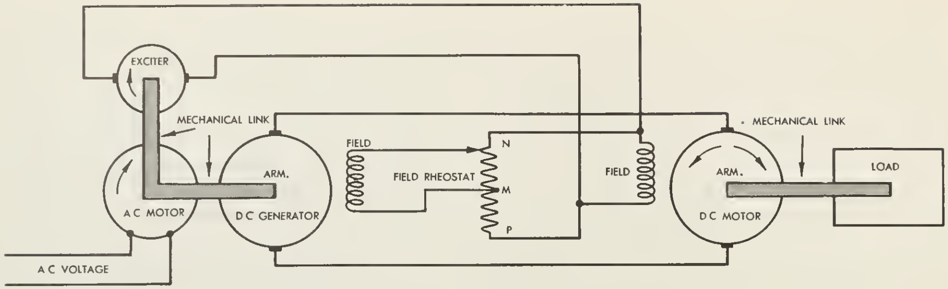
An AC motor, inherently a constant speed device, is not suitable for the requirements of a servomechanism and therefore a DC motor is used. There are two methods of controlling the speed and direction of rotation—controlling the voltage applied to the field coils, and controlling the armature voltage. Controlling the field voltage permits control of speed only from a certain minimum speed upward, since speed varies inversely as the flux set up by the field. However, reduction of speed to zero at times is necessary; hence the control in the servo system is always accomplished by varying the armature voltage.

Placing a variable resistor in series with the armature to control the armature voltage is wasteful since the resistor itself dissipates considerable power. Usually armature voltage is controlled by a DC generator driven by an AC motor. The Ward Leonard drive, discussed later, is an example of this system.

Servo motors operate with constant fields. The field is supplied either by permanent magnets or by DC applied to the field coils. The method of applying DC voltage to the field coils is wasteful of power, since power is consumed even when the motor is not turning. In addition the design of the equipment is much more complicated than the permanent magnet arrangement. Principally, because of this reason, the permanent magnet is the more widely used. One problem in the use of permanent magnets is the danger of demagnetization of the magnet by the field set up by current flow in the armature. This is overcome by the inclusion of a special winding which is connected in series with the armature and wound in opposite direction. In this winding, called the compensating winding, there is set up a compensating field which results from current flow through it. Since the current in the compensating wind varies with the armature current, the compensating field balances out the armature current and protects the permanent magnets.

The Ward Leonard Drive

In the Ward Leonard drive shown on the next page, AC motor turns at a constant speed always in the same direction. It drives the armatures of the large and small DC generators. The small generator provides excitation voltages for both the large DC generator and the motor. The DC motor field is constant, being supplied by field coils. The generator field is not constant. Both its polarity and magnitude are controllable by a field rheostat. When the variable arm is at the point M there is no voltage applied to the generator field coils and consequently there is no output voltage, no voltage applied to the armature of the DC motor, and no motor operation. When the variable arm is between M and N, a voltage of a certain polarity is applied to the field coils of the generator, the generator output has a certain polarity and causes the motor to turn in a certain direction. When the variable tap is located between M and P the generator field voltage is reversed in polarity, the generator

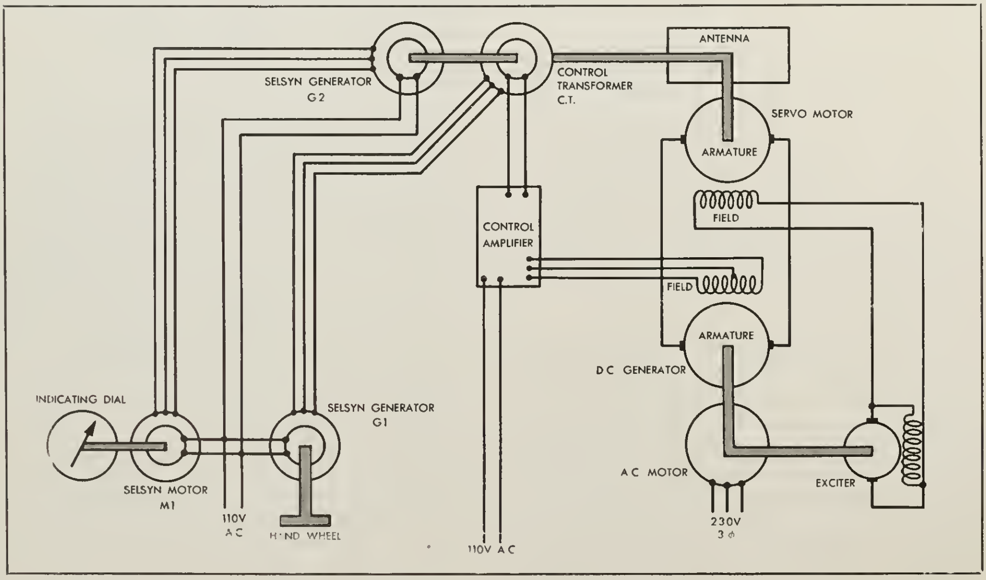


Ward Leonard Drive

output is reversed in polarity and the motor turns in the opposite direction. In either case the motor speed depends upon the voltage applied to the field coils and is controllable by the rheostat. Thus, you can control both the speed and direction of rotation of the motor. It will therefore function very well as a servo mechanism providing that the generator field is controlled by means of an error voltage instead of the field rheostat.

A Servo System Using the Ward Leonard Drive

When the Ward Leonard drive is used for a turning mechanism the servo system operates as shown directly below. The error voltage produced by the control transformer (CT) is fed into the control amplifier the output of which is a DC voltage having a polarity and magnitude that is determined by the phase and magnitude of the CT voltage. The output of the control amplifier is applied to the field of the DC generator of the Ward Leonard drive starting the action described in the last paragraph. The circuits of the control amplifier are explained later. The purpose of selsyn generator G2 and selsyn motor M1 is to furnish information to the operator as to the position of the antenna. In some cases this could be dispensed with since the scope pattern furnishes the same information when the scan is B type or PPI type.

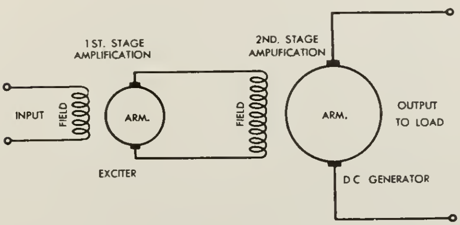


Servo System Using Ward Leonard Drive

The Amplidyne

The amplidyne is a special type DC generator commonly used for the DC generator in a system similar to the one described. Such a system is known as an amplidyne servomechanism. The word amplidyne means dynamo amplifier.

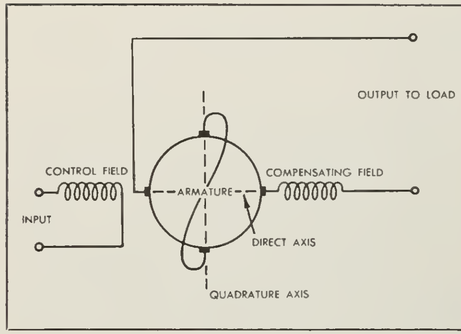
The DC generator may be considered an amplifier since a small amount of power applied to the field coils controls 10 to 100 times as much power in the output of the generator. Of course, the additional power comes from the mechanical power which turns the armature. This mechanical power is comparable to the B voltage of a vacuum tube amplifier. If greater amplification is desired than possible with a single DC generator, then two are used in cascade, that is, the output of the first (the exciter)is the input to the second.



DC Generators as Cascade Amplifiers

While an exciter was used in the Ward Leonard drive system, it may not be considered two stages of amplification, since the input to the exciter was uncontrolled. When two generators are used in cascade as is shown above, the amplification is the product of the two amplifications. For example, if each amplifies the power 30 times, the total amplification is 900.

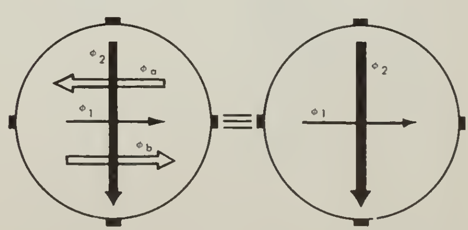
The amplidyne performs like a two-stage DC generator-amplifier with the two stages combined into one armature. The input goes to the control field which is on the direct axis of the machine. Rotation of the armature in the control field sets up a voltage at right angles to the field as shown at the top of the page. Brushes are placed on this axis. If a load were connected to these brushes, the machine would become an ordinary DC generator. However, in the amplidyne these brushes are short circuited and quite high currents flow in the armature producing a magnetic field along the quadrature (vertical) axis. This field is much stronger than the control field. This is known as armature reaction the



The Amplidyne

same as it is in an ordinary generator. The rotation of the armature through this quadrature field sets up a voltage at right angles to it, that is, along the direct (horizontal) axis. Brushes located on the direct axis have the load connected to them. When the load draws current, a magnetic field is set up along the direct axis due to the armature reaction of this second stage. Since the field due to armature reaction in each stage is ninety degrees from the field producing it, the reaction field of the second stage is 180° from the control field and tends to cancel it out, thereby reducing the amplification of the amplidyne. This is comparable to degeneration in a vacuum tube amplifier. In order to prevent this cancellation of the control field, compensating winding on the direct axis and in series with the brushes on the direct axis sets up a field which opposes and cancels out the armature reaction of the second stage.

As shown below the various magnetic fields of an amplidyne occupy certain positions and have certain directions. The symbol $\phi 1$ is the control field, $\phi 2$ the field due to armature reaction of the first stage or the quadrature field, ϕa is the armature reaction of the second stage and ϕb is the field due to the compensating windings.



Magnetic Fields in Amplidyne

Since, by design, $\phi b = \phi a$ but is opposite in direction, the net result is as shown in the effective diagram.

In order to guard against residual magnetism in the direct axis, a demagnetizing winding called a "killer" winding is sometimes found on the stator. It is excited with alternating current from a magneto generator mounted on the same frame. The magneto consists of a bar magnet revolving within a separate field winding to generate the AC voltage. The alternating current flowing in the stator windings forms a demagnetizing system which neutralizes any residual magnetism when the control field current is zero.

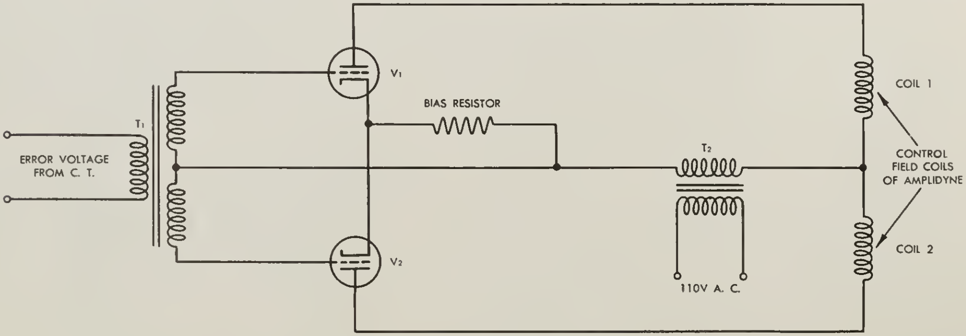
Actually there are two control field windings in an amplidyne when it is used with a control amplifier. The reason is that the vacuum tubes of the control amplifier conduct only in one direction and in order to control the polarity of the output voltage on the amplidyne, it is necessary to reverse the direction of the control field.

A practically constant input ratio makes it necessary that there be no saturation of the frame and the armature core. The use of a *killer* winding assists in making the gain constant. The power gain of the amplidyne is of the order of several hundreds or thousands and may even be tens of thousands. The full load input is about 0.25 to 0.75 watt which can easily be supplied by the control amplifier. The output varies from hundreds of watts to several thousand watts. The speed of response to changes in input depends upon the time constants of the control field, the quadrature windings of the armature and the load circuit or direct axis armature windings. Under the usual operating conditions, the time lag is about 0.1 second.

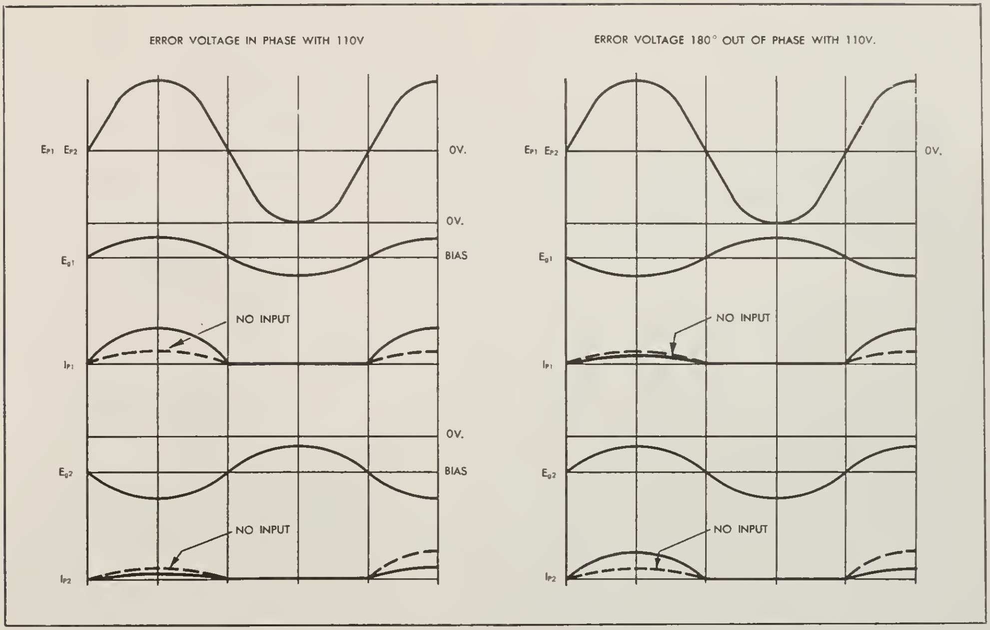
The Control Amplifier

As mentioned previously, the function of the control amplifier is to take the error voltage produced by the control transformer and to convert it into a DC voltage. The magnitude and polarity of this voltage is determined by the magnitude and relative phase of the error voltage. In the control amplifier circuit below, the two triodes are operated in push pull AC voltage from a 110 volt line is applied to the plates through transformer T_2. Bias for the tubes is provided by a common cathode resistor. The error voltage is applied to the grids through the transformer T_1 in such a way that the two grids are 180° out of phase. It is either in phase with or 180° out of phase with the 110 volt line voltage depending upon the relative position of the two shafts (selsyn generator and control transformer.)

To follow through the operation of the control amplifier, assume that the error voltage is in phase with the 110v and that the grid of V_1 goes in a positive direction at the same time that the plate goes positive. This causes an increased current flow through V_1 and this in turn an increase through coil 1. At the same time the grid of V_2 has gone in a negative direction and there is a decrease in current through V_2 and coil 2. Then in the amplidyne the field set up by coil 2 will decrease. There will then be an output voltage of a certain polarity from the amplidyne causing the servo motor to turn in a certain direction. During the alternation of the 110v when the plates are negative no current flows through either tube. Hence the current through coils 1 and 2 is pulsating direct current. (At the top of the next page see the illustration showing current and voltage relationships.)



Basic Control Amplifier Circuit



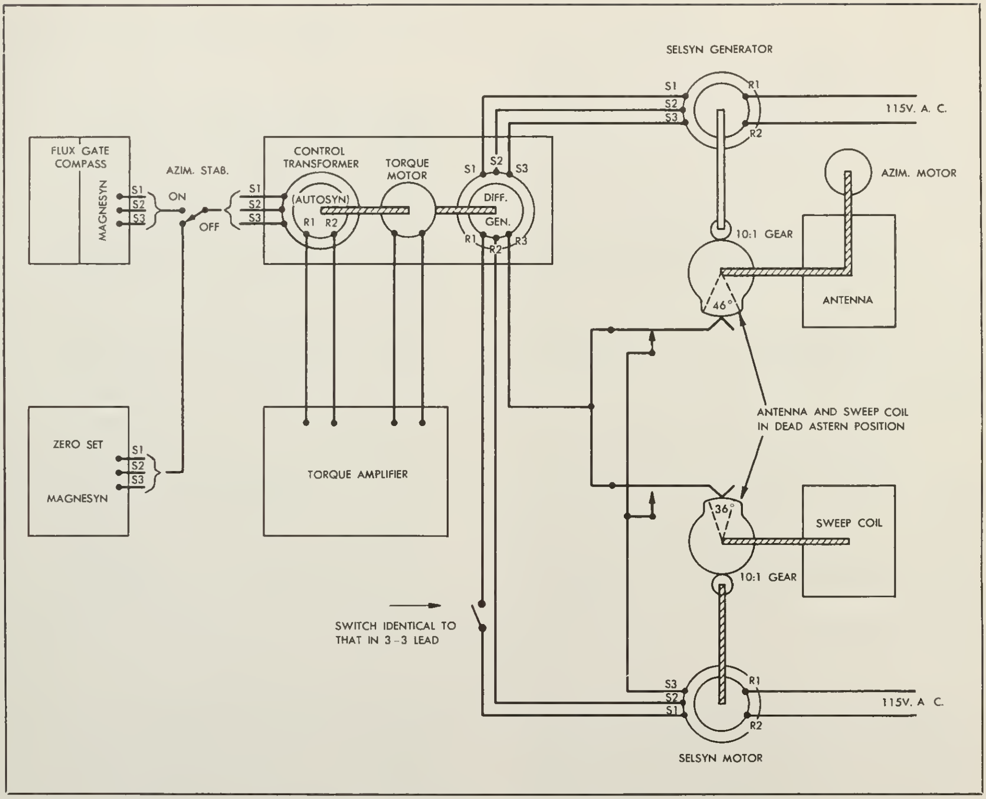
Control Amplifier Currents and Voltages

If a greater voltage is applied to the input transformer T_1, it will result in more current passing through coil 1 and less through coil 2 which causes the servo motor to turn faster.

When the error voltage is 180° out of phase with the 110 volts, the grid of V_2 will go positive and the grid of V_1 negative at the time the plates go positive. As a result, the current flow through coil 2 will increase and that through coil 1 will decrease. This causes the flux in the amplidyne to increase but in the opposite direction so that the output voltage is opposite in polarity. The servo motor will turn in the opposite direction.

In actual practice some sort of filter circuit is employed to smooth out the pulsations of the control amplifier output. This makes the output more steady. One way of accomplishing this is to connect resistors in parallel with the amplidyne control field windings. During the negative alternation of the plate voltage, the self-induced voltage of the inductance of the control field maintains the field current which finds a complete circuit through the shunting resistor. Another way of making the output more nearly pure DC is to put an RC circuit in to act as the load impedance and use this practically constant voltage to feed into a DC amplifier.

Hunting is quite likely to take place in a servomechanism unless some method is devised to prevent it. Hunting means that the servo motor drives the load past the desired position causing an error voltage which in turn reverses its direction and causes the motor to swing back and forth past the final position several times before stopping. Under some conditions continued oscillation may take place. The usual anti-hunt circuit consists of feeding back a portion of the amplidyne output into the input of the control amplifier. This voltage is proportional to the speed of the motor and is applied so that it makes the grid of the conducting tube more negative (in other words, it decreases the gain of the amplifier) and decreases the amplifier output and the motor speed. In this way the motor, in effect, acts as a brake on its own speed. The magnitude of the feedback decreases as the speed decreases and allows the motor to rotate the load to the desired position without unnecessary hunting. The amount of feedback is adjustable by a potentiometer.



Typical Selsyn System Employing Azimuth Stabilization

USE OF SERVOMECHANISMS FOR AZIMUTH STABILIZATION

The heading of an airplane is quite changeable. For this reason, it is desirable sometimes for the presentation of data on the indicator scope to remain in the same position regardless of changes in heading of the plane. For example if the 0° position of the PPI scan is always true north it is much easier to compare the scope picture with the map of the region. In order to accomplish this, it is necessary to change the relative positions of the antenna and sweep on the scope by an amount equal to the change in heading. This can be effected by applying azimuth stabilization to the antenna drive assembly.

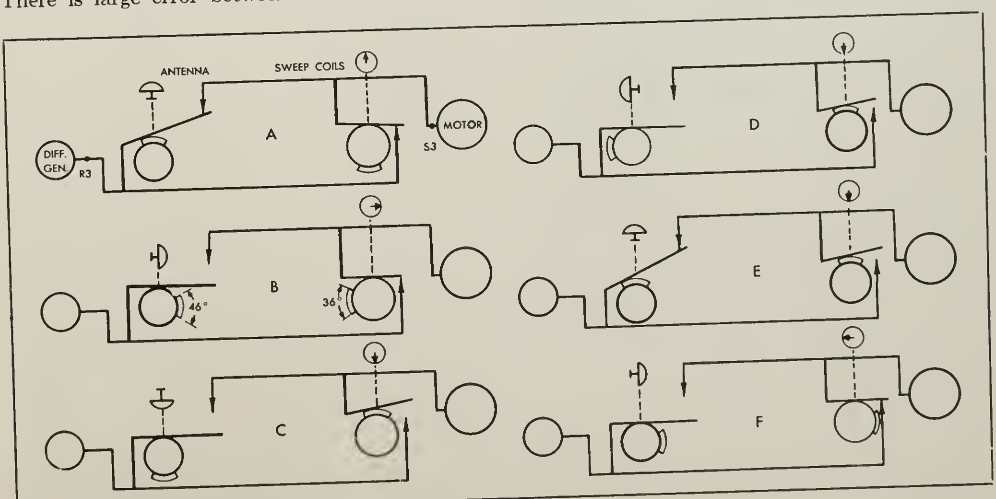
Selsyn System Employing Azimuth Stabilization

The circuits shown above are those of a typical set employing stabilization. Stabilization is obtained by shifting the sweep coils of the CRT whenever the plane changes its heading. The antenna is driven by the azimuth motor at a rate of approximately 24 RPM. A selsyn generator is geared to the antenna through a 10:1 gear. The generator is connected to a differential selsyn which has its output connected to a selsyn motor. This motor is geared, through another 10:1 gear to the sweep coils and causes them to rotate about the neck of the CRT. If the differential generator is in the electrical zero position, the rotation of the sweep coils is in synchronization with the rotation of the antenna. The purpose of the 10:1 gears is to reduce to 1/10 of its value the lag inherent in a selsyn system. In this system, if the selsyn motor lags the generator by 10°, it produces only 1° lag between the antenna and sweep coils. The actual error is a small fraction of 1°. The increased accuracy is gained by the possibility of the antenna and sweep coils being 36°, or some multiple thereof, out of synchronization. Of course the selsyns are always in step but since 36° rotation of the antenna (or sweep coils) will mean one complete rotation of the selsyn motor, there is the possibility that this error could occur. To guard against such an occurrence, cams are attached to the shafts of the antenna and the sweep coils. These are called phasing cams and operate switches in the R1-S1 and R3-S3 leads from the differential generator to the selsyn motor. The cam attached to the antenna shaft *closes* the two switches during 46° of rotation centered about the dead astern position. The cam attached to the sweep coil shaft *opens* switches in the same leads during 36° of rotation, also centered about the astern position. If the two shafts are in synchronization the antenna cam closes the switches 5° ahead of the position where the sweep coil cam opens the switches and holds them closed for 5° past the position where the sweep coil cam again closes its switches. Thus the circuits are complete at all times and the selsyn motor drives the sweep coils in step with the antenna.

To illustrate the action assume that when the system is turned on and the antenna is pointing somewhere in the aft sector, the sweep coils are at a position corresponding to dead ahead. Both sets of phasing switches are closed, and the antenna and sweep coils start rotating. There is large error between the antenna and sweep coils. The illustration below shows the cams and switches at successive steps in the rotation. This out-of-step rotation will continue until both switches are open at the same time (see C), then the sweep coils will stop and remain in that position (180°) until the antenna reaches the position where it will close the switch. This occurs at 23° before the dead astern position. The sweep coil rotor will then electrically interlock and cause the antenna and sweep coils to be in step once more.

The description of the operation thus far assumes that the differential generator was in the electrical zero position. This is true if the AZIM. STAB switch is in the OFF position. In this type operation, the control transformer in the phasing unit has its stator windings connected to a zero set magnesyn. The magnesyn is a device similar in performance to a selsyn generator. The control transformer in this particular set up is known as an autosyn. The output of the control transformer, after being amplified by the torque amplifier, causes the torque motor to turn the CT rotor and differential selsyn rotor to the desired position which is the electrical zero position of the differential generator.

When using azimuth stabilization, turn the switch AZIM. STAB to the ON position. This disconnects the control transformer from the zero set magnesyn and connects it to the flux gate compass magnesyn. The switch also shorts out the phasing switches and connects the differential



Action of Phasing Switches

generator directly to the selsyn motor. The magnesyn in the flux gate compass produces voltages to set up a field in the control transformer which is aligned with the position of the rotor in the magnesyn. This position is dependent upon the position of the compass. The error voltage produced by the control transformer, after amplification, turns the control transformer rotor and that of the differential selsyn to the position corresponding to the compass position. Since the rotor of the differential generator is shifted by an angle equal to the difference between the plane's heading and true north, its output voltages will be such as to cause the sweep coil motor to lead or lag the antenna generator by the same angle. By proper connection and adjustment the sweep coils move in such a way that the 0° position of the scope is true north. In order to determine the plane's heading from the scope pattern a switch connected to the antenna causes the bias of the CRT to decrease when the antenna points dead ahead. This results in a brighter trace which indicates the planes heading. The purpose in shorting out the phasing switches is to permit the antenna and sweep coils to be out of step, as is the case when the plane's heading is other than north.

Index

E

F

Q

R

S

T

U

V

W

Z

☆ U S GOVERNMENT PRINTING OFFICE 1959—520983/15