ELECTROMAGNETIC WAVES AND RADIATING SYSTEMS

EDWARD C. JORDAN

PRENTICE-HALL ELECTRICAL ENGINEERING SERIES

CONVERSION OF UNITS

Symbol	Quantity	MKS units	CGS units	
		The MKS unit below	*is equal to the following number of CGS electromagnetic units:*	*and is also equal to the following number of CGS electrostatic units:*
q	Charge	1 coulomb	10^{-1}	3×10^{9} (stat-coulombs)
I	Current	1 ampere	10^{-1} (ab-amperes)	3×10^{9}
V	Electromotive force	1 volt	10^{8} (ab-volts)	$\frac{1}{3} \times 10^{-2}$ (stat-volts)
R	Resistance	1 ohm	10^{9}	$\frac{1}{9} \times 10^{-11}$
C	Capacitance	1 farad	10^{-9}	9×10^{11} (stat-farads or centimeters)
L	Inductance	1 henry	10^{9}	$\frac{1}{9} \times 10^{-11}$
$\mathbf{F}$	Force	1 newton	10^{5} (dynes)	10^{5} (dynes)
U	Energy	1 joule	10^{7} (ergs)	10^{7} (ergs)
W	Power	1 watt	10^{7} (ergs/sec)	10^{7} (ergs/sec)
σ	Conductivity	1 mho/meter	10^{-11}	9×10^{9}
Φ	Magnetic flux	1 weber	10^{8} (Maxwells)	$\frac{1}{3} \times 10^{-2}$
$\mathbf{B}$	Magnetic flux density	1 weber/square meter	10^{4} (gauss)	$\frac{1}{3} \times 10^{-6}$
$\mathfrak{F}$	Magnetomotive force*	1 ampere (-turn)	$4\pi \times 10^{-1}$ (gilberts)	$12\pi \times 10^{9}$
$\mathbf{H}$	Magnetic intensity*	1 ampere/meter	$4\pi \times 10^{-3}$ (oersteds)	$12\pi \times 10^{7}$

* The starred quantities are affected by rationalization. For these quantities the conversion factors are shown for converting from *rationalized* MKS units to *unrationalized* CGS units.

Meters	Inches	Feet	Miles
.0254	1	.0833	
.3048	12	1	.0001894
1	39.37	3.281	.0006214
1609		5280	1

ELECTROMAGNETIC WAVES AND RADIATING SYSTEMS

EDWARD C. JORDAN
Professor of Electrical Engineering
College of Engineering
University of Illinois

Englewood Cliffs, N. J.
PRENTICE-HALL, INC.

First Printing......September, 1950
Second Printing.........April, 1951
Third Printing..........May, 1955
Fourth Printing......January, 1958
Fifth Printing......November, 1958
Fifth Printing......November, 1958
Sixth Printing..........April, 1960

PRINTED IN THE UNITED STATES OF AMERICA

24998-C

PREFACE

A knowledge of electromagnetic radiation and propagation is now required of virtually all communication and electronic engineers. This book is designed to provide a course in this field for electrical engineers and physicists. It is an outgrowth of courses given by the author at Ohio State University and at the University of Illinois. The level of the first part of the book is suitable for seniors and beginning graduate students; the later chapters are primarily for more advanced graduate students. Although there is sufficient material for a two-semester course, many instructors may prefer to select only certain chapters to be covered in a one-semester or one-quarter course. The division of material among chapters has been made with this fact in mind.

In a text of this scope it is necessary to draw from the writings of many specialists. I am indebted to Professor Erik Hallén for the use of his antenna impedance curves in Chapter 13. For the chapters on propagation, material from the papers of K. A. Norton and C. R. Burrows has been used. The writings of S. A. Schelkunoff are already classics and are largely responsible for many engineering concepts, such as wave impedance and magnetic currents, now in general use. References to his papers and book will be found throughout the text.

It is a pleasure to acknowledge the assistance given by the author's associates at the University of Illinois and elsewhere. W. G. Albright, R. S. Elliott, P. K. Hudson, Ray DuHamel, Edgar Hayden, John Myers, Douglas Royal, John Bell, and many others gave freely of their time in checking the manuscript and reading proof. Discussions with George Sinclair were always helpful. I am especially indebted to J. A. Barkson, who read much of the manuscript and offered many suggestions, and to Nicholas Yaru, who drew the originals for the illustrations.

Several years ago it was my privilege to take a graduate course in

radiation from Professor W. L. Everitt at the Ohio State University. The original set of notes, "Radiation and Radiating Systems," used for that course has formed the nucleus about which this book has been developed. It is my hope that some of the engineering philosophy that was so much a part of that early course may have been carried over into this work.

E. C. JORDAN

Urbana, Illinois

CONTENTS

CHAPTER 1

FUNDAMENTALS OF ELECTROMAGNETIC ENGINEERING

1.01 Circuits and Fields. The rapid advances that have been made in electrical engineering during the past few decades have been due largely to the ability of the engineer to predict with accuracy the performance of complicated electrical networks. The secret of this ability lies chiefly in the use of a simple but powerful tool called *circuit theory.* The power of the circuit approach depends upon its simplicity, and this simplicity is due to the fact that circuit theory is a simplified approximation of a more exact field theory. In chap. 11 familiar circuit relations are derived directly from the more general field relations, and in the process the assumptions and approximations involved in the use of circuit theory are made apparent.

Despite the power and usefulness of the circuit approach the communications engineer concerned with microwaves or with radio transmission problems quickly becomes aware of its limitations. In the over-all design of a radio communication system the engineer can use circuit theory to design the terminal equipment, but between the output terminals of the transmitter and the input terminals of the receiver, circuit theory fails to give him answers, and he must turn to field theory. Electromagnetic field theory deals directly with the field vectors **E** and **H**, whereas circuit theory deals with voltages and currents that are the integrated effects of electric and magnetic fields. Of course voltages and currents are the end results in which the engineer is interested, but the intermediate step, the electromagnetic field, is now a necessary one. It is the purpose of this book to familiarize the student and the engineer with the fundamental relations of the electromagnetic field, and to demonstrate how such relations are used in the solution of engineering problems.

Field theory is more difficult than circuit theory only because of the larger number of variables involved. When current is constant around a circuit, the voltages and currents are functions of one variable—time. In uniform-transmission-line theory, the distance along the line is an added variable, but the engineer has learned to treat this distributed-constants circuit by means of an extension of ordinary circuit theory, which he calls transmission line theory. In the most general electromagnetic field problems there are three space variables involved, and the solutions tend to become correspondingly complex. The additional complexity that results from having to deal with vector quantities in three dimensions can largely be overcome by use of vector analysis. The small amount of effort required to become familiar with vector analysis is soon amply repaid by the simplification that results from its use. For this reason the first topic to be treated will be vector analysis.

1.02 Vector Analysis. The use of vector analysis in the study of electromagnetic field theory results in a real economy of time and thought. Even more important, the vector form helps to give a clearer understanding of the physical laws that mathematics describes. To express these essentially simple physical relations in the longhand scalar form is like trying to sing a song note-by-note, or like sending a code message dot-by-dash, instead of in letter or word groups. The more concise vector form states each relation as a whole, rather than in its component parts. The brief introduction to vector analysis included here is for the benefit of those readers not already familiar with this useful tool. This treatment is adequate for present purposes, but it is expected that the student may later find it desirable to refer to some standard vector analysis text for a more thorough presentation.

Scalar. A quantity that is characterized only by magnitude and algebraic sign is called a scalar. Examples of physical quantities that are scalars are mass, time, temperature, and work. They are represented by italic letters, such as A, B, C, a, b, and c.

Vector. A quantity that has direction as well as magnitude is called a vector. Force, velocity, displacement, and acceleration are examples of vector quantities. They are represented by letters in bold-face roman type, such as **A**, **B**, **C**, **a**, **b**, and **c**. A vector can be represented geometrically by an arrow whose direction is appropriately chosen and whose length is proportional to the magnitude of the vector.

Field. If at each point of a region there is a corresponding value of some physical function, the region is called a field. Fields may be classified as either scalar or vector, depending upon the type of function involved.

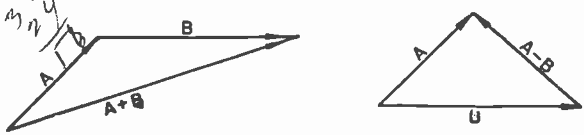
If the value of the physical function at each point is a scalar quantity, then the field is a *scalar field.* The temperature of the atmosphere, the height of the surface of the earth above sea level, and the density of a nonhomogeneous body are examples of scalar fields.

When the value of the function at each point is a vector quantity, the field is a *vector field.* The wind velocity of the atmosphere, the force of gravity on a mass in space, and the force on a charged body placed in an electric field, are examples of vector fields.

Sum and Difference of Two Vectors. The sum of any two vectors **A** and **B** is illustrated in Fig. 1-1a. It is apparent that it makes no difference whether **B** is added to **A** or **A** is added to **B**. Hence

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{1-1}$$

When the order of the operation may be reversed with no effect on the result, the operation is said to obey the *commutative law.*



(a) (b)

Fig. 1-1

Figure 1-1b illustrates the difference of any two vectors **A** and **B**. It is to be remembered that the negative of a vector is a vector of the same magnitude, but with a reversed direction.

Multiplication of a Scalar and a Vector. When a vector is multiplied by a scalar, a new vector is produced whose direction is the same as the original vector and whose magnitude is the product of the magnitudes of the vector and scalar. Thus

$$\mathbf{C} = a\mathbf{B} \tag{1-2}$$

Note the absence of any multiplication symbols between a and **B**. The symbols · and × are reserved for special types of multiplication, which will be discussed later.

The mathematician finds in vector analysis a tool by which relationships can be expressed without reference to a co-ordinate system. The engineer, however, generally needs a reference set of co-ordinates to solve problems. The text will use rectangular or Cartesian co-ordinates, except in those cases where other co-ordinate systems reduce the complexity of the problems. It will be assumed that all vectors and fields are three-dimensional.

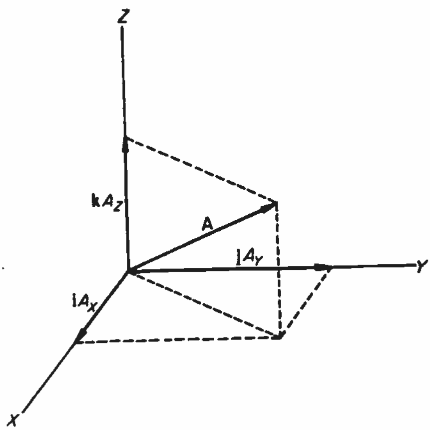


Fig. 1-2

A three-dimensional vector is completely described by its projections on the x, y, and z axes. Therefore it can be said that a three-dimensional vector specifies three scalars (the scalar magnitudes of the three mutually orthogonal vector components). Also, a vector field specifies three scalar fields (the scalar magnitudes of the three component vector fields). This idea of component vectors can be represented by

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k} \tag{1-3}$$

where A_x, A_y, and A_z are the magnitudes of the projections of the vector on the x, y, and z axes respectively, and **i**, **j**, and **k** are unit vectors in the direction of the axes, (Fig. 1-2).

If any two vectors **A** and **B** are added, there results

$$\mathbf{A} + \mathbf{B} = A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k} + B_x\mathbf{i} + B_y\mathbf{j} + B_z\mathbf{k} \tag{1-4}$$

which can be grouped as

$$\mathbf{A} + \mathbf{B} = (A_x + B_x)\mathbf{i} + (A_y + B_y)\mathbf{j} + (A_z + B_z)\mathbf{k} \tag{1-5}$$

This shows that each of the three components of the resultant vector is found by adding the two corresponding components of the individual vectors.

Furthermore, in any vector *equation*, the sum of the **i** components on the left-hand side is equal to the sum of the **i** components on the right-hand side. The same is true also of the **j** and **k** components. *Therefore, a vector equation can be written as three separate and distinct equations.* For example, the equation

$$\mathbf{A} + \mathbf{B} = \mathbf{C} + \mathbf{D} + \mathbf{E} \tag{1-6}$$

could be written as the three equations

$$A_x + B_x = C_x + D_x + E_x \tag{1-6a}$$

$$A_y + B_y = C_y + D_y + E_y \tag{1-6b}$$

$$A_z + B_z = C_z + D_z + E_z \tag{1-6c}$$

The ease with which three component equations can be written as one vector equation makes vector analysis particularly useful in field theory.

Scalar Multiplication. It was just shown that a vector could be multiplied by a scalar. It is also possible to multiply a vector by a vector, but first the meaning of such multiplication must be defined and suitable rules formulated. Two types of vector multiplication have been defined, namely "scalar product" and "vector product." The meaning of such multiplications and the necessary rules are briefly discussed in the following. The *scalar product* of two vectors is a *scalar* quantity whose magnitude is equal to the product of the magnitudes of the two vectors and the cosine of the angle between them. This type of multiplication is often called the *dot product* and is indicated by a · (dot) placed between the two vectors to be multiplied. Hence in Fig. 1-3,

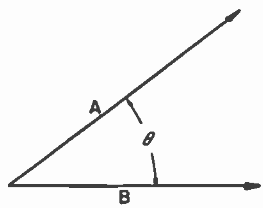


Fig. 1-3

$$\mathbf{A} \cdot \mathbf{B} = AB \cos \theta \tag{1-7}$$

It is seen that the dot product obeys the commutative law, that is

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \tag{1-8}$$

A physical example of the dot product can be found in the relationship between force and distance. If **F** represents a force that acts through the distance **D** (Fig. 1-4), then the work done would be given by the equation

$$\text{Work} = \mathbf{F} \cdot \mathbf{D} \tag{1-9}$$

Again notice that the dot product, which in this case is work, is a scalar quantity.

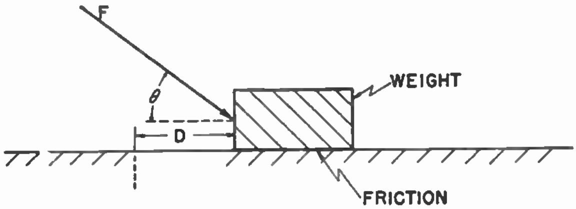


FIG. 1-4

The dot product of two vectors can be found by using ordinary algebraic rules.

Let

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k}$$
$$\mathbf{B} = B_x\mathbf{i} + B_y\mathbf{j} + B_z\mathbf{k}$$

Therefore

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} = & A_xB_x(\mathbf{i} \cdot \mathbf{i}) + A_xB_y(\mathbf{i} \cdot \mathbf{j}) + A_xB_z(\mathbf{i} \cdot \mathbf{k}) \\ & + A_yB_x(\mathbf{j} \cdot \mathbf{i}) + A_yB_y(\mathbf{j} \cdot \mathbf{j}) + A_yB_z(\mathbf{j} \cdot \mathbf{k}) \\ & + A_zB_x(\mathbf{k} \cdot \mathbf{i}) + A_zB_y(\mathbf{k} \cdot \mathbf{j}) + A_zB_z(\mathbf{k} \cdot \mathbf{k})\end{aligned} \tag{1-10}$$

But it can be seen from eq. (7) that

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1 \tag{1-11a}$$
$$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{i} = \mathbf{k} \cdot \mathbf{j} = \mathbf{i} \cdot \mathbf{k} = 0 \tag{1-11b}$$

Therefore eq. (10) reduces to

$$\mathbf{A} \cdot \mathbf{B} = A_xB_x + A_yB_y + A_zB_z \tag{1-12}$$

Vector Multiplication. The *vector product* of two vectors is defined as a *vector* whose magnitude is the product of the magnitudes of the two vectors and the sine of the angle between them, and whose direction is perpendicular to the plane containing the two vectors. If a right-handed screw is rotated from the first vector to the second (through the smaller included angle), it moves in the positive direction of the resultant vector. This type of multiplication is often called the *cross product* and is indicated by a × (cross) placed between the two vectors to be multiplied. Hence, in Fig. (1-3)

$$|\mathbf{A} \times \mathbf{B}| = AB \sin \theta \tag{1-13}$$

where the bars | | indicate "magnitude of."

The direction of the vector **A** × **B** would be into the paper away from the reader. The vector **B** × **A** would have the same magnitude but the opposite direction, that is, toward the reader. Therefore

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{1-14}$$

and the commutative law does *not* apply.

A physical example of vector multiplication can be found in the lifting force of a screw jack. If friction is neglected and a force **f** is applied at the end of a lever arm of length ***l***, then the lifting force **F** produced by the jack will be

$$\frac{p}{2\pi}\mathbf{F} = \mathbf{f} \times \mathbf{l}$$

where the constant p is the pitch of the screw.

The vector product may also be obtained by straightforward algebraic multiplication and a result similar to that of eq. (10) obtained. Thus

$$\begin{aligned} \mathbf{A} \times \mathbf{B} = {} & A_xB_x(\mathbf{i} \times \mathbf{i}) + A_xB_y(\mathbf{i} \times \mathbf{j}) + A_xB_z(\mathbf{i} \times \mathbf{k}) \\ & + A_yB_x(\mathbf{j} \times \mathbf{i}) + A_yB_y(\mathbf{j} \times \mathbf{j}) + A_yB_z(\mathbf{j} \times \mathbf{k}) \\ & + A_zB_x(\mathbf{k} \times \mathbf{i}) + A_zB_y(\mathbf{k} \times \mathbf{j}) + A_zB_z(\mathbf{k} \times \mathbf{k}) \end{aligned} \tag{1-15}$$

By using eqs. (13) and (14) and a right-handed system of co-ordinates (Fig. 1-5) it is found that

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} = -\mathbf{j} \times \mathbf{i} \tag{1-16a}$$

$$\mathbf{j} \times \mathbf{k} = \mathbf{i} = -\mathbf{k} \times \mathbf{j} \tag{1-16b}$$

$$\mathbf{k} \times \mathbf{i} = \mathbf{j} = -\mathbf{i} \times \mathbf{k} \tag{1-16c}$$

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = 0 \tag{1-16d}$$

Therefore eq. (15) reduces to

$$\begin{aligned} \mathbf{A} \times \mathbf{B} = {} & (A_yB_z - A_zB_y)\mathbf{i} \\ & + (A_zB_x - A_xB_z)\mathbf{j} \\ & + (A_xB_y - A_yB_x)\mathbf{k} \end{aligned} \tag{1-17}$$

This result may be remembered easily by noting that the subscripts of the first (positive) part of each term are cyclic with an x-y-z rotation when combined with the axis direction of the associated unit vector. For example, in the first part of the first term, the subscript order is y-z-x (i is in the x direction). The subscript order of the positive part of the second term is z-x-y, and for the positive part of the third term it is x-y-z. The second or negative part of

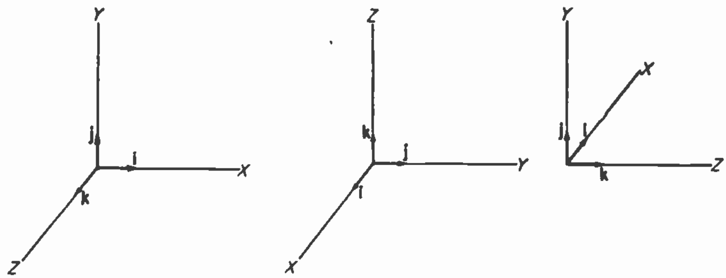


FIG. 1-5. Right-handed co-ordinate system.

each term is obtained by reversing the subscripts of the first part of the term. The correct order also may be found from the determinant

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ \mathbf{i} & \mathbf{j} & \mathbf{k} \end{vmatrix} \quad \text{or} \quad \mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}$$

Differentiation—The ∇ Operator. The differential vector operator ∇, called *del* or *nabla,* has many important applications in physical problems. It is defined as

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k} \qquad (1\text{-}18)$$

A differential operator can be treated in much the same way as any ordinary quantity. For example, with the operator $D = \partial/\partial_x$, the operation Dy means the quantity $\partial y/\partial x$ is to be obtained.

There are three possible operations with ∇ corresponding to the three possible types of vector multiplication, illustrated in eqs. (2), (12), and (17).

1. If V is a scalar function, then by eqs. (2) and (18)

$$\nabla V = \frac{\partial V}{\partial x}\mathbf{i} + \frac{\partial V}{\partial y}\mathbf{j} + \frac{\partial V}{\partial z}\mathbf{k} \qquad (1\text{-}19)$$

This operation is called the *gradient* (for reasons to be explained later), and is abbreviated

$$\nabla V = \text{grad } V \tag{1-20}$$

2. If **A** is a vector function, we can apply eqs. (10), (12), and (18) and get

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{1-21}$$

This operation is called the *divergence* and is abbreviated

$$\nabla \cdot \mathbf{A} = \text{div } \mathbf{A} \tag{1-22}$$

3. If **A** is a vector function, we can use eqs. (15), (17), and (18) to show that

$$\nabla \times \mathbf{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\mathbf{k} \tag{1-23}$$

$$\nabla \times \mathbf{A} = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \\ \mathbf{i} & \mathbf{j} & \mathbf{k} \end{vmatrix}$$

This operation is called the *curl* and can be written as

$$\nabla \times \mathbf{A} = \text{curl } \mathbf{A} \tag{1-24}$$

Identities. The identities that follow are useful in deriving field equations. The student can verify them by direct expansions.

$$\text{div curl } \mathbf{A} = \nabla \cdot (\nabla \times \mathbf{A}) = 0 \tag{1-25}$$

$$\text{curl grad } V = \nabla \times (\nabla V) = 0 \tag{1-26}$$

$$\text{div grad } V = \nabla \cdot (\nabla V) = \nabla^2 V$$

where ∇^2 is defined (in Cartesian co-ordinates) as the operation*

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1-27}$$

* The operator ∇^2 (del squared) is called the *Laplacian*. The Laplacian of a scalar V is given by eq. (27). The Laplacian of a vector **A** is defined as the vector whose Cartesian components are the Laplacians of the Cartesian components of **A**, That is

$$\nabla^2\mathbf{A} = \mathbf{i}\nabla^2 A_x + \mathbf{j}\nabla^2 A_y + \mathbf{k}\nabla^2 A_z \qquad \text{(9 terms)}$$

$$\text{curl curl } \mathbf{A} = \nabla \times (\nabla \times \mathbf{A}) = \text{grad div } \mathbf{A} - \nabla^2\mathbf{A} \quad (1\text{-}28)$$
$$\text{div } \mathbf{A} \times \mathbf{B} = \nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \text{curl } \mathbf{A} - \mathbf{A} \cdot \text{curl } \mathbf{B} \quad (1\text{-}29)$$

Direction Cosines. The *component* of a vector in a given direction is the projection of the vector on a line in that direction. Thus A_x, the x component of **A**, is equal to $A \cos \alpha$, where α is the angle between **A** and the x axis. Then

$$A_x = \mathbf{A} \cdot \mathbf{i}$$

That is, the component of a vector in a given direction is equal to the dot product of the vector and a unit vector in that direction.

If a vector makes angles α, β, γ, with the co-ordinate axes, then

$$l = \cos \alpha, \qquad m = \cos \beta, \qquad n = \cos \gamma$$

are known as the *direction cosines* of the vector.

Problem 1. The scalar product of two vectors may be written in terms of the sum of the products of their direction components.

$$\mathbf{A} \cdot \mathbf{B} = A_xB_x + A_yB_y + A_zB_z$$

Show that the cosine of the angle ψ between the vectors is given by the sum of the products of their direction cosines:

$$\begin{aligned}\cos \psi &= \cos \alpha_A \cos \alpha_B + \cos \beta_A \cos \beta_B + \cos \gamma_A \cos \gamma_B \\ &= l_A l_B + m_A m_B + n_A n_B\end{aligned}$$

1.03 Physical Interpretation of Gradient, Divergence, and Curl. The three operations which can be performed with the operator del have important physical significance in scalar and vector fields. They will be considered in turn.

Gradient. The gradient of any scalar function is the maximum space rate of change of that function. If the scalar function V represents temperature, then $\nabla V = \text{grad } V$ is a temperature gradient, or rate of change of temperature with distance. It is evident that although the temperature V is a scalar quantity—having magnitude but no direction—the temperature gradient ∇V is a vector quantity, its direction being that in which the temperature changes most rapidly. This vector quantity may be expressed in terms of its components in the x, y, and z direction. These are respectively $\dfrac{\partial V}{\partial x}$, $\dfrac{\partial V}{\partial y}$, and $\dfrac{\partial V}{\partial z}$. The resultant temperature gradient is the vector sum of these three components:

$$\nabla V = \frac{\partial V}{\partial x}\mathbf{i} + \frac{\partial V}{\partial y}\mathbf{j} + \frac{\partial V}{\partial z}\mathbf{k}$$

If the scalar V represents electric potential in volts ∇V represents potential gradient or electric intensity in volts per meter (MKS).

Divergence. As a mathematical tool, vector analysis finds great usefulness in simplifying the expressions of the relations that exist in three-dimensional fields. A consideration of fluid motion gives a direct interpretation of divergence and curl.

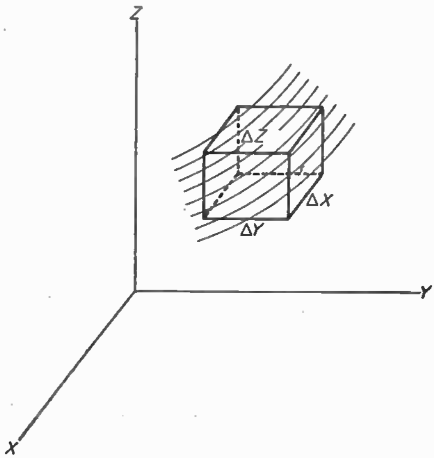


Fig. 1-6

Consider first the flow of an incompressible fluid. (Water is an example of a fluid that is almost incompressible.) In Fig. 1-6 the rectangular parallelepiped Δx, Δy, Δz, is an infinitesimal volume element within the fluid. If ρ_m is the mass density of the fluid, the flow into the volume through the left-hand face is $\rho_m v_y \,\Delta x\, \Delta z$ where v_y is the average of the y component of fluid velocity through the left-hand face. The corresponding velocity through the right-hand face will be $(v_y + (\partial v_y/\partial y)\,\Delta y)$ so that the flow through this face is

$$\left[\rho_m v_y + \frac{\partial(\rho_m v_y)}{\partial y}\,\Delta y\right]\Delta x\,\Delta z$$

The net outward flow in the y direction is therefore

$$\frac{\partial(\rho_m v_y)}{\partial y}\,\Delta x\,\Delta y\,\Delta z$$

Similarly the net outward flow in the z direction is

$$\frac{\partial(\rho_m v_z)}{\partial z}\,\Delta x\,\Delta y\,\Delta z$$

and in the x direction it is

$$\frac{\partial(\rho_m v_x)}{\partial x}\,\Delta x\,\Delta y\,\Delta z$$

The total net outward flow, considering all three directions, is then

$$\left[\frac{\partial(\rho_m v_x)}{\partial x}+\frac{\partial(\rho_m v_y)}{\partial y}+\frac{\partial(\rho_m v_z)}{\partial z}\right]\Delta x\,\Delta y\,\Delta z$$

The net outward flow per unit volume is

$$\frac{\partial(\rho_m v_x)}{\partial x}+\frac{\partial(\rho_m v_y)}{\partial y}+\frac{\partial(\rho_m v_z)}{\partial z}=\text{div}\,(\rho_m \mathbf{v})$$

This is the *divergence* of the fluid at the point x, y, z. Evidently for an *incompressible* fluid the div $(\rho_m\mathbf{v})$ always equals zero. An incompressible fluid cannot diverge from, nor converge toward, a point.

The case of a compressible fluid or gas such as steam is different. When the valve on a steam boiler is opened, there is a value for the divergence at each point within the boiler. There is a net outward flow of steam for each elemental volume. In this case the divergence has a positive value. On the other hand, when an evacuated light bulb is broken, there is momentarily a negative value for divergence in the space that was formerly the interior of the bulb.

Curl. The concept of curl or rotation of a vector quantity is clearly illustrated in the stream flow problems. Figure 1-7 shows a stream on the surface of which floats a leaf (in the x-y plane).

If the velocity at the surface is entirely in the y direction and is uniform over the surface, there will be no rotational motion of the leaf but only a translational motion downstream. However, if there are eddies or vortices in the stream flow, there will in general be a rotational as well as translational motion. The rate of rotation or angular velocity at any point is a measure of the *curl* of the velocity of the water at that point. In this case, where the rotation is about the z axis, the curl of **v** is in the z direction and is designated by $\text{curl}_z\, \mathbf{v}$. A positive value of $\text{curl}_z\, \mathbf{v}$ denotes a rotation from x to y, that is a counterclockwise rotation. From Fig. 1-7b it is seen that a positive value for $\partial v_y/\partial x$ will tend to rotate the leaf in a counterclockwise direction, whereas a positive value for $\partial v_x/\partial_y$ will tend to produce a clockwise rotation. The rate of rotation about the z axis is therefore proportional to the *difference* between these two quantities. By definition of the curl in rectangular co-ordinates,

$$\text{curl}_z\, \mathbf{v} = \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right)$$

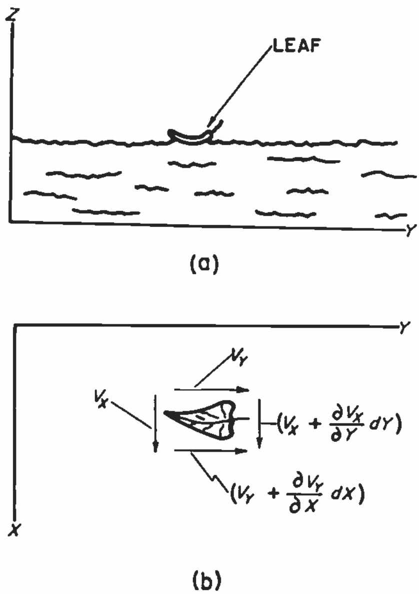


FIG. 1-7. Rotation of a floating leaf.

More generally, considering any point within the fluid, there may be rotations about the x and y axes as well. The corresponding components of the curl are given by

$$\text{curl}_x\, \mathbf{v} = \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}$$

$$\text{curl}_y\, \mathbf{v} = \frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}$$

A rotation about any axis can always be expressed as the sum of the component rotations about the x, y, and z axes. Since the rotations have direction as well as magnitude this will be a vector sum and the resultant rate of rotation or angular velocity will be proportional to

$$\text{curl}\, \mathbf{v} = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right)\mathbf{k}$$

The direction of the resultant curl is the axis of rotation.

It should be observed that it is not necessary to have circular motion or eddies in order to have a value for curl. In the example of Fig. 1-7, if v_x were everywhere zero but v_y were greater in midstream than near the bank (that is, v_y varies in the x direction), the leaf would tend to rotate and there would be a value for curl given by

$$\text{curl}_z\ \mathbf{v} = \frac{\partial v_y}{\partial x}$$

1.04 Vector Relations in Other Co-ordinate Systems. In order to simplify the application of the boundary conditions in particular problems, it is often desirable to express the various vector relations in co-ordinate systems other than the rectangular or Cartesian system. Two other systems are of great importance. They are cylindrical and spherical polar systems. The expressions for gradient, divergence, curl, and so on, in these co-ordinate systems can be obtained directly by setting up a mathematical statement for the particular physical operation to be carried out.

Cylindrical Co-ordinates. The *gradient* of a scalar quantity is the space rate of change of that quantity. In cylindrical co-ordinates the elements of length along the three co-ordinate axes* are $d\rho$, $\rho\, d\phi$, and dz (Fig. 1-8). The respective components of the gradient of a scalar V are therefore

$$\text{grad}_\rho\ V = \frac{\partial V}{\partial \rho}, \qquad \text{grad}_\phi\ V = \frac{\partial V}{\rho\, \partial \phi}, \qquad \text{grad}_z\ V = \frac{\partial V}{\partial z} \tag{1-30}$$

If the unit vectors are designated by $\mathbf{u}_\rho$, $\mathbf{u}_\phi$, and $\mathbf{u}_z$, the gradient may be written in cylindrical co-ordinates as

$$\text{grad}\ V = \frac{\partial V}{\partial \rho}\mathbf{u}_\rho + \frac{\partial V}{\rho\, \partial \phi}\mathbf{u}_\phi + \frac{\partial V}{\partial z}\mathbf{u}_z \tag{1-31}$$

The *divergence* was found to represent the net outward flow per unit volume. The expression for it can be obtained as before by determining the flow through the six surfaces of an elemental

* The symbol ρ is used for radial distance in cylindrical co-ordinates ($\rho = \sqrt{x^2 + y^2}$) in order to distinguish it from r, the radial distance in spherical co-ordinates ($r = \sqrt{x^2 + y^2 + z^2}$). This is necessary because these co-ordinate systems are often used together in problems. No confusion with ρ_m, used for mass density, or ρ, used for volume charge density, is anticipated. If it should ever happen that volume charge density and radial distance in cylindrical co-ordinates appear in the same equation the symbol ρ_v can be used for volume charge density. This is consistent with the notation ρ_s for *surface* charge density, which is used later. When no confusion results, volume charge density is represented by the symbol ρ (without subscript).

volume. Considering an incompressible fluid, the mass density ρ_m will be a constant and so this factor can be dropped from the

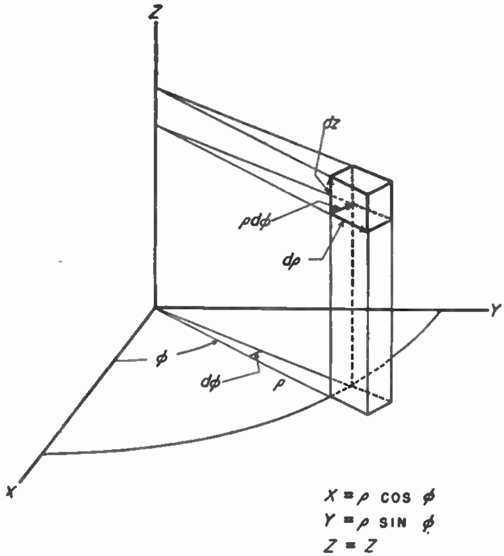


Fig. 1-8. A cylindrical co-ordinate system.

expressions. Then in the ρ direction the flow in through the left-hand face is proportional to

$$v_\rho \rho \, d\phi \, dz$$

The flow out of the right-hand face is proportional to

$$\left(v_\rho + \frac{\partial v_\rho}{\partial \rho} d\rho\right)(\rho + d\rho)\, d\phi \, dz$$

The difference between these two quantities (neglecting the second-order differential) is

$$\frac{\partial v_\rho}{\partial \rho} \rho \, d\rho \, d\phi \, dz + \frac{v_\rho}{\rho} \rho \, d\rho \, d\phi \, dz = \frac{1}{\rho} \frac{\partial(\rho v_\rho)}{\partial \rho} d\rho \, \rho \, d\phi \, dz$$

In the ϕ direction the difference is $(\partial v_\phi/\rho \partial \phi)\, d\rho \, \rho \, d\phi \, dz$, and in the z direction it is $(\partial v_z/\partial z)\, d\rho \, \rho \, d\phi \, dz$. The net flow out is therefore proportional to

$$\left(\frac{1}{\rho} \frac{\partial(\rho v_\rho)}{\partial \rho} + \frac{\partial v_\phi}{\rho \, \partial \phi} + \frac{\partial v_z}{\partial z}\right) d\rho \, \rho \, d\phi \, dz$$

The net flow out per unit volume is proportional to

$$\operatorname{div} \mathrm{v} = \frac{1}{\rho}\frac{\partial(\rho v_\rho)}{\partial \rho} + \frac{\partial v_\phi}{\rho\,\partial \phi} + \frac{\partial v_z}{\partial z}$$

In terms of any vector **A**, the divergence in cylindrical co-ordinates is

$$\operatorname{div} \mathbf{A} = \frac{1}{\rho}\frac{\partial(\rho A_\rho)}{\partial \rho} + \frac{\partial A_\phi}{\rho\,\partial \phi} + \frac{\partial A_z}{\partial z} \tag{1-32}$$

Curl. The three cylindrical components of curl are:

$$\operatorname{curl}_\rho \mathbf{A} = \frac{\partial A_z}{\rho\,\partial \phi} - \frac{\partial A_\phi}{\partial z} \tag{1-33a}$$

$$\operatorname{curl}_\phi \mathbf{A} = \frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho} \tag{1-33b}$$

$$\operatorname{curl}_z \mathbf{A} = \frac{1}{\rho}\left[\frac{\partial}{\partial \rho}(\rho A_\phi) - \frac{\partial A_\rho}{\partial \phi}\right] \tag{1-33c}$$

In chap. 4, the expression for curl in rectangular co-ordinates will be developed in connection with Ampere's law. The expression for curl in cylindrical co-ordinates can be derived in exactly the same manner.

The Laplacian Operator. The operator $\nabla^2 = \nabla \cdot \nabla$ is the divergence of the gradient of the (scalar) quantity upon which ∇^2 operates. Carrying out this operation, it will be found that in cylindrical co-ordinates,

$$\nabla^2 V = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial V}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2} \tag{1-34}$$

For a *vector*, the symbol $\nabla \cdot \nabla\mathbf{A}$ so far has no meaning except in Cartesian co-ordinates where it has been defined (see footnote following eq. 27). The definition for the Laplacian of a vector can be generalized for other orthogonal co-ordinate systems by writing,

$$\nabla^2\mathbf{A} = \nabla \cdot (\nabla\mathbf{A}) \tag{1-35}$$

where $\nabla\mathbf{A}$ is defined to mean* (in cylindrical co-ordinates)

$$\nabla\mathbf{A} = \mathbf{u}_\rho\frac{\partial \mathbf{A}}{\partial \rho} + \mathbf{u}_\phi\frac{\partial \mathbf{A}}{\rho\,\partial \phi} + \mathbf{u}_z\frac{\partial \mathbf{A}}{\partial z} \tag{1-36}$$

* The definitions of the symbol $\nabla\mathbf{A}$, given by eqs. (36) and (42), have significance only when $\nabla\mathbf{A}$ is associated with the divergence operation as in eqs. (35) and (41).

Spherical Polar Co-ordinates. In the spherical polar co-ordinate system the elements of length along the three co-ordinates are dr, $r\,d\theta$, and $r \sin\theta\, d\phi$ (Fig. 1-9).

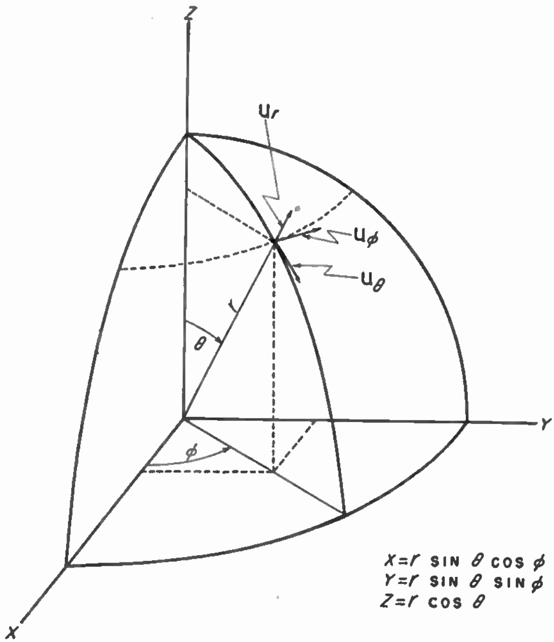


FIG. 1-9. A spherical co-ordinate system.

Gradient. The three components of the gradient in spherical co-ordinates are

$$\text{grad}_r\, V = \frac{\partial V}{\partial r} \qquad \text{grad}_\theta\, V = \frac{\partial V}{r\,\partial\theta} \qquad \text{grad}_\phi\, V = \frac{1}{r\sin\theta}\frac{\partial V}{\partial\phi} \quad (1\text{-}37)$$

Divergence. The expression for divergence in spherical polar co-ordinates is

$$\text{div}\,\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi} \quad (1\text{-}38)$$

Curl. The three spherical polar components for the curl of a vector are

$$\text{curl}_r\,\mathbf{A} = \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta A_\phi) - \frac{\partial A_\theta}{\partial\phi}\right] \quad (1\text{-}39a)$$

$$\text{curl}_\theta\, \mathbf{A} = \left[\frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{1}{r}\frac{\partial(rA_\phi)}{\partial r}\right] \tag{1-39b}$$

$$\text{curl}_\phi\, \mathbf{A} = \frac{1}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right] \tag{1-39c}$$

The Laplacian in Spherical Polar Co-ordinates. For a scalar V,

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} \tag{1-40}$$

The Laplacian of a vector quantity is defined by

$$\nabla^2\mathbf{A} = \nabla\cdot(\nabla\mathbf{A}) \quad (\nabla\cdot\nabla)A \tag{1-41}$$

where in spherical co-ordinates, $\nabla\mathbf{A}$ is defined as

$$\nabla\mathbf{A} = \mathbf{u}_r\frac{\partial\mathbf{A}}{\partial r} + \mathbf{u}_\theta\frac{\partial\mathbf{A}}{r\,\partial\theta} + \mathbf{u}_\phi\frac{1}{r\sin\theta}\frac{\partial\mathbf{A}}{\partial\phi} \tag{1-42}$$

is undefined

Problem 2. In the illustration of the leaf floating on the surface of the water (the x-y plane) show that for a very small circular leaf, curl **v** is equal to twice the angular velocity of rotation of the leaf, that is that

$$\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) = 2\frac{d\theta}{dt}$$

(Suggestion: Assume that the tangential force on the leaf per unit area at any point is a constant times the relative velocity between leaf and water at that point. The sum of all the torques on the leaf must be zero.)

Problem 3. For a two-dimensional system in which $r = \sqrt{x^2 + y^2}$ determine $\nabla^2 V$ (use rectangular co-ordinates and then check in cylindrical co-ordinates) (a) when $V = 1/r$, (b) when $V = \ln 1/r$.

Problem 4. Repeat problem 3 for a three-dimensional system in which $r = \sqrt{x^2 + y^2 + z^2}$ (use rectangular co-ordinates, and check with spherical co-ordinates).

1.05 Units and Dimensions. Although several systems of units are used in electromagnetic theory, most engineers now use some form of the practical meter-kilogram-second (MKS) system. It is to be expected that the marked advantages of this system will prompt its universal adoption.

The existence of the large number of systems of electric and magnetic units requires some explanation. The units used to describe electric and magnetic phenomena can be quite arbitrary, and a complete system of units can be built up from any of a large number of starting points. It is necessary only to define the units of length, mass, time, and one electrical quantity (such as charge, current, permeability or resistance) in order to have the *basic units* from which all other required units can be derived. Unfortunately, in the original CGS (centimeter-gram-second) systems the defined units of length and mass were so small that the derived electrical units were unsuitable for practical use. It was found necessary to set up the so-called practical system with units that were related to the corresponding CGS units by some power of 10 (volt, ampere, ohm, and so on). In 1901 Professor Giorgi showed that this practical series could be made part of a complete system, based upon the meter, kilogram, and second, provided that μ_v, the magnetic permeability of a vacuum or free space, is given the value 10^{-7} instead of unity as in the CGS system. The resulting (MKS) system has the advantage that it utilizes units already in use in electrical engineering. In addition, it is a complete and self-consistent system.

The problem of selecting a suitable system of electric and magnetic units has been further complicated by the question of *rationalization*. As was pointed out by Heaviside, the CGS system is unrationalized in that the factor 4π occurs in the wrong places, that is, where logically it is not expected. It would be expected that 4π would occur in problems having spherical symmetry, 2π in problems having circular or cylindrical symmetry and no π in problems involving rectangular shapes. In the ordinary CGS system that is not the case, and Heaviside proposed to *rationalize* the system. However his proposal involved changing the values of the volt, ampere, ohm, and so on, by nonintegral values and so was not considered feasable for practical reasons. It was pointed out later that, if the permeability μ_v of a vacuum or free space were changed from 1 to 4π in the CGS system, rationalization could be effected without changing the magnitude of the practical units. In the rationalized MKS system of units this requires that μ_v have the value of $4\pi \times 10^{-7}$. In any system of units the product $1/\sqrt{\mu_v \epsilon_v}$ must be equal to c, the velocity of light. This requires that in the rationalized MKS system

$$\epsilon_v = 8.854 \times 10^{-12} \approx \frac{1}{36\pi \times 10^9}$$

In the rationalized system the factor 4π occurs explicitly in Coulomb's law and in Ampere's law for the current element, but it does not occur in Maxwell's field equations. It is for this latter reason that the rationalized system is favored in electromagnetic theory.

In 1935 the Giorgi (MKS) system was adopted as the international standard, with the question of rationalization left unsettled. In this book the *rationalized* MKS system of units will be used. Consequently, the basic or defined electrical unit (the permeability of free space) will have the value $\mu_v = 4\pi \times 10^{-7}$. The common mechanical and electrical quantities as they appear in this system are listed below.

RATIONALIZED MKS SYSTEM OF UNITS

Length. The unit of length is the *meter.*

Mass. The unit of mass is the *kilogram.*

Time t. The unit of time is the *second.*

Force F. The unit of force is the *newton.* It is the force required to accelerate 1 kg at the rate of 1 meter/sec² (1 newton = 10^5 dynes).

Energy. The unit of electrical energy is the same as the unit of mechanical energy. It is the *joule.* A joule is the work done when a force of 1 newton is exerted through a distance of 1 meter (1 joule = 10^7 ergs).

Power. The unit of power is the *watt.* It represents a rate of energy expenditure of 1 joule/sec.

Absolute Permeability of Free Spacc μ_v. This basic electrical unit has the value of $4\pi \times 10^{-7}$ by definition. It has the dimensions of *henry per meter.*

Current I. The unit of current is the *ampere.* The size of the ampere is established through the experimental law of force (Ampere's law) between two very long parallel wires in free space, viz.

$$F = \frac{\mu_v I_1 I_2 L}{2\pi d}$$

where L is the length of the wires, and d is their separation. Thus an ampere is that current (flowing in each conductor) which produces a force of 2×10^{-7} newtons/m length between very long parallel wires spaced 1 meter apart in a vacuum.

Charge Q or q. The unit of charge is the *coulomb.* One ampere of current flowing for 1 sec transports 1 coulomb of charge.

Resistance R. The unit of resistance is the *ohm.* If 1 watt of power is dissipated in a resistance when 1 amp of current flows through it, the value of the resistance is 1 ohm.

Conductance G. Conductance is the reciprocal of resistance. The reciprocal ohm is known as the *mho* (or as the *siemens*).

Resistivity. The resistivity of a medium is the resistance measured between two parallel faces of a unit cube. The unit of resistivity is the *ohm meter*.

Conductivity σ. The conductivity of a medium is the reciprocal of resistivity. The unit of conductivity is the *mho/meter*.

Electromotive Force V. The unit of electromotive force (emf) or voltage is the *volt*, which is defined as 1 watt/amp. It is also equal to 1 joule/coulomb and so has the dimensions of work per unit charge. (It is not a force.)

Electric Intensity **E**. Electric intensity or electric field strength is measured in *volts/meter*. The electric intensity at any point in a medium is the electric force per unit positive charge at that point. It has the dimension newton/coulomb.

Current Density ***i***. The unit of current density is the *ampere/square meter*.

Electric Displacement Ψ. The electric displacement through a closed surface is equal to the charge enclosed by the surface. The unit of electric displacement is the *coulomb*.

Displacement Density **D**. The unit of electric displacement density (usually called just displacement density) is the *coulomb/square meter*.

Magnetic Flux Φ. The voltage V between the terminals of a loop of wire due to a changing magnetic field is related to the magnetic flux through any surface enclosed by the loop by $V = -d\Phi/dt$. The unit of magnetic flux is defined by this relation and is called the *weber*. A weber is 1 *volt · sec.*

Magnetic Flux Density **B**. The unit of magnetic flux density is the *weber/square meter*. (1 weber/sq m = 10^4 gauss)

Magnetic Intensity H. The magnetic intensity or magnetic field strength between two parallel plane sheets carrying equal and oppositely directed currents is equal to the current per meter width (amperes per meter) flowing in the sheets. The unit of magnetic intensity is the *amp/meter*.

Magnetomotive Force $\mathfrak{F}$ (or M). The magnetomotive force between two points a and b is defined as the line integral $\int_a^b \mathbf{H} \cdot d\mathbf{s}$. The unit of magnetomotive force is the *ampere*. The magnetomotive force around a *closed* path is equal to the current enclosed by the path.

Capacitance C. A conducting body has a capacitance of 1 *farad* if it requires a charge of 1 coulomb to raise its potential by 1 volt. A farad is equal to *1 coulomb/volt*.

Inductance L. A circuit has an inductance of 1 *henry* if a changing current of 1 amp/sec induces in the circuit a "back-voltage" of 1 volt. The dimensions of the henry are

$$\frac{\text{volt} \cdot \text{seconds}}{\text{ampere}} = \text{ohm} \cdot \text{seconds}$$

Dielectric Constant ϵ. In a homogeneous medium the electrical quantities **D** and **E** are related by the equation $\mathbf{D} = \epsilon\mathbf{E}$, where ϵ is the dielectric constant of the medium. It has the dimensions *farad/meter*. The dielectric constant of evacuated (free) space is

$$\epsilon_v = 8.854 \times 10^{-12} \approx \frac{1}{36\pi \times 10^9} \text{ farads/m}$$

The dielectric constant of a medium may be written as $\epsilon = \epsilon_r\epsilon_v$ where ϵ_r is a dimensionless constant known as the *relative dielectric constant* of the medium.

Permeability μ. The magnetic flux density and magnetic intensity in a homogeneous medium are related by $\mathbf{B} = \mu\mathbf{H}$ where μ is the magnetic permeability of the medium. It has the dimensions henry/meter. The permeability of free space is $\mu_v = 4\pi \times 10^{-7}$ henry/m. The permeability of a medium may be written as $\mu = \mu_r\mu_v$ where μ_r is the *relative* permeability of the medium.

Table I gives the dimensions of the units of the MKS system. In this table the dimensions of all of the units have been expressed in terms of mass M, length L, time T, and charge Q. By expressing the *dimensions* in terms of charge Q, rather than the defined unit μ, fractional exponents in the dimensional equations are avoided.

A table that can be used for converting from the MKS practical system to the CGS systems or vice versa is shown inside the back cover.

1.06 Order of Magnitude of the Units. A concept of the order of magnitude of the units of the MKS practical system can be obtained from a few examples. A meter is equal to 3.281 ft, and roughly 3 meters equal 10 ft. A kilogram is slightly more than 2 lb (1 kg = 2.205 lb). A newton is approximately the force required to lift ¼ lb. (more accurately 0.225 lb). A joule is the work done in lifting this ¼ lb weight 1 meter. To raise the weight through 1 meter in 1 sec requires the expenditure of 1 watt of power. Whereas the watt is usually thought of as a rather small unit of power (the smallest lamp in general household use requires 15 watts, and it takes 2 or 3 watts to run an electric clock), it represents a considerable amount of mechanical power. A man can do work for a 12-hour day at the rate of about 40 watts, which is less than the power required to run his wife's electric washing machine. The coulomb, which is about the amount of charge passing through a 100-watt lamp in one second would charge a sphere the size of the

TABLE I
DIMENSIONS OF UNITS IN THE MKS SYSTEM

Quantity	Symbol	MKS unit	Dimensional equivalent	Dimensions
Length	l	meter		L
Mass	m	kilogram		M
Time	t	second		T
Charge	q	coulomb		Q
Force	**F**	newton	joule per meter	MLT^{-2}
Energy	U	joule	volt-coulomb	ML^2T^{-2}
Power	W	watt	joule per second	ML^2T^{-3}
Current	I	ampere	coulomb per second	$T^{-1}Q$
Current density	**i**	ampere/square meter		$L^{-2}T^{-1}Q$
Charge density (volume)	ρ (or ρ_v)	coulomb/cubic meter		$L^{-3}Q$
Charge density (surface)	ρ_s	coulomb/square meter		$L^{-2}Q$
Resistance	R	ohm	volt per ampere	$ML^2T^{-1}Q^{-2}$
Conductivity	σ	mho/meter		$M^{-1}L^{-3}TQ^2$
Electromotive force	V	volt	joule per coulomb	$ML^2T^{-2}Q^{-1}$
Electric intensity	**E**	volt/meter	newton per coulomb	$MLT^{-2}Q^{-1}$
Capacitance	C	farad	coulomb per volt	$M^{-1}L^{-2}T^2Q^2$
Dielectric constant	ϵ	farad/meter		$M^{-1}L^{-3}T^2Q^2$
Electric displacement	Ψ	coulomb		Q
Electric displacement density	**D**	coulomb/square meter		$L^{-2}Q$
Magnetic flux	Φ	weber	volt-second	$ML^2T^{-1}Q^{-1}$
Magnetic flux density	**B**	weber/square meter		$MT^{-1}Q^{-1}$
Magnetomotive force	$\mathcal{F}$ (or M)	ampere (turn)		$T^{-1}Q$
Magnetic intensity	**H**	ampere (turn)/meter		$L^{-1}T^{-1}Q$
Inductance	L	henry	ohm-second	ML^2Q^{-2}
Permeability	μ	henry/meter		MLQ^{-2}

earth to about 1400 volts. If it were possible to place a coulomb of charge on each of two small spheres placed 1 meter apart, the force between them would be 9×10^9 newtons, or about the force required to lift a million tons. The farad is a large unit of capacitance, and the terms microfarad (10^{-6} f) and micro-microfarad (10^{-12} f) are in common use. The filter condensers on a radio set are usually 8 or 16 μf (microfarads). The capacitance of a sphere 1 cm in radius is approximately 1 $\mu\mu f$ (micro-microfarads). The inductance of the primary winding of an iron-core audio transformer may be the order of 50 henrys, whereas the inductance of the radio frequency "tuning-coils" for the broadcast band is about 300 μh (microhenrys). A weber per square meter is about one-half the saturation flux density of iron used in transformer cores.

BIBLIOGRAPHY

Kennelly, A. E., "IEC and MKS System of Units," *Elec. Eng.* (*Trans. AIEE*), **54,** 1383 (1935).

Tang, K. Y., "The Giorgi (MKS) System of Electric Units," *Engineering Experiment Station News*, Ohio State University, October, 1939.

Hessler, V. P., "Teaching Electricity and Magnetism," *Trans. AIEE*, **65,** 828–833 (1946).

Pender, H., and S. R. Warren, Jr., *Electric Circuits and Fields*, McGraw-Hill Book Co., Inc., New York, 1943.

Phillips, H. B., *Vector Analysis*, John Wiley and Sons, New York, 1933.

Schelkunoff, S. A., *Applied Mathematics for Engineers and Scientists*, D. Van Nostrand Co., Inc., New York, 1948.

CHAPTER 2

ELECTROSTATICS

2.01 Introduction. The sources of electromagnetic fields are charges, and the strength of a field at any point depends upon the magnitude, position, velocity, and acceleration of the charges involved. An *ele trostatic* field can be considered as a special case of an electromagnetic field in which the sources are stationary,* so that only the magnitude and position of the charges need be considered. The study of this relatively simple case lays the foundations for solving problems of the more general time-varying electromagnetic field. In what follows it is assumed that the reader has had an elementary course covering the subject of electrostatics and has some general knowledge of the experimental facts and their theoretical interpretation. The purposes of this chapter are (1) to review the subject briefly, not as a study in itself but as an introduction to the electromagnetic field, (2) to consider the statement of the laws in the vector form, and finally (3) to state the required relations in the MKS system of units. It is usually much simpler to derive all relations directly in the new unit system rather than to try to use conversion factors to convert from the older esu's and emu's of the CGS system.

2.02 Fundamental Relations of the Electrostatic Field. *Coulomb's Law.* It is found experimentally that between two charged bodies there exists a force that tends to push them apart or pull them together, depending on whether the charges on the bodies are of like or opposite sign. If the two bodies are spheres whose radii are very small compared with their distance apart, and if the spheres are sufficiently remote from conducting surfaces and from other

* Individual charges (e.g., electrons) are of course never stationary, having random velocities, which depend among other things upon the temperature. This statement regarding stationary sources simply means that when any elemental *macroscopic* volume is considered, the *net* movement of charge through any face of the volume is zero.

dielectric media (more technically if the spheres are immersed in an infinite homogeneous insulating medium), the magnitude of the force between them due to their charges obeys an inverse square law. That is

$$F = \frac{q_1 q_2}{kr^2} \tag{2-1}$$

where q_1 is the net charge on one sphere, q_2 the net charge on the other. This is Coulomb's law of force. In the CGS electrostatic system of units the constant k is arbitrarily put equal to unity for a vacuum and relation (1) is used to define the unit of charge for the electrostatic system of units. However, in the MKS system the unit of charge has already been determined from other considerations, and since units of length and force have also been defined, the constant k can be determined from experiment. In order to *rationalize* the units and so leave Maxwell's field equations free from the factor 4π, it is convenient to show a factor 4π explicitly in the constant k and write

$$k = 4\pi\epsilon$$

The "constant" ϵ depends upon the medium or dielectric in which the charges are immersed. It is called the *dielectric constant* (or capacitivity) of the medium. For free space, that is for a vacuum, but also very closely for air, the value of ϵ is

$$\epsilon_v = 8.854 \times 10^{-12} \qquad \text{f/m} \tag{2-2}$$

To a very good approximation (the same approximation involved in writing the velocity of light as $c \approx 3 \times 10^8$ meter/sec) the value of ϵ_v is given by

$$\epsilon_v \approx \frac{1}{36\pi \times 10^9} \tag{2-3}$$

The subscript, v, indicates that this is the dielectric constant of a vacuum or free space. For other media the value of ϵ will be different. Then Coulomb's law in MKS units is

$$F = \frac{q_1 q_2}{4\pi\epsilon r^2} \qquad \text{newtons} \tag{2-4}$$

The direction of the force is along the line joining the two charges.

Electric Intensity **E.** If a small probe charge δq is located at any point near a second fixed charge q, the probe charge experiences a force, the magnitude and direction of which will depend upon its location with respect to the charge q. About the charge q there is said to be a field of electric intensity **E**, and the magnitude of **E** at any point is measured simply as the *force per unit charge* at that point. The direction of **E** is the direction of the force on a positive probe charge, and is along the outward radial from the (positive) charge q.

From equation (4) the magnitude of the force on δq will be

$$SF = \frac{q\,\delta q}{4\pi\epsilon r^2} \tag{2-5}$$

and the magnitude of the electric intensity is

$$E = \frac{q}{4\pi\epsilon r^2} \tag{2-6}$$

The force on the probe charge is dependent upon the strength of the probe charge, but the electric intensity is not. If the charge on the probe is allowed to approach zero, then the force acting on it does also, but the force *per unit charge* remains constant; that is, the electric field due to the charge q is considered to exist, whether or not there is a probe charge to detect its presence.

The direction, as well as the magnitude, of the electric intensity about a point charge is indicated by writing the vector relation

$$\mathbf{E} = \frac{q}{4\pi\epsilon r^2}\,\mathbf{u}_r \tag{2-7}$$

where $\mathbf{u}_r$ is a unit vector along the outward radial from the charge q.

Electric Displacement Ψ *and Displacement Density* **D**. It is seen from eq. (7) that at any particular point the electric intensity **E** depends not only upon the magnitude and position of the charge q, but also upon the dielectric constant of the medium (air, oil, and others) in which the field is measured. It is desirable to associate with the charge q a second electrical quantity that will be *independent* of the medium involved. This second quantity is called *electric displacement* or *electric flux* and is designated by the symbol Ψ. An understanding of what is meant by electric displacement can be gained by recalling Faraday's experiments with concentric spheres. A sphere with charge Q was placed within, but not touching, a larger hollow sphere. The outer sphere was "earthed" momentarily, and then the inner sphere was removed. The charge remaining on the outer sphere was then measured. This charge was found to be equal (and of opposite sign) to the charge on the inner sphere *for all sizes of the spheres* and *for all types of dielectric media* between the spheres. Thus it could be considered that there was an *electric displacement* from the charge on the inner sphere through the medium to the outer sphere, the amount of this displacement depending only upon the magnitude of the charge Q. In MKS units the displacement Ψ is equal in magnitude to the charge that produces it, that is

$$\Psi = Q \qquad \text{coulombs} \qquad (2\text{-}8)$$

For the case of an isolated point charge q remote from other bodies the outer sphere is assumed to have infinite radius. The electric displacement per unit area or *electric displacement density* **D** at any point on a spherical surface of radius r centered at the isolated charge q will be

$$D = \frac{\Psi}{4\pi r^2} = \frac{q}{4\pi r^2} \qquad \text{coulomb/sq m} \qquad (2\text{-}9)$$

The displacement per unit area at any point depends upon the *direction* of the area. Displacement density **D** is therefore a vector quantity, its direction being taken as that direction of the normal to the surface element which makes the displacement through the element of area a maximum. For the case of displacement from an isolated charge this direction is along the radial from the charge and is the same as the direction of **E**. Therefore the vector relation corresponding to (9) is

$$\mathbf{D} = \frac{q}{4\pi r^2}\,\mathbf{u}_r \qquad (2\text{-}10)$$

Comparing eqs. (7) and (10) shows that **D** and **E** are related by the vector relation

$$\mathbf{D} = \epsilon\mathbf{E} \qquad (2\text{-}11)$$

Equation (11) is true in general for all *isotropic* media. For certain crystalline media, the dielectric constant ϵ is different for different directions of the electric field, and for these media **D** and **E** will generally have different directions. Such substances are said to be *anisotropic*. In this book only *homogeneous isotropic* media will be considered. For these ϵ is constant, that is, independent of position (homogeneous) and independent of the magnitude and direction of the electric field.

It is possible to measure the displacement density at a point by the following experimental procedure. Two small thin metallic disks are put in contact and placed together at the point at which **D** is to be determined. They are then separated and removed from the field, and the charge upon them is measured. The charge per unit area is a direct measure of the component of **D** in the direction of the normal to the disks. If the experiment is performed for all possible orientations of the disks at the point in question, the direction (of the normal to the disks) that results in maximum charge on the disks is the direction of **D** at that point, and this maximum value of charge per unit area is the magnitude of **D**.

Lines of Force and Lines of Flux. In an electric field *a line of electric force* is a curve drawn so that at every point it has the direction of the electric intensity. The number of lines per unit area is made proportional to the magnitude of the electric intensity, E. A line of *electric flux* is a curve drawn so that at every point it has the direction of the electric flux density or displacement density. The number of flux lines per unit area is used to indicate the magnitude of the displacement density, D. In homogeneous isotropic media lines of force and lines of flux always have the same direction.

2.03 Gauss's Law. Gauss's law states that *the total displacement or electric flux through any closed surface surrounding charges is equal to the amount of charge enclosed.* This may be regarded as a generalization of a fundamental experimental law (recall Faraday's experiments) or it may be deduced from Coulomb's inverse-square law, and the relation $\mathbf{D} = \epsilon\mathbf{E}$ (now used to define **D**).

Consider a point charge q located in a homogeneous isotropic medium whose dielectric constant is ϵ. The electric intensity at any point a distance r from the charge q will be

$$\mathbf{E} = \frac{q}{4\pi\epsilon r^2}\,\mathbf{u}_r$$

and the displacement density or electric flux density at the same point will be

$$\mathbf{D} = \epsilon\mathbf{E} = \frac{q}{4\pi r^2}\,\mathbf{u}_r$$

Now consider the displacement through some surface enclosing the charge (Fig. 2-1). The displacement or electric flux through the element of surface da is

$$d\Psi = D\,da\,\mathrm{ccs}\,\theta \qquad (2\text{-}12)$$

where θ is the angle between $\mathbf{D}$ and the normal to da. From the figure it is seen that $da \cos\theta$ is the projection of da normal to the radius vector. Therefore, by definition of a solid angle,

$$da \cos\theta = r^2\,d\Omega \qquad (2\text{-}13)$$

where $d\Omega$ is the solid angle subtended at q by the element of area da.

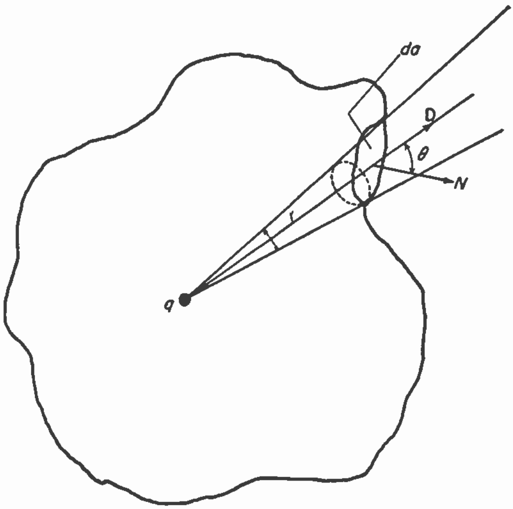


FIG. 2-1. Displacement through a surface enclosing a charge.

The total displacement through the surface is obtained by integrating eq. (12) over the entire surface.

$$\Psi = \oint D\,da\cos\theta \qquad (2\text{-}14)$$

(The circle on the integral sign indicates that the surface of integration is a closed surface.) Using eq. (13) the displacement is given by

$$\Psi = \oint Dr^2\,d\Omega$$

and substituting for D from (9)

$$\Psi = \frac{q}{4\pi}\oint d\Omega \qquad (2\text{-}15)$$

But the total solid angle subtended at q by the *closed surface* is

$$\Omega = \oint d\Omega = 4\pi \text{ solid radians.}$$

Therefore from (15) the total displacement through the *closed surface* will be

$$\Psi = q \tag{2-16}$$

If there are a number of charges within the volume enclosed by the surface the total displacement through the surface will be equal to the sum of all the charges. If the charge is continuously* distributed throughout the volume with a charge density ρ (coulombs per cubic meter), the total displacement through the surface is

$$\Psi = \int_{\text{vol}} \rho \, dV \tag{2-17}$$

where the right-hand side represents the total charge contained within the volume.

It is often desirable to state the above relations in vector form. By definition of the dot product, the expression $D\,da \cos\theta$ in eq. (12) can be written as $\mathbf{D} \cdot \mathbf{da}$. In this case the element of area $\mathbf{da}$ is considered to be a vector quantity having the magnitude da and the direction of the normal to its surface. Then eq. (14) would be written

$$\Psi = \oint \mathbf{D} \cdot \mathbf{da} \tag{2-18}$$

When $\mathbf{da}$ is a part of a *closed* surface as it is here the direction of the *outward* normal is taken to be positive. The right-hand side of eq. (18) is the integration over a closed surface of the normal component of the displacement density, that is, it is the total (outward) electric displacement or electric flux through the surface.

Combining eqs. (17) and (18) the vector statement of Gauss's law is

$$\oint_S \mathbf{D} \cdot \mathbf{da} = \int_{\text{vol}} \rho \, dV \tag{2-19}$$

* Actual charge distributions consist of aggregations of discrete particles or corpuscles. However since there will always be an enormous number of these microscopic particles in any *macroscopic* element of volume ΔV, it is permissable to speak of the charge density ρ where $\rho = \Delta q/\Delta V$ is the charge per unit volume in elemental volume ΔV. Thus by "charge density at a point" is really meant the charge per unit volume in the elemental volume ΔV containing the point. Although ΔV may be made very small, it is always kept large enough to contain many charges.

In words, *the net outward displacement through a closed surface is equal to the charge contained in the volume enclosed by the surface.*

2.04 Electric Field Due to Several Charges. When a test charge δq is located at a point p in the field of a single charge q it experiences a force $\mathbf{F}$ that is given by

$$\mathbf{F} = \frac{q\,\delta q}{4\pi\epsilon r^2}\,\mathbf{u}_r \qquad \text{newtons} \qquad (2\text{-}20)$$

The unit vector $\mathbf{u}_r$ indicates that the direction of the force is along the radius vector from the charge q to the point p. By definition,

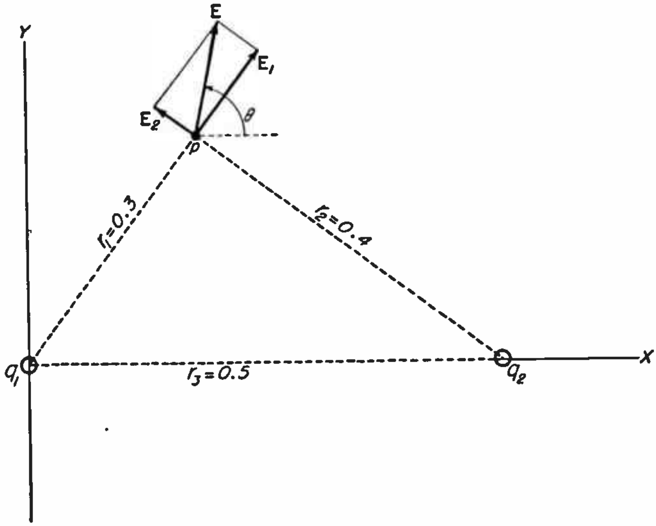


FIG. 2-2. Vector addition of fields.

the electric intensity $\mathbf{E}$ at the point p is the *force per unit charge* and has the same direction as $\mathbf{F}$, so that

$$\mathbf{E} = \frac{q}{4\pi\epsilon r^2}\,\mathbf{u}_r$$

When there are several charges present, each charge will exert a force on the test charge at p, the magnitude and direction of which is given by (20). The resultant force on δq will be the *vector sum* of the individual forces, the addition taking into account the direction as well as magnitude of the forces. Correspondingly the electric intensity at the point p will be the vector sum of the electric intensities because of each charge acting alone. If $q_1, q_2, q_3, \ldots q_n$ are charges located at distances $r_1, r_2, r_3, \ldots r_n$ from the point p the electric intensity at p will be

$$\mathbf{E} = \frac{1}{4\pi\epsilon}\left(\frac{q_1}{r_1^2}\mathbf{u}_{r_1} + \frac{q_2}{r_2^2}\mathbf{u}_{r_2} + \frac{q_3}{r_3^2}\mathbf{u}_{r_3} + \cdots + \frac{q_n}{r_n^2}\mathbf{u}_{r_n}\right)$$

$$= \frac{1}{4\pi\epsilon}\sum_{i=1}^{i=n}\frac{q_i}{r_i^2}\mathbf{u}_{r_i}$$

Example 1: *Electric Field of Two Charges (Method 1).* Determine the electric intensity at the point p in Fig. 2-2 due to the charges q_1 and q_2

$$q_1 = 1 \times 10^{-9} \text{ coulomb}$$
$$q_2 = 8 \times 10^{-10} \text{ coulomb}$$
$$r_1 = 0.3 \text{ meter}$$
$$r_2 = 0.4 \text{ meter}$$

The magnitudes of the individual intensities E_1 and E_2 are

$$|E_1| = \frac{q_1}{4\pi\epsilon r_1^2} = \frac{36\pi \times 10^9 \times 10^{-9}}{4\pi \times 0.3^2} = 100 \text{ volt/m}$$

$$|E_2| = \frac{q_2}{4\pi\epsilon r_2^2} = \frac{36\pi \times 10^9 \times 8 \times 10^{-10}}{4\pi \times 0.4^2} = 45 \text{ volt/m}$$

In order to add these intensities vectorally it is convenient to use the components in the x and y directions. From the geometry of the 3, 4, 5 triangle:

(a) The x component of E_1 is $100 \times 3/5 = 60$
(b) The x component of E_2 is $-45 \times 4/5 = -36$
(c) The y component of E_1 is $100 \times 4/5 = 80$
(d) The y component of E_2 is $45 \times 3/5 = 27$

The total x component of E is

$$E_x = 60 - 36 = 24$$

and the total y component is

$$E_y = 80 \times 27 = 107$$

The resultant electric intensity has a magnitude

$$E = \sqrt{24^2 + 107^2} = 110$$

The angle θ between the direction of E and the x axis is given by

$$\theta = \tan^{-1} 107/24 = 77.4°$$

2.05 The Potential Function. An electric field is a field of force, and a force field can be described in an alternative manner from that given above. If a body being acted upon by a force is moved from one point to another, work will be done on or by the body. If there is no mechanism by which the energy represented by this work can be dissipated, then the field is said to be *conservative*, and the energy must be stored in either the potential or kinetic form. If a charge is moved in a static electric field or a mass is moved in a gravitational field and no friction is present in the region, then no energy is dissipated. Hence these are examples of conservative fields. If some point is taken as a reference or zero point the field of force can be described by the work that must be done in moving the body from the reference point up to any point in the field. A reference point that is commonly used is a point at infinity. For example, if a small body has a charge q and a second body with a small test charge δq is moved from infinity along a radius line to a point p at a distance R from the charge q, then the work done *on* the system in moving the test charge against the force **F** will be

$$\text{Work} = -\int_{\infty}^{R} F\,dr.$$

and since S$F = \dfrac{q\,\delta q}{4\pi\epsilon r^2}$

$$\text{Work on test charge} = -\frac{q\,\delta q}{4\pi\epsilon}\int_{\infty}^{R}\frac{1}{r^2}\,dr = \frac{q\,\delta q}{4\pi\epsilon R}$$

The work done on the test charge *per unit charge* is

$$V = \frac{q \times 1}{4\pi\epsilon R} = \frac{q}{4\pi\epsilon R} \tag{2-21}$$

V is called the *potential* at the point p due to the charge q. Because it is a scalar quantity, having only magnitude and no direction, it is often called the *scalar potential.*

In a *conservative* field the work done in moving from one point to another is independent of the path. This is easily proven. If it were not independent of the path and a charge were moved from point P_1 to point P_2 over one path, and then from point P_2 back to point P_1 over a second path, the work done *on* the body on one path could be different from the work done *by* the body on the second path. If this were true, a net (positive or negative) amount of work would be done when the body returned to its original position P_1. In a conservative field there is no mechanism for dissipating energy corresponding to positive work done and no source from which energy could be absorbed if the work were negative. Hence, it is apparent that the assumption that the work done is different over two paths is untenable, and so the work must be independent of the path. Thus for every point in the static electric field there corresponds one and only one scalar value of the work done in bringing the charge from infinity up to the point in question by any possible path. This scalar value at any point is called the *potential* of that point. The potential* is measured in volts where 1 volt = 1 joule per coulomb.

If two points P_1 and P_2 are separated an infinitesimal distance δ_s and the potential at P_1 is V_1, whereas that at P_2 is $V_1 + \delta V$, it is apparent that the work done in moving a unit charge from point P_1 to point P_2 will be

$$W = V_1 - (V_1 + \delta V) = E_s \, \delta s$$

where E_s is the component of the electric intensity in the direction of δs

$$\therefore \; -\delta V = E_s \, \delta s \tag{2-22}$$

The three components of $\mathbf{E}$ in the x, y, and z direction can be obtained from eq. (22)

$$E_x = -\frac{\partial V}{\partial x}, \qquad E_y = -\frac{\partial V}{\partial y}, \qquad E_z = -\frac{\partial V}{\partial z} \tag{2-23}$$

The three scalar eqs. (23) can be written in one vector equation

$$\mathbf{E} = -\frac{\partial V}{\partial x}\mathbf{i} - \frac{\partial V}{\partial y}\mathbf{j} - \frac{\partial V}{\partial z}\mathbf{k} \tag{2-24}$$

* In electrostatics the term potential or potential difference and voltage are used interchangeably. For time-varying electromagnetic fields, potential, as defined here, has no meaning. However, the voltage between two points a and b, defined by

$$V_{ab} = \int_a^b \mathbf{E} \cdot d\mathbf{s}$$

continues to have meaning as long as the path is specified.

or in the abbreviated vector notation

$$\mathbf{E} = -\operatorname{grad} V \qquad (2\text{-}25)$$

Thus the electric intensity at any point is just the negative of the potential gradient at that point. The direction of the electric field is the direction in which the gradient is greatest or in which the potential changes most rapidly.

Equations (23) give the three components of **E** in rectangular or Cartesian co-ordinates. Very often the conditions of a problem are such that it is more simply solved in cylindrical or spherical co-ordinates. In cylindrical co-ordinates the three mutually perpendicular directions are ρ, ϕ, and z. The elemental increments of length in these directions are $d\rho$, $\rho\, d\phi$, and dz respectively. The space rates of change of potential in these directions will give the corresponding components of electric intensity, viz.,

$$E_\rho = -\frac{\partial V}{\partial \rho}, \qquad E_\phi = -\frac{\partial V}{\rho \partial \phi}, \qquad E_z = -\frac{\partial V}{\partial z} \quad \text{(cylindrical co-ordinates)}$$

In spherical polar co-ordinates, the increments of length are dr, $r\, d\theta$ and $r \sin \theta\, d\phi$ so that the space rates of change of potential and corresponding electric intensities are given by

$$E_r = -\frac{\partial V}{\partial r}, \qquad E_\theta = -\frac{\partial V}{r \partial \theta}, \qquad E_\phi = -\frac{1}{r \sin \theta}\frac{\partial V}{\partial \phi} \quad \text{(spherical co-ordinates)}$$

When the system of charges is specified and the problem is that of determining the resultant electric field due to the charges, it is often simpler to find first the potential field and then determine **E** as the potential gradient according to eqs. (25). This is so because the electric intensity is a vector quantity, and when the electric field produced by several charges is found directly by adding the intensities caused by the individual charges (as was done in the example on page 32), the addition of fields is a vector addition. This relatively complicated operation is carried out by resolving each vector quantity into (generally) three components, adding these components separately, and then combining the total values of the components to obtain the resultant field. On the other hand the potential field is a scalar field and the total potential at any point is found simply as the algebraic sum of the potentials due to each charge. If the potential is known, the electric field can be found from eq. (25).

Example 2: *Electric Field of Two Charges.* As an example let it be required to find the electric intensity at *all* points in the x-y plane due to the charges q_1 and q_2 (Fig. 2-2). By the first method this would be done by writing the expressions for the x and y components of **E** due to each of the charges and adding these separately to obtain the resultant **E**. By the second method the potential due to the charges is found first.

The potential at the point p in (Fig. 2-2) is

$$\begin{aligned} V_p &= \frac{1}{4\pi\epsilon}\left(\frac{q_1}{r_1} + \frac{q_2}{r_2}\right) \\ &= \frac{36\pi \times 10^9}{4\pi}\left(\frac{10^{-9}}{0.3} + \frac{8 \times 10^{-10}}{0.4}\right) \\ &= 48 \text{ volts} \end{aligned}$$

The potential at any point (x, y) due to charges q_1 and q_2, located respectively at (0, 0) and (0.5, 0), is

$$V = \frac{1}{4\pi\epsilon}\left(\frac{q_1}{\sqrt{x^2 + y^2}} + \frac{q_2}{\sqrt{(x - 0.5)^2 + y^2}}\right)$$

The electric intensity is obtained from V by applying eq. (23)

$$\frac{\partial V}{\partial x} = -\frac{1}{4\pi\epsilon}\left\{\frac{q_1 x}{[x^2 + y^2]^{3/2}} + \frac{q_2(x - 0.5)}{[(x - 0.5)^2 + y^2]^{3/2}}\right\}$$

$$\frac{\partial V}{\partial y} = -\frac{1}{4\pi\epsilon}\left\{\frac{q_1 y}{[x^2 + y^2]^{3/2}} + \frac{q_2 y}{[(x - 0.5)^2 + y^2]^{3/2}}\right\}$$

Therefore **E** at any point (x, y) will be

$$\mathbf{E} = \frac{1}{4\pi\epsilon}\left(\left\{\frac{q_1 x}{[x^2 + y^2]^{3/2}} + \frac{q_2(x - 0.5)}{[(x - 0.5)^2 + y^2]^{3/2}}\right\}\mathbf{i} + \left\{\frac{q_1 y}{[x^2 + y^2]^{3/2}} + \frac{q_2 y}{[(x - 0.5)^2 + y^2]^{3/2}}\right\}\mathbf{j}\right)$$

At the particular point p of (Fig. 2-2), $x = 0.18$ and $y = 0.24$. Substituting these values and the values for q_1 and q_2 gives

$$\begin{aligned} \mathbf{E}_p &= \left(\frac{1.62}{0.027} - \frac{2.30}{0.064}\right)\mathbf{i} + \left(\frac{2.16}{0.027} + \frac{1.728}{0.064}\right)\mathbf{j} \\ &= (60 - 36)\mathbf{i} + (80 + 27)\mathbf{j} \\ &= 24\mathbf{i} + 107\mathbf{j} \end{aligned}$$

This checks with the answer obtained in the previous example.

In a simple problem such as the one above there may be little advantage, if any, to using the potential method, but in the more complex problems to be considered later it will be found that the use of the scalar potential results in a real simplification of the problem.

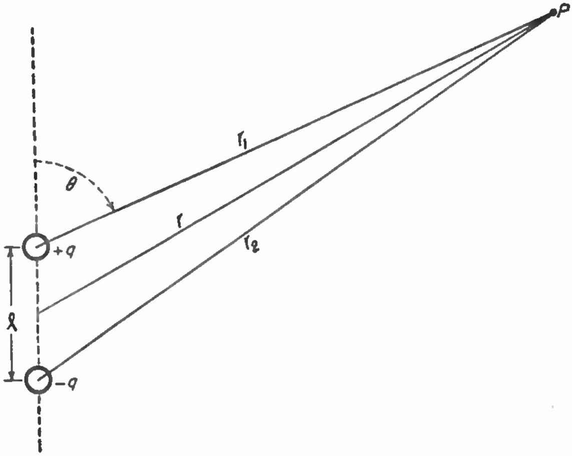


Fig. 2-3. An electric dipole.

Example 3: *Field of an Electric Dipole.* The concept of the *electric dipole* is extremely useful in electromagnetic field theory. Two equal and opposite charges of magnitude q separated by an infinitesimal distance l are said to constitute an electric dipole or *electric doublet.* The electric field due to such an arrangement can be found readily by first finding the potential V. In Fig. 2-3

$$V_p = \frac{1}{4\pi\epsilon}\left\{\frac{q}{r_1} + \frac{-q}{r_2}\right\}.$$

Because l is infinitesimally small

$$r_1 \approx r - \frac{l}{2}\cos\theta$$

$$r_2 \approx r + \frac{l}{2}\cos\theta$$

$$4\pi\epsilon V_p = \frac{q}{r - \dfrac{l}{2}\cos\theta} - \frac{q}{r + \dfrac{l}{2}\cos\theta} = \frac{ql\cos\theta}{r^2 - \dfrac{l^2}{4}\cos^2\theta}$$

$$\approx \frac{ql\cos\theta}{r^2} \quad (\text{for } l^2 \ll r^2)$$

The electric field is found from $\mathbf{E} = -\text{grad } V$. The three components in spherical co-ordinates are

$$E_r = -\frac{\partial V}{\partial r} = \frac{2ql \cos \theta}{4\pi\epsilon r^3}$$

$$E_\theta = -\frac{\partial V}{r\,\partial \theta} = \frac{ql \sin \theta}{4\pi\epsilon r^3}$$

$$E_\phi = -\frac{1}{r \sin \theta}\frac{\partial V}{\partial \phi} = 0$$

2.06 Field Due to a Continuous Distribution of Charge. The potential at a point p due to a number of charges is obtained as a simple algebraic addition of the potentials produced at the point by each of the charges acting alone. If $q_1, q_2, q_3, \ldots q_n$ are charges located at distances $r_1, r_2, r_3, \ldots r_n$, respectively, from the point p, the potential at p is given by

$$V_p = \frac{1}{4\pi\epsilon}\left(\frac{q_1}{r_1} + \frac{q_2}{r_2} + \cdots + \frac{q_n}{r_n}\right)$$

$$= \frac{1}{4\pi\epsilon}\sum_{i=1}^{i=n}\frac{q_i}{r_i}$$

If the charge is distributed continuously throughout a region, rather than being located at a discrete number of points, the region can be divided into elements of volume ΔV each containing a charge $\rho\,\Delta V$, where ρ is the *charge density* in the volume element. The potential at a point p will then be given as before by

$$V_p = \frac{1}{4\pi\epsilon}\sum_{i=1}^{i=n}\frac{\rho_i\,\Delta V_i}{r_i}$$

where r_i is the distance to p from the ith volume element. As the size of volume element chosen is allowed to become very small, the summation becomes an integration, that is

$$V_p = \frac{1}{4\pi\epsilon}\int_{\text{vol}}\frac{\rho\,dV}{r} \qquad (2\text{-}26)$$

The integration is performed throughout the volume where ρ has value.

Example 4: *Potential Distribution about Long Parallel Wires.* Determine the potential distribution about a long parallel pair of wires of

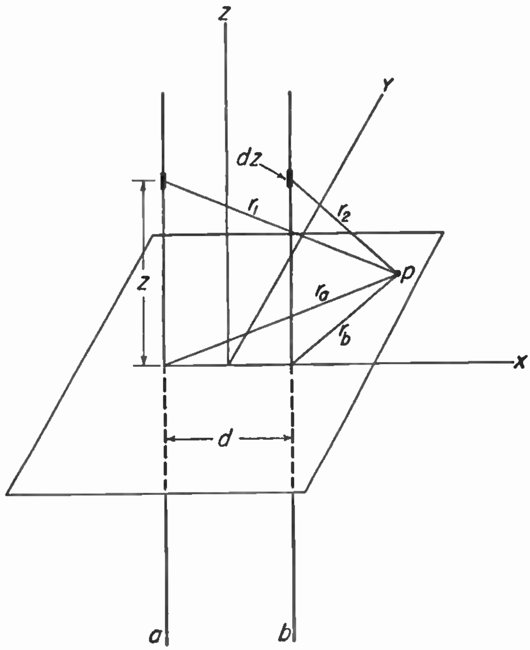


FIG. 2-4. A pair of parallel line charges.

negligible cross section when the wires have equal and opposite charges distributed along their length.

Assume that a linear charge density q' coulombs per meter is distributed along wire a and $-q'$ coulombs per meter along wire b (Fig. 2-4). Then $\rho\,dV$ becomes $q'\,dz$ so that the expression for potential at the point p will be

$$V = \frac{1}{4\pi\epsilon}\int_{-\infty}^{+\infty}\left(\frac{q'}{r_1} - \frac{q'}{r_2}\right)dz$$
$$= \frac{q'}{2\pi\epsilon}\int_0^{\infty}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)dz$$

Substituting $r_1 = \sqrt{r_a^2 + z^2}$ and $r_2 = \sqrt{r_b^2 + z^2}$,

$$V = \frac{q'}{2\pi\epsilon}\int_0^{\infty}\left(\frac{1}{\sqrt{r_a^2 + z^2}} - \frac{1}{\sqrt{r_b^2 + z^2}}\right)dz$$
$$= \frac{q'}{2\pi\epsilon}\left[\ln\left(z + \sqrt{r_a^2 + z^2}\right) - \ln\left(z + \sqrt{r_b^2 + z^2}\right]_0^{\infty}$$
$$= \frac{q'}{2\pi\epsilon}\left[\ln\frac{z + \sqrt{r_a^2 + z^2}}{z + \sqrt{r_b^2 + z^2}}\right]_0^{\infty}$$

As z approaches infinity the fraction $(z + \sqrt{r_a^2 + z^2})/(z + \sqrt{r_b^2 + z^2})$ approaches unity. Therefore, when the limits are inserted, the expression for potential at the point p becomes

$$V = -\frac{q'}{2\pi\epsilon} \ln \frac{r_a}{r_b}$$

$$= \frac{q'}{2\pi\epsilon} \ln \frac{r_b}{r_a} \tag{2-27}$$

It will be observed that in the plane of symmetry between the wires ($r_a = r_b$) the potential is zero.

2.07 Equipotential Surfaces. The solutions to many problems involving electric fields are simplified by making use of equipotential surfaces. An *equipotential surface* is a surface on which the potential is everywhere the same. The movement of charge over such a surface would require no work. Since any two points on the surface have the same potential, there is zero potential difference and therefore zero electric field everywhere along (tangential to) the surface. This means that the electric field must always be perpendicular to an equipotential surface.

A very simple example of equipotential surfaces exists in the case of a point charge. Since $V = q/4\pi r\epsilon$, a surface with a fixed r would have a constant potential. The constant potential surfaces therefore are concentric spherical shells.

In the problem of the parallel line charges the equipotential surfaces can be determined with little difficulty. The locus of a constant potential is obtained by setting the potential of eq. (27) equal to a constant, that is

$$k_1 = \frac{q'}{2\pi\epsilon} \ln \frac{r_b}{r_a}$$

This requires that

$$\left(\frac{r_b}{r_a}\right)^2 = k^2$$

where k^2 is another constant. From Fig. 2-4

$$r_a^2 = \left(x + \frac{d}{2}\right)^2 + y^2$$

$$r_b^2 = \left(x - \frac{d}{2}\right)^2 + y^2$$

Therefore
$$\frac{\left(x - \frac{d}{2}\right)^2 + y^2}{\left(x + \frac{d}{2}\right)^2 + y^2} = k^2$$

$$x^2(1 - k^2) + y^2(1 - k^2) - xd(1 + k^2) + \frac{d^2}{4}(1 - k^2) = 0$$

$$x^2 + y^2 - xd\frac{1 + k^2}{1 - k^2} + \frac{d^2}{4} = 0$$

$$x^2 + y^2 - xd\frac{1 + k^2}{1 - k^2} + \frac{d^2}{4}\frac{(1 + k^2)^2}{(1 - k^2)^2} = \frac{d^2}{4}\left[\frac{(1 + k^2)^2}{(1 - k^2)^2} - 1\right]$$

$$\left[x + \frac{d}{2}\left(\frac{k^2 + 1}{k^2 - 1}\right)\right]^2 + y^2 = \frac{k^2 d^2}{(k^2 - 1)^2}$$

This is the equation of two families of circles with radii $\frac{kd}{k^2 - 1}$ and centers at $\pm\frac{d}{2}\left(\frac{k^2 + 1}{k^2 - 1}\right), 0$. Because of symmetry in the z direction the equipotential surfaces will be cylinders. The cylinders are *not* concentric because k will depend on the potential selected.

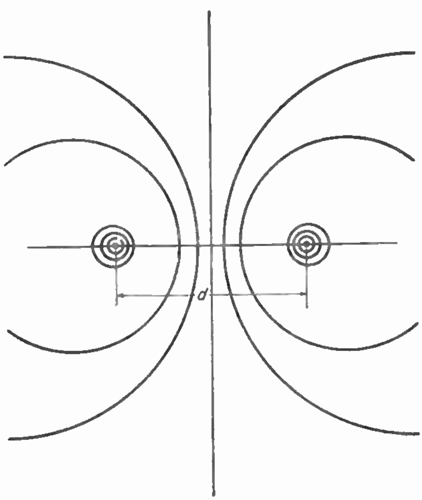


FIG. 2-5. Equipotential surfaces about parallel line charges.

Figure 2-5 shows a plot of the equipotential surfaces about the parallel line charges. It is seen that for *small* values of radius the equipotential cylinders about each line are nearly concentric, with the line charges as the center.

Conductors. A conducting medium is one in which an electric field or difference of potential is always accompanied by a movement of charges. The theory explaining this phenomenon is that a conductor contains *free electrons* or *conduction electrons* that are relatively free to move through the ionic crystal lattice of the conducting medium. It follows that in a conductor there can be no *static* electric field, because any electric field originally present causes the charges to redistribute themselves until the electric field is zero. The electric field being zero within a conductor means that there is no difference of potential between any two points on the conductor. For static electric fields therefore, a conductor surface is always an *equipotential* surface.

It also follows that within a conductor there can be no *net* charge (excess positive or negative charge). If there were a net

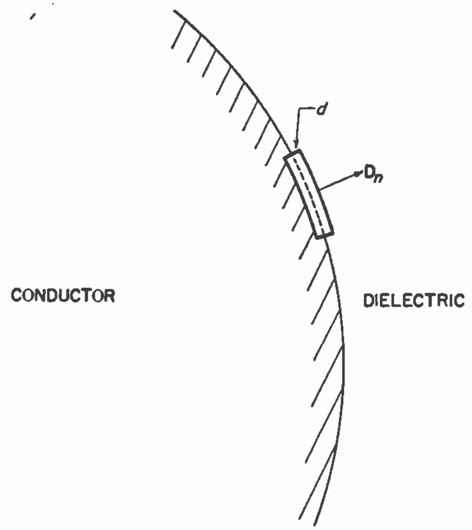


FIG. 2-6. Boundary surface between a conductor and a dielectric.

charge anywhere within the conductor, then by Gauss's law there would be a displacement away from this charge and therefore a displacement density **D** in the conductor. Since $\mathbf{E} = \mathbf{D}/\epsilon$ this requires (for any finite value of dielectric constant ϵ) that there be an electric field **E** in the conductor. But the possibility of this has already been ruled out for the electrostatic case. Therefore, the (net) charge density ρ must be zero within the conductor. There can, however, be a distribution of charge on the surface of the conductor, and this gives rise to a normal component of electric field in the dielectric medium outside the conductor. The strength of this normal component of electric intensity in terms of the surface charge is obtained directly from Gauss's law.

Electric Field Due to Surface Charge. Let the charge per unit area or *surface charge density* on the surface of a conductor be ρ_s coulombs per square meter. Enclose an element of the surface in a volume of "pillbox" shape with its flat surfaces parallel to the conductor surface. Then, if the depth d of the pillbox is made extremely small compared with its diameter, the electric displacement through its edge surface will be negligible compared with any displacement through its flat surfaces. There can be no displacement through the left-hand surface submerged in the conductor (because no **E** exists in the conductor) so all the electric flux must emerge through the right-hand surface. Applying Gauss's law to this case gives

$$D_n\, da = \rho_s\, da$$

where da is the area of one face of the pillbox and D_n is the displacement density normal to the surface. Therefore,

$$D_n = \rho_s \qquad \text{and} \qquad E_n = \frac{\rho_s}{\epsilon}$$

The electric displacement density at the surface of a conductor is normal to the surface and equal in magnitude to the surface charge density. The electric intensity is also normal to the surface and is equal to the surface charge density divided by the dielectric constant.

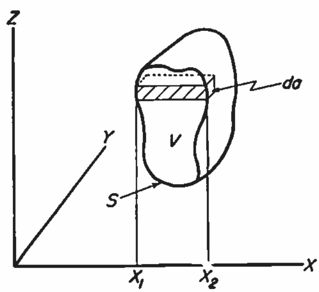


FIG. 2-7. Section of a volume V.

2.08 Divergence Theorem. The divergence theorem (also called Gauss's theorem) relates an integration throughout a volume to an integration over the surface surrounding the volume.

Figure 2-7 shows a closed surface S enclosing a volume V that contains charges (or a charge density) that produce an electric flux density **D**.

By the definition of divergence,

$$\text{div } \mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}$$

so that

$$\int_v \operatorname{div} \mathbf{D}\, dV = \iiint \left(\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}\right) dx\, dy\, dz \tag{2-28}$$

where

$$dV = dx\, dy\, dz$$

Consider now the elemental rectangular volume shown shaded, which has dimensions dy and dz in the y and z directions respectively. Let D_{x_1} and D_{x_2} respectively be the x component of the electric flux entering the left-hand side and leaving the right-hand side of the rectangular volume. The total flux emerging is the algebraic difference of these two. But

$$D_{x_2} - D_{x_1} = \int_{x_1}^{x_2} \frac{\partial D_x}{\partial x}\, dx$$

$$\therefore \iiint \frac{\partial D_x}{\partial x}\, dx\, dy\, dz = \iint (D_{x_2} - D_{x_1})\, dy\, dz \tag{2-29}$$

Now $dy\, dz$ is the x component of the surface element $d\mathbf{a}$, and so (29) is just the integration of the product of D_x times the x component of $d\mathbf{a}$ over the whole surface. (Note that for the right face $D_x\, da_x = D_{x_2}\, dy\, dz$, but for the left face $D_x\, da_x = -D_{x_1}\, dy\, dz$. This is because the direction of $d\mathbf{a}$ is along the *outward* normal and for the left face the x component of $d\mathbf{a}$ has a direction opposite to that of D_{x_1}.)

By definition of a scalar product

$$\mathbf{D} \cdot d\mathbf{a} = D_x\, da_x + D_y\, da_y + D_z\, da_z$$

where da_x indicates the x component of $d\mathbf{a}$, and so on. Then, making use of (29), eq. (28) may be written

$$\int_{\text{vol}} \operatorname{div} \mathbf{D}\, dV = \oint_S \mathbf{D} \cdot d\mathbf{a} \tag{2-30}$$

This is the *divergence theorem.*

Although derived here for the particular case of electric displacement density $\mathbf{D}$ it is a quite general and very useful theorem of vector analysis. For *any* vector, it relates the integral over a closed surface of the normal component of the vector to the integral over the volume (enclosed by the surface) of the divergence of the vector.

Integral Definition of Divergence. The divergence theorem (30) provides a definition of divergence of a vector in the integral form which is easy to put into words. The expression on the right-hand side of (30) is the net outward electric flux through the closed surface S. The expression on the left represents the average divergence of $\mathbf{D}$ multiplied by the volume V that is enclosed by S. Thus the *average divergence* of a vector is the net outward flux of the vector through a closed surface S divided by the volume V enclosed. The limit of the average divergence as S is allowed to shrink to zero about a point is the divergence of the vector at that point; that is,

$$\operatorname{div} \mathbf{D} = \lim_{S \to 0} \frac{\oint_S \mathbf{D} \cdot d\mathbf{a}}{V}$$

In words, the divergence of the vector $\mathbf{D}$ is the net outward flux of $\mathbf{D}$ per unit volume.

Alternative Statement of Gauss's Law. Making use of Gauss's law which states

$$\oint_S \mathbf{D} \cdot d\mathbf{a} = \int_{\text{vol}} \rho \, dV \tag{2-31}$$

and applying the divergence theorem (30), gives

$$\int_{\text{vol}} \operatorname{div} \mathbf{D} \, dV = \int_{\text{vol}} \rho \, dV$$

This holds for any volume whatsoever. As the volume considered is reduced to an elemental volume, this becomes the point relation,

$$\operatorname{div} \mathbf{D} = \rho \tag{2-32}$$

This is the alternative statement of Gauss's law. It states that at every point in a medium the *divergence* of electric displacement density is equal to the charge density. Recalling the physical interpretation of the term divergence, eq. (32) might be stated as follows: The net outward flux of electric displacement per unit volume is equal to the charge per unit volume. Equation (32) will often be found to be a more useful form for mathematical manipulation than the corresponding integral statement (31).

2.09 Poisson's Equation and Laplace's Equation. Equation (32) is a relation between the electric displacement density and the charge density in a medium. If the medium is homogeneous and isotropic so that ϵ is constant and a scalar quantity, eq. (32) can be written as

$$\text{div } \epsilon\mathbf{E} = \epsilon \text{ div } \mathbf{E} = \rho$$

or

$$\text{div } \mathbf{E} = \frac{\rho}{\epsilon} \tag{2-33}$$

Recall that $\mathbf{E}$ is related to the potential V by

$$\mathbf{E} = -\text{ grad } V = -\nabla V$$

Substituting this into (33)

$$\text{div grad } V = -\frac{\rho}{\epsilon} \tag{2-34a}$$

or symbolically

$$\nabla \cdot (\nabla V) = -\frac{\rho}{\epsilon} \tag{2-34b}$$

or

$$\nabla^2 V = -\frac{\rho}{\epsilon} \tag{2-34c}$$

Equation (34) is known as *Poisson's equation.* In free space, that is in a region in which there are no charges ($\rho = 0$), it becomes

$$\nabla^2 V = 0 \tag{2-35}$$

This special case for source-free regions is *Laplace's equation.*

Laplace's Equation. Laplace's equation is a relation of prime importance in electromagnetic field theory. Expanded in rectangular co-ordinates it becomes

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{2-36}$$

This is a second-order partial differential equation relating the rate of change of potential in the three component directions. In any charge-free region the potential distribution must be such that this relation is satisfied. An alternative form of (36) in terms of electric intensity is

$$\text{div } \mathbf{E} = 0 \tag{2-37}$$

In this form the statement is that in a homogeneous charge-free region the number of lines of electric intensity emerging from a unit volume is zero, or (in such a region) lines of electric intensity are continuous.

The Problem of Electrostatics. In a homogeneous charge-free region the potential distribution, whatever it may be, must be a solution of the Laplace equation. The problem is to find a potential distribution that will satisfy (35) as well as the boundary conditions of the particular problem. When the charges are given the potential can be found directly from

$$V = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho}{r}\, dV$$

This is a simple problem and the solution is straightforward. On the other hand, if the potential distribution is given for a certain configuration of conductors, the charge distribution on the conductors can be found from

$$\rho_s = D_n = \epsilon E_n$$

In the general problem as it exists, however, neither the potential distribution nor the charge distribution is known. These are the quantities to be found. A certain configuration of conductors is specified and the voltages or potential differences between conductors are given (or the total charge on each conductor may be given). The charges on the conductors will then distribute themselves to make the conductors equipotential surfaces and at the same time produce a potential distribution between conductors which will satisfy Laplace's equation.

Thus the problem is that of finding a solution to a second-order differential equation (Laplace's equation) that will fit the boundary conditions. The problem is one of integration and therefore straightforward methods of solution are not generally available. In fact, only in a relatively small number of cases, where symmetry or some other consideration makes it possible to specify the charge distribution, can an exact solution be found. Of course, an approximate solution can always be obtained, and the degree of approximation can usually be improved to any desired extent by a systematic method of successive approximations. Unfortunately, this often requires an excessive amount of labor.

A similar situation exists in the more general electromagnetic field problem where the fields and charge distributions are varying in time. Although it is this more general problem that is of primary concern in electromagnetic wave theory, it is helpful to consider some of the special methods and solutions that exist for the electrostatic case. It will be found that some of these special methods can be extended to the general case. Moreover, a knowledge of the actual electrostatic solutions for certain simple configurations is required for later use.

Solutions for Some Simple Cases. It is instructive first to obtain the solutions for the simplest possible cases in which, because of symmetry, the field is constant along two axes of the co-ordinate system and variations occur in one direction only.

EXAMPLE 5(a): *In Rectangular Co-ordinates—Two Parallel Planes.* Two parallel planes of infinite extent in the x and y directions and separated by

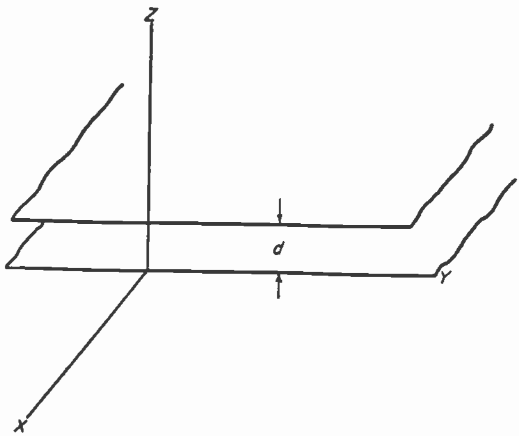


FIG. 2-8. Two parallel planes.

a distance d in the z direction have a potential difference applied between them (Fig. 2-8). It is required to find the potential distribution and electric intensity in the region between the planes.

In rectangular co-ordinates Laplace's equation is

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0$$

From symmetry it is evident that there is no variation of V with x or y, but only with z. For this simple case Laplace's equation reduces to

$$\nabla^2 V = \frac{\partial^2 V}{\partial z^2} = 0$$

which has a solution $V = k_1 z + k_2$

where k_1 and k_2 are arbitrary constants. Substituting the boundary conditions

$$V = V_0 \quad \text{at } z = 0, \qquad V = V_1 \quad \text{at } z = d$$

gives

$$k_2 = V_0 \qquad \text{and} \qquad k_1 = \frac{V_1 - V_0}{d}$$

so that

$$V = \frac{V_1 - V_0}{d} z + V_0$$

The electric intensity is obtained from the relation

$$\mathbf{E} = -\operatorname{grad} V$$
$$= -\frac{\partial V}{\partial z}\mathbf{k}$$
$$= -\frac{V_1 - V_0}{d}\mathbf{k}$$

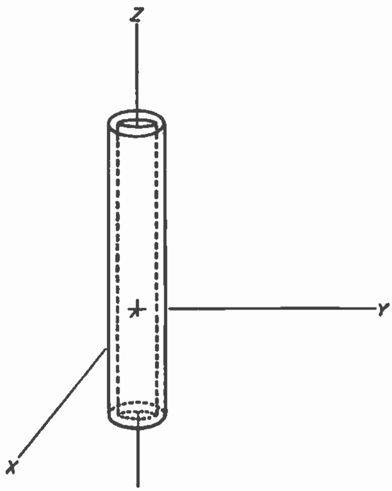


FIG. 2-9. Two concentric cylinders.

The electric intensity is constant in the region between the plates. It is directed along the z axis and toward the plate of lower potential.

EXAMPLE 5(b): *In Cylindrical Co-ordinates—Concentric Cylinders.* In cylindrical co-ordinates Laplace's equation is

$$\nabla^2 V = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial V}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2} = 0 \qquad (2\text{-}38)$$

For the space between two very long concentric cylinders (Fig. 2-9), in which case there will be no variations with respect to either ϕ or z, but only in the ρ direction, eq. (38) becomes

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial V}{\partial \rho}\right) = 0 \qquad (2\text{-}39)$$

A trivial solution to this equation is V equals a constant. A useful solution that fits the boundary condition is

$$V = k_1 \ln \rho + k_2$$

The electric intensity in the region between the cylinders will be

$$\mathbf{E} = -\operatorname{grad} V$$
$$= -\frac{\partial V}{\partial \rho}\mathbf{u}_\rho$$
$$= -\frac{k_1}{\rho}\mathbf{u}_\rho$$

Example 5(c): *In Spherical Co-ordinates—Concentric Spheres.* In spherical co-ordinates with no variations in θ or ϕ directions, Laplace's equation is

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) = 0$$

A solution is

$$V = \frac{k_1}{r} + k_2$$

The electric intensity between the spheres is

$$\mathbf{E} = -\operatorname{grad} V = -\frac{\partial V}{\partial r}\mathbf{u}_r = \frac{k_1}{r^2}\mathbf{u}_r$$

In the three examples just solved, the simplicity of the boundary conditions (due to symmetry) made it possible to guess the solution and write it down from inspection. Only in rare cases is it possible to do this. However, there is an important group of problems that can be solved almost by inspection because their boundary conditions are similar to those of problems which have already been solved. These make use of the principle of the electrical image.

Solution by Means of the Electrical Image. As a simple example of this method of solution consider the problem of a line charge q' coulombs per meter parallel to and at a distance $d/2$ from a perfectly conducting plane of infinite extent. It is required to determine the resulting potential distribution and the electric field. The boundary condition in this case is that the conducting plane must be an equipotential surface. Also if the potential at infinity is considered to be zero, the potential of the conducting plane must be zero since it extends to infinity.

The lines of electric flux, which start on the positive line charge, must terminate on negative charges on the plate and at infinity. These negative charges on the conducting surface are required to distribute themselves so that there is no tangential component of electric field along the surface of the conductor; *i.e.*, so that the conducting plane is an equipotential surface (Fig. 2-10).

If this distribution can be found, the potential at any point can then be determined from the relation $V = \int_{\text{vol}} \frac{\rho}{4\pi\epsilon r} dV$. As it stands, this is not a simple problem. However now recall the problem of two equal and opposite line charges for which the solution has already been obtained (page 40). It will be remembered that the plane of symmetry between the wires is an equipotential plane of zero potential. Hence a conducting surface could be

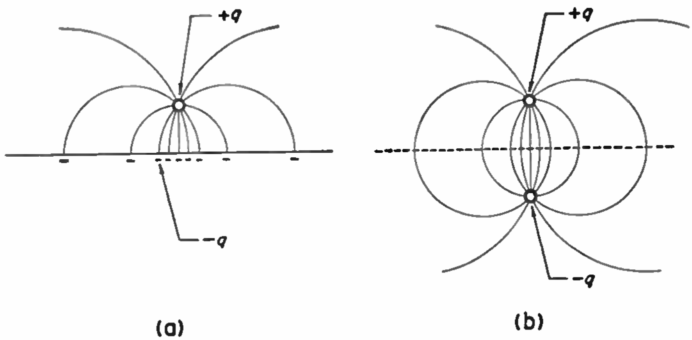


FIG. 2-10. (a) Line charge near a conducting surface. (b) Charges on the conducting surface have been replaced by an appropriately located "image" charge.

placed at the location of this equipotential plane without affecting the potential distribution in any manner whatsoever. If this were done, the negative line charge then has no effect on the field on the opposite side of the conductor and, so as far as *that* field is concerned, can be removed. The problem is now just the one for which a solution is required. The solution can be set down directly. The field due to a line charge at a distance $d/2$ from an infinite conducting plane is exactly the same as the field (on one side of the zero potential plane) produced by that line charge and an equal and opposite line charge located parallel to it and a distance d away. This second (hypothetical) line charge is called the electrical image of the other, from the analogy with optical images.

Thus in any problem involving charges and conductors, if an additional distribution of charges can be found which will make the surfaces to be occupied by the conductors equipotential surfaces having the correct potential, the conductors can be removed and the field in the volume, originally outside the conductors, will not be changed. Then this field can be computed by methods already developed. The problem is now simply that of finding the potential and electric intensity for a distribution of charges without conductors.

As a second example of this method let it be required to find the potential distribution and electric field about a pair of parallel cylindrical conductors which have applied to them a specified voltage or potential difference. Again referring to the line charges on page 40, the equipotential surfaces about these line charges are cylinders. If the conductors are located in this field to coincide with an appropriate pair of equipotentials, their introduction will not change the field configuration outside the volume occupied by the conductors. Thus the potential distribution outside the conductors is just the same as it was before the conductors were introduced and is that produced by a pair of line charges of proper strength located along appropriate axes. The solution to this problem has already been obtained.

It must be observed, that this process of computing the effects of a conductor is an inverse one, i.e., a solution must be found by experience, and there is no straightforward method of finding an analytical solution in every case. This is analogous to the problem of differentiation and integration. In differentiation a straightforward method is available for finding derivatives, but the determination of integrals depends on the experience of the operator, or the recorded experience of those who have gone before him. In the case of electric fields, an analytical expression for the charge distribution that can replace a given conductor is not always known, just as the integral of every function is not known. On the other hand, with a given configuration, there are approximate methods available for determining the effect of a conductor in a field just as there are approximate methods for the integration of any curve that can be graphed.

2.10 Capacitance. The capacitance between two conductors is defined by the relation

$$C = \frac{Q}{V}$$

where V is the voltage or potential difference between the conductors due to equal and opposite charges on them of magnitude Q. When the capacitance of a single conductor is referred to, it is implied that the other conductor is a spherical shell of infinite radius.

EXAMPLE 6: *Parallel Plate Condenser.* Consider the parallel plate condenser having plates of area A and separation d (Fig. 2-11). (d is assumed

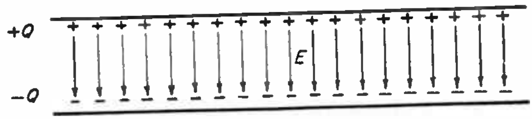


FIG. 2-11. Parallel plate condenser.

to be very small compared with the length and width of the plates so that the effect of flux fringing may be neglected.) If the plates have a charge of magnitude Q, the surface charge density will be

$$\rho_s = \frac{Q}{A}$$

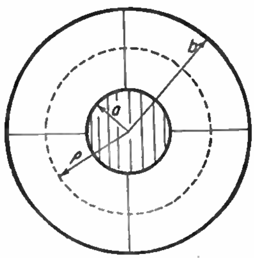


FIG. 2-12. Concentric conductors.

The electric intensity $\mathbf{E}$ between the plates is uniform and of magnitude

$$E = \frac{\rho_s}{\epsilon}$$

where ϵ is the dielectric constant of the medium between the plates.

The voltage between the plates will be

$$V = \int_1^2 \mathbf{E} \cdot d\mathbf{s}$$
$$= Ed$$

The capacitance is $$C = \frac{Q}{V} = \frac{Q}{Ed} = \frac{\epsilon A}{d} \qquad \text{farads}$$

Because of their usefulness in later work the capacitances for two other simple cases will be found.

EXAMPLE 7. *Concentric Conductors.* It is required to determine the capacitance per unit length between two infinitely long concentric conducting cylinders (Fig. 2-12). The outside radius of the inner conductor is a and the inside radius of the outer conductor is b.

Assume a charge distribution q' coulombs per meter on the inner conductor and an equal and opposite charge on the outer conductor. Because of symmetry the lines of electric flux will be radial and the displacement through any cylindrical shell will be q' coulombs per unit length. The magnitude of the displacement density will be

$$D = \frac{q'}{2\pi\rho}$$

and the magnitude of the electric intensity will be

$$E = \frac{q'}{2\pi\rho\epsilon}$$

The voltage between the conductors is

$$V = \int_a^b \mathbf{E} \cdot d\boldsymbol{\rho} = \int_a^b \frac{q'}{2\pi\rho\epsilon}\, d\rho$$

$$= \frac{q'}{2\pi\epsilon} \ln \rho \Big]_a^b = \frac{q'}{2\pi\epsilon} \ln \frac{b}{a}$$

The capacitance per meter will be

$$C = \frac{q'}{V} = \frac{2\pi\epsilon}{\ln b/a} \quad \text{f/m} \qquad (2\text{-}40)$$

For the air dielectric for which $\epsilon = \dfrac{1}{36\pi \times 10^9}$

$$C = \frac{10^{-9}}{18 \ln b/a} \quad \text{f/m} \qquad (2\text{-}41)$$

Example 8: *Parallel Cylindrical Conductors.* The method for determining the electric field for this case has already been considered. A pair

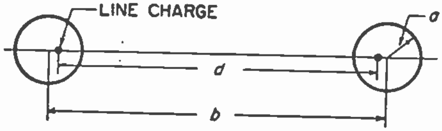


Fig. 2-13. Parallel cylindrical conductors.

of line charges, appropriately located, would make the surfaces occupied by the conductors equipotentials (Fig. 2-13). If the radius of the cylinders is a and the separation between their axes is b, then, in terms of the notation used in connection with Fig. 2-4,

$$b = d\,\frac{k^2 + 1}{k^2 - 1}$$

$$a = \frac{kd}{k^2 - 1}$$

$$\frac{b}{a} = \frac{k^2 + 1}{k}$$

$$ak^2 - bk + a = 0$$

$$k = \frac{b + \sqrt{b^2 - 4a^2}}{2a}$$

But

$$k = \frac{r_b}{r_a}$$

$$\therefore \frac{r_b}{r_a} = \frac{b + \sqrt{b^2 - 4a^2}}{2a}$$

The potential at the surface of one conductor is given by equation (27)

$$V_1 = \frac{q'}{2\pi\epsilon}\ln\frac{b + \sqrt{b^2 - 4a^2}}{2a}$$

where q' is the charge per unit length.

When the separation is large compared with the radius, that is when $b \gg a$, this becomes

$$V_1 = \frac{q'}{2\pi\epsilon}\ln\frac{b}{a}$$

The potential at the other conductor will be equal and opposite. Hence

$$V = V_1 - V_2 = \frac{q'}{\pi\epsilon}\ln\frac{b + \sqrt{b^2 - 4a^2}}{2a}$$

The capacitance per unit length is

$$C = \frac{q'}{V} = \frac{\pi\epsilon}{\ln\dfrac{b + \sqrt{b^2 - 4a^2}}{2a}} \qquad \text{f/m}$$

If $b \gg a$ the capacitance is given very closely by

$$C \approx \frac{\pi\epsilon}{\ln b/a} \qquad \text{f/m} \quad (2\text{-}42)$$

For an air dielectric between the conductors

$$C \approx \frac{10^{-9}}{36 \ln b/a} \qquad \text{f/m} \quad (2\text{-}43)$$

EXAMPLE 9: *Capacitance of a (Finite-length) Wire or Cylindrical Rod.* In the first two of the above three examples the charge distribution was uniform over the conductor surfaces and so the potential distribution could be determined directly and exactly. In the third example the charge distribution was not known but the potential was obtained by showing that the problem was similar to that of a pair of line charges for which the exact solution was known. In practice there are very few problems that can be solved so simply. In most actual problems the charge distribution is unknown and there are no methods available for obtaining an exact solution. It is then necessary to set about finding an approximate solution.

In the present problem it is required to determine the capacitance of a straight horizontal wire or conducting rod elevated at a height h above

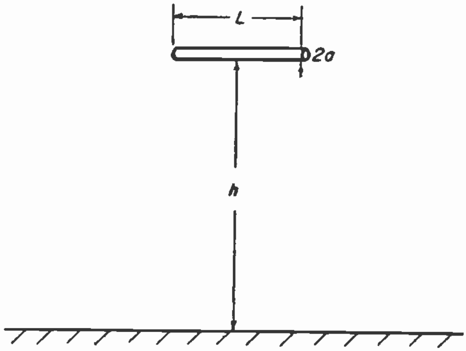


FIG. 2-14. An elevated wire or rod.

the earth. The rod has a length $L = 1$ meter and a radius $a = 0.5$ cm, and is elevated at a height $h = 10$ meters (Fig. 2-14).

For a first attack on the problem it will be assumed that the height above the earth is very great so that the problem is that of determining the capacitance of a cylindrical rod remote from the earth. The boundary condition is that the surface of the rod be an equipotential surface. Obviously the charge distribution cannot be uniform along the length of the rod because such a distribution produces a potential that varies along the length of the wire. Moreover there is apparently no straightforward method available for finding the correct charge distribution, which will make the surface an equipotential. This is a *typical practical* problem.

This particular problem was solved many years ago by G. W. O. Howe, using a method of attack that is now used very frequently in electrostatic and electromagnetic problems. It is first assumed that the charge distribution is uniform (even though such an assumption is known to be incorrect). The potential along the wire due to this uniform charge distribution is calculated. It is then assumed that the true potential, which actually exists along the surface of the wire, is equal to the average value of this calculated potential. Knowing the potential for a given total charge the capacitance of the wire is obtained from $C = Q/V_{avg}$.

Solution: Figure 2-15 shows the rod, which is assumed to have a uniform charge distribution on its surface of amount q' coulombs per meter of length. The surface charge density is the $q'/2\pi a$ coulombs per square meter. The charge on each element of area contributes to the potential at a point p on the surface and the total potential at p can be obtained by integrating these contributions over the surface of the wire. It is possible to simplify this part of the problem in the following manner. It is known that the equipotential surfaces about a line charge of infinite length are cylinders whose axes coincide with the line charge. If a conducting

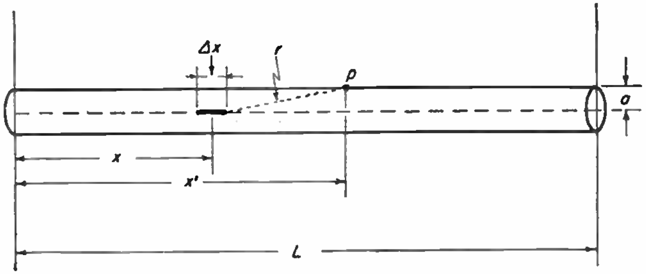


FIG. 2-15. Surface charge is replaced by a line charge along the axis for the purpose of computing potential.

cylinder is made to coincide with one of these equipotential surfaces and is given a charge per unit length equal to that of the line charge, the electric field in the region about the cylinder will be exactly the same as that produced originally by the line charge.

Thus as far as the potential outside of it (and on its surface) is concerned a long charged cylinder may be replaced by a line charge situated along its axis and having the same charge per unit length as the cylinder. Applying this principle in Fig. 2-15 the contribution to the potential at a point x' on the surface due to the charge on an element of length Δx located at point x along the axis will be

$$\Delta V = \frac{q'\,\Delta x}{4\pi\epsilon\sqrt{(x'-x)^2+a^2}} \tag{2-44}$$

where q' is the charge per unit length. The total potential at x' due to the assumed charge distribution along the axis is

$$\begin{aligned} V_{x'} &= \frac{q'}{4\pi\epsilon}\int_0^L \frac{dx}{\sqrt{(x'-x)^2+a^2}} \\ &= \frac{q'}{4\pi\epsilon}\left[-\sinh^{-1}\left(\frac{x'-x}{a}\right)\right]_0^L \\ &= \frac{q'}{4\pi\epsilon}\left[-\sinh^{-1}\left(\frac{x'-L}{a}\right)+\sinh^{-1}\left(\frac{x'}{a}\right)\right] \end{aligned} \tag{2-45}$$

From eq. (45) the potential at the middle of the rod will be

$$V_{x'=L/2} = \frac{q'}{4\pi\epsilon}\left[-\sinh^{-1}\left(-\frac{L}{2a}\right) + \sinh^{-1}\left(\frac{L}{2a}\right)\right]$$
$$= \frac{2q'}{4\pi\epsilon}\sinh^{-1} 100 = \frac{10.6}{4\pi\epsilon}q'$$

and the potential at each end is

$$V_{x'=0} = V_{x'=L} = \frac{5.99q'}{4\pi\epsilon}$$

The potential can be calculated at other points along the length to obtain the resulting distribution, shown by Fig. 2-16.

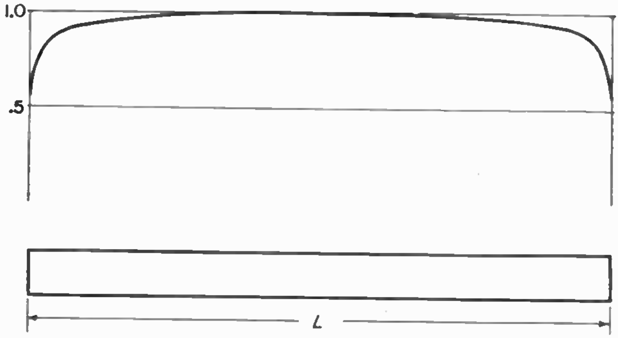


FIG. 2-16. Potential distribution along the rod calculated from assumed uniform-charge distribution.

The *average* potential along the rod may be found by integrating eq. (45) (with respect to x') over the length of the rod and dividing by L.

$$V_{\text{avg}} = \frac{q'}{4\pi\epsilon L}\int_0^L \left[-\sinh^{-1}\left(\frac{x'-L}{a}\right) + \sinh^{-1}\left(\frac{x'}{a}\right)\right] dx'$$
$$= \frac{q'}{4\pi\epsilon L}\Bigg| -(x'-L)\sinh^{-1}\left(\frac{x'-L}{a}\right) + \sqrt{(x'-L)^2 + a^2}$$
$$+ x'\sinh^{-1}\left(\frac{x}{a}\right) - \sqrt{x^2 + a^2}\Bigg|_0^L$$
$$= \frac{q'}{2\pi\epsilon}\left[\frac{a}{L} + \sinh^{-1}\left(\frac{L}{c}\right) - \sqrt{1 + \frac{a^2}{L^2}}\right]$$

Substituting numerical values

$$V_{\text{avg}} = \frac{q'}{2\pi\epsilon}[0.005 + \sinh^{-1} 200 - (1 + 0.00001)]$$
$$= \frac{5.00q'}{2\pi\epsilon}$$

The capacitance of the rod (remote from the earth) will be (approximately)

$$C = \frac{q'L}{V_{\text{avg}}} = 11.11\ \mu\mu\text{f}$$

The effect of the proximity of the earth can be accounted for by means of the image principle. A negative charge $-q'L$ located at the position of the image will decrease the average potential of the rod slightly. With negligible error this negative charge can be considered as being located at a point at the center of the image a distance $2h$ from the rod and the potential at the rod due to this negative charge will be

$$V_{\text{image}} = \frac{-q'L}{4\pi\epsilon \times 2h} = -\frac{0.05q'L}{4\pi\epsilon}$$

The average potential of the rod including the contribution from the image charge is

$$V_{\text{avg}} = \frac{q'}{4\pi\epsilon}(10.0 - 0.05) = \frac{9.95q'}{4\pi\epsilon}$$

The capacitance of the rod including the effect of the presence of the earth will be

$$C = \frac{4\pi\epsilon}{9.95} = 11.16\ \mu\mu\text{f}$$

The proximity of the earth has increased the capacitance by about ½ of 1 per cent. It will be observed that in this case a 50 per cent error in computing the contribution from the image would affect the final answer a negligible amount. Therefore there is usually no justification for seeking a more accurate solution for this part of the problem.

The method outlined above gives an approximate answer for the capacitance of the rod. The degree of approximation can be improved by assuming a second and different charge distribution, which will produce a more nearly uniform potential distribution. (This is easy to do once the potential distribution due to a uniform charge distribution has been found.) It will be found (for this case) that the answers obtained with more nearly correct charge distributions do not differ appreciably from that obtained above. The *correct* value for capacitance will always be a little larger than that calculated from any assumed charge distribution. This is because the actual charge distribution is always such as to make the potential energy of the system, and therefore the potential of the rod, a minimum.

2.11 Energy Stored in an Electric Field. When a condenser is charged so that there exists a voltage V between its plates, there is a storage of energy, which can be converted into heat by discharging the condenser through a resistance. The amount of energy stored can be found by calculating the work done in charging the condenser. Since potential was defined in terms of work per unit charge, the work done in moving a small charge dq against a potential difference V is $V\,dq$. But the voltage V can be expressed in terms of the capacity C and the charge q by

$$V = \frac{q}{C}$$

Therefore the work done in increasing the charge on a condenser by an amount dq is

$$\frac{q}{C}\,dq$$

The total work done in charging a condenser to Q coulombs is

$$\text{Total work} = \int_0^Q \frac{q}{C}\,dq = \frac{1}{2}\frac{Q^2}{C}$$

Therefore the energy stored by a charged condenser is

$$\text{Stored energy} = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}VQ = \frac{1}{2}V^2C$$

This energy is said to be "associated with the electric charge on the conductors," or alternatively, "associated with the electric field in the dielectric between the conductors."*

It is convenient in electromagnetic wave theory, where energy is propagated through space, to use the second of these concepts and associate the energy with the electric field. An expression giving the stored-energy density in terms of the electric field is readily obtained in the case of the parallel-plate condenser where the electric field between the plates is uniform, with a value

$$E = \frac{V}{d}$$

* These statements represent two different points of view or two interpretations of a single set of experimental facts. The question of just where the energy "resides" in this case is similar to the question of where the potential energy is stored when a weight has been raised. The question seems to be one of philosophy or interpretation and as such is unanswerable on the basis of any physical measurements that can be made by the engineer.

V is the voltage between the plates and d is their separation. The expression for stored energy can then be written

$$\text{Stored energy} = \frac{V^2C}{2} = \frac{E^2d^2\epsilon A}{2d} = \frac{\epsilon E^2}{2}(Ad)$$

Since Ad is the volume between the plates the quantity $\epsilon E^2/2$ has the dimensions of energy per unit volume and is said to be the *energy density* of the electric field. Although derived here for the special case of a uniform electric field, it is easily shown for the general case that the quantity $\epsilon E^2/2$, when integrated over the whole volume in which the electric field exists, always gives the correct value for the total stored (electric) energy.

2.12 Conditions at a Boundary between Dielectrics. Consider conditions at the interface between two dielectrics in an electric

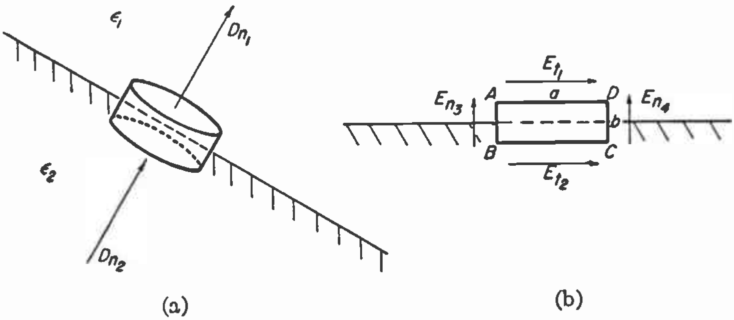


FIG. 2-17. Boundary surface between two dielectric media.

field. The dielectric constants of the media are ϵ_1 and ϵ_2 respectively, and it is assumed that there are no free charges on the boundary surface.

Apply Gauss's law to the shallow pillbox volume that encloses a portion of the boundary (Fig. 2-17). Since there are no charges within the volume the net outward displacement through the surface of the box is zero. As the depth of the box is allowed to approach zero, always keeping the boundary surface between its two flat faces, the displacement through the curved-edge surface becomes negligible. Gauss's law then requires that the displacement through the upper face be equal to the displacement through the lower face. Because the area of the faces are equal, the *normal* *components of the displacement densities must be equal,* that is,

$$D_{n_1} = D_{n_2}$$

Thus there are the same number of lines of displacement flux entering one face as are leaving the other face and the lines of electric displacement are continuous across a boundary surface.

Whereas the normal component of **D** is the same on both sides of a boundary, it is easily shown that the *tangential* component of the electric intensity **E** must be continuous across the boundary. Referring to Fig. 2-17b it is supposed that there are electric intensities $\mathbf{E}_1$ and $\mathbf{E}_2$ respectively in medium (1) and medium (2). In the electrostatic field the voltage around any closed path must be zero, that is,

$$V_{\text{closed path}} \equiv \oint \mathbf{E} \cdot ds \equiv 0$$

Apply this to the rectangular path $ABCD$, in which AD is *just* inside medium (1) and BC *just* inside medium (2). The length of the rectangle is a, and its width is b.

$$\oint \mathbf{E} \cdot ds = -E_{n_3}b + E_{t_2}a + E_{n_4}b - E_{t_1}a \tag{2-46}$$

where E_{t_1} and E_{t_2} are the average tangential components of **E** along paths AD and BC and E_{n_3} and E_{n_4} are the average normal components of **E** along the paths BA and CD. As the sides AD and BC are brought closer together, always keeping the boundary between them, the lengths AB and CD approach zero and the first and third terms in eq. (46) become zero (assuming that the electric field never becomes infinite). Therefore

$$-E_{t_1}a + E_{t_2}a = 0 \qquad \text{and} \qquad E_{t_1} = E_{t_2}$$

The tangential component of **E** *is continuous at the boundary.*

The two conditions (a) Normal D is continuous at the boundary, and (b) Tangential E is continuous at the boundary are used to solve problems involving dielectrics.

Example 10: *Refraction.* Consider the problem of Fig. 2-18 where an infinite slab of dielectric whose dielectric constant is ϵ_2, is immersed in a medium of ϵ_1. Let θ_1 be the angle that the normal to the boundary makes with the lines of electric force in medium (1). Then the lines of **E** and **D** will be refracted in passing through the slab.

Let $\mathbf{D}_1$ and $\mathbf{D}_2$ be the electric displacement density outside and inside the slab respectively, and $\mathbf{E}_1$ and $\mathbf{E}_2$ be the electric intensity outside and inside the slab. Then

$$\mathbf{D}_1 = \epsilon_1\mathbf{E}_1$$
$$\mathbf{D}_2 = \epsilon_2\mathbf{E}_2$$

By the two fundamental principles stated above

$$D_1 \cos \theta_1 = D_2 \cos \theta_2$$
$$E_1 \sin \theta_1 = E_2 \sin \theta_2$$
$$\frac{D_1}{E_1} \cot \theta_1 = \frac{D_2}{E_2} \cot \theta_2$$

Therefore

$$\epsilon_1 \cot \theta_1 = \epsilon_2 \cot \theta_2$$
$$\frac{\tan \theta_1}{\tan \theta_2} = \frac{\epsilon_1}{\epsilon_2} \tag{2-47}$$

Equation (47) gives the relation between the tangents of the angle of incidence θ_1, and the angle of refraction θ_2 in terms of the dielectric constants of the media involved.

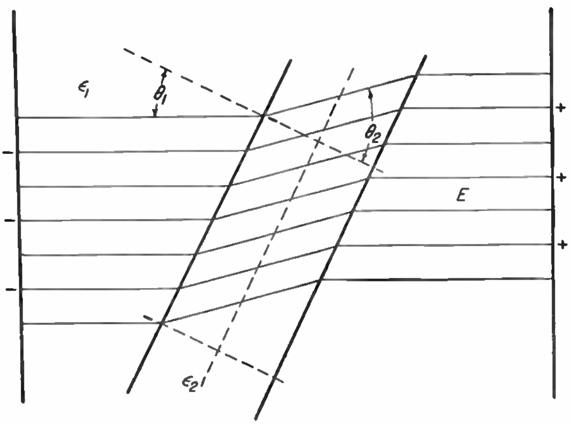


FIG. 2-18. Refraction of an electric field.

2.13 Cylindrical and Spherical Harmonics. It was pointed out in earlier sections of this chapter that, except for a few special cases, the solution of Laplace's equation, subject to the appropriate boundary conditions, was in general a quite difficult problem. There is a group of problems having a certain symmetry that may be solved approximately by use of cylindrical or spherical harmonics.

Because these functions are also required for later use in electromagnetic problems, they will be considered briefly here.

For those problems, which can be set up in cylindrical co-ordinates, and for which there is no variation of the field in the z direction, Laplace's eq. (38) may be written as

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial V}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 V}{\partial\phi^2}=0 \tag{2-48}$$

If a solution of the form $V = \rho^n Q_n$

is assumed (where Q_n is a function of ϕ alone), then substitution of this solution back into (48) shows that Q_n must satisfy the following differential equation.

$$\frac{\partial^2 Q_n}{\partial\phi^2} + n^2 Q_n = 0$$

The solution of this equation is well known and has the form

$$Q_n = A_n \cos n\phi + B_n \sin n\phi$$

where A_n and B_n are arbitrary constants. It will be noted that when $-n$ is substituted for $+n$, the same differential equation for Q results, so that Q_{-n} can be put equal to Q_n. Then, if $\rho^n Q_n$ is a solution of (48), $r^{-n}Q_{-n} = Q_n/r^n$ is also a solution. By inspection it is seen that $V = \ln \rho$ is a solution of (48). Now if a function is a solution of Laplace's equation, each of its partial derivatives with respect to any of the *rectangular* co-ordinates x, y, or z, (but not in general with respect to cylindrical or spherical co-ordinates) is also a solution. That this is so, may be verified by differentiating Laplace's equation partially in rectangular co-ordinates. Differentiating the solution $V = \ln \rho$ with respect to x yields $(\cos \phi)/\rho$ as another solution, while differentiation with respect to y yields $(\sin \phi)/\rho$. Successive differentiation leads to the following set of possible solutions of (48):

$$\ln \rho;\quad \frac{\cos\phi}{\rho};\quad \frac{\sin\phi}{\rho};\quad \frac{\cos 2\phi}{\rho^2};\quad \frac{\sin 2\phi}{\rho^2};\quad \frac{\cos 3\phi}{\rho^3};\quad \frac{\sin 3\phi}{\rho^3};\ \ldots\ .$$

Replacing ρ^{-n} by ρ^n gives a second set, viz.:

$$\rho\cos\phi;\quad \rho\sin\phi;\quad \rho^2\cos 2\phi;\quad \rho^2\sin 2\phi;\quad \rho^3\cos 3\phi;\quad \rho^3\sin 3\phi$$

These solutions of Laplace's equation (48) are known as *circular harmonics* or *cylindrical harmonics.* These harmonic functions may be used to solve problems in which there is no variation of the field in the z direction.

EXAMPLE 11: *Conducting Cylinder in an Electric Field.* A long conducting cylinder is placed in, and perpendicular to, a uniform electric field E_x with the axis of the cylinder coincident with the z axis (Fig. 2-19). Determine the field distribution in the region about the cylinder.

Although the field in the neighborhood of the cylinder will be disturbed by its presence, the distant field will be unaffected and will be just E_x. Therefore, if the potential of the cylinder is taken as zero potential, the potential at a great distance ρ will be $-E_x\rho \cos \phi$. Also the surface of

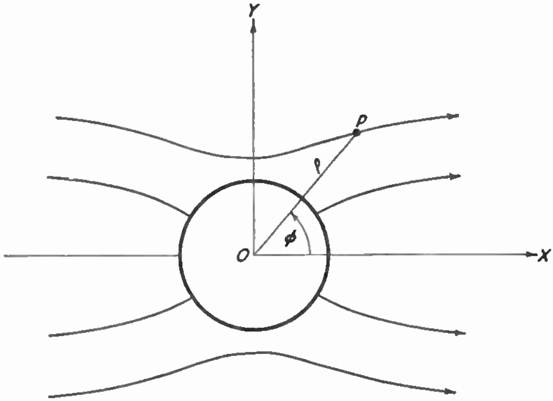


FIG. 2-19. Conducting cylinder in a uniform field.

the cylinder, $\rho = c$, is an equipotential surface, which has arbitrarily been set at zero potential. The problem can be solved by finding that combination of the given cylindrical harmonic solutions that will also satisfy these two boundary conditions. The answer in this case happens to be quite simple, for it is evident that the following combination of cylindrical harmonics, selected from those listed in the table, can be made to satisfy the boundary conditions:

$$V = A\rho \cos \phi + \frac{B \cos \phi}{\rho}$$

For ρ very large $(\rho \to \infty)$

$$V = A\rho \cos \phi = -E_x\rho \cos \phi$$

Therefore $\qquad A = -E_x$

For $\rho = a$, $\qquad V = Aa \cos \phi + \dfrac{B \cos \phi}{a} = 0$

Therefore $$B = -Aa^2 = a^2E_x$$

Then $$V = \left(\frac{a^2}{\rho} - \rho\right) E_x \cos\phi$$

The components of electric field intensity in the region outside the cylinder are given by

$$E_\rho = -\frac{\partial V}{\partial \rho} = \left(\frac{a^2}{\rho^2} + 1\right) E_x \cos\phi$$

$$E_\phi = -\frac{1}{\rho}\frac{\partial V}{\partial \phi} = \left(\frac{a^2}{\rho^2} - 1\right) E_x \sin\phi \tag{2-49}$$

Spherical Harmonics. For problems that can be set up in spherical co-ordinates and for which there is no variation in the ϕ direction, Laplace's equation is

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right) = 0 \tag{2-50}$$

Letting $u = \cos\theta$, so that $du = -\sin\theta\, d\theta$, eq. (50) becomes

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) + \frac{\partial}{\partial u}\left[(1 - u^2)\frac{\partial V}{\partial u}\right] = 0 \tag{2-51}$$

Again assuming that a solution may be found that has the form

$$V = r^n P_n$$

(where P_n is a function of $u = \cos\theta$ alone), substitution into (51) shows that P_n must satisfy the following differential equation:

$$\frac{d}{du}\left[(1 - u^2)\frac{dP_n}{du}\right] + n(n + 1)P_n = 0 \tag{2-52}$$

Equation (52) is known as *Legendre's equation.* This is an important equation in field theory for it is encountered whenever solutions (involving variations with r and θ) are sought to Laplace's equation or the wave equation in spherical co-ordinates. Solutions to eq. (52) may be found by assuming a power series solution, which is inserted back into the differential equation. Equating the coefficients of corresponding powers, relations among these coefficients are found. The result is the following set of solutions for (52)

$$\left.\begin{aligned} P_0 &= 1 \\ P_1 &= u = \cos\theta \\ P_2 &= \tfrac{1}{2}(3u^2 - 1) = \tfrac{1}{2}(3\cos^2\theta - 1) = \tfrac{1}{4}(3\cos 2\theta + 1) \\ P_3 &= \tfrac{1}{2}(5u^3 - 3u) = \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta) \\ &\qquad\qquad\qquad = \tfrac{1}{8}(5\cos 3\theta + 3\cos\theta) \end{aligned}\right\} \quad (2\text{-}53)$$

and so on.

The function P_n is called a *Legendre function* of the order n. Substitution of $-(n + 1)$ for n in (52) results in the same equation, showing that $P_{-(n+1)} = P_n$.

Solutions to Laplace's equation (50) can now be found by trial, using these Legendre functions. Alternatively, the solutions may be found as in the cylindrical harmonic case by a process of partial differentiation. By trial it is found that

$$V = \frac{1}{r}$$

is a solution of (50). Then differentiating partially with respect to z the following solutions are obtained:

$$\left.\begin{array}{ll} \dfrac{1}{r} & 1 \\ \dfrac{1}{r^2}\cos\theta & r\cos\theta \\ \dfrac{1}{r^3}(3\cos^2\theta - 1) & r^2(3\cos^2\theta - 1) \\ \dfrac{1}{r^4}(5\cos^3\theta - 3\cos\theta) & r^3(5\cos^3\theta - 3\cos\theta) \end{array}\right\} \quad (2\text{-}54)$$

The second set has been obtained from the first set by replacing $r^{-(n+1)}$ by r^n. The solutions to Laplace's equation in spherical co-ordinates are called *spherical harmonics.* The particular sets (54), obtained for no variation with ϕ are known more specifically as *zonal harmonics* because the potential is constant in each zone of latitude.

Zonal harmonics can be used to obtain solutions to problems in spherical co-ordinates for which there is no variation in the ϕ direction. A simple example would be that of a conducting sphere placed in a uniform field which is parallel to the z axis. The solution to this problem follows in a manner similar to that of the conducting cylinder and is left as an exercise for the student.

It is important to realize that only certain very special problems yield to an exact solution such as was obtained in the two examples above. In general, an *infinite* number of harmonic solutions would be required to satisfy the boundary conditions. However, just as any periodic function (satisfying certain conditions) may be approximated by a finite number of terms of a Fourier series, so any problem having a geometry suitable for the application of these harmonic functions may be solved approximately by an appropriate combination of a finite number of them.

The methods of this last section are also applicable in the solution of certain electromagnetic problems. Examples of such problems will be encountered in chaps. 13 and 15.

PROBLEMS

1. If a flat conducting surface could have placed on it a surface charge density $\rho_s = 1$ coulomb per square meter, what would be the value of the electric intensity E at its surface?

2. A point charge q is located a distance h above an infinite conducting plane. Using the method of images find the displacement density normal to the plane and hence show that the surface charge density on the plane is

$$\rho_s = -\frac{qh}{2\pi r^3}$$

where r is the distance from the charge q to the point on the plane. Integrate this expression over the plane to show that the total charge on its surface is $-q$.

3. Show that the capacitance of an isolated sphere of radius R is

$$4\pi\epsilon_0 R \text{ farads}$$

4. Verify that the capacitance between two spheres, whose separation d is very much larger than their radii R, is given approximately by

$$C \approx \frac{4\pi\epsilon_0 Rd}{2(d - R)} \approx 2\pi\epsilon_0 R$$

Hence show that the capacitance of a sphere above an infinite ground plane is independent of the height h above the plane when $h \gg R$.

5. In the problem of example 3, section 2.05, derive the expression for $\mathbf{E}$ at any point in the x-y plane directly, that is, by vector addition of the electric fields produced by the two charges.

6. Verify that the expression for the potential due to an electric dipole satisfies the Laplace equation.

7. Verify that the expression obtained for the potential due to two parallel oppositely charged wires, viz.,

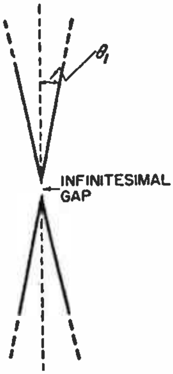


Fig. 2-20

$$V = \frac{q'}{2\pi\epsilon} \ln \frac{r_b}{r_a} = \frac{q'}{2\pi\epsilon} \ln \frac{\sqrt{\left(x - \frac{d}{2}\right)^2 + y^2}}{\sqrt{\left(x + \frac{d}{2}\right)^2 + y^2}}$$

is a solution of the Laplace equation.

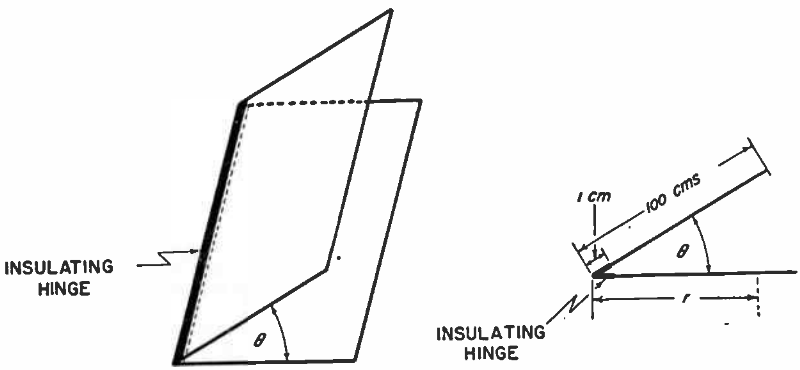
8. A *very long* cylindrical conductor of radius a has a charge q coulombs per meter distributed along its length. Find the electric intensity E in air normal to the surface of the conductor (a) by applying Gauss's law; (b) by finding the potential V and deriving the electric intensity from $E = -\operatorname{grad} V$.

9. (a) Verify that $V = \ln \cot \theta/2$ is a solution of $\nabla^2 V = 0$. (b) Hence show that the capacitance per unit length between two infinitely long coaxial cones (Fig. 2-20), placed tip to tip with an infinitesimal gap between them, is

$$C = \frac{\pi\epsilon}{\ln \cot \theta_1/2} \approx \frac{\pi\epsilon}{\ln 2/\theta_1}$$

for small angles of θ_1.

10. (a) Find the electric field distribution between the hinged plates (Fig. 2-21a) and the charge distribution on the plates in a region not too close to the edges (that is, neglect fringing). The plates are insulated at the hinge.



(a) (b)

Fig. 2-21

(b) If the plates are 1 meter wide and very long (Fig. 2-21b), estimate roughly the capacitance between them per meter length when $\theta = 10$ degrees; when $\theta = 180$ degrees. The insulating hinge extends from $r = 0$ to $r = 1$ cm.

11. The general definition for the voltage between two points in an electromagnetic field is

$$V_{ab} = \int_a^b \mathbf{E} \cdot d\mathbf{s}$$

By taking the point b to infinity, show that in an electrostatic field due to a charge q the voltage at a (with respect to the voltage at infinity) is the same as the potential at a as defined on page 34. That is, show that

$$V_{a-\infty} = \frac{q}{4\pi\epsilon R}$$

where R is the distance of a from the charge q.

12. By the methods of sec. 2.13 derive a set of solutions to Laplace's equation (a) in cylindrical co-ordinates, starting with

$$V = k\phi$$

(b) in spherical co-ordinates starting with

$$V = k \ln \tan \frac{\theta}{2}$$

BIBLIOGRAPHY

Page, L., and N. I. Adams, Jr., *Principles of Electricity*, D. Van Nostrand Co., New York, 1931.

Skilling, H. H., *Fundamentals of Electric Waves*, John Wiley and Sons, New York, 1948.

Jeans, J. H., *The Mathematical Theory of Electricity and Magnetism*, Cambridge University Press, London, 1946.

Smythe, W. R., *Static and Dynamic Electricity*, McGraw-Hill Book Co., Inc., New York, 1939.

Pohl, R. W., *Physical Principles of Electricity and Magnetism*, Blackie and Son, Limited, London, 1930.

Howe, G. W. O., "On the Capacity of Radio-Telegraphic Antennae," *The Electrician*, p. 829, Aug. 28, 1914; also LXXV, p. 870.

CHAPTER 3

THE STEADY MAGNETIC FIELD

Electric charges at rest produce an electric field—the electrostatic field. Electric charges in motion, that is, electric currents, produce a *magnetic field*. This is evidenced by the fact that in the region about a wire carrying a current, each end of a magnetic compass needle experiences a force dependent upon the magnitude of the current. There is said to be a magnetic field about the wire, and the direction of the magnetic field is taken to be that in which the north-seeking pole of the compass needle is urged. The intensity H of the magnetic field was originally defined in a manner similar to that for the electric intensity E. A unit magnetic pole was first defined in terms of the force between two similar poles, and then the magnetic intensity was defined in terms of the force per unit pole. In electromagnetic wave theory, magnetic fields due to electric currents are of chief concern and the effects of permanent magnets are of little importance. Therefore the above approach will be discarded for one that leads more directly to a solution of the type of problems encountered in electromagnetic engineering.

3.01 Theories of the Magnetic Field. It is possible to develop a quantitative theory of the magnetic field from any of several different starting points. Rowland's experiments showed that moving charges produce magnetic effects. Therefore a theory based upon the magnetic forces between individual moving charges would be logical. In this theory permanent-magnet effects are ascribed to the motion of external electrons about the atomic nuclei. This theory is used in modern physics, and can be developed to answer most of the questions that arise in connection with magnetism. Some such fundamental approach is required whenever it is necessary to deal with individual charges, but in most engineering problems where only macroscopic effects are considered such a procedure involves an unnecessary complexity. The motion of a single electron in a wire is an erratic and highly unpredictable affair, subject to forces that vary greatly in the small length of interatomic distances. Yet, the intelligent sophomore experiences little difficulty in predicting with fair accuracy the statistical *average* motion of millions of electrons by a simple application of Ohm's law. For most engineering problems it will be the magnetic effect of *currents* rather than the motion of individual charges that will be of importance, and it would seem reasonable to use the forces between currents as a starting point. Ampere's experiments on the force between current-carrying conductors form a logical starting point for this development and lead to quite satisfactory engineering definitions, especially when the end result desired is in terms of mechanical forces. In electromagnetic wave theory, however, primary interest is in the relations between electric and magnetic fields, and a different starting point proves to be convenient. This starting point is Faraday's induction law, which relates the magnetic flux through a closed path to the voltage induced around the path. This relation, which *defines* magnetic flux in terms of a measurable electric voltage, is the starting point that will be used in the present discussion of magnetic fields.

Still another attack that is often used in electromagnetic theory is to *postulate* a vector potential due to the currents, and then obtain a magnetic field in terms of this potential. This vector-potential method has the marked advantage that it can be readily extended to the general case where the currents vary with time—the electromagnetic field—and in this latter case it will also yield directly the electric field produced by changing currents. In general, use of the vector-potential method simplifies the mathematical analyses and facilitates the solution of electromagnetic problems. Therefore it will be developed and used. However, instead of starting with a postulated potential and deducing from it the electric and magnetic fields, the reverse procedure will be used. The electric and magnetic vectors will be defined in terms of relations derived from experiments, often performed under restricted conditions. These definitions will then be generalized for use in the electromagnetic field, and in the process a potential will be found such that the space and time derivatives of this potential will give the magnetic and electric fields. The generalizations may be considered valid as long as conclusions derived from them agree with subsequent experiment.

3.02 Magnetic Flux Φ.

In the experimental setup indicated in Fig. 3-1, a ballistic galvanometer is connected to a loop placed near a long straight wire, carrying a current I. Probing with a magnetic compass needle shows that there is a magnetic field in the region about the wire. At the position of the loop shown, the direction of the field is out of the plane of the paper for an upward flow of the current I. If now the current I is reduced to zero, the galvanometer is deflected, the amount of the deflection being independent of the rate at which the current is reduced to zero, so long as the time required is short compared with the period of the galvanometer. The current I_g through the galvanometer flows as a result of a voltage V "induced" in the loop and is given by

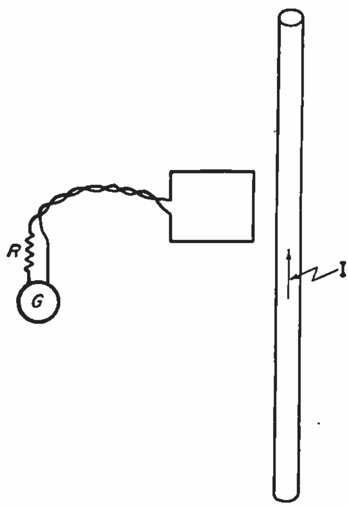


FIG. 3-1. Measurement of magnetic flux.

$$I_g = \frac{V}{R}$$

where R is the total resistance in the galvanometer circuit (R is a very large resistance). The galvanometer deflection is a measure of the charge Q or the time integral of the current through it, so that

$$Q = \int_0^t I_g\,dt = \frac{1}{R}\int_0^t V\,dt$$

is an experimentally determinable quantity. Magnetic flux Φ through the loop is then defined as the time integral of voltage induced in the loop throughout the interval during which the magnetic field is being established; or having been established, as the time integral of voltage throughout the interval in which the field is being reduced to zero. (These quantities are equal but of opposite sign.) That is

$$\Phi = \pm \int_0^t V\,dt \tag{3-1}$$

where the time interval 0 to t is that required to establish the field or reduce it to zero. Differentiating with respect to time gives

$$V = -\frac{\partial \Phi}{\partial t} \tag{3-2}$$

which is Faraday's induction law. Consistent with a right-hand co-ordinate system, the negative sign has been used to indicate that when the flux is increasing in the positive direction (out of the paper through the loop in Fig. 3-1), the induced voltage occurs in a clockwise direction. It is evident from eq. (1) or (2) that the unit of magnetic flux is the volt-second; this unit has been named the *weber*.

3.03 Magnetic-flux Density B. The magnetic flux per unit area through a loop of small area is called the *magnetic-flux density* **B** at the location of the loop. Because the flux through the loop depends upon the orientation of the loop as well as upon its area, magnetic-flux density is a *vector* quantity. The direction of **B** is taken as the normal to the plane of the loop when oriented to enclose maximum flux. The positive sense of **B** is the direction of the magnetic field at the point in question. The unit of magnetic-flux density is the *weber per square meter* or the *volt-second per square meter*. The magnetic flux through any surface is the surface integral of the normal component of **B**, that is

$$\Phi = \int_S B_n\,da = \int_S \mathbf{B} \cdot d\mathbf{a}$$

3.04 Magnetic Intensity H and Magnetomotive Force $\mathfrak{F}$. Using a small probe loop and galvanometer as in Fig. 3-1, it is possible to determine **B** at all points in a region about a long current-carrying wire. Experiment shows that for a homogeneous medium, B is related to the current I through

$$B \propto \frac{\mu I}{r} \tag{3-3}$$

where r is the distance from the wire and μ is a constant that depends upon the medium. The constant μ, called the *permeability* of the medium, may be written as

$$\mu = \mu_r \mu_v$$

where μ_v is the absolute permeability of a vacuum, and μ_r is the *relative permeability* (relative to a vacuum). μ_v is the basic *defined* electrical unit, which has been assigned the value

$$\mu_v = 4\pi \times 10^{-7} \qquad \text{henry/m}$$

in the rationalized MKS system of units. Using this value of μ_v, and probing the field about the wire in a vacuum for which $\mu_r = 1$ (or in air for which $\mu_r \approx 1$), the proportionality factor in (3) is found to be $1/2\pi$, so that the relation becomes

$$B = \frac{\mu I}{2\pi r} = \mu H \tag{3-4}$$

where

$$H = \frac{I}{2\pi r} \qquad \text{amp/m} \tag{3-5}$$

The *magnetic intensity* H is thus defined by this relation in terms of the current which produces it and the geometry of the system. Magnetic intensity is a vector quantity, having the same direction as the magnetic-flux density, so the equality expressed by (4) can be stated as the vector relation

$$\mathbf{B} = \mu \mathbf{H} \tag{3-4a}$$

Under the conditions of the above (long-wire) experiment, H, the magnitude of the magnetic intensity is independent of the permeability of the medium, depending only on the current and distance from it, while B is dependent on the permeability of the medium. In this sense H may be pictured as a magnetic intensity that drives a "resultant" flux density through the medium (but this is not the only possible viewpoint). Although no longer defined in terms of unit poles, the relative value of H at any point may be indicated by the force on one end of a magnetized compass needle.

The line integral

$$\mathfrak{F} = \int_a^b \mathbf{H} \cdot d\mathbf{s}$$

is defined as the *magnetomotive force* between the points a and b. For a circular path about the wire, with the wire at the center, H

has the constant value $I/2\pi r$ and is directed along the path, so that

$$\mathfrak{F} = \oint \mathbf{H} \cdot d\mathbf{s} = I \tag{3-6}$$

It is easily demonstrated that this same result (6) will be obtained for *any* closed path about the current. Equation (6) is *Ampere's work law.* The positive directions (or senses) of magnetomotive force and current are related by the familiar "right-hand rule."

Ampere's work law makes it easy to compute H in certain problems. For example, consider the toroidal coil of Fig. 3-2, consisting

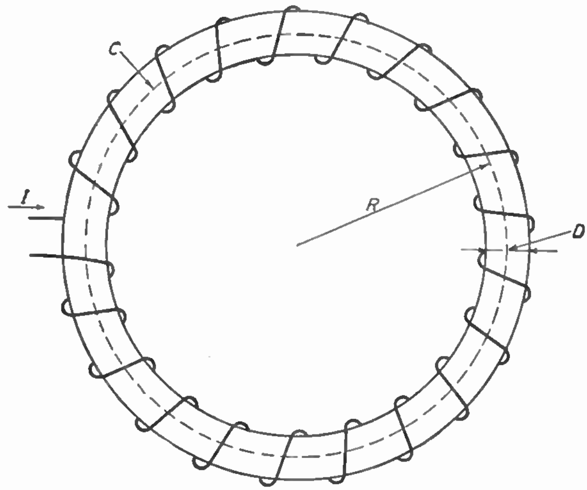


FIG. 3-2. Toroidal coil.

of a large number of closely spaced turns on a tubular core. For any closed path C taken around the core inside the winding, the magnetomotive force will be

$$\mathfrak{F} = nI$$

where n is the number of turns and, therefore, the number of times the path links with the current I. If D, the thickness of the core, is small compared with R, the radius of the ring, the radii of all circular paths through the core are approximately equal to R, so that at any point within the core

$$H \approx \frac{\mathfrak{F}}{2\pi R} = \frac{nI}{2\pi R} = \frac{nI}{l} \quad \text{ampere turns/m} \tag{3-7}$$

with $l = 2\pi R$ denoting the length of the coil. The magnetic intensity is nearly uniform throughout the cross section of the core and is equal to the ampere turns per unit length.

Another example, in which **H** is simply related to the current that produces it, is the case of two very large closely spaced parallel planes carrying equal and oppositely directed currents (Fig. 3-3). The magnetic field is confined to the region between the planes and is found to be uniform (except near the edges) and independent of the distance apart of the planes as long as this distance is small

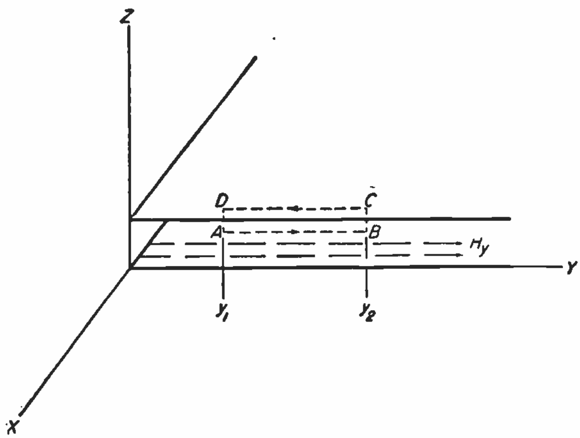


Fig. 3-3. Parallel-plane conductors.

compared with the other dimensions. In Fig. 3-3 the current is assumed to be flowing in the positive x direction (outward) in the upper plate.

Then if J_x represents the current *per meter width* flowing in this plate, Ampere's work law states that

$$\oint_{ABCDA} \mathbf{H} \cdot d\mathbf{s} = H_y(y_2 - y_1) = J_x(y_2 - y_1)$$

from which

$$H_y = J_x \qquad \text{amp/m} \qquad (3\text{-}8)$$

The magnetic intensity is equal in magnitude to the linear current density (amperes per meter width) flowing in each of the planes. It is parallel to the planes, but perpendicular to the direction of current flow.

3.05 Permeability μ. In each of the above examples the magnetic intensity H has been related directly to the current that produces it. Using the toroidal coil example the magnetic flux Φ and therefore the magnetic flux density $\mathbf{B}$ within the core can be measured; the magnetomotive force $\oint \mathbf{H} \cdot d\mathbf{s}$, and therefore H, is known in terms of the current; therefore μ, and hence μ_r, the relative permeability of the medium composing the core, can be determined from the relations

$$\mu = \frac{B}{H}; \qquad \mu = \mu_r \mu_v; \qquad \mu_v = 4\pi \times 10^{-7} \quad \text{henry/m} \tag{3-9}$$

For air and most materials the relative permeability is very nearly unity. For *paramagnetic* substances μ_r is very slightly greater than unity; thus for air it is 1.00000038 and for aluminum it is 1.000023. For *diamagnetic* substances μ_r is slightly less than unity; for copper $(1 - 8.8 \times 10^{-6})$; for water $(1 - 9.0 \times 10^{-6})$. However, for that exceptional class of materials known as *ferromagnetic* materials (iron and certain alloys) the relative permeability may have a value of several hundred or even several thousand. In general, the permeability of these materials is not constant but depends upon the strength of the magnetic field and upon their past magnetic history. However for most applications of interest in electromagnetic wave theory, the range of flux densities involved is small enough that μ may be considered constant.

3.06 Energy Stored in a Magnetic Field. It is found experimentally that a certain amount of work is required to establish a current in a circuit. This work is done in establishing the current against the electromotive force induced in the circuit by the increasing magnetic flux, and the energy thus transferred to the circuit is said to be stored in the magnetic field. The amount of the energy so stored can be determined in terms of the extent and intensity of the magnetic field by considering the elementary example of current flow in a toroidal coil. In this case, when the turns are closely spaced, the magnetic field is confined to the core of the toroid, and the magnetic intensity is given by

$$H = \frac{nI}{l}$$

where n is the number of turns and l is the mean length of path through the core of the coil. The back voltage induced in the coil is, by definition of magnetic flux Φ,

$$\begin{aligned} V &= -n\frac{d\Phi}{dt} \\ &= -nA\frac{dB}{dt} \end{aligned}$$

where B is the magnetic flux density, and A is the cross-sectional area of the core. The work done in establishing the current I in the coil is

$$\begin{aligned} W &= -\int_0^{t_1} VI\,dt \\ &= \int_0^{t_1} lAH\frac{dB}{dt}\,dt \\ &= \int_0^{H_1} \mu lAH\,dH \\ &= lA\left[\frac{\mu H_1^2}{2}\right] \end{aligned} \qquad (3\text{-}10)$$

This is the total energy stored in the field, and since lA is the volume of the region in which the magnetic field exists, it is inferred that the quantity

$$\frac{\mu H_1^2}{2}$$

represents the energy density of the magnetic field. Whether or not it is considered desirable to ascribe a certain energy density to each small volume of space and so "locate" the energy, it is nevertheless true in general that the quantity $\mu H^2/2$, when integrated over the whole volume (in which H has value), does give the correct value for the total stored magnetic energy of the system.

3.07 Ampere's Law for a Current Element. When a current I flows in a closed circuit the magnetic intensity **H** at any point is a result of this flow in the complete circuit. For computational purposes it is convenient to consider the total magnetic intensity at any point as the sum of contributions from elemental lengths ds of the circuit, each carrying the current I. The quantity $I\,ds$ is called a *current element*. It is a vector quantity having the direction of the current, or what amounts to the same thing, the direction of the element ds in which the current flows. This may be indicated by writing $I\,ds$ or alternatively $I\,d\mathbf{s}$. The two notations are used interchangeably and to suit convenience in the particular problem.

The magnitude of the contribution to H from each current element $I\,ds$ cannot be measured directly, but is inferred from experimental results to be

$$dH = \frac{I\,ds \sin\psi}{4\pi r^2} \tag{3-11}$$

r is the distance measured outward from the current element $I\,ds$ to the point p at which H is being evaluated (Fig. 3-4). ψ is the angle between the direction of $I\,d\mathbf{s}$ and the direction of $\mathbf{r}$. The direction of $\mathbf{H}$ is perpendicular to the plane containing $I\,d\mathbf{s}$ and $\mathbf{r}$, in the direction in which a right-hand screw would progress in turning from $I\,d\mathbf{s}$ to $\mathbf{r}$. This complete statement can be written in vector notation simply as

$$d\mathbf{H} = \frac{I\,d\mathbf{s} \times \mathbf{u}_r}{4\pi r^2} \tag{3-12}$$

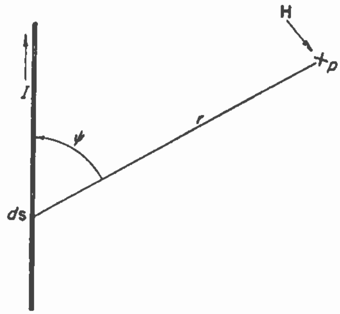


FIG. 3-4

where $\mathbf{u}_r$ is a unit vector in the $\mathbf{r}$ direction. Equation (12) is known as *Ampere's law for a current element* (or sometimes as the *Biot-Savart law*).

The total magnetic intensity $\mathbf{H}$ at a point p will be the sum or integration of the contributions from all the current elements of the circuit and will be

$$\mathbf{H} = \oint \frac{I\,d\mathbf{s} \times \mathbf{u}_r}{4\pi r^2} \tag{3-13}$$

Magnetic Fields of Some Simple Circuits. The magnetic intensity $\mathbf{H}$ at any point due to current flow in a circuit can be obtained by summing the contributions from the current elements that make up the circuit. This is not always a simple task but there are a few problems in which conditions of symmetry make it relatively easy to obtain an answer.

Example 1: *Field at the Center of a Circular Loop.* The contribution to **H** at the center due to any element $I\,ds$ will be directed vertically upward and will have a magnitude

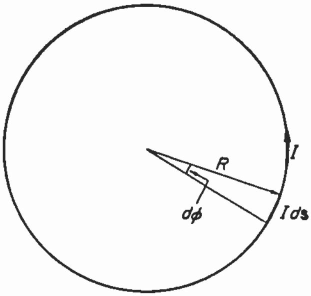


Fig. 3-5

$$dH = \frac{I\,ds}{4\pi R^2} = \frac{IR\,d\phi}{4\pi R^2} = \frac{I\,d\phi}{4\pi R}$$

The total intensity at the center will be

$$H = \int_0^{2\pi} \frac{I\,d\phi}{4\pi R} = \frac{I}{2R}$$

Field about a Long Straight Wire. The magnetic intensity at a distance R from a very long (infinitely long for the purposes of this problem) straight wire carrying a current I can also be obtained by summing the contributions from the individual current elements. This is left as a problem for the student.

3.08 Magnetic Vector Potential. In the electric field it was found desirable to introduce the concept of potential. In that case the electric potential was a space function that depended upon the magnitude and location of the charges, the charges being the *sources* of the electric field. The intensity of the electric field was obtained from the potential V by taking the gradient or space derivative of V. This procedure was often found to be much simpler than that of trying to obtain **E** directly in terms of the magnitude and location of the charges.

Similarly in the case of the magnetic field it would be desirable to be able to set up a *magnetic potential*, the space derivative of which would give the magnetic intensity **H**. Corresponding to the individual charges in the electric field case, the sources of the magnetic field would be the current elements $I\,ds$ of the circuits that produce the field. The magnetic potential being sought would therefore depend upon these current elements. Assuming that a suitable magnetic potential can be found, the properties that such a potential must possess are easily determined by simple reasoning.

Because the magnetic intensity H that is to be derived from the potential is proportional to the strength of the current element $I\,ds$, the potential itself must be proportional to $I\,ds$. Because the magnetic intensity H due to a current element varies inversely as the *square* of the distance r from the element (Ampere's law), the magnetic potential due to the current elements must vary inversely as the *first* power of the distance because H is to be obtained by taking the space derivative of the potential. This is equivalent to dividing by r as far as dimensions are concerned. The electrostatic potential due to charges was a scalar quantity. This was adequate in that case, because the charges themselves were scalars having magnitudes only. In the present case, the current elements have directions as well as magnitudes, and it is necessary that this additional information on the *direction* of the source be contained in the potential due to the source. Therefore the potential in this case must be a *vector* quantity, the direction of which will somehow be related to the direction of the current-element source. If this *vector magnetic potential* is designated by the vector **A**, then it should be possible to obtain **H** as the space derivative of **A**. There are two possible space-derivative operations on a vector quantity, namely the divergence and the curl. The divergence operation yields a scalar quantity, whereas the curl operation yields a vector quantity. Inasmuch as the resulting magnetic intensity **H** is a vector quantity, the curl is the only space-derivative operation which can be used. Therefore, if there is a suitable vector magnetic potential **A**, the magnetic intensity will be derived from it by

$$\mathbf{H} = \text{curl } \mathbf{A} \tag{3-14}$$

As indicated above, the relation between the magnetic vector potential and the current element source must be of the form

$$d\mathbf{A} = k\left(\frac{I\,d\mathbf{s}}{r}\right) \tag{3-15}$$

where the constant k is still to be determined. With one eye on eq. (12) a reasonable guess for the expression for the vector potential of a current element $I\,d\mathbf{s}$ would appear to be

$$d\mathbf{A} = \frac{I\,d\mathbf{s}}{4\pi r} \tag{3-16}$$

3.09 Vector Magnetic Potential of a Current Element. The expression (16) for the magnetic vector potential of a current element was obtained by a combination of logical reasoning and straight guesswork. It remains to be shown that (16) is indeed the required expression. This can be done by inserting (16) in (14) and showing that the result is equivalent to eq. (12). However it is instructive to obtain the expression for $d\mathbf{A}$ directly from a restatement of Ampere's law for the current element.

Consider a current element $I\,d\mathbf{s}$ located at the origin of the co-ordinate system and having components $I\,dx$, $I\,dy$, and $I\,dz$,

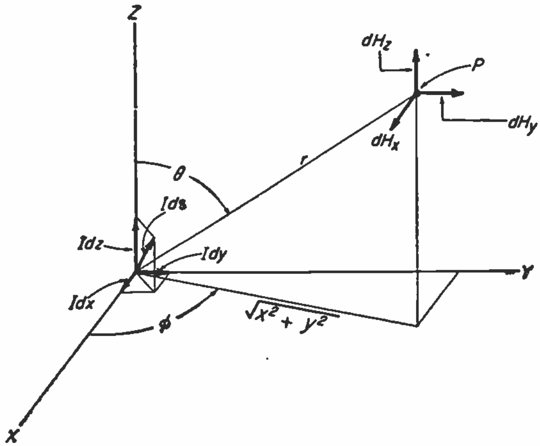


Fig. 3-6

along the respective axes (Fig. 3-6). The magnetic intensity at a point P as given by Ampere's law for the current element is

$$d\mathbf{H} = \frac{I\,d\mathbf{s} \times \mathbf{u}_r}{4\pi r^2} \tag{3-12}$$

As indicated by the cross product, $\mathbf{H}$ is perpendicular to the plane containing $d\mathbf{s}$ and $\mathbf{u}_r$. The magnitude of $d\mathbf{H}$ is given by

$$dH = \frac{I\,ds\sin\psi}{4\pi r^2} \tag{3-17}$$

where ψ is the angle between $d\mathbf{s}$ and $\mathbf{r}$.

The magnetic intensity at P can be considered in terms of its components dH_x, dH_y and dH_z. The magnetic intensity in the x direction is due, in part, to a contribution from $I\,dz$ and in part to a contribution from $I\,dy$. The component $I\,dx$ contributes nothing to the x component of the magnetic intensity since a magnetic field is always perpendicular to the current producing it.

The contribution to dH_x due to $I\,dz$ is, by (17),

$$-\frac{I\,dz}{4\pi r^2}\sin\theta\sin\phi = -\frac{I\,dz}{4\pi r^2}\cdot\frac{\sqrt{x^2+y^2}}{r}\cdot\frac{y}{\sqrt{x^2+y^2}}$$
$$= -\frac{I\,dz}{4\pi}\cdot\frac{y}{r^3}$$

Similarly the contribution to dH_x at P due to $I\,dy$ is

$$\frac{I\,dy}{4\pi}\cdot\frac{z}{r^3}$$

The total dH_x at P due to the current element $I\,ds$ is therefore

$$4\pi\,dH_x = -I\,dz\cdot\frac{y}{r^3} + I\,dy\cdot\frac{z}{r^3} \tag{3-18}$$

Now the first of these terms is the partial derivative with respect to y of $I\,dz/r$ for

$$\frac{\partial}{\partial y}\left(\frac{I\,dz}{r}\right) = I\,dz\,\frac{\partial}{\partial y}\frac{1}{(x^2+y^2+z^2)^{1/2}} = -I\,dz\,\frac{y}{(x^2+y^2+z^2)^{3/2}}$$
$$= -I\,dz\cdot\frac{y}{r^3}$$

Also the second term is the negative of the partial of $I\,dy/r$ with respect to z. Therefore (18) may be written

$$4\pi\,dH_x = \frac{\partial}{\partial y}\left(\frac{I\,dz}{r}\right) - \frac{\partial}{\partial z}\left(\frac{I\,dy}{r}\right) \tag{3-19}$$

In a similar manner the y and z components of $d\mathbf{H}$ can be written in terms of the appropriate derivatives of $I\,d\mathbf{s}/r$. The complete statement would be

$$\begin{aligned}
4\pi\,dH_x &= \frac{\partial}{\partial y}\left(\frac{I\,dz}{r}\right) - \frac{\partial}{\partial z}\left(\frac{I\,dy}{r}\right)\\
4\pi\,dH_y &= \frac{\partial}{\partial z}\left(\frac{I\,dx}{r}\right) - \frac{\partial}{\partial x}\left(\frac{I\,dz}{r}\right)\\
4\pi\,dH_z &= \frac{\partial}{\partial x}\left(\frac{I\,dy}{r}\right) - \frac{\partial}{\partial y}\left(\frac{I\,dx}{r}\right)
\end{aligned} \tag{3-20}$$

The right-hand sides of these equations are the three components of curl $I\,d\mathbf{s}/r$. Therefore eqs. (20) may be written

$$d\mathbf{H} = \operatorname{curl}\frac{I\,d\mathbf{s}}{4\pi r} \tag{3-21}$$

It is evident that for the current element $I\,ds$, the vector magnetic potential is

$$d\mathbf{A} = \frac{I\,d\mathbf{s}}{4\pi r}$$

It is proportional to the current I and to the length of the element ds, and is inversely proportional to the distance r from the current element. It has the *same* direction as the current producing it.

The vector magnetic potential (usually called just vector potential) due to current flow in a complete circuit is obtained as a summation or integration of vector potentials caused by all the current elements that comprise the circuit. That is

$$\mathbf{A} = \int \frac{I\,d\mathbf{s}}{4\pi r} \tag{3-22}$$

where the integration extends over the complete circuit in which I flows. As mentioned previously, the *direction* of a current element can be indicated by making either ds or I the vector quantity. In the latter case the expression for $\mathbf{A}$ would be

$$\mathbf{A} = \int \frac{\mathbf{I}\,ds}{4\pi r}$$

This expression can be written in a more general form by replacing the current I by a current density i and then integrating over the *volume* in which this current density exists. Then the expression for the vector potential $\mathbf{A}$ is

$$\mathbf{A} = \int_V \frac{i\,dV}{4\pi r}$$

This reduces to the previous expression when the current flows in a filamentary circuit.

Example 2: *Magnetic Field about a Long Straight Wire.* Using the vector potential, let it be required to find the magnetic intensity about a long straight wire carrying a current I.

The general expression for vector potential is

$$\mathbf{A} = \int_V \frac{i\,dV}{4\pi r}$$

For this problem the current density i, integrated over the cross section of the wire, gives the total current I. Also the current is entirely in the z direction (Fig. 3-7) so that **A** has only one component, A_z.
Then

$$A_z = \frac{1}{4\pi} \int_{-L}^{+L} \frac{I\,dz}{r}$$

If the point P is taken in the y-z plane, $r = \sqrt{z^2 + y^2}$ and

$$\begin{aligned} A_z &= \frac{1}{2\pi} \int_0^L \frac{I}{\sqrt{y^2 + z^2}}\,dz \\ &= \frac{I}{2\pi} [\ln (z + \sqrt{y^2 + z^2})]_0^L \\ &= \frac{I}{2\pi} [\ln (L + \sqrt{y^2 + L^2}) - \ln y] \end{aligned}$$

For $L \gg y$, the vector potential is given approximately by

$$A_z \approx \frac{I}{2\pi} (\ln 2L - \ln y)$$

Then for a point in the y-z plane

$$H_x = \text{curl}_x\, A_z = \frac{\partial A_z}{\partial y} = -\frac{I}{2\pi y}$$

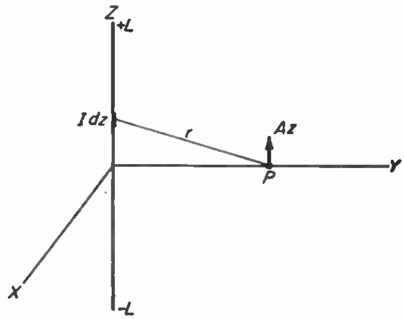


FIG. 3-7. Vector potential about a long straight wire.

The lines of magnetic intensity will be circles about the wire, that is in the ϕ direction. For any arbitrary point P, not necessarily in the y-z plane,

$$H_\phi = \frac{I}{2\pi R}$$

where $R = \sqrt{x^2 + y^2}$ is the distance of the point P from the wire.

EXAMPLE 3: *Magnetic Fields Due to Long Parallel Wires.* Let it be required to derive the expressions for the magnetic field about two long straight parallel wires, carrying equal and oppositely directed currents. Start with $A_z \approx (I/2\pi)(\ln 2L - \ln R)$ for a single wire.

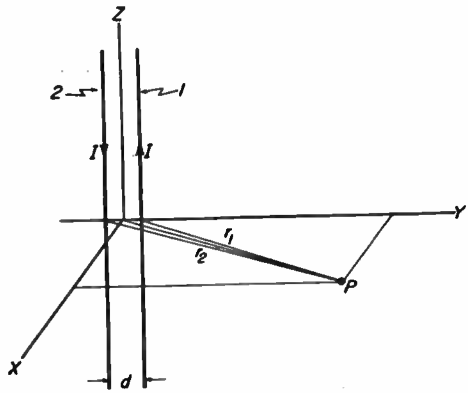


FIG. 3-8. Long parallel wires.

The total vector potential at the point P (Fig. 3-8) will be

$$A_z = A_{z_1} + A_{z_2}$$

where
$$A_{z_1} = \frac{I}{2\pi}(\ln 2L - \ln r_1)$$

and
$$A_{z_2} = -\frac{I}{2\pi}(\ln 2L - \ln r_2)$$

Therefore
$$A_z = \frac{I}{2\pi}(\ln r_2 - \ln r_1)$$

From the figure
$$r_1 = \sqrt{\left(y - \frac{d}{2}\right)^2 + x^2}$$

and
$$r_2 = \sqrt{\left(y + \frac{d}{2}\right)^2 + x^2}$$

so that
$$\frac{\partial r_1}{\partial y} = \frac{y - (d/2)}{r_1} \quad \text{and} \quad \frac{\partial r_2}{\partial y} = \frac{y + (d/2)}{r_2}$$

$$\frac{\partial r_1}{\partial x} = \frac{x}{r_1} \qquad \frac{\partial r_2}{\partial x} = \frac{x}{r_2}$$

Then the x component of $\mathbf{H}$ will be obtained from

$$H_x = \text{curl}_x\, A_z = \frac{\partial A_z}{\partial y}$$
$$= \frac{I}{2\pi}\left[\frac{y + (d/2)}{r_2^2} - \frac{y - (d/2)}{r_1^2}\right]$$

The y component of **H** is

$$H_y = \text{curl}_y\, A_z = -\frac{\partial A_z}{\partial x}$$
$$= -\frac{I}{2\pi}\left(\frac{x}{r_2^2} - \frac{x}{r_1^2}\right)$$

Example 4: *The Magnetic Dipole.* A small circular loop carries a current I. Let it be required to find the magnetic field at distances from the loop that are large compared with the dimensions of the loop. Without loss of generality the point p may be assumed to lie in the y-z plane (Fig. 3-9). The vector potential at the point p will have a component in the

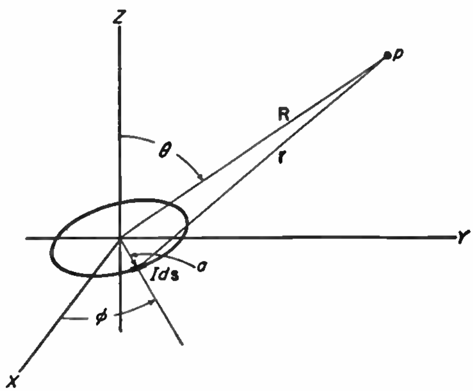


Fig. 3-9. Small circular loop.

ϕ direction only and for p in the y-z plane this means that $A = A_\phi = -A_x$. The contribution to A_x from a current element $I\,ds$ will be

$$dA_x = -\frac{I\,ds \sin\phi}{4\pi r}$$

The total vector potential at p will be

$$A_x = -\oint \frac{I \sin\phi\, ds}{4\pi r} = -\frac{Ia}{4\pi}\int_{\phi=0}^{\phi=2\pi} \frac{\sin\phi\, d\phi}{r}$$

Now $\mathbf{r} = \mathbf{R} - \mathbf{a}$

$$r^2 = \mathbf{r}\cdot\mathbf{r} = (\mathbf{R}-\mathbf{a})\cdot(\mathbf{R}-\mathbf{a}) = R^2 - 2\mathbf{R}\cdot\mathbf{a} + a^2$$

The quantity $\mathbf{R}\cdot\mathbf{a}$ is R times the projection of **a** on **R** and has a value

$$\mathbf{R}\cdot\mathbf{a} = Ra \sin\phi \sin\theta$$

Then
$$\frac{1}{r} = \frac{1}{\sqrt{R^2 - 2Ra \sin \phi \sin \theta + a^2}}$$
$$= R^{-1}\left(1 - \frac{2a}{R} \sin \phi \sin \theta + \frac{a^2}{R^2}\right)^{-\frac{1}{2}}$$

For $R \gg a$, this is given approximately by the first two terms of the binomial expansion, that is

$$\frac{1}{r} \approx R^{-1}\left[1 - \frac{1}{2}\left(-\frac{2a}{R} \sin \phi \sin \theta + \frac{a^2}{R^2}\right)\right]$$

Then
$$A_x = -\frac{Ia}{4\pi} \int_0^{2\pi} \frac{\sin \phi}{R}\left(1 + \frac{a}{R} \sin \phi \sin \theta - \frac{a^2}{2R^2}\right) d\phi$$
$$= -\frac{Ia^2 \sin \theta}{4\pi R^2} \int_0^{2\pi} \sin^2 \phi \, d\phi$$
$$= -\frac{I(\pi a^2) \sin \theta}{4\pi R^2}$$

For an arbitrary location of the point P, not necessarily in the y-z plane, we may write

$$A_\phi = \frac{I(\pi a^2) \sin \theta}{4\pi R^2}$$

There will be two components of $\mathbf{H}$ at the point p. Expanding $\mathbf{H} = \text{curl } \mathbf{A}$ in spherical co-ordinates gives

$$H_\theta = -\frac{1}{r}\frac{\partial}{\partial r}(rA_\phi) = \frac{I(\pi a^2) \sin \theta}{4\pi R^3}$$
$$H_r = \frac{1}{r \sin \theta}\frac{\partial}{\partial \theta}(\sin \theta \, A_\phi) = \frac{I(\pi a^2) \cos \theta}{2\pi R^3}$$

If these expressions are compared with those for the electric dipole (page 39) it will be seen that they are identical when the electric moment ql of the electric dipole is replaced by $\pi a^2 I$ for the loop. πa^2 is the area of the loop, and the product of this area and the current I is known as the *magnetic moment* of the loop. A small loop such as this is often referred to as a *magnetic dipole*.

It will have been observed in the examples above and in the problems at the end of the chapter, that usually little time or labor is saved by using the vector-potential method. Indeed for simple problems the solution can often be obtained more quickly by solving directly for H. This is a common experience encountered in using a new and powerful tool. The simple problem often yields more readily to simple tools. However when the problems become more complex, as they do when time-varying fields are considered, the real power and true worth of the vector-potential method will become apparent.

3.10 Analogies between Electric and Magnetic Fields. It is natural to draw analogies between the electric and magnetic fields. Such analogies are useful in helping to maintain orderly thought processes and often make it possible to arrive at conclusions quickly by comparison with results already obtained in a different but analogous problem. There are several possible analogies that can be drawn between electric and magnetic fields, but two of these are particularly applicable to later work in the (time-varying) electromagnetic field. The first analogy considers **D** and **H** as analogous quantities and **E** and **B** as analogous quantities. This is based on consideration of the fact that displacement density **D** is related directly to its source, the charge, and is independent of the characteristics of the (homogeneous) medium in which the charge is immersed. Similarly the magnetic vector **H** can be related directly to its source, the current, and is independent of the (homogeneous) medium in which the magnetic field exists. The vectors **E** and **B** are also related to their respective sources, charge and current, but show a dependence on the characteristics of the medium, that is on the dielectric constant, and magnetic permeability respectively. This analogy is correct in the sense that it is self-consistent and can be made to give useful interpretations. The second analogy, which is equally valid, considers **E** and **H** as analogous and **D** and **B** as analogous. It is no more "correct" than the first analogy, but has the advantage in electromagnetic field theory that it gives a symmetry to Maxwell's equations that otherwise would be lacking. Inasmuch as these equations form the starting point for every problem of the electromagnetic field, this is a very useful result. Two simple experiments serve to point up this analogy. In Fig. 3-10a voltage V produces an electric field E in the space between the condenser plates. E is equal to V/d and is independent of the dielectric constant ϵ of the dielectric. However the displacement density depends upon ϵ (for a constant applied voltage and therefore constant E) and is given by $D = \epsilon E$. In Fig. 3-10b the current I results in a magnetomotive force nI around the closed path l.

The magnetic intensity H within the core is equal to nI/l. The magnetic flux density B depends upon the permeability μ (for a constant applied mmf) and is given by $B = \mu H$. In this analogy E and H are sometimes pictured as electric and magnetic intensities or forces that result in electric and magnetic flux densities, D and B respectively.

In the above experiments, if the charge Q (instead of the voltage) is held constant in Fig. 3-10a, and the current is held constant as

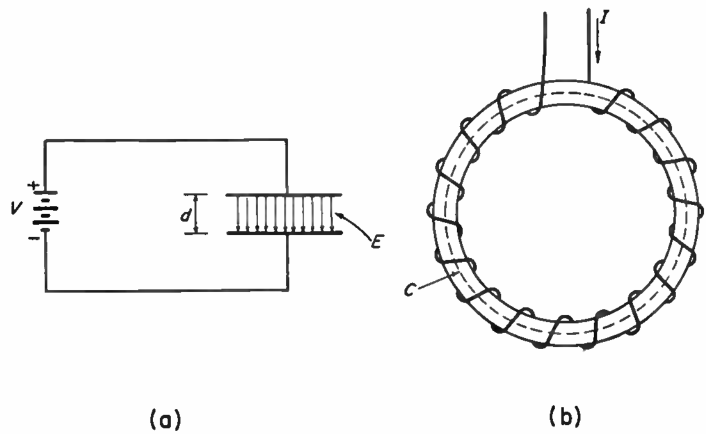


FIG. 3-10. Circuits illustrating analogies between electric and magnetic fields.

before in Fig. 3-10b, then the first analogy results. That is, D and H are the analogous quantities that remain unchanged for different dielectric and core materials.

STEADY MAGNETIC-FIELD PROBLEMS

1. Starting with Ampere's law for a current element, show that the magnetic intensity at a distance R from a very long straight wire carrying a current I amperes is given by

$$H = \frac{I}{2\pi R}$$

2. Verify that *within* a conductor carrying a current I the magnetic intensity at a distance r from the center of the wire is given by

$$H = \frac{Ir}{2\pi R^2}$$

where R is the radius of the wire. The current density is constant across the cross section of the conductor.

3. Verify that expressions (3-14) and (3-16) combine to give the same result for magnetic intensity as is given by (3-12); that is, verify that the curl of the expression for vector potential due to a current element does indeed yield the magnetic intensity as given by Ampere's law. (Suggestion: Solve for the special case of a point in the y-z plane, and then generalize).

4. A very long thin sheet of copper having a width b meters carries a direct current I in the direction of its length. Show that if the sheet is assumed to lie in the x-z plane with the z axis along its center line, the magnetic field about the strip will be given by

$$H_x = -\frac{I}{2\pi b}\left(\tan^{-1}\frac{\frac{b}{2}+x}{y} + \tan^{-1}\frac{\frac{b}{2}-x}{y}\right)$$

$$H_y = \frac{I}{4\pi b}\ln\left[\frac{\left(\frac{b}{2}+x\right)^2 + y^2}{\left(\frac{b}{2}-x\right)^2 + y^2}\right]$$

(NOTE: Solve by first setting up the vector-potential due to long narrow strips.)

5. By setting up the statement of Ampere's work law for elemental areas in cylindrical co-ordinates derive the expansion for curl **H** in these co-ordinates.

6. Show that the answers to Problem 4 agree with the answers to Problem 1 for (a) a point on the y axis when $y \gg b$; (b) a point on the x axis when $x \gg b$.

7. The familiar statement of Ohm's law is $I = V/R$, where the direction of current flow is in the direction of the voltage drop. Show that for an *elemental volume* this law may be written as the vector point relation $\mathbf{i} = \sigma\mathbf{E}$. (Recall that the resistance R of a conductor of length l and cross-sectional area A is given by $R = l/\sigma A$ where σ is the conductivity of the material.)

BIBLIOGRAPHY

See bibliography for chap. 2.

CHAPTER 4

MAXWELL'S EQUATIONS

Up to the present the fields considered have been the static electric field due to charges at rest and the static magnetic field due to steady or unchanging currents. The next step is to determine what modifications will be required when the charge densities and currents are changing with time. Before doing so it is desirable to restate Ampere's work law in the vector form as a point relation.

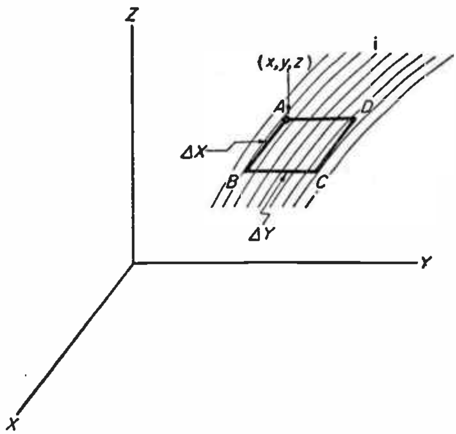


FIG. 4-1

4.01 Ampere's Work Law in the Differential Vector Form. Ampere's work law states that the magnetomotive force around a closed path is equal to the current enclosed by the path. That is

$$\oint \mathbf{H} \cdot d\mathbf{s} = I \qquad \text{amp} \qquad (4\text{-}1)$$

This law may be put into an alternative form as follows: Consider a conducting region in which there is a current density $\boldsymbol{i}$. Let $ABCD$

be an element of area parallel to the x-y plane, and let the co-ordinates of the point A be (x, y, z). The magnetomotive force around the closed path $ABCDA$ can be obtained by summing the magnetomotive forces along the four sides of the rectangle. If the average value of H_x over the path AB is represented by $\bar{H}_x$ and the average value of H_y over the path AD is represented by $\bar{H}_y$, then the following relations will hold:

$$\text{mmf from } A \text{ to } B = \bar{H}_x \,\Delta x$$
$$\text{mmf from } B \text{ to } C = \left(\bar{H}_y + \frac{\partial \bar{H}_y}{\partial x}\,\Delta x\right)\Delta y$$
$$\text{mmf from } C \text{ to } D = -\left(\bar{H}_x + \frac{\partial \bar{H}_x}{\partial y}\,\Delta y\right)\Delta x$$
$$\text{mmf from } D \text{ to } A = -\bar{H}_y \,\Delta y$$

Adding on both sides,

$$\text{mmf around closed path} = \left(\frac{\partial \bar{H}_y}{\partial x} - \frac{\partial \bar{H}_x}{\partial y}\right)\Delta x\,\Delta y$$

The current flowing through this rectangle is

$$dI = i_z\,\Delta x\,\Delta y$$

Therefore by Ampere's law,

$$\left(\frac{\partial \bar{H}_y}{\partial x} - \frac{\partial \bar{H}_x}{\partial y}\right)\Delta x\,\Delta y = i_z\,\Delta x\,\Delta y$$

As Δx and Δy are allowed to approach zero, $\bar{H}_x$ becomes H_x, and $\bar{H}_y$ becomes H_y, so that in the limit

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = i_z \tag{4-2a}$$

Next if the element of area is taken parallel to the y-z plane, and then parallel to the z-x plane, the following relations are obtained:

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = i_x \tag{4-2b}$$
$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = i_y \tag{4-2c}$$

The three scalar eqs. (2a), (2b), and (2c) can be combined into the single vector equation

$$\text{curl } \mathbf{H} = \boldsymbol{i} \tag{4-2}$$

This is an alternative statement (in the differential vector form) of Ampere's law. Equations (1) and (2) are stated correctly for a right-hand set of co-ordinate axes, and it is seen that the right-hand rule for determining the direction of H is included in both of these statements. The differential forms (2) of the equation require a *homogeneous* medium because the space derivative has no meaning at a discontinuity of the medium. When the path under consideration crosses a discontinuity the integral form (1) is suitable.

Interpretation of Curl **H.** Equation (2) relates the curl of the magnetic intensity to the current density that exists at any point in a region. A study of this relation is helpful in obtaining a physical picture of the curl of a vector. The picture can be made clearer if eq. (2) is integrated over an area to give

$$\int_S \text{curl } \mathbf{H} \cdot d\mathbf{a} = \int_S \boldsymbol{i} \cdot d\mathbf{a} \tag{4-3}$$

The right-hand side of (3), being the current density integrated over a surface S, is just the total current I flowing through the surface. Recalling the original form of the statement of Ampere's law in eq. (1) shows that the following relation must be true:

$$\int_S \text{curl } \mathbf{H} \cdot d\mathbf{a} = \oint \mathbf{H} \cdot d\mathbf{s} \tag{4-4}$$

This relates the integral of curl **H** over a surface to the line integral of **H**, or magnetomotive force, around the closed path bounding the surface. If the surface is reduced to an element of area $d\mathbf{a}$, the left-hand side becomes curl $\mathbf{H} \cdot d\mathbf{a}$. Dividing through by da, the result is $|\text{curl } \mathbf{H}| = (\oint \mathbf{H} \cdot d\mathbf{s})/da$, which may be interpreted as: "curl **H** equals the magnetomotive force per unit area." The *direction* of curl **H** is that direction of the area $d\mathbf{a}$ that results in a maximum magnetomotive force around its edge.

Stoke's theorem. The relation (4) obtained above for a magnetic field **H** is in fact a perfectly general relation true for any vector. That is, for any vector **A**,

$$\int_S \text{curl } \mathbf{A} \cdot d\mathbf{a} = \oint \mathbf{A} \cdot d\mathbf{s}$$

This equation is known as *Stokes' theorem.* It provides a very useful relation between an integration over a surface and an integration around the closed path bounding the surface. As was seen above it also provides a definition in the integral form for the curl of a vector.

4.02 Error in Simple Statement of Ampere's Law for Time-Varying Fields. It was proved in chap. 1 that for any vector A

$$\text{div curl } A = 0$$

Applying this to eq. (2) above necessitates that

$$\text{div } \boldsymbol{i} = 0 \tag{4-5}$$

Equation (5) states that there is no net outward flow of current from an elemental volume; that is, the current has no sources or sinks in the sense that it does not start or stop anywhere in a circuit. In other words, there must be a continuous flow of current throughout the entire circuit.

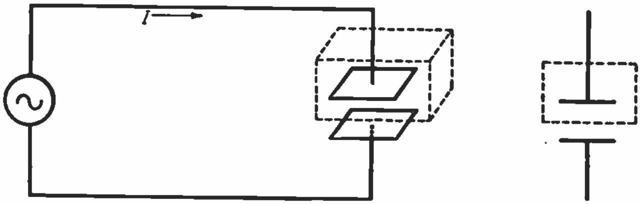


Fig. 4-2

This is true in the steady or direct-current case, and indeed, this is just a statement of Kirchoff's law for currents. However, eq. (5) is not necessarily true if the circuit contains condensers and the current is varying with time. Observe, for example the simple situation of Fig. 4-2.

In this case a voltage V, which is changing with time, will cause charges to flow onto the plates of the condenser. However, no charge will move across the region between the two plates. Hence current must start and stop on the condenser plates and i must have a divergence there. It will be necessary to modify eq. (5) to take care of this.

Equation of Continuity. Consider the diagram of Fig. 4-3, where an element of volume may have a different movement of charge through one face as compared with another. Let ρ be the charge density as a function of the co-ordinates.

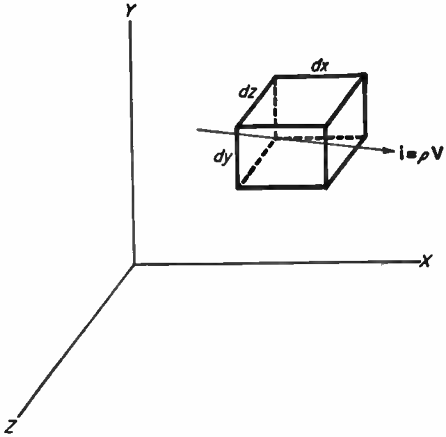


FIG. 4-3

The current density is $\rho\mathbf{v}$, where $\mathbf{v}$ is the velocity with which charges are moving.

Then in the elemental cube

$$\text{The current flowing in the left face} = \rho v_x \, dy \, dz$$

$$\text{That flowing out the right face} = \left(\rho v_x + \frac{\partial(\rho v_x)}{\partial x} dx\right) dy \, dz$$

The increase in charge within the volume per unit time due to movement in the x direction is therefore $-\dfrac{\partial(\rho v_x)}{\partial x} dx \, dy \, dz$.

The total increase due to movement in and out of all faces of the cube will be

$$\text{Total increase in charge per unit time} = \left[-\frac{\partial(\rho v_x)}{\partial x} - \frac{\partial(\rho v_y)}{\partial y} - \frac{\partial(\rho v_z)}{\partial z}\right] dx \, dy \, dz$$

$$\text{Increase in charge per unit time per unit volume} = -\left[\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right]$$

$$= -\operatorname{div} \rho\mathbf{v} = -\operatorname{div} \boldsymbol{i}$$

$$\text{But increase in charge per unit volume per unit time} = \frac{\partial \rho}{\partial t}$$

Now, since increase in charge per unit time per unit volume equals increase in charge per unit volume per unit time

$$\frac{\partial \rho}{\partial t} = - \operatorname{div} \boldsymbol{i} \tag{4-6}$$

Equation (6) is called the equation of continuity.

Although the equation of continuity has been developed here in connection with the flow of charges, it has quite general application in many fields, being the fundamental law of fluid motion. Under such circumstances ρ stands for the density of the fluid and the equation of continuity then states that the rate at which the quantity of fluid in a unit volume is increasing is equal to the rate at which the fluid is flowing *into* the volume from outside.

4.03 The Generalized Magnetomotive Force Equation. Ampere's work law stated in the vector point-relation form is

$$\operatorname{curl} \mathbf{H} = \boldsymbol{i} \tag{4-2}$$

It has been seen that taking the divergence of both sides of this equation leads to the conclusion that

$$\operatorname{div} \boldsymbol{i} = 0$$

or that current must be continuous. Since this is evidently not true for the alternating current case shown in Fig. 4-2, where current flows (momentarily) into the dotted rectangular enclosure without any corresponding outward flow, it follows that Ampere's law (eq. 2) must be in error for this case. An application of the equation of continuity has shown that a correct statement regarding the divergence of current density would be

$$\operatorname{div} \boldsymbol{i} = - \frac{\partial \rho}{\partial t} \tag{4-6}$$

Using this relation it is easy to arrive at a more general statement of Ampere's law which will be true in all cases. Recall Gauss's law as a vector point relation

$$\operatorname{div} \mathbf{D} = \rho$$

Take the time derivative of both sides

$$\frac{\partial}{\partial t} \operatorname{div} \mathbf{D} = \frac{\partial \rho}{\partial t} \tag{4-7}$$

Because space and time are independent variables, the order of differentiation may be reversed, that is,

$$\frac{\partial}{\partial t} \operatorname{div} \mathbf{D} = \operatorname{div} \frac{\partial \mathbf{D}}{\partial t}$$

so that eq. (7) becomes

$$\operatorname{div} \frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \rho}{\partial t}$$

Then using eq. (6)

$$\operatorname{div} \frac{\partial \mathbf{D}}{\partial t} = -\operatorname{div} i$$

or

$$\operatorname{div} \left(i + \frac{\partial \mathbf{D}}{\partial t} \right) = 0 \tag{4-8}$$

Equation (8), rather than eq. (5), is the correct statement when time changing (alternating) currents and fields are considered. In the direct-current case, where there is no change with time, eq. (8) reduces to (5).

It is evident that the term $\partial \mathbf{D}/\partial t$ has the dimensions of a current density. Now if $\partial \mathbf{D}/\partial t$ is considered as being a kind of current density then it would be possible to write (8) as

$$\operatorname{div} i_T = 0$$

where

$$i_T = \left(i + \frac{\partial \mathbf{D}}{\partial t} \right)$$

is the *total* current density. Under these circumstances it would be true that the (total) current is continuous and Ampere's law would hold even for the alternating current case where there are condensers in the circuit. Maxwell first observed the error in the original statement of Ampere's law (eq. 2) and modified the statement by replacing the conduction current density i by the total current density $[i + (\partial \mathbf{D}/\partial t)]$. The term $\partial \mathbf{D}/\partial t$ is called the *displacement current density.* The generalized statement of Ampere's law becomes

$$\operatorname{curl} \mathbf{H} = \left(i + \frac{\partial \mathbf{D}}{\partial t} \right) \tag{4-9}$$

Equation (9) is called the first of *Maxwell's equations.*

Maxwell's assumption that a changing displacement density (that is, a changing electric field) was equivalent to an electric current density, and as such would produce a magnetic field, has had most far-reaching effects. Combined with Faraday's law which indicates that a changing magnetic field will produce an electric field it leads directly to the "wave equations." This result enabled Maxwell to predict electromagnetic wave propagation some thirty years before Hertz's brilliant researches gave experimental verification. It should be observed that the assumption was made as a result of recognition of an error that was pointed up by the mathematics. This is an interesting example of one of those rather rare cases where the mathematical reasoning has preceded and pointed the way for experiment.

4.04 Faraday's Law and Maxwell's Second Equation. Faraday's induction law is analogous to Ampere's law. It states that the electromotive force or voltage around a closed path is equal to the negative of the time rate of change of magnetic flux enclosed by the path. That is

$$\oint \mathbf{E} \cdot ds = -\frac{\partial \Phi}{\partial t}$$

But

$$\Phi = \int_S \mathbf{B} \cdot da$$

where the integration of the magnetic-flux density is over a surface bounded by the closed path.

Then*

$$\oint \mathbf{E} \cdot ds = -\frac{\partial}{\partial t}\int_S \mathbf{B} \cdot da \cdot = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot da$$

But, by Stoke's theorem,

$$\oint \mathbf{E} \cdot ds = \int_S \text{curl } \mathbf{E} \cdot da$$

Therefore

$$\int_S \text{curl } \mathbf{E} \cdot da = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot da \tag{4-10}$$

If the surface S is now reduced to an elemental surface, eq. (10) becomes the point relation

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4-11}$$

* The partial derivative with time is used throughout to indicate that only variations of magnetic flux with time through a *fixed* closed path or at a fixed region in space are being considered. For a discussion of induced emf under other conditions refer to any text on electricity and magnetism. A thorough treatment is given in E. G. Cullwick, *The Fundamentals of Electromagnetism*, The Macmillan Co., Cambridge, England, 1939.

This is known as the *second Maxwell equation.*

Again recalling that the divergence of the curl of any vector is zero, it is evident from eq. (11) that div $\partial\mathbf{B}/\partial t = 0$, and therefore (for time varying fields)

$$\text{div } \mathbf{B} = 0 \tag{4-12}$$

Equation (12) states that there are no sources* of **B** and that lines of magnetic flux are *continuous.* This is in agreement with the assumption that there are no isolated magnetic poles and, consequently, no (physical) magnetic conduction current.

An interpretation of the curl of **E** is obtained from eq. (11) and the integral definition of curl. From the integral definition of curl the left-hand side of (11) is the line integral of **E** per unit area, and (11) states that this is equal to the negative time rate of change of magnetic flux per unit area. Now the line integral of **E** around any path is simply the voltage around the path, and so (11) is just Faraday's law stated for the closed path about an element of area. The voltage around the small closed path could be measured by a loop of wire connected to a voltmeter. The voltmeter reading divided by the area of the loop is a direct measure of the curl of **E**. As the loop is oriented in various directions, the direction of the axis of the loop that results in maximum voltage around the loop is the direction of the curl of **E**. In a region in which there is no time-changing magnetic flux, the voltage around the loop would be zero, and curl $\mathbf{E} = 0$. The electric field is then said to *have no curl,* or to be *irrotational* in that region. Evidently in electrostatics the electric field is always irrotational or without curl.

4.05 The Field Equations in Vector Form. The two Maxwell equations together with the expressions relating **D** and **B** to their sources are generally known as the *electromagnetic field equations* or just the *field equations.* In the differential vector form the field equations are:

$$\text{curl } \mathbf{H} = \dot{\mathbf{D}} + i \qquad \text{I}$$
$$\text{curl } \mathbf{E} = -\dot{\mathbf{B}} \qquad \text{II}$$
$$\text{div } \mathbf{D} = \rho \qquad \text{III}$$
$$\text{div } \mathbf{B} = 0 \qquad \text{IV}$$

* As used here *source* is a mathematical term. In the vector analysis of fluid fields a *source* is a point at which fluid is emitted or introduced into a region, and a *sink* is a point at which the fluid is absorbed or removed. The term source also has a broader use as the "cause of a phenomenon." In this latter sense, the sources of magnetic fields are electric currents.

The dot over a quantity indicates the time derivative of that quantity. These relations will be referred to so often that they have been labeled with Roman numerals and will be indicated in that manner throughout the remainder of this text.

The Field Equations in Differential Scalar Form. The field equations I–IV appear above in the abbreviated vector form. Written in the expanded scalar form in rectangular co-ordinates they are

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \frac{\partial D_x}{\partial t} + i_x \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \frac{\partial D_y}{\partial t} + i_y \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \frac{\partial D_z}{\partial t} + i_z\end{aligned}\right\} \qquad \text{I}$$

$$\left.\begin{aligned}\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t}\end{aligned}\right\} \qquad \text{II}$$

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \qquad \text{III}$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \qquad \text{IV}$$

The Field Equations in Integral Form. The field equations are often written in the integral form. The differential vector or scalar forms above are more convenient in the actual solution of problems, but the integral form is easier to interpret and to state in words. In the integral form

$$\oint \mathbf{H} \cdot ds = \int_s (\dot{\mathbf{D}} + \mathbf{i}) \cdot da \qquad \text{I}$$

$$\oint \mathbf{E} \cdot ds = -\int_s \dot{\mathbf{B}} \cdot da \qquad \text{II}$$

$$\int_{\substack{\text{closed}\\\text{surface}}} \mathbf{D} \cdot da = \int_{\text{vol}} \rho \, dV \qquad \text{III}$$

$$\int_{\substack{\text{closed}\\\text{surface}}} \mathbf{B} \cdot da = 0 \qquad \text{IV}$$

Word Statement of the Field Equations. A word statement of the significance of the field equations is readily obtained from their mathematical statement in the integral form. It would be somewhat as follows:

I. The magnetomotive force around a closed path is equal to the conduction current plus the time derivative of the electric displacement through any surface bounded by the path.

II. The electromotive force around a closed path is equal to the time derivative of the magnetic displacement through any surface bounded by the path.

III. The total electric displacement through the surface enclosing a volume is equal to the total charge within the volume.

IV. The net magnetic flux emerging through any *closed* surface is zero.

As indicated previously the time derivative of electric displacement is called displacement current. The term electric current is then generalized in meaning to include both conduction currents and displacement currents.* Furthermore, if the time derivative of electric displacement is called an electric current, the time derivative of magnetic displacement can be considered as being a magnetic current. Finally, electromotive force is called electric voltage, so that magnetomotive force may be called magnetic voltage.

The first two Maxwell equations can then be stated:

I. *The magnetic voltage around a closed path is equal to the electric current through the path.*

II. *The electric voltage around a closed path is equal to the magnetic current through the path.*

4.06 Conditions at a Boundary Surface. Maxwell's equations in the differential vector or scalar form express the relationship that must exist between the four field vectors **E, D, H,** and **B** at any point within a continuous medium. In this form, because they involve space derivatives, they cannot be expected to yield information at points of discontinuity in the medium. However, the integral form of statement can always be used to determine what happens at the boundary surface between different media.

* Also *convection* currents (e.g., electron beam currents). Conduction currents obey Ohm's law, $i = \sigma \mathbf{E}$; convection currents do not.

The following statements can be made regarding the electric and magnetic fields at any surface of discontinuity:

(a) The *tangential* component of **E** is continuous at the surface. That is, it is the same *just* outside the surface as it is *just* inside the surface.

(b) The *tangential* component of **H** is continuous across a surface except at the surface of a *perfect conductor*. At the surface

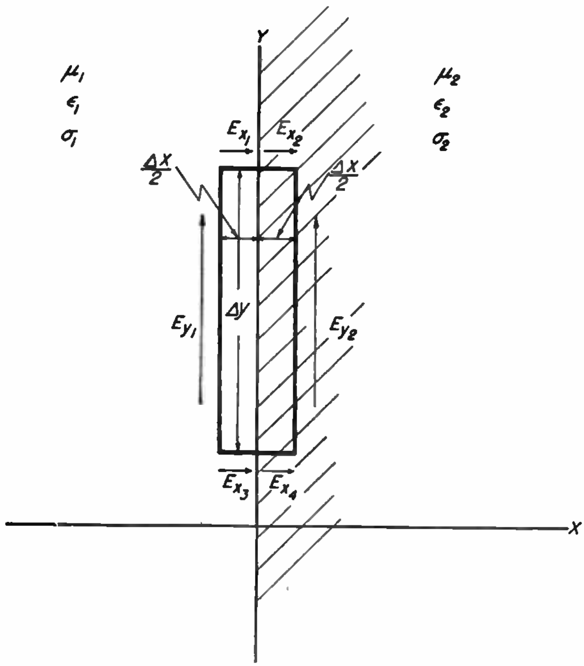


FIG. 4-4. A boundary surface between two media.

of a perfect conductor the *tangential* component of **H** is discontinuous by an amount equal to the surface current per unit width.

(c) The *normal* component of **B** is continuous at the surface of discontinuity.

(d) The *normal* component of **D** is continuous if there is no surface charge density. Otherwise **D** is discontinuous by an amount equal to the surface charge density.

The proof of these boundary conditions is obtained by a direct application of Maxwell's equations at the boundary between the media. Suppose the surface of discontinuity to be parallel to the y-z plane. Consider the small rectangle of width Δx and length Δy enclosing a small portion of each of media (1) and (2).

The integral form of the second Maxwell equation (II) is

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\int_S \dot{\mathbf{B}} \cdot d\mathbf{a}$$

For the elemental rectangle of Fig. 4-4 this becomes

$$E_{y_2}\,\Delta y - E_{x_2}\frac{\Delta x}{2} - E_{x_1}\frac{\Delta x}{2} - E_{y_1}\,\Delta y + E_{x_3}\frac{\Delta x}{2} + E_{x_4}\frac{\Delta x}{2} = -\dot{\bar{B}}_z\,\Delta x\,\Delta y \quad (4\text{-}13)$$

where $\bar{B}_z$ is the average magnetic-flux density through the rectangle $\Delta x\,\Delta y$. Now consider conditions as the area of the rectangle is made to approach zero by reducing the width Δx of the rectangle, always keeping the surface of discontinuity between the sides of the rectangle. If it is assumed that B is always finite, then the right-hand side of eq. (13) will approach zero. If E is also assumed to be everywhere finite, then the $\Delta x/2$ terms of the left-hand side will reduce to zero, leaving

$$E_{y_2}\,\Delta y - E_{y_1}\,\Delta y = 0$$

for $\Delta x = 0$. Therefore

$$E_{y_1} = E_{y_2}$$

That is, *the tangential component of E is continuous.*

Similarly the integral statement of eq. I is

$$\oint \mathbf{H} \cdot d\mathbf{s} = \int_S (\dot{\mathbf{D}} + \mathbf{i}) \cdot d\mathbf{a}$$

which becomes

$$H_{y_2}\,\Delta y - H_{x_2}\frac{\Delta x}{2} - H_{x_1}\frac{\Delta x}{2} - H_{y_1}\,\Delta y + H_{x_3}\frac{\Delta x}{2} + H_{x_4}\frac{\Delta x}{2} = (\dot{\bar{D}}_z + \bar{\imath}_z)\,\Delta x\,\Delta y \quad (4\text{-}14)$$

If the rate of change of electric displacement $\dot{D}$ and current density $\imath$ are both considered to be finite, then as before (14) reduces to

$$H_{y_2}\,\Delta y - H_{y_1}\,\Delta y = 0$$

or

$$H_{y_1} = H_{y_2}$$

The *tangential component of H is continuous* (for *finite* current densities; that is, for any *actual* case).

Note on a Perfect Conductor. A *perfect* conductor is one which has infinite conductivity. In such a conductor the electric intensity **E** is zero for any finite current density. All actual conductors have a finite value for conductivity. However, the actual conductivity may be very large and for many practical applications it is useful to assume it to be infinite. Such an assumption will lead to difficulties (because of indeterminacy) in formulating the boundary conditions unless care is taken in setting them up. As will be shown later, the depth of penetration into a conductor of an alternating electric field and of the current produced by the field decreases as the conductivity increases. Thus in a good conductor a high-frequency current will flow in a thin sheet near the surface, the depth of this sheet approaching zero as the conductivity approaches infinity. This gives rise to the useful concept of a *current sheet.* In a current sheet a finite current per unit width, J amperes per meter, flows in a sheet of vanishingly small depth Δx, but with the required infinitely large current density $\bar{\imath}$, such that

$$\lim_{\Delta x \to 0} \bar{\imath}\, \Delta x = J \qquad \text{amp/m}$$

Consider again the above example of the magnetomotive force around the small rectangle. If the current density $\bar{\imath}_z$ becomes infinite as Δx approaches zero, the right-hand side of eq. (14) will not become zero. Let J amperes per meter be the actual current per unit width flowing along the surface. Then as $\Delta x \to 0$ the eq. (14) for H becomes

$$H_{y_2}\, \Delta y - H_{y_1}\, \Delta y = J_z\, \Delta y$$

Hence

$$H_{y_1} = H_{y_2} - J_z \tag{4-15}$$

(Note that $\mathbf{D} = \epsilon\mathbf{E}$ remains finite and therefore $\bar{D}_z\, \Delta x$ is zero for $\Delta x = 0$.)

Now, if the electric field is zero within a perfect conductor, the magnetic field must also be zero (for *alternating* fields) as the second Maxwell equation II shows. Then in eq. (15), H_{y_2} must be zero and so

$$H_{y_1} = -J_z \tag{4-16}$$

Equation (16) states that the current per unit width along the surface of a perfect conductor is equal to the magnetic intensity

H just outside the surface. The magnetic field and surface current will be parallel to the surface, but perpendicular to each other. In vector notation this is written

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}$$

where $\mathbf{n}$ is the unit vector along the outward normal to the surface.

Conditions on the Normal Components of **B** *and* **D**. The remaining boundary conditions are concerned with the *normal* components of **B** and **D**. The integral form of the third field equation is

$$\oint_S \mathbf{D} \cdot d\mathbf{a} = \int_{\text{vol}} \rho \, dV \qquad \text{III}$$

When applied to the elementary "pill-box" volume of Fig. 4-5, eq. III becomes

$$D_{n_1} \, da - D_{n_2} \, da + \Psi_{\text{edge}} = \bar{\rho} \, \Delta x \, da \qquad (4\text{-}17)$$

In this expression da is the area of each of the flat surfaces of the pillbox, Δx is their separation, and $\bar{\rho}$ is the average charge density

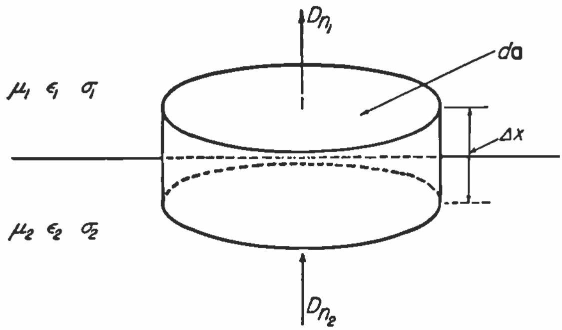


FIG. 4-5. A "pill-box" volume encloses a portion of a boundary surface.

within the volume $\Delta x \, da$. Ψ_{edge} is the outward electric flux through the curved-edge surface of the pillbox. As $\Delta x \to 0$, that is, as the flat surfaces of the box are squeezed together, always keeping the boundary surface between them, $\Psi_{\text{edge}} \to 0$, for finite values of displacement density. Also for *finite* values of average charge density $\bar{\rho}$, the right-hand side of (17) approaches zero, and (17) reduces to

$$D_{n_1} \, da - D_{n_2} \, da = 0$$

(for $\Delta x = 0$). Then for the case of no surface charge the condition on the normal components of **D** is

$$D_{n_1} = D_{n_2} \tag{4-18}$$

That is, if there is no surface charge the normal component of **D** is continuous across the surface.

In the case of a metallic surface, the charge is considered to reside "on the surface." If this layer of surface charge has a *surface charge density* ρ_s coulombs per *square* meter, the charge density ρ of the surface layer is given by

$$\rho = \frac{\rho_s}{\Delta x} \qquad \text{coulomb/cu m}$$

where Δx is thickness of the surface layer. As Δx approaches zero, the charge density approaches infinity in such a manner that

$$\lim_{\Delta x \to 0} \rho\,\Delta x = \rho_s$$

Then in Fig. 4-5, if the surface charge is always kept between the two flat surfaces as the separation between them is decreased, the right-hand side of eq. (17) approaches $\rho_s\,da$ as Δx approaches zero. Equation (17) then reduces to

$$D_{n_1} - D_{n_2} = \rho_s \tag{4-19}$$

When there is a surface charge density ρ_s, the normal component of displacement density is *discontinuous* across the surface by the amount of the surface charge density.

For any metallic conductor the displacement density $\mathbf{D} = \epsilon\mathbf{E}$ within the conductor will be a very small quantity (it will be zero in the electrostatic case, or in the case of a perfect conductor). Then if medium 2 is a metallic conductor $D_{n_2} = 0$ and eq. (19) becomes

$$D_{n_1} = \rho_s \tag{4-20}$$

The normal component of displacement density in the dielectric is equal to the surface charge density on the conductor.

In the case of magnetic-flux density **B**, since there are no isolated "magnetic charges," a similar analysis leads at once to

$$B_{n_1} = B_{n_2}$$

The normal component of magnetic flux density is always continuous across a boundary surface.

PROBLEMS

1. Show that the displacement current through the condenser is equal to the conduction current I (Fig. 4-6).

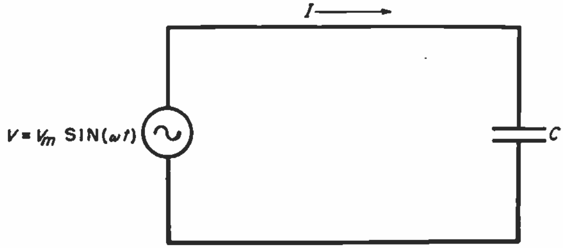


Fig. 4-6

2. *Within* a perfect conductor **E** is always zero. Using Maxwell's equations, show that **H** must also be zero for time varying fields. Can a steady (unchanging) magnetic field exist within a perfect conductor? Show that the normal component of **B** (and therefore **H**) must be zero at the surface of a perfect conductor.

3. A "transmission line" consists of two parallel perfectly conducting planes of large extent, separated by a distance d meters. The conducting

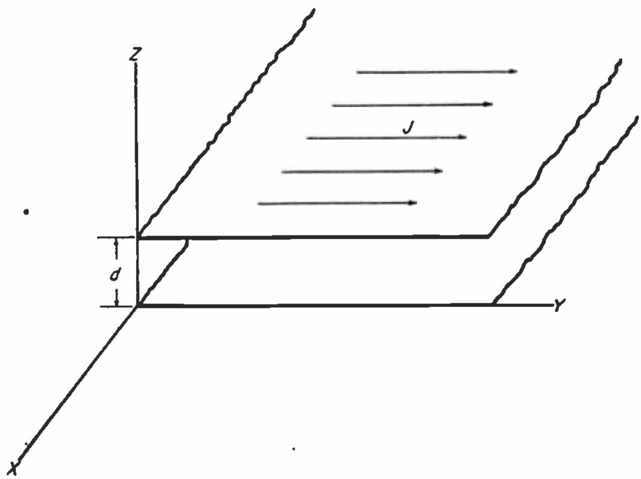


Fig. 4-7. Parallel-plane "transmission" line.

planes carry an alternating linear current density J amp/m in the y direction, that is,

$$J = J_0 e^{j\omega[t-(y/c)]}$$

or

$$J = J_0 \cos \omega \left(t - \frac{y}{c}\right)$$

Applying Maxwell's first equation in the region between the conductors find the electric intensity, and hence the voltage between the planes, when $d = 1$ meter and the effective linear current density is $J_{\text{eff}} = 1$ amp/m.

4. A square loop of wire, 20 cm by 20 cm, has a voltmeter (of infinite impedance) connected in series with one side. Determine the voltage indicated by the meter when the loop is placed in an alternating magnetic field, the maximum intensity of which is 1 ampere per meter. The plane of the loop is perpendicular to the magnetic field; the frequency is 10 mc.

5. A No. 10 copper wire carries a conduction current of 1 amp at 60 cps. What is the displacement current in the wire? For copper assume $\epsilon = \epsilon_v$, $\mu = \mu_v$, $\sigma = 5.8 \times 10^7$.

6. The electric vector **E** of an electromagnetic wave in free space is given by the expressions

$$E_x = E_z = 0 \qquad E_y = A\, e^{j\omega[t-(x/v)]}$$

Using Maxwell's equations for free space conditions determine expressions for the components of the magnetic vector **H**.

BIBLIOGRAPHY

Abraham, M., and R. Becker, *The Classical Theory of Electricity and Magnetism*, Blackie and Son, Ltd., London and Glasgow, 1937.

Skilling, H. H., *Fundamentals of Electric Waves*, John Wiley and Sons, New York, 1948.

Ramo, S., and J. R. Whinnery, *Fields and Waves in Modern Radio*, John Wiley and Sons, New York, 1944.

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand, New York, 1943.

Sarbacher, R. J., and W. A. Edson, *Hyper and Ultrahigh Frequency Engineering*, John Wiley and Sons, New York, 1943.

Stratton, J. A., *Electromagnetic Theory*, McGraw-Hill, New York, 1941.

CHAPTER 5

ELECTROMAGNETIC WAVES

PART I—ELECTROMAGNETIC WAVES IN A HOMOGENEOUS MEDIUM

In the solution of any electromagnetic problem the fundamental relations that must be satisfied are the four field equations

$$\text{curl } \mathbf{H} = \dot{\mathbf{D}} + i \qquad \text{I}$$
$$\text{curl } \mathbf{E} = -\dot{\mathbf{B}} \qquad \text{II}$$
$$\text{div } \mathbf{D} = \rho \qquad \text{III}$$
$$\text{div } \mathbf{B} = 0 \qquad \text{IV}$$

In addition there are three relations that concern the characteristics of the medium in which the fields exist. These are

$$\mathbf{D} = \epsilon \mathbf{E} \qquad (5\text{-}1)$$
$$\mathbf{B} = \mu \mathbf{H} \qquad (5\text{-}2)$$
$$i = \sigma \mathbf{E} \qquad (5\text{-}3)$$

where ϵ, μ, and σ are the permittivity, permeability, and conductivity of the medium, which is assumed to be homogeneous, isotropic, and sourcefree. A *homogeneous* medium is one for which the quantities ϵ, μ, and σ are constant throughout the medium. The medium is *isotropic* if ϵ is a scalar constant, so that **D** and **E** have everywhere the same direction. The form of Maxwell's equations, given by I and II, is for *sourcefree* regions, that is, regions in which there are no *impressed* voltages or currents (no generators). The relations of the fields to their sources will be considered in chap. 10 and subsequent chapters.

When the relations (1), (2), and (3) are inserted in I and II, Maxwell's equations become differential equations relating the electric and magnetic intensities **E** and **H**. If they are then solved as simultaneous equations, they will determine the laws which both **E** and **H** must obey.

5.01 Solution for Free-space Conditions. Before obtaining the solution for the general case it is instructive to consider the simple, but important, particular case of electromagnetic phenomena in free space—that is in a perfect dielectric containing no charges and no conduction currents. For this case the field equations become

$$\text{curl } \mathbf{H} = \dot{\mathbf{D}} \tag{5-4}$$
$$\text{curl } \mathbf{E} = -\dot{\mathbf{B}} \tag{5-5}$$
$$\text{div } \mathbf{D} = 0 \tag{5-6}$$
$$\text{div } \mathbf{B} = 0 \tag{5-7}$$

Differentiate (4) with respect to time. Since the curl operation is a differentiation with respect to space, the order of differentiation may be reversed, that is,

$$\frac{\partial \text{ curl } \mathbf{H}}{\partial t} = \text{curl } \dot{\mathbf{H}}$$

Also since ϵ and μ are independent of time

$$\dot{\mathbf{D}} = \epsilon\dot{\mathbf{E}} \tag{5-8}$$
$$\dot{\mathbf{B}} = \mu\dot{\mathbf{H}} \tag{5-9}$$

so that there results

$$\text{curl } \dot{\mathbf{H}} = \epsilon\ddot{\mathbf{E}} \tag{5-10}$$

The symbol $\ddot{\mathbf{E}}$ means $\dfrac{\partial^2 \mathbf{E}}{\partial t^2}$.

Take the curl of both sides of (5) and using (9), obtain

$$\text{curl curl } \mathbf{E} = -\mu \text{ curl } \dot{\mathbf{H}} \tag{5-11}$$

Substitute eq. (10) into (11)

$$\text{curl curl } \mathbf{E} = -\mu\epsilon\ddot{\mathbf{E}} \tag{5-12}$$

It was shown in identity (1-28) that

$$\text{curl curl } \mathbf{E} = \text{grad div } \mathbf{E} - \nabla^2\mathbf{E}$$

Combine this equation with (12) to obtain

$$\text{grad div } \mathbf{E} - \nabla^2\mathbf{E} = -\mu\epsilon\ddot{\mathbf{E}} \tag{5-13}$$

but

$$\text{div } \mathbf{E} = \frac{1}{\epsilon} \text{ div } \mathbf{D} = 0$$

therefore eq. (13) becomes

$$\nabla^2 \mathbf{E} = \mu\epsilon\ddot{\mathbf{E}} \tag{5-14}$$

This is the law that **E** must obey.

Differentiating (5) with respect to time and taking the curl of (4) it will be found on combining that **H** obeys the same law, viz.

$$\nabla^2 \mathbf{H} = \mu\epsilon\ddot{\mathbf{H}} \tag{5-15}$$

Equations (14) and (15) are known as the *wave equations*. Thus the first condition on either **E** or **H** is that it must satisfy the wave equation. (Note that although **E** and **H** obey the same law, **E** is not equal to **H**.)

5.02 Uniform Plane Wave Propagation. The wave equation reduces to a very simple form in the special case where **E** and **H** are considered to be independent of two dimensions, say y and z. Then

$$\nabla^2 \mathbf{E} = \frac{\partial^2 \mathbf{E}}{\partial x^2}$$

so that (14) becomes

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{5-16}$$

Vector eq. (16) is equivalent to three scalar equations, one for each of the scalar components of **E**. In general, for uniform plane wave propagation in the x direction, **E** may have components E_y and E_z, but (as will be seen later) not E_x. Without loss of generality attention can be restricted to one of the components, say E_y, knowing that results for E_z will be similar to those obtained for E_y. Then the equation to be solved has the form

$$\frac{\partial^2 E_y}{\partial x^2} = \mu\epsilon \frac{\partial^2 E_y}{\partial t^2} \tag{5-16a}$$

Equation (16a) is a second-order partial differential equation, which occurs frequently in mechanics and engineering. For example it is the differential equation for the displacement from equilibrium along a uniform string. Electrical engineers will recognize it as the differential equation for voltage or current along a lossless transmission line. Its general solution is of the form

$$E = f_1(x - v_0 t) + f_2(x + v_0 t) \tag{5-17}$$

where $v_0 = 1/\sqrt{\mu\epsilon}$ and f_1 and f_2 are *any* functions (not necessarily the same) of $(x - v_0t)$ and $(x + v_0t)$ respectively. The expression $f(x - v_0t)$ means a function f of the variable $(x - v_0t)$. Examples are, $A \cos \beta(x - v_0t)$, $C\, e^{k(x-v_0t)}$, $\sqrt{x - v_0t}$, etc. All of these expressions represent wave motion.

A wave* may be defined in the following way: If a physical phenomenon that occurs at one place at a given time is reproduced

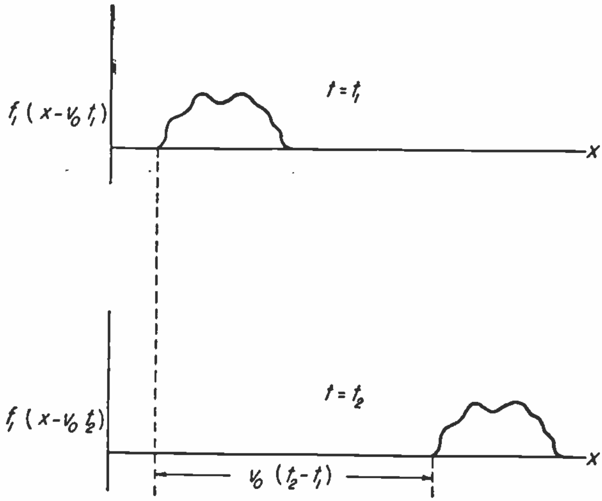


FIG. 5-1. A wave traveling in the positive x direction.

at other places at later times, the time delay being proportional to the space separation from the first location, then the group of phenomena constitute a wave. Note that a wave is not necessarily a repetitive phenomena in time. Those who survive a tidal wave are thankful for this.

The functions $f_1(x - v_0t)$ and $f_2(x + v_0t)$ describe such a wave mathematically, the variation of the wave being confined to one dimension in space. This is shown by Fig. 5-1.

If a fixed time is taken, say t_1, then the function $f_1(x - v_0t_1)$ becomes a function of x since v_0t_1 is a constant. Such a function is represented by the first curve. If another time, say t_2, is taken,

* The term *wave* also has an entirely different usage, viz.: a *recurrent function of time at a point*, as in the expression *sinusoidal voltage wave*. Usually there will be no doubt as to which kind of wave is meant.

another function of x is obtained, exactly the same shape as the first except that the second curve is displaced to the right by a distance $v_0(t_2 - t_1)$. This shows that the phenomenon has traveled in the positive x direction with a velocity v_0.

On the other hand, the function $f_2(x + v_0t)$ corresponds to a wave traveling in the negative x direction. Thus the general solution of the wave equation in this case is seen to consist of two waves, one traveling to the right (away from the source), and the other traveling to the left (back toward the source). If there is no reflecting surface present to reflect the wave back to the source, the second term of (17) is zero and the solution is given by

$$E = f_1(x - v_0t) \tag{5-18}$$

Problem 1. Does the function $e^{k(x-v_0t)}$ represent a wave if k is a *real* number? Sketch it as a function of x for several instants of time.

5.03 Sinusoidal Time Variations. In solving a one-dimensional wave equation, such as (16a), no restriction is put upon how **E** and **H** might vary with time, and the functions f_1 and f_2 of eq. (17) can be any functions of $(x - v_0t)$. In practice most generators produce voltages and currents, and hence electric and magnetic fields, which vary sinusoidally with time (at least approximately). Even where this is not the case any periodic variation can always be analysed in terms of sinusoidal variations with fundamental and harmonic frequencies, so it is customary in most problems to assume sinusoidal time variations. This can be expressed by writing, for example,

$$E = E_0 \cos \omega t \tag{5-19a}$$

or

$$E = E_0 \sin \omega t \tag{5-19b}$$

where $f = \omega/2\pi$ is the frequency of the variation. In electrical engineering it is more usual to express sinusoidal time variations in the exponential form

$$E = E_0\, e^{j\omega t} \tag{5-20}$$

where $E_0\, e^{j\omega t} = E_0(\cos \omega t + j \sin \omega t)$.

It is seen that the real part of (20) is equal to (19a), and the imaginary part is equal to (19b).* In working a problem, the exponential form (that is, both real and imaginary parts) is carried through to the end, but only the real part *or* the imaginary part of the final answer is used, the other part being discarded. Use of the

* This, of course, assumes that E_0 is real.

real part corresponds to starting with $E_0 \cos \omega t$, while use of the imaginary part corresponds to starting with (19b). The use of this exponential form will be treated more fully in the next chapter.*

It will be observed that if

$$E = E_0 e^{j\omega t}$$

then

$$\dot{E} = \frac{\partial E}{\partial t} = j\omega E_0 e^{j\omega t}$$

$$= j\omega E \qquad (5\text{-}21)$$

and

$$\ddot{E} = -\omega^2 E \qquad (5\text{-}22)$$

Also

$$\int E\, dt = \frac{E}{j\omega} \qquad (5\text{-}23)$$

Assuming that the variation with time of all fields and currents is represented by $e^{j\omega t}$, eqs. (4) and (5) can be written

$$\text{curl } \mathbf{H} = j\omega\epsilon\mathbf{E} \qquad (5\text{-}24)$$

$$\text{curl } \mathbf{E} = -j\omega\mu\mathbf{H} \qquad (5\text{-}25)$$

Differentiating (24) with respect to time and taking the curl of (25) and combining gives

$$\text{curl curl } \mathbf{E} = \omega^2\mu\epsilon\mathbf{E}$$

Making use of the identity (1-28) of chap. 1, and the fact that div $\mathbf{E} = 0$ for this case

$$\nabla^2\mathbf{E} = -\omega^2\mu\epsilon\mathbf{E} \qquad (5\text{-}26)$$

For the case of no variation of $\mathbf{E}$ with respect to y or z this results

* The correctness of the results obtained in carrying both real and imaginary parts through the problem, and using only the real (or the imaginary) part of the final solution, depends on the *linearity* of the equations. In *power* calculations the relations are no longer linear, and caution must be observed if correct results are to be obtained. For example, writing

$$V = V_0 e^{j\omega t} \qquad I = I_0 e^{j\omega t}$$

it is *not* correct to say, where V_0 and I_0 are real,

$$W = V_0 e^{j\omega t} I_0 e^{j\omega t} = V_0 I_0 e^{2j\omega t}$$

the *real* part of which would be

$$W = Re\,(V_0 I_0 e^{2j\omega t}) = V_0 I_0 \cos 2\omega t$$

This equation indicates that the *average* power over a complete cycle would be zero, an incorrect result. However, it can be shown (see chap. 6) that when V_0 and I_0 are expressed in the complex form the correct value for real power is given by

$$W\ (\text{real}) = \tfrac{1}{2}\, Re\,(V_0 I_0^*)$$

where I_0^* is the complex conjugate of I_0.

in an equation corresponding to (16)

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = -\omega^2 \mu \epsilon \mathbf{E} \tag{5-27}$$

Again considering only the E_y component, a solution may be written in the form

$$\begin{aligned} E_y &= E' \, e^{-j\omega\sqrt{\mu\epsilon}x} + E'' \, e^{+j\omega\sqrt{\mu\epsilon}x} \\ &= E' \, e^{-j\beta x} + E'' \, e^{+j\beta x} \end{aligned} \tag{5-28}$$

where

$$\beta = \omega \sqrt{\mu\epsilon}$$

Showing the time variation explicitly by writing

$$E' = E_0' \, e^{j\omega t} \qquad E'' = E_0'' \, e^{j\omega t}$$

eq. (28) becomes

$$E_y = E_0' \, e^{j(\omega t - \beta x)} + E_0'' \, e^{j(\omega t + \beta x)} \tag{5-29}$$

This equation represents the sum of two waves traveling in opposite directions. If only the real part of the expression is used, the solution has the form

$$E_y = E_0' \cos(\omega t - \beta x) + E_0'' \cos(\omega t + \beta x) \tag{5-30}$$

whereas, if only the imaginary part is used, there results

$$E_y = E_0' \sin(\omega t - \beta x) + E_0'' \sin(\omega t + \beta x) \tag{5-31}$$

Equation (30) or (31) is a special case of eq. (17), which is obtained when a sinusoidal time variation is assumed. It is seen that in a homogeneous lossless medium the assumption of sinusoidal time variations results in space variations that are also sinusoidal.

The wave represented by the first term of eq. (30) is sketched in Fig. 5-2 for successive instants of time.

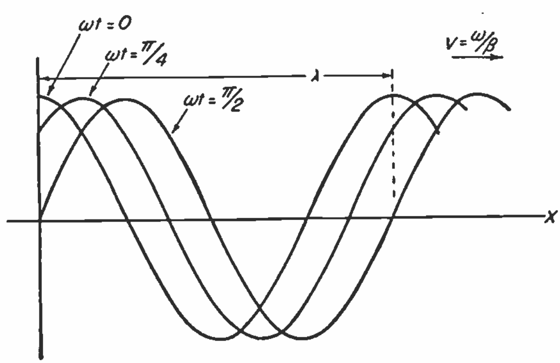


Fig. 5-2. A sinusoidal traveling wave.

It progresses in the positive x direction with a velocity ω/β. This becomes apparent by noting that a wave crest or maximum value of E_y occurs when

$$\omega t - \beta x = 0 \quad \text{(or any even multiple of } \pi) \tag{5-32}$$

In order to always remain with a crest, it is necessary to move in the positive x direction with a velocity

$$v = \frac{x}{t} = \frac{\omega}{\beta} \tag{5-33}$$

so that (32) is always satisfied. The wave, represented by the first term of eq. (30) and sketched in Fig. 5-2, is called a *traveling wave* (in this case it is an *unattenuated* traveling wave) to distinguish it from a *standing wave*, which does not progress.

The distance between adjacent crests or any two corresponding points on adjacent waves is the wavelength λ, and the frequency with which the crests appear at a given point is the frequency f. It is evident that the velocity with which the wave is propagating in the x direction is also given by

$$v = \lambda f \tag{5-34}$$

Combining (33) and (34)

$$\frac{\omega}{\beta} = \lambda f$$

showing that

$$\beta = \frac{2\pi}{\lambda} \tag{5-35}$$

β is the *phase shift constant* and is a measure of the phase shift (in radians) per unit length. Expression (35) is a statement of the fact that the phase shifts 2π radians, or one complete cycle, in a distance of one wavelength.

5.04 Uniform Plane Waves. Equation (18) is a solution of the wave equation for the particular case where the electric intensity is independent of y and z and is a function of x and t only. Such a wave is called a *uniform plane wave*. A *plane wave* is one for which the phase is the same for all points on a plane surface. If the amplitude is also constant over this plane surface, it is a *uniform* plane wave. Although this is a special case of electromagnetic wave propagation, it is a very important one practically and will be considered further.

The plane-wave equation

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

may be written in terms of the components of **E** as

$$\frac{\partial^2 E_x}{\partial x^2} = \mu\epsilon \frac{\partial^2 E_x}{\partial t^2} \tag{5-36a}$$

$$\frac{\partial^2 E_y}{\partial x^2} = \mu\epsilon \frac{\partial^2 E_y}{\partial t^2} \tag{5-36b}$$

$$\frac{\partial^2 E_z}{\partial x^2} = \mu\epsilon \frac{\partial^2 E_z}{\partial t^2} \tag{5-36c}$$

In a region in which there is no charge density

$$\text{div } \mathbf{E} = \frac{1}{\epsilon} \text{div } \mathbf{D} = 0$$

That is

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0$$

For a uniform plane wave in which **E** is independent of y and z, the last two terms of this relation are equal to zero so that it reduces to

$$\frac{\partial E_x}{\partial x} = 0$$

Therefore there is no variation of E_x in the x direction. From eq. (36a) it is seen that the second derivative with respect to time of E_x must then be zero. This requires that E_x be either zero, constant in time, or increasing uniformly with time. A field satisfying either of the last two of these conditions would not be a part of the wave motion, and so E_x can be put equal to zero. Therefore a uniform plane wave progressing in the x direction has no x component of **E**. A similar analysis would show that there is no x component of **H**. It follows, therefore, that uniform plane electromagnetic waves are transverse and have components of **E** and **H** only in directions perpendicular to the direction of propagation.

Relation between **E** *and* **H** *in a Uniform Plane Wave.* For a uniform plane wave traveling in the x direction **E** and **H** are both independent of y and z, and **E** and **H** have no x component. In this case

$$\text{curl } \mathbf{E} = -\frac{\partial E_z}{\partial x}\mathbf{j} + \frac{\partial E_y}{\partial x}\mathbf{k}$$

$$\text{curl } \mathbf{H} = -\frac{\partial H_z}{\partial x}\mathbf{j} + \frac{\partial H_y}{\partial x}\mathbf{k}$$

Then the first Maxwell Equation (I) can be written

$$-\frac{\partial H_z}{\partial x}\mathbf{j} + \frac{\partial H_y}{\partial x}\mathbf{k} = \epsilon\left(\frac{\partial E_y}{\partial t}\mathbf{j} + \frac{\partial E_z}{\partial t}\mathbf{k}\right)$$

and the second equation (II) becomes

$$-\frac{\partial E_z}{\partial x}\mathbf{j} + \frac{\partial E_y}{\partial x}\mathbf{k} = -\mu\left(\frac{\partial H_y}{\partial t}\mathbf{j} + \frac{\partial H_z}{\partial t}\mathbf{k}\right)$$

Equating **j** terms and then the **k** terms yields the four relations

$$-\frac{\partial H_z}{\partial x} = \epsilon\frac{\partial E_y}{\partial t} \tag{5-37a}$$

$$\frac{\partial H_y}{\partial x} = \epsilon\frac{\partial E_z}{\partial t} \tag{5-37b}$$

$$\frac{\partial E_z}{\partial x} = \mu\frac{\partial H_y}{\partial t} \tag{5-37c}$$

$$\frac{\partial E_y}{\partial x} = -\mu\frac{\partial H_z}{\partial t} \tag{5-37d}$$

Now if $E_y = f_1(x - v_0t)$, where $v_0 = 1/\sqrt{\mu\epsilon}$, then

$$\frac{\partial E_y}{\partial t} = \frac{\partial f_1}{\partial(x - v_0t)}\frac{\partial(x - v_0t)}{\partial t} = -v_0\frac{\partial f_1}{\partial(x - v_0t)}$$

This is generally written as

$$\frac{\partial E_y}{\partial t} = f_1'(x - v_0t)\frac{\partial(x - v_0t)}{\partial t} = -v_0f_1'(x - v_0t)$$

where $f_1'(x - v_0t)$ means

$$\frac{\partial f_1(x - v_0t)}{\partial(x - v_0t)}$$

Substituting for $\frac{\partial E_y}{\partial t}$ in (37a) above gives

$$\frac{\partial H_z}{\partial x} = v_0\epsilon f_1'(x - v_0t)$$

Then

$$H_z = \sqrt{\frac{\epsilon}{\mu}}\int f_1'(x - v_0t)\,dx + C$$

Now

$$\frac{\partial f_1(x - v_0t)}{\partial x} = f_1'(x - v_0t)\frac{\partial(x - v_0t)}{\partial x} = f_1'(x - v_0t)$$

Hence

$$H_z = \sqrt{\frac{\epsilon}{\mu}} \int \frac{\partial f_1(x - v_0 t)}{\partial x}\, dx + C$$
$$= \sqrt{\frac{\epsilon}{\mu}} f_1(x - v_0 t) + C$$
$$= \sqrt{\frac{\epsilon}{\mu}} E_y + C \qquad (5\text{-}38)$$

The constant of integration C that appears indicates that a field independent of x could be present. Inasmuch as this field would not be a part of the wave motion, it will be neglected and the relation between H_z and E_y becomes

$$H_z = \sqrt{\frac{\epsilon}{\mu}} E_y$$

or

$$\frac{E_y}{H_z} = \sqrt{\frac{\mu}{\epsilon}} \qquad (5\text{-}39)$$

Similarly it can be shown that

$$\frac{E_z}{H_y} = -\sqrt{\frac{\mu}{\epsilon}} \qquad (5\text{-}40)$$

Since

$$E = \sqrt{E_y^2 + E_z^2} \qquad \text{and} \qquad H = \sqrt{H_z^2 + H_y^2}$$

where E and H are the total electric and magnetic intensities, there also results

$$\frac{E}{H} = \sqrt{\frac{\mu}{\epsilon}} \qquad (5\text{-}41)$$

Equation (41) states that in a traveling* plane electromagnetic wave there is a definite ratio between the amplitudes of E and H and that this ratio is equal to the square root of the ratio of permeability to the dielectric constant of the medium. Since the units of E are volts per meter and the units of H are amperes per meter, the ratio

$$\frac{E}{H} = \sqrt{\frac{\mu}{\epsilon}}$$

* The term *traveling* wave is used to indicate that the wave is progressing in one direction and there is no *standing* wave (see section on reflection). When there is a reflected wave resulting in a standing-wave distribution, the ratio E/H can have any value between zero and infinity.

will have the dimensions of impedance or ohms. For this reason it is customary to refer to the ratio $\sqrt{\frac{\mu}{\epsilon}}$ as the *characteristic impedance* or *intrinsic impedance* of the (nonconducting) medium. For free space

$$\mu = \mu_v = 4\pi \times 10^{-7} \qquad \text{henrys/m}$$

$$\epsilon = \epsilon_v \approx \frac{1}{36\pi \times 10^9} \qquad \text{f/m}$$

so that

$$\sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_v}{\epsilon_v}} \approx 120\pi = 377 \text{ ohms}$$

For any medium, whether conducting or not, the intrinsic impedance is designated by the symbol η. When the medium is free space or a vacuum, the subscript v is used. That is, the intrinsic impedance of free space is

$$\eta_v = \sqrt{\frac{\mu_v}{\epsilon_v}} = 377 \text{ ohms}$$

Polarization. The plane wave just considered has no x component of electric field (that is, no component of **E** in the direction of propagation), but in general would have components E_y and E_z. If $E_z = 0$ and only E_y has value the wave is said to be *polarized in the y direction.* If $E_y = 0$ but E_z has value the wave would be polarized in the z direction. If both E_y and E_z components are present and are *in time phase,* the resultant electric field has a direction dependent on the relative magnitude of E_y and E_z. The angle which this direction makes with the y axis is $\tan^{-1} E_z/E_y$ and this angle will be constant with time. In all of the above cases in which the direction of the resultant vector is constant with time the wave is said to be *linearly polarized,* and the direction of polarization is just the direction of the electric vector.*

If the E_y and E_z components are not in time phase, that is, if at a given point they reach their maximum values at different instants of time, then the direction of the resultant electric vector will vary with time. In this case the locus of the end point of the resultant **E** will be an ellipse and the wave is said to be *elliptically polarized.*

* In optics the *plane of polarization* is taken as being that plane in the direction of propagation that contains the *magnetic vector H.*

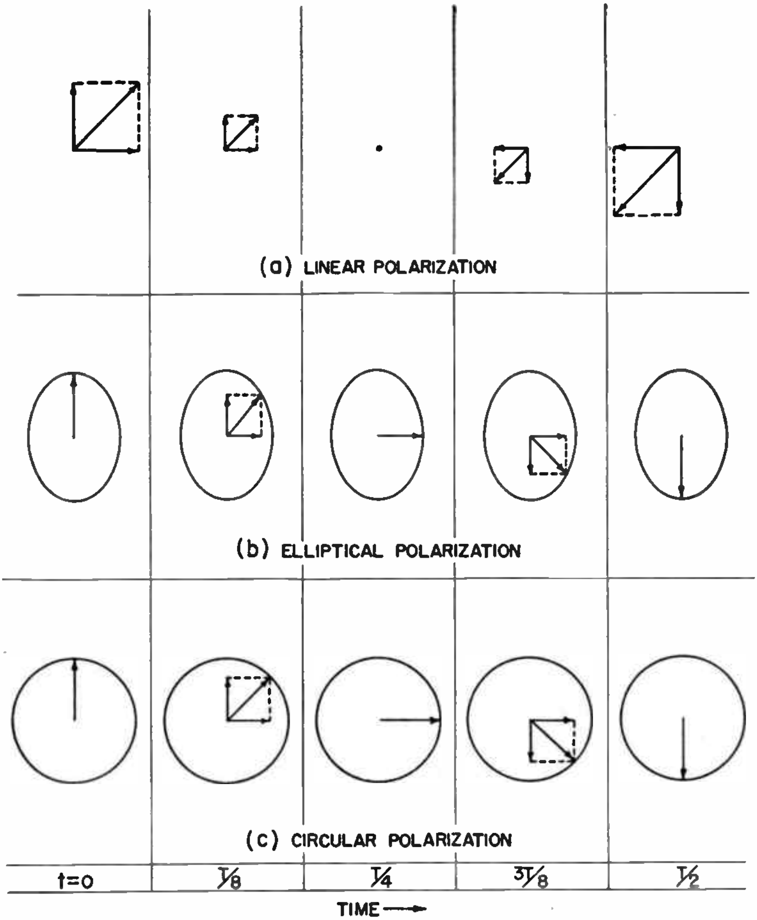


FIG. 5-3. Linear, elliptical, and circular polarization.

In the particular case where E_y and E_z have equal magnitudes and a 90° time phase difference, the locus of the resultant **E** is a circle and the wave is *circularly polarized.*

5.05 The Wave Equations for a Conducting Medium. In the foregoing sections Maxwell's equations were solved for the particular case of a perfect dielectric, such as free space, in which there were neither charges nor conduction currents. For regions in which the conductivity is not zero and conduction currents may exist, the more general solution must be obtained. It follows in a manner similar to the simpler case already considered.

Recall Maxwell's equations:

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}} + i \qquad \text{I}$$

$$\text{curl } \mathbf{E} = -\mu\dot{\mathbf{H}} \qquad \text{II}$$

If the medium has a conductivity σ (mhos/m), the conduction current density will be given by Ohm's law:*

$$i = \sigma\mathbf{E} \qquad (5\text{-}42)$$

so that eq. I becomes

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}} + \sigma\mathbf{E} \qquad (5\text{-}43)$$

Again assuming that all fields and currents vary with time as $e^{j\omega t}$ so that, for example,

$$\dot{\mathbf{E}} = j\omega\mathbf{E}$$

eq. (43) becomes

$$\text{curl } \mathbf{H} = (\sigma + j\omega\epsilon)\mathbf{E} \qquad (5\text{-}44)$$

Differentiating with respect to time gives

$$\text{curl } \dot{\mathbf{H}} = j\omega(\sigma + j\omega\epsilon)\mathbf{E} \qquad (5\text{-}45)$$

Take the curl of both sides of equation II and then substitute into it eq. (45)

$$\begin{aligned}\text{curl curl } \mathbf{E} &= -\mu \text{ curl } \dot{\mathbf{H}} \\ &= -j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E}\end{aligned}$$

Recall that

$$\text{curl curl } \mathbf{E} = \text{grad div } \mathbf{E} - \nabla^2\mathbf{E}$$

Combining these last two equations, there results

$$\nabla^2\mathbf{E} - j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E} = \text{grad div } \mathbf{E} \qquad (5\text{-}46)$$

Now for any homogeneous medium in which ϵ is constant

$$\text{div } \mathbf{E} = \frac{1}{\epsilon}\text{ div } \mathbf{D}$$

But div $\mathbf{D} = \rho$, and since there is no *net* charge within a conductor

* Equation (42) is the vector statement (applicable to an elemental volume) of the more familiar relation $I = V/R$ (see problem 7, chap. 3).

(although there may be a charge on the surface), the charge density ρ equals zero* and therefore

$$\text{div } \mathbf{D} = 0$$

Equation (46) then becomes

$$\nabla^2\mathbf{E} - j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E} = 0 \tag{5-47}$$

This is the wave equation for **E**. The wave equation for **H** is obtained in a similar manner.

$$\text{curl curl } \mathbf{H} = (\sigma + j\omega\epsilon) \text{ curl } \mathbf{E}$$

$$\text{curl } \mathbf{E} = -j\omega\mu\mathbf{H}$$

$$\text{grad div } \mathbf{H} - \nabla^2\mathbf{H} = (-j\omega\mu)(\sigma + j\omega\epsilon)\mathbf{H}$$

But
$$\text{div } \mathbf{H} = \frac{1}{\mu} \text{ div } \mathbf{B} = 0$$

Therefore
$$\nabla^2\mathbf{H} - (j\omega\mu)(\sigma + j\omega\epsilon)\mathbf{H} = 0 \tag{5-48}$$

This is the wave equation for **H**.

Equations (47) and (48) are the general wave equations for a homogeneous conducting medium and sinusoidal time variations.

Wave Propagation in a Conducting Medium. The solutions of general wave eqs. (47) and (48) in a conducting medium will as before yield expressions for a wave. In this case, however, on account of the finite conductivity, there will be loss in the medium and the wave will be attenuated as it progresses. It is desirable to know the value of the attenuation constant in terms of the constants of the medium.

Equation (47) may be written in the form

$$\nabla^2\mathbf{E} - \gamma^2\mathbf{E} = 0 \tag{5-49}$$

where

$$\gamma^2 = (j\omega\mu)(\sigma + j\omega\epsilon)$$

* The statement of no *net* charge within a conductor is consistent with our notion of current flow as a drift of free negative electrons through the positive atomic lattice of the conductor. Within any macroscopic element of volume the positive and negative charges are equal in number (on the average), and the net charge is zero. It is easily shown for steady-state sinusoidal time variations that div D = 0 (and therefore $\rho = 0$) in conductors is a direct consequence of Maxwell's equations and Ohm's law (see problem 3). It can also be shown that if a charge ever were placed within a conductor (in some manner not explained) the "transient time" or "relaxation time" required for this charge to appear on the surface would be exceedingly small for any materials considered to be conductors (see problem 4).

In general, the constant γ is complex and has real and imaginary parts designated by α and β respectively. That is, $\gamma = \alpha + j\beta$.

Again consider a uniform plane wave traveling in the x direction. For this case, (49) becomes

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = \gamma^2 \mathbf{E} \tag{5-50}$$

A possible solution for (50) would be

$$\mathbf{E} = \mathbf{E}' \, e^{\pm \gamma x}$$

For reasons, which will become apparent, use the minus sign and consider the solution

$$\mathbf{E} = \mathbf{E}' \, e^{-\gamma x} \tag{5-51}$$

When $\mathbf{E}'$ is expressed explicitly as a function of time as for example

$$\mathbf{E}' = E_0 \, e^{j\omega t}$$

eq. (51) can be written

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_0 \, e^{j\omega t} \, e^{-\gamma x} \\ &= \mathbf{E}_0 \, e^{j\omega t} \, e^{-\alpha x} \, e^{-j\beta x} \\ &= \mathbf{E}_0 \, e^{-\alpha x} \, e^{j(\omega t - \beta x)} \end{aligned} \tag{5-52}$$

Equation (52) is the equation of a wave moving in the x direction with a velocity ω/β. The wave is attenuated by the factor $e^{-\alpha x}$. α and β are the real and imaginary parts respectively of

$$\gamma = \sqrt{(j\omega\mu)(\sigma + j\omega\epsilon)}$$

The constant γ is known as the propagation constant for the wave. As is seen from eq. (52), α, the real part of γ, is a measure of the rate at which the wave is attenuated as it progresses through the medium. β, the imaginary part of γ, is the phase shift per unit length for the wave. Since the phase shifts through a complete cycle, or 2π radians, for each wave length,

$$\beta = \frac{2\pi}{\lambda}$$

The velocity of propagation of the wave, or the phase velocity is given by

$$v = \lambda f = \frac{\omega}{\beta}$$

In terms of the "primary" constants of the medium, that is σ, μ, and ϵ, the values of α and β are

$$\alpha = \text{real part of } \sqrt{(j\omega\mu)(\sigma + j\omega\epsilon)}$$

$$= \omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} - 1\right)} \tag{5-53}$$

$$\beta = \omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} + 1\right)} \tag{5-54}$$

Problem 2. From the expression $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$ derive expressions (53) and (54).

Problem 3. Using Maxwell's equation I show that

$$\text{div } D = 0 \text{ in a conductor,}$$

if Ohm's law and sinusoidal time variations (i.e., as $e^{j\omega t}$) are assumed.

Problem 4. Using div D $= \rho$, Ohm's law, and the equation of continuity show that if at any instant a charge density ρ existed within a conductor, it would decrease to $1/e$ times this value in a time ϵ/σ seconds. Calculate this time for a copper conductor.

5.06 Conductors and Dielectrics. In electromagnetics, materials are divided roughly into two classes; conductors and dielectrics or insulators. The dividing line between the two classes is not sharp and some media (for example the earth) are considered as conductors in one part of the radio frequency range, but as dielectrics (with loss) in another part of the range.

In Maxwell's first equation:

$$\text{Curl } \mathbf{H} = \sigma\mathbf{E} + j\omega\epsilon\mathbf{E}$$

the first term on the right is conduction current density and the second term is displacement current density. The ratio $\sigma/\omega\epsilon$ is therefore just the ratio of conduction current density to displacement current density in the medium. Hence, $\sigma/\omega\epsilon = 1$ can be considered to mark the dividing line between conductors and dielectrics. For *good conductors* such as metals $\sigma/\omega\epsilon$ is very much greater than unity over the entire radio frequency spectrum. For example for copper, even at the relatively high frequency of 30,000 mc, $\sigma/\omega\epsilon$ is about $3 \cdot 5 \times 10^8$. For *good dielectrics* or insulators $\sigma/\omega\epsilon$ is very much less than unity in the radio frequency range. For example, for mica at audio or radio frequencies $\sigma/\omega\epsilon$ is of the order of 0.0002. For good conductors σ and ϵ are nearly independent of frequency, but for most materials classed as dielectrics the "constants" σ and ϵ are functions of frequency. It has been found for these materials that the ratio $\sigma/\omega\epsilon$ is often relatively constant over the frequency range of interest. For this and other reasons the properties of dielectrics are usually given in terms of the dielectric "constant" ϵ and the ratio $\sigma/\omega\epsilon$. Under these circumstances the ratio $\sigma/\omega\epsilon$ is known as the *dissipation factor D* of the dielectric. For reasonably good dielectrics, that is those having small values of D, the dissipation factor is practically the same as the *power factor* of the dielectric. Actually, power factor is given by

$$\text{P.F.} = \sin \phi$$

where

$$\phi = \tan^{-1} D$$

Dissipation factor and power factor differ by less than 1 per cent when their values are less than 0.15.

Most materials used in radio are required either to pass conduction currents readily or to prevent the flow of conduction current as completely as possible. For this reason most materials met with in practice will fall into either the good conductor or the good insulator class. The important practical exception is the earth, which occupies an in-between position throughout most of the radio frequency spectrum. This case will be treated in detail in the chapter on propagation. For both good conductors and good dielectrics certain approximations are valid which simplify considerably the expressions for α and β.

Wave Propagation in Good Dielectrics. For this case $\sigma/\omega\epsilon \ll 1$ so that it is possible to write to a very good approximation

$$\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} \cong \left(1 + \frac{\sigma^2}{2\omega^2\epsilon^2}\right)$$

where only the first two terms of the binomial expansion have been used. Then expression (53) for α becomes

$$\alpha \cong \omega \sqrt{\frac{\mu\epsilon}{2}\left[\left(1 + \frac{\sigma^2}{2\omega^2\epsilon^2}\right) - 1\right]} = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon}} \qquad (5\text{-}55)$$

This expression may be compared with the expression for the attenuation factor of a low-loss transmission line having zero series resistance. In that case the expression for α is

$$\alpha = \frac{G}{2}\sqrt{\frac{L}{C}} = \frac{G}{2} Z_0$$

The expression for β reduces in a similar manner

$$\beta \cong \omega \sqrt{\frac{\mu\epsilon}{2}\left[\left(1 + \frac{\sigma^2}{2\omega^2\epsilon^2}\right) + 1\right]} = \omega \sqrt{\mu\epsilon}\left(1 + \frac{\sigma^2}{8\omega^2\epsilon^2}\right) \quad (5\text{-}56)$$

$\omega \sqrt{\mu\epsilon}$ is the phase shift factor for a perfect dielectric. The effect of a small amount of loss is to add the second term of (56) as a small correction factor. The velocity of the wave in the dielectric is given by

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\epsilon}\left(1 + \frac{\sigma^2}{8\omega^2\epsilon^2}\right)}$$

$$\cong v_0\left(1 - \frac{\sigma^2}{8\omega^2\epsilon^2}\right) \quad (5\text{-}57)$$

Where $v_0 = 1/\sqrt{\mu\epsilon}$ is the velocity of the wave in the dielectric when the conductivity is zero. The effect of a small amount of loss is to reduce slightly the velocity of propagation of the wave. It will be shown later that the general expression for the intrinsic or characteristic impedance of a medium which has a finite conductivity is

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}}$$

Using the same approximations as above, this becomes for a good dielectric

$$\eta = \sqrt{\frac{\mu}{\epsilon}\left(\frac{1}{1 + \frac{\sigma}{j\omega\epsilon}}\right)}$$

$$\cong \sqrt{\frac{\mu}{\epsilon}}\left(1 + j\frac{\sigma}{2\omega\epsilon}\right)$$

Since $\sqrt{\mu/\epsilon}$ is the intrinsic impedance of the dielectric when $\sigma = 0$, it is seen that the chief effect of a small amount of loss is to add a small reactive component to the intrinsic impedance.

Wave Propagation in a Good Conductor. For this case $\frac{\sigma}{\omega\epsilon} \gg 1$ so that the expression for γ may be written

$$\gamma = \sqrt{(j\omega\mu\sigma)\left(1 + j\frac{\omega\epsilon}{\sigma}\right)}$$

$$\cong \sqrt{j\omega\mu\sigma} = \sqrt{\omega\mu\sigma}\ \underline{/45^\circ}$$

Therefore
$$\alpha = \beta = \sqrt{\frac{\omega\mu\sigma}{2}}$$

The velocity of the wave in the conductor will be

$$v = \frac{\omega}{\beta} = \sqrt{\frac{2\omega}{\mu\sigma}}$$

and the intrinsic impedance of the conductor is

$$\eta \cong \sqrt{\frac{j\omega\mu}{\sigma}} = \sqrt{\frac{\omega\mu}{\sigma}}\ \underline{/45^\circ}$$

It is seen that in good conductors where σ is very large, both α and β are also large. This means that the wave is attenuated greatly as it progresses through the conductor and the phase shift per unit length is also great. The velocity of the wave, being inversely proportional to β, is very small in a good conductor, and is of the same order of magnitude as that of a sound wave in air. The characteristic impedance is also very small and has a *reactive* component. The angle of this impedance is always 45 degrees for good conductors.

Depth of Penetration. In a medium which has conductivity the wave is attenuated as it progresses owing to the losses which occur. In a good conductor at radio frequencies the rate of attenuation is very great and the wave may penetrate only a very short distance before being reduced to a negligibly small percentage of its original strength. A term that has significance under such circumstances is the *depth of penetration.* The depth of penetration, δ, is defined as that depth in which the wave has been attenuated to $1/e$ or approximately 37 per cent of its original value. Since the amplitude decreases by the factor $e^{-\alpha x}$ it is apparent that at that distance x, which makes $\alpha x = 1$, the amplitude is only $1/e$ times its value at $x = 0$. By definition this distance is equal to δ, the depth of penetration; so

$$\alpha\delta = 1 \qquad \text{or} \qquad \delta = \frac{1}{\alpha}$$

The general expression for depth of penetration is

$$\delta = \frac{1}{\alpha} = \frac{1}{\omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} - 1\right)}}$$

For a *good* conductor the depth of penetration is

$$\delta = \frac{1}{\alpha} \cong \sqrt{\frac{2}{\omega\mu\sigma}}$$

As an example of the order of magnitude of δ in metals, the depth of penetration of a megacycle wave into copper which has a conductivity $\sigma = 5.8 \times 10^7$ mhos per meter and a permeability approximately equal to that of free space is

$$\delta = \sqrt{\frac{2 \times 10^7}{2\pi \times 10^6 \times 4\pi \times 5.8 \times 10^7}} = 0.0667 \text{ mm}$$

At 100 mc it is 0.00667 mm, whereas at 60 cps, it is 8.67 mm.

Problem 5. Earth is considered to be a good conductor when $\omega\epsilon/\sigma \ll 1$. Determine the highest frequencies for which earth can be considered a good conductor if $\ll 1$ means less than 0.1. Assume the following constants:

$$\sigma = 5 \times 10^{-3} \text{ mho/meter} \qquad \epsilon = 10\epsilon_v$$

Problem 6. A copper wire carries a conduction current of 1 amp. Determine the displacement current in the wire at 100 mc. (Assume that copper has about the same permittivity as free space, that is $\epsilon = \epsilon_v$. For copper $\sigma = 5.8 \times 10^7$ mhos/m.)

PART II—REFLECTION AND REFRACTION OF PLANE WAVES

5.07 Reflection by a Perfect Conductor—Normal Incidence. When an electromagnetic wave traveling in one medium impinges upon a second medium having a different dielectric constant, permeability, or conductivity, the wave in general will be partially transmitted and partially reflected. In the case of a plane wave in air incident normally upon the surface of a perfect conductor, the wave is entirely reflected. For fields that vary with time neither **E** nor **H** can exist within a perfect conductor so that none of the energy of the incident wave can be transmitted. Since there can be no loss within a perfect conductor, none of the energy is absorbed. As a result the amplitudes of **E** and **H** in the reflected wave are the same as in the incident wave, and the only difference is in the direction of power flow. If the expression for the electric field of the *incident* wave is

$$E_i\, e^{j(\omega t - \beta x)}$$

and the surface of the perfect conductor is taken to be the $x = 0$ plane, the expression for the reflected wave will be

$$E_r\, e^{j(\omega t + \beta x)}$$

where E_r must be determined from the boundary conditions. Inasmuch as the tangential component of **E** must be continuous across the boundary and **E** is zero within the conductor, the tangential component of **E** just outside the conductor must also be zero. This requires that the sum of the electric intensities in the initial and reflected waves add to give zero resultant intensity in the plane $x = 0$. Therefore

$$E_r = -E_i$$

The amplitude of the reflected electric intensity is equal to that of the initial electric intensity, but its phase has been reversed on reflection.

The resultant electric field at any point a distance $-x$ from the $x = 0$ plane will be the sum of the intensities of the incident and reflected waves at that point and will be given by

$$\begin{aligned} E_T &= E_i\, e^{j(\omega t - \beta x)} + E_r\, e^{j(\omega t + \beta x)} \\ &= E_i[e^{j(\omega t - \beta x)} - e^{j(\omega t + \beta x)}] \\ &= E_i\, e^{j\omega t}(e^{-j\beta x} - e^{+j\beta x}) \\ &= -\,2jE_i \sin \beta x\, e^{j\omega t} \end{aligned} \tag{5-58}$$

Equation (58) shows that the incident and reflected waves combine to produce a *standing wave*, which does not progress. The magnitude of the electric field varies sinusoidally with distance from the reflecting plane. It is zero at the surface and at multiples of half wavelength from the surface. It has a maximum value of twice the electric intensity of the incident wave at distances from the surface that are odd multiples of a quarter wavelength.

Inasmuch as the boundary conditions require that the electric intensity be reversed in phase on reflection in order to produce zero resultant field at the surface, it follows that the magnetic intensity must be reflected without reversal of phase. If both magnetic and electric intensities were reversed, there would be no reversal of direction of energy propagation, which is required in

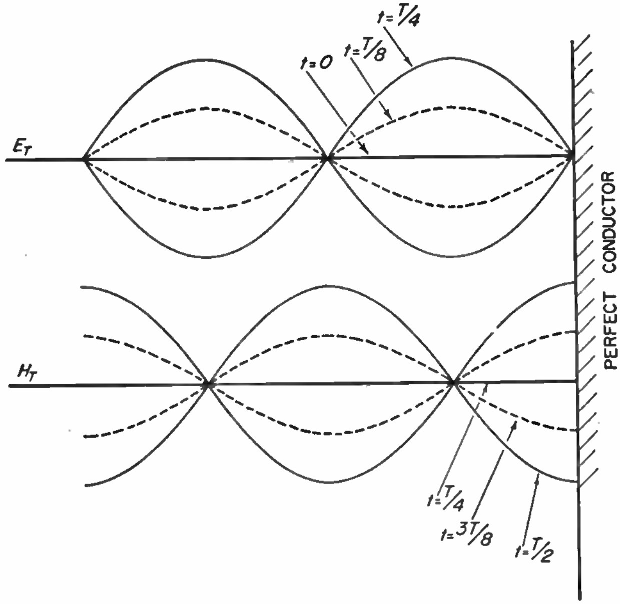


FIG. 5-4. Standing waves of E and H.

this case. Therefore, the phase of the reflected magnetic intensity H_r is the same* as that of the incident magnetic intensity H_i at the surface of reflection $x = 0$. The expression for the resultant mag-

* An alternative way of arriving at this same result is from a consideration of current flow in the conductor. If it is assumed for the incident wave, which is traveling to the right in the positive x direction, that E_i is in the positive y direction and H_i is in the positive z direction (it will be seen later that the direction of energy propagation is always the direction of the vector $\mathbf{E} \times \mathbf{H}$), the current flow in the conductor will be in the same direction as the incident electric field, that is, in the positive y direction. This current flow produces an electric field $-E_y$ to oppose the incident field (Lenz's law) and produces a magnetic field, which is shown by application of the right hand rule to be in the positive z direction. Therefore the magnetic field of the reflected wave has the *same* direction as in the incident wave.

netic field will be

$$\begin{aligned} H_T &= H_i\, e^{j(\omega t - \beta x)} + H_r\, e^{j(\omega t + \beta x)} \\ &= H_i[e^{j(\omega t - \beta x)} + e^{j(\omega t + \beta x)}] \\ &= 2H_i \cos \beta x\, e^{j\omega t} \end{aligned} \qquad (5\text{-}59)$$

The resultant magnetic intensity H also has a standing wave distribution. In this case, however, it has maximum value at the surface of the conductor and at multiples of a half wavelength from the surface, whereas the zero points occur at odd multiples of a quarter wavelength from the surface. From the boundary conditions for H it follows that there must be a surface current of J amperes per meter, such that $J = H_T$ (at $x = 0$).

Since E_i and H_i were in time phase in the incident plane wave, a comparison of (58) and (59) shows that E_T and H_T are 90 degrees out of time phase because of the factor j in (58). This is as it should be, for it indicates no average flow of power. This is the case when the energy transmitted in the forward direction is equalled by that reflected back.

That E_T and H_T are 90 degrees apart in time phase can be seen more clearly by rewriting (58) and (59). Replacing $-j$ by its equivalent $e^{-j(\pi/2)}$ and combining this with the $e^{j\omega t}$ term to give $e^{j[\omega t - (\pi/2)]}$ eq. (58) becomes

$$E_T = 2E_i \sin \beta x\, e^{j[\omega t - (\pi/2)]} \qquad (5\text{-}58a)$$

Recalling that only the real (or only the imaginary) part of the $e^{j[\omega t - (\pi/2)]}$ term is to be used finally, (58a) means

$$E_T = 2E_i \sin \beta x \cos \left(\omega t - \frac{\pi}{2}\right) \qquad (5\text{-}58b)$$

Likewise rewriting (59),

$$H_T = 2H_i \cos \beta x \cos (\omega t) \qquad (5\text{-}59a)$$

Comparison of (58b) and (59a) shows that E_T and H_T differ in time phase by $\pi/2$ radians or 90 degrees.

5.08 Reflection by a Perfect Dielectric—Normal Incidence. When a plane electromagnetic wave is incident normally on the surface of a perfect dielectric, part of the energy is transmitted and part of it is reflected. A *perfect* dielectric is one with zero conductivity, so that there is no loss or absorption of power in propagation through the dielectric.

As before, consider the case of a plane wave traveling in the x direction incident on a boundary that is parallel to the $x = 0$ plane. Let E_i be the electric intensity of the incident wave striking the boundary, E_r be the electric intensity of the reflected wave leaving the boundary in the first medium, and E_t be the electric intensity of the transmitted wave propagated into the second medium. Similar subscripts will be applied to the magnetic intensity H. Let ϵ_1 and μ_1 be the constants of the first medium and ϵ_2 and μ_2 be the constants of the second medium. Designating by η_1 and η_2, the ratios $\sqrt{\mu_1/\epsilon_1}$ and $\sqrt{\mu_2/\epsilon_2}$ the following relations will hold

$$E_i = \eta_1 H_i$$
$$E_r = -\eta_1 H_r$$
$$E_t = \eta_2 H_t$$

The continuity of the tangential components of **E** and **H** require that

$$H_i + H_r = H_t$$
$$E_i + E_r = E_t$$

Combining these

$$H_i + H_r = \frac{1}{\eta_1}(E_i - E_r) = H_t = \frac{1}{\eta_2}(E_i + E_r)$$
$$\eta_2(E_i - E_r) = \eta_1(E_i + E_r)$$
$$E_i(\eta_2 - \eta_1) = E_r(\eta_2 + \eta_1)$$
$$\frac{E_r}{E_i} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{5-60}$$

Also

$$\frac{E_t}{E_i} = \frac{E_i + E_r}{E_i} = 1 + \frac{E_r}{E_i}$$
$$= \frac{2\eta_2}{\eta_2 + \eta_1} \tag{5-61}$$

Furthermore

$$\frac{H_r}{H_i} = -\frac{E_r}{E_i} = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2} \tag{5-62}$$
$$\frac{H_t}{H_i} = \frac{\eta_1}{\eta_2}\frac{E_t}{E_i} = \frac{2\eta_1}{\eta_1 + \eta_2} \tag{5-63}$$

The permeabilities of all known insulators do not differ appreciably from that of free space, so that $\mu_1 = \mu_2 = \mu_v$. Inserting this relation the above expressions can be written in terms of the dielectric constants as follows:

$$\frac{E_r}{E_i} = \frac{\sqrt{\mu_v/\epsilon_2} - \sqrt{\mu_v/\epsilon_1}}{\sqrt{\mu_v/\epsilon_2} + \sqrt{\mu_v/\epsilon_1}}$$

$$\frac{E_r}{E_i} = \frac{\sqrt{\epsilon_1} - \sqrt{\epsilon_2}}{\sqrt{\epsilon_1} + \sqrt{\epsilon_2}} \tag{5-64}$$

Similarly

$$\frac{E_t}{E_i} = \frac{2\sqrt{\epsilon_1}}{\sqrt{\epsilon_1} + \sqrt{\epsilon_2}} \tag{5-65}$$

$$\frac{H_r}{H_i} = \frac{\sqrt{\epsilon_2} - \sqrt{\epsilon_1}}{\sqrt{\epsilon_1} + \sqrt{\epsilon_2}} \tag{5-66}$$

$$\frac{H_t}{H_i} = \frac{2\sqrt{\epsilon_2}}{\sqrt{\epsilon_1} + \sqrt{\epsilon_2}} \tag{5-67}$$

5.09 Reflection by a Perfect Insulator—Oblique Incidence. If a plane wave is incident upon a boundary surface that is not parallel

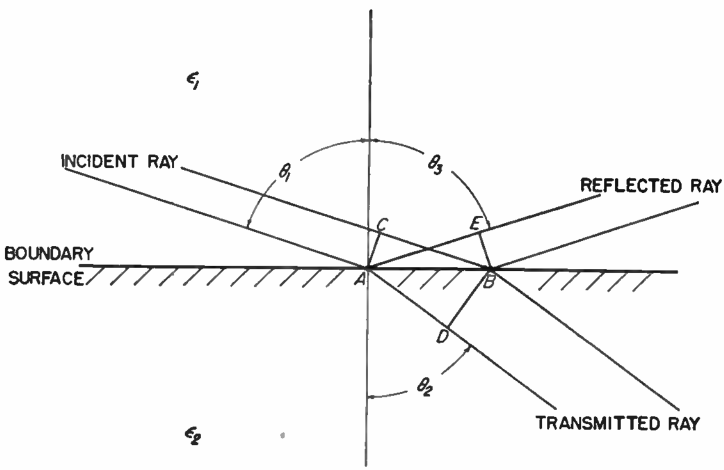


FIG. 5-5. Reflection and refraction.

to the plane containing **E** and **H**, the boundary conditions are more complex. Again part of the wave will be transmitted and part of it reflected, but in this case the transmitted wave will be refracted; that is the direction of propagation will be altered. Consider Fig. 5-5, which shows a ray of the wave. (A ray is a line drawn normal to the equiphase surfaces.)

In the diagram the one side of the incident ray travels the distance CB, whereas the other side of the transmitted ray travels the distance AD and the left side of the reflected ray travels from A to E. If v_1 is the velocity of the wave in medium (1) and v_2 is the velocity in medium (2), then

$$\frac{CB}{AD} = \frac{v_1}{v_2}$$

Now $CB = AB \sin \theta_1$ and $AD = AB \sin \theta_2$, so that

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2}$$

In terms of the constants of the media, v_1 and v_2 are given by

$$v_1 = \frac{1}{\sqrt{\mu_1 \epsilon_1}} = \frac{1}{\sqrt{\mu_v \epsilon_1}}$$
$$v_2 = \frac{1}{\sqrt{\mu_2 \epsilon_2}} = \frac{1}{\sqrt{\mu_v \epsilon_2}}$$

Therefore

$$\frac{\sin \theta_1}{\sin \theta_2} = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \qquad (5\text{-}68)$$

Furthermore

$$AE = CB$$

and as a result, $\sin \theta_1 = \sin \theta_3$, or

$$\theta_1 = \theta_3 \qquad (5\text{-}69)$$

The angle of incidence is equal to the angle of reflection; the angle of incidence is related to the angle of refraction by eq. (68), which in optics is known as the *law of sines*, or *Snell's law*.

In a later section it will be shown that the power transmitted per square meter in a wave is the vector product of **E** and **H**. Since **E** and **H** are at right angles to each other, in this case the power transmitted per square meter is equal to E^2/η. The power in the incident wave striking AB will be proportional to $(1/\eta_1)E_i^2 \cos \theta_1$, that reflected will be $(1/\eta_1)E_r^2 \cos \theta_1$ and that transmitted through the boundary will be $(1/\eta_2)E_t^2 \cos \theta_2$. By the conservation of energy

$$\frac{1}{\eta_1} E_i^2 \cos \theta_1 = \frac{1}{\eta_1} E_r^2 \cos \theta_1 + \frac{1}{\eta_2} E_t^2 \cos \theta_2$$

$$\frac{E_r^2}{E_i^2} = 1 - \frac{\eta_1 E_t^2 \cos \theta_2}{\eta_2 E_i^2 \cos \theta_1}$$

$$\frac{E_r^2}{E_i^2} = 1 - \frac{\sqrt{\epsilon_2}\, E_t^2 \cos \theta_2}{\sqrt{\epsilon_1}\, E_i^2 \cos \theta_1} \tag{5-70}$$

It is necessary to consider separately two cases. The first of these is the case in which the electric vector is parallel to the

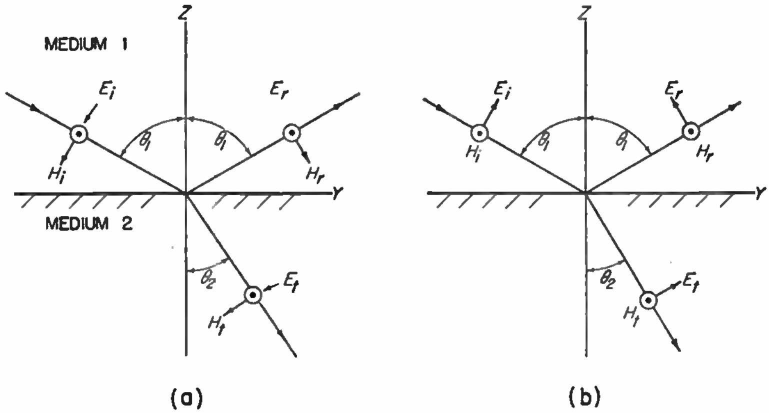


FIG. 5-6. Reflection and refraction waves that have (a) perpendicular (horizontal) polarization, and (b) parallel (vertical) polarization.

boundary surface or perpendicular to the plane of incidence. (The plane of incidence is the plane containing the incident ray and the normal to the surface.) This case is often termed *horizontal polarization*. In the second case the magnetic vector is parallel to the boundary surface, and the electric vector is parallel to the plane of incidence. This case is often termed vertical polarization. The two cases are shown in Fig. 5-6. The terms "horizontally and vertically polarized waves" refer to the fact that waves from horizontal and vertical antennas, respectively, would produce these particular orientations of electric and magnetic vectors in waves striking the surface of the earth. However, it is seen that, whereas the electric vector of a "horizontally" polarized wave is horizontal, the electric vector of a "vertically" polarized wave is not wholly vertical but has some horizontal component. More significant designations are the terms "perpendicular" and "parallel" polarization to indicate that the electric vector is perpendicular or parallel to the plane of incidence. In wave guide work the terms transverse electric (TE) and transverse magnetic (TM) are used to indicate that the electric or magnetic vector respectively is parallel to the boundary plane. The reason for this will be discussed later.

Case I: *Perpendicular (Horizontal) Polarization.* In this case the electric vector **E** is perpendicular to the plane of incidence and parallel to the reflecting surface. Let the electric intensity E_i of the incident wave be in the positive x direction (outward in Fig. 5-6a), and let the assumed positive directions for E_r and E_t in the reflected and transmitted waves also be in the positive x direction. Then, applying the boundary condition that the tangential component of **E** is continuous across the boundary,

$$E_i + E_r = E_t$$

$$\frac{E_t}{E_i} = 1 + \frac{E_r}{E_i} \tag{5-71}$$

Insert this in eq. (70)

$$\frac{E_r^2}{E_i^2} = 1 - \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 + \frac{E_r}{E_i}\right)^2 \frac{\cos\theta_2}{\cos\theta_1}$$

$$1 - \left(\frac{E_r}{E_i}\right)^2 = \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 + \frac{E_r}{E_i}\right)^2 \frac{\cos\theta_2}{\cos\theta_1}$$

$$1 - \frac{E_r}{E_i} = \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 + \frac{E_r}{E_i}\right) \frac{\cos\theta_2}{\cos\theta_1}$$

$$\frac{E_r}{E_i} = \frac{\sqrt{\epsilon_1}\cos\theta_1 - \sqrt{\epsilon_2}\cos\theta_2}{\sqrt{\epsilon_1}\cos\theta_1 + \sqrt{\epsilon_2}\cos\theta_2}$$

Now from eq. (68)

$$\sqrt{\epsilon_2}\cos\theta_2 = \sqrt{\epsilon_2(1 - \sin^2\theta_2)} = \sqrt{\epsilon_2 - \epsilon_1\sin^2\theta_1}$$

therefore

$$\frac{E_r}{E_i} = \frac{\sqrt{\epsilon_1}\cos\theta_1 - \sqrt{\epsilon_2 - \epsilon_1\sin^2\theta_1}}{\sqrt{\epsilon_1}\cos\theta_1 + \sqrt{\epsilon_2 - \epsilon_1\sin^2\theta_1}} \tag{5-72}$$

$$= \frac{\cos\theta_1 - \sqrt{(\epsilon_2/\epsilon_1) - \sin^2\theta_1}}{\cos\theta_1 + \sqrt{(\epsilon_2/\epsilon_1) - \sin^2\theta_1}} \tag{5-72a}$$

Equation (72) gives the ratio of reflected to incident electric intensity for the case of a perpendicularly polarized wave.

CASE II: *Parallel (Vertical) Polarization.* In this case **E** is parallel to the plane of incidence and **H** is parallel to the reflecting surface. Again applying the boundary condition that the tangential component of **E** is continuous across the boundary in this case gives (Fig. 5-6b),

$$(E_i - E_r)\cos\theta_1 = E_t\cos\theta_2$$

$$\frac{E_t}{E_i} = \left(1 - \frac{E_r}{E_i}\right)\frac{\cos\theta_1}{\cos\theta_2}$$

Insert this in eq. (70)

$$\left(\frac{E_r}{E_i}\right)^2 = 1 - \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 - \frac{E_r}{E_i}\right)^2\frac{\cos\theta_1}{\cos\theta_2}$$

$$1 - \frac{E_r^2}{E_i^2} = \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 - \frac{E_r}{E_i}\right)^2\frac{\cos\theta_1}{\cos\theta_2}$$

$$1 + \frac{E_r}{E_i} = \sqrt{\frac{\epsilon_2}{\epsilon_1}}\left(1 - \frac{E_r}{E_i}\right)\frac{\cos\theta_1}{\cos\theta_2}$$

$$\frac{E_r}{E_i}\left(1 + \sqrt{\frac{\epsilon_2}{\epsilon_1}}\frac{\cos\theta_1}{\cos\theta_2}\right) = \sqrt{\frac{\epsilon_2}{\epsilon_1}}\frac{\cos\theta_1}{\cos\theta_2} - 1$$

$$\frac{E_r}{E_i} = \frac{\sqrt{\epsilon_2}\cos\theta_1 - \sqrt{\epsilon_1}\cos\theta_2}{\sqrt{\epsilon_2}\cos\theta_1 + \sqrt{\epsilon_1}\cos\theta_2}$$

$$= \frac{\sqrt{\epsilon_2}\cos\theta_1 - \sqrt{\epsilon_1(1 - \sin^2\theta_2)}}{\sqrt{\epsilon_2}\cos\theta_1 + \sqrt{\epsilon_1(1 - \sin^2\theta_2)}}$$

Recall that $\sin^2\theta_2 = \dfrac{\epsilon_1}{\epsilon_2}\sin^2\theta_1$

$$\frac{E_r}{E_i} = \frac{(\epsilon_2/\epsilon_1)\cos\theta_1 - \sqrt{(\epsilon_2/\epsilon_1) - \sin^2\theta_1}}{(\epsilon_2/\epsilon_1)\cos\theta_1 + \sqrt{(\epsilon_2/\epsilon_1) - \sin^2\theta_1}} \tag{5-73}$$

Equation (73) gives the reflection coefficient for parallel or vertical polarization, that is, the ratio of reflected to incident electric intensity when **E** is parallel to the plane of incidence.

Brewster Angle. Of particular interest is the possibility in eq. (73) of obtaining no reflection at a particular angle. This occurs when the numerator is zero. For this case

$$\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1} = \frac{\epsilon_2}{\epsilon_1}\cos\theta_1$$

$$\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_1 = \frac{\epsilon_2^2}{\epsilon_1^2} - \frac{\epsilon_2^2}{\epsilon_1^2}\sin^2\theta_1$$

$$(\epsilon_1^2 - \epsilon_2^2)\sin^2\theta_1 = \epsilon_2(\epsilon_1 - \epsilon_2)$$

$$\sin^2\theta_1 = \frac{\epsilon_2}{\epsilon_1 + \epsilon_2}$$

$$\cos^2\theta_1 = \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}$$

$$\tan\theta_1 = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{5-74}$$

At this angle, which is called the *Brewster angle,* there is no reflected wave when the incident wave is parallel (or vertically) polarized. If the incident wave is not entirely parallel polarized, there will be some reflection, but the reflected wave will be entirely of perpendicular (or horizontal) polarization.

Examination of eq. (72), which is for perpendicular polarization, shows that there is no corresponding Brewster angle for this polarization.

5.10 Direction Cosines. Sometimes it is necessary to write the expression for a plane wave that is traveling in some arbitrary direction with respect to a fixed set of axes. This is most conveniently done in terms of the direction cosines of the normal to the plane of the wave. By definition of a uniform plane wave the equiphase surfaces are planes. Thus in the expression

$$E = E_1\, e^{j\omega[t-(x/v)]}$$

for a wave traveling in the x direction, the planes of constant phase are given by the equation

$$x = \text{a constant}$$

For a plane wave traveling in some arbitrary direction, say the s direction, it is necessary to replace x with an expression that, when put equal to a constant, gives the equiphase surfaces.

The equation of a plane is given by

$$\mathbf{N} \cdot \mathbf{r} = \text{a constant}$$

where **r** is the radius vector from the origin to any point P on the plane and **N** is the *unit vector* normal to the plane.* That this is so can be seen from Fig. 5-7, in which a plane perpendicular to the unit vector **N** intersects the y-z plane along the line A-A. The dot product $\mathbf{N} \cdot \mathbf{r}$ is the projection of the radius vector **r** along the normal to the plane, and it is apparent that this will have the constant value OM for all points on the plane. Now the dot product of two vectors is a scalar equal to the sum of the products of the components of the vectors along the axes of the coordinate system. Therefore

$$\mathbf{N} \cdot \mathbf{r} = lx + my + nz \qquad (5\text{-}75)$$

where x, y, z are the components of the vector **r** and l, m, n are the components of the unit vector **N** along the x, y, and z axes. The components l, m, and n are the cosines of the angles that the unit vector **N** makes with the positive x, y, and z axes, respectively, and are termed the *direction cosines* or *direction components* of the vector.

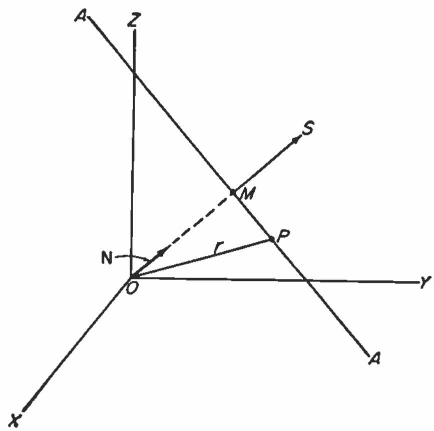


FIG. 5-7

The equation of a plane wave traveling in the direction **N**, normal to the planes of constant phase, can now be written as

$$E = E_1 e^{j\omega\left(t - \frac{\mathbf{N}\cdot\mathbf{r}}{v}\right)}$$

$$= E_1 e^{j\omega\left(t - \frac{lx+my+nz}{v}\right)} \qquad (5\text{-}76)$$

As an example of the use of such expressions for plane wave propagation, the reflection of a plane wave obliquely incident upon a perfect conductor will be considered.

5.11 Reflection by a Perfect Conductor—Oblique Incidence. When a plane wave is incident upon a perfect conductor at an oblique angle, the wave is totally reflected with the angle of inci-

* In this section capital **N** rather than lower case **n** is used for the unit normal in order to avoid possible confusion with the direction cosine n, defined later.

dence equal to the angle of reflection. As in the case of reflection from a dielectric, there will be two cases to consider, viz., **E** perpendicular to the plane of incidence and **E** parallel to the plane of incidence.

CASE I: **E** *Perpendicular to the Plane of Incidence.* Let the incident and reflected waves make angles $\theta_i = \theta_r = \theta$ with the z axes as in Fig. 5-8. Because the directions of these two waves have oppositely directed components along the z axis, there must be a standing wave distribution along this axis. In the y direction the incident and reflected waves both progress

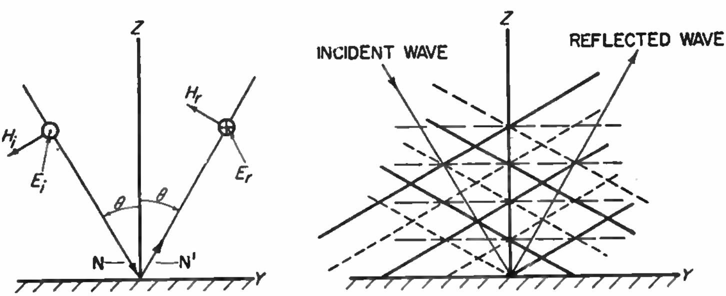


FIG. 5-8. Field pattern above a reflecting plane when the wave is incident at an oblique angle. (Perpendicular or horizontal polarization.)

to the right with the same velocity so there will be a traveling wave in the positive y direction. That these conclusions are correct can be seen by adding the expressions representing the two waves.

With the co-ordinate system chosen as shown in Fig. 5-8, the expression for the reflected wave is

$$E \text{ (reflected)} = E_r e^{j\omega\left(t - \frac{\mathbf{N}'\cdot\mathbf{r}}{v}\right)}$$

$$= E_r e^{j\omega\left(t - \frac{l'x + m'y + n'z}{v}\right)} \tag{5-77}$$

where E_r is the amplitude of the electric intensity of the reflected wave at the plane of reflection and l', m', n' are the direction cosines of the normal (N') to the wave front of the reflected wave. For the wave normal of the reflected wave

$$l' = 0, \qquad m' = \cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta, \qquad n' = \cos\theta$$

so that (77) becomes

$$E = E_r e^{j\omega\left(t - \frac{y \sin\theta + z\cos\theta}{v}\right)} \tag{5-78}$$

For the incident wave

$$l = 0, \qquad m = \cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta, \qquad n = \cos(\pi - \theta) = -\cos\theta$$

and

$$E_{\text{incident}} = E_i\, e^{j\omega\left(t - \frac{y \sin\theta - z\cos\theta}{v}\right)} \tag{5-79}$$

From the boundary conditions

$$E_r = -E_i$$

Therefore the total electric intensity (sum of incident and reflected intensities) will be

$$\begin{aligned} E &= E_i\left[e^{j\omega\left(t - \frac{y\sin\theta - z\cos\theta}{v}\right)} - e^{j\omega\left(t - \frac{y\sin\theta + z\cos\theta}{v}\right)}\right] \\ &= E_i\left(e^{j\frac{\omega z\cos\theta}{v}} - e^{-j\frac{\omega z\cos\theta}{v}}\right) e^{j\omega\left(t - \frac{y\sin\theta}{v}\right)} \\ &= 2jE_i \sin(\beta z\cos\theta)\, e^{j(\omega t - \beta y\sin\theta)} \\ &= 2jE_i \sin\beta_z z\, e^{j(\omega t - \beta_y y)} \end{aligned} \tag{5-80}$$

where $\beta = \omega/v = 2\pi/\lambda$ is the phase shift constant of the incident wave, $\beta_z = \beta\cos\theta$ is the phase shift constant in the z direction, and $\beta_y = \beta\sin\theta$ is the phase shift constant in the y direction. Equation (80) shows a standing-wave distribution of electric intensity along the z axis. The wavelength λ_z (twice the distance between nodal points), measured along this axis, is greater than the wavelength λ of the incident wave. The relation between the wavelengths is

$$\lambda_z = \frac{2\pi}{\beta_z} = \frac{2\pi}{\beta\cos\theta} = \frac{\lambda}{\cos\theta}$$

The planes of zero electric intensity occur at multiples of $\lambda_z/2$ from the reflecting surface. The planes of maximum electric intensity occur at odd multiples of $\lambda_z/4$ from the surface.

The whole standing wave distribution of electric intensity is seen from eq. (80) to be traveling in the y direction with a velocity

$$v_y = \frac{\omega}{\beta_y} = \frac{\omega}{\beta\sin\theta} = \frac{v}{\sin\theta}$$

This is the velocity with which a crest of the incident wave moves along the y axis. The wavelength in this direction is

$$\lambda_y = \frac{\lambda}{\sin\theta}$$

These relations between the velocities and wavelengths in the various directions are shown more clearly in Fig. 5-9, which shows successive crests of an incident wave intersecting with the y and z axes. For small

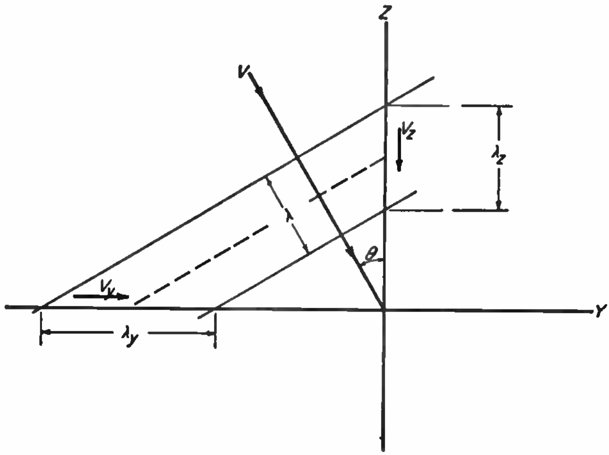


FIG. 5-9. Relations between wavelengths and velocities in different directions.

angles of θ it is seen that the velocity v_y, with which a crest moves along the y axis, becomes very great, approaching infinity as θ approaches zero.

CASE II: **E** *Parallel to the Plane of Incidence.* In this case E_i and E_r will have the instantaneous directions shown because the components parallel to the perfectly conducting boundary must be equal and opposite. The magnetic intensity vector **H** will be reflected without phase reversal as an examination of the direction of current flow will show. The magnitudes of **E** and **H** will be related by

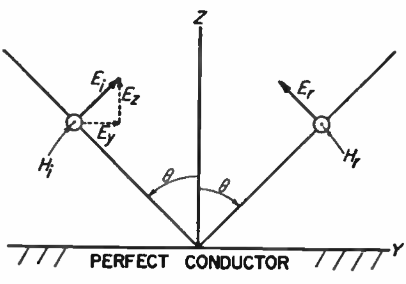


FIG. 5-10. Reflection of a parallel—or vertically—polarized wave.

$$\left|\frac{E_i}{H_i}\right| = \left|\frac{E_r}{H_r}\right| = \eta$$

For the incident wave the expression for magnetic intensity would be

$$H = H_i\, e^{j\omega\left(t - \frac{y \sin \theta - z \cos \theta}{v}\right)}$$

and for the reflected wave

$$H = H_r\, e^{j\omega\left(t - \frac{y \sin \theta + z \cos \theta}{v}\right)}$$

and, since $H_i = H_r$, the total magnetic intensity is

$$H = 2H_i \cos \beta_z z \, e^{j(\omega t - \beta_y y)} \tag{5-81}$$

where, as before
$$\beta_z = \frac{\omega}{v} \cos \theta$$
$$\beta_y = \frac{\omega}{v} \sin \theta$$

The magnetic intensity has a standing-wave distribution in the z direction with the planes of maximum H located at the conducting surface and at multiples of one-half λ_z from the surface. The planes of zero magnetic intensity occur at odd multiples of $\lambda_z/4$ from the surface.

In adding together the electric intensities of the incident and reflected waves it is necessary to consider separately the components in the y and z directions. For the initial wave

$$E_i = \eta H_i, \qquad E_z = \eta \sin \theta H_i, \qquad E_y = \eta \cos \theta H_i$$

For the reflected wave

$$H_r = H_i, \qquad E_z = \eta \sin \theta H_r, \qquad E_y = -\eta \cos \theta H_r$$

The total z component of electric intensity is

$$E_z = 2\eta \sin \theta H_i \cos \beta_z z \, e^{j(\omega t - \beta_y y)} \tag{5-82}$$

The total y component of electric intensity is

$$E_y = 2j\eta \cos \theta H_i \sin \beta_z z \, e^{j(\omega t - \beta_y y)} \tag{5-83}$$

where $$\beta_z = \frac{\omega}{v} \cos \theta \qquad \text{and} \qquad \beta_y = \frac{\omega}{v} \sin \theta$$

Both components of the electric intensity have a standing-wave distribution above the reflecting plane. However, for the normal or z component of E the maxima occur at the plane and multiples of $\lambda_z/2$ from the plane, whereas for the component of **E** parallel to the reflecting plane the minima occur at the plane and at multiples of $\lambda_z/2$ from the plane.

Problem 7. Sketch the planes of zero magnetic intensity, zero E_z, and zero E_y for the case of oblique reflection with **E** parallel to plane of incidence (Fig. 5-10).

5.12 The Transmission Line Analogy. The student familiar with ordinary transmission line theory cannot have failed to notice the similarity between the equations of wave propagation developed in this chapter and those giving voltage and current distributions along uniform transmission lines. The similarity is especially marked in the expressions for the reflection coefficients in the two cases. This similarity is more than a coincidence. There exists a close analogy between the propagation of plane waves in a homogeneous medium and the propagation of voltage and current along a uniform transmission line. This analogy is so close that it can be used not only as an aid in obtaining an understanding of a new subject, but also to obtain the solutions to actual problems. Because of his background in transmission-line theory the engineer often finds himself able to write directly the solutions to electromagnetic wave problems, or at any rate to set them up in terms of familiar circuit concepts. For these reasons the analogy will be considered step-by-step in some detail in order that the similarities, and the differences, may be fully understood.

For a uniform transmission line having the constants R, L, C and G per unit length, the voltage and current equations may be written in the differential form as

$$\left.\begin{aligned} \frac{\partial e}{\partial x} + L\frac{\partial i}{\partial t} + Ri &= 0 \\ \frac{\partial i}{\partial x} + C\frac{\partial e}{\partial t} + Ge &= 0 \end{aligned}\right\} \tag{5-84a}$$

For a homogeneous medium Maxwell's equations are

$$\text{curl } \mathbf{E} = -\mu\frac{\partial \mathbf{H}}{\partial t}$$

$$\text{curl } \mathbf{H} = \epsilon\frac{\partial \mathbf{E}}{\partial t} + \sigma\mathbf{E}$$

For a uniform plane wave propagating in the x direction and having only components E_y and H_z these become

$$\left.\begin{aligned} \frac{\partial E_y}{\partial x} + \mu\frac{\partial H_z}{\partial t} &= 0 \\ \frac{\partial H_z}{\partial x} + \epsilon\frac{\partial E_y}{\partial t} + \sigma E_y &= 0 \end{aligned}\right\} \tag{5-84b}$$

Inspection of these equations shows that the following quantities are analogous:

e (volt)	E (volt/m)
i (amp)	H (amp/m)
C (f/m)	ϵ (f/m)
L (henry/m)	μ (henry/m)
G (mho/m)	σ (mho/m)
R (ohm/m)	

In this analogy there appears to be nothing corresponding to R. The reason for this will be seen later. If the voltages and currents, and electric and magnetic intensities are assumed to vary sinusoidally with time, so that

$$e = V\,e^{j\omega t}, \qquad i = I\,e^{j\omega t}, \quad E_y = E\,e^{j\omega t}.\ H_z = H\,e^{j\omega t}$$

eqs. (84) become*

$$\left.\begin{aligned} \frac{\partial V}{\partial x} + (R + j\omega L)I &= 0 \\ \frac{\partial I}{\partial x} + (G + j\omega C)V &= 0 \end{aligned}\right\} \quad (5\text{-}85a) \qquad \left.\begin{aligned} \frac{\partial E}{\partial x} + (0 + j\omega\mu)H &= 0 \\ \frac{\partial H}{\partial x} + (\sigma + j\omega\epsilon)E &= 0 \end{aligned}\right\} \quad (5\text{-}85b)$$

Differentiating with respect to x, these equations combine to give the following second-order differential equations:

$$\left.\begin{aligned} \frac{\partial^2 V}{\partial x^2} - (R + j\omega L)(G + j\omega C)V &= 0 \\ \frac{\partial^2 I}{\partial x^2} - (R + j\omega L)(G + j\omega C)I &= 0 \end{aligned}\right\} \quad (5\text{-}86a) \qquad \left.\begin{aligned} \frac{\partial^2 E}{\partial x^2} - (j\omega\mu)(\sigma + j\omega\epsilon)E &= 0 \\ \frac{\partial^2 H}{\partial x^2} - (j\omega\mu)(\sigma + j\omega\epsilon)H &= 0 \end{aligned}\right\} \quad (5\text{-}86b)$$

A possible solution for any of these equations would be of the form

$$V,\ I,\ E, \text{ or } H = A\epsilon^{-\gamma x} + B\epsilon^{\gamma x} \tag{5-87}$$

where $$\gamma^2 = (R + j\omega L)(G + j\omega C)$$

* In writing eqs. (85) each term really should be multiplied by the factor $e^{j\omega t}$ to express the variation with time. However, it is customary—and quite correct—to cancel out or divide through by the common time factor ($e^{j\omega t}$), leaving an equation expressing relations between amplitudes. Thus in eqs. (85), V, I, E, and H, as written, are functions of x but not of time. In order to differentiate with respect to time it is first necessary to reinsert the time variation by multiplying through the $e^{j\omega t}$. After differentiating, the common time factor can again be dropped.

Because the symbols V, I, E and H in this case are not functions of time, it would be correct to use total derivatives with respect to x instead of partial derivatives. However eqs. (85) are still true when the symbols V, I, E, and H (without subscripts) represent functions of time, as for example $V = V_o\,e^{j\omega t}$. In this latter case partial derivatives would be required. Since the symbols without subscripts are used interchangeably to represent both instantaneous values and amplitudes of the quantities (for reasons discussed in chap. 6), the partial derivative signs will be used throughout.

for the eqs. (86a), and

$$\gamma^2 = (j\omega\mu)(\sigma + j\omega\epsilon)$$

for the eqs. (86b). When the variation with time is expressed by reinserting the factor $e^{j\omega t}$, the first term of expression (87) represents a wave traveling to the right and the second expression represents a wave traveling to the left.

An alternative solution to eqs. (86) is often used in transmission line theory. In this solution the exponentials are combined differently and the solution appears in terms of hyperbolic functions, and can be written

$$\left.\begin{aligned} V &= A_1 \cosh \gamma x + B_1 \sinh \gamma x \\ I &= A_2 \cosh \gamma x + B_2 \sinh \gamma x \end{aligned}\right\} \qquad (5\text{-}88a)$$

Let $V = V_R$, $I = I_R$ at $x = 0$ and $V = V_S$, $I = I_S$ at $x = x_1$

Substitute these in (88a) and find for the coefficients the values

$$A_1 = V_R,\ B_1 = -\sqrt{\frac{R + j\omega L}{G + j\omega C}}\, I_R$$

$$A_2 = I_R,\ B_2 = -\sqrt{\frac{G + j\omega C}{R + j\omega L}}\, V_R$$

$$\left.\begin{aligned} E &= A_1 \cosh \gamma x + B_1 \sinh \gamma x \\ H &= A_2 \cosh \gamma x + B_2 \sinh \gamma x \end{aligned}\right\} \qquad (5\text{-}88b)$$

Let $E = E_R$, $H = H_R$ at $x = 0$ and $E = E_S$, $H = H_S$ at $x = x_1$

Substitute these in (88b) and evaluate the coefficients.

$$A_1 = E_R,\ B_1 = -\sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}}\, H_R$$

$$A_2 = H_R,\ B_2 = -\sqrt{\frac{\sigma + j\omega\epsilon}{j\omega\mu}}\, E_R$$

In transmission line theory it is customary to write

$$Z = R + j\omega L \qquad Y = G + j\omega C$$

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

where Z_0 is called the *characteristic impedance* of the transmission line.

Similarly in wave theory it is customary to write

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}}$$

where η is called the *characteristic impedance* or *intrinsic impedance* of the medium.

In terms of these quantities, and writing $L = -x_1$ eqs. (88)

become*

$$\left.\begin{aligned} V_S &= V_R \cosh \gamma L \\ &\quad + Z_0 I_R \sinh \gamma L \\ I_S &= I_R \cosh \gamma L \\ &\quad + \frac{V_R}{Z_0} \sinh \gamma L \end{aligned}\right\} \quad (5\text{-}89a) \qquad \left.\begin{aligned} E_S &= E_R \cosh \gamma L \\ &\quad + \eta H_R \sinh \gamma L \\ H_S &= H_R \cosh \gamma L \\ &\quad + \frac{E_R}{\eta} \sinh \gamma L \end{aligned}\right\} \quad (5\text{-}89b)$$

When the line is very long (or the homogeneous medium very thick) so that γl is a large quantity,

$$\cosh \gamma l \approx \frac{e^{\gamma l}}{2} \approx \sinh \gamma l$$

and the ratios of voltage to current and E to H are seen from eqs. (89) to be

$$\frac{V_S}{I_S} = Z_0 \qquad \frac{E_S}{H_S} = \eta$$

The characteristic impedance Z_0, and intrinsic impedance η are, respectively, the ratios of V to I on a transmission line and E to H in a uniform plane wave under conditions where there is no reflected wave from the termination, or, in other words, when the wave along the line or in the medium is a *traveling* wave. Equations (89) are the general equations for the propagation of waves along uniform transmission lines or plane waves in homogeneous media. For the special cases of a "lossless" line or a "lossless" (nonconducting) medium the following simplifications occur

	$R = G = 0$ so that	
	$Z_0 = \sqrt{L/C}$	
	$\gamma = \sqrt{(j\omega L)(j\omega C)}$	
	$= j\omega\sqrt{LC}$	
but	$\gamma = \alpha + j\beta$	
Therefore	$\alpha = 0$	
	$\beta = \omega\sqrt{LC}$	

	$\sigma = 0$	so that
	$\eta = \sqrt{\mu/\epsilon}$	
	$\gamma = \sqrt{(j\omega\mu)(j\omega\epsilon)}$	
	$= j\omega\sqrt{\mu\epsilon}$	
but	$\gamma = \alpha + j\beta$	
Therefore	$\alpha = 0$	
	$\beta = \omega\sqrt{\mu\epsilon}$	

Under these circumstances, since $\cosh j\beta = \cos \beta$, and $\sinh j\beta = j \sin \beta$, the general eqs. (89) reduce to

$$\left.\begin{aligned} V_S &= V_R \cos \beta L + jZ_0 I_R \sin \beta L \\ I_S &= I_R \cos \beta L + j\frac{V_R}{Z_0} \sin \beta L \end{aligned}\right\} \quad (5\text{-}90a) \qquad \left.\begin{aligned} E_S &= E_R \cos \beta L + j\eta H_R \sin \beta L \\ H_S &= H_R \cos \beta L + j\frac{E_R}{\eta} \sin \beta L \end{aligned}\right\} \quad (5\text{-}90b)$$

* **The details are given in section 8.06.**

The quantities $1/\sqrt{LC}$ and $1/\sqrt{\mu\epsilon}$ have the dimensions of velocity and are in fact the velocities of wave propagation along the lossless line and in the lossless medium respectively. In either case, when the dielectric is air, so that $\mu = \mu_v$ and $\epsilon = \epsilon_v$,

$$\frac{1}{\sqrt{LC}} = c \approx 3 \times 10^8 \qquad \text{meter/sec}$$

$$\frac{1}{\sqrt{\mu_v \epsilon_v}} = c \approx 3 \times 10^8 \qquad \text{meter/sec}$$

It has been seen that propagation of a uniform plane wave in a homogeneous medium is analogous to propagation along a uniform

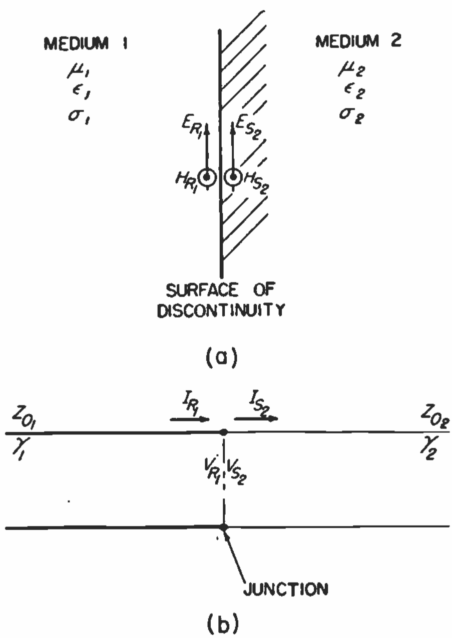


FIG. 5-11. Reflection and transmission (a) at a boundary surface between two media and (b) at a junction between two transmission lines.

transmission line. If the uniform plane wave passes abruptly from one medium to another, the surface of discontinuity being a plane perpendicular to the direction of propagation, the analogy will continue to hold. This is so because the boundary conditions at the surface of discontinuity are the same as those existing at the junction between two transmission lines having different constants. For the latter the continuity requirements are that (1) the voltage be continuous across the junction and (2) the current be continuous across the junction. These are the same requirements that hold for the analogous quantities E and H across a boundary surface. The usefulness of the above analogy may be shown best by means of examples.

EXAMPLE 1: *Reflection at the Surface of a Medium Having Arbitrary Constants.* A uniform plane electromagnetic wave, propagating in a medium which has constants μ, ϵ, and σ, impinges normally upon a second medium of infinite depth having the constants μ_2, ϵ_2, and σ_2 (Fig. 5-11a). Determine

(a) The amplitudes and phases of the reflected electric and magnetic intensities relative to the intensities of the incident wave.

(b) The amplitudes of the electric and magnetic intensities transmitted into the second medium.

For the transmission lines of Fig. 5-11b let V'_{R_1}, V''_{R_1} be the initial and reflected voltages on line (1) at the junction, and let I'_{R_1}, I''_{R_1} be the initial and reflected currents. It will be recalled that these voltages and currents are related by $V'_{R_1}/I'_{R_1} = Z_0$, $V''_{R_1}/I''_{R_1} = -Z_0$. Then, as in eqs. (60) and (62),

$$\frac{V''_{R_1}}{V'_{R_1}} = \frac{Z_R - Z_{0_1}}{Z_R + Z_{0_1}} = \frac{Z_{0_2} - Z_{0_1}}{Z_{0_2} + Z_{0_1}} \tag{5-91}$$

$$\frac{I''_{R_1}}{I'_{R_1}} = \frac{Z_{0_1} - Z_R}{Z_{0_1} + Z_R} = \frac{Z_{0_1} - Z_{0_2}}{Z_{0_1} + Z_{0_2}} \tag{5-92}$$

In the above equations Z_R is the terminating impedance for line (1). In this case line (1) is terminated by the input impedance to line (2), which is just the characteristic impedance Z_{0_2} of line (2) since it has been assumed that this second line is infinitely long. By analogy for Fig. 5-11a

$$\frac{E''_{R_1}}{E'_{R_1}} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{5-60}$$

$$\frac{H''_{R_1}}{H'_{R_1}} = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2} \tag{5-62}$$

where

$$\eta_1 = \sqrt{\frac{j\omega\mu_1}{\sigma_1 + j\omega\epsilon_1}}$$

$$\eta_2 = \sqrt{\frac{j\omega\mu_2}{\sigma_2 + j\omega\epsilon_2}}$$

and where, as in the transmission line case,

$$\frac{E'_R}{H'_{R_1}} = -\frac{E''_{R_1}}{H''_{R_1}} = \eta_1$$

Equations (60) and (62) give the reflected electric and magnetic field intensities in terms of the field intensities of the incident wave.

In Fig. 5-11b the voltage and current entering line (2) (the transmitted waves) are given by

$$V_{S_2} = V_{R_1} = V'_{R_1} + V''_{R_1}$$
$$I_{S_2} = I_{R_1} = I'_{R_1} + I''_{R_1}$$

In terms of the voltage and current of the initial or incident wave,

$$\frac{V_{S_2}}{V'_{R_1}} = 1 + \frac{V''_{R_1}}{V'_{R_1}} = \frac{2Z_{0_2}}{Z_{0_2} + Z_{0_1}} \tag{5-93}$$

$$\frac{I_{S_2}}{I'_R} = 1 + \frac{I''_{R_1}}{I'_{R_1}} = \frac{2Z_{0_1}}{Z_{0_2} + Z_{0_1}} \tag{5-94}$$

Similarly in Fig. 5-11a, the electric and magnetic intensities transmitted into the second medium are related to the E and H of the incident wave, by

$$\frac{E_{S_2}}{E'_{R_1}} = \frac{2\eta_2}{\eta_2 + \eta_1} \tag{5-61}$$

$$\frac{H_{S_2}}{H'_{R_1}} = \frac{2\eta_1}{\eta_2 + \eta_1} \tag{5-63}$$

It is interesting to evaluate expressions (60), (62), (61), and (63) for the case of an electromagnetic wave in air incident normally upon a copper sheet. A frequency of 1 mc will be assumed. For this example

$$\mu_1 = \mu_v \qquad \mu_2 = \mu_v$$
$$\epsilon_1 = \epsilon_v \qquad \epsilon_2 = \epsilon_v$$
$$\sigma_1 = 0 \qquad \sigma_2 = 5.8 \times 10^7 \text{ mhos/m}$$

so that

$$\eta_1 = \sqrt{\frac{\mu_v}{\epsilon_v}} = 377 \text{ ohms}$$

$$\eta_2 = \sqrt{\frac{j2\pi \times 10^6 \times 4\pi \times 10^{-7}}{5.8 \times 10^7 + j2\pi \times 10^6 \times 8.854 \times 10^{-12}}} = 0.000369 \underline{/45^\circ} \text{ ohms}$$

Then the ratio of reflected to incident electric intensities, as given by eq. (60), is

$$\frac{E''_{R_1}}{E'_{R_1}} = \frac{3.69 \times 10^{-4} \underline{/45^\circ} - 377}{3.69 \times 10^{-4} \underline{/45^\circ} + 377}$$
$$= -0.9999986 \underline{/-0.000079^\circ}$$

Similarly $$\frac{H''_{R_1}}{H'_{R_1}} = +0.9999986 \underline{/-0.000079^\circ}$$

It is seen that differences between these reflection coefficients for copper and the coefficients of minus and plus unity, which would be obtained for a *perfect* reflector, are indeed negligible. For most practical purposes, copper can be considered a perfect reflector of radio waves.

The relative strengths of the transmitted intensities for this case are

$$\frac{E_{S_2}}{E'_{R_1}} = \frac{7.38 \times 10^{-4} \underline{/45^\circ}}{3.69 \times 10^{-4} \underline{/45^\circ} + 377} = 0.00000196 \underline{/45^\circ}$$

$$\frac{H_{S_2}}{H'_{R_1}} = \frac{2 \times 377}{377 + 3.69 \times 10^{-4} \underline{/45^\circ}} = 1.9999986 \underline{/-0.00004^\circ}$$

The electric intensity just inside the metal is approximately 2×10^{-6} times that of the initial wave; the magnetic intensity just inside the metal is approximately twice the magnetic intensity of the initial wave. This last result could be inferred from the fact that, since the magnetic intensity is reflected without phase reversal, the total magnetic intensity just outside the surface of the copper is approximately double that of the initial wave and therefore, because of continuity requirements, H just inside the copper is also approximately twice the magnetic intensity of the incident wave. The ratio of E to H just inside the metal is equal to η_2, the characteristic impedance of the copper. That is

$$\frac{E_{S_2}}{H_{S_2}} = \eta_2 = 0.000369 \underline{/45^\circ} \text{ ohms}$$

For many practical purposes this is sufficiently close to zero to consider the copper sheet to be a zero-impedance surface.

5.13 Surface Impedance. It has been seen that at high frequencies the current is confined almost entirely to a very thin sheet at the surface of the conductor. In many applications it is convenient to make use of a *surface impedance* defined by

$$Z_S = \frac{E_{\tan}}{J} \tag{5-95}$$

where $E_{\tan}$ is the electric intensity parallel to, and at the surface of, the conductor and J is the linear current density that flows as a result of this $E_{\tan}$. The linear current density J represents the total conduction current per meter width flowing in the thin sheet. If it is assumed that the conductor is a flat plate with its surface at the $y = 0$ plane (Fig. 5-12), the current distribution in the y direction will be given by

$$i = i_0 e^{-\gamma y}$$

where i_0 is the current density at the surface.

It is assumed that the thickness of the conductor is very much greater than the depth of penetration, so that there is no reflection

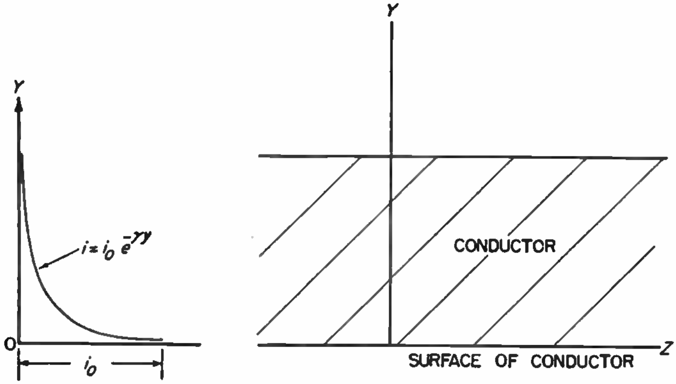


FIG. 5-12. Current distribution in a thick flat-plate conductor.

from the back surface of the conductor. The total conduction current per meter width, that is, the linear current density is

$$\begin{aligned} J &= \int_0^\infty i\,dy = i_0 \int_0^\infty e^{-\gamma y}\,dy \\ &= -\frac{i_0}{\gamma}\,[e^{-\gamma y}]_0^\infty = \frac{i_0}{\gamma} \end{aligned} \tag{5-96}$$

But i_0, the current density at the surface, is

$$i_0 = \sigma E$$

Therefore
$$Z_S = \frac{E}{J} = \frac{\gamma}{\sigma}$$

The constant γ for propagation in a conducting medium was found to be

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} \approx \sqrt{j\omega\mu\sigma}$$

This gives for a thick conductor

$$Z_S = \sqrt{\frac{j\omega\mu}{\sigma}} = \eta \text{ (for the conducting medium)} \tag{5-97}$$

It is seen that *the surface impedance of a plane conductor that is very much thicker than the skin depth is just equal to the characteristic impedance of the conductor.* This is also the input impedance of the conductor when viewed as a transmission line conducting energy into the interior of the metal. When the thickness of the plane conductor is not great compared with the depth of penetration, reflection of the wave occurs at the back surface of the conductor. Under these conditions, the input impedance is approximately equal to the input impedance of a lossy line terminated in an open circuit, viz.,

$$Z_{\text{in}} = \eta \coth \gamma l \tag{5-98}$$

where l is the thickness of the conductor, and η and γ are its intrinsic impedance and propagation constant respectively. The approximation is ordinarily valid because the actual termination $\eta_v = 377$ ohms is very much greater than η of the conductor.

Surface Impedance of Good Conductors. For any material normally classed as a good conductor $\sigma \gg \omega\epsilon$, and if the conductor thickness is very much greater than the depth of penetration, the surface impedance of such a conductor is

$$Z_s \approx \sqrt{\frac{j\omega\mu}{\sigma}} = \sqrt{\frac{\omega\mu}{\sigma}}\ \underline{/45^\circ} \tag{5-97}$$

The *surface resistance* is

$$R_s \approx \sqrt{\frac{\omega\mu}{2\sigma}} \tag{5-97a}$$

and the surface reactance has the same magnitude as R_s at all frequencies

$$X_s \approx \sqrt{\frac{\omega\mu}{2\sigma}} \tag{5-97b}$$

The surface resistance defined by (97a) as the real part of the surface impedance is the high-frequency or *skin-effect resistance* per unit length of a flat conductor of unit width. (It has the dimension of ohms and its value does not depend upon the units used to measure length and width as long as they are the same.) Recalling that the expression for depth of penetration in a conductor is

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} \tag{5-99}$$

it is seen that

$$R_s = \frac{1}{\sigma\delta} \tag{5-100}$$

The surface resistance of a flat conductor at any frequency is equal to the d-c resistance of a thickness δ of the same conductor, where δ is the depth of penetration or *skin depth*. This means that the conductor, having a thickness very much greater than δ and having the exponential current distribution throughout its depth, has the same resistance as would a thickness δ of the conductor with the current distributed uniformly throughout this thickness. From this it follows that the power loss per unit area of the plane conductor will be given by $J_{eff}^2 R_s$, where R_s is its surface resistance and J_{eff} is the linear current density or current per meter width (effective value) flowing in the conductor. This same conclusion can be obtained from consideration of power flow, a subject that will be taken up in the next chapter.

ADDITIONAL PROBLEMS

8. From the boundary conditions that E_{tan} and H_{tan} are continuous, derive the reflection and transmission coefficients for a uniform plane wave incident normally on the boundary surface between any two media; i.e.,

$$\frac{E''}{E'} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \qquad \frac{H''}{H'} = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}$$

$$\frac{E_2}{E'} = \frac{2\eta_2}{\eta_1 + \eta_2} \qquad \frac{H_2}{H'} = \frac{2\eta_1}{\eta_1 + \eta_2}$$

E', E'', and E_2 represent the electric intensity of the incident, reflected and transmitted waves respectively, with a similar notation for the magnetic intensity.

9. The electric intensity of a uniform plane electromagnetic wave in free space is 1 volt per meter, and the frequency is 300 mc. If a very large thick flat copper plate is placed normal to the direction of wave propagation, determine (a) the electric intensity at the surface of the plate; (b) the magnetic intensity at the surface of the plate; (c) the depth of penetration; (d) the conduction current density at the surface; (e) the conduction current density at a distance of 0.01 mm below the surface; (f) the linear current density J; (g) the surface impedance; (h) the power loss per square meter of surface area. For copper use $\sigma = 5.8 \times 10^7$, $\epsilon \approx \epsilon_v$, $\mu \approx \mu_v$.

10. A uniform plane electromagnetic wave is incident normally upon a sheet of dielectric material, which has the following constants: $\epsilon = 4\epsilon_v$, $\mu = \mu_v$, $\sigma = 10^{-6}$ mhos per meter. If the sheet is 2 cm thick and the amplitude of the electric intensity of the incident wave is 100 mv/m, determine the electric intensity of the wave after passing through the sheet (a) if the frequency is 3000 mc; (b) if the frequency is 30 mc. (NOTE:

It is assumed that σ and ϵ are independent of frequency. In general, for the so-called dielectric materials, this is not true.)

11. Determine the reflection coefficients for an electromagnetic wave incident normally on (a) a sheet of copper; (b) a sheet of iron. Use $f = 1$ mc. Assume $\sigma = 1 \times 10^6$ mhos/m, $\mu = 1000\mu_v$ for the iron.

12. In the analogy between plane wave propagation in a homogeneous conducting medium and wave propagation along transmission lines, there appears to be nothing corresponding to R. Discuss. [Suggestion: Compare eqs. (5.85b).]

13. A thick brass plate is plated with a 0.0005 inch thickness of silver. What is the surface impedance at (a) 10 kc, (b) 1 mc, (c) 100 mc? Compare the surface impedance of the plated brass with that of a solid silver plate and a solid brass plate. (For silver $\sigma = 6.2 \times 10^7$; for brass $\sigma = 1 \times 10^7$; for both assume that $\mu = \mu_v$, $\epsilon = \epsilon_v$.)

14. A sheet of glass, having a relative dielectric constant of 8 and negligible conductivity, is coated with a silver plate. Show that at a frequency of 100 mc the surface impedance will be *less* for a 0.001 cm coating than it is for a 0.002 cm coating, and explain why.

15. Determine the voltmeter reading by two different methods. Assume that all the conductors are perfect and that the coaxial cable is lossless.

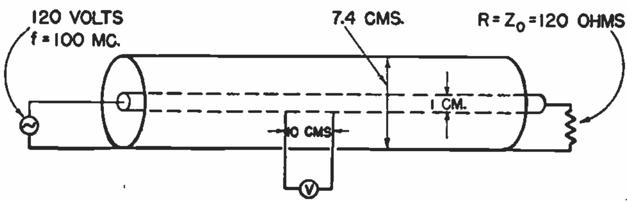


FIG. 5-13. Coaxial cable for Problem 15.

16. "Free-space cloth" consists of a cloth coated with conducting material that has a surface impedance of 377 ohms per square. Show that if the thickness of the coating is much greater than the depth of penetration, the surface impedance will be complex, with a reactance equal to the resistance (assuming $\sigma/\omega\epsilon \gg 1$ for the conducting material). However, if the coating is made sufficiently thin, show that the surface impedance will be almost a pure resistance. Determine appropriate values for σ and l, where l is the thickness of the coating.

BIBLIOGRAPHY

See bibliography for chap. 4.

CHAPTER 6

POYNTING VECTOR AND THE FLOW OF POWER

6.01 Poynting's Theorem. As electromagnetic waves propagate through space from their source to distant receiving points, there is a transfer of energy from the source to the receivers. There exists a simple and direct relation between the rate of this energy transfer and the amplitudes of electric and magnetic intensities of the electromagnetic wave. This relation can be obtained from Maxwell's equations as follows.

The magnetomotive force equation I can be written

$$i = \text{curl } \mathbf{H} - \epsilon\dot{\mathbf{E}} \qquad \text{I}$$

This expresses a relation between quantities which have the dimensions of current density. If it is multiplied through by **E**, there will result a relation between quantities which will have the dimensions of power per unit volume. That is

$$\mathbf{E} \cdot i = \mathbf{E} \cdot \text{curl } \mathbf{H} - \epsilon\mathbf{E} \cdot \dot{\mathbf{E}} \qquad (6\text{-}1)$$

Recall that for any vectors the following identity holds

$$\text{div } \mathbf{E} \times \mathbf{H} = \mathbf{H} \cdot \text{curl } \mathbf{E} - \mathbf{E} \cdot \text{curl } \mathbf{H}$$

Therefore

$$\mathbf{E} \cdot i = \mathbf{H} \cdot \text{curl } \mathbf{E} - \text{div } \mathbf{E} \times \mathbf{H} - \epsilon\mathbf{E} \cdot \dot{\mathbf{E}} \qquad (6\text{-}2)$$

Introducing the second field equation,

$$\text{curl } \mathbf{E} = -\mu\dot{\mathbf{H}} \qquad \text{II}$$

obtain

$$\mathbf{E} \cdot i = -\mu\mathbf{H} \cdot \dot{\mathbf{H}} - \epsilon\mathbf{E} \cdot \dot{\mathbf{E}} - \text{div } \mathbf{E} \times \mathbf{H} \qquad (6\text{-}3)$$

Now

$$\mathbf{H} \cdot \dot{\mathbf{H}} = \frac{1}{2}\frac{\partial}{\partial t}H^2 \qquad \text{and} \qquad \mathbf{E} \cdot \dot{\mathbf{E}} = \frac{1}{2}\frac{\partial}{\partial t}E^2$$

so that

$$\mathbf{E} \cdot \boldsymbol{i} = -\frac{\mu}{2}\frac{\partial}{\partial t} H^2 - \frac{\epsilon}{2}\frac{\partial}{\partial t} E^2 - \operatorname{div} \mathbf{E} \times \mathbf{H}$$

Integrating over a volume V,

$$\int_{\text{vol}} \mathbf{E} \cdot \boldsymbol{i}\, dV = -\frac{\partial}{\partial t}\int_{\text{vol}} \left(\frac{\mu}{2} H^2 + \frac{\epsilon}{2} E^2\right) dV - \int_{\text{vol}} \operatorname{div} \mathbf{E} \times \mathbf{H}\, dV \tag{6-4}$$

Using the divergence theorem the last term can be changed from a volume integral to a surface integral, that is,

$$\int_{\text{vol}} \operatorname{div} \mathbf{E} \times \mathbf{H}\, dV = \oint_S \mathbf{E} \times \mathbf{H} \cdot d\mathbf{a}$$

Then eq. (4) can be written

$$\int_{\text{vol}} \mathbf{E} \cdot \boldsymbol{i}\, dV = -\frac{\partial}{\partial t}\int_{\text{vol}} \left(\frac{\mu}{2} H^2 + \frac{\epsilon}{2} E^2\right) dV - \oint_S \mathbf{E} \times \mathbf{H} \cdot d\mathbf{a} \tag{6-5}$$

A physical interpretation of eq. (5) leads to some interesting conclusions. It will be considered term by term.

The term on the left-hand side represents (instantaneous) power dissipated in the volume V. This result is obtained as a generalization of Joule's law. A conductor of cross-sectional area A, carrying a current I and having a voltage drop E per unit length will have a power loss of EI watts per unit length. The power dissipated per unit volume would be

$$\frac{EI}{A} = Ei \text{ watts per unit volume}$$

In this case $\mathbf{E}$ and $\boldsymbol{i}$ are in the same direction. In general, where this may not be true, the power dissipated per unit volume would still be given by the product of $\boldsymbol{i}$ and the component of $\mathbf{E}$ having the same direction as $\boldsymbol{i}$. That is, the power dissipated per unit volume would always be given by

$$\mathbf{E} \cdot \boldsymbol{i}$$

and the total power dissipated in a volume V would be

$$\int_{\text{vol}} \mathbf{E} \cdot \boldsymbol{i}\, dV \tag{6-6}$$

When the **E** in this expression represents the electric intensity required to produce the current density i in the conducting medium, the expression (6) represents power dissipated as ohmic (I^2R) loss. However, if the **E** is an electric intensity due to a source of power, for example due to a battery, then the power represented by the integral expression (6) would be used up in driving the current against the battery voltage and hence charging the battery. If the direction of **E** were opposite to that of i, the "dissipated" power represented by (6) would be negative. In this case, the battery would be generating electric power.

Consider next the first term on the right-hand side of eq. (5). In the electrostatic field it was found that the quantity $\frac{1}{2}\epsilon E^2$ could be considered to represent the *energy density* or the stored electric energy per unit volume of the electric field. Also for the steady magnetic field the quantity $\frac{1}{2}\mu H^2$ represented the stored energy density of the magnetic field. If it is assumed that these quantities continue to represent stored energy densities when the fields are changing with time (and there seems to be no real reason for considering otherwise), the integral represents the total stored energy in the volume V. The negative time derivative of this quantity then represents the rate at which the stored energy in the volume is decreasing.

The interpretation of the remaining term follows from the application of the law of conservation of energy. The rate of energy dissipation in the volume V must equal the rate at which the stored energy in V is decreasing, plus the rate at which energy is entering the volume V from outside. The term

$$-\oint_S \mathbf{E} \times \mathbf{H} \cdot da$$

therefore must represent the rate of flow of energy inward through the surface of the volume. Then this expression without the negative sign,

$$\oint_S \mathbf{E} \times \mathbf{H} \cdot da \qquad (6\text{-}7)$$

represents rate of flow of energy *outward* through the surface enclosing the volume.

The interpretation of eq. (5) leads to the conclusion that the integral of $\mathbf{E} \times \mathbf{H}$ over any closed surface gives the rate of energy flow through that surface. It is seen that the vector

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \tag{6-8}$$

has the dimensions of watts per square meter. It is *Poynting's theorem* that the vector product $\mathbf{P} = \mathbf{E} \times \mathbf{H}$ at any point is a measure of the rate of energy flow per unit area at that point. The direction of flow is perpendicular to $\mathbf{E}$ and $\mathbf{H}$ in the direction of the vector $\mathbf{E} \times \mathbf{H}$.

EXAMPLE 1: *Power Flow for a Plane Wave.* The expression for rate of energy flow per unit area is checked very easily in the case of a uniform plane wave traveling with a velocity

$$v_0 = \frac{1}{\sqrt{\mu\epsilon}}$$

The total energy density due to electric and magnetic fields is given by

$$\tfrac{1}{2}(\epsilon E^2 + \mu H^2)$$

For a wave moving with a velocity $\mathbf{v}_0$ the rate of flow of energy per unit area would be

$$\mathbf{P} = \tfrac{1}{2}(\epsilon E^2 + \mu H^2)\mathbf{v}_0 \tag{6-9}$$

Recalling that for a plane wave the magnitudes of E and H are related by

$$\frac{E}{H} = \sqrt{\frac{\mu}{\epsilon}}$$

eq. (9) becomes

$$\mathbf{P} = \frac{1}{2}\left(\epsilon\sqrt{\frac{\mu}{\epsilon}}\,EH + \mu\sqrt{\frac{\epsilon}{\mu}}\,EH\right)\mathbf{v}_0$$

$$= \left(\frac{EH}{v_0}\right)\mathbf{v}_0$$

$$= \mathbf{E} \times \mathbf{H}$$

EXAMPLE 2: *Power Flow in a Concentric Cable.* Consider the transfer of power to a load resistance R along a concentric cable which has a d-c voltage V between conductors and a steady current I flowing in the inner and outer conductors. The conductors are assumed to have negligible resistance. The radius of the inner conductor is a and the (inside) radius of the outer conductor is b. The magnetic intensity H will be directed in circles about the axis. By Ampere's law the magnetomotive force around any of these circles will be equ..l to the current enclosed, that is,

$$\oint \mathbf{H} \cdot d\mathbf{s} = \mathbf{I}$$

in the region between the conductors.

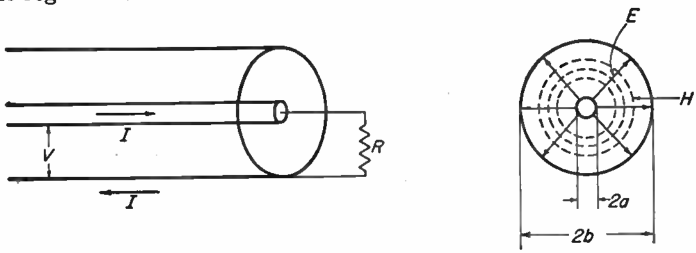


FIG. 6-1

For this case H is constant along any of the circular paths so

$$\oint \mathbf{H} \cdot d\mathbf{s} = 2\pi r H$$

where r is the radius of the circle being considered.
Hence

$$H = \frac{I}{2\pi r}$$

The electric intensity **E** will be directed radially. In the example on page 55 it was shown that

$$V = \frac{q}{2\pi\epsilon} \log \frac{b}{a}$$

where q was the charge per unit length. Also it was shown that

$$E = \frac{q}{2\pi\epsilon r}$$

Therefore the magnitude of E will be given by

$$E = \frac{V}{r \log \frac{b}{a}}$$

The Poynting vector is $\mathbf{P} = \mathbf{E} \times \mathbf{H}$

It is directed parallel to the axis of the cable. Since **E** and **H** are everywhere at right angles, the magnitude of **P** is simply

$$P = EH$$

The total power flow along the cable will be given by the integration of the Poynting vector over any cross-sectional surface. If the conductors are considered to be perfect, E will have value only in the region between them and the Poynting vector will have value only in the same region. Let the element of area be $2\pi r dr$. Then

$$\begin{aligned} W &= \int_S \mathbf{E} \times \mathbf{H} \cdot da \\ &= \int_a^b \frac{V}{r \log b/a} \left(\frac{I}{2\pi r}\right) 2\pi r\, dr \\ &= \frac{VI}{\log b/a} \int_a^b \frac{dr}{r} \\ &= VI \end{aligned}$$

This is the well-known result that the power flow along the cable is the product of the voltage and current. It is interesting to observe that this result was obtained by an integration over an area that did not include the conductors. According to this picture, for the perfect conductor case the flow of power is entirely external to the conductors. Even when the conductors have resistance, there is no contribution within the conductors to the Poynting vector in the direction parallel to the axis, for there is no value of E within a conductor at right angles to the direction of current flow. In the case of the open-wire transmission line, the fields extend throughout all space and there is a value of Poynting vector everywhere in space, except within conducting bodies. Therefore the rather remarkable conclusion is reached that when a transmission line is used to deliver power from a generator to a load, the power transmission takes place through all the nonconducting regions of space and none of the power flows through the conductors that make up the transmission line.

Example 3: *Conductor Having Resistance.* When a conductor having resistance carries a direct current I, there will be a value of E within the conductor. It will be parallel to the direction of the current ($\mathbf{E} = \mathbf{i}/\sigma$), so there will still be no radial component of $\mathbf{E}$. Hence there will still be no value of Poynting vector within the wire parallel to the axis, but there will now be a radial component of $\mathbf{P}$. Consider a wire of length L having a voltage drop V_L along the wire. Let the wire be parallel to the z axis. Then in the wire and at its surface

$$E_z = \frac{V_L}{L}$$

The magnetic intensity $\mathbf{H}$ will be in the φ direction and at the surface of the wire it will have a value

$$H_\varphi = \frac{I}{2\pi a}$$

where a is the radius of the wire. E_z and H_φ are at right angles, so the Poynting vector will have a magnitude

$$P = E_z H_\varphi$$

and will be directed radially into the wire. The total power flowing into the wire through the surface will be

$$\begin{aligned} W &= \int_0^L E_z H_\varphi 2\pi a \, dz \\ &= \frac{V_L I}{L} \int_0^L dz \\ &= V_L I \end{aligned}$$

which is the usual expression for loss due to ohmic resistance. This derivation shows that the power required to supply this loss may be considered as coming from the field outside the wire, entering it through the surface of the wire.

It is interesting to observe how the power flow continues inward. Inside the wire the value of H does not vary with the radius in the same way as outside, because the current enclosed varies with r in this case. If i is the current density, the current enclosed at a radius r will be

$$I_{\text{enc}} = \pi r^2 i$$

For a wire of radius a having a total current I

$$I_{\text{enc}} = \frac{\pi r^2 I}{\pi a^2} = \frac{r^2}{a^2} I$$

Therefore inside the wire ($r < a$)

$$H = \frac{r^2 I}{2\pi r a^2}$$

The power flowing inward through an imaginary cylindrical shell of radius $r < a$ will be

$$\begin{aligned} W &= \frac{V_L}{L} 2\pi r L H \\ &= V_L I \frac{r^2}{a^2} \end{aligned} \qquad (6\text{-}10)$$

Equation (10) shows that the power dissipated within any shell is proportional to the volume enclosed by the shell through which the power is flowing. Hence the power dissipated per unit volume is uniform throughout the wire.

The configuration of the electric field about a two-wire line will appear somewhat as illustrated in Fig. 6-2 when there is a resistance drop in the conductors. The curvature near the surface of the wire is due to the voltage drop along the wire.

EXAMPLE 4: *Poynting Vector about A-C Lines.* When a transmission line delivers a-c power, the voltage, and therefore the electric and magnetic fields, vary with time. Also, if it is a long line, the phases of voltage and current (and E and H) will vary along the length of the line. For the simple case of a lossless line terminated in its characteristic impedance

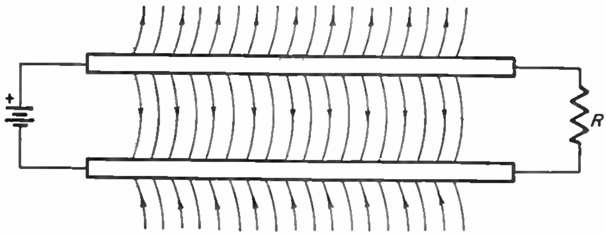


FIG. 6-2. Electric-field configuration about a two-wire transmission line which has resistance.

which is a pure resistance, the variation in time and along the line of both voltage and current will be given by the expression for a traveling wave, that is, they are proportional to

$$\cos \omega \left(t - \frac{z}{v}\right)$$

For any value of z and t there will be a certain distribution of the Poynting vector over a plane parallel to the x and y axes. At every point in this plane, **P** will be parallel to the z axis. The Poynting vector will be given by an expression of the form

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} = A \cos^2 \omega \left(t - \frac{z}{v}\right) f(x, y)$$

The function $f(x, y)$ will not vary with z or t. For a fixed value of time, the total power passing through a plane will vary with the position of the plane, that is with z, whereas for a fixed value of z the power through the plane will vary with time. It will be noted that the power flow past a given plane is in pulses of double frequency, a fact readily appreciated when observing the flicker of a 25-cycle electric light bulb.

In a polyphase line a study of the Poynting vector shows that the power passing through a plane of fixed z will not vary as a function of time. In this case the Poynting vector distribution spirals about the line as it is propagated forward. The value of P integrated over a plane of constant z will be found to be independent of time. In such a plane where z = constant, the distribution of the Poynting vector would appear to be revolving about the line.

6.02 Note on the Interpretation of E × H. The interpretation of **E** × **H** as the power flow per unit area is an extremely useful concept, especially in radiation problems. For example, an integration of **E** × **H** over a surface enclosing a transmitting antenna gives the power radiated by the antenna. Although this interpretation of **E** × **H** never gives an answer which is known to be erroneous, it sometimes leads to a picture which the engineer is loathe to accept. Most engineers find acceptable the concept of energy transmission through space, either with or without guiding conductors, when wave motion is present. However for many engineers this picture becomes disturbing for transmission line propagation in the dc case. When **E** and **H** are static fields produced by unrelated sources, the picture becomes even less credible. The classic illustration of a bar magnet on which is placed an electric charge is one which is often cited. In this example a static electric field is crossed with a steady magnetic field and a strict interpretation of Poynting's theorem seems to require a continuous circulation of energy around the magnet. This is a picture that the engineer generally is not willing to accept (although he usually does not question the theory of permanent magnetism, which requires a continuous circulation of electric currents within the magnet). Fortunately, there exists an easy way out of the dilemma posed by this last example.

First, it is observed that the surface integral in eq. (5) is over the *closed* surface surrounding the volume. If any closed surface is taken about the bar magnet, it is found that **E** × **H** integrated over this closed surface is always zero. In other words, the *net* power flow away from the magnet is zero as it should be. Secondly, it is noted that, even though the power flow through any closed surface is correctly given by eq. (7), it does not necessarily follow that **P** = **E** × **H** represents correctly the power flow at each point. For, to the vector **E** × **H**, could be added any other vector having zero divergence (that is, any vector that is the curl of another vector) without changing the value of the integral in (7). This can be shown by applying the divergence theorem. Suppose the correct value for power flow at any point is not **E** × **H**, but rather **P** = **E**

$\times$ **H** + **F**, where **F** is the curl of some other vector, say **G**. Then the net power flow through any closed surface would be

$$\oint_S (\mathbf{E} \times \mathbf{H} + \mathbf{F}) \cdot da = \oint_S (\mathbf{E} \times \mathbf{H}) \cdot da + \int_{\text{vol}} \text{div } \mathbf{F} \, dv$$
$$= \oint_S (\mathbf{E} \times \mathbf{H}) \cdot da$$

because div curl $\mathbf{G} \equiv 0$.

It is seen that even though it may be possible to write an expression that gives correctly the net flow of power through a closed surface, it is still not possible to state just where the energy is. This problem is by no means peculiar to the electromagnetic field. The total potential energy of a raised weight is a readily calculable quantity but the "distribution" of this energy is not known. Just where the potential energy of a raised weight or a charged body "resides" is a question for philosophic speculation only. It cannot be answered on the basis of any measurements that the engineer can make.

6.03 Instantaneous, Average, and Complex Poynting Vector. In an ac circuit, the instantaneous power is always given by the product of the instantaneous voltage and the instantaneous current.

$$W_{\text{inst}} = V_{\text{inst}} I_{\text{inst}}$$

The real power or average over a cycle is

$$W_{\text{av}} = VI \cos \theta \tag{6-11}$$

where θ is the time-phase angle between voltage and current, and V and I are *effective* values. The reactive power, or reactive volt-amperes, is

$$W_{\text{react}} = VI \sin \theta \tag{6-12}$$

When the voltage and current are written in the *complex* form,† that is,

$$V = V_{\text{re}} + jV_{\text{im}}$$
$$I = I_{\text{re}} + jI_{\text{im}}$$

† A familiarity with complex notation as used in the solution of a-c circuits is assumed. In accordance with the 1948 IRE Standards on Symbols, when it is necessary or desirable to differentiate between real and complex scalars, the complex scalars (or phasors as they are now known) will be shown in bold italic type as $\boldsymbol{V}$. For complex vectors bold roman type is used, the same as for real vectors.

it is easily shown, using Fig. 6-3, eqs. (11) and (12), and a little trigonometry, that the real or average power is given by

$$W_{av} = V_{re}I_{re} + V_{im}I_{im} \tag{6-13}$$

while the reactive power is

$$W_{react} = V_{im}I_{re} - V_{re}I_{im} \tag{6-14}$$

This is a result well known to all electrical engineers. By multi-

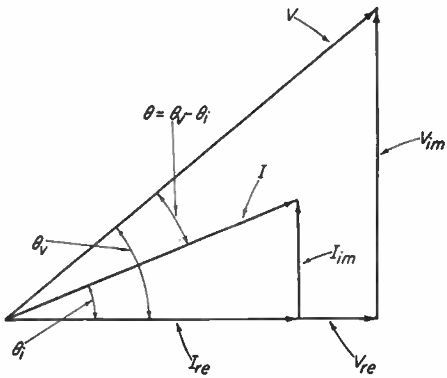


FIG. 6-3. Phasor diagram for voltage and current.

plying out it is seen that the real power of eq. (13) is given by the real part of the product.

$$VI^*,$$

where I^* is the complex conjugate of I. That is,

$$I^* = I_{re} - jI_{im}$$

Also the reactive power is given by the imaginary part of VI^*. Therefore it is possible to write†

$$W_{real} = Re\ VI^* \tag{6-15}$$
$$W_{react} = Im\ VI^* \tag{6-16}$$
$$W_{complex} = W_{real} + jW_{react} = VI^*$$

where Re and Im indicate that only the real or only the imaginary part is to be used.

† The student is reminded that the real and imaginary parts of the product VI (where V and I are in the complex form) do *not* represent real and reactive power.

In expressions (15) and (16), V and I represent effective values of (complex) voltage and current. In terms of peak values (maximum in time), which for sinusoidal variations are $\sqrt{2}$ times the effective values, eqs. (15) and (16) would be

$$W_{\text{real}} = \tfrac{1}{2}\, Re\, V_0 I_0^* \tag{6-17}$$
$$W_{\text{react}} = \tfrac{1}{2}\, Im\, V_0 I_0^* \tag{6-18}$$

In electromagnetic field theory there are relations similar to the above between the Poynting vector **P** (watts/sq m) and **E** (volts/m) and **H** (amp/m). The instantaneous power flow per square meter is

$$\mathbf{P}_{\text{inst}} = \mathbf{E}_{\text{inst}} \times \mathbf{H}_{\text{inst}}$$

The real and reactive power per square meter is†

$$\mathbf{P}_{\text{real}} = Re\,(\mathbf{E}_{\text{eff}} \times \mathbf{H}_{\text{eff}}^*)$$
$$\mathbf{P}_{\text{react}} = Im\,(\mathbf{E}_{\text{eff}} \times \mathbf{H}_{\text{eff}}^*)$$
$$\mathbf{P}_{\text{complex}} = \mathbf{E}_{\text{eff}} \times \mathbf{H}_{\text{eff}}^*$$

where **E** and **H** are expressed in the complex form and are effective values. In field theory peak values, rather than effective values, are used most, so it is usual to let the symbols **E** and **H** (without subscripts) represent peak or maximum values in time. Then

$$\mathbf{P}_{\text{real}} = \tfrac{1}{2}\, Re\,(\mathbf{E} \times \mathbf{H}^*) \tag{6-19}$$
$$\mathbf{P}_{\text{react}} = \tfrac{1}{2}\, Im\,(\mathbf{E} \times \mathbf{H}^*) \tag{6-20}$$
$$\mathbf{P}_{\text{complex}} = \tfrac{1}{2}\, \mathbf{E} \times \mathbf{H}^*$$

where **E** and **H** are now peak values in the complex form. The first of these expressions represents the average or real power flow per unit area. The second represents a flow of reactive power, a surging back and forth of the energy in the field.

The product of **E** and **H** in equations (19) and (20) is a vector product. Only mutually perpendicular components of **E** and **H** contribute anything to power flow, and the direction of the flow is normal to the plane containing **E** and **H**. Thus in rectangular co-ordinates, the complex flow of power per unit area normal to the y-z plane is

$$P_x = \tfrac{1}{2}(E_y H_z^* - E_z H_y^*) \tag{6-21}$$

† For *vectors* the form corresponding to expression (11), i.e., $\mathbf{P}_{\text{real}} = \mathbf{E} \times \mathbf{H} \cos\theta$, would require special interpretation, because the time phase angles between the scalar components may be different for the different components.

with corresponding expressions for the other directions. In spherical co-ordinates, the outward (radial) flow of complex power per unit area is

$$P_r = \tfrac{1}{2}(E_\theta H_\phi^* - E_\phi H_\theta^*) \tag{6-22}$$

Problem 1. Verify that

$$VI \cos \theta = V_{re}I_{re} + V_{im}I_{im} = Re\, VI^*$$

and

$$VI \sin \theta = V_{im}I_{re} - V_{re}I_{im} = Im\, VI^*$$

where the symbols represent effective values of voltage and current.

Problem 2. A concentric cable (assumed perfectly conducting) is one wavelength long and is terminated in its characteristic impedance, a pure resistance.

(a) Indicate the magnitude and direction of the Poynting vector along the line at successive one-eighth period intervals of time throughout a cycle.

(b) Repeat part (a) for the case where the line is terminated by a short circuit.

Problem 3. A short vertical transmitting antenna erected on the surface of a perfectly conducting earth produces an effective field intensity

$$E_{\text{eff}} = E_{\theta\text{eff}} = 100 \sin \theta \qquad \text{mv/m}$$

at points a distance of 1 mile from the antenna (θ is the polar angle). Compute the Poynting vector and the total power radiated. (For the *distant* field, $H = H_\phi = \dfrac{E_\theta}{\eta_v}$.)

6.04 Power Loss in a Plane Conductor. An evaluation of the normal component of Poynting vector at the surface of a conductor will give the power flow per unit area through the surface and hence the power loss in the conductor.

Let there be a tangential component of magnetic intensity $\mathbf{H}_{\tan}$ at the surface of a metallic conductor (assumed for the present to be an infinitely large flat plate having a thickness very much greater than the skin depth δ). From the continuity requirements across the boundary surface the tangential component of **H** just inside the conductor will have this same value $\mathbf{H}_{\tan}$. Inside the conductor $\mathbf{E}_{\tan}$, the tangential component of **E** is related to $\mathbf{H}_{\tan}$ by

$$\frac{E_{\tan}}{H_{\tan}} = \eta_m$$

and

$$\eta_m = \sqrt{\frac{j\omega\mu_m}{\sigma_m + j\omega\epsilon_m}} \approx \sqrt{\frac{j\omega\mu_m}{\sigma_m}} = \sqrt{\frac{\omega\mu_m}{\sigma_m}} \quad \underline{/45°}$$

where n_m is the intrinsic impedance of the conductor. (The subscript m has been used to indicate that the quantities inside the metallic conductor are meant.) Just inside the surface of the conductor $E_{tan} = n_m H_{tan}$ and, from the continuity requirements across the boundary, the tangential component of electric intensity just outside the surface will also be E_{tan}. Then the average (or real) power flow per unit area normal to the surface will be

$$\mathbf{P}_n \text{ (real)} = \tfrac{1}{2} Re\, (\mathbf{E}_{tan} \times \mathbf{H}_{tan}^*) \tag{6-23}$$

When $\mathbf{E}_{tan}$ and $\mathbf{H}_{tan}$ are at right angles, and since for any good conductor E_{tan} leads H_{tan} by 45 degrees in time phase, (23) becomes

$$\begin{aligned} P_n &= \tfrac{1}{2}|E_{tan}||H_{tan}| \cos 45° \\ &= \left(\frac{1}{2\sqrt{2}}\right) |n_m||H_{tan}|^2 \\ &= \left(\frac{1}{2\sqrt{2}}\right) \frac{|E_{tan}|^2}{|n_m|} \end{aligned} \tag{6-24}$$

where the bars | | indicate the absolute magnitude of the complex quantity. For a conductor which has a thickness very much greater than the skin depth δ, the surface impedance Z_s is equal to the intrinsic impedance n_m of the conductor, so that

$$P_n = \frac{1}{2\sqrt{2}} |Z_s||H_{tan}|^2 = \frac{1}{2\sqrt{2}} \frac{|E_{tan}{}^2|}{|Z_s|} \quad \text{watt/sq m} \tag{6-25}$$

In a conductor the linear current density $\mathbf{J}$ is equal in magnitude to the tangential magnetic intensity at the surface, so

$$P_n = \left(\frac{1}{2\sqrt{2}}\right) |Z_s||J|^2 \quad \text{watt/sq m} \tag{6-26}$$

In expressions (25) and (26), E_{tan}, H_{tan}, and J are peak values. In terms of effective values

$$\begin{aligned} P_n &= \frac{1}{\sqrt{2}} \frac{|E_{t(\text{eff})}|^2}{|Z_s|} = \frac{1}{\sqrt{2}} |Z_s||H_{t(\text{eff})}|^2 \\ &= \frac{1}{\sqrt{2}} |Z_s||J_{\text{eff}}|^2 \\ &= R_s J_{\text{eff}}{}^2 \qquad \text{watt/sq m} \end{aligned} \tag{6-27}$$

This result agrees with that previously obtained in chap. 5.

[NOTE: In this chapter, where the notions of complex power and complex Poynting vector were introduced, special type (bold italic) was used to distinguish complex scalars from real scalars. After the student or engineer has become familiar with the use of complex quantities, the need for making a distinction between the two hardly ever exists, and so most engineering texts use the same type (light-face italic) for complex scalars as for real scalars. This practice will be followed in the remainder of the text, except in special cases where it is desired to emphasize that the quantities are complex rather than real.]

Problem 4. A uniform plane wave having field components E_x and H_y is guided in the z direction between a pair of parallel copper planes. If the frequency is 100 mc and the field intensity of the transmitted wave is $E_x = 1$ volt/m, determine by two methods the power loss per square meter in each of the conducting planes.

BIBLIOGRAPHY

See bibliography for chap. 4.

CHAPTER 7

GUIDED WAVES

In the wave propagation so far discussed, only uniform plane waves, remote from any guiding surfaces, have been considered. In many actual cases, propagation is by means of *guided* waves, that is, waves that are guided along or over conducting or dielectric surfaces. Common examples of guided electromagnetic waves are the waves along ordinary parallel-wire and coaxial transmission

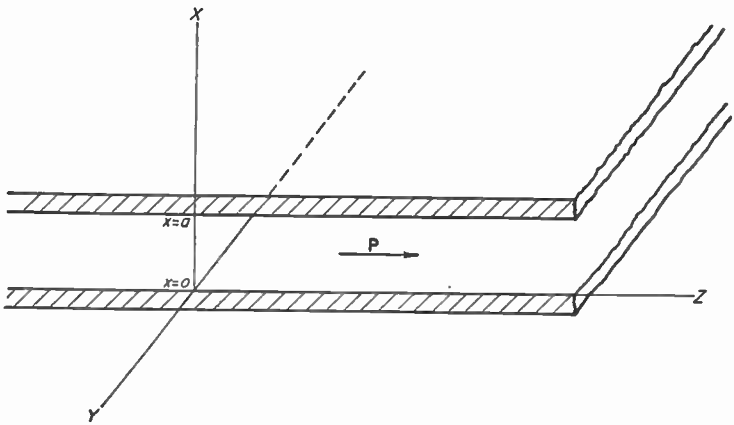


FIG. 7-1. Parallel conducting planes.

lines, waves in wave guides, and waves that are guided along the earth's surface from a radio transmitter to the receiving point. The study of such guided waves will now be undertaken.

7.01 Waves between Parallel Planes. For purposes of study a simple illustrative example is that of an electromagnetic wave, propagating between a pair of parallel perfectly conducting planes of infinite extent in the y and z directions (Fig. 7-1). In order to determine the electromagnetic field configurations in the region between the planes, Maxwell's equations will be solved subject to the appropriate boundary conditions. Because perfectly conducting planes have been assumed, these boundary conditions are very simple, being

$$E_{\text{tangential}} = 0, \qquad H_{\text{normal}} = 0$$

at the surfaces of the conductors.*

In general, and assuming that all time variations are as $e^{j\omega t}$, Maxwell's equations and the wave equations are

$$\text{curl } \mathbf{H} = (\sigma + j\omega\epsilon)\mathbf{E} \qquad \text{curl } \mathbf{E} = -j\omega\mu\mathbf{H} \tag{7-1}$$

$$\nabla^2\mathbf{E} = \gamma^2\mathbf{E} \qquad \nabla^2\mathbf{H} = \gamma^2\mathbf{H} \tag{7-2}$$

where

$$\gamma = \sqrt{(\sigma + j\omega\epsilon)(j\omega\mu)} \tag{7-3}$$

In rectangular co-ordinates, and for the nonconducting region between the planes, these equations become

$$\left.\begin{aligned}
\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= j\omega\epsilon E_x & \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -j\omega\mu H_x \\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= j\omega\epsilon E_y & \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -j\omega\mu H_y \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= j\omega\epsilon E_z & \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -j\omega\mu H_z
\end{aligned}\right\} \tag{7-4}$$

$$\left.\begin{aligned}
\frac{\partial^2\mathbf{E}}{\partial x^2} + \frac{\partial^2\mathbf{E}}{\partial y^2} + \frac{\partial^2\mathbf{E}}{\partial z^2} &= -\omega^2\mu\epsilon\mathbf{E} \\
\frac{\partial^2\mathbf{H}}{\partial x^2} + \frac{\partial^2\mathbf{H}}{\partial y^2} + \frac{\partial^2\mathbf{H}}{\partial z^2} &= -\omega^2\mu\epsilon\mathbf{H}
\end{aligned}\right\} \tag{7-5}$$

It will be assumed that propagation is in the z direction, and that the variation of all field components in this direction may be expressed in the form $e^{-\bar{\gamma}z}$, where in general

$$\bar{\gamma} = \bar{\alpha} + j\bar{\beta} \tag{7-6}$$

is a complex propagation constant,† whose value is to be determined.

* It is easy to show for actual conductors such as copper or brass (which have a very high, but not infinite, conductivity) that the finite conductivity has negligible effect on the field configuration. Therefore it is possible to use the fields calculated on the basis of perfectly conducting planes to determine the surface currents that must flow in these planes. The currents so calculated may then be used to compute the losses, and hence the attenuation, which occur with finitely conducting planes. This is a standard engineering approach.

† In general $\bar{\gamma}$ will not be equal to γ, defined by equation (3), but $\bar{\gamma}$ reduces to γ in the special case of uniform plane waves.

This is a quite reasonable assumption because (as will be shown later) for any uniform transmission line or guide the fields must obey an exponential law along the line. When the time variation factor is combined with the z-variation factor, it is seen that the combination

$$e^{j\omega t} \cdot e^{-\bar{\gamma}z} = e^{(j\omega t - \bar{\gamma}z)} = e^{-\bar{\alpha}z} \cdot e^{j(\omega t - \bar{\beta}z)} \tag{7-7}$$

represents a wave propagating in the z direction. If $\bar{\gamma}$ happens to be an imaginary number, that is if $\bar{\alpha} = 0$, expression (5) represents a wave without attenuation. On the other hand, if $\bar{\gamma}$ is real so that $\bar{\beta} = 0$, there is no wave motion but only an exponential decrease in amplitude.

Since the space between the planes is infinite in extent in the y direction, there are no boundary conditions to be met in this direction, and it can be assumed that the field is uniform or constant in the y direction. This means that the derivatives with respect to y in (4) can be put equal to zero. In the x direction however, there are certain boundary conditions which must be met. Therefore it is not possible to specify arbitrarily what the distribution of fields in this direction will be. This answer must come out of the solution of the differential equations when the boundary conditions are applied.

When the variation in the z direction of each of the field components is shown explicitly by writing, for example,

$$H_y = H_y^0 \, e^{-\bar{\gamma}z}$$

it is seen immediately that

$$\frac{\partial H_y}{\partial z} = -\bar{\gamma} H_y^0 \, e^{-\bar{\gamma}z} = -\bar{\gamma} H_y$$

with similar results for the z derivatives of the other components. Making use of this result and remembering that the y derivative of any component is zero, eqs. (4) and (5) become

$$\left.\begin{aligned} +\bar{\gamma}H_y &= j\omega\epsilon E_x & \bar{\gamma}E_y &= -j\omega\mu H_x \\ -\bar{\gamma}H_x - \frac{\partial H_z}{\partial x} &= j\omega\epsilon E_y & -\bar{\gamma}E_x - \frac{\partial E_z}{\partial x} &= -j\omega\mu H_y \\ \frac{\partial H_y}{\partial x} &= j\omega\epsilon E_z & \frac{\partial E_y}{\partial x} &= -j\omega\mu H_z \end{aligned}\right\} \tag{7-8}$$

$$\left.\begin{aligned}\frac{\partial^2 \mathbf{E}}{\partial x^2} + \bar{\gamma}^2\mathbf{E} &= -\omega^2\mu\epsilon\mathbf{E} \\ \frac{\partial^2 \mathbf{H}}{\partial x^2} + \bar{\gamma}^2\mathbf{H} &= -\omega^2\mu\epsilon\mathbf{H}\end{aligned}\right\} \tag{7-9}$$

In eqs. (9) it should be remembered that each of these equations is really three equations, one for each of the components of **E** or **H**. Equations (8) can be solved simultaneously to yield the following equations

$$\left.\begin{aligned} H_x &= -\frac{\bar{\gamma}}{h^2}\frac{\partial H_z}{\partial x} & E_x &= -\frac{\bar{\gamma}}{h^2}\frac{\partial E_z}{\partial x} \\ H_y &= -\frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x} & E_y &= +\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x}\end{aligned}\right\} \tag{7-10}$$

where

$$h^2 = \bar{\gamma}^2 + \omega^2\mu\epsilon \tag{7-11}$$

In eqs. (10) the various components of electric and magnetic intensities are expressed in terms of E_z and H_z. With the exception of one possibility, to be discussed later, it will be observed that there must be a z component of either E or H; otherwise all the components would be zero and there would be no fields at all in the region considered. Although in the general case both E_z and H_z could be present at the same time, it is convenient and desirable to divide the solutions into two sets. In the first of these, there is a component of **E** in the direction of propagation (E_z), but no component of **H** in this direction. Such waves are called *E waves*, or more commonly, *transverse magnetic* (TM) waves, because the magnetic intensity **H** is entirely transverse. The second set of solutions has a component of **H** in the direction of propagation, but no E_z component. Such waves are called *H waves* or *transverse electric* (TE) waves. The solutions to eqs. (8) and (9) for these two cases will now be obtained. Since the differential equations are linear, the sum of these two sets of solutions yields the most general solution.

7.02 Transverse Electric Waves ($E_z = 0$). Inspection of eqs. (10) shows that when $E_z \equiv 0$, but H_z does not equal zero, the field components H_y and E_x will also equal zero, whereas, in general, there will be nonzero values for the components H_x and E_y. Since each of the field components obeys the wave equation as given by eqs. (9), the wave equation can be written for the component E_y

$$\frac{\partial^2 E_y}{\partial x^2} + \bar{\gamma}^2 E_y = -\omega^2 \mu\epsilon E_y$$

This can be written as

$$\frac{\partial^2 E_y}{\partial x^2} = -h^2 E_y \tag{7-12}$$

Recalling that $E_y = E_y{}^0(x)\, e^{(j\omega t - \bar{\gamma} z)}$, eq. (12) reduces to

$$\frac{d^2 E_y{}^0}{dx^2} = -h^2 E_y{}^0 \tag{7-12a}$$

where as before $$h^2 = \bar{\gamma}^2 + \omega^2 \mu\epsilon$$

Equation (12a) is the differential equation of simple harmonic motion. Its solution can be written in the form

$$E_y{}^0 = C_1 \sin hx + C_2 \cos hx \tag{7-13}$$

where C_1 and C_2 are arbitrary constants.

Showing the variation with time and in the z direction the expression for E_y is

$$E_y = (C_1 \sin hx + C_2 \cos hx)\, e^{(j\omega t - \bar{\gamma} z)} \tag{7-13a}$$

The arbitrary constants C_1 and C_2 can be determined from the boundary conditions. For the parallel-plane wave guide of Fig. (7-1) the boundary conditions are quite simple. They require that the tangential component of **E** be zero at the surface of the (perfect) conductors for all values of z and time. This requires that

$$\left.\begin{array}{ll} E_y = 0 & \text{at } x = 0 \\ E_y = 0 & \text{at } x = a \end{array}\right\} \text{for all values of } z \text{ and } t \qquad \text{(boundary conditions)}$$

In order for the first of these conditions to be true, it is evident that C_2 must be zero. Then the expression for E_y is

$$E_y = C_1 \sin hx\, e^{(j\omega t - \bar{\gamma} z)}$$

Application of the second boundary condition imposes a restriction on h. In order for E_y to be zero at $x = a$ for all values of z and t it is necessary that

$$h = \frac{m\pi}{a} \tag{7-14}$$

where $m = 1, 2, 3, \ldots$

(The special case of $m = 0$ will be discussed later.) Therefore

$$E_y = C_1 \sin\left(\frac{m\pi}{a}x\right)e^{(j\omega t - \bar{\gamma}z)} \tag{7-15}$$

The other components of E and H can be obtained by inserting eq. (15) in eqs. (10). When this is done, it is seen that the expressions for the field intensities for transverse electric waves between parallel planes are

$$\begin{aligned} E_y &= C_1 \sin\left(\frac{m\pi}{a}x\right)e^{(j\omega t - \bar{\gamma}z)} \\ H_z &= -\frac{m\pi}{j\omega\mu a} C_1 \cos\left(\frac{m\pi}{a}x\right)e^{(j\omega t - \bar{\gamma}z)} \\ H_x &= -\frac{\bar{\gamma}}{j\omega\mu} C_1 \sin\left(\frac{m\pi}{a}x\right)e^{(j\omega t - \bar{\gamma}z)} \end{aligned} \tag{7-16}$$

Each value of m specifies a particular field configuration or *mode*, and the wave associated with the integer m is designated as the $TE_{m,0}$ wave or $TE_{m,0}$ mode. The second subscript (equal to zero in this case) refers to another factor which varies with y, which is found in the general case of rectangular guides. It will be noticed that the smallest value of m that can be used in eqs. (16) is $m = 1$, because $m = 0$ makes all the fields identically zero. That is, the lowest order mode that can exist in this case is the $TE_{1,0}$ mode.

In writing expressions for the field components as in eq. (16), the variation of all the fields with time and in the z direction is the same for any particular value of m and is shown by the factor $e^{j\omega t - \bar{\gamma}z}$. Rather than carry this factor through the entire analysis, it is customary to drop it, putting it back in for the final result. Thus eqs. (16) can be written

$$\begin{aligned} E_y &= C_1 \sin\frac{m\pi}{a}x \\ H_x &= -\frac{\bar{\gamma}}{j\omega\mu} C_1 \sin\frac{m\pi}{a}x \\ H_z &= \frac{-m\pi}{j\omega\mu a} C_1 \cos\frac{m\pi}{a}x \end{aligned} \tag{7-17}$$

where now the factor $e^{(j\omega t - \bar{\gamma}z)}$ is understood.* Whenever it is desired

* As shown in (17), E_y, H_x, and H_z are functions of x only, and not of time or z. They are the *crest* values or amplitudes of the waves. Most relations

to show the time and z variations explicitly, the expressions are multiplied by $e^{(j\omega t - \bar{\gamma} z)}$.

The factor $\bar{\gamma}$ is the propagation constant, which is ordinarily complex, the real part $\bar{\alpha}$ being the attenuation constant and the imaginary part $\bar{\beta}$ being the phase shift constant. However, it will be shown in section 7.04 that for the present problem of waves guided by perfectly conducting walls, $\bar{\gamma}$ is either a pure real or a pure imaginary. In that range of frequencies where $\bar{\gamma}$ is real, $\bar{\alpha}$ has value but $\bar{\beta}$ is zero, so that there is attenuation but no phase shift and, therefore, no wave motion. In the range of frequencies where $\bar{\gamma}$ is imaginary, $\bar{\alpha}$ is zero but $\bar{\beta}$ has value, so that there is propagation by wave motion without attenuation. It is this latter range of frequencies that is of chief interest in wave guide propagation. Writing $\bar{\gamma} = j\bar{\beta}$, eqs. (16) for $TE_{m,0}$ waves in the propagation range may be written as

$$\begin{aligned} E_y &= C_1 \sin\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \\ H_x &= -\frac{\bar{\beta}}{\omega\mu} C_1 \sin\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \\ H_z &= \frac{jm\pi}{\omega\mu a} C_1 \cos\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \end{aligned} \qquad (7\text{-}16a)$$

A sketch of these field distributions at some particular instant of time is shown in Fig. 7-2 for the $TE_{1,0}$ mode.

7.03 Transverse Magnetic Waves ($H_z \equiv 0$). The case of transverse magnetic waves between parallel planes can be solved in a manner similar to that used for TE waves. In this instance H_z will be zero, and inspection of eqs. (10) shows that H_x and E_y will also be zero, while in general, E_z, E_x, and H_y will have value. Solving the wave equation for H_y, gives as before

$$H_y = C_3 \sin hx + C_4 \cos hx \qquad (7\text{-}18)$$

where the factor $e^{(j\omega t - \bar{\gamma} z)}$ is understood. The boundary conditions cannot be applied directly to H_y to evaluate the constants C_3 and

between the field components, including those which involve determination of power, can be made in terms of these crest values. However, when it is desired to differentiate or integrate any of the field components with respect to time or z, the factor $e^{(j\omega t - \gamma z)}$ is put back in until the desired operation has been performed, and then this factor is again dropped.

C_4, because in general the tangential component of H is not zero at the surface of a conductor. However from eqs. (10) the expressions for E_z can be obtained in terms of H_y, and then the boundary conditions applied to E_z. From eqs. (10) and (18)

$$E_z = \frac{h}{j\omega\epsilon}[C_3 \cos hx - C_4 \sin hx]$$

Applying the boundary conditions that E_z must be zero at $x = 0$ shows that $C_3 = 0$. The second condition that E_z must be zero

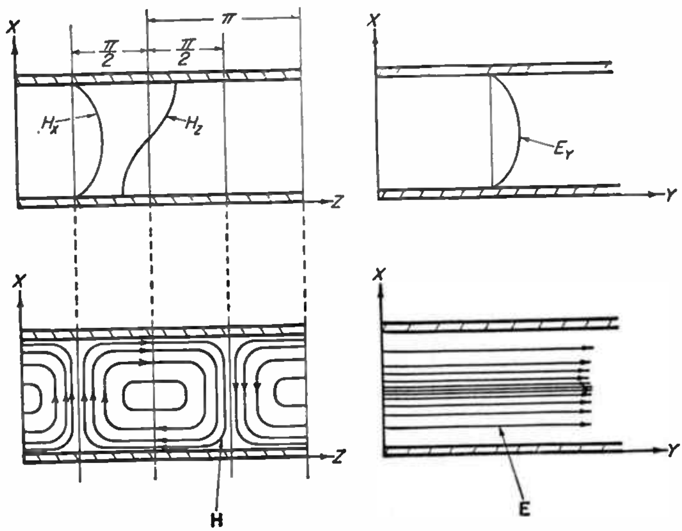


FIG. 7-2. Electric and magnetic fields between parallel planes for the $TE_{1,0}$ wave.

at $x = a$ requires that $h = m\pi/a$ where m is any integer. Then the expressions for E_z, H_y, and E_x become

$$\left.\begin{aligned} E_z &= -\frac{m\pi C_4}{j\omega\epsilon a} \sin \frac{m\pi}{a} x \\ H_y &= C_4 \cos \frac{m\pi}{a} x \\ E_x &= \frac{\bar{\gamma} C_4}{j\omega\epsilon} \cos \frac{m\pi}{a} x \end{aligned}\right\} \qquad (7\text{-}19)$$

Multiplying by the factor $e^{(j\omega t - \bar{\gamma} z)}$ to show the variation with time and in the z direction, and putting $\bar{\gamma} = j\bar{\beta}$ for the range of frequencies in which wave propagation occurs, the expressions for TM waves between parallel perfectly conducting planes are

$$\left.\begin{aligned} H_y &= C_4 \cos\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \\ E_x &= \frac{\bar{\beta}}{\omega\epsilon} C_4 \cos\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \\ E_z &= \frac{jm\pi}{\omega\epsilon a} C_4 \sin\left(\frac{m\pi}{a}x\right) e^{j(\omega t - \bar{\beta} z)} \end{aligned}\right\} \tag{7-19a}$$

As in the case of transverse electric waves, there is an infinite number of modes corresponding to the various values of m from 1 to infinity. However in this case of transverse magnetic waves there

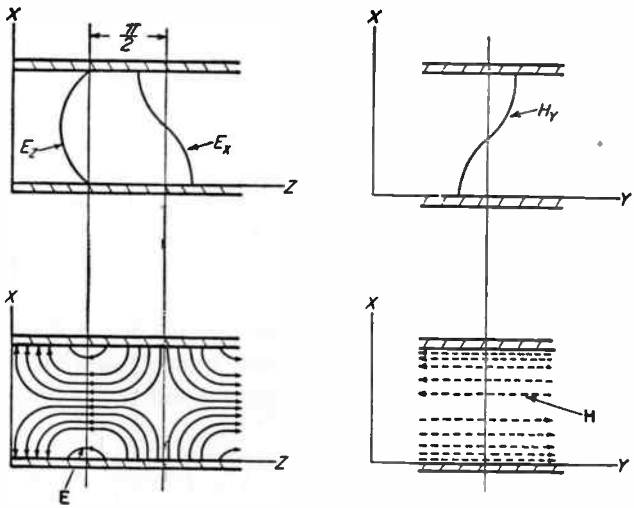


FIG. 7-3. The $TM_{1,0}$ wave between parallel planes.

is also the possibility of $m = 0$, because $m = 0$ in the above equations does not make all the fields vanish. This particular case of $m = 0$ will be discussed in detail in a later section. A sketch of the $TM_{1,0}$ wave between parallel planes is shown in Fig. 7-3.

7.04 Characteristics of TE and TM Waves. The transverse electric and transverse magnetic waves between parallel conducting planes exhibit some interesting and rather surprising properties that seem quite different from those of uniform plane waves in free space. These properties can be studied by investigating the propagation constant $\bar{\gamma}$ for these waves.

Examination of eqs. (16) for TE waves and eqs. (19a) for TM waves shows that for each of the components of **E** or **H** there is a sinusoidal or cosinusoidal standing-wave distribution across the guide in the x direction. That is, each of these components varies in magnitude, but not in phase, in the x direction. In the y direction, by assumption, there is no variation of either magnitude or phase of any of the field components. Thus any x-y plane is an equiphase plane for each of the field components (that is, any particular component, E_y for example, reaches its maximum value in time at the same instant for all points on the plane). Also these equiphase surfaces progress along the guide in the z direction with a velocity $\bar{v} = \omega/\bar{\beta}$, where $\bar{\beta}$, the phase shift constant, is the imaginary part of the propagation constant $\bar{\gamma}$. Now from eq. (11), $\bar{\gamma}$ can be expressed in terms of h and frequency and the constants of the medium by

$$\bar{\gamma} = \sqrt{h^2 - \omega^2\mu\epsilon} \tag{7-20}$$

Inserting the restrictions on h imposed by eq. (14), this becomes

$$\bar{\gamma} = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\epsilon} \tag{7-21}$$

Inspection of eq. (21) shows that at frequencies sufficiently high so that $\omega^2\mu\epsilon > (m\pi/a)^2$, the quantity under the radical will be negative and $\bar{\gamma}$ will be a pure imaginary equal to $j\bar{\beta}$, where

$$\bar{\beta} = \sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2} \tag{7-22}$$

Under these conditions the fields will progress in the z direction as waves, and the attenuation of these waves will be zero (for perfectly conducting planes).

As the frequency is decreased, a critical frequency $f_c = \frac{\omega_c}{2\pi}$ will be reached at which

$$\omega_c{}^2\mu\epsilon = \left(\frac{m\pi}{a}\right)^2 \tag{7-23}$$

For all frequencies less than f_c, the quantity under the radical will be positive and the propagation constant will be a real number. That is, $\bar{\gamma}$ will have value but $\bar{\beta}$ will equal zero. This means that the fields will be attenuated exponentially in the z direction and that there will be no wave motion, since the phase shift per unit length is now zero. The frequency f_c, at which wave motion ceases, is called the cut-off frequency of the guide. From eq. (23).

$$f_c = \frac{m}{2a\sqrt{\mu\epsilon}} \tag{7-24}$$

It is seen that for each value of m, there is a corresponding cut-off frequency below which wave propagation cannot occur. Above the cut-off frequency, wave propagation does occur and the attenuation of the wave is zero (for perfectly conducting planes). The phase shift constant $\bar{\beta}$, in the range where wave propagation occurs, is given by eq. (22). It is seen that $\bar{\beta}$ varies from zero at the cut-off frequency up to the value $\omega\sqrt{\mu\epsilon}$ as the frequency approaches infinity. The distance required for the phase to shift through 2π radians is a wavelength, so that the wavelength $\bar{\lambda}$ is given in terms of $\bar{\beta}$ by

$$\bar{\lambda} = \frac{2\pi}{\bar{\beta}} \tag{7-25}$$

Also the velocity of propagation of the *wave* is given by the wavelength times the frequency, so that

$$\bar{v} = \bar{\lambda}f = \frac{2\pi f}{\bar{\beta}} = \frac{\omega}{\bar{\beta}} \tag{7-26}$$

When the expression for $\bar{\beta}$ is put in eqs. (25) and (26), the wavelength and wave velocity are given by

$$\bar{\lambda} = \frac{2\pi}{\sqrt{\omega^2\mu\epsilon - (m\pi/a)^2}} \tag{7-27}$$

$$\bar{v} = \frac{\omega}{\sqrt{\omega^2\mu\epsilon - (m\pi/a)^2}} \tag{7-28}$$

It is seen that at the cut-off frequency both $\bar{\lambda}$ and $\bar{v}$ are infinitely large. As the frequency is raised above the cut-off frequency, the velocity decreases from this very large value. It approaches a lower limit.

$$\bar{v} \to v_0 = \frac{1}{\sqrt{\mu\epsilon}} \tag{7-29}$$

as the frequency becomes high enough so that $(m\pi/a)^2$ is negligible compared with $\omega^2\mu\epsilon$. When the dielectric medium between the plates is air, μ and ϵ have their free space values μ_v and ϵ_v, and the lower limit of velocity, given by (29), is just the free-space velocity c, where as usual

$$c = \frac{1}{\sqrt{\mu_v\epsilon_v}} \approx 3 \times 10^8 \qquad \text{meter/sec}$$

Therefore the velocity of the wave varies from a value equal to the velocity of light in free-space up to an infinitely large value as the frequency is reduced from extremely high values down to the cut-off frequency. This velocity is the *wave* velocity or *phase* velocity, and is different from the velocity with which the energy propagates. The distinction between these velocities will be considered in a later section of this chapter.

7.05 Transverse Electromagnetic Waves. For transverse electric (TE) waves between the parallel planes, it was seen that the lowest value of m that could be used without making all the field components zero was $m = 1$. That is, the lowest-order TE wave is the $TE_{1,0}$ wave. For transverse magnetic (TM) waves however, a value of m equal to zero does not necessarily require that all the fields be zero. Putting $m = 0$ in eqs. (19a) leaves

$$\left.\begin{aligned} H_y &= C_4\, e^{j(\omega t - \bar{\beta} z)} \\ E_x &= \frac{\bar{\beta}}{\omega\epsilon} C_4\, e^{j(\omega t - \bar{\beta} z)} \\ E_z &= 0 \end{aligned}\right\} \tag{7-30}$$

For this special case of transverse magnetic waves the component of **E** in the direction of propagation, that is E_z, is also zero so that the electromagnetic field is *entirely transverse*. Consistent with previous notation this wave is called the transverse electromagnetic (TEM) wave. Although it is a special case of guided-wave propagation, it is an extremely important one, because it is the familiar type of wave propagated along all ordinary two-conductor transmission lines when operating in their customary (low-frequency) manner. It is usually called the *principal wave*.

There are several interesting properties of TEM waves which follow as special cases of the more general TE or TM types of waves. For the TEM waves between the parallel planes it is seen from eqs. (30) that not only are the fields entirely transverse, but they are constant in amplitude across a cross-section normal to the direction of propagation; and, of course, their ratio is also constant. For $m = 0$ and an air dielectric, the expressions for $\bar{\gamma}$, $\bar{\beta}$, $\bar{v}$, and $\bar{\lambda}$ reduce to

$$\left.\begin{aligned} \bar{\gamma} \rightarrow \gamma &= j\omega\sqrt{\mu_v\epsilon_v} \\ \bar{\beta} \rightarrow \beta &= \omega\sqrt{\mu_v\epsilon_v} \\ \bar{v} \rightarrow v &= \frac{1}{\sqrt{\mu_v\epsilon_v}} = c \\ \bar{\lambda} \rightarrow \lambda &= \frac{2\pi}{\omega\sqrt{\mu_v\epsilon_v}} = \frac{c}{f} \end{aligned}\right\} \tag{7-31}$$

Unlike TE and TM waves, the velocity of the TEM wave is independent of frequency and has the familiar free-space value, $c \approx 3 \times 10^8$ meter/sec. (It has this value only when the planes are perfectly conducting and the space between them is a vacuum. The effect of finite conductivity for the conducting planes is to reduce the velocity slightly. This effect will be considered in a later section.) Also from eq. (24), the cut-off frequency for the TEM wave is zero. This means that for transverse electromagnetic waves, all frequencies down to zero can propagate along the guide. The ratio of E to H between the parallel planes for a traveling wave is

$$\left|\frac{E_x}{H_y}\right| = \frac{E_x}{H_y} = \frac{\bar{\beta}}{\omega\epsilon} = \sqrt{\frac{\mu_v}{\epsilon_v}} \tag{7-32}$$

which is just the intrinsic impedance, η_v, of free space.

A sketch of the TEM wave between parallel planes is shown in Fig. 7-4.

7.06 Velocities cf Propagation. It was seen that except for the TEM wave, the velocity with which an electromagnetic wave propagates (in an air dielectric) between a pair of parallel planes is always greater than c, the free-space velocity of electromagnetic waves. In actual rectangular or cylindrical wave guides (to be considered in chap. 9), the TEM wave cannot exist and the *wave* or *phase velocity* is always greater than the free-space velocity. On the other hand, the *velocity* with which the *energy* propagates along a guide is always less than the free space velocity. The relation between these velocities is made clear by consideration of a simple and well-known illustration. Figure 7-5 might be considered to represent water waves approaching the shore line or a breakwater a-a at an angle θ. The velocity of the waves could be determined

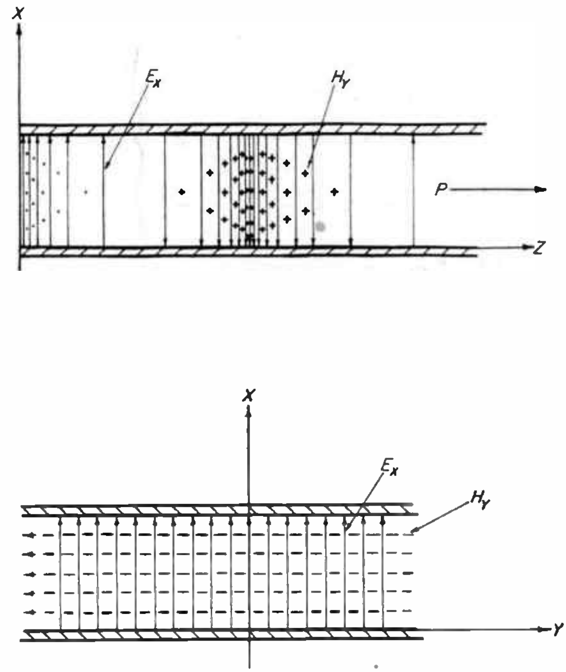


FIG. 7-4. The TEM wave between parallel planes.

by measuring the distance λ between successive crests and recording the frequency f with which the crests passed a given observation point. The velocity c with which the waves are traveling would be given by

$$c = \lambda f$$

Alternatively, if one wished to determine the velocity c without going into the water, this could be done by measuring the angle θ and the velocity $\bar{v} = v_z$ with which the crests move *along the shore* *line* (in the z direction). This velocity would be given by

$$\bar{v} = \bar{\lambda} f$$

where $\bar{\lambda}$ is now the distance between crests *along the shore line.* Evidently $\bar{v}$ and $\bar{\lambda}$ are greater than c and λ respectively, and are related to them by

$$\bar{\lambda} = \frac{\lambda}{\cos\theta} \tag{7-33a}$$

$$\bar{v} = \frac{c}{\cos\theta} \tag{7-33b}$$

When the direction of wave travel is nearly parallel to the shore, that is, when the angle θ is small, the velocity $\bar{v}$ with which the crests move along the shore line is very nearly equal to c, the free-space velocity of the waves. However, when the angle θ is near

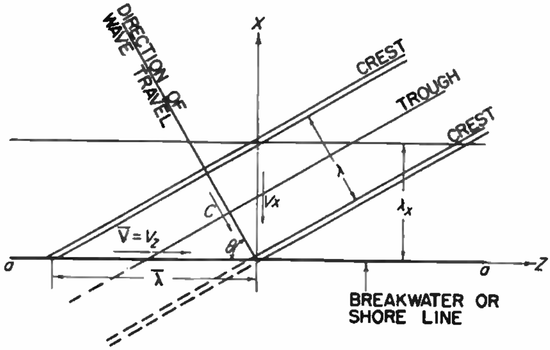


FIG. 7-5. Water wave approaching a breakwater.

90 degrees the velocity with which the crests advance along the shore line is very great, and approaches infinity as θ approaches 90 degrees.

Consider now wave propagation within a wave guide. It is always possible, though sometimes not too practical, to obtain the field configuration within a rectangular guide by superposing two or more plane waves in a suitable manner. For the $TE_{m,0}$ waves in rectangular guides and for these same waves between parallel planes as already considered, this separation into component waves is quite simple. It is left for the student to show (problem 2) that two uniform plane waves having the same amplitude and frequency, but opposite phases can be added to produce the field distributions of the $TE_{m,0}$ waves. The direction of the component waves are as shown in Fig. 7-6, where the angle θ between the walls of the guide and the direction of the waves depends upon the frequency and the dimension a. For each of the component waves the electric vector **E** will be in the y direction and the magnetic vector **H** will lie in the x-z plane and will be perpendicular to the direction of travel of that wave. In order to satisfy the boundary conditions at the walls of the guide, the electric fields due to the two component waves must add to zero at those surfaces. The only

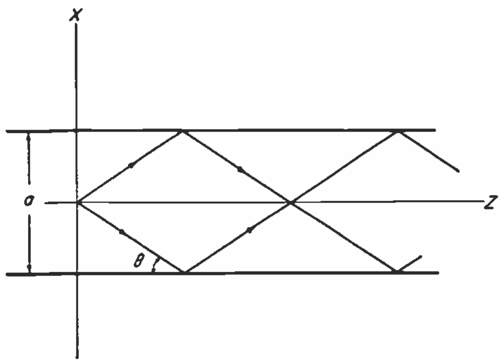


FIG. 7-6. Direction of travel of the component uniform-plane waves between parallel planes.

way in which it is possible to have E_y equal to zero at the walls and still have values of E_y at points between the walls is to have a standing wave distribution of E_y across the guide, with the nodal points of the standing wave occurring at the wall surfaces. This condition requires that a, the separation between the walls, must be some multiple of a half-wavelength measured in the direction perpendicular to the walls. Referring again to Fig. 7-5 the required condition is that

$$\frac{m\lambda_x}{2} = a$$

where m is an integer and where λ_x is the distance between crests measured in the x direction. Since $\lambda_x = \lambda/\sin\theta$, it is seen that the condition on θ is

$$\sin\theta = \frac{m\lambda}{2a} \tag{7-34}$$

Because the sine cannot be greater than unity, it is apparent that a, the separation between the walls must be greater than $\lambda/2$, where λ is the free-space wavelength of the wave. The wavelength for which

$$\lambda = \frac{2a}{m} \tag{7-35}$$

is the cut-off wavelength for that value of m. At the cut-off wavelength $\sin\theta$ is unity and θ is 90 degrees. That is, the waves bounce back and forth between the walls of the guide, and there is no wave motion parallel to the axis. As λ is decreased from the cut-off value, θ also decreases, so that at wavelengths much shorter than cut-off (very high frequency) the waves travel almost parallel to the axis of the guide.

The wavelength $\bar{\lambda} = \lambda_z$, *parallel* to the walls of the guide, which is the wavelength ordinarily measured in wave guide work, is given by

$$\bar{\lambda} = \frac{\lambda}{\cos\theta} = \frac{\lambda}{\sqrt{1 - (m\lambda/2a)^2}} \tag{7-36}$$

This is the distance between equiphase points in the direction of the axis of the guide. The phase velocity in this direction is

$$\bar{v} = \frac{c}{\cos\theta} = \frac{c}{\sqrt{1 - (m\lambda/2a)^2}} \tag{7-37}$$

It is evident that because of the zig-zag path traveled by each of the component waves, the velocity,* v_g, with which the energy propagates along the axis of the guide will be less than the free-space velocity c. In terms of the angle θ, for a guide with an air dielectric, it will be

$$v_g = c\cos\theta$$

In terms of the width dimension a in wavelengths, it is

$$v_g = c\sqrt{1 - \left(\frac{m\lambda}{2a}\right)^2} \tag{7-38}$$

It will be noted that the product of the phase velocity and the velocity with which the energy propagates is equal to the square of the

* This velocity, v_g is the group velocity. The terms *phase* velocity, *group* velocity, and *signal* velocity are discussed in more detail in Appendix I.

free-space velocity, that is,

$$\bar{v} \times v_g = c^2 \tag{7-39}$$

As the frequency is reduced toward the cut-off frequency, the angle θ approaches 90 degrees, so that the phase velocity $\bar{v}$ becomes very large, and the velocity with which the energy propagates becomes very small. At the cut-off frequency $\bar{v}$ is infinite, but v_g is zero, that is propagation of energy along the guide *by wave motion* ceases.*

For a (lossless) dielectric in the guide having permittivity ϵ and permeability μ, different from ϵ_v and μ_v, the velocity c must be replaced by $v_0 = 1/\sqrt{\mu\epsilon}$.

7.07 Attenuation in Parallel Plane Guides. The problem of wave propagation between parallel conducting planes has been solved for the theoretical case of *perfect* conductors, and the solutions appear as eqs. (16a), (19a), and (30) for the $TE_{m,0}$, $TM_{m,0}$ and TEM modes respectively. In actual wave guides the conductivity of the walls is usually very large, but it is never infinite, and there are always some losses. These losses will modify the results obtained for the lossless case by the introduction of the multiplying factor $e^{-\alpha z}$ in eqs. (16a), (19a), and (30). The problem now is to determine this attenuation factor α that is caused by losses in the walls of the guide.

In order to see how α may be evaluated for wave guides, consider the familiar problem of attenuation in ordinary two-conductor transmission lines. For any line with uniformly distributed constants, the amplitudes of voltage and current along the line (when the line is terminated in its characteristic impedance) are

$$V = V_0 e^{-\alpha z} \tag{7-40}$$

$$I = I_0 e^{-\alpha z} \tag{7-41}$$

and the average power transmitted is

$$\begin{aligned} W &= \tfrac{1}{2} VI \cos\theta \\ &= \tfrac{1}{2} V_0 I_0 e^{-2\alpha z} \cos\theta \end{aligned} \tag{7-42}$$

* This is not to say that there are no fields within the guide. In section 7.04 it was seen that below cut-off frequency $\bar{\gamma}$ is real, so that $\bar{\alpha}$ has value and $\bar{\beta}$ is zero. This means that the fields then penetrate into the guide with an exponential decrease in amplitude, and with no phase shift (for the infinitely-long guide with perfectly-conducting walls). A wave guide operated in this manner is known as an *attenuator*.

The rate of decrease of transmitted power along the line will be

$$-\frac{\partial W}{\partial z} = +2\alpha W \tag{7-43}$$

The decrease of transmitted power per *unit length* of line is

$$-\Delta W = 2\alpha W$$

and this must be equal to the power lost or dissipated per unit length. Therefore

$$\frac{\text{Power lost per unit length}}{\text{Power transmitted}} = \frac{2\alpha W}{W} = 2\alpha$$

so that

$$\alpha = \frac{\text{Power lost per unit length}}{2 \times \text{power transmitted}} \tag{7-44}$$

Using eq. (44), the attenuation factor can be determined for more general cases of guided wave transmission where the terms "voltage" and "current" may no longer apply.

The computation of power loss in a wave guide appears at first glance to be a rather difficult problem, because the loss depends upon the field configuration within the guide, and the field configuration, in turn, depends to some extent upon the losses. The attack on this problem is one that is used quite often in engineering. It is first assumed that the losses will have negligible effect upon the field distribution within the guide. Using the field distributions calculated for the lossless case, the magnetic-intensity tangential to each conducting surface is used to determine the current flow in that surface. Using this value of current and the known resistance of the walls, the losses are computed and α is determined from (44). If desired, a second and closer approximation could then be made, using a field distribution corrected to account for the calculated losses. However, for metallic conductors of high conductivity such as copper or brass, the first approximation yields quite accurate results, and a second approximation is rarely necessary.

EXAMPLE 1: *Attenuation Factor for the TEM Wave.* The expressions obtained for magnetic and electric fields between parallel perfectly conducting planes (Fig. 7-1) in the case of the TEM mode were

$$H_y = C_4 e^{j(\omega t - \beta z)}$$
$$E_x = \eta C_4 e^{j(\omega t - \beta z)} \tag{7-30a}$$

The linear current density in each of the conducting planes will be given by

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}$$

so the amplitude of the linear current density in each plane is

$$J = C_4$$

The loss per square meter in each conducting plane is

$$\tfrac{1}{2}J^2R_s = \tfrac{1}{2}C_4{}^2R_s$$

where

$$R_s = \sqrt{\frac{\omega\mu_m}{2\sigma_m}}$$

is the resistive component of the surface impedance given by the expression

$$Z_s = \sqrt{\frac{j\omega\mu_m}{\sigma_m}}$$

μ_m and σ_m refer of course to values in the metallic conductor. The total loss in the upper and lower conducting surfaces per meter length for a width b meters of the guide is

$$C_4{}^2R_sb$$

The power transmitted down the guide per unit cross-sectional area is

$$\tfrac{1}{2}\, Re\, (\mathbf{E} \times \mathbf{H}^*). \tag{7-45}$$

E_x and H_y are right angles and in time phase and $|E_x| = \eta|H_y|$, so (45) reduces to

$$\tfrac{1}{2}\eta C_4{}^2$$

For a spacing a meters between the planes the cross section area of a width b meters of the guide is ba square meters and the power transmitted through this area is

$$\text{Power transmitted} = \tfrac{1}{2}\eta C_4{}^2ba$$

From (44) the attenuation factor is

$$\alpha = \frac{C_4{}^2R_sb}{2 \times \tfrac{1}{2}\eta C_4{}^2ba}$$

$$= \frac{R_s}{\eta a} = \frac{1}{\eta a}\sqrt{\frac{\omega\mu_m}{2\sigma_m}} \qquad \text{nepers/meter} \tag{7.46}$$

This expression should be compared with the corresponding expression for the attenuation factor of an ordinary transmission line (eq. 8-65), which is

$$\alpha = \frac{R}{2Z_0}$$

where R is the resistance per unit length of the line (that is *twice* the conductor resistance).

EXAMPLE 2: *Attenuation of TE Waves.* The expressions for E and H for the transverse electric modes between perfectly conducting parallel planes (Fig. 7-1) are

$$\left.\begin{aligned} E_y &= C_1 \sin\left(\frac{m\pi}{a}x\right) e^{j(\omega t-\bar{\beta}z)} \\ H_x &= -\frac{\bar{\beta}}{\omega\mu} C_1 \sin\left(\frac{m\pi}{a}x\right) e^{j(\omega t-\bar{\beta}z)} \\ H_z &= \frac{jm\pi}{\omega\mu a} C_1 \cos\left(\frac{m\pi}{a}x\right) e^{j(\omega t-\bar{\beta}z)} \end{aligned}\right\} \qquad (7\text{-}16a)$$

The amplitude of linear current density in the conducting planes will be equal to the tangential component of H (i.e., H_z) at $x = 0$ and $x = a$

$$\begin{aligned} |J_y| &= |H_z| \qquad (\text{at } x = 0,\ x = a) \\ &= \frac{m\pi C_1}{\omega\mu a} \end{aligned}$$

It is interesting to note in passing that for these modes there is no flow of current in the direction of wave propagation. The loss in each plate is

$$\frac{1}{2} J_y^2 R_s = \frac{m^2\pi^2 C_1^2 \sqrt{\omega\mu_m/2\sigma_m}}{2\omega^2\mu^2 a^2} \qquad (7\text{-}47)$$

The power transmitted in the z direction through an element of area $da = dx \cdot dy$ is

$$\begin{aligned} \text{Power transmitted per unit area} &= \tfrac{1}{2}\, Re\, (\mathbf{E} \times \mathbf{H}^*) \cdot da \\ &= -\tfrac{1}{2}(E_y H_x)\, dx\, dy \\ &= \frac{\bar{\beta} C_1^2}{2\omega\mu} \sin^2\left(\frac{m\pi}{a}x\right) dx\, dy \end{aligned}$$

Power transmitted in the z direction for a guide 1 meter wide with a spacing between conductors of a meters is

$$\int_{x=0}^{x=a} \frac{\bar{\beta} C_1^2}{2\omega\mu} \sin^2\left(\frac{m\pi}{a}x\right) dx = \frac{\bar{\beta} C_1^2 a}{4\omega\mu} \qquad (7\text{-}48)$$

Dividing twice expression (47) by twice expression (48), the attenuation factor is

$$\alpha = \frac{2m^2\pi^2 \sqrt{\omega\mu_m/2\sigma_m}}{\bar{\beta}\omega\mu a^3}$$

Recalling that $\bar{\beta} = \sqrt{\omega^2\mu\epsilon - (m\pi/a)^2}$ the expression for attenuation factor for TE waves between parallel conducting planes for frequencies above cut-off is

$$\alpha = \frac{2m^2\pi^2\sqrt{\omega\mu_m/2\sigma_m}}{\omega\mu a^3\sqrt{\omega^2\mu\epsilon - (m\pi/a)^2}} \tag{7-49}$$

The value of this expression decreases from infinity at cut-off to quite low values at higher frequencies. For frequencies very much higher than cut-off the attenuation varies inversely as the three-halves power of the frequency.

Attenuation Factor for TM Waves. The expression for the attenuation factor for TM waves between parallel conducting planes can be obtained in a similar manner. It differs from expression (49)

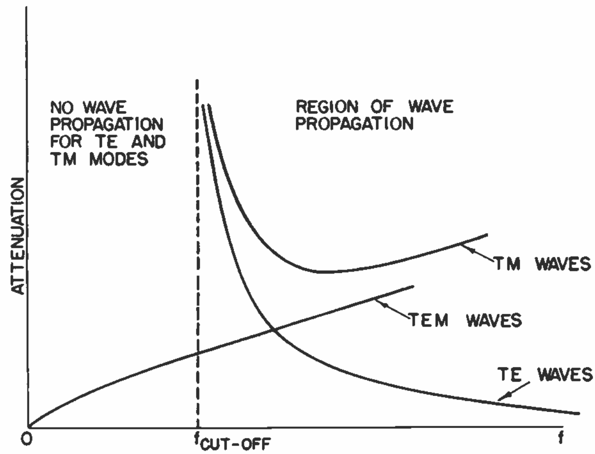


Fig. 7-7. Attenuation versus frequency characteristics of waves guided between parallel conducting planes.

in that the attenuation reaches a minimum at a frequency that is $\sqrt{3}$ times the cut-off frequency and then increases with frequency. At frequencies much higher than cut-off the attenuation of the TM modes increases directly as the square root of frequency.

A sketch of variation of attenuation with frequency for different modes propagating between parallel conducting planes is shown in Fig. 7-7.

7.08 Wave Impedances. In ordinary transmission line theory, a brief discussion of which is given in the next chapter, extensive use is made of the "characteristic impedance," Z_0, of the line. This impedance gives the ratio of voltage to current (for an infinitely long line), and its real part is a measure of the power transmitted for a given amplitude of current. In transmission line theory power is propagated along one axis only, and only one impedance constant is involved. However in three-dimensional wave propagation power may be transmitted along any or all of the three axes of the co-ordinate system, and consequently three impedance constants must be defined. For example, in the Cartesian co-ordinate system the complex power per unit area transmitted in the x, y, and z directions respectively is given by

$$P_x = \tfrac{1}{2}(E_yH_z^* - E_zH_y^*) \qquad P_y = \tfrac{1}{2}(E_zH_x^* - E_xH_z^*)$$
$$P_z = \tfrac{1}{2}(E_xH_y^* - E_yH_x^*)$$

The real or average Poynting vector in any of the three directions is given by the real part of the appropriate expression. It is now convenient to define the *wave impedances* at a point by the following ratios of electric to magnetic intensities:

$$\left.\begin{aligned} Z^+_{xy} &= \frac{E_x}{H_y} & Z^+_{yz} &= \frac{E_y}{H_z} & Z^+_{zx} &= \frac{E_z}{H_x} \\ Z^+_{yx} &= -\frac{E_y}{H_x} & Z^+_{zy} &= -\frac{E_z}{H_y} & Z^+_{xz} &= -\frac{E_x}{H_z} \end{aligned}\right\} \qquad (7\text{-}50)$$

These are the wave impedances looking along the *positive* directions of the co-ordinates, and this fact is indicated by the superscript plus sign. The impedances in the opposite directions are the negative of those given above, and the negative direction is indicated by a superscript minus sign. Thus in the directions of *decreasing* co-ordinates

$$\begin{aligned} Z^-_{xy} &= -\frac{E_x}{H_y} & Z^-_{yz} &= -\frac{E_y}{H_z} & Z^-_{zx} &= -\frac{E_z}{H_x} \\ Z^-_{yx} &= \frac{E_y}{H_x} & Z^-_{zy} &= \frac{E_z}{H_y} & Z^-_{xz} &= \frac{E_x}{H_z} \end{aligned}$$

Corresponding definitions would obtain for any orthogonal co-ordinate system. In terms of these wave impedances the x, y, and z components of complex Poynting vector are

$$P_x = \tfrac{1}{2}(Z^+_{yz}H_zH^*_z + Z^+_{zy}H_yH^*_y) = -\tfrac{1}{2}(Z^-_{yz}H_zH^*_z + Z^-_{zy}H_yH^*_y)$$
$$P_y = \tfrac{1}{2}(Z^+_{zx}H_xH^*_x + Z^+_{xz}H_zH^*_z) = -\tfrac{1}{2}(Z^-_{zx}H_xH^*_x + Z^-_{xz}H_zH^*_z)$$
$$P_z = \tfrac{1}{2}(Z^+_{xy}H_yH^*_y + Z^+_{yx}H_xH^*_x) = -\tfrac{1}{2}(Z^-_{xy}H_yH^*_y + Z^-_{yx}H_xH^*_x)$$

The subscripts on the wave impedances indicate the particular components of **E** and **H** involved, and the algebraic signs of the wave impedances have been chosen so that, if the real part of any given impedance is positive, the corresponding average power flow is in the direction indicated by the impedance.

Applying these definitions to waves propagating between parallel planes the wave impedance in the direction of propagation can be found. For the TEM wave (the exceptional case where both E and H are transverse), the wave impedance is given by eq. (32), and it is seen to be equal to η_v, the same as for a uniform plane wave in free space. For TE waves, the wave impedance can be obtained from eqs. (17). It is

$$Z^+_{yx} = -\frac{E_y}{H_x} = \frac{j\omega\mu}{\bar{\gamma}} \qquad \text{(7-51)}$$

where $$\bar{\gamma} = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \omega^2\mu\epsilon}$$

The wave impedance in the z direction is constant over the cross section of the guide. For frequencies below cut-off for which $\bar{\gamma}$ is real, the impedance is a pure reactance indicating no acceptance of power by the guide and therefore no transmission down the guide. For frequencies above cut-off $\bar{\gamma}$ is a pure imaginary (under the assumption of perfectly conducting walls) and can be written

$$\bar{\gamma} = j\bar{\beta} = j\sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2}$$

so that $$Z^+_{yx} = \frac{\omega\mu}{\sqrt{\omega^2\mu\epsilon - (m\pi/a)^2}} = \frac{\omega\mu}{\bar{\beta}} \qquad \text{(7-52)}$$

The wave impedance is real and decreases from an infinitely large value at cut-off toward the asymptotic value of $\eta = \sqrt{\mu/\epsilon}$ as the frequency increases to values much higher than cut-off.

These results could equally well have been obtained by considering the TE wave as being made of two uniform plane waves reflected back and forth between the conducting planes and making an angle θ with the axis of propagation (Fig. 7-6). For the TE wave propagating in the positive z direction the transverse component of **E** will be E_y, whereas the transverse component of **H** will be $-H_x = -H \cos \theta$; therefore the wave impedance in the z direction is

$$Z^+_{yx} = -\frac{E_y}{H_x} = -\frac{E}{H\cos\theta} = \frac{\eta}{\cos\theta} \tag{7-53}$$

Making use of eq. (37), this may be written as

$$Z^+_{yx} = \frac{\bar{v}}{c}\eta = \frac{\omega\mu}{\bar{\beta}}$$

which is the same result as was obtained in (52).

For TM waves the transverse component of **E** will be $E_x = E\cos\theta$, whereas the transverse component of **H** will be H_y. The wave impedance for this case is

$$Z^+_{xy} = \frac{E_x}{H_y} = \frac{E\cos\theta}{H} = \eta\cos\theta \tag{7-54}$$

It varies from zero at the cut-off frequency up to the asymptotic value η for frequencies much higher than cut-off.

There is a marked resemblance between the properties of these wave impedances and the characteristic impedances of the prototype T or π sections in ordinary filter theory. For example, the wave impedance for TE waves between parallel planes may be written as

$$Z^+_{yx} = \frac{\eta}{\cos\theta} = \frac{\eta}{\sqrt{1-\sin^2\theta}} \tag{7-55}$$

Making use of the relations

$$\sin\theta = \frac{m\lambda}{2a}, \qquad \lambda_c = \frac{2a}{m}, \qquad f_c = \frac{1}{\lambda_c\sqrt{\mu\epsilon}}$$

where λ_c and f_c are the cut-off wavelength and cut-off frequency, eq. (55) becomes

$$Z^+_{yx} = \frac{\eta}{\sqrt{1-(f_c/f)^2}} \tag{7-56}$$

This is similar to the expression for the characteristic impedance of the prototype π section of a high pass filter, which is

$$Z_{0\pi} = \frac{\sqrt{L/C}}{\sqrt{1-(f_c/f)^2}} \tag{7-57}$$

Similarly the expression for the wave impedance for TM waves between parallel planes may be written as

$$Z^{+}_{xy} = \eta \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{7-58}$$

which corresponds to the expression for characteristic impedance of the prototype T section of a high-pass filter,

$$Z_{0T} = \sqrt{\frac{L}{C}} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{7-59}$$

The wave impedances for waves between parallel planes are shown as functions of frequency in Fig. 7-8. In Chap. 9 a general transmission line analogy will be developed for TM and TE waves in cylindrical guides of any cross-sectional shape.

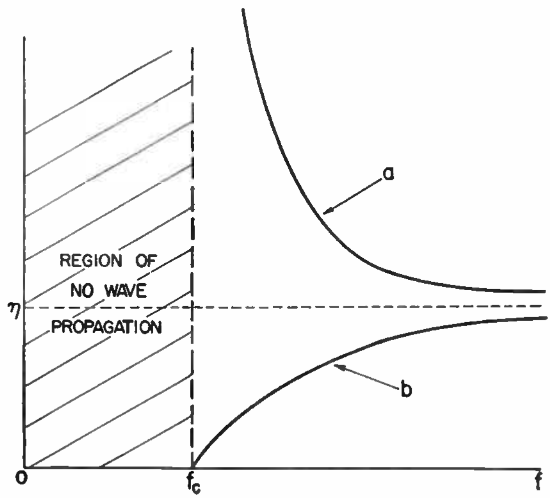


Fig. 7-8. Wave impedances for waves between parallel conducting planes: a, TE waves; b, TM waves.

In this chapter the characteristics of waves propagating between two parallel planes have been considered in some detail. The concepts developed in the treatment of this simple illustrative system are quite general and may be extended to apply to all guided systems. In chap. 8 these general principles will be applied to "ordinary" two-conductor transmission lines, and in chap. 9 application will be made to practical forms of waveguides. Before leaving this simple system some consideration will be given to the electric field configuration and current flow within the metal walls of the guiding system.

7.09 Electric Field and Current Flow Within the Conductor. When an electromagnetic wave is guided along the surface of a conductor, currents flow in the conductor and charges appear and disappear on its surface. The current distribution within the conductor and the charge distribution on the surface can be obtained from a straightforward solution of Maxwell's equations, subject

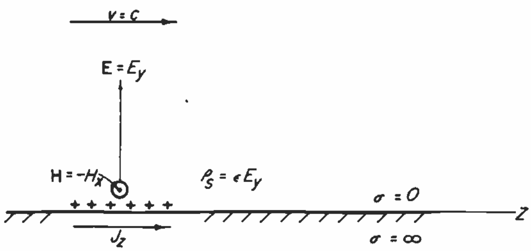


FIG. 7-9. Current and surface charge on a perfect conductor that is guiding an electromagnetic wave.

to the appropriate boundary conditions at the boundary surface between the dielectric and the conductor. However the results are somewhat complex and require interpretation. For this reason, before obtaining the exact solution, it is advantageous to consider in a qualitative manner, and from facts already known, certain features of the problem.

In Fig. 7-9 a TEM wave is guided along the surface (in the x-z plane) of a conductor which, for the moment, will be assumed to be perfectly conducting. For the case considered the electric intensity, $\mathbf{E} = \mathbf{j}E_y$, will be normal to the surface, and the magnetic intensity $\mathbf{H} = -\mathbf{i}H_x$ will be parallel to the surface. There will be a surface current J_z, flowing in the z direction, and related to the magnetic intensity by the vector relation $\mathbf{J} = \mathbf{n} \times \mathbf{H}$, which in this case becomes $J_z = -H_x$. Since $E_y = -\eta_v H_x$, it follows that

$$J_z = \frac{E_y}{\eta_v} \tag{7-60}$$

A surface charge density appears on the surface, the value of which is given by

$$\rho_s = D_y = \epsilon_v E_y \tag{7-61}$$

From (60) and (61) it is seen that both ρ_s and J_z are proportional to E_y, so that at any instant of time the position of maximum charge occurs at the same value of z as the position of maximum surface current.

When the conductivity of the conductor is reduced from infinity to a large but finite value such as obtains for ordinary metallic conductors, the situation is modified in several respects. The chief effect is the introduction of a small tangential component of **E**, which is required to drive the linear current density J against the surface impedance Z_s of the conductor. Making the assumption (known to be very good) that H will not be changed appreciably by the finite rather than infinite conductivity, the tangential component of **E** can be obtained from

$$\begin{aligned} E_z &= J_z Z_s = -H_x Z_s \\ &= -H_x \sqrt{\frac{j\omega\mu}{\sigma}} = -H_x \sqrt{\frac{\omega\mu}{\sigma}} \quad \underline{/45^\circ} \end{aligned}$$

The horizontal or tangential component of E is seen to lead $-H_x$ and therefore E_y by an angle of 45 degrees. The conductor is considered to be sufficiently good that the inequality $\sigma \gg \omega\epsilon$ holds for all frequencies considered. The depth of penetration, although small for good conductors, is not zero, and the linear current density J_z is now distributed throughout the thickness of the conductor, with approximately two-thirds of it concentrated within the "skin depth" δ. The linear current density J_z is still in phase with the magnetic intensity $-H_x$, but the current density i_z at the surface is in phase with E_z, and so leads $-H_x$ by 45 degrees. The penetration of the electric field and current waves into the conductor can be visualized by employing an artifice which yields an approximate but simple picture of the phenomena.

Since the electric field and current penetrate into the conductor by means of wave motion, it is convenient to think of the metallic medium as a large number of transmission lines, side by side guiding energy into the interior of the conductor* (Fig. 7-10). The

* G. W. O. Howe, "Wireless Currents at the Earth's Surface," *Wireless Engineer*, Vol. 17, No. 204, p. 385, September 1940.

picture is particularly simple if it is imagined that there are perfectly conducting strips, parallel to the x-y plane, imbedded in the metallic medium and serving as the "conductors" of the transmission lines.

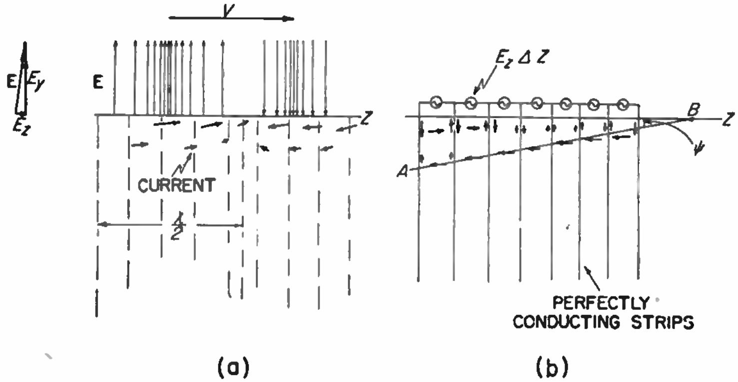


FIG. 7-10. Penetration of the electric field and current in a conductor that is not perfect (approximate representation).

Considering now a square vertical column of unit width (in the z direction and unit depth (in the x direction) as a transmission line, it will have the following constants per unit length (in the y direction):

$$G = \sigma \quad \text{mhos/meter} \qquad L = \mu \quad \text{henries/meter}$$
$$R = 0 \quad \text{ohms/meter} \qquad C = \epsilon \quad \text{farads/meter}$$

where σ,μ and ϵ are the constants of the metallic medium. The "input voltage" to each of these lines will be $E_z\,\Delta z = E_z$ (for $\Delta z = 1$). The "input current" per unit depth (in the x direction) will be equal to this voltage divided by the input impedance of the line. Assuming that the line is long enough so that any reflected wave has negligible amplitude (that is, conductor thickness $\gg \delta$), the input impedance will be equal to the "characteristic impedance" of the line.

$$Z_{\text{in}} = Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$
$$= \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} \approx \sqrt{\frac{\omega\mu}{\sigma}} \quad \underline{/45^\circ}$$

The "input current" to each "transmission line" flows down in one conductor, through the medium, and back up the second conductor.

If the input voltage were the same magnitude and phase for all lines, the vertical currents of adjacent lines would cancel and the current flow in the medium would be entirely in the horizontal direction.

This is the case when a uniform plane wave is incident normally on the surface of a conductor. In this case there is no current flow in the conducting strips and they may be removed without in any way affecting the current flow in the medium. However, when E_z is caused by a radio wave traveling parallel to the surface of the conducting medium, there will be a phase difference between the input voltages of adjacent transmission lines equal to $2\pi\,\Delta z/\lambda$, where $\Delta z/\lambda$ is the width in wavelengths of the vertical columns. In this case, there will be currents in the vertical strips as indicated in Fig. 7-10(b), and, of course, this simple analysis is no longer exact. However for a metallic medium where the depth of penetration is very small, the error in this approximate approach is also small. The picture could be improved upon by sloping the conducting strips to be normal to the phase front of the wave advancing into the metal. As the wave external to the surface advances in the z direction with a velocity $v_z = \omega/\beta_0$, the wave penetrates into the metal with a much slower velocity $v_y \approx \omega/\beta_1$. The first of these velocities is approximately equal to the velocity of light in free space, whereas the second is of the same order of magnitude as the velocity of sound in air. The slope of the line AB, parallel to an equiphase surface within the metal, is given by

$$\tan\psi_1 = \frac{v_y}{v_z} \approx \frac{\beta_0}{\beta_1} = \frac{\omega\sqrt{\mu\epsilon}}{\sqrt{\omega\mu\sigma/2}}$$

$$= \sqrt{\frac{2\omega\epsilon}{\sigma}}$$

For any good conductor the angle ψ_1 is very nearly zero. For example, for copper at 100 mc, $\psi_1 = 0.000079$ degrees. Since the direction of propagation of a wave is normal to the equiphase surfaces, the statement, that in the metal the wave propagates almost perpendicularly to the surface, is well justified.

Rigorous Solution. Having obtained a qualitative picture of what happens within a conducting medium as an electromagnetic wave is guided along its surface, it is now in order to set up and obtain a more rigorous solution as a boundary value problem. The problem is that of finding solutions to Maxwell's equations in regions O and I (Fig. 7-11), which will fit the boundary conditions at the surface of the conductor.

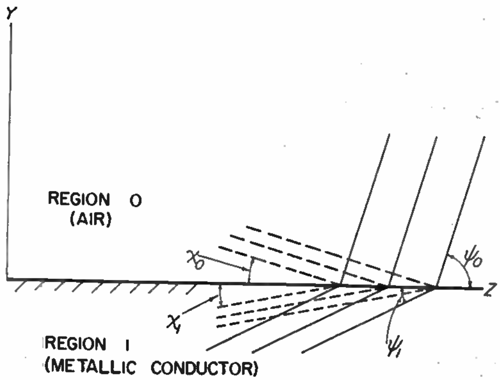


FIG. 7-11. Equiphase lines (solid) and equiamplitude lines (dashed) for an electromagnetic wave guided along a conducting plane.

The following assumptions will be made:

(1) No variations in the x direction. Therefore $\partial/\partial x \equiv 0$.

(2) Variations in the z direction can be represented by $e^{-\gamma_0 z}$ in the dielectric and by $e^{-\gamma_1 z}$ in the metal. The values of γ_0 and γ_1 must come out of the solution.

(3) Variations in the y direction are as yet unknown and must be solved for.

Then, again representing all time variations by $e^{j\omega t}$, Maxwell's equations become:

Above the surface (Region O):

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} + \gamma_0 E_y &= -j\omega\mu_v H_x \\ -\gamma_0 H_x &= j\omega\epsilon_v E_y \\ -\frac{\partial H_x}{\partial y} &= j\omega\epsilon_v E_z \end{aligned}\right\} \quad (7\text{-}62a)$$

Within the conductor (Region I):

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} + \gamma_1 E_y &= -j\omega\mu_1 H_x \\ -\gamma_1 H_x &= (\sigma_1 + j\omega\epsilon_1) E_y \\ -\frac{\partial H_x}{\partial y} &= (\sigma_1 + j\omega\epsilon_1) E_z \end{aligned}\right\} \quad (7\text{-}62b)$$

combining gives

$$\frac{\partial^2 H_x}{\partial y^2} + \gamma_0^2 H_x = -\omega^2 \mu_v \epsilon_v H_x \qquad \frac{\partial^2 H_x}{\partial y^2} + \gamma_1^2 H_x = j\omega\mu_1(\sigma_1 + j\omega\epsilon_1) H_x$$

or

$$\frac{\partial^2 H_x}{\partial y^2} = h_0^2 H_x \qquad \frac{\partial^2 H_x}{\partial y^2} = h_1^2 H_x$$

where $h_0^2 = (-\gamma_0^2 - \omega^2\mu_v\epsilon_v)$

where $h_1^2 = (-\gamma_1^2 + \gamma_m^2)$
and $\gamma_m^2 = j\omega\mu_1(\sigma_1 + j\omega\epsilon_1)$

Solutions to these differential equations may be written as

$$H_x = C_1 e^{h_0 y} + C_2 e^{-h_0 y} \quad (7.63a) \qquad H_x = C_3 e^{h_1 y} + C_4 e^{-h_1 y} \quad (7.63b)$$

In taking the square root of h^2, if it is agreed that that root which has a positive real part will be used, then only the second term of (63a) need be considered. The first term represents a field which becomes infinitely large at $y = \infty$. Since this could not represent a physical field, this first term will be discarded by putting $C_1 = 0$. Similarly within the conductor, the second term represents a nonphysical field that becomes infinite at $y = -\infty$. Therefore C_4 can be put equal to zero. Showing the variations with time and in the z direction, the expressions for magnetic intensity can now be written

Above the surface:

$$H_x = C_2 e^{-h_0 y} e^{(j\omega t - \gamma_0 z)}$$

Below the surface:

$$H_x = C_3 e^{+h_1 y} e^{(j\omega t - \gamma_1 z)}$$

At the surface ($y = 0$), these expressions must be equal at all instants of time and for all values of z, because H_x must be continuous across the boundary. This requires that $C_2 = C_3$ and $\gamma_0 = \gamma_1$. Then the expressions for vertical and horizontal components of electric intensity can be written:

Above the surface:

$$E_y = \frac{-\gamma_0 C_2}{j\omega\epsilon_v} e^{-h_0 y} e^{(j\omega t - \gamma_0 z)}$$

$$E_z = \frac{h_0 C_2}{j\omega\epsilon_v} e^{-h_0 y} e^{(j\omega t - \gamma_0 z)}$$

In the conductor:

$$E_y = -\frac{\gamma_0 C_2}{\sigma_1 + j\omega\epsilon_1} e^{h_1 y} e^{(j\omega t - \gamma_0 z)}$$

$$E_z = -\frac{h_1 C_2}{\sigma_1 + j\omega\epsilon_1} e^{h_1 y} e^{(j\omega t - \gamma_0 z)}$$

At $y = 0$ the expressions for E_z must be equal. Therefore

$$\frac{h_0}{j\omega\epsilon_v} = \frac{-h_1}{\sigma_1 + j\omega\epsilon_1} \approx \frac{-h_1}{\sigma_1} \qquad \text{(for metallic conductors)}$$

$$h_0 = -\frac{j\omega\epsilon_v h_1}{\sigma_1} \qquad h_0^2 = \frac{-\omega^2\epsilon_v^2}{\sigma_1^2} h_1^2$$

$$\gamma_0^2 = -\omega^2\mu_v\epsilon_v - h_0^2 = -\omega^2\mu_v\epsilon_v + \frac{\omega^2\epsilon_v^2}{\sigma_1^2}(-\gamma_0^2 + \gamma_m^2)$$

From this,

$$\gamma_0 \approx \sqrt{-\omega^2\mu_v\epsilon_v\left(1 - \frac{j\omega\mu_1\epsilon_v}{\mu_v\sigma_1}\right)}$$

$$\approx \frac{j\omega}{c}\sqrt{1 - \frac{j\omega\mu_1\epsilon_v}{\mu_v\sigma_1}}.$$

For nonferrous metallic conductors $\mu_1 \approx \mu_v$, $\epsilon_1 \approx \epsilon_v$, so that

$$\gamma_0 = \gamma_1 \approx \frac{j\omega}{c}\sqrt{1 - \frac{j\omega\epsilon_v}{\sigma_1}}$$

$$h_0^2 \approx -\frac{\omega^2}{c^2}\left(\frac{j\omega\epsilon_v}{\sigma_1}\right) \qquad h_0 \approx -\frac{j\omega}{c}\sqrt{\frac{j\omega\epsilon_v}{\sigma_1}}$$

$$h_1 \approx \sqrt{j\omega\mu_v\sigma_1}$$

The resultant expressions for the fields in the two regions are

Above the conductor:

$$\left.\begin{aligned} H_x &= C_2 e^{-h_0 y} e^{(j\omega t - \gamma_0 z)} \\ E_y &= \frac{-\gamma_0}{j\omega\epsilon_v} H_x \approx -\eta_v H_x \\ E_z &= \frac{h_0}{j\omega\epsilon_v} H_x \approx -\eta_v\sqrt{\frac{j\omega\epsilon_v}{\sigma_1}} H_x \\ \frac{E_y}{E_z} &= -\frac{\gamma_0}{h_0} \approx \sqrt{\frac{\sigma_1}{\omega\epsilon_v}} \quad \underline{/-45^\circ} \end{aligned}\right\} \quad (7.64a)$$

Within the conductor:

$$\left.\begin{aligned} H_x &= C_2 e^{h_1 y} e^{(j\omega t - \gamma_0 z)} \\ E_y &\approx \frac{-j\omega\epsilon_v}{\sigma_1}\eta_v H_x \\ E_z &\approx -\sqrt{\frac{j\omega\mu_v}{\sigma_1}} H_x \\ \frac{E_y}{E_z} &\approx \sqrt{\frac{\omega\epsilon_v}{\sigma_1}} \quad \underline{/45^\circ} \end{aligned}\right\} \quad (7.64b)$$

It is seen that in the region of the air dielectric, outside the conductor, the electric intensity is almost normal to the surface. The field is elliptically polarized, the small horizontal component of **E** leading the vertical component by 45 degrees. Within the conductor the field is almost horizontal or parallel to the surface, the very small vertical component leading the horizontal component by 45 degrees.

The equiphase and equiamplitude surfaces can be obtained from the first of eqs. (64a and b). By letting

$$\gamma_0 = \alpha_0 + j\beta_0, \qquad h_0 = p_0 + jq_0, \qquad h_1 = p_1 + jq_1$$

these equations can be written as

$$H_x = C_2 e^{(-p_0 y - \alpha_0 z)} e^{j(\omega t - q_0 y - \beta_0 z)} \text{ (in the dielectric)}$$

and

$$H_x = C_2 e^{(p_1 y - \alpha_0 z)} e^{j(\omega t + q_1 y - \beta_0 z)} \text{ (in the conductor)}$$

Equiamplitude surfaces are obtained by setting the real exponents equal to a constant. This leads to

$$\tan \chi_0 = -\frac{y}{z} = \frac{\alpha_0}{p_0} \approx \sqrt{\frac{\omega\epsilon_v}{2\sigma_1}} \text{ (in the dielectric)} \qquad (7\text{-}65)$$

$$\tan \chi_1 = \frac{y}{z} = \frac{\alpha_0}{p_1} \approx \frac{\omega\epsilon_v}{\sigma_1}\sqrt{\frac{\omega\epsilon_v}{2\sigma_1}} \text{ (in the conductor)} \qquad (7\text{-}66)$$

Equiphase surfaces are obtained by setting the imaginary exponents equal to a constant. The slopes of the equiphase surfaces are given by

$$\tan \psi_0 = \frac{y}{z} = -\frac{\beta_0}{q_0} \approx \sqrt{\frac{2\sigma_1}{\omega\epsilon_v}} \text{ (in the dielectric)} \quad (7\text{-}67)$$

$$\tan \psi_1 = \frac{y}{z} = \frac{\beta_0}{q_1} \approx \sqrt{\frac{2\omega\epsilon_v}{\sigma_1}} \text{ (in the conductor)} \quad (7\text{-}68)$$

The angles χ_0, χ_1, ψ_0, and ψ_1 are shown in Fig. 7-11 where the equiphase lines are shown solid and the equiamplitude lines are shown dotted. In order to show the angles, their sizes have been very much exaggerated in

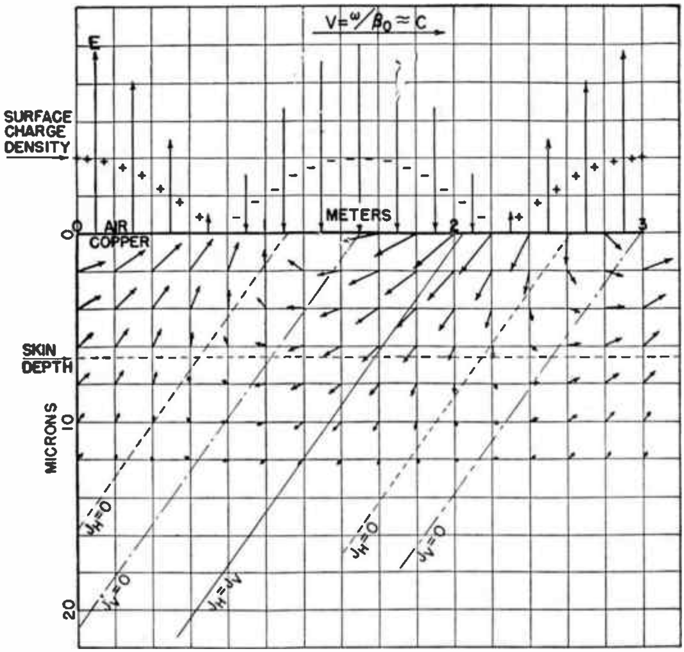


FIG. 7-12. Instantaneous current distribution within a copper conductor as a 100-mc wave is guided over its surface. (The vertical scale has been expanded by a factor of 10^5, and the vertical current scale is 10^5 times the horizontal current scale. Lengths of arrows refer to magnitudes at tail of arrow.)

this diagram. It is seen from eqs. (65) and (67) that in the dielectric the equiphase and equiamplitude surfaces are mutually perpendicular. In the conductor both equiamplitude and equiphase surfaces are nearly parallel to the surface, with the equiamplitude surface making a much smaller angle than the equiphase surface.

Within the conductor the conduction current density, given by

$$i = \sigma E$$

is seen to have both horizontal and vertical components. Using eqs. (64b) it is possible to make an instantaneous plot of current flow in the conductor as the electromagnetic wave is guided along its surface. This has been done in Fig. 7-12 for a copper conductor at 100 mc. In order to show the current flow adequately the vertical scale has been expanded by a factor of 100,000. The current magnitudes, indicated by the lengths of the arrows, are drawn to scale, but the vertical current scale is 10^5 times the horizontal current scale. Thus, if a horizontal current density of 1 *ampere* per square meter is represented by an arrow of unit length, an arrow of the same length in the vertical direction represents only 10 *microamperes* per square meter. It is apparent from the figure that the vertical currents are very small compared with the horizontal currents. However, it is these minute vertical currents that bring to the surface the charges on which the external electric flux terminates. Since total current normal to the surface must be continuous across the boundary surface, the vertical conduction current within the conductor at the surface is equal to the displacement current normal to the surface in the dielectric (the displacement current in the conductor is negligible). These vertical currents are a maximum at those places where the charge density on the surface is zero.

The plot of Fig. 7-12 is for a single instant of time. As time passes, the entire field configuration shown sweeps to the right with a velocity approximately equal to the velocity of light in free space.

PROBLEMS

1. A TEM wave is guided between two perfectly conducting parallel planes (Fig. 7-13). The frequency is 300 mc. Determine the voltage

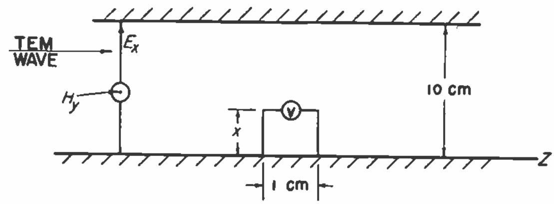


Fig. 7-13

reading of the (infinite impedance) voltmeter (a) by using Maxwell's electromotive force law (Faraday's induction law); (b) in terms of voltages induced in conductors which are parallel to the electric field.

2. Show that the field configuration of the $TE_{m,0}$ wave between parallel planes can be obtained by superposing two plane waves that are reflected back and forth between the walls of the guide as indicated in Fig. 7-6.

3. (a) Derive an expression for the attenuation factor for the $TM_{1,0}$ wave between parallel conducting planes.

(b) Verify that the attenuation is a minimum at a frequency which is $\sqrt{3}$ times the cut-off frequency.

4. For any uniform transmission line, for which R, L, C, and G per unit length are independent of position along the line (and, of course, independent of the magnitude of voltage and current), show that variation along the line of V and I can always be represented by an exponential law.

5. Use Maxwell's equations to show that it is impossible for the TEM wave to exist within any single-conductor wave guide (such as an ordinary rectangular or circular guide).

HINT: For $\oint \mathbf{H} \cdot d\mathbf{s}$ to have value in the transverse plane, there must be a longitudinal flow of current (conduction or displacement).

6. A plane wave propagating in a dielectric medium of permittivity ϵ_1 and permeability $\mu_1 = \mu_v$ is incident at an angle θ_1 upon a second dielectric of permittivity ϵ_2 and permeability $\mu_2 = \mu_v$. The wave is polarized parallel to the plane of incidence. Then, if the electric and magnetic intensities of the incident wave are E_1 and H_1, the component of E_1 parallel to the boundary surface will be $E_1 \cos \theta_1$ and the component of H_1 parallel to the surface will be H_1, so that the "wave impedance" of medium (1) in a direction normal to the surface would be $E_1 \cos \theta_1/H_1 = \eta_1 \cos \theta_1$. Similarly the "wave impedance" for the refracted ray in medium (2) in the direction normal to the surface would be $E_2 \cos \theta_2/H_2 = \eta_2 \cos \theta_2$. It would be expected when these impedances normal to the boundary surface are equal that there would be no reflection at the surface. Show that the condition that these impedances be equal is the same condition that led to the Brewster angle in eq. (5-73).

7. (a) In Chap. 9 (eq. 9-56), the expression for phase velocity in a rectangular guide of any cross-section is shown to be $\bar{v} = v_0/\sqrt{1 - \omega_c^2/\omega^2}$ where ω_c is a constant which depends upon the dimensions of the guide. Show that the group velocity defined by $v_g = d\omega/d\bar{\beta}$ is given by $v_g = v_0 \sqrt{1 - \omega_c^2/\omega^2}$.

(b) Using the definition $v_g = d\omega/d\bar{\beta}$, show that eq. (7-38) follows from (7-37).

BIBLIOGRAPHY

See bibliography for Chap. 4.

CHAPTER 8

TRANSMISSION LINES

8.01 Introduction. In the study of wave propagation between parallel planes, it was found that there were many possible modes or types of waves which could be propagated. Except for the special case of the transverse electromagnetic (TEM) wave, however, all of these modes require a certain minimum separation (in wavelengths) between the conductors for propagation to be possible. Only for the TEM wave could the conductor separation be small compared with a wavelength. This statement also holds for practical transmission lines, such as coaxial or parallel-wire lines, and it is for this reason that only the TEM mode need be considered at low frequencies, that is at power, audio, and radio frequencies below 200 or 300 mc. All other modes would require impractically large cross-sectional dimensions of the guiding systems. If a system of conductors guides this low-frequency-type TEM wave, it is called a transmission line, whereas if it supports TE or TM waves, it is called a wave guide. Transmission lines are considered in this present chapter and wave guides will be studied in chap. 9. Transmission lines always consist of at least two separate conductors between which a voltage can exist, but wave guides may, and often do, involve only one conductor; for example, a hollow rectangular or circular cylinder within which the wave propagates.

Although the TEM transmission line wave is but one special case of guided wave propagation, it is so important practically, that it is usually treated as "ordinary transmission line theory" quite early in the training of the electrical engineer. In this treatment, circuit concepts are extended to cover this distributed-constants circuit. It is the purpose of this chapter to show how the circuit approach follows directly from Maxwell's equations, and also to review briefly transmission line theory, especially as it applies in the case of low-loss lines. It will be found that many of the results and conclusions of ordinary transmission line theory may be applied with slight modifications to the more general cases of wave propagation. In particular, the concept of impedance developed in circuit theory can be carried over to transmission lines, and then extended to general wave propagation. This makes it possible, instead of working separately with V and I or E and H, to deal with their ratio, which is usually the important quantity in engineering.

Actual two-conductor transmission lines usually take the form of parallel-wire or coaxial lines. Before considering these practical cases, however, circuit concepts will be developed for the simpler case of a parallel plane transmission line carrying the TEM wave.

8.02 Circuit Representation of the Parallel-plane Transmission Line. In communication engineering a transmission line carrying the principal (TEM) wave is represented as a distributed-constants network having a series impedance $Z = R + j\omega L$ per unit length and a shunt admittance $Y = G + j\omega C$ per unit length. It is instructive to draw the equivalent circuit and evaluate the constants for the parallel plane transmission line of Fig. 7-1. For the special case of perfectly conducting planes and a perfect (lossless) dielectric, the series resistance and shunt conductance are both zero, so that the equivalent circuit representation is that of Fig. 8-1, where there is an inductance L per unit length and a capacitance C per unit length. The values of these constants in terms of the line dimensions and the constants of the medium between the planes can be obtained directly from Maxwell's equations.

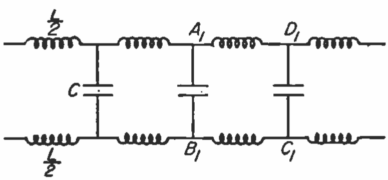


FIG. 8-1. Circuit representation of a lossless line.

Consider the various sections of a parallel-plane transmission line shown in Fig. 8-2. It is assumed that the line is carrying the TEM mode in the positive z direction, so that $\mathbf{E} = \mathbf{i}E_x$ and $\mathbf{H} = \mathbf{j}H_y$. The linear surface current density in the lower plane is $J_z = H_y$. The separation between the planes is a meters and, although they are infinite in extent in the y direction, a section b meters wide will be considered as being the transmission line. (By making this section a part of planes of infinite extent the field will not depend on y, and edge effects are eliminated.) Applying the emf equation to the closed path $ABCDA$

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\int_S \dot{\mathbf{B}} \cdot d\mathbf{a}$$

becomes

$$V_{AB} + V_{BC} + V_{CD} + V_{DA} = -j\omega B_y a \, \Delta z \tag{8-1}$$

where, as usual, time variations as $e^{j\omega t}$ are assumed. For perfectly conducting planes the tangential component of $\mathbf{E}$ is zero and so

$$V_{BC} = V_{DA} = 0$$

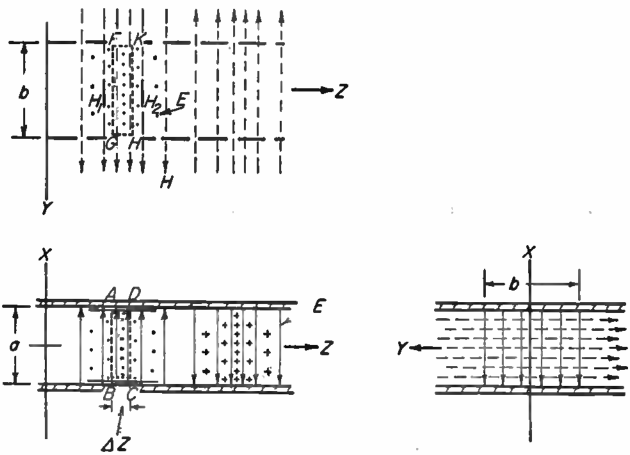


Fig. 8-2. Parallel-plane transmission line.

which leaves

$$V_{CD} - V_{BA} = -j\omega B_y a \, \Delta z$$

Dividing through by Δz and expressing in the differential form

$$\frac{dV}{dz} = -j\omega B_y a \tag{8-2}$$

It will be seen that

$$B_y = \mu H_y = \mu J_z = \frac{\mu I}{b}$$

where I is the current flowing in the strip of width b meters. Therefore, eq. (2) becomes

$$\frac{dV}{dz} = -\frac{j\omega\mu a}{b} I \tag{8-3}$$

Comparison with the ordinary circuital form of the transmission line equation

$$\frac{dV}{dz} = -j\omega LI \tag{8-4}$$

shows that for the parallel plane transmission line of width b meters and spacing a meters

$$L = \mu \frac{a}{b} \tag{8-5}$$

Similarly by writing the mmf equation for the path $FGHK$ in the y-z plane gives

$$bH_{FG} - bH_{KH} = j\omega\epsilon E_x b \, \Delta z \tag{8-6}$$

which becomes

$$\frac{d(bH_y)}{dz} = -j\omega\epsilon E_x b$$

Replace bH_y by $bJ_z = I$ and E_x by V/a. Then

$$\frac{dI}{dz} = -\frac{j\omega\epsilon b}{a} V \tag{8-7}$$

Comparison with the usual equation

$$\frac{dI}{dz} = -j\omega CV \tag{8-8}$$

shows that for the parallel-plane transmission line

$$C = \epsilon \frac{b}{a} \tag{8-9}$$

It is seen that for a parallel plane transmission line the inductance per unit length is simply the permeability μ of the medium multiplied by a geometry factor a/b, which in this case is proportional to the spacing and inversely proportional to the width of the line. Also the capacitance per unit length is the dielectric constant ϵ of the medium, multiplied by a geometry factor which, in this case, is proportional to the width and inversely proportional to the spacing. The reciprocal of the square root of the product of L and C gives the velocity of wave propagation along the line. That is,

$$v = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu\epsilon}} \tag{8-10}$$

For lines of different cross section the geometry factors will of course be different. However, since the velocity of propagation is given by $v = 1/\sqrt{LC}$ for all uniform unloaded lossless transmission lines, and since the velocity is independent of the line geometry (whether parallel-wire, coaxial, etc.), it follows that the geometry factors for L and C must always be reciprocal (for lossless lines). For example, for parallel-wire lines it was found in chap. 2 that the capacitance per unit length was

$$C = \frac{\pi\epsilon}{\ln \dfrac{b + \sqrt{b^2 - 4a^2}}{2a}} = \frac{\pi\epsilon}{\cosh^{-1}\left(\dfrac{b}{2a}\right)} \tag{8-11}$$

Therefore, the inductance per unit length must be

$$L = \frac{\mu \ln \dfrac{b + \sqrt{b^2 - 4a^2}}{2a}}{\pi} = \frac{\mu \cosh^{-1}\left(\dfrac{b}{2a}\right)}{\pi} \tag{8-12}$$

It is, of course, more than just a coincidence that the geometry factors for the L and C of a line are reciprocal. The significance of this relation is discussed in section (8.05).

A clear concept of the meaning of the permeability constant μ and dielectric constant ϵ is obtained from the parallel-plane transmission line of Fig. (8-2). If this line has unit width and unit separation, so that $a = b = 1$, then

$$L = \mu \qquad \text{and} \qquad C = \epsilon$$

Thus ϵ is the capacitance between conductors of 1 meter length of the parallel plane line, which is 1 meter wide and has a separation of 1 meter. Similarly, μ is the inductance per meter length of the same line. In terms of voltage and current, μ is a measure of the change per unit length of the transverse voltage when the current is changing at the rate of 1 amp/sec. Also the dielectric constant ϵ is a measure of the capacitive (displacement) current flow per unit length when the voltage between the planes is changing at the rate of 1 volt/sec. In terms of electric and magnetic fields, μ is a measure of the rate of change of E with distance owing to a change of H with time. Similarly, ϵ is a measure of the rate of change of H with distance owing to a change of E with time. Of course this is just the information conveyed by Maxwell's equations in their differential form.

In one respect the equivalent circuit representation of a transmission line may be misleading. In the equivalent circuit there exists a voltage drop $L(dI/dt)$ along each unit length of line. In the actual line, since E tangential to the surface of a perfect conductor is always zero, the voltage drop *along the surface of the line* is necessarily zero. Even if the conductors are imperfect so that an E parallel to the surface of the conductors is possible, the only voltage drop along the line would be that due to the current flow through the surface impedance, and this is ordinarily very small as has already been seen. The $L(dI/dt)$ drop in the equivalent circuit represents in the actual line the change per unit length of the *transverse* voltage between conductors. With a zero voltage drop along paths tangential to the (perfect) conductors, the difference of the transverse voltages AB and DC is equal to the induction voltage $-d\Phi/dt$ around the closed path $ABCDA$ (Fig. 8-2). But in the equivalent *circuit* representation of Fig. 8-1, where *fields are not considered,* the voltage around the closed path $A_1B_1C_1D_1A_1$ is zero. Therefore the induction voltage $-d\Phi/dt$ (which is responsible for the change in transverse voltage along the line) is shown as a series voltage, drop, $-L\, dI/dt$, across a lumped inductive reactance.

The characteristic impedance of the lossless parallel plane transmission line is

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{\mu}{\epsilon}}\frac{a}{b} = \eta\frac{a}{b} \qquad (8\text{-}13)$$

For the line of unit dimensions, $a = b = 1$, the characteristic impedance is just the intrinsic impedance of the dielectric medium between the plates.

8.03 Parallel-plane Transmission Lines with Loss. If the parallel plane transmission lines have loss, the results obtained above must be modified. The loss in the line will be due to the resistance of the conductors and to any conductivity of the dielectric between them. Again applying the electromotive force equation around the path $ABCDA$ of Fig. 8-2, the voltages V_{BC} and V_{DA} will now not be zero but will each have a value

$$V_{BC} = V_{DA} = J_z Z_s \Delta z$$

This is the voltage drop in length Δz of each conductor due to J_z flowing against the surface impedance Z_s. Then eq. (1) becomes

$$V_{CD} - V_{BA} = -j\omega B_y a\,\Delta z - 2J_z Z_s\,\Delta z \tag{8-14}$$

Writing $B_y = \mu H_y = \mu J_z = \mu I/b$, and putting (14) in the differential form,

$$\frac{dV}{dz} = -j\omega LI - Z'I = -(j\omega L + Z')I \tag{8-15}$$

where, as before, $$L = \frac{\mu a}{b}$$

and $$Z' = \frac{2Z_s}{b} \tag{8-16}$$

is the series impedance per unit length of the line (that is twice the surface impedance of a width b of each conductor). The impedance Z' is complex and can be written as $Z' = R' + j\omega L'$ where R' will be the series resistance per unit length and $j\omega L'$ will be the surface or internal reactance per unit length. Then eq. (15) can be written

$$\frac{dV}{dz} = -[R' + j\omega(L' + L)]I \tag{8-15a}$$

If the dielectric between the conducting plates is not perfect, but has a value σ, then there will be a transverse conduction current density $\sigma\mathbf{E}$, which will modify the magnetomotive force around the rectangle $FGHK$. Instead of (6) the mmf equation will now be

$$-(bH_{KH} - bH_{FG}) = (\sigma E_x + j\omega\epsilon E_x)b\,\Delta z$$

Then $$\frac{d(bH_y)}{dz} = -b(\sigma + j\omega\epsilon)E_x$$

Replacing bH_y by $bJ_z = I$ and E_x by $\dfrac{V}{a}$

$$\frac{dI}{dz} = -\left(\frac{b\sigma}{a} + j\frac{\omega\epsilon b}{a}\right)V$$
$$= -(G + j\omega C)V \tag{8-17}$$

where $C = \epsilon b/a$ is the capacitance per unit length and $G = b\sigma/a$ is the conductance per unit length of line.

Equations (15) and (17) are in the circuital form, familiar to engineers, and may be solved to yield the well-known "transmission line equations." Before carrying out the solution, Maxwell's equations will be applied to two practical transmission lines, the coaxial line and the parallel-wire line, to show the development of the same relations [eqs. (15) and (17)] for them.

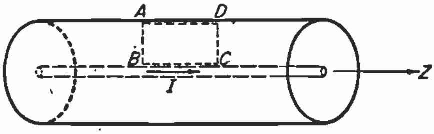


FIG. 8-3. Coaxial transmission line.

8.04 Coaxial and Parallel-wire Lines. The circuit constants for the equivalent circuit of a coaxial or parallel-wire line can be obtained in the same manner as in the case of parallel planes. In the coaxial line of Fig. 8-3 we can apply Maxwell's emf equation to the closed path $ABCDA$, for which $AD = BC$ has unit length.

$$V_{AB} + V_{BC} + V_{CD} + V_{DA} = -\frac{d\Phi}{dt} = -j\omega\Phi$$

$$V_{CD} - V_{BA} = -j\omega\Phi - Z'I - Z''I \quad (8\text{-}18)$$

where IZ' and IZ'' are the voltage drops per unit length along the inner and outer conductors, respectively. For perfect conductors these would be zero. If the magnetic flux per unit length of line is related to I by

$$\Phi = LI \quad (8\text{-}19)$$

eq. (18) may be written

$$\frac{dV}{dz} = -(Z' + Z'' + j\omega L)I \quad (8\text{-}20)$$

Z' and Z'' are the surface or internal impedances per unit length of the inner and outer conductors. If the depth of penetration is small compared with the radii of the conductors, these are given by

$$Z' = \frac{Z_s}{2\pi a} \qquad Z'' = \frac{Z_s}{2\pi b}$$

where Z_s is the surface impedance of a plane conductor of unit length and unit width, and a and b are the radii of the inner and outer conductors. The resistance per unit length of line will be the real part of the sum of Z' and Z''. That is

$$R = \sqrt{\frac{\omega\mu}{2\sigma}}\left(\frac{1}{2\pi a} + \frac{1}{2\pi b}\right)$$

$$= \sqrt{\frac{f\mu}{4\pi\sigma}}\left(\frac{1}{a} + \frac{1}{b}\right) \qquad \text{ohms/m} \quad (8\text{-}21)$$

The surface or internal reactance of the conductors will have this same value. The internal reactance of the conductors should be added to the external reactance $j\omega L$ to obtain the total reactance per unit length. The voltage equation may be written

$$\frac{dV}{dz} = -(R + j\omega L)I \tag{8-20a}$$

where R is given by (21) and ωL is the sum of the external inductive reactance and the surface reactance just determined.

To obtain the current equation apply Maxwell's emf equation to the closed path $ABCDEFA$ on the surface of the inner conductor and let the length $FA = DC$ be unity (Fig. 8-4). Designating the magnetomotive force by $\mathfrak{F}$,

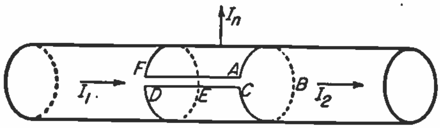


FIG. 8-4

$$\mathfrak{F}_{ABC} + \mathfrak{F}_{CD} + \mathfrak{F}_{DEF} + \mathfrak{F}_{FA} = I_n \tag{8-22}$$

I_n is the current normal to the surface enclosed by the path $ABCD$-EFA; that is, I_n is the transverse current per unit length from the inner to the outer conductor. In general, I_n will consist of a leakage or conduction current, I_c, proportional to the voltage, and a displacement current, I_d, proportional to the rate of change of voltage. Let G and C be proportionality factors such that

$$I_c = GV \qquad \text{and} \qquad \Psi = CV \tag{8-23}$$

Ψ is the electric displacement from the surface enclosed by the path, and the displacement current will be

$$I_d = \frac{d\Psi}{dt} = C\frac{dV}{dt}$$

Therefore, the right-hand side of (22) becomes

$$I_n = (G + j\omega C)V$$

Considering terms on the left-hand side of (22),

$$\mathfrak{F}_{CD} = -\mathfrak{F}_{FA} \qquad \mathfrak{F}_{DEF} = I_1 \qquad \mathfrak{F}_{CBA} = I_2$$

so that the left-hand side reduced to

$$\mathfrak{F}_{ABC} + \mathfrak{F}_{DEF} = I_1 - I_2$$
$$= -\frac{dI}{dz}$$

The current equation is then

$$\frac{dI}{dz} = -(G + j\omega C)V \qquad (8\text{-}24)$$

Equations (20a) and (24) are the familiar circuital form of the transmission line equations. L and C are the inductance and capacitance per unit length of line. For a coaxial line, having perfect conductors, L and C are defined by (19) and (23). For conductors having large but finite conductivity, the value of L will be slightly greater than that obtained for the ideal case, although the difference is usually negligible for efficient (low-loss) transmission lines.

The equivalent circuit and differential equation for a parallel-wire line are derived in a similar manner. It is left for the student to carry this through, and to derive the expressions for L and C for this case.

8.05 E and H about Long Parallel Cylindrical Conductors. In section 8.02 it was found that the geometry factors for the L and C of parallel perfectly conducting cylinders were always reciprocal. As might be suspected, this interesting result is not just a coincidence, but follows as a logical consequence of the similarity that exists between all two-dimensional electric and magnetic field distributions. It is well known that lines of **E** and **H** about long parallel circular cylinders are always orthogonal, and that the magnitudes of **E** and **H** are related at all points by a constant factor that is dependent on the charge on the conductors and the current flowing through them. It is easy to show that this same correspondence between electric and magnetic fields must hold even in the more

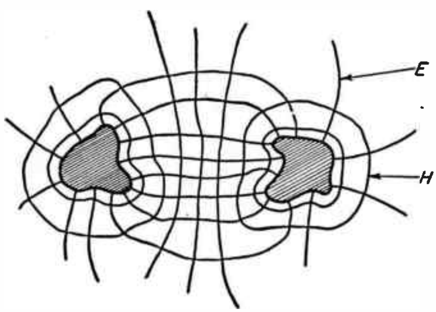


FIG. 8-5. Parallel cylinders of arbitrary cross section.

general case where the parallel cylinders have any arbitrary cross section as in Fig. 8-5.

The static electric field configuration is obtained as a solution to Laplace's equation subject to the boundary conditions of the problem. In rectangular co-ordinates, for two dimensional fields that are independent of the z co-ordinate, Laplace's equation is

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \qquad (8\text{-}25)$$

where V is the (electrostatic) potential, the gradient of which gives the electric field. Similarly, the magnetic field configuration can be obtained as the curl of a magnetic (vector) potential that has the direction of the current producing it. When the conductors are entirely in one direction, say the z direction, the vector potential has only one component A_z, and the components of magnetic intensity lie in the x-y plane and are given by

$$H_x = \frac{\partial A_z}{\partial y} \qquad H_y = -\frac{\partial A_z}{\partial x} \qquad (8\text{-}26)$$

Under these conditions it can be readily shown that A_z (which now may be treated as a scalar quantity) also satisfies eq. (25). In a region in which there are no currents, Ampere's law indicates that the line integral of $\mathbf{H}$ around every closed path is zero. That is $\oint \mathbf{H} \cdot d\mathbf{s} = 0$. The differential vector statement of this law is

$$\text{curl } \mathbf{H} = 0 \qquad (8\text{-}27)$$

For this case where there is no z component of $\mathbf{H}$, relation (27) becomes

$$\frac{\partial H_x}{\partial y} - \frac{\partial H_y}{\partial x} = 0$$

Inserting relations (26) gives

$$\frac{\partial^2 A_z}{\partial x^2} + \frac{\partial^2 A_z}{\partial y^2} = 0 \qquad (8\text{-}28)$$

Thus for two-dimensional magnetic fields the potential A_z satisfies Laplace's equation, and the configuration of the magnetic field, obtained from (26), is always such that relation (28) is satisfied. In addition, of course, the boundary conditions of the particular problem must also be satisfied.

Now consider the problem of two parallel cylindrical conductors (assumed perfectly conducting) that carry equal and oppositely directed currents I. For the d-c case the current is uniformly distributed throughout the conductor, but for rapidly alternating fields, because of the phenomenon of skin effect, the current exists only near the surface of the conductor. Although this is a d-c field analysis, the results will be applied chiefly to the alternating field cases, so the assumption of current concentrated in a thin sheet at the surface of the conductor will be used. Except when the spacing between conductors is large compared with their diameters, the current distribution around the circumference of the conductor will not be uniform. The actual current distribution will be such that the boundary conditions at the surface of the conductor are satisfied. This is similar to the electrostatic problem where the charge distribution around the cylinders was such as to make the cylinders equipotential surfaces, and satisfy the condition that $E_{\text{tang}} = 0$.

The corresponding boundary condition for the magnetic intensity is that $H_{\text{norm}} = 0$ (for a perfect conductor). That is, the magnetic intensity at the surface is entirely *tangential.* Equations (26) indicate that if the magnetic intensity normal to the surface is zero, there can be no change of A_z in a direction tangential to the surface. Therefore the conductor must also be an "equipotential" surface for the magnetic potential A_z. Because in this case both A_z and V satisfy Laplace's equation, and in addition satisfy the same boundary conditions, it follows that the expressions for A_z and V will always be identical, except for some constant factor. Then, because

$$E_x = -\frac{\partial V}{\partial x} \qquad H_x = \frac{\partial A_z}{\partial y}$$

$$E_y = -\frac{\partial V}{\partial y} \qquad H_y = -\frac{\partial A_z}{\partial x} \tag{8-29}$$

it follows that $\mathbf{E}$ and $\mathbf{H}$ will always be orthogonal, and that their magnitudes will be related to each other by the same factor that related V and A_z.

The electric and magnetic field configuration obtained from solutions of Laplace's equation are for the electrostatic and steady current cases, respectively. In general, it would not be expected that these same solutions would hold for alternating fields, especially at high frequencies. It turns out that for two-dimensional fields however, the field configurations obtained for the static cases also hold for the alternating cases. This is because the general Maxwell emf and mmf equations reduce to their steady-field counterparts in the two-dimensional case. For example, the mmf equation in the x-y plane for the region outside the conductors (no conduction current) is

$$\text{curl}_z\, \mathbf{H} = \epsilon \dot{E}_z \tag{8-30}$$

But for two dimensional fields in the x-y plane, E_z is zero so that for any path in this plane this equation reduces to eq. (27), which yielded the Laplace equation (28). Similarly the Maxwell emf equation

$$\text{curl}_z\, \mathbf{E} = -\mu \dot{H}_z \tag{8-31}$$

reduces to

$$\text{curl}_z\, \mathbf{E} = 0 \tag{8-32}$$

for the two-dimensional problem. Equation (32), the integral form of which is

$$\oint \mathbf{E} \cdot ds = 0 \tag{8-33}$$

states that for any path in the x-y plane the electric field is conservative. Therefore, $\mathbf{E}$ is derivable as the gradient of a scalar potential V, and in a region in which there are no charges, the relation

$$\text{div}\, \mathbf{E} = 0$$

leads directly to Laplace's equation.

8.06 Transmission Line Theory. The differential equations (20a) and (24) relating voltage and current along a transmission line may be solved to yield the transmission line equations.

$$\frac{\partial V}{\partial z} = -(R + j\omega L)I \tag{8-20a}$$

$$\frac{\partial I}{\partial z} = -(G + j\omega C)V \tag{8-24}$$

Differentiating and combining gives

$$\frac{\partial^2 V}{\partial z^2} = \gamma^2 V \tag{8-34}$$

$$\frac{\partial^2 I}{\partial z^2} = \gamma^2 I \tag{8-35}$$

where $\gamma^2 = (R + j\omega L)(G + j\omega C)$

Solutions to eqs. (34) and (35) may be written in either exponential- or hyperbolic-function form. In the exponential form, viz.,

$$V = V' e^{-\gamma z} + V'' e^{+\gamma z} \tag{8-36}$$

$$I = I' e^{-\gamma z} + I'' e^{+\gamma z} \tag{8-37}$$

the solutions are shown as the sum of two waves, one traveling in the positive z direction and the other traveling in the negative z direction. The ratio of voltage to current for the wave traveling in the positive z direction is

$$\frac{V'}{I'} = Z_0 \tag{8-38}$$

whereas for the "reflected" wave traveling in the opposite direction

$$\frac{V''}{I''} = -Z_0 \tag{8-39}$$

Z_0 is the characteristic impedance of the line and is related to the so-called primary constants R, L, C, and G by

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{8-40}$$

If the line is terminated in an impedance Z_R located at $z = 0$, the ratio of V to I at this point will be equal to Z_R so that

$$Z_R = \frac{V}{I} = \frac{V' + V''}{I' + I''} = \frac{Z_0(I' - I'')}{I' + I''}$$

These relations can be recombined to give the *reflection coefficients*,

$$\frac{V''}{V'} = \frac{Z_R - Z_0}{Z_R + Z_0} \qquad \frac{I''}{I'} = \frac{Z_0 - Z_R}{Z_0 + Z_R} \tag{8-41}$$

In the hyperbolic-function form the solutions to (34) and (35) are

$$V = A_1 \cosh \gamma z + B_1 \sinh \gamma z$$
$$I = A_2 \cosh \gamma z + B_2 \sinh \gamma z \tag{8-42}$$

The constants A_1, A_2, B_1 and B_2 are evaluated by applying the boundary conditions. Let

$$V = V_R, \qquad I = I_R \qquad \text{at } z = 0$$
$$V = V_s, \qquad I = I_s \qquad \text{at } z = z_1$$

substituting these relations in (42) and using eqs. (20a) and (24),

$$V_s = V_R \cosh \gamma z_1 - Z_0 I_R \sinh \gamma z_1$$

$$I_s = I_R \cosh \gamma z_1 - \frac{V_R}{Z_0} \sinh \gamma z_1 \qquad (8\text{-}43)$$

It is usual to make the location of the terminating impedance Z_R the reference point ($z = 0$), and to consider the sending end as being to the left of this reference point, that is, in the $-z$ direction as in Fig. 8-6. Then letting $l = -z_1$, eqs. (43) become

$$V_s = V_R \cosh \gamma l + Z_0 I_R \sinh \gamma l \qquad (8\text{-}44)$$

$$I_s = I_R \cosh \gamma l + \frac{V_R}{Z_0} \sinh \gamma l \qquad (8\text{-}45)$$

where l is measured from the receiving end of the line.

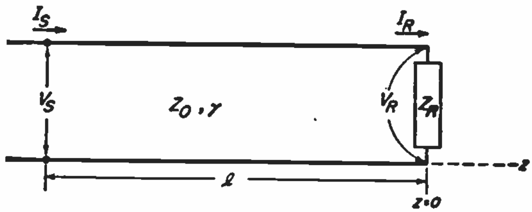


FIG. 8-6

These are the general transmission line equations that relate the voltages and currents at the two ends of the line. The general expression for the input impedance of the line is obtained by dividing (44) by (45), that is,

$$Z_{\text{in}} = \frac{V_s}{I_s} = \frac{V_R \cosh \gamma l + Z_0 I_R \sinh \gamma l}{I_R \cosh \gamma l + (V_R/Z_0) \sinh \gamma l} \qquad (8\text{-}46)$$

Certain special cases are of interest. For a line short-circuited at the receiving end, $Z_R = 0$, and therefore $V_R = 0$, and the input impedance is

$$Z_{sc} = Z_0 \tanh \gamma l \qquad (8\text{-}47)$$

On the other hand, for an open-circuited line $Z_R = \infty$, $I_R = 0$, so that the input impedance is

$$Z_{\infty} = Z_0 \coth \gamma l \qquad (8\text{-}48)$$

8.07 Low-loss Radio Frequency and UHF Transmission Lines. The low-loss transmission line is of special interest to the engineer concerned with the transmission of energy at radio and ultrahigh frequencies. There are two reasons for this. First, most practical lines designed for use at these frequencies will be low-loss lines. Second, at ultrahigh frequencies, sections of low-loss line are used as circuit elements, and a knowledge of the operation of such "distributed-constants circuits" is of considerable importance.

A *low-loss transmission line* is one for which

$$\begin{aligned} R &\ll \omega L \\ G &\ll \omega C \end{aligned} \tag{8-49}$$

where R, L, C, and G are the resistance, inductance, capacitance, and conductance per unit length of the line. When the above inequalities hold, the following approximations are valid:

$$Z = R + j\omega L \approx j\omega L$$

$$Y = G + j\omega C \approx j\omega C$$

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \tag{8-50}$$

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \approx j\omega \sqrt{LC} \tag{8-51}$$

Since $\gamma = \alpha + j\beta$, this last expression gives

$$\alpha \approx 0 \tag{8-52}$$

$$\beta \approx \omega \sqrt{LC} \tag{8-53}$$

The approximation for β is very good for low-loss lines, but occasionally the approximation of zero for α may not be good enough, even though α is very small compared with β. A closer approximation for α may be obtained by rearranging the expression for γ and using the binomial expansion. Thus

$$\begin{aligned} \gamma &= j\omega \sqrt{LC} \sqrt{\left(1 + \frac{R}{j\omega L}\right)\left(1 + \frac{G}{j\omega C}\right)} \\ &\approx j\omega \sqrt{LC} \left(1 + \frac{R}{2j\omega L}\right)\left(1 + \frac{G}{2j\omega C}\right) \\ &\approx j\omega \sqrt{LC} \left(1 + \frac{R}{2j\omega L} + \frac{G}{2j\omega C}\right) \\ &\approx \frac{R}{2\sqrt{L/C}} + \frac{G\sqrt{L/C}}{2} + j\omega \sqrt{LC} \end{aligned}$$

which gives
$$\alpha \approx \frac{1}{2}\left(\frac{R}{Z_0} + GZ_0\right) \qquad (8\text{-}54)$$
$$\beta \approx \omega \sqrt{LC} \qquad (8\text{-}55)$$

The more correct value for α given by (54) need only be used in place of (52) when the line losses are being considered. As far as voltage and current distributions are concerned, the attenuation of most low-loss ultrahigh frequency lines is so small that the approximation $\alpha = 0$ gives satisfactory results. This may seem strange in view of the fact that R, and therefore α, *increases* with frequency, and α is not usually neglected at low (power and audio) frequencies. The explanation for this apparent paradox is that although α, the attenuation *per unit length*, increases approximately as the square root of frequency, the attenuation *per wavelength* decreases as the square root of the frequency. Transmission lines are ordinarily a few wavelengths long at most, and αl can usually be neglected (compared with βl) at the ultrahigh frequencies. Thus for many purposes, low-loss lines may be treated as though they were lossless; that is, as if $R = G = \alpha = 0$.

Using the approximate values for the secondary constants given by (50), (51), (52), and (53), the general transmission lines become for this low-loss, high-frequency case

$$V_s = V_R \cos \beta l + jI_R Z_0 \sin \beta l \qquad (8\text{-}56)$$
$$I_s = I_R \cos \beta l + j\frac{V_R}{Z_0} \sin \beta l \qquad (8\text{-}57)$$

where now $Z_0 \approx \sqrt{L/C}$ is a pure resistance.

The input impedance of such a line is

$$\begin{aligned} Z_s &= \frac{V_s}{I_s} \\ &= Z_R\left(\frac{\cos \beta l + j(Z_0/Z_R) \sin \beta l}{\cos \beta l + j(Z_R/Z_0) \sin \beta l}\right) \\ &= Z_0\left(\frac{Z_R \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_R \sin \beta l}\right) \qquad (8\text{-}58) \end{aligned}$$

The voltage and current distributions along the line are obtained from eqs. (56) and (57) by replacing l, the length of line, by x, the distance from the terminating impedance Z_R. Since voltmeters and ammeters read magnitude without regard to phase, the absolute magnitude of expressions (56) and (57) have been used in Fig. 8-7 to show the standing wave distributions for various conditions of the terminating impedance Z_R.

In general the terminating impedance Z_R will be a complex impedance having both resistance and reactance, but it will be shown later that the results for the general case may be inferred

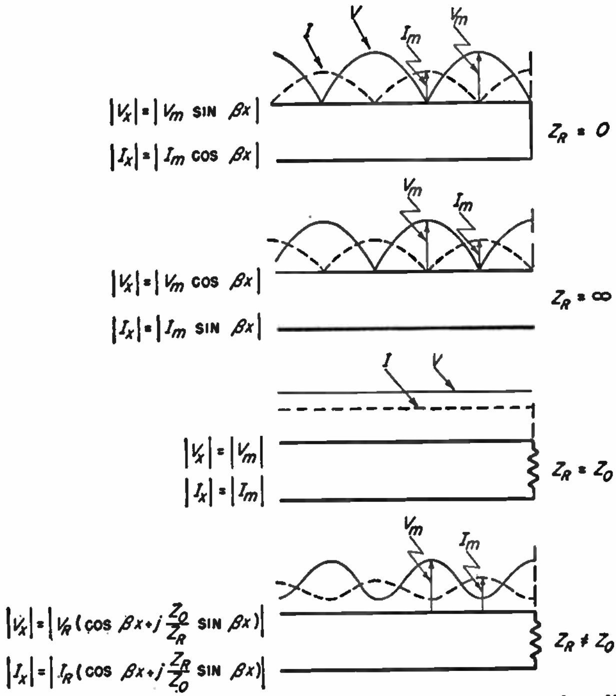


FIG. 8-7. Voltage and current distribution along a lossless line.

from those obtained for the particular case of a pure resistance termination. For this latter case where $Z_R = R$, eqs. (56) and (57) may be written as

$$|V_x| = V_R \sqrt{\cos^2 \beta x + (R_0/R)^2 \sin^2 \beta x} \tag{8-59}$$

$$|I_x| = I_R \sqrt{\cos^2 \beta x + (R/R_0)^2 \sin^2 \beta x} \tag{8-60}$$

For the lossless line being considered Z_0 is a pure resistance

$$Z_0 = R_0 = \sqrt{\frac{L}{C}}$$

Examination of eqs. (59) and (60) shows that the voltage and current distributions are given by the square root of the sum of a cosine-squared term and a sine-squared term. It is evident that the maximum value of voltage or current will occur at that value of x that makes the larger of these terms a maximum. In the particular case of a line terminated in R_0, that is, for which $R = R_0$, the sine and cosine terms have equal amplitudes and the square root of the sum of their squares has constant value for all values of x. That is, there are no standing waves on the line. For all other cases, however, the magnitude will vary along the length of the line. When R is less than R_0, the amplitude of the sine terms of (59) will be larger than that of the cosine term and the voltage maxima will occur at those values of x that make sin βx a maximum, viz., at $x = \lambda/4$, $3\lambda/4$, and so on. Also the voltage minima will occur at those values of x that make the sine term a minimum, viz., $x = 0$, $\lambda/2$, and so on, also for this case of $R < R_0$, the *current* maxima will occur at $x = 0$, $\lambda/2$, and so on, and the current minima at $x = \lambda/4$, $3\lambda/4$, and so on. When the terminating resistor is larger than R_0, the conditions for both voltage and current are reversed.

One of the important measurable quantities on a transmission line is the standing-wave ratio of voltage or current. When R is less than R_0, eq. (59) shows that the voltage maximum, which occurs when $\sin \beta x = 1$, will have a value

$$V_{max} = V_R \cdot \frac{R_0}{R}$$

Also the voltage minimum, which occurs when $\sin \beta x = 0$, will have a value

$$V_{min} = V_R$$

The ratio of maximum voltage to minimum voltage is therefore

$$\frac{V_{max}}{V_{min}} = \frac{R_0}{R} \qquad \text{(for } R < R_0\text{)}$$

Similarly the standing wave of current ratio is given by

$$\frac{I_{max}}{I_{min}} = \frac{R_0}{R} \qquad \text{(for } R < R_0\text{)}$$

For $R > R_0$ these expressions are just reversed, that is

$$\frac{V_{max}}{V_{min}} = \frac{I_{max}}{I_{min}} = \frac{R}{R_0} \qquad \text{(for } R > R_0\text{)}$$

Using these expressions, the value of a terminating resistance may be determined in terms of R_0 from relative measurements of voltage or current along the line. R_0 is readily calculable from the line dimensions.

Case where Z_R is not a Pure Resistance. When the terminating impedance Z_R is *not* a pure resistance, standing-wave measurements can be still used, and in this case will yield values of both resistance and reactance of the termination. From eqs. (59) and (60) it was seen that with a resistance termination a voltage maximum or minimum always occurred right at the termination ($x = 0$). However,

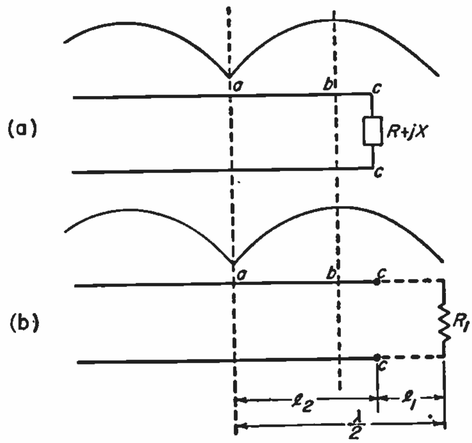


FIG. 8-8. A complex terminating impedance in (a) is replaced by a pure resistance termination in (b).

when the terminating impedance has reactance as well as resistance, the maximum or minimum is always displaced from the position $x = 0$, and the direction and amount of this displacement can be used to determine the sign and magnitude of the reactance of the load.

Figure 8-8 shows a transmission line terminated in an impedance that has a reactive component. The voltage distribution along the line is shown. Because the impedance is not a pure resistance, the voltage maximum (or minimum) does not occur at the termination. Now any complex impedance can be obtained by placing a pure resistance of proper value at the end of an appropriate length of (lossless) transmission line. In part (b) of Fig. 8-8, the complex impedance $R + jX$ has been replaced by the proper value of resistance R_1 at the end of a length l_1 of line, such that the impedance at $c - c$ looking towards R_1 is equal to $R + jX$. The standing wave back from $c - c$ toward the source will be unchanged and that toward R_1 (shown dotted) will be just a continuation of it. Quite evidently the proper position for R_1 is at a distance of one-half wavelength from the minimum point a (or the maximum point b if R_1 is greater than R_0), and the proper value of R_1 is given by the standing-wave ratio on the line, that is, by

$$\frac{R_1}{R_0} = \frac{V_{\min}}{V_{\max}} \qquad \text{or} \qquad \frac{R_1}{R_0} = \frac{V_{\max}}{V_{\min}}$$

Because any resistance greater than R_0 can be obtained by a resistance less than R_0 at the end of a quarter wave section of line (see below), it is really only necessary to consider for R_1 resistances less than or equal to R_0. It is then possible to state that any impedance whatsoever can be obtained by means of a pure resistance R_1 (not greater than R_0) at the end of a length l_1 of lossless transmission line, less than one half wavelength long.

The value of the impedance $Z = R + jX$ is given in terms of R_1 and l_1 by eq. (58). Rationalizing and separating into real and imaginary parts, eq. (58) becomes

$$R = \frac{R_0^2 R_1}{R_0^2 \cos^2 \beta l_1 + R_1^2 \sin^2 \beta l_1} \tag{8-61}$$

$$X = \frac{R_0(R_0^2 - R_1^2) \sin \beta l_1 \cos \beta l_1}{R_0^2 \cos^2 \beta l_1 + R_1^2 \sin^2 \beta l_1} \tag{8-62}$$

Equations (61) and (62) make it possible to determine both the resistance and reactance values of a terminating impedance from standing-wave measurements on the transmission line. The sign of the reactance, that is, whether inductive (positive) or capacitive (negative) can be obtained by inspection as shown in Fig. 8-9.

Considering the value of R_1 to be less than R_0, eq. (62) shows that when l_1 is less than one-quarter wavelength, the reactance X is positive (i.e., inductive), whereas if l_1 is between one-quarter and one-half wavelength, X will be negative (capacitive). From this results the conclusion that, if the standing wave of voltage slopes down toward the terminating impedance (Fig. 8-9a), the impedance is inductive; if the slope is up toward the impedance

(Fig. 8-9b), the impedance is capacitive. Of course, if the slope is zero at the termination, the terminating impedance is a pure resistance.

In practice the measurable quantities are l_2, the distance from the termination to the minimum point a, and the *standing-wave ratio*

$$\rho = \frac{R_0}{R_1} = \frac{V_{\max}}{V_{\min}}$$

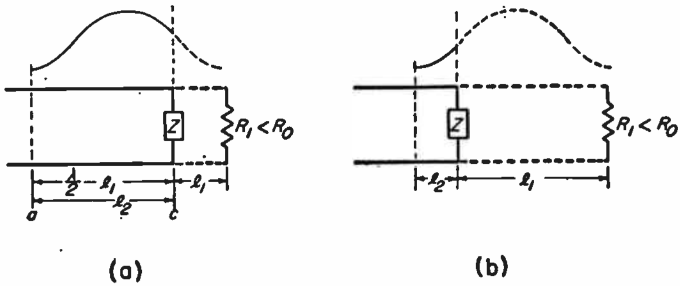


FIG. 8-9. A terminating impedance that is inductive (a) or capacitive (b).

In terms of these measurable quantities, the resistance and reactance of the terminating impedance is given by

$$R = \frac{\rho R_0}{\rho^2 \cos^2 \beta l_2 + \sin^2 \beta l_2} \tag{8-63}$$

$$X = \frac{-R_0(\rho^2 - 1) \sin \beta l_2 \cos \beta l_2}{\rho^2 \cos^2 \beta l_2 + \sin^2 \beta l_2} \tag{8-64}$$

8.08 UHF Lines as Circuit Elements. The transfer of energy from one point to another is only one use of transmission lines. At the ultrahigh frequencies an equally important application is the use of sections of lines as circuit elements. Above 150 mc the ordinary lumped-circuit elements become difficult to construct and, at the same time, the required physical size of sections of transmission lines has become small enough to warrant their use as circuit elements. They can be used in this manner up to about 3000 mc where their physical size then becomes too small and wave guide technique begins to take over.

In Fig. 8-10 are shown some line sections and their low-frequency equivalents. The magnitude of the input reactance of the first four of these sections is given by eq. (58) when the appropriate value of Z_R is inserted; that is, $Z_R = 0$ for the shorted section and $Z_R = \infty$ for the open sections. The resistive component of the input impedance is negligible for the usual low-loss lines used at UHF. Thus it is seen that for line lengths less than a quarter of a wavelength the shorted section is equivalent to an inductance, and the open section to a capacitance. For length of line between a quarter and a half wavelength, the shorted section is capacitive and the open section is inductive. However, it should be noted that

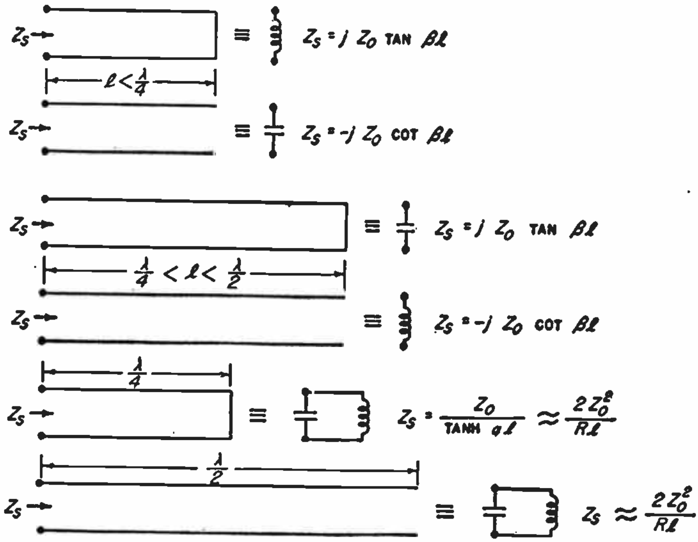


FIG. 8-10. Input impedance of various transmission line sections.

unlike their low-frequency equivalents, these "inductances" and "capacitances" change value with frequency.

The Quarter-wave and Half-wave Sections. For the particular case of the shorted quarter-wave line or the open half-wave line, the input reactance, given by (58), goes to infinity, and the resistive component of the input impedance must be taken into account. This corresponds to conditions in the parallel-resonant circuit (the low-frequency analogue), which has an infinite impedance if resistance is neglected. In both cases (the quarter-wave line and the parallel-resonant circuit) the actual input impedance when the series resistance is not neglected is a pure resistance of very high value. In the case of the line its value is given approximately by

$$R_{ar} \approx \frac{2Z_0^2}{Rl}$$

where R_{ar} is the input resistance of the line at a resonant length and R is the series resistance per unit length of the line. l is the length of the resonant section, which will be an odd multiple of a quarter wavelength for a shorted line or an even multiple of one-quarter wavelength for an open line. This expression is obtained directly from eqs. (44) and (45) in which the actual line loss is not neglected as follows:

For a shorted line for which $V_R = 0$, eqs. (44) and (45) become

$$V_s = I_R Z_0 \sinh \gamma l$$
$$I_s = I_R \cosh \gamma l$$

Dividing the voltage equation by the current equation gives the input impedance of a short-circuited line as

$$Z_s = Z_0 \tanh \gamma l$$
$$= Z_0 \cdot \frac{\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l}{\cosh \alpha l \cos \beta l + j \sinh \alpha l \sin \beta l}$$

For line lengths that are an odd multiple of a quarter wavelength, $\sin \beta l = \pm 1$ and $\cos \beta l = 0$. Under these conditions the input impedance becomes

$$Z_s = Z_0 \frac{\cosh \alpha l}{\sinh \alpha l}$$

If αl is very small, as is generally true for sections of low-loss line, $\cosh \alpha l \approx 1$ and $\sin \alpha l \approx \alpha l$ so that

$$Z_s \approx \frac{Z_0}{\alpha l}$$

When $\omega L \gg R$ and $\omega C \gg G$, α is given in terms of the line constants by

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{8-54}$$

For the air dielectric lines commonly used the losses due to the conductance G are negligible, so that G can be neglected and

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} = \frac{R}{2Z_0} \tag{8-65}$$

Substituting this in the above expression for input impedance of a short-circuited line, whose length is an odd multiple of a quarter wavelength, gives

$$Z_s = \frac{Z_0}{\alpha l} = \frac{2Z_0^2}{Rl} \tag{8-66}$$

An identical expression is obtained for an open-ended section that is a multiple of a half wave long.

Resonance in Line Sections. The shorted quarter-wave section has other properties of the parallel-resonant circuit. It is a *resonant* circuit and produces the resonant rise of voltage or current which exists in such circuits. The mechanism of resonance is particularly easy to visualize in this case. If it is assumed that a small voltage is induced into the line near the shorted end, there will be a voltage wave sent down the line and reflected without change of phase at the open end. This reflected wave travels back and is reflected again at the shorted end with reversal of phase. Because it required one-half cycle to travel up and back the line, this twice-reflected wave now will be in phase with the original induced voltage and so adds directly to it. Evidently those additions continue to increase the voltage (and current) in the line until the I^2R loss is equal to the power being put into the line. A voltage step-up of several hundred times is possible depending upon the Q of the line.

Input Impedance of the Tuned Line. When the quarter-wave section is tapped at some point x along its length, a further correspondence between this circuit and the simple low-frequency parallel resonant circuit is seen. The reactance looking toward the shorted end will be inductive and of value $Z_{sc} = jZ_0 \tan \beta x$. The reactance looking toward the open end will be of equal magnitude but opposite sign, i.e., a capacitive reactance. Its value is given by $Z_\infty = jZ_0 \cot \beta(\lambda/4 - x) = -jZ_0 \tan \beta x$. The equal but opposite reactances are in parallel just as they are in Fig. 8-11b and the input impedance will be purely resistive. As the tap point is moved from the shorted end toward the open end of the line, the impedance seen at the tap point is a pure resistance that varies from zero to the quite high value already given ($R_S = 2Z_0^2/Rl$). This corresponds in the circuit of Fig. 8-11b to varying the reactances X_L and X_C from low to high values, meanwhile always keeping the circuit tuned (i.e., $X_L = X_C$).

It is of interest to know how the input resistance varies as the tap point is moved along the quarter-wave section. For the relatively high Q circuits used in such applications the voltage distribution along the line may be considered sinusoidal and it is a simple matter to determine the input resistance at any point a distance x from the shorted end. For a given magnitude of voltage and current on the quarter-wave section a certain fixed amount of power

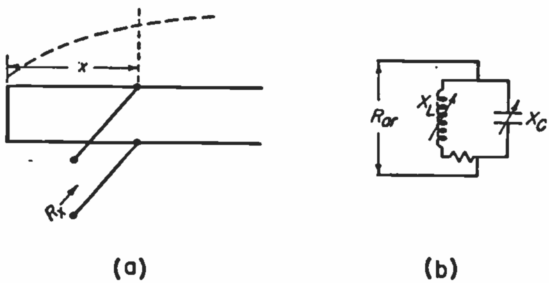


FIG. 8-11. (a) Tapped quarterwave line and (b) its equivalent circuit.

input will be required to supply the I^2R losses, regardless of where this power is fed in. This power input is equal to

$$\frac{V_s^2}{R_s} = \frac{V_s^2 Rl}{2Z_0^2}$$

where V_S and R_S are, respectively, the voltage and input resistance at the open end of the section. When the tap point of the feed line is at a distance x from the shorted end (Fig. 8-11a), the power input is given by V_x^2/R_x, where R_x is the input resistance at the point x. V_x is the voltage at this point and equals $V_S \sin \beta x$. Therefore

$$\frac{V_x^2}{R_x} = \frac{V_S^2 \sin^2 \beta x}{R_x} = \frac{V_S^2 Rl}{2Z_0^2}$$

which gives

$$R_x = \frac{2Z_0^2}{Rl} \cdot \sin^2 \beta x$$

Thus the input resistance varies as the square of the sine of the angular distance from the shorted end.

Q of Resonant Transmission Line Sections. One of the important properties of any resonant circuit is its selectivity or its ability to pass freely some frequencies, but to discriminate against others.

The selectivity of a resonant circuit may be conveniently stated in terms of the ratio $\Delta f/f_0$, where f_0 is the resonant frequency and $\Delta f = f_2 - f_1$ is the frequency difference between the "half-power" frequencies. In the case of a series resonant circuit $\Delta f/2$ represents the amount the frequency must be shifted away from the resonant frequency in order to reduce the current to 70.7 per cent of I_0, its value at the resonant frequency. (A constant voltage source is assumed.) Evidently this occurs when the reactance of the circuit becomes equal to the resistance and the phase angle of the circuit is 45°. For the parallel-resonant case $\Delta f/2$ represents the frequency shift away from unity power factor resonance necessary to reduce the voltage across the parallel circuit to 70.7 per cent of its value at resonance. (A constant current source is assumed.) This occurs when the absolute magnitude of the impedance is 70.7 per cent of the impedance at resonance.

The ratio $f_0/\Delta f$ may be used to define the Q of a resonant circuit. The Q of a resonant transmission line section can be determined as follows:

The input impedance of any shorted line section is given by

$$
\begin{aligned}
Z_s &= Z_0 \tanh \gamma l \\
&= Z_0 \frac{\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l}{\cosh \alpha l \cos \beta l + j \sinh \alpha l \sin \beta l}
\end{aligned}
$$

When the frequency is a resonant frequency f_0, then $\beta l = n\pi/2$ (where n is an odd integer), $\cos \beta l = 0$ and $\sin \beta l = \pm 1$, the expression for the input impedance becomes

$$Z_s = Z_0 \frac{\cosh \alpha l}{\sinh \alpha l} = \frac{Z_0}{\tanh \alpha l} \approx \frac{Z_0}{\alpha l}$$

When the frequency is shifted off resonance by a small amount δf, that is when $f = f_0 + \delta f$, then

$$\beta l = \frac{2\pi f}{v} l = \frac{2\pi(f_0 + \delta f)}{v} l = \frac{n\pi}{2} + \frac{2\pi\, \delta f l}{v}$$

Under these conditions

$$
\begin{aligned}
\cos \beta l &= -\sin \frac{(2\pi\, \delta f l)}{v} \\
\sin \beta l &= \cos \frac{(2\pi\, \delta f l)}{v}
\end{aligned}
$$

$$\text{and} \quad Z_s = Z_0 \frac{-\sinh \alpha l \sin\left(\frac{2\pi\ \delta f l}{v}\right) + j \cosh \alpha l \cos\left(\frac{2\pi\ \delta f l}{v}\right)}{-\cosh \alpha l \sin\left(\frac{2\pi\ \delta f l}{v}\right) + j \sinh \alpha l \cos\left(\frac{2\pi\ \delta f l}{v}\right)}$$

For moderately high Q circuits the first term in the numerator is the product of two small quantities and may be neglected in comparison with other terms. Putting

$$\cosh \alpha l \approx 1, \qquad \sinh \alpha l \approx \alpha l, \qquad \cos\left(\frac{2\pi\ \delta f l}{v}\right) \approx 1,$$

$$\sin\left(\frac{2\pi\ \delta f l}{v}\right) \approx \left(\frac{2\pi\ \delta f l}{v}\right)$$

gives

$$Z_s = \frac{Z_0}{\alpha l + j\left(\frac{2\pi\ \delta f l}{v}\right)}$$

When the imaginary term in the denominator is equal to the real term, the impedance Z_s will be 70.7 per cent of its value for a resonant length, and the frequency shift required to make this true will be $\Delta f/2$. Therefore

$$\frac{2\pi\ \Delta f l}{2v} = \alpha l$$

$$\Delta f = \frac{\alpha v}{\pi} = \frac{2\alpha f_0}{\beta}$$

The Q of the resonant section is

$$Q = \frac{f_0}{\Delta f} = \frac{\beta}{2\alpha} \tag{8-67}$$

Alternative forms of this expression are

$$Q = \frac{\pi f_0}{\alpha v} = \frac{2\pi f_0 Z_0}{R v} = \frac{\omega L}{R} \tag{8-68}$$

The Q is independent of the number of quarter wavelengths in the resonant section as long as αl is a small quantity. It is interesting to observe that the Q of a resonant section of transmission line is equal to the ratio of inductive reactance per unit length to resistance per unit length.

A similar analysis could be carried through for an open-ended resonant section (for which the length would be some multiple of a half wavelength). The expression for Q in this case would be identical with the above.

The Quarter Wave Line as a Transformer. When a section of transmission line is used as a reactance, or as a resonant circuit, it is a two-terminal network. The input terminals of the section are connected across the generator or load and the other terminals are left open or shorted as the case may be. However, a section of line is often used as a four-terminal network, in which case it is inserted in series between generator and load. Because the input impedance is in general different from the load impedance connected across the output terminals, the line section is an impedance-transforming network. This is true for all lengths of line, but the quarter-wave section has certain particular properties that make it very useful in this respect.

For any impedance termination Z_r, the input impedance of a section of lossless line is given by eq. (58) as

$$Z_s = Z_r \left(\frac{\cos \beta l + jZ_0/Z_r \sin \beta l}{\cos \beta l + jZ_r/Z_0 \sin \beta l} \right)$$

For the particular case of a quarter-wave section, $\beta l = \pi/2$, and this reduces to

$$Z_s = \frac{Z_0^2}{Z_r}$$

For the case under consideration, where Z_0 is a pure resistance R_0, this is

$$Z_s = \frac{R_0^2}{Z_r} \tag{8-69}$$

Thus the quarter-wave section is an impedance transformer, or more correctly an impedance inverter. Whatever the terminating impedance may be, the inverse impedance will appear at the input. If the output impedance consisted of a resistance R_2 in series with an inductive reactance X_{L_2}, the input impedance would be given by a resistance R_1 in *parallel* with a capacitive reactance X_{C_1}, where

$$R_1 = \frac{R_0^2}{R_2} \qquad \text{and} \qquad X_{C_1} = \frac{R_0^2}{X_{L_1}}$$

A pure resistance termination R is transformed into a pure resistance of value R_0^2/R.

This property of matching any two impedances Z_1, Z_2 such that $Z_1Z_2 = Z_0^2$ finds many practical applications. It can be used to join together, without impedance mismatch, lines having different characteristic impedances; it is only necessary to make the characteristic impedance of the quarter-wave matching section the geometric mean of the Z_0's to be matched. By means of the quarter-wave section a pure resistance load can be matched to a generator having a generator impedance that is resistive so long as the geometric mean between the resistances gives a value for the required characteristic impedance that is practicable to obtain.

Voltage Step-up of the Quarter-wave Transformer. As long as the quarter-wave transforming section is considered as being lossless, the ratio between input and output voltages will just be the square root of the ratio of the input and output impedances being matched. From the voltage equation (56), for the quarter-wave section

$$\frac{V_s}{V_r} = \frac{jI_rZ_0}{V_r} = \frac{jZ_0}{Z_r} = j\sqrt{\frac{Z_s}{Z_r}}$$

or calling V_r/V_s the voltage step-up

$$\left|\frac{V_r}{V_s}\right| = \sqrt{\frac{Z_r}{Z_s}}$$

For the infinite impedance termination, that is an open circuit, this simple relation indicates an infinite voltage step-up, and it becomes necessary to resort to the exact eqs. (44) and (45) for the correct answer in this case. For the quarter-wave section the voltage equation of (44) becomes

$$V_s = jV_r \sinh \alpha l + jI_r Z_0 \cosh \alpha l$$

In open circuit I_r is zero and the voltage step-up is

$$\frac{V_r}{V_s} = \frac{1}{\sinh \alpha l} \approx \frac{1}{\alpha l} = \frac{2Z_0}{Rl}$$

For the quarter-wave section this may be written

$$\left|\frac{V_r}{V_s}\right| = \frac{8Z_0}{R\lambda} = \frac{8Z_0f}{Rv}$$

while for a three-quarterwave section the voltage step-up would be

$$\frac{V_r}{V_s} = \frac{8Z_0f}{3R_v}$$

8.09 Impedance Matching by Means of Stub Lines. When a line is terminated in an impedance other than its characteristic impedance Z_0, reflection will occur and there will be standing waves of voltage and current along the line which may be very large if there is considerable "mismatch." In general, these standing waves are undesirable because they increase the line losses. It is possible to obtain an impedance match between the line and its load by use of a properly located "stub line."

Consider the UHF line, shown in Fig. 8-12, terminated in a resistance R, different from R_0. At a point $l = \lambda/4$ from the termination the input impedance will be a pure resistance of value $R_{in} = R_0^2/R$. If R is less than R_0, R_{in} will be greater than R_0,

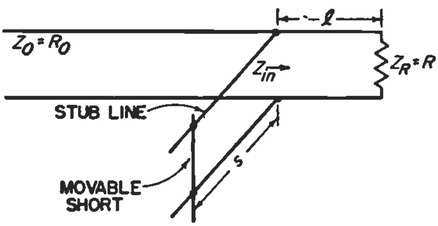


Fig. 8-12

whereas if R is greater than R_0, R_{in} will be less than R_0. Somewhere between $l = 0$ and $l = \lambda/4$, the resistance component of the input impedance will equal R_0. However, there will also be a reactive component at such a point; but if this is tuned out by means of an equal and opposite reactance (the stub line), only the resistance component $R_{in} = R_0$ will remain and the line coming up to this point will be properly terminated. At any point l the input impedance is, from (58),

$$Z_{in} = R_0\left(\frac{R \cos \beta l + jR_0 \sin \beta l}{R_0 \cos \beta l + jR \sin \beta l}\right)$$

This impedance can be considered as a resistance in series with a reactance or resistance in parallel with a reactance. Because it is desired to tune out the reactance component with another reactance in parallel (the stub line), the parallel representation will be used. The input admittance at a point l will be

$$Y_{in} = G_{in} + jB_{in} = \frac{1}{Z_{in}} = \frac{R_0 \cos \beta l + jR \sin \beta l}{R_0(R \cos \beta l + jR_0 \sin \beta l)}$$

$$= \frac{RR_0}{R_0(R^2 \cos^2 \beta l + R^2 \sin^2 \beta l)}$$

$$+ \frac{j(R^2 - R_0^2) \sin \beta l \cos \beta l}{R_0(R^2 \cos^2 \beta l + R_0^2 \sin^2 \beta l)}$$

and l should be chosen so that

$$G_{in} = \frac{1}{R_0}$$

That is,

$$\frac{R_0 R}{R_0(R^2 \cos^2 \beta l + R_0^2 \sin^2 \beta l)} = \frac{1}{R_0}$$

This gives

$$\frac{R}{R_0} \cos^2 \beta l + \frac{R_0}{R} \sin^2 \beta l = \cos^2 \beta l + \sin^2 \beta l$$

$$\cos^2 \beta l \left(1 - \frac{R}{R_0}\right) = \sin^2 \beta l \left(\frac{R_0}{R} - 1\right)$$

$$\tan^2 \beta l = \frac{R_0 - R}{R_0} \cdot \frac{R}{R_0 - R} = \frac{R}{R_0}$$

$$\tan \beta l = \sqrt{\frac{R}{R_0}} \qquad (8\text{-}70)$$

This equation gives the distance back from the termination R to the point where the input conductance is equal to $1/R_0$ and determines the correct location of the stub.

The length of the stub line required can be calculated by making its reactance equal and opposite to that at the tap point. Assuming the stub has the same characteristic impedance as the line, it will have a reactance equal to

$$jZ_0 \tan \beta S = jR_0 \tan \beta S$$

where S is the stub length. Then

$$jR_0 \tan \beta S = -\frac{1}{jB_{in}}$$

$$jB_{in} = j\frac{\sin \beta l \cos \beta L(R^2 - R_0^2)}{R_0(R^2 \cos^2 \beta l + R_0^2 \sin^2 \beta l)} = -\frac{1}{jR_0 \tan \beta S}$$

$$\tan \beta S = \frac{R^2 \cos^2 \beta l + R_0^2 \sin^2 \beta l}{\sin \beta l \cos \beta l (R^2 - R_0^2)}$$

$$= \frac{\dfrac{R}{R_0} + \dfrac{R_0}{R} \tan^2 \beta l}{\tan \beta l \left(\dfrac{R}{R_0} - \dfrac{R_0}{R}\right)} = \frac{\dfrac{R}{R_0} + 1}{\sqrt{\dfrac{R}{R_0}} \left(\dfrac{R}{R_0} - \dfrac{R_0}{R}\right)} \qquad (8\text{-}71)$$

This can be reduced to

$$\tan \beta S = \frac{\sqrt{R/R_0}}{(R/R_0) - 1} \qquad (8\text{-}72)$$

Stub Line Matching in Terms of Maximum and Minimum Voltages. In an actual experimental set-up the value of the terminating resistance is generally unknown, so that it is desirable to determine the dimensions for the stub line match directly from the measurable standing-wave ratio that exists on the transmission line.

The standing-wave ratio in terms of the terminating resistance and characteristic resistance is

$$\frac{V_M}{V_m} = \frac{R}{R_0} \qquad (\text{for } R > R_0)$$

so that the position and length S of the matching stub are given by

$$\tan \beta l = \sqrt{\frac{V_M}{V_m}} \qquad (8\text{-}73)$$

$$(\text{for } R > R_0)$$

$$\tan \beta S = \frac{\sqrt{V_M/V_m}}{(V_M/V_m) - 1} \qquad (8\text{-}74)$$

In these expressions the subscripts M and m have been used to indicate maximum and minimum respectively.

Stub Matching a Line to a Complex Impedance. The formulas just derived give the stub adjustments required to match a line to a resistance load. They may also be used to match a line to a complex load impedance if the proper reference point is taken. When a line is terminated in a complex impedance, there will be a standing wave on the line and there will be neither a voltage maximum nor a voltage minimum at the termination. However, at some point down the line there will be a voltage maximum and at this point the input impedance will be a pure resistance greater than

R_0; at another point there will be a voltage minimum where the impedance will be a pure resistance less than R_0. *Therefore, if measurements are made from either a voltage maximum or a voltage minimum, the problem will have been reduced to that of matching a line to a pure resistance.* The solution for this case has been given above. Therefore, the experimental procedure for matching a line to *any* complex impedance would be as follows:

Experimental Procedure for Stub Line Matching. 1. Measure the standing wave of voltage back from the termination. Note the position of the maxima and minima and the ratio of maximum to minimum voltages V_M/V_m.

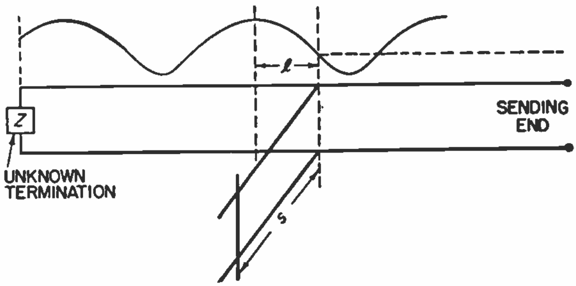


FIG. 8-13. Stub matching a transmission line.

2. Place a shorted stub line of length S a distance l back from a maximum point toward the sending end. The line is then properly terminated in its characteristic resistance.

The length l and S are given by

$$\tan \beta l = \sqrt{\frac{V_M}{V_m}}$$

$$\tan \beta S = \frac{\sqrt{V_M/V_m}}{(V_M/V_m) - 1}$$

Of course, the stub can be placed on the other side of the voltage maximum, but then a *capacitive* stub would be required. For a shorted stub this means a length greater than a quarter wavelength, which is usually undesirable. The procedure outlined above ensures a stub length less than a quarter wavelength.

Double-stub Tuner. The single-stub matching unit just discussed is satisfactory for impedance matching on open parallel-wire lines, but proves inconvenient to use on coaxial lines, because on these lines it is difficult to vary the position of the stub along the line. For matching on coaxial lines the double-stub tuner arrangement of Fig. 8-14 is usually used. This arrangement consists of two adjustable tuners that have movable shorting plungers, but are fixed in position on the line.

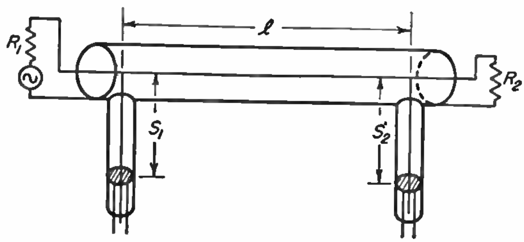


FIG. 8-14. Double-stub tuner.

It is easy to show that with the double-stub tuner of Fig. 8-14 it is possible to match any two impedances within a certain specified range of values. The analysis will be carried through for matching pure resistances only, but the results are general, because any complex admittance can be expressed as a conductance in parallel with

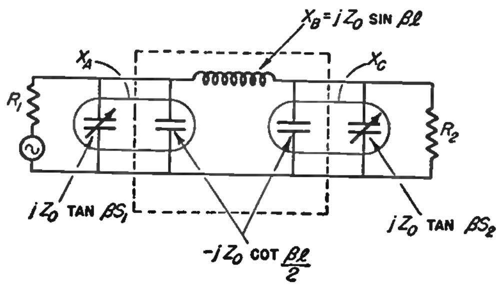


FIG. 8-15. Circuit representation of a double-stub turner.

a susceptance. The susceptance can then be tuned out by means of the adjustable susceptance of the stub, which is in parallel with the load, leaving a pure resistance to be matched.

The distributed-constants circuit of Fig. 8-14 can be represented at one frequency by the lumped-constants circuit of Fig. 8-15. The portion of the network within the dotted enclosure is the π-section representation of the length l of lossless transmission line between the two tuners. The reactances of the three arms are*

$$X_B = Z_0 \sin \beta l \qquad X_A' = X_C' = -Z_0 \cot \frac{\beta l}{2}$$

In parallel with the vertical arms of the π network are the reactances of the tuning sections. These reactances are

$$X_A'' = Z_0 \tan \beta S_1 \qquad X_B'' = Z_0 \tan \beta S_2$$

These reactances are shown negative (capacitive) in the figure, but of course they may have any value from plus infinity to minus infinity. Combining these parallel reactances giving a matching π-section that has one fixed element, X_B, and two adjustable elements, X_A and X_C, which may have any reactance desired. (The reactance X_A is formed by X_A' and X_A'' in parallel. Similarly X_C is formed by X_C' and X_C'' in parallel.) The values of X_A and X_C required to match two resistances R_1 and R_2 are†

$$X_A = \frac{-R_1 X_B}{R_1 \pm \sqrt{R_1 R_2 - X_B^2}}$$

$$X_C = \frac{-R_2 X_B}{R_2 \pm \sqrt{R_1 R_2 - X_B^2}}$$

It is possible to match any two resistances R_1 and R_2 as long as X_B is less than $\sqrt{R_1 R_2}$. It would appear, by suitable choosing the length l of the section near zero or some multiple of $\lambda/2$, that X_B could be made as small as is desired. However, if such a length is chosen, it will be found that the required values of X_A and X_C will come out very nearly equal to zero, and large currents will flow in the section. The elements have been assumed lossless, but every physical set-up has some loss, and large circulating currents mean inefficient operation. For this reason the length l should be selected between about $\lambda/8$ and $3\lambda/8$, or $5\lambda/8$ and $7\lambda/8$. The range of resistances that can be matched will be any resistances such that

$$\sqrt{R_1 R_2} > Z_0 \sin \beta l$$

* See for example, W. L. Everitt, *Communication Engineering*, McGraw-Hill, New York, 1932, p. 173.

† *Communication Engineering*, p. 265.

If resistances outside this range (that is, low resistances) are to be matched, a triple-stub matching unit can be used. The triple-stub unit can be analyzed by considering it as two double-stub units in tandem. The first double-stub unit matches one of the low resistances R_1 to any suitable high resistance. Then the second double-stub unit matches this high resistance to the second low resistance R_2. Thus with a triple-stub tuner, any two impedances whatsoever can be matched.

Resonance with l Variable. When resonance phenomena are considered in the ordinary lumped-constant circuits, two things are of interest: the operation of the circuit with L and C fixed and the frequency variable or, secondly, the operation with a fixed frequency

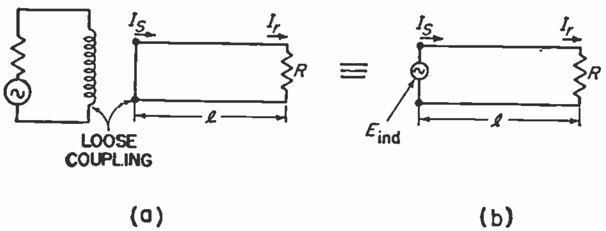


FIG. 8-16. "Series" feed.

but with either L or C variable. Similarly with the line circuits, interest may center around a line of fixed length under conditions of varying frequency, or the frequency may be fixed and the line length varied to obtain resonance. The first of these conditions was considered under selectivity of line sections. The second will be considered now.

(*a*) *"Series" Resonance.* When the voltage is introduced "in series" in the circuit, usually at one end of the line, the adjustments made for maximum current from a low-impedance source, the line may be said to be in series resonance. This is illustrated by the circuit of Fig. 8-16, which shows a constant voltage V_{ind}, induced into one end of the line or inserted directly as shown in the equivalent circuit.

The first problem is to determine the length l that will make the sending-end current I_s a maximum.

Since $|I_s| = E_{ind}/|Z_{in}|$, the problem is simply that of determining for what length l, $|Z_{in}|$ will be a minimum. From eq. (58) for the lossless line case

$$Z_{\text{in}} = R_0 \left(\frac{R \cos \beta l + jR_0 \sin \beta l}{R_0 \cos \beta l + jR \sin \beta l} \right)$$

so that

$$|Z_{\text{in}}|^2 = R_0^2 \left(\frac{R^2 \cos^2 \beta l + R_0^2 \sin^2 \beta l}{R_0^2 \cos^2 \beta l + R^2 \sin^2 \beta l} \right)$$

$$= R_0^2 \left(\frac{k^2 \cos^2 \beta l + \sin^2 \beta l}{\cos^2 \beta l + k^2 \sin^2 \beta l} \right)$$

where

$$k = \frac{R}{R_0}$$

When $|Z_{\text{in}}|^2$ is a minimum, $|Z_{\text{in}}|$ will be a minimum. Taking the derivative with respect to l and putting this equal to zero gives

$$\begin{aligned} 0 &= \frac{\partial}{\partial l} |Z_{\text{in}}|^2 \\ &= (\cos^2 \beta l + k^2 \sin^2 \beta l)(-2k^2 \cos \beta l \sin \beta l + 2 \sin \beta l \cos \beta l) \\ &\quad - (k^2 \cos^2 \beta l + \sin^2 \beta l)(-2 \cos \beta l \sin \beta l + 2k^2 \sin \beta l \cos \beta l) \end{aligned}$$

which simplifies to

$$-2k^4 \sin \beta l \cos \beta l + 2 \sin \beta l \cos \beta l = 0$$

For $k = 1$, that is, when R is equal to R_0, this expression is identically zero and the input impedance is independent of length. For all other values of k the expression can be satisfied only if $\sin \beta l$ or $\cos \beta l$ equals zero. For R less than R_0 the input impedance is a minimum when $\sin \beta l$ is zero, that is, when the length of the line is a multiple of a half wavelength. For R greater than R_0 the input impedance is a minimum when $\cos \beta l$ is zero, that is, when l is an odd multiple of a quarter wavelength.

A second, perhaps more important problem, is to determine the line lengths which make the current I_r through the resistance R a maximum.

Putting $V_r = I_r R$ for a resistive termination, the voltage equation for lossless transmission lines may be written

$$\frac{V_s}{I_r} = R \cos \beta l + jZ_0 \sin \beta l$$

or

$$\left| \frac{V_s}{I_r} \right| = \sqrt{R^2 \cos^2 \beta l + Z_0^2 \sin^2 \beta l} \tag{8-75}$$

For a constant voltage V_s, the current I_r will be a maximum when this expression is a minimum. This will occur for sin $\beta l = 0$ when R is less than Z_0, and for cos $\beta l = 0$ when R is greater than Z_0. Therefore, for values of R less than Z_0, the current through R, will be greatest for lengths l, which are multiples of $\lambda/2$. For values of R greater than Z_0, the current through R will be a maximum when l is an odd multiple of $\lambda/4$.

Case when Terminating Impedance is not a Pure Resistance. When the terminating impedance is not a pure resistance, the length l for "resonance" will not be a multiple of $\lambda/4$ or $\lambda/2$ as occurred above. However, the required length is readily determined by representing the complex impedance by means of a pure resistance

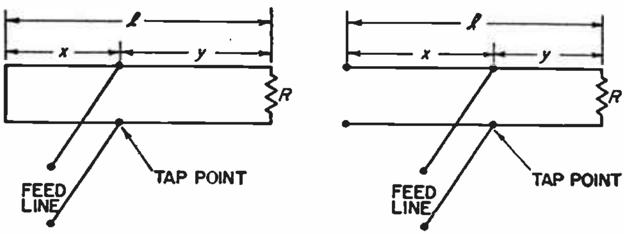


FIG. 8-17. "Parallel feed."

R_1 at the end of the appropriate length l_1 of line [eqs. (61) and (62)]. Evidently the required length l will be given by

$$l = \frac{n\lambda}{2} - l_1$$

(*b*) *"Parallel" Resonance.* When the voltage is applied across the line as shown in Fig. 8-17 rather than in series as in Fig. 8-16, the circuit corresponds to the ordinary low-frequency parallel circuit. The adjustment that is of interest is the length l, which will present a unity power factor load to the feed line tapped on at point x. The unity power factor condition requires that the reactive component of the impedance seen looking to the right from the tap point be equal and opposite to the reactance seen looking to the left from the tap point. For a given length y from the tap point to the load, the reactance seen to the right of the tap point will vary with the value of load resistance R. This means that the required length x will depend upon R, and the total length $l = (x + y)$ for unity power factor resonance will depend upon the value of the load impedance. This is similar to the analogous low-frequency circuit case in which the reactances required for unity power factor resonance depend upon the resistance in the circuit.

The case of most practical interest is the one in which the feed line is tapped on to the resonant section at a point, such that the feed line is terminated in its characteristic resistance. If the characteristic resistance of the feed line is the same as that of the resonant section, as is usually the case, this problem reduces to one for which the solution is already known, viz., the stub-line match problem. The section of length x can be considered as the stub, located at a distance y from the load R. For any load resistance R the required length y will be given by

$$\tan \beta y = \sqrt{\frac{R}{R_0}} \tag{8-76}$$

whereas the required length x would be given by

$$\tan \beta x = \frac{\sqrt{R/R_0}}{(R/R_0) - 1} \tag{8-77}$$

The total length of the section l for unity power factor resonance will be

$$l = x + y$$

Except for R equal zero or infinity l will *not* equal $\lambda/2$ or $\lambda/4$.

8.10 Graphical Representation of Transmission Line Phenomena. It is possible to solve most of the transmission-line problems considered above, and many others as well, by means of a simple graphical procedure. If impedances are plotted in the complex plane in the form of the R-X diagram of Fig. 8-18, then it turns out that for a lossless line terminated in some fixed impedance Z_R, the locus of the input impedance Z_s (as βl is varied) is a circle through Z_R. Since the center and radius of the circle are easily calculable, this leads to a very simple and rapid method for making transmission line calculations.

Because, as has already been seen, any complex impedance Z_R can be obtained by use of a pure resistance at the end of an appropriate length of lossless line, it is only necessary to discuss the pure resistance termination in order to solve the most general case. Consider a lossless transmission line terminated in a pure resistance R.

If the standing-wave ratio ρ, is expressed as a number greater than unity, then

$$\rho = \frac{R}{R_0} \qquad \text{for } R > R_0$$

$$\rho = \frac{R_0}{R} \qquad \text{for } R < R_0$$

It is convenient in a diagram such as that of Fig. 8-18 to normalize the impedances by giving their values with respect to Z_0, which is taken as unity. In the present application Z_0 will always be a pure resistance R_0. The normalized impedances are designated by lower case letters as

$$r + jx = \frac{Z}{R_0} = \frac{R}{R_0} + j\frac{X}{R_0}$$

and in the diagram R_0 will occur at the point (1,0). Now referring to the figure, it can be shown that a circle drawn through some resistance value r, with center at $\frac{\rho^2 + 1}{2\rho}$ and radius equal to $\frac{\rho^2 - 1}{2\rho}$, will be the locus of the input impedances as βl is allowed to vary.

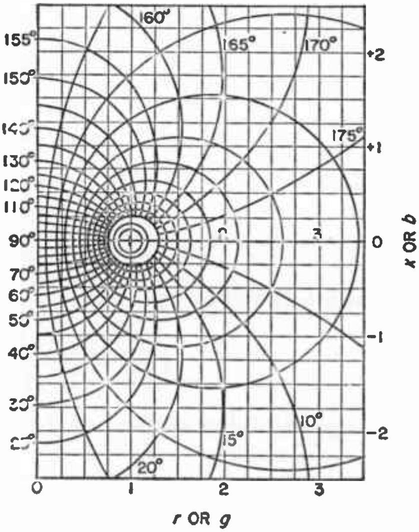


FIG. 8-18. Impedance or admittance diagram.

From eq. (58), the normalized input impedance is

$$r_s + jx_s = \frac{Z_s}{R_0} = \frac{r \cos \beta l + j \sin \beta l}{\cos \beta l + jr \sin \beta l} \tag{8-78}$$

Letting $\theta = \beta l$, and replacing sin and cos by their exponential equivalents, eq. (78) becomes

$$r_s + jx_s = \frac{r(e^{+j\theta} + e^{-j\theta}) + (e^{j\theta} - e^{-j\theta})}{(e^{j\theta} + e^{-j\theta}) + r(e^{j\theta} - e^{-j\theta})}$$

which reduces to

$$r_s + jx_s = \frac{(r+1) + (r-1)\,e^{-j2\theta}}{(r+1) - (r-1)\,e^{-j2\theta}} \tag{8-79}$$

Now the reflection coefficient, which in general is a complex quantity, given by

$$K = \frac{V''}{V'} = \frac{Z_R - Z_0}{Z_R + Z_0}$$

will, in this case, be a real number equal to $\frac{r-1}{r+1}$. Its absolute value will be

$$|K| = \frac{\rho - 1}{\rho + 1}$$

Then from eq. (79)

$$r_s + jx_s = \frac{1 + K\,e^{-j2\theta}}{1 - K\,e^{-j2\theta}}$$

so that

$$K\,e^{-j2\theta} = \frac{(r_s - 1) + jx_s}{(r_s + 1) + jx_s} \tag{8-80}$$

Equating the absolute values of (80),

$$|K|^2 = \frac{r_s^2 - 2r_s + 1 + x^2}{r_s^2 + 2r_s + 1 + x^2}$$

Cross multiplying and completing the square leads to

$$\left(r_s - \frac{1 + |K|^2}{1 - |K|^2}\right)^2 + x^2 = \left(\frac{2|K|}{1 - |K|^2}\right)^2 \tag{8-81}$$

This is the equation of a circle having a radius

$$\frac{2|K|}{1 - |K|^2} = \frac{\rho^2 - 1}{2\rho}$$

and a center located on the real axis at

$$\frac{1 + |K|^2}{1 - |K|^2} = \frac{\rho^2 + 1}{2\rho}$$

In the diagram of Fig. 8-18 circles corresponding to various values of terminating resistance (and therefore to various constant values of ρ) have been drawn in. It is necessary to mark off on these circles points indicating the values of βl. If the points on various circles having the same value of βl are joined, it is found that the curves so formed are circles orthogonal to the ρ = constant set. The equation for the βl = constant circles is derivable, in a manner similar to the derivation of (81), by equating the tangents of the angles in (80), instead of the absolute magnitudes. The result yields circles with their centers located on the imaginary axis at $1/\tan 2\theta$ and having radii equal to $1/\sin 2\theta$. These circles have been drawn in on Fig. 8-18 and labeled with the appropriate values of βl.

Using Fig. 8-18 it is a simple matter to make various computations as the following examples will indicate.

(a) A 300-ohm line is terminated in a 600-ohm resistive impedance. What is the input impedance if the line is one-eighth wavelength long?

Following the $\rho = 2$ circle in clockwise direction for 45°, the normalized impedance is found to be $0.8 - j0.6$. Therefore, $Z_s = 240 - j180$ ohms.

(b) The line of part (a) is terminated in an impedance $Z_R = 180 + j200$ ohms. Find the input impedance. *Answer:* $z = 2.5 + j0.4$, $Z_s = 750 + j120$.

Because the transmission line equations have the same form when expressed for admittances as they do for impedances, the impedance diagram can be used as an admittance diagram by simply changing the labels. That is, g replaces r, and b replaces x, with G_0 replacing R_0. Using the admittance diagram, the stub matching problem is readily solved in the following manner. The chart is entered at the (normalized) value of the load admittance,

$$y = g + jb = \frac{G + jB}{G_0}$$

to determine a ρ = constant circle. This circle is followed around to its intersection with the vertical axis at $g = 1$. At this point the input conductance has the value $g = 1$ (actual value $G = G_0$), and the input susceptance has a value given by the length of the vertical intercept. The electrical distance βl, measured along the transmission line from the load point, at which the stub should be located is given by the angle through which it was necessary to turn to reach the vertical axis. The required susceptance of the stub to cancel the susceptance on the line at this point is just the negative of the susceptance indicated at the intersection of the circle and the vertical ($g = 1$) axis. The electrical stub length that will produce this susceptance (assuming a shorted stub) is given by the value of the βl = constant curve which intersects the imaginary axis ($g = 0$) at the particular value of susceptance required.

EXAMPLE 1: A 300-ohm line is terminated in an admittance $Y_R = 0.002 - j0.0033$, which corresponds to a normalized admittance

$$g + jb = \frac{Y_R}{G_0} = 0.6 - j1.0$$

Determine the location and length of a shorted stub to match this load to the line.

Entering the diagram at the admittance value $0.6 - j1.0$, it is found by interpolation that this point marks the intersection of $\rho = 3.45$ and $\beta l = 40°$ circles. Following the ρ = constant circle clockwise around to its intersection with the vertical ($g = 1$) line, a value of $\beta l = 152$ degrees is found. The susceptance at this point is 1.35, so the stub must have an input susceptance of -1.35. The shorted electrical stub length that has this value of susceptance is given by the βl = constant curve, which intersects the $g = 0$ axis at $b = -1.35$. This is found to be 37 degrees. Therefore the electrical stub length is 37 degrees and its location on the transmission line is $152 - 40 = 112$ degrees, from the load.

It is interesting to note in passing how the impedance value corresponding to a certain admittance (or vice versa) can be obtained from the circle diagram. For example, to obtain the (normalized) impedance corresponding to the (normalized) admittance $y = 0.6 - j1.0$, it is merely necessary to follow the $\rho = 3.45$ circle around from 40 degrees (through another 90 degrees) to 130 degrees, at which point the normalized impedance can be read as $z = 0.44 + j0.74$. That this should be so follows directly from the impedance inversion properties of the quarter-wave line which "inverts" impedances about Z_0 (which in this case is unity). That is, $Z_s = Z_0^2/Z_R$, and $z_s = 1/z_R = y_R$ (for $\beta l = 90$ degrees).

The settings for double-stub tuners are also obtained with ease by this graphical method. Because the spacing between tuners is fixed (say at 135 degrees), the procedure is to adjust the admittance of the first stub to such a value that when 135 degrees is turned off on the ρ = constant circle the intersection with the ($g = 1$) line will have just been reached. Then the second stub is used to cancel the remaining susceptance at that point on the line.

The examples given will serve to indicate the effectiveness of the graphical approach. The diagram of Fig. 8-18 is only one of several such diagrams that can be used. While this particular chart is probably the simplest, other diagrams based on the same principles, and derivable from the diagram of Fig. 8-18 by suitable transforma-

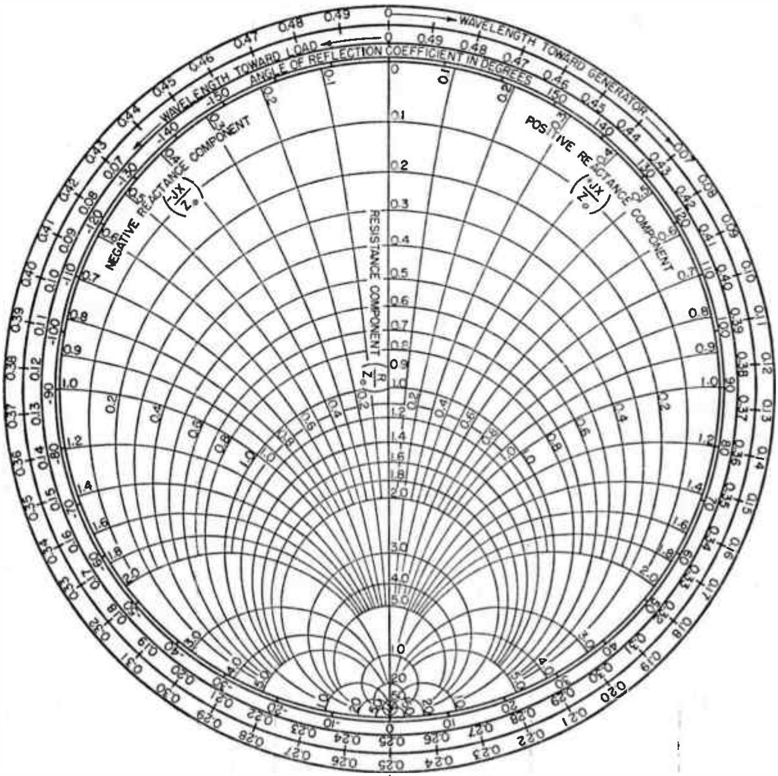


Fig. 8-19. Smith chart for transmission line calculations.

tion of variables, are more convenient for actual computations. Several such charts are available commercially. Among these, the polar impedance chart, described by P. H. Smith,* has come into extensive use. This is a circular chart illustrated qualitatively in Fig. 8-19. In it, the r = constant and x = constant lines have been

* P. H. Smith, "An Improved Transmission-Line Calculator," *Electronics*, **17**, p. 130 (Jan. 1944); also **12**, p. 29 (Jan. 1939).

TABLE II
NUMERICAL DATA ON UHF TRANSMISSION LINES

	Parallel Wire Lines Conductor radius = a Conductor spacing (between centers) = b	*Coaxial Lines* Outer radius of inner conductor = a Inner radius of outer conductor = b
Inductance L (henry/m length of line)	$\frac{\mu_v}{\pi}\cosh^{-1}\frac{b}{2a}$ or $\frac{\mu_v}{\pi}\ln\frac{b}{a}$ for $b \gg a$	$\frac{\mu_v}{2\pi}\ln\frac{b}{a}$
Capacitance C (f/m length of line)	$\frac{\pi\epsilon}{\cosh^{-1} b/2a}$ or $\frac{\pi\epsilon}{\ln b/a}$ for $b \gg a$	$\frac{2\pi\epsilon}{\ln b/a}$
Resistance R (ohms/unit length of line)	$\frac{R_s}{\pi a} = \frac{1}{\pi a}\sqrt{\frac{\omega\mu_v}{2\sigma}}$ ohms/m For copper lines $R = \frac{8.31 \times 10^{-8} f^{1/2}}{a}$ ohms/m $\approx \frac{\sqrt{f_{mc}}}{a_{in.}}$ ohms/1000 ft	$\frac{R_s}{2\pi}\left(\frac{1}{a}+\frac{1}{b}\right) = \frac{1}{2\pi}\sqrt{\frac{\omega\mu_v}{2\sigma}}\left(\frac{1}{a}+\frac{1}{b}\right)$ ohms/m For copper lines $R = 4.16 \times 10^{-8} f^{1/2}\left(\frac{1}{a}+\frac{1}{b}\right)$ ohms/m
Conductance G (mhos/m length of line)	ωC(dissipation factor) $\approx \omega C$(power factor of dielectric)	

TABLE II—*Continued*

	Parallel Wire Lines Conductor radius $= a$ Conductor spacing (between centers) $= b$	*Coaxial Lines* Outer radius of inner conductor $= a$ Inner radius of outer conductor $= b$
Characteristic impedance $Z_0 = \sqrt{\dfrac{R + j\omega L}{G + j\omega C}} \approx \dfrac{L}{\sqrt{C}}$	(air dielectric) $120 \cosh^{-1} \dfrac{b}{2a}$ or $276 \log_{10} \dfrac{b}{a}$ for $b \gg a$	$\dfrac{\eta}{2\pi} \ln \dfrac{b}{a} = \dfrac{138}{\sqrt{\epsilon_r}} \log_{10} \dfrac{b}{a}$
Attenuation constant α (neper/m)	$\alpha = \dfrac{R}{2Z_0} + \dfrac{GZ_0}{2}$	
Phase shift constant β (radians/m)	$\beta = \dfrac{2\pi}{\lambda} = \dfrac{\omega}{v_0}$	
Phase velocity v_0 (meter/sec)	$v_0 \approx \dfrac{1}{\sqrt{LC}} \approx \dfrac{3 \times 10^8}{\sqrt{\mu_r \epsilon_r}}$ for low-loss lines	

For air $\mu_r \approx 1$; $\epsilon_r \approx 1$

For copper $\mu_r \approx 1$; $\epsilon_r \approx 1$; $\sigma = 5.8 \times 10^7$

$\mu_v = 4\pi \times 10^{-7}$ $\epsilon_v \approx \dfrac{1}{36\pi \times 10^9}$

transformed into orthogonal sets of circles in such a manner that the ρ = constant lines now appear as concentric circles about R_0 ($r = 1$), which is at the center of the chart. The chart is equipped with a rotating radial arm so that computations can be performed as simply as computations on a slide rule. For the communication engineer who is required to work extensively with transmission line problems, the transmission line calculator is almost as indispensable as his slide rule.

PROBLEMS

1. Derive an expression for the inductance and capacitance per unit length of a coaxial transmission line.

2. Repeat for a parallel wire line (assume perfect conductors).

3. Compute the line "constants" per unit length, R, L, C, G, α, and β, and the characteristic impedance, Z_0, for each of the following lines at the frequencies indicated.

(a) No. 12 wires (diameter = 0.0808 in.) spaced 3 in. apart at 10 mc; at 100 mc.

(b) ⅜-in. diameter rods spaced 1 in. at 100 mc; at 1000 mc.

(c) A coaxial line having a ⅛-in. diameter inner conductor and a ⅜-in. outer conductor, at 1000 mc.

4. (a) A dipole antenna is fed by a transmission line consisting of No. 12 wires at 3-in. spacing. The measured ratio $V_{max}/V_{min} = 4$, and the location of a voltage minimum is 2.8 meters from the antenna feed point. $f = 112$ mc. Determine the antenna impedance.

(b) If a current indicator is used instead of a voltage indicator, where will the maximum and minimum readings be obtained, and what will be their ratio?

5. A shorted length of a parallel-rod transmission line is connected between grid and plate of a tube to make a UHF oscillator. What should be the length of the line to tune to 300 mc, if the effective capacitance between grid and plate is 3 $\mu\mu$fd? The rods are ⅜ in. in diameter and are spaced 1 in. apart.

6. A lossless transmission line has a characteristic impedance of 300 ohms and is one quarter wavelength long. What will be the voltage at the open-circuited receiving end, when the sending end is connected to a generator which has 50-ohm internal impedance and a generated voltage of 10 volts.

7. For low-loss transmission line sections which are much shorter than one quarter wavelength show that the input reactance can be represented by

$$X_{in} \approx \omega L_{in} = \omega L l$$

when the line is shorted, and

$$X_{in} \approx \frac{1}{\omega C_{in}} = \frac{1}{\omega Cl}$$

when the line is open.

$$L = \frac{Z_0}{v} \qquad \text{and} \qquad C = \frac{1}{vZ_0}$$

are the inductance and capacitance per unit length of the line.

8. Derive the expression for Q (eq. 67) for an open-end half-wave line.

9. Show that a coaxial line having an outer conductor of radius b will have minimum attenuation when the radius a of the inner conductor satisfies the ratio $b/a = 3.6$.

BIBLIOGRAPHY

Everitt, W. L., *Communication Engineering*, McGraw-Hill, New York, 1932.

Slater, J. C., *Microwave Transmission*, McGraw-Hill, New York, 1942.

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand, New York, 1943.

CHAPTER 9

WAVE GUIDES

9.01 Rectangular Guides. Practical wave guides usually take the form of rectangular or circular cylinders. Other cross sectional shapes are possible, but in general these other shapes offer no electrical advantages over the simpler forms in use and are more expensive to manufacture.

In order to determine the electromagnetic field configuration within the guide, Maxwell's equations are solved subject to the appropriate boundary conditions at the walls of the guide. Again assuming perfect conductivity for the walls of the guide, the boundary conditions are simply that E_{tan} and H_{norm} will be zero at the surface of the conductors. For rectangular guides Maxwell's equations and the wave equations are expressed in rectangular co-ordinates and the solution follows almost exactly as for waves between parallel planes. Assuming that time variations are given by $e^{j\omega t}$, and that variations in the z direction may be expressed as $e^{-\bar{\gamma}z}$, where $\bar{\gamma} = \bar{\alpha} + j\bar{\beta}$, Maxwell's equations become (for the loss-free region within the guide)

$$\begin{aligned}
\frac{\partial H_z}{\partial y} + \bar{\gamma} H_y &= j\omega\epsilon E_x & \frac{\partial E_z}{\partial y} + \bar{\gamma} E_y &= -j\omega\mu H_x \\
\frac{\partial H_z}{\partial x} + \bar{\gamma} H_x &= -j\omega\epsilon E_y & \frac{\partial E_z}{\partial x} + \bar{\gamma} E_x &= j\omega\mu H_y \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= j\omega\epsilon E_z & \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -j\omega\mu H_z
\end{aligned} \tag{9-1}$$

and the wave equations for E_z and H_z are

$$\left.\begin{aligned}
\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \bar{\gamma}^2 E_z &= -\omega^2\mu\epsilon E_z \\
\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \bar{\gamma}^2 H_z &= -\omega^2\mu\epsilon H_z
\end{aligned}\right\} \tag{9-2}$$

Equations (1) can be combined into the form

$$
\begin{aligned}
H_x &= -\frac{\bar{\gamma}}{h^2}\frac{\partial H_z}{\partial x} + j\frac{\omega\epsilon}{h^2}\frac{\partial E_z}{\partial y} \\
H_y &= -\frac{\bar{\gamma}}{h^2}\frac{\partial H_z}{\partial y} - j\frac{\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x} \\
E_x &= -\frac{\bar{\gamma}}{h^2}\frac{\partial E_z}{\partial x} - j\frac{\omega\mu}{h^2}\frac{\partial H_z}{\partial y} \\
E_y &= -\frac{\bar{\gamma}}{h^2}\frac{\partial E_z}{\partial y} + j\frac{\omega\mu}{h^2}\frac{\partial H_z}{\partial x}
\end{aligned}
\tag{9-3}
$$

where $$h^2 = \bar{\gamma}^2 + \omega^2\mu\epsilon$$

These equations give the relationships among the fields within the guide. It will be noticed that, if E_z and H_z are both zero, *all* the

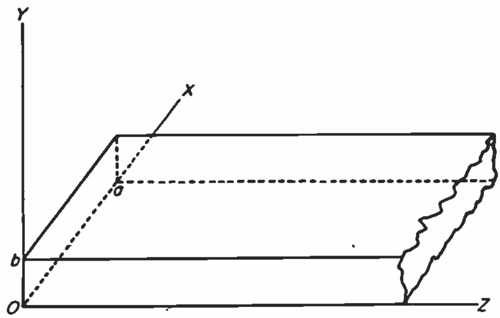


FIG. 9-1. A rectangular guide.

fields within the guide will vanish. Therefore, for waveguide transmission (no inner conductor) there must exist either an E_z or an H_z component. The TEM wave cannot exist inside a single-conductor wave guide. As in the case of waves between parallel planes, it is convenient to divide the possible field configurations within the guide into two sets, transverse magnetic (TM) waves for which $H_z \equiv 0$, and transverse electric (TE) waves for which $E_z \equiv 0$. For the rectangular guide shown in Fig. 9-1 the boundary conditions are:

$$
\begin{aligned}
E_x &= E_z = 0 \qquad \text{at } y = 0 \text{ and } y = b \\
E_y &= E_z = 0 \qquad \text{at } x = 0 \text{ and } x = a
\end{aligned}
$$

9.02 Transverse Magnetic Waves in Rectangular Guides. The wave equations (2) are partial differential equations that can be solved by the usual technique of assuming a product solution. This procedure leads to two ordinary differential equations, the solutions of which are known. Let

$$E_z = XY \tag{9-4}$$

where X is a function of x alone, Y is a function of y alone, and the factor $e^{(j\omega t - \bar{\gamma}z)}$ is understood. Inserting (4) in (2) gives

$$Y\frac{d^2X}{dx^2} + X\frac{d^2Y}{dy^2} + \bar{\gamma}^2 XY = -\omega^2\mu\epsilon XY$$

Putting $h^2 = \bar{\gamma}^2 + \omega^2\mu\epsilon$ as before, this becomes

$$Y\frac{d^2X}{dx^2} + X\frac{d^2Y}{dy^2} + h^2XY = 0$$

Divide by XY,

$$\frac{1}{X}\frac{d^2X}{dx^2} + h^2 = -\frac{1}{Y}\frac{d^2Y}{dy^2} \tag{9-5}$$

Equation (5) equates a function of x alone to a function of y alone. The only way in which such a relation can hold for all values of x and y is to have each of these functions equal to some constant, say A^2. Then

$$\frac{1}{X}\frac{d^2X}{dx^2} + h^2 = A^2 \tag{9-6}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -A^2 \tag{9-7}$$

A solution of eq. (6) is

$$X = C_1 \cos Bx + C_2 \sin Bx$$

where

$$B^2 = h^2 - A^2$$

The solution of eq. (7) is

$$Y = C_3 \cos Ay + C_4 \sin Ay$$

This gives

$$E_z = XY = C_1C_3 \cos Bx \cos Ay + C_1C_4 \cos Bx \sin Ay + C_2C_3 \sin Bx \cos Ay + C_2C_4 \sin Bx \sin Ay \tag{9-8}$$

The constants C_1, C_2, C_3, C_4, A, and B must now be selected to fit the boundary conditions, viz.:

$$E_z = 0 \qquad \text{when } x = 0,\ x = a,\ y = 0,\ y = b$$

If $x = 0$, the general expression (8) becomes

$$E_z = C_1C_3 \cos Ay + C_1C_4 \sin Ay$$

For E_z to vanish (for all values of y) it is evident that C_1 must be zero. Then the general expression for E_z will be

$$E_z = C_2C_3 \sin Bx \cos Ay + C_2C_4 \sin Bx \sin Ay \tag{9-9}$$

When $y = 0$, eq. (9) reduces to

$$E_z = C_2C_3 \sin Bx$$

For this to be zero for all values of x it is possible to have either C_2 or C_3 equal to zero (assuming $B \neq 0$). Putting $C_2 = 0$ in (9) would make E_z identically zero, so instead C_3 will be put equal to zero. Then the general expression (9) for E_z reduces to

$$E_z = C_2C_4 \sin Bx \sin Ay \tag{9-10}$$

In addition to the amplitude constant $C = C_2C_4$, there are still two unknown constants, A and B. However, there are two more boundary conditions to be applied.

If $x = a$

$$E_z = C \sin Ba \sin Ay$$

In order for this to vanish for all values of y (and assuming $A \neq 0$, because $A = 0$ would make E_z identically zero) the constant B must have the value

$$B = \frac{m\pi}{a} \qquad \text{where } m = 1, 2, 3, \cdots$$

Again if $y = b$,

$$E_z = C \sin \frac{m\pi}{a} x \sin Ab$$

and for this to vanish for all values of x, A must have the value

$$A = \frac{n\pi}{b} \qquad \text{where } n = 1, 2, 3, \cdots$$

Therefore the final expression for E_z is

$$E_z = C \sin \frac{m\pi}{a} x \sin \frac{n\pi}{b} y \tag{9-11}$$

Making use of eqs. (3) and putting $\bar{\gamma} = j\bar{\beta}$ (as in section 7.02) for frequencies above the cut-off frequency, the following expressions are obtained:

$$\left.\begin{aligned} E_x &= \frac{-j\bar{\beta}C}{h^2} B \cos Bx \sin Ay \\ E_y &= \frac{-j\bar{\beta}C}{h^2} A \sin Bx \cos Ay \\ H_x &= \frac{j\omega\epsilon C}{h^2} A \sin Bx \cos Ay \\ H_y &= \frac{-j\omega\epsilon C}{h^2} B \cos Bx \sin Ay \end{aligned}\right\} \tag{9-12}$$

where
$$B = \frac{m\pi}{a} \text{ and } A = \frac{n\pi}{b} \tag{9-13}$$

These expressions show how each of the components of electric and magnetic intensities varies with x and y. The variation with time and along the axis of the guide, that is in the z direction, is shown by putting back into each of these expressions the factor $e^{j\omega t - \bar{\gamma}z}$.

In the derivation of the fields it was found necessary to restrict the constants A and B to the values given by expressions (12). In these expressions a and b are the width and height of the guide, and m and n are integers. Now, by definition,

$$A^2 + B^2 = h^2$$

and
$$h^2 = \bar{\gamma}^2 + \omega^2\mu\epsilon$$

Therefore,

$$\begin{aligned} \bar{\gamma} &= \sqrt{h^2 - \omega^2\mu\epsilon} \\ &= \sqrt{A^2 + B^2 - \omega^2\mu\epsilon} \\ &= \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega^2\mu\epsilon} \end{aligned} \tag{9-14}$$

Equation (14) defines the propagation constant for a rectangular guide for TM waves. For low frequencies, where $\omega^2\mu\epsilon$ is small, $\bar{\gamma}$ will be a real number. The propagation constant met with in ordinary transmission line theory is a complex number, that is $\bar{\gamma} = \bar{\alpha} + j\bar{\beta}$, where $\bar{\alpha}$ is the attenuation constant (attenuation per unit length) and $\bar{\beta}$ is the phase shift constant (phase shift per unit length). If $\bar{\gamma}$ is real, $\bar{\beta}$ must be zero, and there can be no phase shift along the tube. This means there can be no wave motion along the tube for low frequencies. However, as the frequency is increased, a value for ω will be reached that will make the expression under the radical in (14) equal to zero. If this value of ω is called ω_c, then for all values of ω greater than ω_c, the propagation constant $\bar{\gamma}$ will be imaginary and will have the form $\bar{\gamma} = j\bar{\beta}$. For the case under consideration (perfectly conducting walls) the attenuation constant $\bar{\alpha}$ is zero for all frequencies such that $\omega > \omega_c$. For these frequencies

$$\bar{\beta} = \sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \tag{9-15}$$

The value of ω_c is given by

$$\omega_c = \frac{1}{\sqrt{\mu\epsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{9-16}$$

The cut-off frequency, that is the frequency below which wave propagation will not occur, is

$$f_c = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{9-17}$$

and the corresponding cut-off wavelength is

$$\lambda_c = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} \tag{9-18}$$

The velocity of wave propagation will be given by

$$\bar{v} = \frac{\omega}{\bar{\beta}} = \frac{\omega}{\sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}} \tag{9-19}$$

This last expression indicates that the velocity of propagation of the wave in the guide is greater than the phase velocity in free space. As the frequency is increased above cut-off, the phase velocity decreases from an infinitely large value and approaches c, the velocity in free space, as the frequency increases without limit.

Since the wavelength in the guide is given by $\bar{\lambda} = \bar{v}/f$, it will be longer than the corresponding free-space wavelength. From the expression for $\bar{v}$

$$\bar{\lambda} = \frac{2\pi}{\sqrt{\omega^2\mu\epsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}} \tag{9-20}$$

In the above expressions the only restriction on m and n is that they be integers. However from eqs. (12) and (13) it is seen that if either m or n is zero the fields will all be identically zero. Therefore the lowest possible value for either m or n (for TM waves) is unity. From eq. (17) it is evident that the lowest cut-off frequency will occur for $m = n = 1$. Substituting these values in eqs. (13) gives the fields for the lowest frequency TM wave which can be propagated through the guide. This particular wave is called the $TM_{1,1}$ wave for obvious reasons. Higher order waves (larger values of m and n) require higher frequencies in order to be propagated along a guide of given dimensions.

9.03 Transverse Electric Waves in Rectangular Guides. The equations for transverse electric waves ($E_z = 0$) can be derived in a manner similar to that for transverse magnetic waves. This is left as an exercise for the student. H_z will be found to have the same general form as eq. (8). This is differentiated with respect to x and y to find E_x, E_y, H_x, and H_y. The boundary conditions are then applied to E_x and E_y to give the resulting expressions:

$$\left.\begin{aligned}
H_z &= C \cos Bx \cos Ay \\
H_x &= \frac{j\bar{\beta}}{h^2} CB \sin Bx \cos Ay \\
H_y &= \frac{j\bar{\beta}}{h^2} CA \cos Bx \sin Ay \\
E_x &= \frac{j\omega\mu}{h^2} CA \cos Bx \sin Ay \\
E_y &= -\frac{j\omega\mu}{h^2} CB \sin Bx \cos Ay
\end{aligned}\right\} \tag{9-21}$$

$$B = \frac{m\pi}{a} \qquad A = \frac{n\pi}{b}$$

In the above expressions $\bar{\gamma}$ has been put equal to $j\bar{\beta}$, which is valid for frequencies above cut-off.

For TE waves the equations for $\bar{\beta}$, f_c, λ_c, $\bar{v}$, and $\bar{\lambda}$ are found to be identical to those for TM waves. However, in eqs. (21) for TE waves it will be found possible to make either m or n (but not both) equal to zero without causing all the fields to vanish. That is, a lower order is possible than in the TM wave case. The lowest order TE wave in rectangular guides is therefore the $TE_{1,0}$ wave. This wave which has the lowest cut-off frequency is called the *dominant* wave.

It is seen that the subscripts m and n represent the number of half-period variations of the field along the x and y co-ordinates respectively. By convention,* the x co-ordinate is assumed to coincide with the larger transverse dimension, so the $TE_{1,0}$ wave has a lower cut-off frequency than the $TE_{0,1}$.

For practical reasons in most experimental work with rectangular guides the dominant $TE_{1,0}$ wave is used. For this wave, substituting $m = 1$ and $n = 0$, the fields are

$$\left.\begin{aligned} H_z &= C \cos \frac{\pi x}{a} \\ H_x &= \frac{j\bar{\beta} a C}{\pi} \sin \frac{\pi x}{a} \\ E_y &= \frac{-j\omega\mu a C}{\pi} \sin \frac{\pi x}{a} \\ E_x &= H_y = 0 \end{aligned}\right\} \qquad (9\text{-}22)$$

$$\bar{\beta} = \sqrt{\omega^2 \mu \epsilon - \left(\frac{\pi}{a}\right)^2}$$

$$f_c = \frac{c}{2a} \qquad \lambda_c = 2a \qquad h = \frac{\pi}{a}$$

For the $TE_{1,0}$ wave the cut-off frequency is that frequency for which the corresponding (free-space) half wavelength is equal to the width of the guide. For the $TE_{1,0}$ wave the cut-off frequency is independent of the dimension b.

In Fig. 9-2 are sketched the field configurations for the lower order TE and TM waves in rectangular guides.

Possible methods for feeding rectangular guides so that these waves may be initiated are shown in Fig. 9-3. In order to launch a particular mode, a type of probe is chosen which will produce

* "Definition of Terms Relating to Wave Guides," *IRE Standards on Radio Wave Propagation*, 1945.

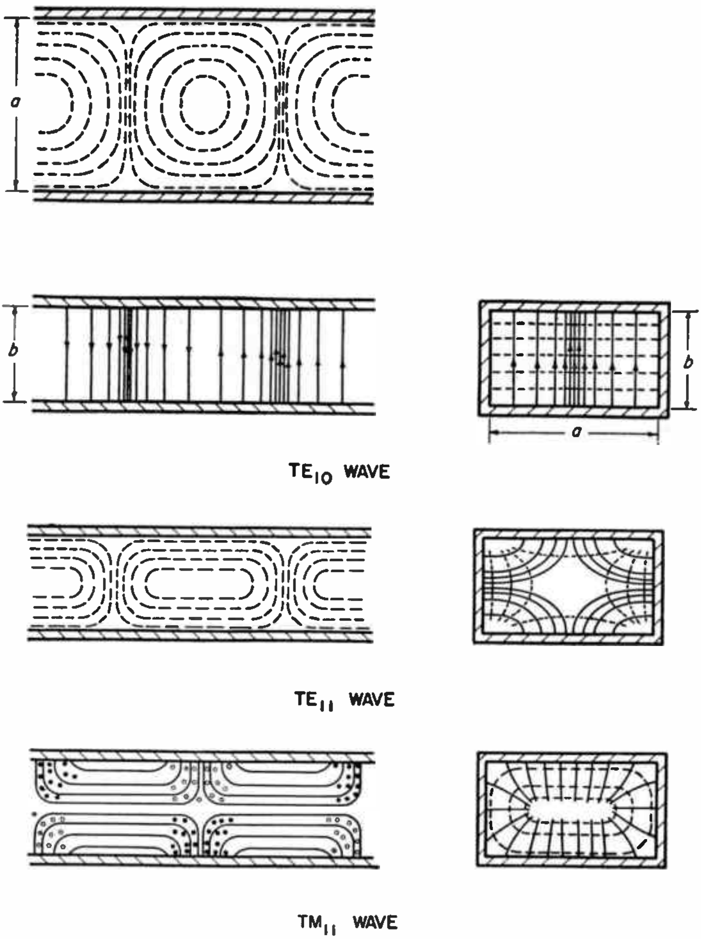


FIG. 9-2. Electric (solid) and magnetic (dashed) field configurations for the lower-order modes in a rectangular guide.

lines of E and H that are roughly parallel to the lines of E and H for that mode. Thus in Fig. 9-3*a* the probe is parallel to the y axis and so produces lines of E in the y direction and lines of H which lie in the x-z plane. This is the correct field configuration for the $TE_{1,0}$ mode. In (*b*), the parallel probes fed with opposite phase tend to set up the $TE_{2,0}$ mode. In (*d*) the probe parallel to the z axis produces magnetic field lines in the x-y plane, which is correct for the TM modes. The field configuration due to probes and antennas is the subject of chap. 10.

It is possible for several modes to exist simultaneously in a guide if the frequency is above cut-off for those particular modes. However the guide dimensions are often chosen so that only the dominant mode can exist.

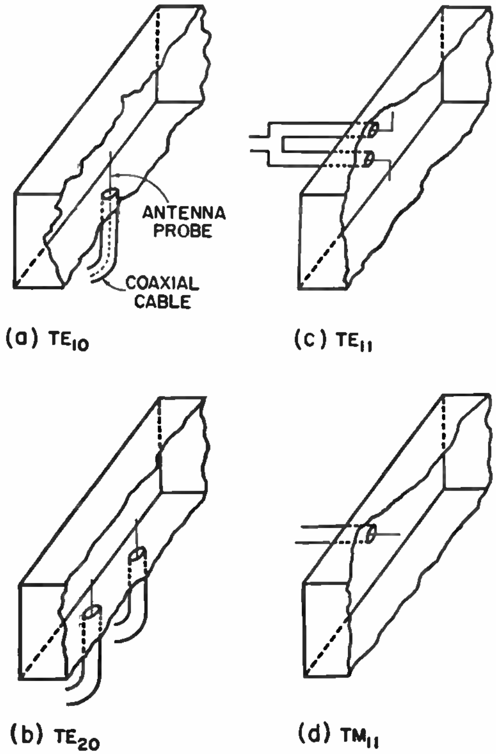


Fig. 9-3. Excitation methods for various modes.

Problem 1. A rectangular guide has cross section dimensions

$$a = 7 \text{ cm} \qquad b = 4 \text{ cm}$$

Determine all the modes which will propagate at a frequency of (a) 3000 mc, (b) 5000 mc.

Problem 2. Starting with expressions (9-16) and (9-20) derive the relations

$$\lambda = \frac{\bar{\lambda}\lambda_c}{\sqrt{\bar{\lambda}^2 + \lambda_c^2}}$$

where λ is the free-space wavelength ($\lambda = c/f$), $\bar{\lambda}$ is the wavelength measured in the guide, and λ_c is the cut-off wavelength.

Problem 3. Show by means of arrows the directions of the instantaneous Poynting vector for the $TE_{1,0}$ wave in a rectangular guide.

Problem 4. (a) Indicate the (instantaneous) directions of current flow in all the walls of a rectangular guide carrying a $TE_{1,0}$ wave.

(b) Where in the guide could slots be cut without affecting operation?

Problem 5. Starting with eqs. (9-2) and (9-3) derive expressions (9-21) for TE waves.

9.04 The TEM Wave in Wave Guides. The waves that will propagate inside hollow rectangular cylinders have been divided into two sets: the transverse magnetic waves of eqs. (11) and (12) which have no z component of H, and the transverse electric waves of eqs. (21) that have no z component of E. It will be found that corresponding sets of TM and TE waves can also propagate within circular wave guides, or indeed, in cylindrical guides of any cross sectional shape. It is easily shown, however, that the familiar TEM wave, for which there is no axial component of either E or H, cannot possibly propagate within a single-conductor wave guide.

Suppose a TEM wave is assumed to exist within a hollow guide of any shape. Then lines of H must lie entirely in the transverse plane. Also in a nonmagnetic material.

$$\text{div } H = 0$$

which requires that the lines of H be *closed* loops. Therefore, if a TEM exists inside the guide, the lines of H will be closed loops in plane perpendicular to the axis. Now by Maxwell's first equation the magnetomotive force around each of these closed loops must equal the axial current (conduction or displacement) through the loop. In the case of a "guide" with an inner conductor, e.g., a coaxial transmission line, this axial current through the H loops is the conduction current in the inner conductor. However for a hollow wave guide having no inner conductor, this axial current must be a displacement current. But an axial-displacement current requires an axial component of E, something not present in a TEM wave. Therefore the TEM wave cannot exist in a single-conductor wave guide.

9.05 Bessel Functions. In solving for the electromagnetic fields within guides of circular cross section, a differential equation known as Bessel's equation is encountered. The solution of the equation leads to *Bessel Functions.* These functions will be considered briefly in this section in preparation for the following section on circular wave-guides. These same functions can be expected to appear in any two-dimensional problem in which there is circular symmetry. Examples of such problems are the vibrations of a circular membrane, the propagation of waves within a circular cylinder, and the electromagnetic field distribution about an infinitely long wire.

The differential equation involved in these problems has the form

$$\frac{d^2P}{d\rho^2} + \frac{1}{\rho}\frac{dP}{d\rho} + \left(1 - \frac{n^2}{\rho^2}\right)P = 0 \tag{9-23}$$

where n is any integer.* One solution to this equation can be obtained by assuming a power series solution

$$P = a_0 + a_1\rho + a_2\rho^2 + \cdots \tag{9-24}$$

Substitution of this assumed solution back into (23) and equating the coefficients of like powers leads to a series solution for the differential equation. For example in the special case where $n = 0$, eq. (23) is

$$\frac{d^2P}{d\rho^2} + \frac{1}{\rho}\frac{dP}{d\rho} + P = 0 \tag{9-25}$$

When the power series (24) is inserted in (25) and the sums of the coefficients of each power of ρ are equated to zero, the following series is obtained

$$\begin{aligned} P = P_1 &= C_1\left[1 - \left(\frac{\rho}{2}\right)^2 + \frac{(\frac{1}{2}\rho)^4}{(2!)^2} - \frac{(\frac{1}{2}\rho)^6}{(3!)^2} + \cdots\right] \\ &= C_1\left[1 - \frac{\rho^2}{2^2} + \frac{\rho^4}{2^2\cdot 4^2} - \frac{\rho^6}{2^2\cdot 4^2\cdot 6^2} + \cdots\right] \\ &= C_1 \sum_{r=0}^{\infty} (-1)^r \frac{(\frac{1}{2}\rho)^{2r}}{(r!)^2} \end{aligned} \tag{9-26}$$

This series is convergent for all values of ρ, either real or complex. It is called Bessel's function of the *first kind* of *order zero* and is denoted by the symbol

$$J_0(\rho)$$

* If n is not restricted to integral values, the symbol ν is used. *See* Appendix II.

The zero order refers to the fact that it is the solution of (23) for the case of $n = 0$. The corresponding solutions for $n = 1, 2, 3$, etc., are designated $J_1(\rho)$, $J_2(\rho)$, $J_3(\rho)$, where the subscript n denotes the order of the Bessel function. Since eq. (23) is a second-order differential equation, there must be two linearly independent solutions for each value of n. The second solution may be obtained in a manner somewhat similar to that used for the first, but starting with a slightly different series that is suitably manipulated to yield a solution.* This second solution is known as Bessel's function of the *second kind*, or Neumann's function, and is designated by the symbol†

$$N_n(\rho)$$

where again n indicates the order of the function. For the zero order of this solution of the second kind, the following series is obtained

$$N_0(\rho) = \frac{2}{\pi}\left\{\ln\left(\frac{\rho}{2}\right) + \gamma\right\} J_0(\rho) - \frac{2}{\pi}\sum_{r=1}^{\infty}(-1)^r \frac{(\tfrac{1}{2}\rho)^{2r}}{(r!)^2}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{r}\right) \quad (9\text{-}27)$$

The complete solution of (25) is then

$$P = AJ_0(\rho) + BN_0(\rho) \quad (9\text{-}28)$$

A plot of $J_0(\rho)$ and $N_0(\rho)$ is shown in Fig. 9-4. Because all the Neumann functions become infinite at $\rho = 0$, these second solutions cannot be used for any physical problem in which the origin is included, as for example the hollow wave guide problem.

It is apparent that [except near the origin for $N_0(\rho)$] these curves bear a marked similarity to damped cosine and sine curves. Indeed, for large values of (ρ) these functions do approach the sinusoidal forms. As ρ becomes very large

$$J_0(\rho) \rightarrow \sqrt{\frac{2}{\pi\rho}}\cos\left(\rho - \frac{\pi}{4}\right) \quad (9\text{-}29)$$

$$N_0(\rho) \rightarrow \sqrt{\frac{2}{\pi\rho}}\sin\left(\rho - \frac{\pi}{4}\right) \quad (9\text{-}30)$$

* N. W. McLachlan, *Bessel Functions for Engineers*, Oxford University Press, New York, 1934.

† The symbol $Y(\rho)$ is used in some texts and tables. It should be noted that there are other forms for this second solution which differ by a constant from the one given.

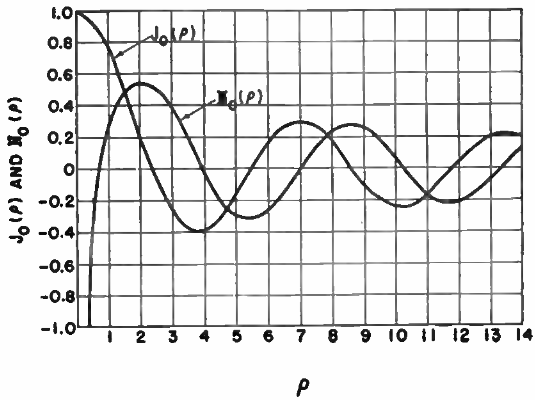


FIG. 9-4. Zero-order Bessel functions of the first and second kinds.

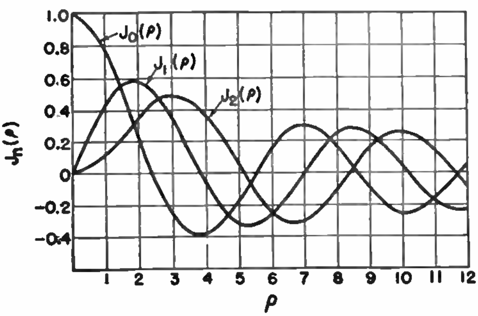


FIG. 9-5a. Higher-order Bessel functions of the first kind.

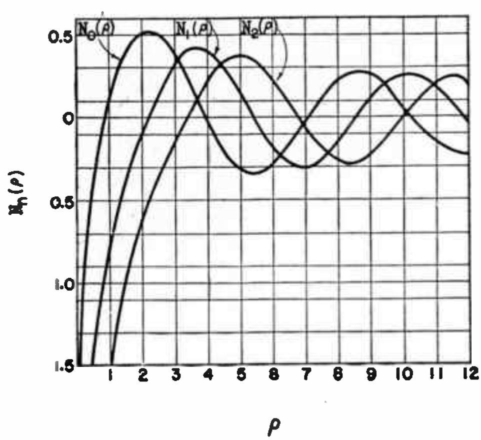


FIG. 9-5b. Higher-order Bessel functions of the second kind (Neumann functions).

Figs. 9-5(a) and (b) show Bessel functions of the first and second kinds for the higher orders. A further discussion of these functions is given in Appendix II.

9.06 Solution of the Field Equations: Cylindrical Co-ordinates. The method of solution of the electromagnetic equations for guides of circular cross section is similar to that followed for rectangular guides. However, in order to simplify the application of the boundary conditions (electric field tangential to the surface equals zero), it is expedient to express the field equations and the wave equations in the cylindrical co-ordinate system.

In cylindrical co-ordinates in a nonconducting region (and again assuming variations with time and z to be given by $e^{j\omega t - \bar{\gamma} z}$), Maxwell's equations are

$$\left.\begin{aligned}
\frac{\partial H_z}{\rho\partial\phi} + \bar{\gamma}H_\phi &= j\omega\epsilon E_\rho \\
\frac{\partial E_z}{\rho\partial\phi} + \bar{\gamma}E_\phi &= -j\omega\mu H_\rho \\
-\bar{\gamma}H_\rho - \frac{\partial H_z}{\partial\rho} &= j\omega\epsilon E_\phi \\
-\bar{\gamma}E_\rho - \frac{\partial E_z}{\partial\rho} &= -j\omega\mu H_\phi \\
\frac{1}{\rho}\left(\frac{\partial(\rho H_\phi)}{\partial\rho} - \frac{\partial H_\rho}{\partial\phi}\right) &= j\omega\epsilon E_z \\
\frac{1}{\rho}\left(\frac{\partial(\rho E_\phi)}{\partial\rho} - \frac{\partial E_\rho}{\partial\phi}\right) &= -j\omega\mu H_z
\end{aligned}\right\} \tag{9-31}$$

These equations can be combined to give

$$\left.\begin{aligned}
h^2H_\rho &= j\frac{\omega\epsilon}{\rho}\frac{\partial E_z}{\partial\phi} - \bar{\gamma}\frac{\partial H_z}{\partial\rho} \\
h^2H_\phi &= -j\omega\epsilon\frac{\partial E_z}{\partial\rho} - \frac{\bar{\gamma}}{\rho}\frac{\partial H_z}{\partial\phi} \\
h^2E_\rho &= -\bar{\gamma}\frac{\partial E_z}{\partial\rho} - j\frac{\omega\mu}{\rho}\frac{\partial H_z}{\partial\phi} \\
h^2E_\phi &= -\frac{\bar{\gamma}}{\rho}\frac{\partial E_z}{\partial\phi} + j\omega\mu\frac{\partial H_z}{\partial\rho} \\
h^2 &= \bar{\gamma}^2 + \omega^2\mu\epsilon
\end{aligned}\right\} \tag{9-32}$$

The wave equation in cylindrical co-ordinates is

$$\frac{\partial^2 E}{\partial\rho^2} + \frac{1}{\rho^2}\frac{\partial^2 E}{\partial\phi^2} + \frac{\partial^2 E}{\partial z^2} + \frac{1}{\rho}\frac{\partial E}{\partial\rho} = \mu\epsilon\ddot{E} \tag{9-33}$$

Proceeding in a manner similar to that followed in the rectangular case, let

$$E = P(\rho) \cdot Q(\phi) \cdot e^{-\bar{\gamma}z+j\omega t} \tag{9-34}$$

where $P(\rho)$ is a function of ρ alone and $Q(\phi)$ is a function of ϕ alone. Substituting the expression for E_z in the wave equation gives

$$Q\frac{d^2P}{d\rho^2} + \frac{Q}{\rho}\frac{dP}{d\rho} + \frac{P}{\rho^2}\frac{d^2Q}{d\phi^2} + PQ\bar{\gamma}^2 + \omega^2\mu\epsilon PQ = 0$$

Divide by PQ,

$$\frac{1}{P}\frac{d^2P}{d\rho^2} + \frac{1}{\rho P}\frac{dP}{d\rho} + \frac{1}{Q\rho^2}\frac{d^2Q}{d\phi^2} + h^2 = 0 \tag{9-35}$$

As before, eq. (35) can be broken up into two ordinary differential equations

$$\frac{d^2Q}{d\phi^2} = -n^2Q \tag{9-36}$$

$$\frac{d^2P}{d\rho^2} + \frac{1}{\rho}\frac{dP}{d\rho} + \left(h^2 - \frac{n^2}{\rho^2}\right)P = 0 \tag{9-37}$$

where n is a constant. The solution of eq. (36) is

$$Q = (A_n \cos n\phi + B_n \sin n\phi) \tag{9-38}$$

Dividing through by h^2, eq. (37) is transformed into

$$\frac{d^2P}{d(\rho h)^2} + \frac{1}{(\rho h)}\frac{dP}{d(\rho h)} + \left[1 - \frac{n^2}{(\rho h)^2}\right]P = 0 \tag{9-39}$$

This is a standard form of Bessel's equation in terms of (ρh). Using only the solution that is finite at $(\rho h) = 0$, gives

$$P(\rho h) = J_n(\rho h) \tag{9-40}$$

where $J_n(\rho h)$ is Bessel's function of the first kind of order n. Substituting the solutions (38) and (40) in (34),

$$E_z = J_n(\rho h)(A_n \cos n\phi + B_n \sin n\phi)\, e^{-\bar{\gamma}z+j\omega t} \tag{9-41}$$

The solution for H_z will have exactly the same form as for E_z and can therefore be written

$$H_z = J_n(\rho h)(C_n \cos n\phi + D_n \sin n\phi)\, e^{-\bar{\gamma}z+j\omega t} \tag{9-42}$$

For TM waves the remaining field components can be obtained by inserting (41) into (32). For TE waves (42) must be inserted into the set corresponding to (32).

9.07 TM and TE Waves in Circular Guides. As in the case of rectangular guides, it is convenient to divide the possible solutions for circular guides into transverse magnetic and transverse electric waves. For the TM waves H_z is identically zero and the wave equation for E_z is used. The boundary conditions require that E_z must vanish at the surface of the guide. Therefore, from (41)

$$J_n(ha) = 0 \tag{9-43}$$

where a is the radius of the guide. There is an infinite number of possible *TM* waves corresponding to the infinite number of roots of (43). As before $h^2 = \bar{\gamma}^2 + \omega^2\mu\epsilon$, and, as in the case of rectangular guides, h^2 must be less than $\omega^2\mu\epsilon$ for transmission to occur. This means that h must be small or else extremely high frequencies will be required. This in turn means that only the first few roots of (43) will be of practical interest. The first few roots are

$$\left.\begin{matrix} (ha)_{0,1} = 2.405 & (ha)_{1,1} = 3.85 \\ (ha)_{0,2} = 5.52 & (ha)_{1,2} = 7.02 \end{matrix}\right\} \tag{9-44}$$

The first subscript refers to the value of n and the second refers to the roots in their order of magnitude. The various TM waves will be referred to as $TM_{0,1}$, $TM_{1,2}$, etc.

Since $\bar{\gamma} = \sqrt{h^2 - \omega^2\mu\epsilon}$, this gives for $\bar{\beta}$

$$\bar{\beta}_{nm} = \sqrt{\omega^2\mu\epsilon - h^2{}_{nm}}$$

The cut-off or critical frequency below which transmission of a wave will not occur is

$$f_c = \frac{h_{nm}}{2\pi\sqrt{\mu\epsilon}}$$

where

$$h_{nm} = \frac{(ha)_{nm}}{a}$$

The phase velocity is

$$\bar{v} = \frac{\omega}{\bar{\beta}} = \frac{\omega}{\sqrt{\omega^2\mu\epsilon - h^2{}_{nm}}}$$

From eqs. (32) the various components of TM waves can be computed in terms of E_z. The expressions for TM waves in circular guides are

$$E_z = A_n J_n(h\rho) \cos n\phi$$

$$\left.\begin{aligned} H_\rho &= -\frac{jA_n\omega\epsilon n}{h^2\rho} J_n(\rho h) \sin n\phi \\ H_\phi &= -\frac{jA_n\omega\epsilon}{h} J_n'(\rho h) \cos n\phi \\ E_\rho &= \frac{\bar\beta}{\omega\epsilon} H_\phi \\ E_\phi &= -\frac{\bar\beta}{\omega\epsilon} H_\rho \end{aligned}\right\} \tag{9-45}$$

The variations of each of these field components with time and in the z direction are shown by multiplying each of the expressions of (45) by the factor $e^{j(\omega t - \bar\beta z)}$. In the original expression (41) for E_z, the arbitrary constant B_n has been put equal to zero. The relative amplitudes of A_n and B_n determine the orientation of the fields in the guide, and for a circular guide and any particular value of n, the $\phi = 0$ axis can always be oriented to make either A_n or B_n equal to zero.

For *transverse electric* waves E_z is identically zero and H_z is given by eq. (42). By substituting (42) into eqs. (32), the remaining field components can be found. The expressions for TE waves in circular guides are

$$\left.\begin{aligned} H_z &= C_n J_n(h\rho) \cos n\phi \\ H_\rho &= \frac{-j\bar\beta C_n}{h} J_n'(h\rho) \cos n\phi \\ H_\phi &= \frac{jn\bar\beta C_n}{h^2\rho} J_n(h\rho) \sin n\phi \\ E_\rho &= \frac{\omega\mu}{\bar\beta} H_\phi \\ E_\phi &= -\frac{\omega\mu}{\bar\beta} H_\rho \end{aligned}\right\} \tag{9-46}$$

where the factor $e^{j(\omega t - \bar\beta z)}$ is understood.

The boundary conditions to be met for TE waves are that $E_\phi = 0$ at $\rho = a$. From (32) E_ϕ is proportional to $\partial H_z/\partial\rho$, and therefore to $J_n'(h\rho)$, where the prime denotes the derivative with respect to $(h\rho)$. Therefore, for TE waves the boundary conditions require that

$$J_n'(ha) = 0 \tag{9-47}$$

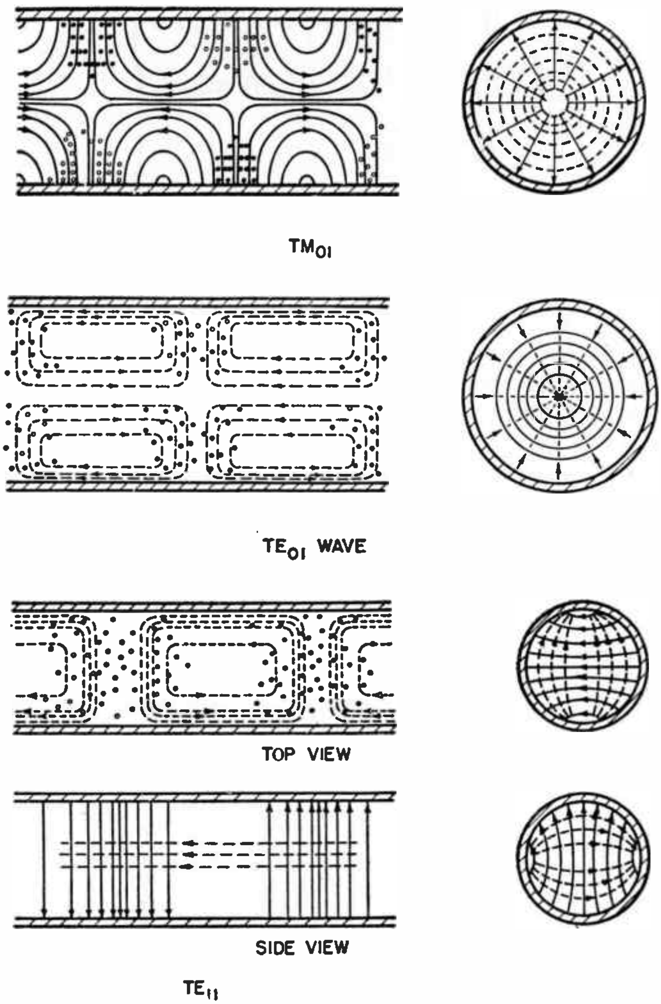


Fig. 9-6. TE and TM waves in circular guides.

and it is the roots of (47), which must be determined. The first few of these roots are

$$\left.\begin{array}{ll} (ha)'_{0\,1} = 3.83 & (ha)'_{1,1} = 1.84 \\ (ha)'_{c,2} = 7.02 & (ha)'_{1,2} = 5.33 \end{array}\right\} \quad (9\text{-}48)$$

The corresponding TE waves are referred to as $TE_{0,1}$, $TE_{1,1}$, and so on. The equations for f_c, $\bar{\beta}$, $\bar{\lambda}$, and $\bar{v}$ are identical to those for the TM waves. It is understood, of course, that the roots of eq. (47) are to be used in connection with TE waves only.

Inspection of eqs. (44) and (48) shows that the wave having the lowest cut-off frequency is the $TE_{1,1}$ wave. The wave having the next lowest cut-off frequency is the $TM_{0,1}$. Some representative TM and TE waves are shown in Fig. 9-6.

9.08 Wave Impedances and Characteristic Impedances. The wave impedances at a point have been defined by eqs. (7-50). For waves guided by transmission lines or wave guides, interest centers on the wave impedance which is seen when looking in the direction of propagation, that is, along the z axis. Inspection of expressions (12) for the transverse field components of a TM wave in a rectangular guide shows that

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\sqrt{E_x^2 + E_y^2}}{\sqrt{H_x^2 + H_y^2}} = \frac{\bar{\beta}}{\omega\epsilon}$$

Therefore
$$Z_{xy} = Z_{yx} = \frac{\bar{\beta}}{\omega\epsilon} = Z_z \tag{9-49}$$

The wave impedances looking in the z direction are equal and may be put equal to Z_z, where

$$Z_z = \frac{E_{\text{trans}}}{H_{\text{trans}}} = \frac{\sqrt{E_x^2 + E_y^2}}{\sqrt{H_x^2 + H_y^2}} \tag{9-50}$$

is the ratio of the total transverse electric intensity to the total transverse magnetic intensity.

A similar inspection of eqs. (45) for TM waves in circular guides shows that for them also

$$Z_z = Z_{\rho\phi} = -Z_{\phi\rho} = \frac{\bar{\beta}}{\omega\epsilon} \tag{9-51}$$

It is seen that for TM waves in rectangular or circular guides, or indeed in cylindrical guides of any cross section, the wave impedance in the direction of propagation is *constant over the cross section of the guide,* and is the same for guides of different shapes. Recalling that

$$\bar{\beta} = \sqrt{\omega^2\mu\epsilon - h^2}$$

and that the cut-off angular frequency ω_c has been defined as that frequency that makes

$$\omega_c^2\mu\epsilon = h^2$$

it follows that $\bar{\beta}$ can be expressed in terms of the cut-off frequency by

$$\bar{\beta} = \omega\sqrt{\mu\epsilon}\sqrt{1 - (\omega_c^2/\omega^2)} \tag{9-52}$$

Then from (49) or (51) the wave impedance in the z direction for TM waves is

$$\begin{aligned} Z_z(\text{TM}) &= \sqrt{\frac{\mu}{\epsilon}}\sqrt{1 - (\omega_c^2/\omega^2)} \\ &= \eta\sqrt{1 - (\omega_c^2/\omega^2)} \end{aligned} \tag{9-53}$$

Thus for any cylindrical guide the wave impedance for TM waves is dependent only on the intrisic impedance of the dielectric and the ratio of the frequency to the cut-off frequency.

For TE waves the same conclusion can be reached. However for TE waves it is found that

$$Z_z(\text{TE}) = \frac{\omega\mu}{\bar{\beta}} = \frac{\eta}{\sqrt{1 - (\omega_c^2/\omega^2)}} \tag{9-54}$$

For TEM waves between parallel planes or on ordinary parallel-wire or coaxial transmission lines the cut-off frequency is zero, and the wave impedance reduces to

$$Z_z(\text{TEM}) = \eta \tag{9-55}$$

The dependence of $\bar{\beta}$ on the ratio of frequency to cut-off frequency as shown by (52) affects the phase velocity and the wavelength in a corresponding manner. Thus the phase or wave velocity in a cylindrical guide of any cross section is given by

$$\bar{v} = \frac{\omega}{\bar{\beta}} = \frac{1}{\sqrt{\mu\epsilon}\sqrt{1 - (\omega_c^2/\omega^2)}} = \frac{v_0}{\sqrt{1 - (\omega_c^2/\omega^2)}} \tag{9-56}$$

where $v_0 = 1/\sqrt{\mu\epsilon}$, and μ and ϵ are the constants of the dielectric. The wavelength in the guide, measured in the direction of propagation, is

$$\begin{aligned} \bar{\lambda} = \frac{\bar{v}}{f} = \frac{2\pi}{\bar{\beta}} &= \frac{1}{f\sqrt{\mu\epsilon}\sqrt{1 - (\omega_c^2/\omega^2)}} \\ &= \frac{\lambda_0}{\sqrt{1 - (\omega_c^2/\omega^2)}} \end{aligned} \tag{9-57}$$

where λ_0 is the wavelength of a TEM wave of frequency f in a dielectric having the constants μ and ϵ. Since $\omega_c^2/\omega^2 = \lambda_0^2/\lambda_c^2$ it follows that

$$\bar{\lambda} = \frac{\lambda_0\lambda_c}{\sqrt{\lambda_c^2 - \lambda_0^2}}$$

or

$$\lambda_0 = \frac{\bar{\lambda}\lambda_c}{\sqrt{\lambda_c^2 + \bar{\lambda}^2}} \tag{9-58}$$

A quantity of great usefulness in connection with ordinary two-conductor transmission lines is the (integrated) *characteristic impedance*, Z_0, of the line. For such lines, Z_0 can be defined in terms of the voltage-current ratio or in terms of the power transmitted for a given voltage or a given current. That is, for an infinitely long line

$$Z_0 = \frac{V}{I}; \qquad Z_0 = \frac{2W}{II^*}; \qquad Z_0 = \frac{VV^*}{2W} \tag{9-59}$$

where V and I are peak values in time. For ordinary transmission lines these definitions are equivalent, but for wave guides they lead to three values that depend upon the guide dimensions in the same way, but which differ by a constant.

For example, consider the three definitions given by (59) for the case of the $TE_{1,0}$ mode in a rectangular guide (Fig. 9-1). The voltage will be taken as the maximum voltage from the lower face of the guide to the upper face. This occurs at $x = a/2$ and has a value

$$V_m = \int_0^b E_y(\max)\, dy = bE_y(\max) = \frac{-j\omega\mu baC}{\pi} \tag{9-60}$$

The longitudinal linear current density in the lower face is

$$J_z = -H_x = -\frac{j\beta aC}{\pi}\sin\frac{\pi x}{a} \tag{9-61}$$

The total longitudinal current in the lower face is

$$I = \int_0^a J_z\, dx = \frac{-j2a^2\beta C}{\pi^2}$$

Then the "integrated" characteristic impedance by the first definition is

$$Z_0(V, I) = \frac{\pi b}{2a}\frac{\omega\mu}{\beta} = \frac{\pi b}{2a}Z_z = \frac{\pi b\eta}{2a\sqrt{1 - (f_c^2/f^2)}} \tag{9-62}$$

In terms of the second definition, the characteristic impedance for the $TE_{1,0}$ wave in a rectangular guide is found to be

$$Z_0(W, I) = \frac{\pi^2 b}{8a} Z_z = \frac{\pi}{4} Z_0(V, I) \tag{9-63}$$

In terms of the third definition the integrated characteristic impedance is

$$Z_0(W, V) = \frac{2b}{a} Z_z = \frac{4}{\pi} Z_0(V, I) \tag{9-64}$$

In the next section the utility of the concept of characteristic impedance for cylindrical wave guides will be demonstrated.

9.09 Transmission Line Analogy for Waveguides. There exists a useful analogy between the electric and magnetic field intensities of TM and TE waves and the voltages and currents on suitably loaded transmission lines. This analogy enables the engineer to draw "equivalent circuits," which are often helpful to him in dealing with unfamiliar electromagnetic problems.

For TM waves ($H_z = 0$) in rectangular co-ordinates the field equations are

$$\left.\begin{aligned} \frac{\partial H_y}{\partial z} &= -j\omega\epsilon E_x & \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -j\omega\mu H_x \\ \frac{\partial H_x}{\partial z} &= j\omega\epsilon E_y & \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -j\omega\mu H_y \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= j\omega\epsilon E_z & \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -j\omega\mu H_z \\ & \frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} = 0 \end{aligned}\right\} \tag{9-65}$$

Now since $H_z = 0$,

$$\text{curl}_{xy}\, \bar{E} = 0$$

That is, in the x-y plane the electric field has no curl (the voltage around a closed path is zero) and so in this plane E may be written as the gradient of some scalar potential V. Then

$$E_x = -\frac{\partial V}{\partial x} \qquad E_y = -\frac{\partial V}{\partial y} \tag{9-66}$$

From the first equations of (65) and (66) and using (3)

$$\frac{\partial}{\partial z}\left(\frac{j\omega\epsilon}{h^2}\frac{\partial E_z}{\partial x}\right) = -j\omega\epsilon\frac{\partial V}{\partial x}$$

whence

$$\frac{\partial}{\partial z}\left(\frac{j\omega\epsilon}{h^2}E_z\right) = -j\omega\epsilon V \tag{9-67}$$

From the fifth of (65) and the first of (66) and using (3)

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\omega^2\mu\epsilon}{h^2}\frac{\partial E_z}{\partial x}$$

whence

$$\frac{\partial V}{\partial z} = \left(\frac{\omega^2\mu\epsilon}{h^2} - 1\right)E_z$$

$$= -\left(j\omega\mu + \frac{h^2}{j\omega\epsilon}\right)\left(\frac{j\omega\epsilon}{h^2}E_z\right) \tag{9-68}$$

The quantity $j\omega\epsilon E_z$ is the longitudinal displacement current density and $1/h^2$ has the dimensions of area, so $j\omega\epsilon E_z/h^2$ represents

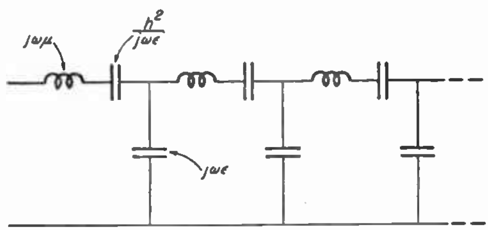


Fig. 9-7. Equivalent transmission line circuit representation for TM waves.

a current in the z direction and will be designated by I_z. Then (67) and (68) become

$$\frac{\partial I_z}{\partial z} = -j\omega\epsilon V \qquad \frac{\partial V}{\partial z} = -\left[j\omega\mu + \frac{h^2}{j\omega\epsilon}\right]I_z \tag{9-69}$$

These are the differential equations for a lossless transmission line having a series impedance per unit length $Z = j\omega\mu + (h^2/j\omega\epsilon)$ and a shunt admittance per unit length $Y = j\omega\epsilon$. The "equivalent circuit" for such a transmission line is that shown in Fig. 9-7.

For TE waves the two equations of interest from the set corresponding to (65) are

$$\frac{\partial E_y}{\partial z} = j\omega\mu H_x \qquad \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\epsilon E_x \tag{9-70}$$

Since $E_z = 0$, $\text{curl}_{xy} H = 0$; then in the x-y plane it is possible to define a scalar (magnetic) potential U such that

$$H_x = -\frac{\partial U}{\partial x} \qquad H_y = -\frac{\partial U}{\partial y} \tag{9-71}$$

From (70) and (71) and using eqs (3)

$$\frac{\partial}{\partial z}\left(\frac{j\omega\mu}{h^2}\frac{\partial H_z}{\partial x}\right) = -j\omega\mu\frac{\partial U}{\partial x} \qquad \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \frac{\omega^2\mu\epsilon}{h^2}\frac{\partial H_z}{\partial y}$$

whence

$$\frac{\partial}{\partial z}\left(\frac{j\omega\mu}{h^2}H_z\right) = -j\omega\mu U \qquad \frac{\partial U}{\partial z} = -\left(\frac{h^2}{j\omega\mu} + j\omega\epsilon\right)\left(\frac{j\omega\mu}{h^2}H_z\right) \tag{9-72}$$

The quantity $j\omega\mu H_z/h^2$ has the dimensions of voltage and U has the dimensions of current, so (72) may be written

$$\frac{\partial V_1}{\partial z} = -ZI_1 \qquad \frac{\partial I_1}{\partial z} = -YV_1 \tag{9-73}$$

where now $\qquad V_1 = \dfrac{j\omega\mu H_z}{h^2} \qquad I_1 = U$

$$Z = j\omega\mu \qquad Y = j\omega\epsilon + \frac{h^2}{j\omega\mu}$$

The "equivalent circuit" for TE waves is shown in Fig. 9-8.

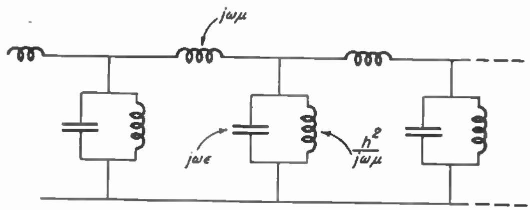


FIG. 9-8. Equivalent transmission line circuit representation for TE waves.

The "loaded" transmission line circuits of Figs. 9-7 and 9-8 have high-pass filter characteristics. The cut-off frequency for the line of Fig. 9-7 occurs when the series reactance equals zero, whereas for the line of Fig. 9-8, the cut-off frequency is that which makes the shunt susceptance equal to zero. Both of these equalities require that

$$h^2 = \omega_c{}^2\mu\epsilon$$

as was already obtained from wave theory. The characteristic impedance of the line of Fig. 9-7 is

$$Z_0(\mathrm{TM}) = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{j\omega\mu + (h^2/j\omega\epsilon)}{j\omega\epsilon}} = \sqrt{\frac{\mu}{\epsilon}}\sqrt{1 - \frac{\omega_c^2}{\omega^2}}$$
$$= Z_z(\mathrm{TM}) \qquad (9\text{-}74)$$

The characteristic impedance of the line of Fig. 9-8 is

$$Z_0(\mathrm{TE}) = \sqrt{\frac{j\omega\mu}{j\omega\epsilon + (h^2/j\omega\mu)}}$$
$$= \sqrt{\frac{\mu}{\epsilon}}\sqrt{\frac{1}{1 - (\omega_c^2/\omega^2)}} = Z_z(\mathrm{TE}) \qquad (9\text{-}75)$$

The characteristic impedances of the equivalent transmission lines are equal to the corresponding wave impedances as would be expected.

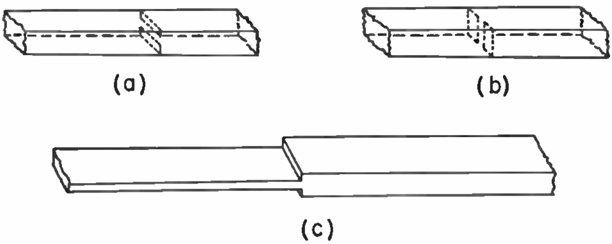


FIG. 9-9. Typical discontinuities in wave guides: (a) Iris with edges perpendicular to E. (b) Iris with edges parallel to E. (c) Change of wave-guide dimensions.

The concept of a waveguide as an equivalent transmission line with a certain characteristic impedance and propagation constant is a powerful tool in the solution of many wave guide problems because it enables the engineer to obtain the solution by means of well-known circuit and transmission line theory. For example, the wave-guide problems, illustrated in Fig. 9-9, can be solved in terms of the "equivalent circuits," shown in 9-10. Thus an iris in a wave guide behaves as a shunt reactance on the equivalent line. The reactance is positive or inductive when the edges of the iris are parallel to **E** (Fig. 9-9b); it is negative or capacitive when the edges are perpendicular to **E** (Fig. 9-9a). An abrupt change in wave-guide dimensions (Fig. 9-9c) is represented by the equivalent circuit of Fig. 9-10c, in which two equivalent transmission lines are joined together, with an appropriate reactance shunted across the junction.

In these examples the calculation of the actual value of shunting reactance to be used in any particular case is, of course, a field problem. However it is a field problem which can be solved in a fairly straightforward manner by matching solutions at the boundary.* The procedure is to represent the field at the junction or discontinuity by the sum of principal and higher order waves, the relative

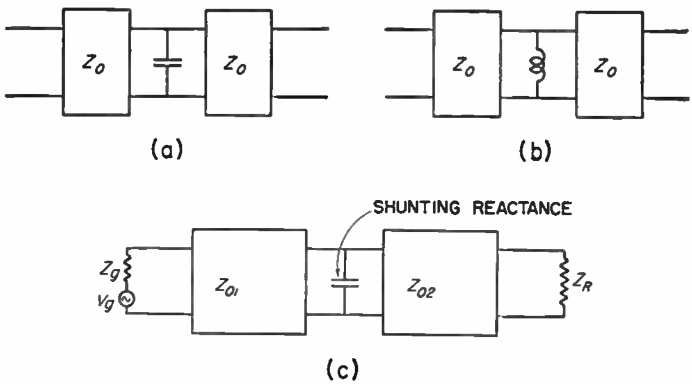


FIG. 9-10. "Equivalent circuits" for the wave-guide discontinuities illustrated in Fig. 9-9.

amplitudes of which are obtained by matching the tangential components of **E** and **H** at the boundary. The higher order waves are set up by the discontinuity and are required in order to meet the boundary conditions. However, in general, they have cut-off frequencies higher than the frequency of transmission and so are attenuated rapidly. The load impedance and the generator are assumed to be sufficiently far removed from the iris or junction to be out of the field of these higher order waves. It is for this reason that the problem can be treated in terms of the effect of the discontinuity on the *principal wave only*, which fact, in turn, makes valid the circuit representation by means of an ordinary transmission line. (Otherwise, a transmission line having a different set of

* J. R. Whinnery and H. W. Jamieson, "Equivalent Circuits for Discontinuities in Transmission Lines," *Proc. IRE*, **32**, 98–114, February, 1944; S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand, New York, 1943, p. 492.

"constants" for each mode would be required.) It is found that the higher order modes make no contribution to the voltage at the junction, and therefore the total voltage is just the voltage of the principal wave. However, in the region of the junction the higher order waves do make contributions to the current. Although the *total* current must be continuous across the junction, the principal-wave current is discontinuous by the amount of the higher order mode current. This discontinuity of principal-wave current is accounted for by the effect of the equivalent shunting reactance.

The values for equivalent shunting reactances have been calculated in terms of the iris or junction dimensions for many cases and may be found in handbooks.* Using this known reactance, together with the known characteristic impedance for the guide [as given, for example, by eq. (62) for the $TE_{1,0}$ wave], the wave-guide problems of Fig. 9-9 are readily solved in terms of the well-known circuit problems of Fig. 9-10.

9.10 Attenuation Factor and *Q* of Wave Guides. In solving Maxwell's equations for the region within rectangular or circular wave guides, the assumptions were made that the dielectric was lossless and that the walls of the guide were perfectly conducting. Under these conditions an expression was obtained for the propagation constant $\bar{\gamma}$, which was

$$\bar{\gamma} = \sqrt{h^2 - \omega^2\mu\epsilon}$$

The quantity h^2 is a real number, the value of which depends upon the guide dimensions and the order of the mode being considered. For example, for rectangular guides h^2 is given by

$$h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2$$

For frequencies below cut-off $\omega^2\mu\epsilon$ is less than h^2, and $\bar{\gamma}$ is a real number, which is then put equal to $\bar{\alpha}$. That is, below cut-off,

$$\bar{\gamma} = \bar{\alpha} = \sqrt{h^2 - \omega^2\mu\epsilon}$$

In general, for frequencies well below cut-off, $\bar{\alpha}$ is a large number, and the fields decrease exponentially at a rapid rate. At the cut-off frequency $\bar{\gamma}$ becomes equal to zero; for all frequencies above

* N. Marcuvitz, *Waveguide Handbook*, Radiation Laboratory Series, Vol. 10, McGraw-Hill, New York, 1948.

cut-off $\bar{\gamma}$ is a pure imaginary and the attenuation constant $\bar{\alpha}$ is zero. This result is correct for the assumed conditions. However, whereas the dielectric within the guide may be very nearly lossless (air for example), the walls of an actual guide do have some loss. Therefore a finite, though perhaps small, value of attenuation would be expected in the range of frequencies above cut-off.

The actual attenuation factor for waves propagating within cylindrical guides may be calculated to a very good approximation by the method already outlined for parallel-plane guides. In this approach it is assumed that the finite conductivity of the walls will have only a small effect on the configurations within the guide; in particular the magnetic field tangential to the wall is expected to depend only slightly on the conductivity of the walls. This is very nearly true as long as the conductivity is high, as it is for metals. Then the tangential magnetic intensity computed for perfectly conducting walls is used to determine the linear current density in the walls. This linear current density, squared and multiplied by the actual surface resistance of the walls, gives the actual power loss per unit area in the walls. The attenuation factor *in the range of propagation* is then given by

$$\alpha = \frac{\text{power lost per unit length}}{2 \times \text{power transmitted.}} \tag{9-76}$$

The power transmitted is obtained by integrating the axial component of the Poynting vector over the cross section of the guide. Because the transverse components of E and H have been found to be in phase and normal to each other, the axial Poynting vector is given simply as

$$P_z = \tfrac{1}{2}|E_{\text{trans}}||H_{\text{trans}}|$$

Using (50), this may be written

$$P_z = \tfrac{1}{2}Z_z|H_{\text{trans}}|^2$$

or

$$P_z = \frac{1}{2Z_z}|E_{\text{trans}}|^2$$

The total power transmitted is

$$W = \tfrac{1}{2}Z_z \int_{\text{area}} |H_{\text{trans}}|^2 \, da \tag{9-77}$$

where the integration is over the cross-section area of the guide. The power lost per unit length of guide is

$$W_{\text{lost}} = \tfrac{1}{2}R_s \int_{\text{surf}} |J|^2 \, da$$
$$= \tfrac{1}{2}R_s \int_{\text{surf}} |H_{\text{tang}}|^2 \, da \qquad (9\text{-}78)$$

where the integration is taken over the wall surface of a unit length of the guide.

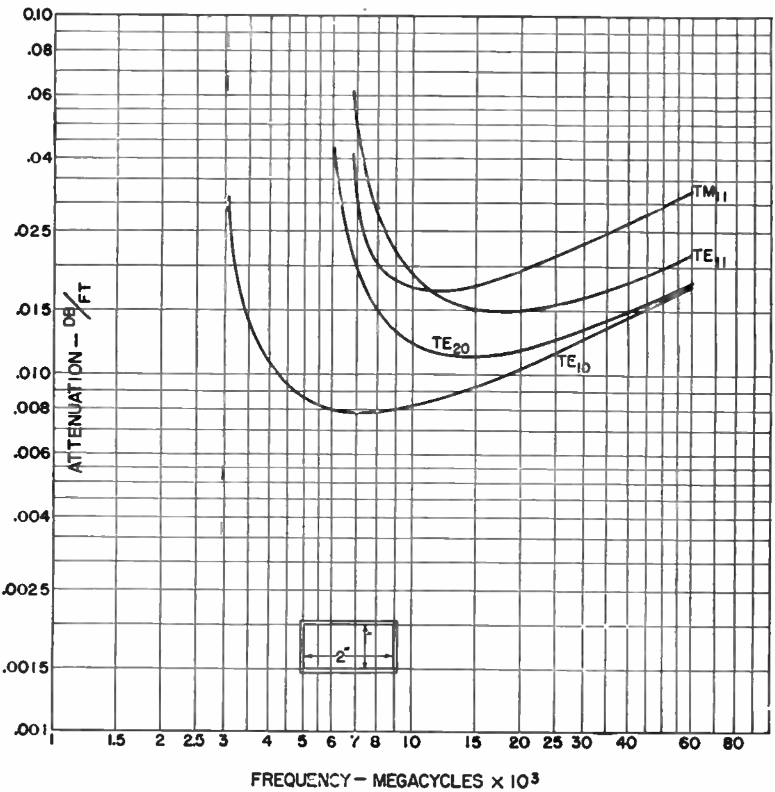


FIG. 9-11. Attenuation vs. frequency curves for various modes in a typical rectangular brass guide.

Formulas for attenuation factors for rectangular and circular guides computed by this or equivalent methods can be found in many textbooks and handbooks. For the dominant $TE_{1,0}$ mode in rectangular guides the result is

$$\left.\begin{aligned}\alpha &= \frac{R_s}{b\eta_0 K}\left[1+\frac{2bf_c^2}{af^2}\right] \quad \text{nepers/m}\\ &= \frac{8.7R_s}{b\eta_0 K}\left[1+\frac{2bf_c^2}{af^2}\right] \quad \text{db/m}\end{aligned}\right\} \tag{9-79}$$

where $K = \sqrt{1-\frac{f_c^2}{f^2}} \qquad R_s = \sqrt{\frac{\omega\mu_m}{2\sigma_m}}$

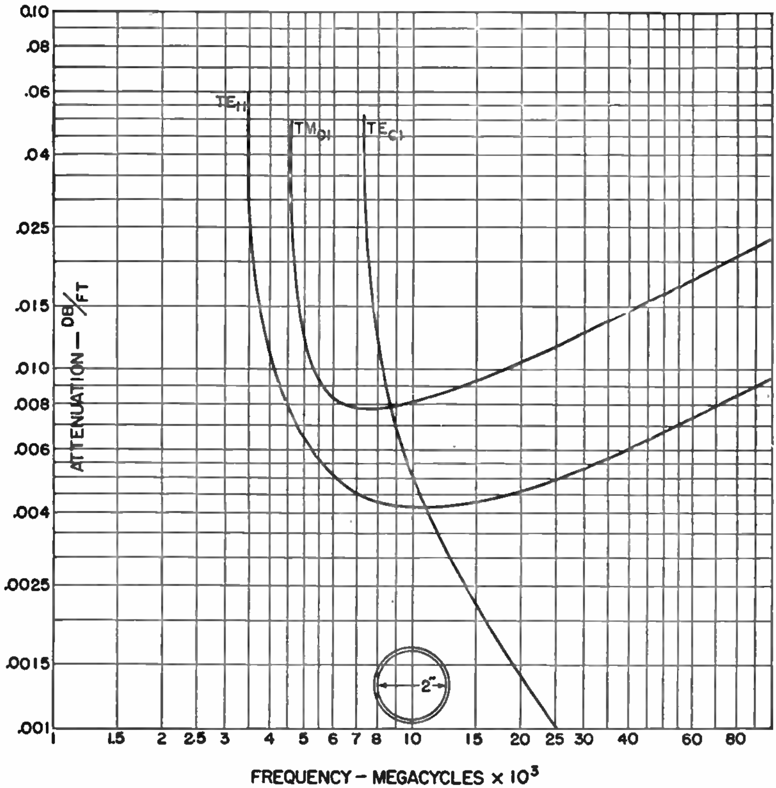


FIG. 9-12. Attenuation vs. frequency curves for various modes in a circular guide.

Attenuation *vs.* frequency curves are sketched for typical rectangular brass guides in Fig. 9-11, and for circular guides in Fig. 9-12. The attenuation is very high near the cut-off frequency, but decreases to a quite low value at frequencies somewhat above cut-off. For TM waves in cylindrical guides of any shape, there is a frequency of minimum attenuation that occurs at $\sqrt{3}$ times the critical frequency. In general, for still higher frequencies, the attenuation again increases (approximately as the square root of frequency for very high frequencies). An exception to this last statement appears to occur for the $TE_{0,m}$ waves in circular guides. For these waves in perfectly circular guides the wall currents decrease as the frequency increases, and the attenuation theoretically decreases indefinitely with increasing frequency. Unfortunately, slight deformations of the guide produce additional wall currents that nullify this desirable characteristic.

A quantity closely related to the attenuation factor α is the "quality" factor Q. For ordinary transmission lines carrying the TEM wave Q was found to be expressible in terms of the secondary line constants by

$$Q = \frac{\beta}{2\alpha}$$

Making use of relation (76), the expression for Q for ordinary transmission lines becomes

$$Q = \beta\left(\frac{\text{power transmitted}}{\text{power lost per unit length}}\right)$$

$$= \frac{\omega}{v}\left(\frac{\text{energy transmitted per second}}{\text{energy lost per second per unit length}}\right)$$

But, energy transmitted per second equals stored energy per unit length times v.

$$Q = \omega\,\frac{\text{energy stored per unit length}}{\text{energy lost per unit length per second}} \tag{9-80}$$

Expression (80) may be considered as a general definition for Q, applicable to wave guides as well as to ordinary transmission lines. It should be compared with the circuit definition of Q, which may be stated as

$$Q = \omega\,\frac{\text{energy stored in circuit}}{\text{energy lost per second}} \tag{9-81}$$

For the TEM wave on the lossless or distortionless transmission line,* the velocity v represented both the phase velocity and the

* When the transmission line has distortion, phase velocity and group

group velocity. For wave guides these two velocities are different, the group velocity, v_g, being related to the phase velocity $\bar{v}$ by

$$v_g = \frac{v_0^2}{\bar{v}} \qquad \text{where } v_0 = \frac{1}{\sqrt{\mu\epsilon}} \tag{9-82}$$

For a wave-guide,

energy transmitted per second
$= v_g \times$ (energy stored per unit length)

or

$$\text{energy stored per unit length} = \frac{1}{v_g} \times \text{power transmitted} \tag{9-83}$$

Using (80), (83), and (76), the Q of a wave guide is given by

$$Q = \frac{\omega}{v_g}\left(\frac{\text{power transmitted}}{\text{power lost per unit length}}\right)$$
$$= \frac{\omega}{2\alpha v_g} \tag{9-84}$$

This also may be written in the following equivalent forms

$$Q = \frac{\omega\bar{v}}{2\alpha v_0^2} = \frac{\omega}{2\alpha v_0\sqrt{1 - \omega_c^2/\omega^2}} \tag{9-84a}$$

Because of the low attenuation factors obtainable with wave guides compared to transmission lines, it is possible to construct wave-guide sections having extremely high Q's. This is of importance when such sections are used as resonators, or as the elements of wave-guide filters.

ADDITIONAL PROBLEMS

6. Verify the results obtained in eqs. (63) and (64).

7. Show that for a coil the definition for Q given by (81) reduces to $Q = \omega L/R$.

8. Show that at frequencies much higher than the cut-off frequency, the Q of a rectangular guide carrying the dominant $TE_{1,0}$ wave approaches the value

$$Q \to b\alpha_m$$

where $\alpha_m = \sqrt{\omega\mu_m\sigma_m/2}$ is the attenuation factor for a wave propagating in the metal of the guide walls. (NOTE: Assume $\mu_m \approx \mu_v$.)

---

velocity are not equal. For a thorough discussion of this case see E. A. Guillemin, *Communication Networks*, John Wiley & Sons, Inc., New York, 1935, Vol. II. Also see Appendix I.

BIBLIOGRAPHY

Southworth, G. C., "Hyper-frequency Wave Guides—General Considerations and Experimental Results," *BSTJ*, **15**, 284 (1936).

Barrow, W. L., "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," *Proc. IRE*, **24**, 1298 (1936).

Schelkunoff, S. A., "Transmission Theory of Plane Electromagnetic Waves," *Proc. IRE*, **25**, 1457 (1937).

Chu, L. J. and W. L. Barrow, "Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section," *Proc. IRE*, **26**, 1520 (1938).

Clavier, A. G., "Attenuation and *Q* Factors in Wave Guides," *Electrical Communication*, **23**, No. 4 (1946).

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand Co., New York, 1943.

Bronwell, A. B. and R. E. Beam, *Theory and Application of Microwaves*, McGraw-Hill, New York, 1947.

CHAPTER 10

RADIATION

10.01 Vector Potential in the Electromagnetic Field. In the previous chapters, relations which exist among the electromagnetic field vectors have been studied, but so far no consideration has been given to the sources* of the fields. It is now in order to consider how these fields are related to their *sources*, that is, the charges and currents that produce them. Although it is possible in theory to obtain expressions for the electric and magnetic intensities **E** and **H** directly in terms of the charge and current densities ρ and $\boldsymbol{i}$, such a derivation is, in general, quite difficult. It will be recalled that in the study of the electrostatic field and the steady magnetic field it was found possible, and often simpler, to first set up *potentials* in terms of the charges or currents, and then to obtain the electric or magnetic fields from these potentials. In a similar manner, in the electromagnetic field it turns out to be much simpler first to set up potentials that are related to the charges and currents, and then to obtain **E** and **H** from these potentials.

The first step in this process is that of finding a suitable potential or potentials that satisfy the conditions of the problem. There are several possible ways of doing this, and all of these attacks require a certain amount of educated guesswork. The method chosen here will use the heuristic approach in which the expressions for potentials are guessed. If the guess proves to be correct, it is then possible to demonstrate that these potentials do indeed meet all the requirements of the field equations and of the problem.

The electromagnetic field is produced by charge and current distributions, which vary with time. The electrostatic field and the steady magnetic field can be considered as special cases of the electromagnetic field for which the time variations are reduced to zero. It is reasonable, therefore, to suppose that as the frequency

* The term source is here used in its broader sense (*see* footnote on page 102).

approaches zero, the potential(s) developed for the electromagnetic field will reduce to the scalar potential V of the electrostatic field and the vector potential **A** of the steady magnetic field. Conversely, it is reasonable to expect that the potentials for the electromagnetic field might be obtained by generalizing the static field potentials to account for time variations. This proves to be the case.

In the *electrostatic* field a scalar potential V was set up, from which the electric intensity **E** could be obtained by taking a space derivative, that is,

$$\mathbf{E} = -\operatorname{grad} V \tag{10-1}$$

The potential V was related to the sources of the field (the charges) by

$$V = \frac{1}{4\pi\epsilon} \sum_{i=1}^{i=m} \frac{q_i}{r_i}$$

When the charges were considered to be distributed continuously throughout a volume, the potential was expressed as a function of the charge density ρ by

$$V = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho}{r}\, dV \tag{10-2}$$

where the integration was extended over the whole of the volume* containing charge which contributes to the potential V.

In the *steady magnetic field*, produced by direct currents, a vector potential **A** was set up, and the magnetic field intensity **H** was obtained from **A** by again taking a space derivative, this time the curl.

$$\mathbf{H} = \operatorname{curl} \mathbf{A} \tag{10-3}$$

For a current element $I\,ds$, the contribution to the vector potential was

$$d\mathbf{A} = \frac{1}{4\pi} \frac{I\,ds}{r} \tag{10-4}$$

For a complete (or closed) circuit, the expression for vector potential was

$$\mathbf{A} = \frac{1}{4\pi} \oint \frac{I\,ds}{r} \tag{10-5}$$

* It is not anticipated that any confusion will result from the use of dV for volume element and V for voltage or potential.

When the current is considered to be distributed throughout the cross section of a conductor as a current density $\mathbf{i}$ (rather than concentrated in a thin filament), the expression for vector potential becomes

$$\mathbf{A} = \frac{1}{4\pi} \int_{\text{vol}} \frac{\mathbf{i}}{r}\, dV \tag{10-6}$$

where the integration is extended over the entire volume of conductors containing current density $\mathbf{i}$ (Fig. 10-1).

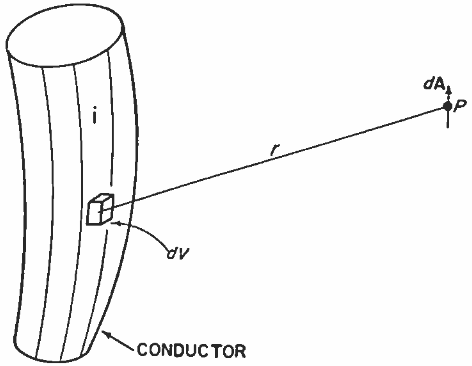


Fig. 10-1

For *alternating* or *time-changing* currents, the current density can be written

$$\mathbf{i} = \mathbf{i}_0 \cos \omega t \tag{10-7}$$

and the expression expected for vector potential would be

$$\mathbf{A} = \frac{1}{4\pi} \int_{\text{vol}} \frac{\mathbf{i}_0 \cos \omega t}{r}\, dV \tag{10-8}$$

In expression (8), r is the distance from the volume element dV to the point P at which the potential is being evaluated. Expression (8) assumes that the vector potential at P will be in phase with the current density $\mathbf{i}$, regardless of how large r may be. In effect, an infinite velocity of propagation has been assumed. Actually, electromagnetic disturbances propagate with a finite velocity v, which in free space is equal to $c \approx 3 \times 10^8$ meter/sec. This finite time of propagation requires that there be a time delay between any change in the current density and the effect of this change at P.

The amount of this time delay will be r/v seconds, and at an angular frequency ω, this corresponds to a phase delay $\omega r/v$ radians. When the finite time of propagation is taken into account the expression for vector potential is

$$\mathbf{A} = \frac{1}{4\pi} \int_{\text{vol}} \frac{i_0 \cos \omega \left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-9}$$

$$= \frac{1}{4\pi} \int_{\text{vol}} \frac{i_0 \cos (\omega t - \beta r)}{r} \, dV \tag{10-10}$$

where

$$\beta = \frac{\omega}{v} = \frac{2\pi}{\lambda} \tag{10-11}$$

In exactly the same way, the generalized expression for scalar potential due to a time-varying charge density distribution $\rho_0 \cos \omega t$ would be

$$V = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho_0 \cos \omega \left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-12}$$

Expressions (9) and (12) indicate that time variations of $\mathbf{A}$ and V at the point P correspond to time variations in current and charge densities that occurred at an earlier time $t' = [t - (r/v)]$. Thus the potential variations are retarded in time, and for this reason the potentials given by (9) and (12) are known as the *retarded potentials.*

Instead of eq. (7), the time variations of current density may be shown more generally by

$$i = i(t)$$

where $i(t)$ can represent any specified function of time. Similarly the time variations of charge density can be shown as $\rho(t)$. The corresponding expressions for the retarded potentials are

$$\mathbf{A} = \frac{1}{4\pi} \int_{\text{vol}} \frac{i\left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-13}$$

$$V = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho\left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-14}$$

where $i[t - (r/v)]$ and $\rho[t - (r/v)]$ are the same functions of $(t - r/v)$ as i and ρ are of t.

Having obtained expressions for the retarded potentials in terms of the currents and charges, the magnetic and electric field intensities can be obtained from **A** and V by taking the appropriate space derivatives. However, the resulting **E** and **H** are not entirely independent, but are in fact related through Maxwell's equations. Therefore, if **H** is obtained through the relation $\mathbf{H} = \text{curl } \mathbf{A}$, it should be possible to find **E** from **H**, through either of Maxwell's equations.

Using the second Maxwell equation

$$\text{curl } \mathbf{E} = -\mu\dot{\mathbf{H}}$$
$$= -\mu \text{ curl } \dot{\mathbf{A}}$$

and transposing,
$$\text{curl } (\mathbf{E} + \mu\dot{\mathbf{A}}) = 0 \tag{10-15}$$

Equation (15) is a differential equation, a particular solution of which is

$$(\mathbf{E} + \mu\dot{\mathbf{A}}) = 0$$

However, a more general solution to (15) is

$$(\mathbf{E} + \mu\dot{\mathbf{A}}) = \pm \text{ grad } V_0 \tag{10-16}$$

where V_0 is any scalar function. That this is a solution becomes evident when vector identity number (1-26) is recalled. Using the minus sign in (16) gives

$$\mathbf{E} = -\mu\dot{\mathbf{A}} - \text{grad } V_0 \tag{10-17}$$

If (17) is a perfectly general expression for **E**, it will give the electric field in the particular case of no time variations, that is, the static field. For no variations with time (17) reduces to

$$\mathbf{E} = -\text{ grad } V_0$$

and it is seen that the scalar function V_0 [which was arbitrary in (16)] is, in fact, just the scalar potential V due to the charge distribution ρ. Therefore, the electric intensity can be obtained from

$$\mathbf{E} = -\mu\dot{\mathbf{A}} - \text{grad } V \tag{10-18}$$

where **A** and V are given by (13) and (14).

Alternatively **E** can be obtained by using the first Maxwell equation instead of the second. In free space, that is outside the conducting region carrying current density i, use of Maxwell's first equation gives

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}}$$

$$\mathbf{E} = \frac{1}{\epsilon}\int \text{curl } \mathbf{H}\, dt$$

$$= \frac{1}{\epsilon}\int \text{curl curl } \mathbf{A}\, dt \tag{10-19}$$

From (18) and (19) it follows that the following relation between **A** and V must hold

$$\frac{1}{\epsilon}\int \text{curl curl } \mathbf{A}\, dt + \mu\dot{\mathbf{A}} = -\text{ grad } V \tag{10-20}$$

Problem 1. Show that equation (20) is satisfied if

$$\text{div } \mathbf{A} = -\epsilon\dot{V}$$

(NOTE: Since **H** obeys the wave equation, its time and space derivatives and integrals will also obey the wave equation. Therefore $\nabla^2\mathbf{A} = \mu\epsilon\ddot{\mathbf{A}}$.)

Equation (20) indicates that the potentials **A** and V are not independent, but are related to each other, and problem 1 shows that a suitable relation between them would be

$$\text{div } \mathbf{A} = -\epsilon\dot{V} \tag{10-21}$$

Actually, it can be shown that (21) is a relation that must always hold in any real problem because it follows directly from the equation of continuity

$$\text{div } i = -\dot{\rho} \tag{10-22}$$

which always holds in a *physical* problem. Equation (22) states that the charge and current densities *cannot be specified independently*, but must always be such as to satisfy the equation of continuity. Therefore, in any real problem it is not necessary to specify *both* charges and currents, because one can be obtained in terms of the other through (22). Correspondingly, the complete electromagnetic field can be obtained by use of the vector potential alone, because **A** and V are always related through (21).

Proof that (21) follows from (13), (14), and (22) can be found in various papers and books,* and will not be carried through here.

Alternative Method of Deriving the Electromagnetic Potentials. In the foregoing portion of this section the electromagnetic potentials were derived as generalizations from the scalar and vector potentials of the static electric and magnetic fields. An alternative approach, and the one usually adopted in most texts on electromagnetic theory, is to use the field equations to set up differential equations that the potentials must obey. The potentials are then written down as solutions of these partial differential equations. Because the vector potential is an important concept in the electromagnetic field, and because it is a new and unfamiliar concept to the engineer at this stage of his training, it is desirable to examine it from several angles in order to gain a more thorough understanding. For these reasons this alternative approach will also be carried through.

In the electrostatic field the electric intensity was related to the scalar potential by

$$\mathbf{E} = -\operatorname{grad} V \tag{10-1}$$

Also $\mathbf{E}$ was related directly to the charge density through Gauss' law

$$\operatorname{div} \mathbf{E} = \frac{1}{\epsilon} \operatorname{div} \mathbf{D} = \frac{\rho}{\epsilon} \tag{10-23}$$

Combining (23) with (1) gives Poisson's equation for the electrostatic field

$$\operatorname{div} \operatorname{grad} V = \nabla^2 V = -\frac{\rho}{\epsilon} \tag{10-24}$$

Equation (24) is a differential equation relating the scalar potential to the charge density. Now it is known that at point P, the potential due to a distribution of charge density ρ is

$$V_p = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho}{r}\, dV \tag{10-25}$$

Therefore (25) must be a solution of (24).

* For example see J. Grosskopf, "On the Application of the two Methods of Solution of Maxwell's Equations in the Calculation of the Electromagnetic Fields of Radiating Conductors," *Hochfrequenz. Tech. und Elektro-akustik*, **49,** 205–211 (1937); also, J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., New York, 1941, p. 429.

In the electromagnetic field it will be assumed that **H** can be obtained from some vector potential **A**, yet to be determined, through the relation

$$\mathbf{H} = \text{curl } \mathbf{A} \tag{10-03}$$

Then using the second Maxwell equation

$$\text{curl } \mathbf{E} = -\mu\dot{\mathbf{H}} = -\mu \text{ curl } \dot{\mathbf{A}}$$

or

$$\text{curl } (\mathbf{E} + \mu\dot{\mathbf{A}}) = 0$$

from which

$$\mathbf{E} = -\mu\dot{\mathbf{A}} - \text{grad } V \tag{10-26}$$

where, as before, V can be shown to be the scalar potential due to charges, and **A** is still to be determined. Writing the first Maxwell equation for a region containing a conduction current density $\boldsymbol{i}$

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}} + \boldsymbol{i} \tag{10-27}$$

Combining (26), (27), and (3) gives

$$\text{curl curl } \mathbf{A} = -\mu\epsilon\ddot{\mathbf{A}} - \epsilon \text{ grad } \dot{V} + \boldsymbol{i} \tag{10-28}$$

Use of vector identity number (1-28) changes this to

$$\text{grad div } \mathbf{A} - \nabla^2\mathbf{A} = -\mu\epsilon\ddot{\mathbf{A}} - \epsilon \text{ grad } \dot{V} + \boldsymbol{i} \tag{10-29}$$

It has been indicated that the relation between **A** and V given by (21) must always be satisfied in any real problem, so (29) reduces to

$$\nabla^2\mathbf{A} - \mu\epsilon\ddot{\mathbf{A}} = -\boldsymbol{i} \tag{10-30}$$

Taking the time derivative of (21) and combining with the divergence of (26) gives

$$\text{div } \mathbf{E} = \mu\epsilon\ddot{V} - \text{div grad } V \tag{10-31}$$

Application of Gauss' law, (23), finally yields

$$\nabla^2 V - \mu\epsilon\ddot{V} = -\frac{\rho}{\epsilon} \tag{10-32}$$

Equations (30) and (32) may be regarded as partial differential equations, the solutions of which will give **A** and V. The form of the solution for (32) may be inferred by considering the equation in two limiting cases. First, for the static case in which there are no variations with time, (32) reduces to Poisson's equation (24), for which (25) is a solution. Second, for the time-varying case, but outside the volume occupied by the charge density ρ, (32) reduces to the free-space wave equation, viz.,

$$\nabla^2 V = \mu\epsilon \ddot{V} \tag{10-33}$$

A particular solution for (33) (suitable, for example, for the case of no variation of the field with θ or ϕ) is of the form

$$V_p = \frac{KF\left(t - \frac{r}{v}\right)}{r} \tag{10-34}$$

where K depends upon the source, and F is any function of $\left(t - \frac{r}{v}\right)$.

Now the potential in free space (outside the region where there are charges) is nevertheless due to the distribution in the region occupied by the charge. If the charge distribution is a function of time, the effect of a change in ρ is not felt immediately in all space, but the disturbance propagates outward as a wave traveling with a finite velocity v. If (25) is the solution of (24), and (34) is a solution of (33), a logical guess for the solution of the general differential equation (32) might be

$$V_p = \frac{1}{4\pi\epsilon} \int_{\text{vol}} \frac{\rho\left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-35}$$

Similarly a possible solution for (30), which is the same equation as (32) from a mathematical point of view, would be

$$\mathbf{A}_p = \frac{1}{4\pi} \int_{\text{vol}} \frac{i\left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-36}$$

That (35) and (36) are indeed solutions of (32) and (30) can be checked by reinserting the former into the latter.*

Problem 2. Show that for a current along the z axis the expression $\mathbf{H} = \text{curl } \mathbf{A}$ reduces to

$$H_\phi = -\sin\theta \frac{\partial A_z}{\partial r}$$

when only the distant field is considered.

(NOTE: This is an important result, for it covers the majority of practical antennas.)

* This process, which is usually quite straightforward, is rather involved in this case, and reference should be made to one of the following texts: H. A. Lorentz, *Theory of Electrons*, pp. 17–19; M. Mason and W. Weaver, *The Electromagnetic Field*, University of Chicago Press, Chicago, 1929 (p. 282).

10.02 The Alternating Current Element (or Oscillating Electric Dipole). An excellent example of the use of the retarded vector potential occurs in the calculation of the electromagnetic field of an alternating current element (or oscillating electric dipole). A *current element Idl* refers to a filamentary current I flowing through an elemental length dl. This is approximated when a current I flows in a very short length of thin wire, if the length dl considered is so short that the current is essentially constant along the length. Although an isolated current element may appear to be a very unreal concept, any physical circuit or antenna carrying current may be considered to consist of a large number of such elements joined end to end. Therefore, if the electromagnetic field of this "building block" is known, the electromagnetic field of any actual antenna having a specified current distribution may be calculated.

Figure 10-2 shows an alternating-current element $Idl \cos \omega t$ located at the origin of a spherical co-ordinate system. The problem is to calculate the electromagnetic field at an arbitrary point P.

The first step is to obtain the vector potential $\mathbf{A}$ at P. The general expression for $\mathbf{A}$ is given by

$$\mathbf{A} = \frac{1}{4\pi} \int_{\text{vol}} \frac{i\left(t - \frac{r}{v}\right)}{r} \, dV \tag{10-36}$$

The integration over the volume in (36) consists of an integration over the cross-sectional area of the wire and an integration along its length. The current density $\boldsymbol{i}$, integrated over the cross-sectional area of the wire, is just the current I, and because this is assumed to be constant along the length, integration over the length gives Idl. Therefore in this simple example the expression for vector potential becomes

$$A_z = \frac{1}{4\pi} \frac{Idl \cos \omega \left(t - \frac{r}{v}\right)}{r} \tag{10-37}$$

The vector potential has the same direction as the current element, in this case the z direction, and is retarded in time by r/v seconds.

The magnetic intensity $\mathbf{H}$ is obtained through the relation

$$\mathbf{H} = \text{curl } \mathbf{A}$$

Reference to the expressions in chap. 1, showing the curl in spherical co-ordinates, gives the components of **H** in terms of A_r, A_θ, and A_ϕ. From Fig. (10-2) it is seen that for this case

$$A_r = A_z \cos\theta; \qquad A_\theta = -A_z \sin\theta; \qquad A_\phi = 0 \quad (10\text{-}38)$$

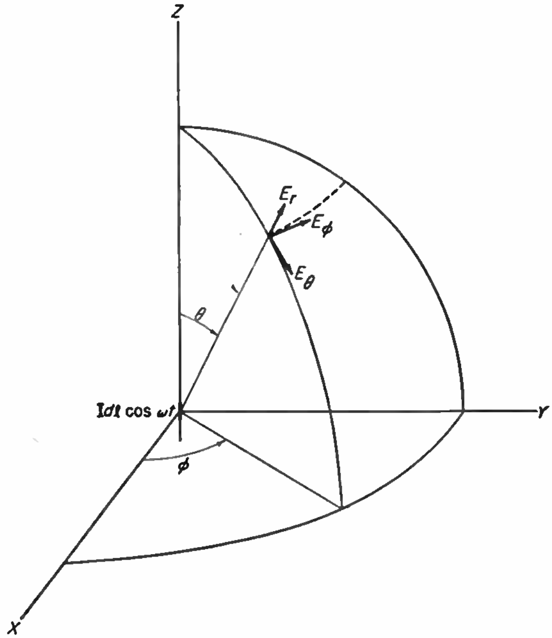


FIG. 10-2. A current element at the center of a spherical co-ordinate system.

Then from expressions (1-39a, b, and c) and noting that because of symmetry, $\partial/\partial\phi = 0$,

$$H_r = \text{curl}_r\, \mathbf{A} = 0$$
$$H_\theta = \text{curl}_\theta\, \mathbf{A} = 0$$

$$H_\phi = \frac{1}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right]$$
$$= \frac{Idl}{4\pi r}\left\{\frac{\partial}{\partial r}\left[-\sin\theta\cos\omega\left(t - \frac{r}{v}\right)\right] - \frac{\partial}{\partial\theta}\left[\frac{\cos\theta}{r}\cos\omega\left(t - \frac{r}{v}\right)\right]\right\}$$
$$= \frac{Idl\sin\theta}{4\pi}\left[\frac{-\omega\sin\omega\left(t - \frac{r}{v}\right)}{rv} + \frac{\cos\omega\left(t - \frac{r}{v}\right)}{r^2}\right] \quad (10\text{-}39)$$

The electric intensity **E** can be obtained from **H** through Maxwell's first equation, which at the point P (in free space) is

$$\text{curl } \mathbf{H} = \epsilon \dot{\mathbf{E}}$$

$$\mathbf{E} = \frac{1}{\epsilon} \int \text{curl } \mathbf{H} \, dt \qquad (10\text{-}40)$$

Taking the curl of (39) and then integrating with respect to time (the order is immaterial) gives for the components of **E**,

$$E_\theta = \frac{Idl \sin \theta}{4\pi\epsilon} \left(\frac{-\omega \sin \omega t'}{rv^2} + \frac{\cos \omega t'}{r^2 v} + \frac{\sin \omega t'}{\omega r^3} \right) \qquad (10\text{-}41)$$

$$E_r = \frac{2Idl \cos \theta}{4\pi\epsilon} \left(\frac{\cos \omega t'}{r^2 v} + \frac{\sin \omega t'}{\omega r^3} \right) \qquad (10\text{-}42)$$

and rewriting (39)

$$H_\phi = \frac{Idl \sin \theta}{4\pi} \left(\frac{-\omega \sin \omega t'}{rv} + \frac{\cos \omega t'}{r^2} \right) \qquad (10\text{-}43)$$

where $t' = \left(t - \frac{r}{v} \right)$.

It is somewhat surprising to find that something so apparently simple as a current element should give rise to an electromagnetic field as complicated as that given by (41), (42), and (43). However a study of these expressions soon shows the significance and necessity for each of the terms.

Consider first the expression for H_ϕ. It is seen to consist of two terms, one of which varies inversely as r and the other inversely as r^2. The second of these, called the *induction* field, will predominate at points close to the current element where r is small, whereas at great distances, where r is large, the second term becomes negligible compared with the first. This first inverse-distance term is called the *radiation* or *distant* field. The two fields will have *equal amplitudes* at that value of r, which makes

$$\frac{1}{r^2} = \frac{\omega}{rv}$$

that is, at

$$r = \frac{v}{\omega} = \frac{\lambda}{2\pi} \approx \frac{\lambda}{6}$$

Except for the fact that t' has replaced t, the induction term

$$\frac{Idl \sin \theta \cos \omega t'}{4\pi r^2}$$

is just the magnetic field intensity that would be given by a direct application of the Biot-Savart law (or Ampère's law for the current element) if this were extended to cover the case of an alternating current $I \cos \omega t$. The fact that the true field is a function of $t' = t - (r/v)$ instead of t, accounts for the finite time of propagation. However, at points close to the element where the induction term predominates, r/v is a very small quantity and $t' \approx t$.

The inverse-distance (radiation) term is an extra term, not present for steady currents. It results from the fact of the finite time of propagation, which is of no account in the steady-field case. It will be shown later that this radiation term contributes to a flow of energy away from the source (the current element), whereas the induction term contributes to energy that is stored in the field during one quarter of a cycle and returned to the circuit during the next.

Examination of the expressions for E_θ and E_r shows that E_θ has an induction $(1/r^2)$ term and a radiation $(1/r)$ term, and E_r has a $1/r^2$ term. In addition both components of **E** have a term that varies as $1/r^3$. From their similarity with the components of the field of an electrostatic dipole (see chap. 2, example 3), these $1/r^3$ terms are called the *electrostatic* field terms (or sometimes just *electric* field terms).

Relation between a Current Element and an Electric Dipole. It is not just a coincidence that the expressions for the electric field of the alternating current element should contain terms that correspond to the field of an oscillating electric dipole. Although only *current* was specified in setting up the hypothetical current element, the equation of continuity (or conservation of charge) requires that there be an accumulation of charge at the ends of the element, which is given by

$$\frac{\partial q}{\partial t} = I \cos \omega t$$

That is, the charge at one end is increasing and at the other end decreasing, by the amount of the current flow (coulombs per second). In order to obtain a physical approximation of an isolated current element, one could terminate the current element in two small spheres or disks on which the charges could accumulate. If the wire is very thin compared with the radius of the spheres (so that its distributed capacitance is negligible compared with the capacitance between the spheres), the current in the wire will be uniform. In addition, the radii of the spheres should be small compared with dl, their distance apart, and in turn dl should be very much shorter

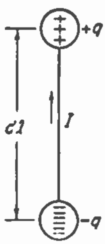


FIG. 10-3. Hertzian dipole.

FIG. 10-4. Chain of Hertzian dipoles.

than a wavelength. The arrangement then is essentially that of the original Hertzian oscillating electric dipole (Fig. 10-3).

$$\frac{\partial q}{\partial t} = I \cos \omega t$$

$$q = \frac{I \sin \omega t}{\omega}$$

From comparison with the electrostatic dipole, the electric intensity that would be expected to result from the separated charges at the ends of the current element would be

$$E_\theta = \frac{q\,dl \sin \theta}{4\pi\epsilon r^3} = \frac{I dl \sin \theta \sin \omega t'}{4\pi\epsilon\omega r^3} \tag{10-44}$$

$$E_r = \frac{2q\,dl \cos \theta}{4\pi\epsilon r^3} = \frac{2I dl \cos \theta \sin \omega t'}{4\pi\epsilon\omega r^3} \tag{10-45}$$

These are exactly the $1/r^3$ terms that automatically appeared in the solution for the electromagnetic field of the current element.

When a current element forms part of a complete circuit there is no accumulation of charge at its ends if the current is uniform throughout the circuit, for the current from one element flows into the next. In this case, as would be expected, the $1/r^3$ terms due to accumulated charge vanish, leaving only induction and radiation fields. In terms of a chain of Hertzian dipoles (Fig. 10-4), the positive charge at the end of one dipole is just cancelled by an equal amount of negative charge at the opposite end of the adjacent dipole. However, if the current along the circuit or antenna is not uniform along its length, but is distributed as, for example, in Fig. 10-5, this could be represented as a chain of current elements, or Hertzian dipoles, having slightly different amplitudes. In this case the adjacent charges do not completely cancel, and there is an accumulation of charge on the surface of the wire, as indicated in Fig. 10-5. These surface charges are responsible for a relatively strong component of electric intensity normal to the surface of the wire.

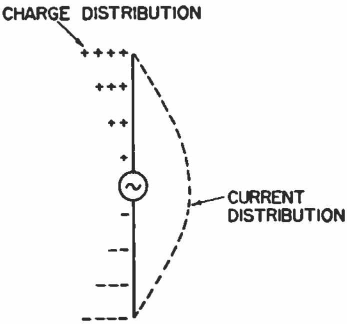


FIG. 10-5. Current and charge distribution on a linear antenna.

Problem 3. Obtain expressions (41) and (42) for E_θ and E_r due to a current element through the alternative relation

$$\mathbf{E} = -\mu\dot{\mathbf{A}} - \text{grad } V$$

(NOTE: Obtain V from $\mathbf{A}$ through div $\mathbf{A} = -\epsilon\dot{V}$; alternatively write V directly from the charges that can be obtained from I through the equation of continuity.)

Problem 4. Starting with the expression $Idl\, e^{j\omega t}$ for a current element, show that the expressions for vector potential and field intensities will be

$$A_z = \frac{Idl}{4\pi r} e^{-j\beta r}$$

$$H_\phi = \frac{Idl \sin\theta\, e^{-j\beta r}}{4\pi r}\left(j\beta + \frac{1}{r}\right)$$

$$E_\theta = \frac{\eta Idl \sin\theta\, e^{-j\beta r}}{4\pi r}\left(j\beta + \frac{1}{r} + \frac{1}{j\beta r^2}\right)$$

$$E_r = \frac{\eta Idl \cos\theta\, e^{-j\beta r}}{4\pi r}\left(\frac{2}{r} + \frac{2}{j\beta r^2}\right)$$

where $\beta = \omega/v$, $\eta = \sqrt{\mu/\epsilon} = \mu v$ and the time factor $e^{j\omega t}$ is understood.

10.03 Power Radiated by a Current Element. The power flow per unit area at the point P will be given by the Poynting vector at that point. The instantaneous Poynting vector is given by $\mathbf{E}_{\text{inst}} \times \mathbf{H}_{\text{inst}}$ and it will have both θ and r components. Replacing v by $c \approx 3 \times 10^8$ for free-space propagation, the θ component of instantaneous Poynting vector will be

$$
\begin{aligned}
P_\theta &= -E_r H_\phi \\
&= \frac{I^2 dl^2 \sin 2\theta}{16\pi^2\epsilon}\left(\frac{\sin^2 \omega t'}{r^4 c} - \frac{\cos^2 \omega t'}{r^4 c} - \frac{\sin \omega t' \cos \omega t'}{\omega r^5} + \frac{\omega \sin \omega t' \cos \omega t'}{r^3 c^2}\right) \\
&= \frac{I^2 dl^2 \sin 2\theta}{16\pi^2\epsilon}\left(\frac{-\cos 2\omega t'}{r^4 c} - \frac{\sin 2\omega t'}{2\omega r^5} + \frac{\omega \sin 2\omega t'}{2r^3 c^2}\right) \qquad (10\text{-}46)
\end{aligned}
$$

The average value of $\sin 2\omega t'$ or $\cos 2\omega t'$ over a complete cycle is zero. Therefore, for any value of r, the average of P_θ over a complete cycle is zero. P_θ represents only a surging back and forth of power in the θ direction without any net or average flow. The radial Poynting vector is given by

$$
\begin{aligned}
P_r &= E_\theta H_\phi \\
&= \frac{I^2 dl^2 \sin^2 \theta}{16\pi^2\epsilon}\left(\frac{\sin \omega t' \cos \omega t'}{\omega r^5} + \frac{\cos^2 \omega t'}{r^4 c} - \frac{\omega \sin \omega t' \cos \omega t'}{r^3 c^2}\right. \\
&\qquad \left. - \frac{\sin^2 \omega t'}{r^4 c} - \frac{\omega \sin \omega t' \cos \omega t'}{r^3 c^2} + \frac{\omega^2 \sin^2 \omega t'}{r^2 c^3}\right) \\
&= \frac{I^2 dl^2 \sin^2 \theta}{16\pi^2\epsilon}\left(\frac{\sin 2\omega t'}{2\omega r^5} + \frac{\cos 2\omega t'}{r^4 c} - \frac{\omega \sin 2\omega t'}{r^3 c^2} + \frac{\omega^2(1 - \cos 2\omega t')}{2r^2 c^3}\right) \\
&\qquad (10\text{-}47)
\end{aligned}
$$

The *average* value of radial Poynting vector over a cycle will be due to part of the final term only and is

$$
\begin{aligned}
P_{r(\text{av})} &= \frac{\omega^2 I^2 dl^2 \sin^2 \theta}{32\pi^2 r^2 c^3 \epsilon} \\
&= \frac{\eta}{2}\left(\frac{\omega I dl \sin \theta}{4\pi r c}\right)^2 \qquad \text{watts/sq m} \quad (10\text{-}48)
\end{aligned}
$$

None of the terms in the expressions for Poynting vector represent an *average* power flow except that of eq. (48). The only terms of **E** and **H** that contribute to this average power flow are the radiation or inverse-distance terms. At a large distance from the source these radiation terms are the only ones that have appreciable value, but even close to the current element, where the induction and electric fields predominate, only the $1/r$ terms contribute to an average outward flow of power.

That it is only the inverse-distance terms that can contribute to an outward flow of power from the source can be proven by simple reasoning. If one considers two concentric shells or radii r_1 and r_2 enclosing the source, then the average outward rate of energy flow through shell r_2 must be the same as through shell r_1, if there is to be no continuous accumulation (or decrease) of energy stored in the region between them. This requires that the power density decrease as $1/r^2$ since the area of the shells increases as r^2. The power density is proportional to E times H (or to E^2 or H^2), so E and H, which contribute to an average radial power flow, must be proportional to $1/r$. Components of E or H, which are inversely proportional to r^2 or r^3, can contribute to an instantaneous flow of energy into the region between the shells, but this energy must later be returned to the source because it cannot be stored permanently in a finite volume of the medium.

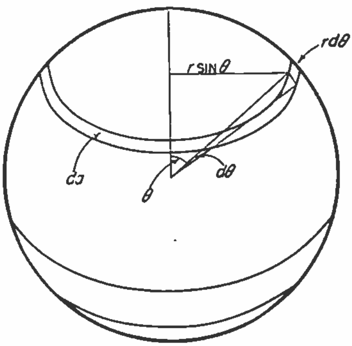


FIG. 10-6. Element of area on a spherical surface.

From eqs. (41), (42), and (43) the amplitudes of the *radiation* fields of an electric current element Idl are

$$E_\theta = \frac{\omega Idl \sin\theta}{4\pi\epsilon v^2 r} = \frac{\eta Idl \sin\theta}{2\lambda r} = \frac{60\pi Idl \sin\theta}{r\lambda} \tag{10-49}$$

$$H_\phi = \frac{\omega Idl \sin\theta}{4\pi v r} = \frac{Idl \sin\theta}{2\lambda r} \tag{10-50}$$

The radiation terms of E_θ and H_ϕ are in time phase and are related by

$$\frac{E_\theta}{H_\phi} = \eta \tag{10-51}$$

The total power radiated by the current element can be computed by integrating the radial Poynting vector over a spherical surface centered at the element. P is independent of the azimuthal angle ϕ, so the element of area on the spherical shell will be taken as the strip da where

$$da = 2\pi r^2 \sin\theta\, d\theta$$

Then the total power radiated is

$$\begin{aligned}
Power &= \oint_{\text{surface}} P_{r(\text{av})}\, da = \int_0^\pi \frac{\eta}{2}\left(\frac{\omega I dl \sin\theta}{4\pi r c}\right)^2 2\pi r^2 \sin\theta\, d\theta \\
&= \frac{\eta\omega^2 I^2 dl^2}{16\pi c^2}\int_0^\pi \sin^3\theta\, d\theta \\
&= \frac{\eta\omega^2 I^2 dl^2}{16\pi c^2}\left[\frac{-\cos\theta}{3}(\sin^2\theta + 2)\right]_0^\pi \\
&= \frac{\eta\omega^2 I^2 dl^2}{12\pi c^2} \qquad \text{watts}
\end{aligned} \tag{10-52}$$

In this expression I is maximum or peak current. In terms of effective current the power radiated is

$$\begin{aligned}
\text{Power} &= \frac{\eta\omega^2 I^2_{\text{eff}} dl^2}{6\pi c^2} \\
&= 80\pi^2 \left(\frac{dl}{\lambda}\right)^2 I_{\text{eff}}^2
\end{aligned}$$

The coefficient of I_{eff}^2 has the dimensions of resistance and is called the *radiation resistance* of the current element. Then, for a current element,

$$R_{\text{rad}} = 80\pi^2\left(\frac{dl}{\lambda}\right)^2 \tag{10-53}$$

10.04 Application to Short Antennas. The hypothetical current element is a useful tool for theoretical work, but it is not a practical antenna. The *practical* "elementary dipole" is a center-fed antenna having a length that is very short in wavelengths. The current amplitude on such an antenna decreases uniformly from a maximum at the center to zero at the ends (Fig. 10-7a). For the same current I (at the terminals) the (short) practical dipole of length l will radiate only one-quarter as much power as the current element of the same length, which has the current I throughout its entire length. (The field intensities at every point are reduced to one-half, and the power density will be reduced to one-quarter.) Therefore, the radiation resistance of a practical short dipole is one-quarter that of the current element of the same length. That is

$$R_{rad}\text{ (short dipole)} = 20\pi^2\left(\frac{l}{\lambda}\right)^2$$

$$\approx 200\left(\frac{l}{\lambda}\right)^2 \quad \text{ohms} \quad (10\text{-}53)$$

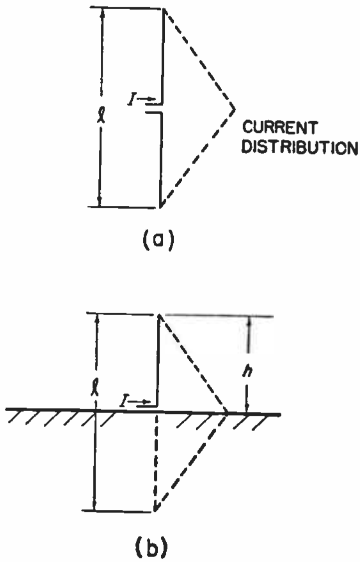


FIG. 10-7. Current distribution on short antennas: (a) short dipole; (b) short monopole.

The monopole of height h (Fig. 10-7b), or short vertical antenna mounted on a reflecting plane, produces the same field intensities above the plane as does the dipole of length $l = 2h$ when both are fed with the same current. However, the short vertical antenna radiates only through the semispherical surface above the plane, so its radiated power is only one-half that of the corresponding dipole. Therefore, the radiation resistance of the monopole of height $h = l/2$ is

$$R_{rad}\text{ (monopole)} = 10\pi^2\left(\frac{l}{\lambda}\right)^2$$

$$= 40\pi^2\left(\frac{h}{\lambda}\right)^2$$

$$\approx 400\left(\frac{h}{\lambda}\right)^2 \quad \text{ohms} \quad (10\text{-}54)$$

These formulas hold strictly for very short antennas* only, but they are good approximations for dipoles of lengths up to quarter wavelength, and monopoles of heights up to one-eighth wavelength.

* Lower-case symbols, l and h, have been used in equation (54) for these *short* antennas. For general-length antennas, discussed in the next section, the symbols L and H are used.

10.05 Assumed Current Distribution. In order to calculate the electromagnetic fields of longer antennas, it is necessary to know the current distribution along the antennas. This information should be obtainable by solving Maxwell's equations subject to the appropriate boundary conditions along the antenna. However, for the cylindrical antenna, this is a comparatively difficult problem, and it is only in quite recent years that satisfactory solutions have been obtained. One of these will be considered in chap. 13. In the absence of a known antenna current, it is possible to *assume* a certain distribution and from that to calculate approximate field distributions. The accuracy of the fields so calculated will, of course, depend upon how good an assumption was made for current distribution. By thinking of the center-fed antenna as an open-circuited transmission line that has been opened out, a sinusoidal current distribution with current nodes at the ends is suggested. The fact that it is known from Abraham's work on thin ellipsoids that the current will be truly sinusoidal for the infinitely thin case, makes this assumption more justifiable. It turns out to be a very good assumption for thin antennas, sufficiently good in fact, that even with more accurate (and much more complicated) formulas available, the sinusoidal distribution is still used for much of the work in the antenna field. When greater accuracy is desired, and in those particular cases where the sinusoidal assumption breaks down entirely, it is necessary to use a distribution that is closer to the true one.

Figure 10-8a shows a center-fed dipole with a sinusoidal current distribution, and Fig. 10-8b shows the corresponding monopole. "A *dipole antenna** is a straight radiator, usually fed in the center, and producing a maximum of radiation in the plane normal to the axis. The length specified is the over-all length." The vertical antenna (of height $H = L/2$) fed against an infinitely-large perfectly-conducting plane has the same radiation characteristics above the plane as does the dipole antenna of length L in free space. This is because the fields due to a current element $I\,dz$ when reflected from the plane, appear to originate at an image element located beneath the plane. Moreover, the impedance of the vertical antenna fed against the reflecting plane is just one-half that of the corresponding dipole of length $L = 2H$. Thus the dipole of Fig. 10-8a and the corresponding base-fed vertical antenna of Fig. 10-8b

* *IRE Standards on Antennas*, 1948.

can be solved conveniently as one problem. The base-fed vertical antenna of Fig. 10-8b will be referred to simply as a *monopole* of height H, the infinitely large perfectly reflecting plane being understood, unless otherwise stated. The corresponding center-fed dipole will be referred to as a dipole of *length* $L = 2H$ or as a dipole of *half-length* H.

The image principle used here is discussed further in chap. 12.

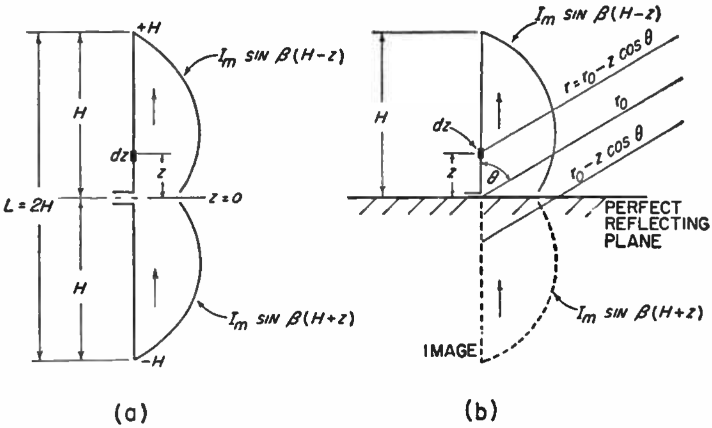


FIG. 10-8. (a) Center-fed dipole with assumed sinusoidal current distribution, (b) corresponding monopole.

10.06 Radiation from a Quarter-wave Monopole or Half-wave Dipole. It will be assumed that the current is sinusoidally distributed as shown in Fig. 10-8 and that time variations will be indicated by $e^{j\omega t}$. Then

$$I = I_m \sin \beta(H - z)\, e^{j\omega t} \qquad z > 0$$
$$I = I_m \sin \beta(H + z)\, e^{j\omega t} \qquad z < 0$$

where I_m is the value of current at the current *loop* or current *maximum*. Then the expression for the vector potential at a point P due to the current element $I\,dz$ will be

$$dA_z = \frac{I\,dz\,e^{j(\omega t - \beta r)}}{4\pi r} = \frac{I\,e^{-j\beta r}\,dz}{4\pi r}$$

where r is the distance from the current element to the point P, and where the time factor has been dropped in the second expression. The total vector potential at P due to all the current elements will be

$$A_z = \frac{1}{4\pi}\int_{-H}^{0} \frac{I_m \sin\beta(H+z)\, e^{-j\beta r}\, dz}{r} + \frac{1}{4\pi}\int_{0}^{H} \frac{I_m \sin\beta(H-z)\, e^{-j\beta r}\, dz}{r} \qquad (10\text{-}55)$$

Because only the distant or radiation fields are required in this problem, it is possible to make some simplifying approximations. For the inverse-distance factor (the r in the denominator) it is valid to write

$$r \approx r_0$$

However for the r in the phase factor in the numerator, it is the *difference* between r and r_0 that is important. For very large values of r the lines to the point P are essentially parallel ($\theta \approx \theta_0$) and for the r in the phase factor one can write approximately

$$r = r_0 - z\cos\theta$$

Then the expression for A_z becomes

$$A_z = \frac{I_m e^{-j\beta r_0}}{4\pi r_0}\left[\int_{-H}^{0} \sin\beta(H+z)e^{j\beta z\cos\theta}\, dz + \int_{0}^{H} \sin\beta(H-z)e^{j\beta z\cos\theta}\, dz\right] \qquad (10\text{-}56)$$

For the particular case of $H = \lambda/4$,

$$\sin\beta(H+z) = \sin\beta(H-z) = \cos\beta z$$

and the integral becomes

$$\begin{aligned}
A_z &= \frac{I_m e^{-j\beta r_0}}{4\pi r_0}\int_0^H \cos\beta z\,(e^{j\beta z\cos\theta} + e^{-j\beta z\cos\theta})\, dz \\
&= \frac{I_m e^{-j\beta r_0}}{4\pi r_0}\int_0^H [\cos\{\beta z(1+\cos\theta)\} + \cos\{\beta z(1-\cos\theta)\}]\, dz \\
&= \frac{I_m e^{-j\beta r_0}}{4\pi r_0}\left\{\frac{\sin[\beta z(1+\cos\theta)]}{\beta(1+\cos\theta)} + \frac{\sin[\beta z(1-\cos\theta)]}{\beta(1-\cos\theta)}\right\}_0^{\lambda/4} \\
&= \frac{I_m e^{-j\beta r_0}}{4\pi\beta r_0}\left[\frac{(1-\cos\theta)\cos\left(\frac{\pi}{2}\cos\theta\right) + (1+\cos\theta)\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta}\right] \\
&= \frac{I_m e^{-j\beta r_0}}{2\pi\beta r_0}\left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta}\right] \qquad (10\text{-}57)
\end{aligned}$$

Recalling from problem (2) that when the current is entirely in the z direction,

$$H_\phi = \frac{-\partial A_z}{\partial r} \sin\theta$$

The expression for magnetic intensity at a distant point will be

$$H_\phi = \frac{jI_m\, e^{-j\beta r_0}}{2\pi r_0} \left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \right] \tag{10-58}$$

where only the inverse-distance term has been retained. The electric intensity for the radiation field will be

$$\begin{aligned} E_\theta &= \eta H_\phi \\ &= \frac{j60I_m\, e^{-j\beta r_0}}{r_0} \left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \right] \end{aligned} \tag{10-59}$$

The magnitude of the electric intensity for the radiation field of a half-wave dipole or quarter-wave monopole is

$$E_\theta = \frac{60I_m}{r_0} \left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \right] \quad \text{volt/m} \tag{10-60}$$

E_θ and H_ϕ are in time phase so the maximum value in time of the Poynting vector is just the product of the peak values of E_θ and H_ϕ, and the average value in time of the Poynting vector will be one-half the peak value. Then

$$P_{av} = \frac{\eta I_m^2}{8\pi^2 r_0^2} \left[\frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta} \right]$$

The total power radiated through a semispherical surface of radius r_0 (Fig. 10-6) will equal

$$\oint P_{av}\, da = \frac{\eta I_m^2}{4\pi} \int_0^{+\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\, d\theta \tag{10-61}$$

It is necessary to evaluate this integral. Most of the difficulty in radiation problems is usually in connection with the evaluation of an integral. The following substitutions are typical.

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\,d\theta = \frac{1}{2}\int_0^{\pi/2} \frac{1+\cos(\pi\cos\theta)}{\sin\theta}\,d\theta$$

Let

$$u = \cos\theta$$
$$du = -\sin\theta\,d\theta$$
$$\frac{d\theta}{\sin\theta} = \frac{-du}{\sin^2\theta} = -\frac{du}{1-u^2}$$

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\,d\theta = -\frac{1}{2}\int_1^0 \frac{(1+\cos\pi u)}{1-u^2}\,du$$
$$= \frac{1}{4}\int_0^1 (1+\cos\pi u)\left(\frac{1}{1+u}+\frac{1}{1-u}\right)du$$
$$= \frac{1}{4}\int_{-1}^{+1} \frac{1+\cos\pi u}{1+u}\,du$$

Let

$$v = \pi(1+u)$$
$$dv = \pi\,du$$
$$\frac{dv}{v} = \frac{du}{1+u}$$
$$\pi u = v - \pi$$
$$\cos\pi u = \cos v\cos\pi + \sin v\sin\pi = -\cos v$$

Therefore

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\,d\theta = \frac{1}{4}\int_0^{2\pi} \frac{1-\cos v}{v}\,dv$$
$$= \frac{1}{4}\int_0^{2\pi}\left(\frac{v^1}{2!}-\frac{v^3}{4!}+\frac{v^5}{6!}-\frac{v^7}{8!}+\cdots\right)dv$$
$$= \frac{1}{4}\left(\frac{v^2}{2\cdot 2!}-\frac{v^4}{4\cdot 4!}+\frac{v^6}{6\cdot 6!}-\frac{v^8}{8\cdot 8!}+\cdots\right)_0^{2\pi} \quad (10\text{-}62)$$

The series of eq. (10-62) can be evaluated by substitution. It does not converge rapidly and so a number of terms must be used. This evaluation is shown below.

$$v = 2\pi = 6.2832 \qquad \log_{10} v = 0.79818$$
$$2 \cdot 2! = 4 \qquad \log_{10} v^2 = 1.59636$$
$$\log_{10} 2 \cdot 2! = 0.60206$$
$$\log_{10} \frac{v^2}{2 \cdot 2!} = .99430 \qquad \frac{v^2}{2 \cdot 2!} = 9.870$$

The other terms are found in a similar manner. Using eight terms, the sum of the positive terms is 26.878 and the sum of the negative terms is 24.441. Therefore

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\, d\theta = 0.6093 \tag{10-63}$$

It is also possible to integrate such a function as

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\, d\theta$$

graphically or by Simpson's or the trapezoidal rule. For example by the trapezoidal rule if θ is taken in increments of 5^0, then the following table is constructed:

θ in degrees	0	5	10	15	20	25	30	35	40	45
$\frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$	0	0	.003	.011	.028	.050	.086	.138	.201	.280

θ in degrees	50	55	60	65	70	75	80	85	90
$\frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$	.369	.468	.578	.683	.788	.875	.942	.980	1.00

Now

$$\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\,d\theta = \frac{\pi}{2} \times \frac{1}{18}\left[\frac{1.000 + 0}{2} + \sum_{\theta=5°}^{\theta=85°} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\right]$$

$$= \frac{\pi}{36} \times 6.987$$

$$= 0.609$$

This method of numerical integration shows that in a given antenna problem, if the current distribution is known, the radiation resistance may always be found by straightforward methods, although the integration may be tedious. The power radiated through the semispherical surface is obtained by inserting the value of the integral in (61). Then

$$\text{Radiated power} = \frac{0.609\eta I_m^2}{4\pi}$$

In this expression I_m is peak current. In terms of effective current the radiated power would be

$$\text{Radiated power} = \frac{0.609\eta I^2_{m(\text{eff})}}{2\pi}$$

$$= 36.5 I^2_{m(\text{eff})} \tag{10-64}$$

Therefore the radiation resistance of a quarter-wave monopole antenna is 36.5 ohms.

For the half-wave dipole antenna in free space, power would be radiated through a complete spherical surface. Therefore, for the same current the power radiated would be twice as much, and the radiation resistance for the half-wave dipole is

$$R_{\text{rad}} = 73 \text{ ohms}$$

Problem 5. Derive the expression for the radiation term of the electric field of a half-wave dipole [eq. (10-60)] without the use of the vector potential; that is, by adding directly the (distant) fields owing to the current elements.

Problem 6. Derive the general expression corresponding to eq. (59) for the (distant) electric field of a dipole antenna of any half-length H. It is

$$E_\theta = \frac{j60 I_m e^{-j\beta r_0}}{r_0}\left[\frac{\cos(\beta H\cos\theta) - \cos\beta H}{\sin\theta}\right]$$

Problem 7. Write the expression for the power radiated through a spherical surface by the dipole of half-length H.

10.07 Sine Integral and Cosine Integral. The series of eq. (62) that resulted from the integral

$$\int_0^x \frac{1 - \cos v}{v}\, dv.$$

has been evaluated and can be found in tables. It is designated as $S_1(x)$. This series also occurs in the integral

$$\text{Ci}\ (x) = -\int_x^\infty \frac{\cos v}{v}\, dv \tag{10-65}$$

This latter integral is called the *cosine integral* of x, and is abbreviated as shown. A companion integral defined by

$$\text{Si}\ (x) = \int_0^x \frac{\sin v}{v}\, dv \tag{10-66}$$

is known as the *sine integral of* x.

The cosine integral of x is related to $S_1(x)$ by

$$\text{Ci}\ (x) = \ln x + C - S_1(x)$$

where $C = 0.5772157$ is Euler's constant, and

$$S_1(x) = \int_0^x \frac{1 - \cos v}{v}\, dv = \left(\frac{x^2}{2 \cdot 2!} - \frac{x^4}{4 \cdot 4!} + \frac{x^6}{6 \cdot 6!} - + \cdots\right) \tag{10-67}$$

These integrals occur frequently in radiation problems. They have been studied extensively, and tables giving their values may be found in several books.*

Problem 8. Integrate the expression of problem 7 and show that the general expression for the radiation resistance of a dipole of half-length H is

$$R_{\text{rad}} = 30\{S_1(b) - [S_1(2b) - S_1\ (b)] \cos b + [S_1(2b) - \text{Si}\ (b)] \sin b + [1 + \cos b]S_1(b) - \sin b\ \text{Si}\ (b)\}$$

where $b = 2\beta H$

10.08 Electromagnetic Field Close to an Antenna. In section (10.06) and problem 6 expressions for the radiation or distant fields

* E. Jahnke and F. Emde, *Tables of Functions*, B. G. Teubner, Leipzig, 1933; F. E. Terman, *Radio Engineers Handbook*, McGraw-Hill, New York, 1943.

of an antenna were derived. For some purposes, for example to determine the mutual impedance between antennas, it is necessary to know the electric and magnetic intensities in *the neighborhood* of the antenna. In this region the field is often called the *near* field, in contrast to the distant or radiation field. Because the near field will include induction and electric as well as radiation fields, it can be expected to be more complex than the distant field. The answer of most interest will be the component of electric intensity parallel to the antenna, that is E_z. For this reason it is convenient to use a *cylindrical* co-ordinate system, or actually, a combination of cylindrical, spherical, and rectangular co-ordinate systems. This is shown in Fig. 10-9, where the antenna is assumed to extend along the z axis. In this figure the following relations will hold:

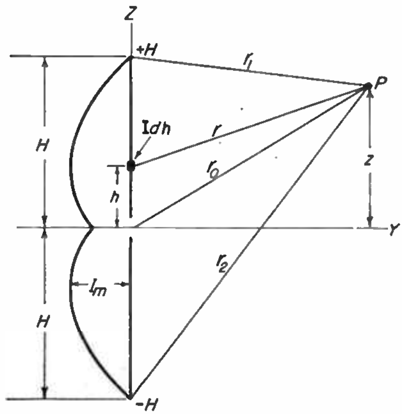


FIG. 10-9. Geometry for fields near the antenna.

$$r = \sqrt{(z - h)^2 + y^2}$$
$$r_1 = \sqrt{(z - H)^2 + y^2}$$
$$r_2 = \sqrt{(z + H)^2 + y^2}$$
$$r_0 = \sqrt{z^2 + y^2}$$

The co-ordinates of the point P are (ρ, ϕ, z) in cylindrical co-ordinates (or r_0, θ, ϕ in spherical co-ordinates). However, because of symmetry, there are no variations in the ϕ direction and so, without loss of generality, the point P may be located in the y-z plane ($\phi = 90°$ plane). In this case $\rho = y$, and the distances from various points along the antenna will be as indicated in the figure.

Again assuming a sinusoidal distribution of current, the antenna current will be

$$I = I_m \sin \beta(H - h) \qquad h > 0$$
$$I = I_m \sin \beta(H + h) \qquad h < 0$$

where again the exponential time factor $e^{j\omega t}$ has been dropped. The expression for vector potential at the point P will be

$$A_z = \frac{I_m}{4\pi}\left[\int_0^H \frac{\sin\beta(H-h)e^{-j\beta r}}{r}\,dh + \int_{-H}^0 \frac{\sin\beta(H+h)e^{-j\beta r}}{r}\,dh\right]$$

Replacing $\sin\beta(H-h)$ with

$$\frac{e^{j\beta(H-h)} - e^{-j\beta(H-h)}}{2j}$$

and using a corresponding expression for $\sin\beta(H+h)$, there results

$$A_z = \frac{I_m}{8\pi j}\left[e^{j\beta H}\int_0^H \frac{e^{-j\beta(r+h)}}{r}\,dh - e^{-j\beta H}\int_0^H \frac{e^{-j\beta(r-h)}}{r}\,dh + e^{j\beta H}\int_{-H}^0 \frac{e^{-j\beta(r-h)}}{r}\,dh - e^{-j\beta H}\int_{-H}^0 \frac{e^{-j\beta(r+h)}}{r}\,dh\right] \quad (10\text{-}68)$$

In cylindrical co-ordinates the magnetic intensity at the point P will be given by

$$H_\phi = \text{curl}_\phi\,\mathbf{A} = -\frac{\partial A_z}{\partial\rho}$$

With the point P in the y-z plane, this can be written as

$$H_\phi = -H_x = -\frac{\partial A_z}{\partial y}$$

$$H_\phi = -\frac{I_m}{8\pi j}\left[e^{j\beta H}\int_0^H \frac{\partial}{\partial y}\left(\frac{e^{-j\beta(r+h)}}{r}\right)dh - e^{-j\beta H}\int^H \frac{\partial}{\partial y}\left(\frac{e^{-j\beta(r-h)}}{r}\right)dh + e^{j\beta H}\int_{-H}^0 \frac{\partial}{\partial y}\left(\frac{e^{-j\beta(r-h)}}{r}\right)dh - e^{-j\beta H}\int_{-H}^0 \frac{\partial}{\partial y}\left(\frac{e^{-j\beta(r+h)}}{r}\right)dh\right] \quad (10\text{-}69)$$

Consider the first term only,

$$e^{j\beta H}\int_0^H \frac{\partial}{\partial y}\left(\frac{e^{-j\beta(r+h)}}{r}\right)dh = e^{j\beta H}\int_0^H \left[\frac{-j\beta y\, e^{-j\beta(r+h)}}{r^2} - \frac{y\, e^{-j\beta(r+h)}}{r^3}\right]dh \quad (10\text{-}70)$$

The integrand turns out to be a perfect differential. Integrating gives

$$e^{j\beta H}\left[\frac{y\,e^{-j\beta(r+h)}}{r(r+h-z)}\right]_{h=0}^{h=H} = y\,e^{j\beta H}\left[\frac{e^{-j\beta(r_1+H)}}{r_1(r_1+H-z)} - \frac{e^{-j\beta r_0}}{r_0(r_0-z)}\right]$$

$$= y\,e^{j\beta H}\left[\frac{(r_1-H+z)e^{-j\beta(r_1+H)}}{r_1[r_1^2-(H-z)^2]} - \frac{(r_0+z)e^{-j\beta r_0}}{r_0(r_0^2-z^2)}\right]$$

But $\qquad r_1^2 - (H-z)^2 = r_0^2 - z^2 = y^2$

so that the first term becomes

$$\frac{e^{j\beta H}}{y}\left[\left(1-\frac{H-z}{r_1}\right)e^{-j\beta(r_1+H)}-\left(1+\frac{z}{r_0}\right)e^{-j\beta r_0}\right]$$

Similarly the second, third, and fourth terms of equation (10-69) are

$$\frac{e^{-j\beta H}}{y}\left[\left(1+\frac{H-z}{r_1}\right)e^{-j\beta(r_1-H)}-\left(1-\frac{z}{r_0}\right)e^{-j\beta r_0}\right]$$

$$\frac{e^{j\beta H}}{y}\left[\left(1-\frac{H+z}{r_2}\right)e^{-j\beta(r_2+H)}-\left(1-\frac{z}{r_0}\right)e^{-j\beta r_0}\right]$$

$$\frac{e^{-j\beta H}}{y}\left[\left(1+\frac{H+z}{r_2}\right)e^{-j\beta(r_2-H)}-\left(1+\frac{z}{r_0}\right)e^{-j\beta r_0}\right]$$

Adding these four terms, the magnetic field strength can be obtained. It is

$$H_\phi = -\frac{I_m}{4\pi j}\left(\frac{e^{-j\beta r_1}}{y}+\frac{e^{-j\beta r_2}}{y}-\frac{2\cos\beta H\, e^{-j\beta r_0}}{y}\right) \qquad (10\text{-}71)$$

The electric field can be obtained from the magnetic field by recalling that in free space

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}}$$

so that

$$\mathbf{E} = \frac{1}{j\omega\epsilon}\text{curl } \mathbf{H}$$

In the $x = 0$ plane

$$E_z = \frac{1}{j\omega\epsilon}(\text{curl } H_\phi)_z = \frac{1}{j\omega\epsilon y}\frac{\partial}{\partial y}(yH_\phi)$$

$$E_y = \frac{1}{j\omega\epsilon}(\text{curl } H_\phi)_y = -\frac{1}{j\omega\epsilon}\frac{\partial}{\partial z}(H_\phi)$$

Substituting the expression for H_ϕ in these equations gives

$$E_z = \frac{-j\beta I_m}{4\pi\omega\epsilon y}\left(\frac{y\,e^{-j\beta r_1}}{r_1}+\frac{y\,e^{-j\beta r_2}}{r_2}-2\cos\beta H\,\frac{y\,e^{-j\beta r_0}}{r_0}\right)$$

which reduces to

$$E_z = -j30I_m\left(\frac{e^{-j\beta r_1}}{r_1}+\frac{e^{-j\beta r_2}}{r_2}-2\cos\beta H\,\frac{e^{-j\beta r_0}}{r_0}\right) \qquad (10\text{-}72)$$

Similarly,

$$E_y = j30I_m$$

$$\left(\frac{z-H}{y}\cdot\frac{e^{-j\beta r_1}}{r_1}+\frac{z+H}{y}\cdot\frac{e^{-j\beta r_2}}{r_2}-\frac{2z\cos\beta H}{y}\,\frac{e^{-j\beta r_0}}{r_0}\right) \qquad (10\text{-}73)$$

and rewriting the expression for magnetic intensity

$$H_\phi = \frac{j30I_m}{\eta y}\left(e^{-j\beta r_1} + e^{-j\beta r_2} - 2\cos\beta H\, e^{-j\beta r_0}\right) \qquad (10\text{-}74)$$

Equations (72), (73), and (74) give the electric and magnetic intensities both near to and far from an antenna carrying a sinusoidal current distribution. It is quite remarkable that something so complex as the electromagnetic field close to an antenna should be expressible by such simple relations. The secret of this result lies in the integral of eq. (70), the integrand of which turned out to be a perfect differential. This happy circumstance occurred only because the current distribution was stipulated to be *sinusoidal* (although it also occurs for an unattenuated traveling-wave distribution, two of which can be combined to give the sinusoidal distribution). The result becomes even more remarkable when the expression for the important parallel component E_z is considered and interpreted. Examining eq. (72), it is seen that the first term represents a spherical wave originating at the top of the antenna. The numerator is the phase factor ($e^{j\omega t}$ is understood) and the denominator is the inverse-distance factor. Similarly, the second term represents a spherical wave of equal amplitude originating at the other end of the antenna, or, in the case of a monopole antenna on a reflecting plane, at the image point of the top of the antenna. Finally, the third term represents a wave originating at the center of the antenna (at the base in the case of a monopole antenna). The amplitude of this latter wave depends upon the antenna half-length H. It is zero for $H = \lambda/4$, that is, for a half-wave dipole or quarter-wave monopole. The sources of the spherical waves represented by the terms of eq. (72) are isotropic, that is they radiate uniformly in all directions, as is the case, for example, for a point source of sound. If a point source of sound is situated above a perfect (acoustic) reflecting plane, an equal image source is automatically obtained. Thus it is seen that the pressure field of a single, point source of sound located one-quarter wavelength above a reflecting plane will give a true representation of both magnitude and phase of the parallel component of electric intensity about a quarter-wave monopole or half-wave dipole carrying a sinusoidally distributed current. Antennas of other lengths may be represented by two point sources. This fact has been used* to give experimental data from which the (electrical) mutual impedance between antennas can be computed.

Equations (72) and (73) can be used to calculate the parallel and normal components of electric intensity in the immediate neighborhood of an antenna. Figures 10-10 and 10-11 show the calculated values† of the components of **E** at the surface of a half-wave dipole for the assumed sinusoidal current distribution. Fig-

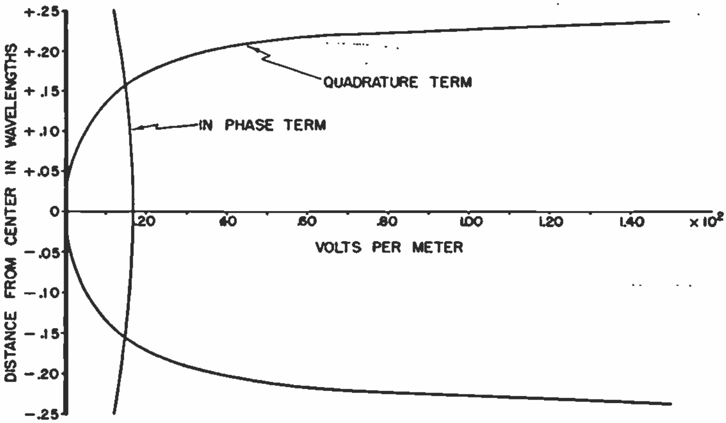


FIG. 10-10. In-phase and quadrature terms of parallel component of electric intensity along a half-wave dipole. (Calculated for a number 4 wire at 20 mc carrying an assumed sinusoidal current distribution.)

ure 10-10 shows the relative magnitudes of the in-phase and quadrature terms of E_z, the parallel components of **E**. Figure 10-11 compares the relative magnitudes of the quadrature terms of parallel and perpendicular components of **E**. It is seen that except very near the ends of the antenna the normal component is very large compared with the parallel component.

Problem 9. Verify that

$$\int \left[\frac{-j\beta y \, e^{-j\beta(r+h)}}{r^2} - \frac{y \, e^{-j\beta(r+h)}}{r^3} \right] dh = \frac{y \, e^{-j\beta(r+h)}}{r(r + h - z)}$$

where r, h, z, and y are as indicated in Fig. 10-9.

* E. C. Jordan and W. L. Everitt, "Acoustic Models of Radio Antennas," *Proc. IRE*, **29**, 4, 186 (1941).

† P. S. Carter, "Circuit Relations in Radiating Systems," *Proc. IRE*, **20**, 6, 1004 (1932).

10.09 Network and Antenna Theorems. In ordinary circuit theory certain network theorems have proven very useful in simplifying the solution of many problems. The validity of these theorems is based upon the linearity and/or the bilaterialism of the networks. In electromagnetic field theory the solution of any antenna problem can be obtained (at least in theory) by application of Maxwell's equations and the appropriate boundary conditions. The field equations themselves are linear and as long as the "constants" μ, ϵ, and σ of the media involved are truly constant, that is, do not vary with the magnitude of the signal (linearity) nor with direction (bilateralism), the same theorems can be applied. The usefulness of such theorems in antenna work is evidenced by the fact that with their aid nearly all the properties of a receiving antenna can be deduced from the known transmitting properties of the same antenna. A few of these theorems that find most usefulness in antenna problems are the following:

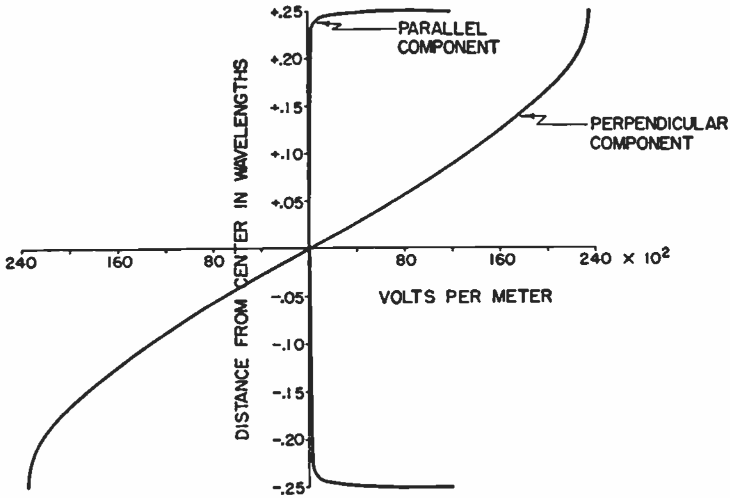


FIG. 10-11. Quadrature terms of both parallel and perpendicular components of electric intensity along a half-wave dipole. (Calculated for a number 4 wire at 20 mc carrying an assumed sinusoidal current distribution.)

Superposition Theorem. "In a network of generators and linear impedances, the current flowing at any point is the sum of the currents that would flow if each generator were considered separately, all other generators being replaced at the time by impedances equal to their internal impedances."

This fundamental principle follows directly from the *linearity* of the field equations and Ohm's law. When an impedance or network of impedances is *linear*, a given increase of voltage produces an increase of current that is independent of the magnitude of the current already flowing. Therefore the effect of each generator can be considered separately and independently of whether or not other generators are generating.

Thevenin's Theorem. "If an impedance Z_R be connected between any two terminals of a linear network containing one or more generators, the current which flows through Z_R will be the same as it would be if Z_R were connected to a simple generator whose generated voltage is the open circuit voltage that appeared at the terminals in question 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."

This theorem follows from the principle of superposition. A proof of it can be found in any of the references on circuit theory.

Maximum Power Transfer Theorem. "An impedance connected to two terminals of a network will absorb maximum power from the network when the impedance is equal to the conjugate of the impedance seen looking back into the network from the two terminals."

Corollary: "The maximum power that can be absorbed from a network equals $V_{oc}^2/4R$, where V_{oc} is the open-circuit voltage at the output terminals and R is the resistive component of the impedance looking back from the output terminals."

Compensation Theorem. "Any impedance in a network may be replaced by a generator of zero internal impedance, whose generated voltage at every instant is equal to the instantaneous potential difference that existed across the impedance because of the current flowing through it."

Reciprocity Theorem. "In any system composed of linear bilateral impedances, if an electromotive force V is applied between any two terminals and the current I is measured in any branch, the ratio of V to I, called the transfer impedance, will be the same as the ratio of V to I obtained when the positions of generator and ammeter are interchanged."

In the above statement of this theorem the generator and ammeter are assumed to have zero impedance. It is readily shown that the theorem also holds if the generator and ammeter have impedances which are equal.* The reciprocity theorem is one of the most powerful theorems in both circuit and field theory. It was originally stated by Rayleigh in a form somewhat similar to the above. A generalized statement of the theorem, suitable for application to fields as well as circuits, has been made and proven by Carson.†

Generalized Reciprocal Theorem. "Let a distribution of impressed periodic electric intensity $\mathbf{E}' = \mathbf{E}'(x, y, z)$, produce a corresponding distribution of current intensity $\boldsymbol{i}' = \boldsymbol{i}'(x, y, z)$, and let a second distribution of equiperiodic impressed electric intensity $\mathbf{E}'' = \mathbf{E}''(x, y, z)$ produce a second distribution of current density $\boldsymbol{i}'' = \boldsymbol{i}''(x, y, z)$, then

$$\int_{\text{vol}} (\mathbf{E}' \cdot \boldsymbol{i}'')\, dV = \int_{\text{vol}} (\mathbf{E}'' \cdot \boldsymbol{i}')\, dV \qquad (10\text{-}75)$$

the volume integration being extended over all conducting and dielectric media." $\mathbf{E}$ and $\boldsymbol{i}$ are vectors and $\mathbf{E} \cdot \boldsymbol{i}$ denotes the scalar product. This statement of the theorem was derived from Maxwell's equations and the only restriction is that magnetic matter be excluded (that is $\mu = \mu_v$). The statement for networks is a particular case of this more general statement and can be derived from it.

Fig. 10-12

Consider the application of this general theorem to the simple problem of two antennas (1) and (2) far removed from other conducting bodies (Fig. 10-12). Assume a voltage V_1' to be applied

* An alternative statement of the theorem, which is sometimes very useful, can be made in terms of an infinite-impedance (constant-current) generator and an infinite-impedance voltmeter. This alternative statement is: "In any system composed of linear bilateral impedances, the positions of a constant-current generator and an infinite-impedance voltmeter may be interchanged without affecting the voltage across the voltmeter, either in magnitude or phase, relative to the generator current."

† J. R. Carson, "Reciprocal Theorems in Radio Communications," *Proc. I.R.E.* 17, 6, 952 (1929).

at the terminals of (1), as a result of which a current I_2' flows in the zero impedance ammeter connected between the terminals of (2). Then apply a voltage V_2'' at (2) and obtain a current I_1'' at (1).

Now applying the general theorem, consider the term

$$\int_{\text{vol}} \mathbf{E}' \cdot i'' \, dV$$

For perfectly conducting antennas, E' will be zero everywhere along the antennas, except between the terminals of (1) where the voltage V_1 is applied. Therefore the first integral reduces to

$$\int_{\text{vol}} \mathbf{E}' \cdot i'' \, dV = V_1' I_1''$$

Similarly the second integral becomes

$$\int_{\text{vol}} \mathbf{E}'' \cdot i' \, dV = V_2'' I_2'$$

Thus the general statement reduces to

$$V_1' I_1'' = V_2'' I_2'$$

or

$$\frac{V_1'}{I_2'} = \frac{V_2''}{I_1''} \tag{10-76}$$

which is the simple statement for networks.

The reciprocity theorem has proven a powerful and useful tool in circuit and field theory, and many corollary statements have been derived from it. Some of these corollaries, especially those concerning reciprocity of powers, have been derived under special conditions. If applied in circumstances where these special conditions are violated, the corollary statements may break down. The reciprocity theorem itself, in the form of eqs. (75) or (76) is perfectly general, and always gives the correct answer as long as only linear bilateral circuits or media are involved.

Antenna Theorems. From the above theorems can be deduced several very useful antenna theorems, relating the properties of transmitting and receiving antennas. So far, most of the analysis has been concerned with transmitting antennas for which the assumption of sinusoidal current distribution is known to yield results of usable accuracy. On the other hand, for antennas excited as receiving antennas, the current distribution varies with the direction of arrival of the received field and is not even approximately sinusoidal, except for the resonant lengths (half-wave dipole, etc.). This being the case, it is by no means obvious that the directional and impedance properties of an antenna should be identical for the transmitting and receiving conditions. Because, in the general case, the current distribution is not sinusoidal for the receiving antenna, direct computation of its properties is usually a relatively complicated problem. The following theorems make it possible to infer the properties of a receiving antenna from its properties as a transmitting antenna, and vice versa.

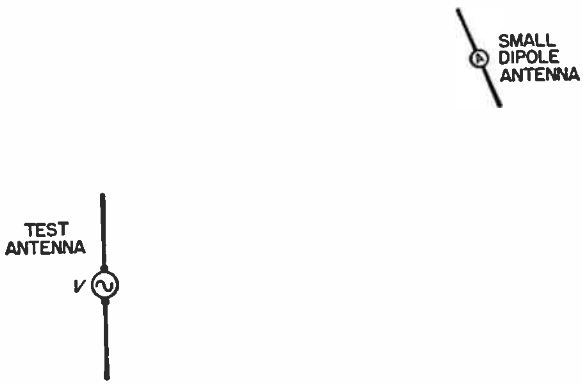


Fig. 10-13

Equality of Directional Patterns. "The directional pattern of a receiving antenna is identical with its directional pattern as a transmitting antenna."

Proof: This theorem results directly from an application of the reciprocity theorem. The directional pattern of a transmitting antenna is the polar characteristic that indicates the intensity of the radiated field at a fixed distance in different directions in space. The directional pattern of a receiving antenna is the polar characteristic that indicates the response of the antenna to unit field intensity from different directions. The pattern as a transmitting antenna could be measured as indicated in Fig. (10-13) by means of a short exploring dipole moved about on the surface of a large sphere centered at the antenna under test. (The exploring dipole is always oriented so as to be perpendicular to the radius vector and in the plane containing the electric vector). A voltage V is applied to the test antenna, and the current I flowing in the short dipole antenna will be a measure of the electric field at the position of the dipole antenna. If then the voltage V is applied to the dipole and the test antenna current is measured, the receiving pattern of the test antenna can be obtained. But by the reciprocity theorem, for every location of the probe antenna, the ratio of V to I is the same as before. Therefore the radiation pattern as a receiving antenna will be identical with the pattern as a transmitting antenna.

When the test antenna radiates an elliptically polarized wave, that is, when the radiated electric intensity has two components, E_θ and E_ϕ, that are not in time phase, the radiation patterns for the θ polarization and ϕ polarization are shown separately. The

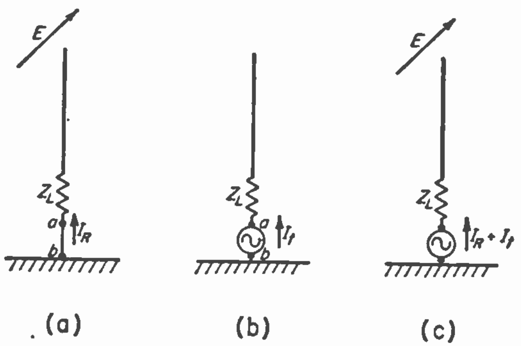


Fig. 10-14

pattern for the particular polarization specified is obtained by keeping the exploring dipole parallel to that polarization. It follows as before that for *each* of the polarizations the radiation patterns for transmitting and receiving will be the same.

Equivalence of Transmitting and Receiving Antenna Impedances. "The impedance of an antenna referred to a given pair of terminals is independent of the method of excitation. In particular, the antenna impedance for receiving is the same as for transmitting."

Proof: This theorem follows directly from the principle of superposition. Consider the antenna of Fig. 10-14 connected in series with an impedance Z_L and excited as a receiving antenna (a), a transmitting antenna (b), or both (c). As a receiving antenna (a), a field **E** is assumed to be incident upon the antenna, and a current I_R flows through the load impedance Z_L and the shorting link connecting terminals *a-b*. Applying Thevenin's theorem, the value of I_R will be

$$I_R = \frac{V_\infty}{Z_r + Z_L}$$

where V_∞ is the open-circuit voltage at *a-b* and Z_r is the antenna impedance for the receiving condition.

Now insert at terminals *a-b* an opposing voltage V such that

$$V = -V_\infty$$

This will reduce the current through Z_L and at *a-b* to zero (Fig. 10-14*c*).

Finally remove the exciting field **E**, but continue to apply the voltage V at *a-b*. The current I_t, which now flows, will be given by

$$I_t = \frac{V}{Z_t + Z_L}$$

where Z_t is the antenna impedance for the transmitting condition.

By the principle of superposition the current which flows when both excitations are applied simultaneously is the sum of the currents which flow when the excitations are applied separately. In this case the sum of the currents is zero, that is

$$I_R + I_t = 0$$

or

$$I_R = -I_t$$

From this it follows that

$$\frac{V_\infty}{Z_r + Z_L} = -\frac{V}{Z_t + Z_L} \qquad (10\text{-}77)$$

and since $V = -V_\infty$, the only way (77) can be satisfied is to have

$$Z_r = Z_t = Z_a \qquad (10\text{-}78)$$

Thus the impedance of an antenna is the same, whether it be used as a transmitting antenna or as a receiving antenna. The relation (78) will hold whatever may be the value of Z_L. In actual operation Z_L would be the load impedance in the receiving case and the generator impedance in the transmitting case.

In the discussion above, no restriction was put upon the exciting field **E**. It could be obtained as the result field owing to any number and combination of radiators. However, it is necessary for these radiators to be *distant* radiators; otherwise the impedance Z_r or Z_t will not represent the self-impedance of the antenna. If the radiators producing the excitation **E** are sufficiently close that the mutual impedance between them and the antenna being considered is not negligible (compared with the impedance of this antenna) then Z_r will actually be the self-impedance of the antenna plus the impedance coupled-in owing to the presence of the other antennas. However, even under these conditions the relation $Z_r = Z_t$ will hold if in the transmitting case the radiators that produced **E** are left in position and connected to impedances equal to the impedances of the generators that excited them.

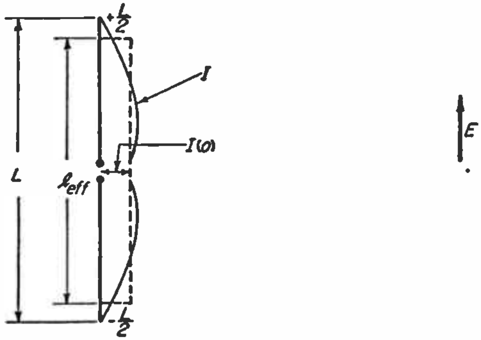


Fig. 10-15

Equality of Effective Lengths. The *effective length,* l_{eff}, of an antenna is a term used to indicate the effectiveness of the antenna as a radiator or collector of electromagnetic energy. The significance of the term as applied to transmitting antennas is illustrated in Fig. 10-15. The effective length of a transmitting antenna is that length of an equivalent linear antenna that has a current $I(0)$ at all points along its length and that radiates the same field intensity as the actual antenna in the direction perpendicular to its length. $I(0)$ is the current at the terminals of the actual antenna. That is, for transmitting

$$I(0)l_{eff}(\text{trans}) = \int_{-L/2}^{+L/2} I\,dl$$

or

$$l_{eff}(\text{trans}) = \frac{1}{I(0)} \int_{-L/2}^{+L/2} I\,dl \qquad (10\text{-}79)$$

The effective length of a receiving antenna is defined in terms cf the open-circuit voltage developed at the terminals of the antenna for a given received field intensity **E**. That is, for receiving

$$l_{\text{eff}}(\text{rec}) = \frac{V_{\infty}}{E} \tag{10-80}$$

where V_{∞} is the open-circuit voltage produced at the terminals of the antenna because of a uniform exciting field E volts per meter parallel to the antenna.

Using the reciprocity theorem it is readily shown that these two definitions yield the same value cf effective length for a given

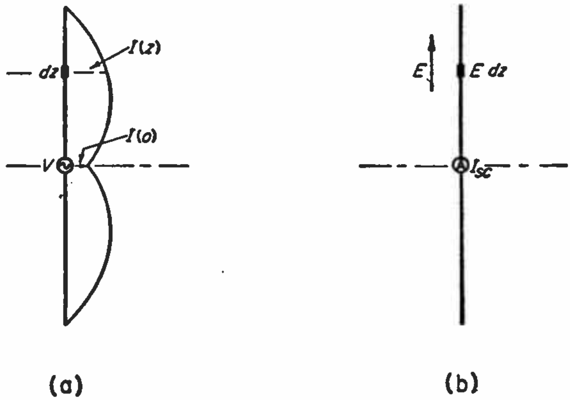


FIG. 10-16

antenna. First consider the transmitting antenna case (Fig. 10-16a). A voltage V applied at the terminals produces a current $I(0) = V/Z_a$ at the terminals and a current $I(z)$ at any point z along the antenna. Z_a is the antenna impedance measured at the terminals.

Next consider the same antenna as a receiving antenna (Fig. 10-16b) where an exciting field E is parallel to the vertical portion of the antenna. In each elemental length dz there will be induced a voltage $E\,dz$. Now by the reciprocity theorem a voltage $E\,dz$ at z will produce a current at the terminals (with terminals short-circuited) of amount

$$dI = \frac{E\,dz}{V}\,I(z)$$

The total current at the short-circuited terminals due to all the induced voltages $E\,dz$ along the antenna will be

$$I_{sc} = \int \frac{E\,dz}{V} I(z)$$
$$= \frac{E}{V} \int I(z)\,dz$$

since E is uniform along the antenna. The open-circuit voltage at the base due to this induced voltage will be

$$V_{oc} = I_{sc} Z_a$$
$$= \frac{EZ_a}{V} \int I(z)\,dz$$

But $$\frac{V}{Z_a} = I(0)$$

Therefore $$\frac{V_{oc}}{E} = \frac{1}{I(0)} \int I(z)\,dz = l_{eff}(\text{trans})$$

Therefore $l_{eff}(\text{rec}) = l_{eff}(\text{trans})$. That is, the effective length of an antenna for receiving is equal to its effective length as a transmitting antenna.

At low and medium frequencies the antenna usually consists of a vertical radiator (with or without top-loading), fed against ground. Under these conditions the "image" forms half of the antenna system, and it becomes convenient to speak of the half-length of the antenna system. In this terminology a vertical antenna a quarter-wavelength high has a half-length of $\lambda/4$. Its *effective half-length** is the length of linear antenna, also fed against ground, that will produce the same field intensity **E** when carrying a current $I(0)$ at all points along its length. The effective half-length of a quarter-wave vertical antenna with a sinusoidal current distribution is $2/\pi \times \lambda/4 = \lambda/2\pi$. A curve of effective length *vs.* length for straight radiators having the sinusoidal current distribution is plotted in Fig. 10-17.

* This was originally termed the *effective height* of the ground-based antenna, but this usage has been discontinued because of confusion with the height above the ground of elevated antennas. The *effective height* of an (elevated) antenna is now defined as the height of its center of radiation above the effective ground level.

Generalized Effective Length. As ordinarily used, the effective length of an antenna is a measure of the effectiveness of the antenna as a radiator or receptor in directions perpendicular to its axis. However, the term can be generalized to give a measure of radiated field intensity in any direction, and in this case the effective length of an antenna will be a function of direction (θ and ϕ). The equivalent linear antenna, having the constant current $I(0)$ along its length, is always oriented to be perpendicular to the direction

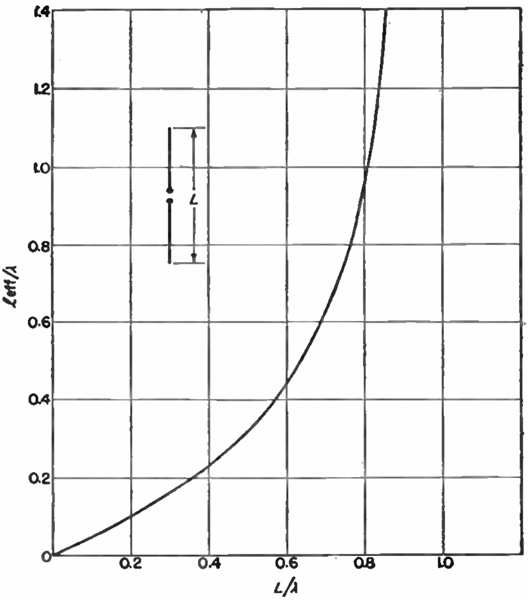


FIG. 10-17. Effective length vs. length for dipole antennas.

being considered. (It lies in the plane containing **E** and the radius vector.) In the most general case the radiated field intensity from an actual antenna may be elliptically polarized. This can be handled by specifying separately the effective lengths of the antenna for two components of the electric intensity **E**; for example, an effective length l^{θ}_{eff} for the θ component and l^{ϕ}_{eff} for the ϕ component. Since the θ and ϕ components of **E** will, in general, not have the same time phase, the resultant generalized effective length l'_{eff} will be a complex vector. That is

$$\mathbf{l}'_{eff} = \mathbf{u}_{\theta} l^{\theta}_{eff} + K\mathbf{u}_{\phi} l^{\phi}_{eff} \tag{10-81}$$

where K is a unit complex number, and l^{θ}_{eff} and l^{ϕ}_{eff} are functions (usually different) of θ and ϕ. Using the reciprocity theorem it is easy to show that this generalized effective length will be the same for receiving as transmitting.

The problem of elliptical polarization will be considered further in sec. (12.12).

Field Intensity in Terms of Effective Length. A current I amperes flowing in an elemental length dl produces a radiation field intensity at a distance r meters in a direction normal to the current element of amount

$$E = \frac{60\pi I\,dl}{r\lambda} \qquad \text{volts/meter}$$

Therefore the field intensity produced by an antenna having an effective length l_{eff} will be

$$E = \frac{60\pi I(0) l_{\text{eff}}}{r\lambda} \qquad \text{volts/meter} \qquad (10\text{-}82)$$

where $I(0)$ is the current at the terminals. If the generalized effective length is used in expression (82), the radiation field will be completely specified in all directions. Conversely, if the field intensity E is known, the effective length is given in terms of E by

$$l_{\text{eff}} = \frac{r\lambda E}{60\pi I(0)} \qquad (10\text{-}83)$$

From eq. (82) it is seen that the effective length of an antenna gives the field intensity E at a given distance for a given current flowing at the antenna terminals (or some other specified reference point). However, no indication is given of the power required to produce the field intensity. A term that can be used to specify the field intensity at a given distance (and therefore the power radiated per unit solid angle in that direction) for a given total power is antenna gain. The gain of an antenna will be considered in chap. 12.

ADDITIONAL PROBLEMS

10. (a) Using the reciprocity principle, show that the current at the center of a half-wave dipole is

$$I = \frac{\lambda E}{\pi Z_a}$$

when excited by a uniform field E parallel to the antenna and when the terminals are short-circuited. Z_a is the antenna impedance measured at the terminals. Assume that the current distribution as a transmitting antenna is sinusoidal, that is $I(z) = I_m \sin \beta(H - |z|) = I_m \cos \beta z$ (for the half-wave dipole).

(b) Under the same conditions what will be the open-circuit voltage at the antenna terminals? How much current would flow in a load impedance Z_L connected to those terminals?

11. If the dipole receiving antenna of problem 10 has any length $L = 2H$, show that the short-circuit current at the terminals would be

$$I = \frac{E\lambda(1 - \cos \beta H)}{Z_a \pi \sin \beta H}$$

Assume that, as a transmitting antenna, the current distribution would be sinusoidal, i.e.,

$$I(z) = I_m \sin \beta(H - |z|)$$

12. The current distribution in a half-wave dipole transmitting antenna is known to be nearly sinusoidal. The current distribution of a half-wave dipole receiving antenna is also nearly sinusoidal when the terminals are short-circuited. Using the superposition and compensation theorems, verify that the current distribution of the half-wave dipole as a receiving antenna must be approximately sinusoidal when the terminals are connected to any load impedance Z_L.

13. A half-wave dipole antenna is excited by a uniform electric field **E**, which is parallel to a plane through the antenna, but which arrives from a polar angle θ. (In problem 10, θ was 90°.) Using the reciprocity principle, determine the expression for the short-circuit current. Do it *two* ways.

$$\textit{Answer: } I_{sc} = \frac{\lambda E}{\pi Z_a} \frac{\cos\left(\frac{\pi}{2} \cos \theta\right)}{\sin \theta}.$$

14. Assuming the sinusoidal current distribution

$$I(z) = I_m \sin \beta(H - |z|)$$

derive the expression for the effective length of a dipole antenna of length $L = 2H$.

$$\textit{Answer: } l_{eff} = \frac{\lambda(1 - \cos \beta H)}{\pi \sin \beta H}.$$

15. (a) At what distance from a 60-cycle circuit is the radiation field approximately equal to the induction field?

(b) Approximately what are the relative amplitudes of the radiation, induction, and electric fields at a distance of 1 wavelength from a Hertzian dipole?

16. (a) Show that the unattenuated radiation field at the surface of the earth of a quarter-wave monopole is given by the formula

$$E = \frac{6.14}{r}\sqrt{W} \qquad \text{mv/m effective}$$

where r is in miles and W is the power radiated in watts. The *unattenuated* field is the value of field intensity that would exist if the earth were perfectly conducting.

(b) Derive the corresponding expression for a short monopole ($H \ll \lambda$).

17. An "inverted L" receiving antenna has vertical and horizontal legs which are both 10 meters long. The field intensity of the received wave is vertical and has a measured value of 1 mv/m. Frequency is 1 mc. What is the open-circuit voltage at the base of the antenna? (NOTE: as a transmitting antenna the current distribution on an "inverted L" is approximately sinusoidal, with the current node at the open end of the horizontal arm.)

18. A half-wave dipole is located parallel to and one-quarter wavelength from a plane metallic reflecting sheet. Sketch the lines of current flow in the sheet. (Suggestion: Use the image principle and the relation $\mathbf{J} = \mathbf{n} \times \mathbf{H}$.)

19. Short vertical monopole antennas that are suitably "top-loaded" with a capacitive load, have an essentially uniform current along their whole length. Set up the vector potential and derive an expression for the average value (in time) of the Poynting vector at large distances from such an antenna.

20. (a) Set up an expression for the vector potential due to a traveling-wave current distribution

$$[I(z) = I_m e^{-j\beta z}]$$

along a terminated wire antenna of length L.

(b) Show that the distant field of such an antenna is given by

$$E_\theta = \frac{30 I_m \sin\theta}{r(1 - \cos\theta)}\,[2 - 2\cos\beta(1 - \cos\theta)L]^{1/2}$$

21. (a) A toroidal coil has a large number of closely wound turns on a core of high permeability so that at d-c virtually all of the magnetic flux is confined to the core (Fig. 10-18). When an alternating current flows in the winding, is there at the point P (1) a value of vector potential $\mathbf{A}$? (2) a value of magnetic intensity $\mathbf{H}$? (3) a value of electric intensity $\mathbf{E}$?

Assuming a current of 1 amp at 60 cps through a 1000-turn winding, and $\mu = 1000\mu_v$

(b) approximately what is the voltage around the path s?

(c) What is the order of magnitude of each of the vectors of parts (1), (2) and (3)?

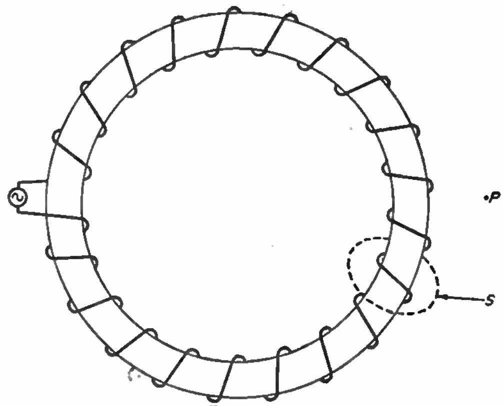


FIG. 10-18. A toroidal coil on a high-permeability core.

22. It is possible to define a single function, called the Hertzian vector, from which both electric and magnetic intensities may be derived. This vector is

$$\mathbf{Z} = \int \mathbf{A}\, dt$$

Using this function, show that:

(a) $\mathbf{E} = -\mu\ddot{\mathbf{Z}} + \frac{1}{\epsilon}\nabla(\nabla \cdot \mathbf{Z})$

(b) $\mathbf{H} = \nabla \times \dot{\mathbf{Z}}$

(c) $\nabla^2\mathbf{Z} = \mu\epsilon\ddot{\mathbf{Z}}$ (in free space, where $\nabla \times \mathbf{H} = \dot{\mathbf{D}}$)

BIBLIOGRAPHY

Pistolkors, A. A., "The Radiation Resistance of Beam Antennas," *Proc. IRE*, **17**, 562 (March, 1929).

Carter, P. S., "Circuit Relations in Radiating Systems," *Proc. IRE*, **20**, 1004 (June, 1932).

Brown, G. H., "Directional Antennas," *Proc. IRE*, **25**, 78 (January 1937).

Burgess, R. E., "Aerial Characteristics," *Wireless Engineer*, **21**, 247, 154 (April, 1944).

Moullin, E. B., "The Radiation Resistance of Aerials Whose Length is Comparable with the Wavelength," *J. IEE*, **78**, 540 (1936); *also Wireless Section, IEE*, **11**, 93 (June, 1936).

Carson, J. R., "Electromagnetic Theory and the Foundations of Electric Circuit Theory," *Bell System Technical Journal*, **6**, 1–17 (January, 1927).

Bechmann, R., "Calculation of Electric and Magnetic Field Strengths of any Oscillating Straight Conductors," *Proc. IRE*, **19**, 461 (March, 1931).

Bechmann, R., "On the Calculation of the Radiation Resistance of Beam Antennas," *Proc. IRE*, **19**, 1471 (August, 1931).

Labus, J., "Rechnerische Ermittlung der Impedanz von Antennen," *Hochfrequenz und Elektroakustik*, **41**, 17 (January, 1933).

CHAPTER 11

IMPEDANCE

PART I

Induced-EMF Method of Calculating Impedances

In chap. 10 the radiation resistance of an antenna was obtained by the Poynting vector method. In that method the distant or radiation field of the antenna was determined, and then the power radiated was computed by integrating the Poynting vector over a large spherical surface. The same method can be used to determine the total power radiated by an array of antennas, but it will give no information as to how much power is contributed by the individual elements of the array. Moreover, whereas the distant field determines the power radiated, and hence the radiation resistance of an antenna, the reactive power and reactance of an antenna are determined by the near fields in close to the antenna, and so the distant-field method cannot be used to obtain reactance.

These limitations are not present in an alternative method, which was developed for obtaining the radiation resistance of single antennas and then extended to obtain the impedance of antennas in a multi-element array. In this second method, which is known as the *induced-emf method*, the electric intensity produced at any point P near or far from the antenna is first determined. Then, if the radiation resistance of a single antenna is required, the point P is brought back to the surface of the antenna and the voltage induced in each element of length of the antenna is determined. The power required to produce the assumed current against the opposition of this induced voltage is computed for each element of length, and then the total power is obtained by integrating over the length of the antenna. This gives the total power required to establish the assumed current against the self-induced emf and hence gives the power radiated by the antenna (ohmic losses are neglected). From this the radiation resistance can be determined. The mutual impedance between antennas can be obtained in a similar manner by calculating the power required to drive the assumed current in one antenna against the voltage induced in it by current flow in a second antenna.

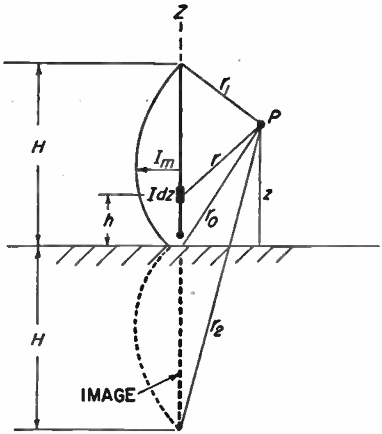


FIG. 11-1

11.01 Radiation Resistance by the Induced-emf Method. The exact expressions for the electric intensity about the antenna of Fig. 11-1, which is assumed to have a sinusoidal current distribution, are given by eqs. (10-72) and (10-73). For the component of E parallel to the axis of the antenna the expression is

$$E_z = 30I_m\left(\frac{-j\,e^{-j\beta r_1}}{r_1} + \frac{-j\,e^{-j\beta r_2}}{r_2} + \frac{2j\cos\beta H\,e^{-j\beta r_0}}{r_0}\right) \quad (11\text{-}1)$$

When the point P is on the antenna:

$$r_0 = z \qquad r_1 = H - z \qquad r_2 = H + z$$

The power required to produce the assumed current against the induced voltage will be

$$P = -\tfrac{1}{2}\int_0^H |E_z||I(z)|\cos\psi\,dz \quad (11\text{-}2)$$

where $|E_z|$ and $|I(z)|$ are the magnitudes of E_z and $I(z)$ and ψ is the phase angle between them. In eq. (1) the factor $-j\,e^{-j\beta r_1}$ in the first term indicates that the electric intensity represented by this term lags the current I_m by the angle $[(\pi/2) + \beta r_1]$. The power factor for this term is then

$$\cos\left(\frac{\pi}{2} + \beta r_1\right) = -\sin\beta r_1$$

Similarly the power factor for the second and third terms is established and, remembering that $I(z) = I_m \sin\beta(H - z)$, the expression for power becomes

$$P = 15I_m^2 \int_0^H \left[\frac{\sin\beta(H-z)\sin\beta(H-z)}{H-z} + \frac{\sin\beta(H-z)\sin\beta(H+z)}{H+z} - \frac{2\cos\beta H\sin\beta(H-z)\sin\beta z}{z} \right] dz \qquad (11\text{-}3)$$

This expression can be integrated term by term. Consider the first term and let $u = \beta(H - z)$, $du = -\beta\, dz$, $du/u = -dz/(H - z)$ so that when $z = 0$, $u = \beta H$, and when $z = H$, $u = 0$. Then

$$\int_0^H \frac{\sin\beta(H-z)\sin\beta(H-z)}{H-z}\,dz = -\int_{\beta H}^0 \frac{\sin^2 u}{u}\,du$$
$$= \frac{1}{2}\int_0^{2\beta H} \frac{1-\cos(2u)}{(2u)}\,d(2u) = \frac{1}{2}S_1(2\beta H)$$

where the function $S_1(x)$ is defined by eq. (10-67).

To integrate the second term of eq. (3) let

$$w = 2\beta(H+z) \qquad dw = 2\beta dz \qquad 2\beta z = w - 2\beta H$$

Then $\displaystyle\int_0^H \frac{\sin\beta(H-z)\sin\beta(H+z)}{H+z}\,dz$

$$= \frac{1}{2}\int_0^H \frac{\cos 2\beta z - \cos 2\beta H}{(H+z)}\,dz$$
$$= -\frac{1}{2}\int_{2\beta H}^{4\beta H} \frac{\cos 2\beta H - \cos(w - 2\beta H)}{w}\,dw$$
$$= -\frac{1}{2}\int_{2\beta H}^{4\beta H} \frac{\cos 2\beta H(1-\cos w) - \sin w \sin 2\beta H}{w}\,dw$$
$$= -\frac{1}{2}\cos 2\beta H \int_{2\beta H}^{4\beta H} \frac{1-\cos w}{w}\,dw + \frac{1}{2}\sin 2\beta H \int_{2\beta H}^{4\beta H} \frac{\sin w}{w}\,dw$$
$$= -\frac{1}{2}\cos 2\beta H \left(\int_0^{4\beta H} \frac{1-\cos w}{w}\,dw - \int_0^{2\beta H} \frac{1-\cos w}{w}\,dw \right)$$
$$+ \frac{1}{2}\sin 2\beta H \left(\int_0^{4\beta H} \frac{\sin w}{w}\,dw - \int_0^{2\beta H} \frac{\sin w}{w}\,dw \right)$$
$$= -\frac{1}{2}\cos 2\beta H[S_1(4\beta H) - S_1(2\beta H)] + \frac{1}{2}\sin 2\beta H[\mathrm{Si}\,(4\beta H) - \mathrm{Si}\,(2\beta H)]$$

where Si (x) is the sine integral of x and is defined by eq. (10-66)

The third term of eq. (3) can be integrated almost directly. It is

$$-2 \cos \beta H \int_0^H \frac{\sin \beta(H - z) \sin \beta z}{z} dz$$
$$= - \cos \beta H \int_0^H \frac{\cos \beta(H - 2z) - \cos \beta H}{z} dz$$
$$= \cos \beta H \int_0^{2\beta H} \frac{\cos \beta H(1 - \cos 2\beta z) - \sin \beta H \sin 2\beta z}{2\beta z} d(2\beta z)$$
$$= \cos^2 \beta H S_1(2\beta H) - \cos \beta H \sin \beta H \text{ Si } (2\beta H)$$

Using the notation $2\beta H = b$, the power radiated as given by equation (3) is

$$P = {}^{15}\!/_{2} I_m{}^2[S_1(b) - (S_1(2b) - S_1(b)) \cos b + (\text{Si } (2b) - \text{Si } (b)) \sin b + (1 + \cos b)S_1(b) - \sin b \text{ Si } (b)] \quad (11\text{-}4)$$

Rearranging terms and dividing by $I_m{}^2/2$, an expression for radiation resistance is obtained.

$$R_{rad} = 15[(2 + 2 \cos b)S_1(b) - \cos bS_1(2b) - 2 \sin b \text{ Si } (b) + \sin b \text{ Si } (2b)] \quad (11\text{-}5)$$

Expression (5) gives the radiation resistance of the antenna, *referred to the current loop*, I_m. This resistance is plotted in Fig. 11-2 for monopole antennas as a function of antenna height H. For center-fed dipole antennas of half-length H, the values of resistance should be multiplied by 2. For example, for a quarter-wave monopole mounted on a reflecting plane the radiation resistance is shown as 36.5 ohms. For a half-wave dipole (in free space) it would be 73 ohms. For a half-wave monopole and a center-fed full-wave dipole the values of radiation resistance are 99.5 and 199 ohms respectively.

11.02 Radiation Resistance Referred to the Base. The radiation resistance of an antenna is that value of resistance which, when multiplied by the square of the antenna current (effective value), gives the power radiated. If the loop current I_m is used, the radiation resistance is said to be *referred to the loop*. If the base current $I(0)$ (or the current at the terminals in the case of a center-fed dipole) is used, the radiation resistance is said to be *referred to the* *base* (or referred to the feed point). Since the radiated power is the same in both cases

$$\tfrac{1}{2}I_m^2 R_r \text{ (loop)} = \tfrac{1}{2}I^2(0) R_r \text{ (base)} \tag{11-6}$$

where I_m and $I(0)$ are both peak values in time. For an antenna having the sinusoidal current distribution, the base or terminal

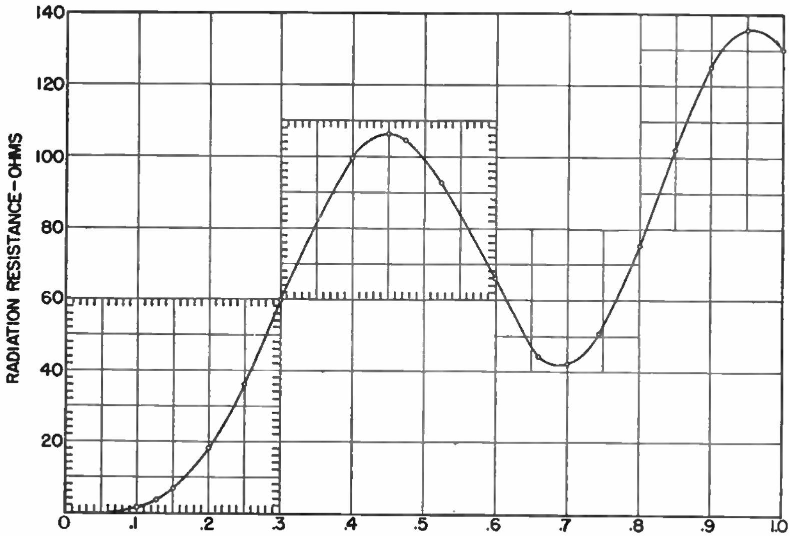


FIG. 11-2. Radiation resistance (referred to the current loop) of monopole antenna as a function of antenna height H/λ.

current is related to the loop current by

$$I(0) = I_m \sin \beta H$$

Therefore, for this case,

$$R_r \text{ (base)} = \frac{I_m^2}{I^2(0)} R_r \text{ (loop)}$$

$$= \frac{R_r \text{ (loop)}}{\sin^2 \beta H} \tag{11-7}$$

For antenna lengths for which H is a multiple of a half-wavelength, the assumed sinusoidal distribution gives a value of zero for the current at the feed point, and eq. (7) indicates that the input resistance will be infinite. For these lengths the actual input current will be small but not zero, and the input resistance will be large but not infinite. Although a value of infinity may be regarded as the first approximation to the actual resistance, it is a worthless approximation for practical purposes. For these cases it becomes necessary to use a method involving a better approximation than the sinusoidal for current distribution. Such methods are considered in chap. 13. It is seen that the sinusoidal current distribution can be used to give useful answers over a certain range of antenna lengths, but there are other ranges in which the approximation fails. In general this is true of every approximate method, and it is necessary for the engineer always to consider the limitations as well as the capabilities of any method he may employ.

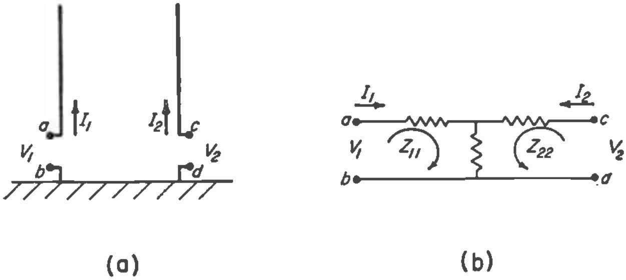


FIG. 11-3

11.03 Mutual Impedance between Antennas. When two or more antennas are used in an array, the driving-point impedance of each antenna depends upon the self-impedance of that antenna and in addition upon the mutual impedance between that antenna and each of the others. For example, consider the two-element array of Fig. 11-3 in which base currents I_1 and I_2 flow as a result of voltages V_1 and V_2 applied at the bases. As far as voltages and currents at the terminals *a-b* and *c-d* are concerned the two antennas of Fig. 11-3a can be represented by the general four-terminal network of Fig. 11-3b.

Z_{11} is the impedance measured at the terminals *a-b* with the terminals *c-d* open; that is, Z_{11} is the *mesh impedance* of mesh 1. Similarly Z_{22} is the mesh impedance of mesh 2, and is the impedance that would be measured at terminals *c-d* with *a-b* open. $Z_{12} = Z_{21}$ is the mutual impedance between the antennas and is defined in both figures by

$$Z_{21} = \frac{V_{21}}{I_1} \qquad Z_{12} = \frac{V_{12}}{I_2}$$

where V_{21} is the open-circuit voltage induced across terminals *c-d* of antenna 2 owing to current I_1, flowing (at the base) in antenna 1. Similarly V_{12} is the open-circuit voltage across the terminals of antenna 1 owing to current I_2, flowing in antenna 2. Under most conditions the impedance Z_{11} is approximately equal to the *self-impedance* Z_{S_1} of antenna 1. The self-impedance of an antenna is its input impedance with all other antennas entirely removed. Except when antenna 2 is very near a resonant length (that is $H \approx \lambda/2$) or when it is very close to antenna 1, the input impedance of antenna 1 will be nearly the same with antenna 2 open-circuited as it would be with antenna 2 entirely removed from the field of antenna 1.

FIG. 11-4

The mesh equations for Fig. 11-3 are

$$\left.\begin{aligned} V_1 &= I_1Z_{11} + I_2Z_{12} \\ V_2 &= I_1Z_{21} + I_2Z_{22} \end{aligned}\right\} \tag{11-8}$$

Let $r = I_1/I_2$, where in general r is a complex number. Then

$$\frac{V_1}{I_1} = Z_{11} + \frac{1}{r}Z_{12} \tag{11-9}$$

$$\frac{V_2}{I_2} = rZ_{21} + Z_{22} \tag{11-10}$$

It is seen that the input impedances, V_1/I_1 and V_2/I_2 are dependent upon the current ratio r. It is these impedances that any impedance-transforming networks must be designed to feed, and in order to calculate them the mutual impedance must be known.

In the Fig. 11-3 and the corresponding mesh eqs. (8), the voltages V_1 and V_2 are assumed to be supplied by zero-impedance generators. When the generators have finite impedances, Z_{g1} and Z_{g2}, equations (8) still apply if the generator impedances are included in the mesh impedances. The representation then becomes that of Fig. 11-4 for which the mesh equations would be

$$\begin{aligned} V_1 &= I_1\bar{Z}_{11} + I_2Z_{12} \\ V_2 &= I_1Z_{21} + I_2\bar{Z}_{22} \end{aligned} \qquad (11\text{-}11)$$

where $\bar{Z}_{11} = Z_{11} + Z_{g1}$ and $\bar{Z}_{22} = Z_{22} + Z_{g2}$

and where V_1 and V_2 are now supplied by zero-impedance generators, which produce voltages equal to the open-circuit voltages of the actual generators.

11.04 Computation of Mutual Impedance. The mutual impedance Z_{21} between the two antennas of Fig. 11-3 or 11-4 is defined by

$$Z_{21} = \frac{V_{21}}{I_1} \qquad (11\text{-}12)$$

where V_{21} is the open-circuit voltage at the terminals of antenna 2 due to a base current I_1 in antenna 1. Now the electric intensity at all points along antenna 2 due to a current in antenna 1 can be calculated, and the problem is that of determining the open-circuit voltage at the terminals of antenna 2, which results from the voltages induced in all the elemental lengths of the antenna. This result may be obtained by an application of the reciprocity theorem.

Consider antenna 2 with antenna 1 in place, but V_1 not generating (Fig. 11-4). A voltage $V_2 = I_2(0)Z_2'$ applied at the terminals ($z = 0$) of antenna 2 will produce a base current $I_2(0)$ and current at any point z, which will be indicated by $I_2(z)$. The impedance Z_2' is the impedance looking into the terminals of 2 and is given by

$$Z_2' = \bar{Z}_{22} - \frac{Z^2_{12}}{\bar{Z}_{11}}$$

Applying the reciprocity theorem, if a voltage $I_2(0)Z_2'$ applied at the base produces a current $I_2(z)$ at a point z along antenna 2, then a voltage $E_{z1}\,dz$, induced at z, will produce at the base a (short-circuit) current

$$dI_{sc} = \frac{E_{z1}\,dz}{I_2(0)Z_2'}\,I_2(z)$$

The total short-circuit current at the base due to voltage induced along the entire length of the antenna will be

$$I_{sc} = \frac{1}{I_2(0)Z_2'}\int_0^{H_2} E_{z1}I_2(z)\,dz$$

By Thevenin's theorem the open-circuit voltage at the base then will be

$$\begin{aligned} V_{21} &= -I_{sc}Z_2' \\ &= -\frac{1}{I_2(0)}\int_0^{H_2} E_{z1}I_2(z)\,dz \end{aligned}$$

The minus sign results from the fact that a current through the short-circuited terminals in the assumed positive direction (upwards) requires on open circuit a voltage across *c-d* of negative polarity (that is, *d* positive with respect to *c*, whereas the assumed positive polarity is *c*, positive with respect to *d*).

The expression for mutual impedance between the antennas is

$$Z_{21} = \frac{V_{21}}{I_1(0)} = -\frac{1}{I_1(0)I_2(0)}\int_0^{H_2} E_{z1}I_2(z)\,dz \qquad (11\text{-}13)$$

In expression (13), $I_2(z)$ is the current distribution along antenna 2 when fed by a voltage at the base and with antenna 1 closed through the generator impedance Z_{g_1}; also E_{z1} is the voltage induced along antenna 2 due to a base current $I_1(0)$ in antenna 1, and with antenna 2 closed through the generator impedance Z_{g_2}. Now because of *linearity* the mutual impedance between the antennas will be independent of the generator impedances Z_{g_1} and Z_{g_2}. By making the latter very large the evaluation of the expression for mutual impedance can be greatly simplified. Thus, if Z_{g_1} is very large, antenna 1 is effectively open-circuited when V_1 is not generating, and, except for the special case of $H_1 \approx \lambda/2$, the current distribution $I_2(z)$ along antenna 2 due to a voltage at the base of antenna 2 will be nearly the same as it would be with antenna 1 removed. The distribution $I_2(z)$ can therefore be represented by the sinusoidal approximation for an isolated transmitting antenna

$$I_2(z) = I_{2m}\sin\beta(H_2 - z) \qquad (11\text{-}14)$$

Again with Z_{g_2} very large and V_2 not generating, antenna 2 is essentially open-circuited and the distribution along it of the induced voltage $E_{z_1}\,dz$ due to current in antenna 1 is just that already given in equation (1).

Inserting eqs. (1) and (14) in (13), the expression for mutual impedance becomes

$$Z_{21} = -\frac{30 I_{1m} I_{2m}}{I_1(0) I_2(0)} \int_0^{H_2} \left(\frac{-j\, e^{-j\beta r_1}}{r_1} - \frac{j\, e^{-j\beta r_2}}{r_2} + \frac{2j \cos \beta H\, e^{-j\beta r_0}}{r_0} \right) \sin \beta (H_2 - z)\, dz \qquad (11\text{-}15)$$

Equation (15) gives the mutual impedance *referred to the base.* The mutual impedance referred to the loop currents will be given by expression (15) multiplied by the ratio of the product of base currents to the product of loop currents, that is by

$$\frac{I_1(0) I_2(0)}{I_{1m} I_{2m}}$$

Therefore the mutual impedance referred to loop currents will be

$$Z_{21} = -30 \int_0^{H_2} \left(\frac{-j\, e^{-j\beta r_1}}{r_1} - \frac{j\, e^{-j\beta r_2}}{r_2} + \frac{2j \cos \beta H\, e^{-j\beta r_0}}{r_0} \right) \sin \beta (H_2 - z)\, dz \qquad (11\text{-}16)$$

It is usually this mutual impedance referred to the current loops that is plotted and shown in curves. When this impedance is known, the mutual impedance referred to the base or terminal currents can be easily calculated.

In Fig. 11-5 are shown two monopole antennas of height H mounted on a perfect reflecting plane and spaced apart a distance d. For this case

$$\left. \begin{aligned} r_0 &= \sqrt{d^2 + z^2} \\ r_1 &= \sqrt{d^2 + (H - z)^2} \\ r_2 &= \sqrt{d^2 + (H + z)^2} \end{aligned} \right\} \qquad (11\text{-}17)$$

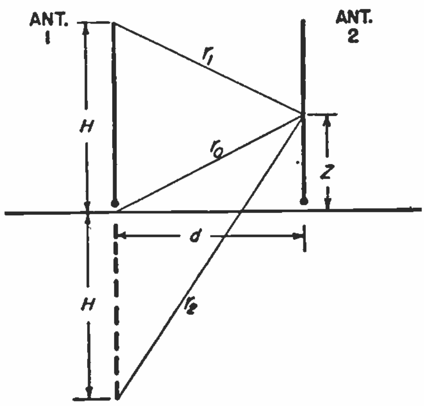


Fig. 11-5

Expression (16) for mutual impedance is complex. The real part gives the mutual resistance, and the imaginary part gives the mutual reactance. Substituting the relations (17) in the real part of (16) gives an expression for mutual resistance.

$$R_{21} = 30 \int_0^H \sin \beta(H - z) \left[\frac{\sin \beta \sqrt{d^2 + (H - z)^2}}{\sqrt{d^2 + (H - z)^2}} + \frac{\sin \beta \sqrt{d^2 + (H + z)^2}}{\sqrt{d^2 + (H + z)^2}} - 2 \cos \beta H \frac{\sin \beta \sqrt{d^2 + z^2}}{\sqrt{d^2 + z^2}} \right] dz \quad (11\text{-}18)$$

Similarly the imaginary part of (16) yields the expression for mutual reactance

$$X_{21} = 30 \int_0^H \sin \beta(H - z) \left[\frac{\cos \beta \sqrt{d^2 + (H - z)^2}}{\sqrt{d^2 + (H - z)^2}} + \frac{\cos \beta \sqrt{d^2 + (H + z)^2}}{\sqrt{d^2 + (H + z)^2}} - \frac{2 \cos \beta H \cos \beta \sqrt{d^2 + z^2}}{\sqrt{d^2 + z^2}} \right] dz \quad (11\text{-}19)$$

The integrations indicated in (18) and (19) can be carried out in a manner similar to that employed in evaluating eq. (3). The case of mutual resistance between quarter-wave monopoles is not too difficult and is left as an exercise. The result for quarter-wave monopoles spaced at distance d is

$$R_{12} = 15[2 \text{ Ci } \beta d - \text{Ci } (\sqrt{(\beta d)^2 + \pi^2} - \pi) - \text{Ci } (\sqrt{(\beta d)^2 + \pi^2} + \pi)] \quad (11\text{-}20)$$

The expression for mutual reactance between quarter-wave monopoles is

$$X_{12} = 15[\text{Si } (\sqrt{(\beta d)^2 + \pi^2} - \pi) + \text{Si } (\sqrt{(\beta d)^2 + \pi^2} + \pi) - 2 \text{ Si } (\beta d)] \quad (11\text{-}21)$$

The general expressions for the mutual impedance between antennas of (equal) height H and a distance d apart are*

$$R_{12} = 30[\sin \beta H \cos \beta H (\text{Si } u_2 - \text{Si } v_2 - 2 \text{ Si } v_1 + 2 \text{ Si } u_1) - \frac{\cos 2\beta H}{2} (2 \text{ Ci } u_1 - 2 \text{ Ci } u_0 + 2 \text{ Ci } v_1 - \text{Ci } u_2 - \text{Ci } v_2) - (\text{Ci } u_1 - 2 \text{ Ci } u_0 + \text{Ci } v_1)] \quad (11\text{-}22)$$

* P. S. Carter, "Circuit Relations in Radiating Systems and Applications to Antenna Problems," *Proc. IRE*, **20,** 1004 (1932); J. Labus, "Mathematical Calculation of the Impedance of Antennas," *Hochfrequenz. Technik*, **41,** 17 (1933); G. H. Brown and R. King, "High Frequency Models in Antenna Investigations," *Proc. IRE*, **22,** 457 (1934).

$$X_{12} = -30\left[\sin \beta H \cos \beta H(2 \text{ Ci } v_1 - 2 \text{ Ci } u_1 + \text{Ci } v_2 - \text{Ci } u_2) - \frac{\cos 2\beta H}{2}(2 \text{ Si } u_1 - 2 \text{ Si } u_0 + 2 \text{ Si } v_1 - \text{Si } u_2 - \text{Si } v_2) - (\text{Si } u_1 - 2 \text{ Si } u_0 + \text{Si } v_1)\right] \quad (11\text{-}23)$$

where $u_0 = \beta d$

$$u_1 = \beta(\sqrt{d^2 + H^2} - H) \qquad u_2 = \beta(\sqrt{d^2 + (2H)^2} + 2H$$
$$v_1 = \beta(\sqrt{d^2 + H^2} + H) \qquad v_2 = \beta(\sqrt{d^2 + (2H)^2} - 2H)$$

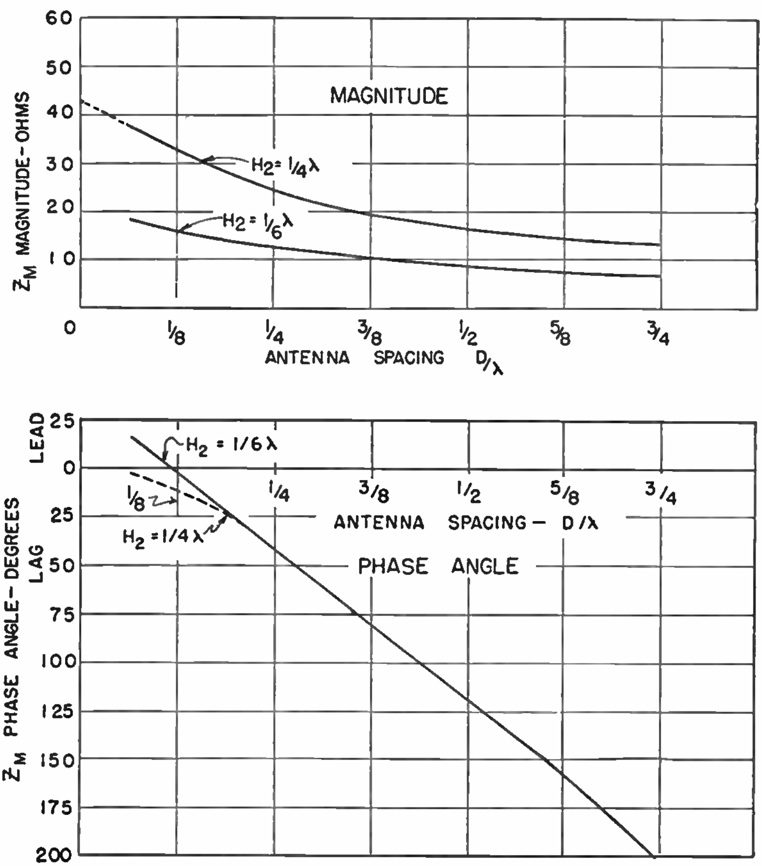


FIG. 11-6. The mutual impedance (referred to the current loops) between a quarter-wave monopole, H_1, and a monopole antenna of height H_2.

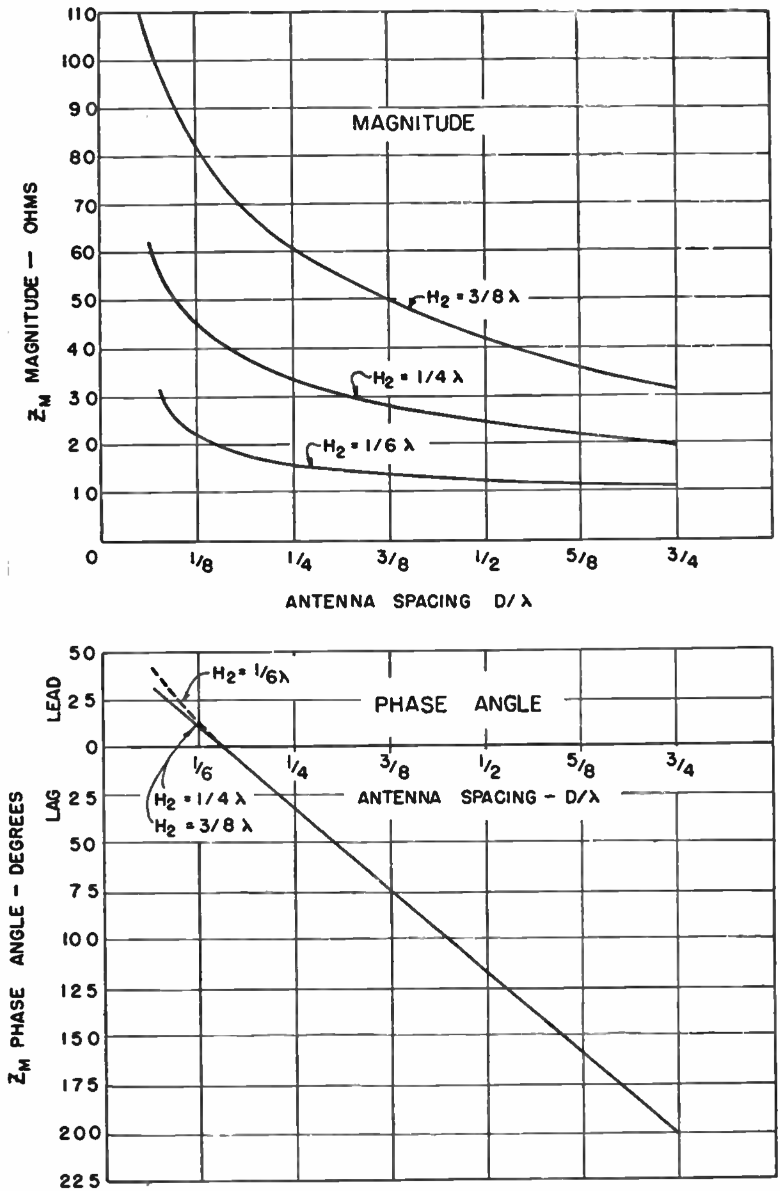


FIG. 11-7. The mutual impedance (referred to the current loops) between a three-eighth-wave monopole, H_1, and a monopole antenna of height H_2.

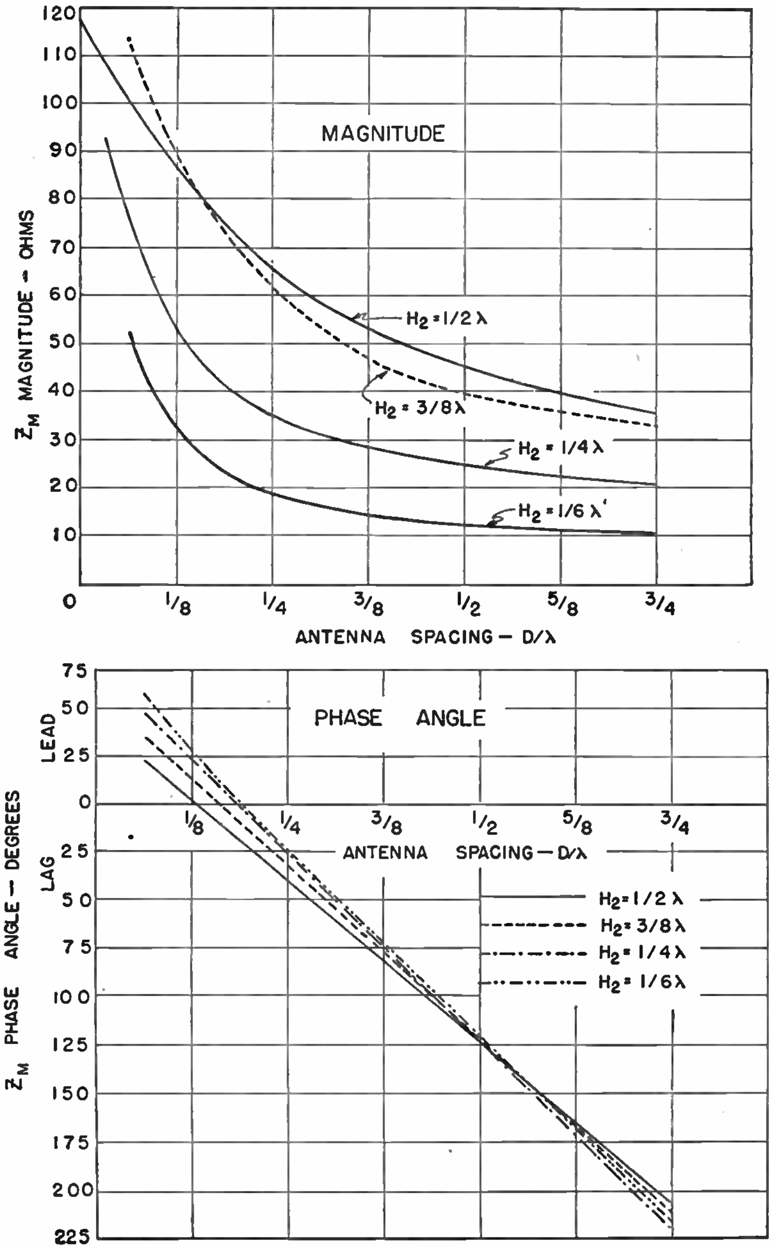


FIG. 11-8. The mutual impedance (referred to the current loops) between a half-wave monopole, H_1, and a monopole antenna of height H_2.

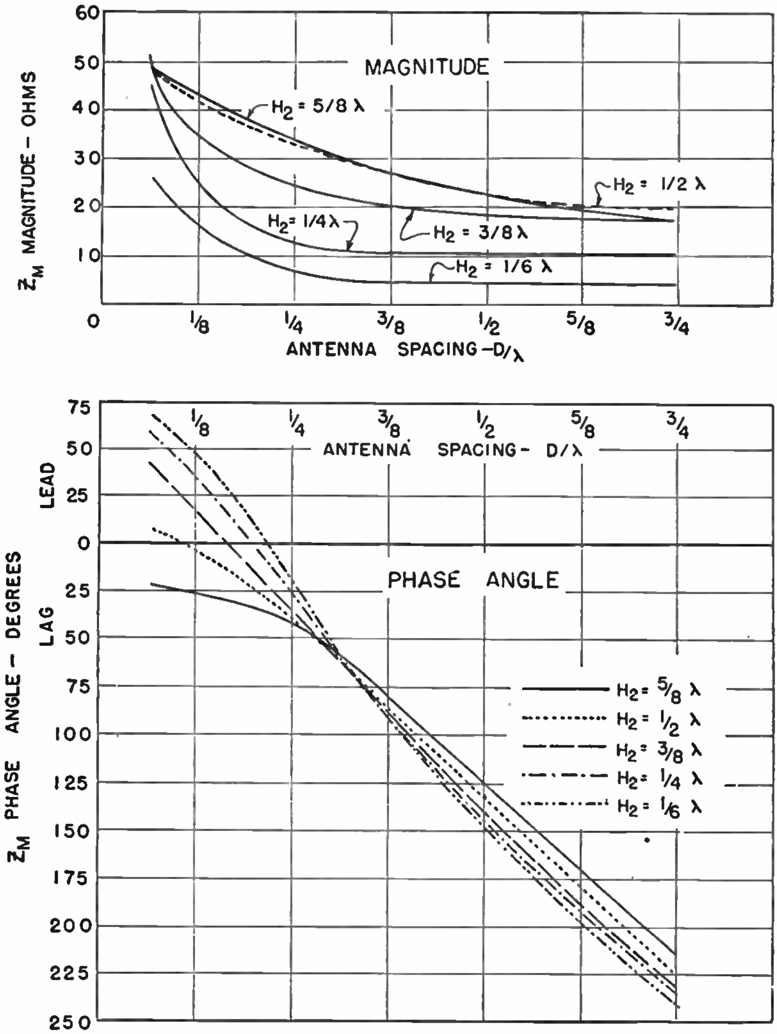


FIG. 11-9. The mutual impedance (referred to the current loops) between a five-eighth-wave monopole, H_1, and a monopole antenna of height H_2.

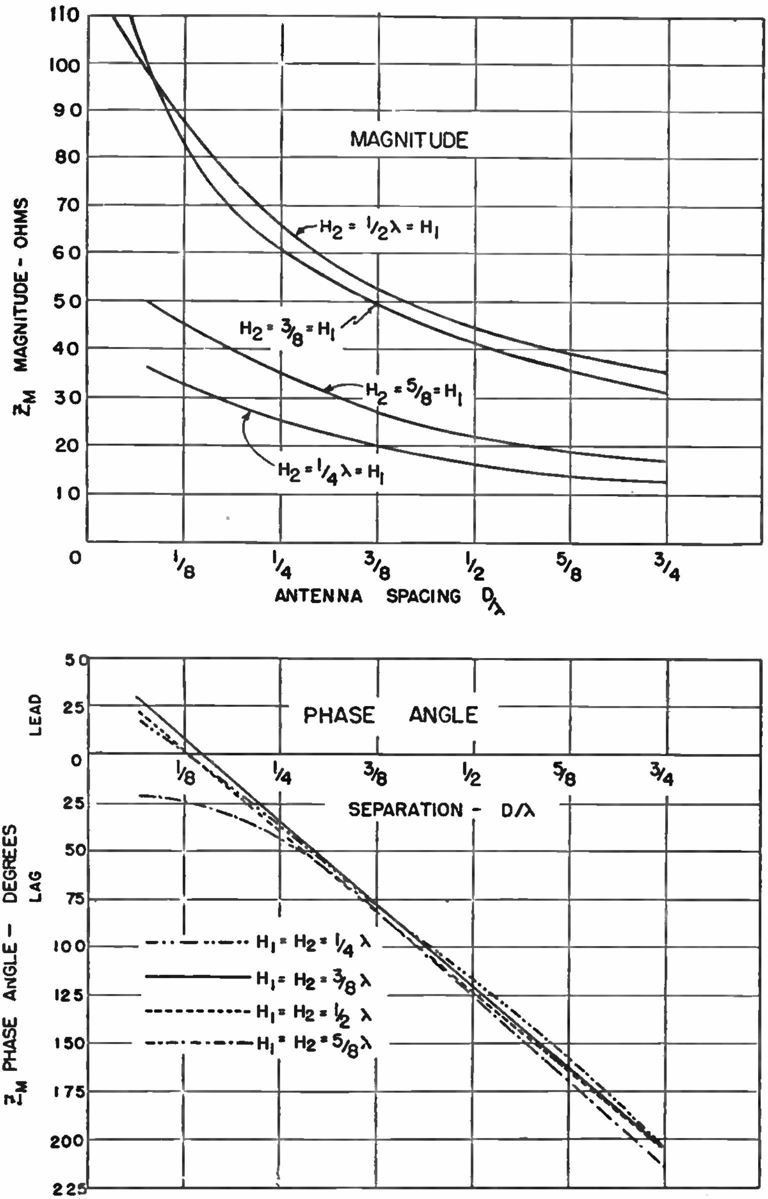


FIG. 11-10. The mutual impedance (referred to the current loops) between monopole antennas of equal height.

Since the evaluation of these expressions is a laborious task, graphical and other methods* of evaluating them have been employed. Curves for the mutual impedance between antennas of unequal as well as equal heights are given in Figs. 11-6 through 11-10.

Problem 1a. Derive the expression (11-20) for the mutual resistance between two quarter-wave monopoles. (NOTE: Use of the substitutions $x = H - z$ and $u = \beta(\sqrt{d^2 + x^2} - x)$, $v = \beta(\sqrt{d^2 + x^2} + x)$ in the appropriate places will aid in the derivation).

Problem 1b. Derive the expression (11-21) for the mutual reactance between two quarter-wave antennas.

Problem 2. Evaluate the expressions for R_{12} and X_{12} to obtain the mutual impedance between two quarter-wave antennas at half-wavelength separation. **Ans.:** $Z_{12} = -6.3 - j14.9$

Problem 3. (a) Derive the expressions for the mutual resistance and reactance between two half-wave vertical antennas at the surface of a perfect earth; (b) Evaluate these expressions to obtain the magnitude and phase angle of the mutual impedance when the antennas are spaced one half-wavelength apart.

Problem 4. Evaluate eq. (11-5) to obtain the radiation resistance of a ⅜-wave monopole antenna. **Ans.:** 93 ohms (referred to loop current)

Problem 5. A broadcast antenna array consists of two quarter-wave vertical towers spaced a quarter-wavelength apart. The antennas are to be fed with equal currents which differ by 90 degrees in phase. The radiated power is 5000 watts. Determine the magnitude and relative phases of the driving-point voltages.

Assume
$$Z_{11} = Z_{22} = 37 + j21 \text{ ohms}$$
$$Z_{12} = 21 - j14 \text{ ohms.}$$

Problem 6. A broadcast array is to consist of 3 quarter-wave radiators in a line spaced ¼ wavelength apart. They are to be fed with equal currents of such phases that

$$I_1 = I_2\underline{/-120^\circ} \qquad I_3 = I_2\underline{/+120^\circ}$$

where antenna 2 is the center element. Assuming that $Z_{11} = Z_{22} = Z_{33} = 37 + j21$, and using mutual-impedance values picked from curves as being $Z_{12} = Z_{23} = 20 - j16$, $Z_{13} = -6 - j15$, determine the input impedances of each of the antennas, and the required magnitude and phases of the voltages at the feed points.

* G. H. Brown and R. King, *loc. cit.;* E. C. Jordan and W. L. Everitt, "Acoustic Models of Radio Antennas," *Proc. IRE*, **29**, 4, 186 (1941).

Problem 7. An array consisting of three vertical towers in a line is to be fed with currents

$$I_3 = I_1 = 1.5I_2\underline{/-114°}$$

Bridge measurements on the towers (of identical dimensions) give $Z_{11} = 20.2 - j11$, $Z_{22} = 19.5 - j11$, $Z_{33} = 20.2 - j21.2$. The impedance of No. 1, measured with No. 2 open-circuited, but No. 3 *resonated* to ground ($X_{33} = 0$) is $19.0 - j13$. Similarly, the impedance of No. 2, with No. 3 open but No. 1 resonated to ground, measured $23.5 - j4.2$ ohms. The impedance of No. 2, with No. 1 open but No. 3 resonated to ground, measured $23.1 - j4.0$ ohms. From this set of measurements determine probable values of mutual impedances. Calculate the input or driving point impedances when the array is in operation, and the power fed to each element for a total of 1000 watts radiated.

11.05 Reactance of an Antenna. By definition the mutual impedance between two antennas is a measure of the voltage induced in the second antenna for 1 amp of current flow in the first. As two antennas of equal height are brought closer together until they coincide, the voltage induced in the second antenna becomes equal to the back or self-induced voltage against which the current in the first antenna must be driven. Therefore it would be expected that the mutual impedance between two antennas of equal length would approach the self-impedance of one of them as the antenna spacing approaches zero. This is indeed the case, and it will be found, if antenna spacing d is put equal to zero, that the formula for mutual resistance between equal-length antennas reduces to the expression (11-5) for the radiation resistance of a single antenna. In the case of the mutual-reactance formula it is found that, except for the special case of H equal to an odd multiple of $\lambda/4$, the mutual reactance becomes infinite as the antenna spacing approaches zero. This gives a value of infinity for the self-reactance of an antenna. The answer is correct for the conditions stated, namely for the sinusoidally distributed current in a *filamentary-* or zero-diameter antenna. (It will be shown in section 11.08 that the inductive reactance of a wire becomes infinitely large as the wire diameter approaches zero.) It is evident that in computing the *reactance* of an antenna, its finite diameter will have to be considered. The reason why the approximation of an infinitely thin antenna is valid for computing radiation resistance but not reactance, becomes apparent when it is recalled that the radiated power, which determines the value of radiation resistance, depends only on the distant fields. On the other hand, the reactive power, and hence the reactance of the antenna, depends upon the induction and electrostatic fields close to the antenna. The strength of these near fields depends to a marked degree upon the shape and thickness of the antenna, whereas these same factors affect the distant field only slightly.

The computation of the reactance of a finite-diameter antenna can be accomplished by using the induced-emf method in a manner similar to that employed for computing antenna resistance. The electric intensity E_z due to the assumed sinusoidal current distribution is evaluated at the surface of the antenna. Then the magnitude of E_z at each point is multiplied by the current in the element of area at that point (and by the sine of the phase angle between them) to give the reactive power associated with the element. This product is then integrated over the surface of the antenna to obtain the total reactive power. Dividing the reactive power by the square of the current at the reference point gives the reactance of the antenna (referred to the current at the reference point).

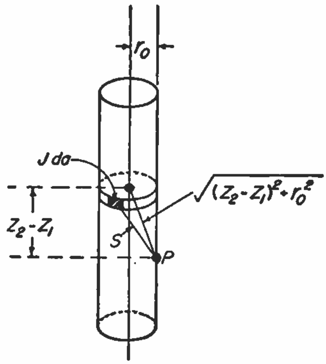


Fig. 11-11

Figure 11-11 shows a length of a cylindrical antenna of radius r_0, which is assumed to be carrying a sinusoidal distribution of current along its length. The current is uniformly distributed around the circumference of the cylinder, the major portion of it flowing in a very small thickness of conductor adjacent to the outer surface. For purposes of making an approximate computation it may be assumed that the electric field at the surface, calculated from the actual current distributed around the cylinder, would be the same as that which can be calculated by considering the current to be concentrated along a filament at the axis of the cylinder. Actually, the "average" distance S from a point P on the surface to a current element Jda on a typical ring located a distance $Z_2 - Z_1$ from P is somewhat greater than the distance of P from the corresponding point on the axis given by $\sqrt{(Z_2 - Z_1)^2 + r_0^2}$. It has been shown* that if an effective radius a, equal to $\sqrt{2}\, r_0$, is used as the radius of the cylindrical surface on which E is calculated, the results of this simplified analysis will be very close to those obtained by a rigorous analysis. In view of the fact that the current distribution along the length of the cylinder is not known, and a sinusoidal current has been assumed, a more rigorous analysis for this case would hardly seem to be justified.

Now referring to Fig. (11-11), both E_z and J are constant around the circumference of the antenna at a given height, say z_2. Since the linear current density J, integrated around the circumference, is just the total antenna current I at that cross section, the reactive power associated with a ring of height dz is

$$-\tfrac{1}{2}|E_z||I(z)| \sin \psi \, dz$$

The total reactive power for the whole antenna is

$$-\tfrac{1}{2}\int_0^H |E_z||I(z)| \sin \psi \, dz$$

and the antenna reactance referred to the loop current is

$$-\frac{1}{I_m^2}\int_0^H |E_z||I(z)| \sin \psi \, dz \qquad (11\text{-}24)$$

In this expression $I(z)$ is the (assumed sinusoidal) antenna current and E_z is the electric intensity due to such a current at a distance from the current of $a = \sqrt{2}\, r_0$. Now except for the point of reference (24) is equivalent to the imaginary part of (13), which gives the *mutual* reactance between two filamentary antennas. Thus within the limits of the assumptions and approximations made, the reactance of an antenna of finite radius r_0 is equal to the mutual reactance between two filamentary antennas of the same length spaced $\sqrt{2}\, r_0$ apart. Substituting $a = \sqrt{2}\, r_0$ for d in expression (23) for mutual reactance gives for the reactance of a monopole antenna of height H and radius $r_0 = a/\sqrt{2}$.

$$\begin{aligned} X = -30[&\sin \beta H \cos \beta H (2 \text{ Ci } v_1 - 2 \text{ Ci } u_1 + \text{Ci } v_2 - \text{Ci } u_2) \\ &- \frac{\cos 2\beta H}{2} (2 \text{ Si } u_1 - 2 \text{ Si } u_0 + 2 \text{ Si } v_1 - \text{Si } u_2 - \text{Si } v_2) \\ &- (\text{Si } u_1 - 2 \text{ Si } u_0 + \text{Si } v_1)] \end{aligned} \qquad (11\text{-}25)$$

* O. Zinke, "Fundamentals of Voltage and Current Distributions along Antennas," *Arch. Elektrotech.*, **35**, 67–84 (1941).

where $u_0 = \beta a$

$$u_1 = \beta(\sqrt{H^2 + a^2} - H) \qquad u_2 = \beta(\sqrt{(2H)^2 + z^2} + 2H)$$
$$v_1 = \beta(\sqrt{H^2 + a^2} + H) \qquad v_2 = \beta(\sqrt{(2H)^2 + a^2} - 2H)$$

The radius of the antenna will normally be a very small fraction of a wavelength so that

$$\beta a \ll 1$$

and the following approximations may be used.

$$\text{Si}\ (\beta a) = \beta a - \frac{(\beta a)^3}{3.3!} + - \cdots \approx \beta a \approx 0$$
$$\text{Ci}\ \beta(\sqrt{H^2 + a^2} - H) \approx \text{Ci}\ \beta H\left(1 + \frac{a^2}{2H^2} - 1\right) \approx \text{Ci}\left(\frac{\beta a^2}{2H}\right)$$
$$\text{Ci}\ \beta(\sqrt{(2H)^2 + a^2} - 2H) \approx \text{Ci}\left(\frac{\beta a^2}{4H}\right)$$
$$\text{Si}\ u_1 \approx 0 \qquad \text{Si}\ v_1 \approx \text{Si}\ 2\beta H$$
$$\text{Si}\ v_2 \approx 0 \qquad \text{Si}\ u_2 \approx \text{Si}\ (4\beta H)$$
$$\text{Ci}\ v_1 \approx \text{Ci}\ (2\beta H)$$
$$\text{Ci}\ u_2 \approx \text{Ci}\ (4\beta H)$$

Using these approximations, expression (25) becomes

$$X = -30\left\{\sin \beta H \cos \beta H \left[2\ \text{Ci}\ 2\beta H - 2\ \text{Ci}\left(\frac{\beta a^2}{2H}\right) + \text{Ci}\left(\frac{\beta a^2}{4H}\right) - \text{Ci}\ 4\beta H\right] - \frac{\cos 2\beta H}{2}[2\ \text{Si}\ 2\beta H - \text{Si}\ 4\beta H] - \text{Si}\ 2\beta H\right\}$$

Now when x is very small, Ci $x \approx \gamma + \ln x$ where $\gamma = 0.5772 \cdots$ is Euler's constant. Using this substitution, the second and third terms of the above expression may be combined

$$-2\ \text{Ci}\left(\frac{\beta a^2}{2H}\right) + \text{Ci}\left(\frac{\beta a^2}{4H}\right) = -\gamma + \ln\left(\frac{H}{\beta a^2}\right)$$

so that the final expression for the reactance of a monopole antenna of radius $r_0 = a/\sqrt{2}$ and length H becomes

$$X = -15 \left\{ \sin 2\beta H \left[-\gamma + \ln \left(\frac{H}{\beta a^2} \right) + 2 \text{ Ci } (2\beta H) - \text{Ci } (4\beta H) \right] \right.$$

$$\left. - \cos 2\beta H [2 \text{ Si } (2\beta H) - \text{Si } (4\beta H)] - 2 \text{ Si } (2\beta H) \right\} \quad (11\text{-}26)$$

For the particular case of a quarter-wave antenna, $\sin 2\beta H = 0$, $\cos 2\beta H = -1$, and the expression for reactance reduces to

$$X = 15 \text{ Si } (4\beta H)$$
$$= 15 \text{ Si } (2\pi) = 21.25 \text{ ohms}$$

Expression (26) gives the reactance (referred to the current loop) of a monopole antenna of length H and radius $r_0 = a/\sqrt{2}$, as given by the induced-emf method, using the sinusoidal current distribution assumption. The reactance referred to the base can be obtained from (26) by dividing by $\sin^2 \beta H$. Figure 11-12 shows resistance and reactance values computed by this method for short monopole antennas of different thicknesses. The resistance or reactance of the corresponding dipole antennas of length $L = 2H$ is just double that of the monopole antenna of length H.

It is seen that under the assumed conditions of sinusoidal current distribution, a quarter-wavelength antenna has a positive reactance of 21.25 ohms, and this value of reactance is independent of antenna diameter as long as the latter is small in wavelengths. For lengths other than multiples of the quarter-wavelength, the reactance depends very greatly on the antenna diameter, being very large for thin antennas. This fact indicates the desirability of using fat antennas for broad-band applications such as television, where a low ratio of antenna reactance to resistance (low Q) is required.

It will also be seen that, as the antenna length is varied, the reactance goes through zero for some length shorter than a quarter-wavelength. This means that the "resonant" length is always somewhat less than a quarter wavelength, being shorter for fat antennas.

Problem 8. Verify that as the spacing d approaches zero, the expression (11-22) for mutual resistance between two antennas of equal height reduces to the expression for the radiation resistance of a single antenna.

11.06 Note on the Induced-emf Method. The power radiated from an antenna has been calculated by two methods. The first of these methods, called the Poynting vector method, was covered in chap. 10. The second, known as the induced-emf method, was the subject of this chapter. Because of certain questions that

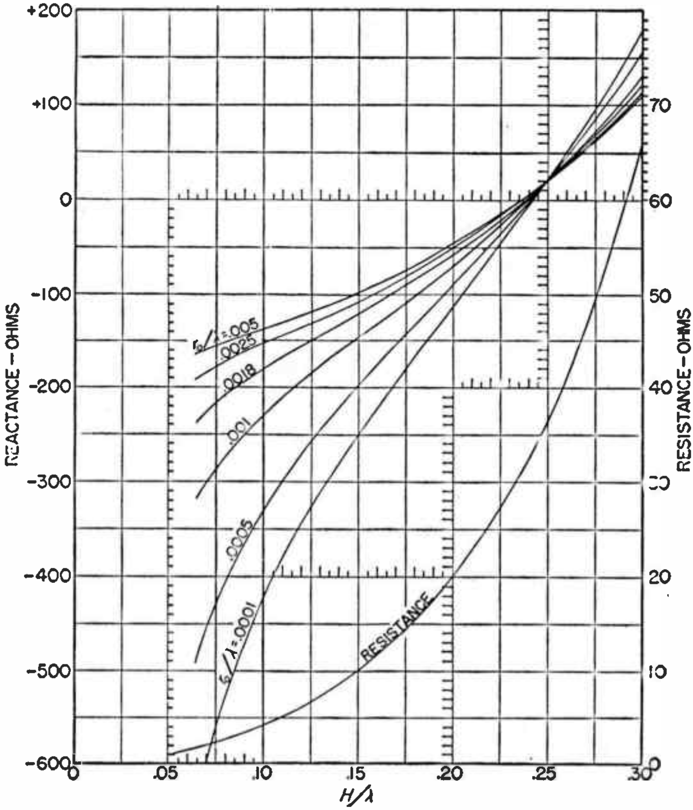


FIG. 11-12. The resistance and reactance (referred to base) of short monopole antennas as computed by the induced-emf method.

inevitably are raised concerning this latter method, it is desirable to compare the two methods in some detail.*

(*a*) *Poynting Vector Method.* A certain current distribution is assumed to exist along the antenna. The electric and magnetic field intensities due to the assumed current distributions are com-

* The clarification of the induced-emf method is due to R. E. Burgess, "Aerial Characteristics," *Wireless Engineer*, **21**, 247, 154 (1944).

puted at a point P on some surface enclosing the antenna. The net outward flow of power through this surface is obtained by integrating the Poynting vector $\mathbf{E} \times \mathbf{H}$ over the entire surface and over a cycle in time. In practice, the enclosing surface is usually chosen to be a sphere of very large radius, under which condition the difference in distance to various points on the antenna affects only the phase, and not the magnitude, of the contributions to the total field and hence the computation is correspondingly simplified. A sinusoidal current distribution is usually assumed, and the method is in error only by the amount that the *radiation* fields, produced by the actual current distribution, differ from the radiation fields calculated from this assumed sinusoidal distribution. Inasmuch as the actual current distribution is known to be very nearly sinusoidal for thin transmitting antennas, the answer obtained is a good approximation to the true power radiated.

The calculation is usually made assuming a filamentary current, but the results hold for finite diameter antennas as long as the diameter is very small compared with the length and compared with a wavelength.

(*b*) *Induced-emf Method.* In the second method a filamentary current distribution is assumed as before and the electric and magnetic intensities resulting therefrom computed. However, in this case the point P_1, at which the fields are computed, is taken right at the filament. Each current element $I\,dl$ is multiplied by the component E_z of the electric field parallel to it at that point, to obtain the power required to drive the current against the electric field. The real part of the total power, obtained by integrating $|E_z||I|\,dl \cos \psi$ over the length of the antenna, represents the total power radiated (ohmic losses assumed negligible). ψ is the time phase angle between E_z and I at the point in question. This method gives exactly the same value for power radiated as the previous Poynting vector method. This is as it should be because, as will be shown, this method can be derived directly from the Poynting vector method. The approximation involved in this method is the same as in the Poynting vector method and is that of assuming a sinusoidal current distribution, whereas the actual current distribution is only approximately sinusoidal.

Although the induced-emf method is essentially the same as the Poynting vector method and gives exactly the same results, its validity is sometimes questioned when it is applied to an actual antenna having a finite diameter. The reason for this is as follows: If the antenna is assumed to be a perfect conductor (the usual assumption), the boundary conditions require that the total electric intensity **E** along the surface of the antenna be zero. In the case of a transmitting antenna, excited by a lumped voltage across a gap, the only electric field existing along the surface of the antenna is the field E_z, induced by the currents and charges along the antenna. The boundary conditions require that this electric intensity be zero everywhere on the surface, and therefore the product $|E_z||I|\,dl \cos\psi$ is zero at every point along the antenna. Then $|E_z||I|\,dl \cos\psi$ integrated along the antenna is zero and the power radiated from *the conducting part of the antenna* is zero. This also is as it should be, because the conductor contains in itself no source of electromagnetic energy, the energy coming from the generator. However, there are two questions raised that require clarification.

1. Since the actual $|E_z||I|\,dl \cos\psi$ that exists along the surface is zero and, therefore, not even approximately the same as $|E_z||I|\,dl \cos\psi$, computed from the assumed sinusoidal distribution, is there any justification for expecting that the value given by the computed $\int|E_z||I|\,dl \cos\psi$ is even approximately correct?
2. Since the actual $\int|E_z||I|\,dl \cos\psi$ over the surface of the conductor is zero, an incidental question is "from where is the power radiated?"

The answer to the first question regarding the validity of the method can be obtained readily by considering initially a receiving antenna of resonant length that has the load terminals *a-b* short-circuited and which, therefore, reradiates all the received energy. Assume first that the current flowing in the antenna owing to the received electric field has a true sinusoidal distribution. The self-induced electric intensity or "back voltage" due to this current flow (and the corresponding charge distribution) can be calculated in the usual manner and will be designated by E_s. (The subscript *s* indicates that this is the electric intensity computed from the assumed sinusoidal distribution.) Then, *if* the received or applied tangential field—which will be designated by E'—were exactly equal and opposite to E_s at all points along the surface of the antenna, the assumed current would flow in the antenna. The boundary conditions at the surface of the antenna would be satisfied because the total electric intensity parallel to it would be $E = E' + E_s = 0$. The power reradiated by the antenna is obtained by integrating the Poynting vector over the surface of the antenna. As pointed out above and in the next section, this is equal to $|E_s||I|\, dl \cos \psi$ integrated along the length of the antenna. Similarly, the power per unit length flowing into the antenna from the received field is $-|E'||I|\, dl \cos \psi$ (outward flow of energy is assumed positive). The net flow of power out of the antenna, which is the difference between these two, is equal to zero.

Next consider the same short-circuited receiving antenna under conditions where the received field E' does not have the particular configuration required in the above case, but instead has some arbitrary value along the length of the antenna. In particular, consider the case where E' is *uniform*, as it would be for reception of a plane wave at $\theta = 90$ degrees. Then the current distribution will *not* be sinusoidal, and the actual current distribution will be such as to produce a self-induced field E'' along the antenna, such that $E' + E'' = 0$. That is, E'' will be uniform or constant along the antenna and will have a value $E'' = -E'$. Now, although the actual current distribution cannot be sinusoidal, it is known to be very closely sinusoidal for the resonant length. Evidently then, it requires but a very small change in current distribution from the sinusoidal to change the self-induced parallel component of E from that calculated for the sinusoidal current cases, E_s, to the value $E'' = -E'$ that must exist in the actual case. Since the current distribution is but little changed from the sinusoidal, the power radiated for a given loop current must be very nearly equal to the case for the true sinusoidal distribution. (Small changes in current produce only small changes in the *radiation* terms of the electric field). The actual power reradiated in this case is $\int |E''||I_a|\, dl \cos \psi$, where I_a is the actual current and E'' is the self-induced parallel component of electric field due to it. But from the previous statement this must be very nearly equal to the power reradiated in the sinusoidal case, which is $\int |E_s||I_s|\, dl \cos \psi$. That is

$$\int |E''||I_a|\, dl \cos \psi = \int - |E'||I_a|\, dl \cos \psi \approx \int |E_s||I_s|\, dl \cos \psi$$

This means that, although the actual current is not sinusoidal and the actual self-induced voltage $E'' = -E'$ differs *greatly* from that calculated from a sinusoidal distribution, nevertheless, the radiated power computed from an assumed sinusoidal distribution with its resulting E_s gives an answer that is very close to that which would be obtained from $\int -|E'||I_a|\,dl \cos\psi$ if the actual current I_a were known. However, it should be noted that this is true only because the actual current distribution is nearly sinusoidal.

Finally, consider the case of a transmitting antenna in which the applied electric field is concentrated over a short section at the center.

If V is the applied voltage and S is the separation of the terminals a-b (Fig. 11-13), then the field across a-b can be considered to be $E' = V/S$. The applied field is zero everywhere else along the antenna. The actual current that flows in the antenna as a result of the applied voltage V must be such as to produce a self-induced electric field opposite to the applied field everywhere along the antenna. That is, the self-induced field must be zero everywhere along the antenna except between a and b, where it has a value of $-V/S$. It is an experimental fact that the actual antenna current that flows and necessarily produces the above electric field distribution, is *very closely sinusoidal* for thin transmitting antennas. Therefore, as in the discussion of receiving antennas, the radiated power computed from $\int|E_s||I_s|\,dl \cos\psi$ must be very nearly equal to the actual power radiated. In this case, the actual power radiated is

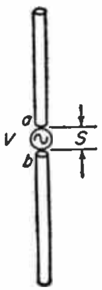


FIG. 11-13

$$\int_0^L -|E'||I_a|\,dl \cos\psi = \int_a^b -|E'||I_a|\,dl \cos\psi = |V||I_0| \cos\theta$$

where I_0 is the current at the feed point and θ is the angle between V and I. Therefore, actual power radiated $= |V||I_0| \cos\theta \approx \int_0^L |E_s||I_s|\,dl \cos\psi$.

It should be noted in passing that this latter integration should be performed over the whole of the antenna including the section between a and b. However, since E_s between a and b is of the same order of magnitude as E_s at adjacent points on the antenna, the error incurred in neglecting the section a-b becomes very small when the gap length is small compared with the length of the antenna. However, the situation is very different in the case of the actual current distribution with the resulting actual distribution of the self-induced field. In this latter case, the integral is zero everywhere, except at the gap or generator. As the gap is made very small, the actual E' across it becomes very large for a given applied voltage V and the gap or generator section cannot be neglected. Indeed, it may be said that all the power flows out from this generator section, being guided into space by the antenna conductors.

11.07 Equivalence of Induced-emf and Poynting Vector Methods. It is easy to show that the induced-emf and Poynting vector methods for computing radiated power are one and the same method when the surface of integration coincides with the surface of the antenna. Consider an antenna of length L and radius R, which has some arbitrary current distribution I_z. The components of $\mathbf{E}$ and $\mathbf{H}$, tangential to the surface along the length of the antenna, are E_z and H_ϕ. At the top and bottom ends the tangential components are E_r and H_ϕ. If the real Poynting vector is integrated over the surface of the antenna, the following result is obtained:

$$Re \int_s \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{a} = \int_0^L |E_z||H_\phi| \cos\psi \, 2\pi R \, dz + 2\int_0^R |E_r||H_\phi| \cos\psi \, 2\pi r \, dr.$$

where ψ is the time phase angle between the tangential components of $\mathbf{E}$ and $\mathbf{H}$.

The first integral covers the entire surface of the antenna except the end caps. The second integral covers these end caps. The quantity $2\pi RH_\phi$ is equal to the line integral of H_ϕ about the antenna and, by Maxwell's first equation, this is equal to the total current flowing through the closed path, so that

$$2\pi RH_\phi = \oint H \, dl = I_z$$

where I_z is the current along the antenna. Using this relation, the first integral becomes

$$\int_0^L |E_z||I_z| \cos\psi \, dz$$

The angle ψ is now the time phase angle between E_z and I_z, because I_z and H_ϕ are in time phase.

In the end caps the current flows radially and must be zero at the center. Denoting by J_r the radial surface current density in the caps, the relation $\mathbf{H} = \mathbf{J} \times \mathbf{n}$ becomes $H_\phi = J_r$ for the top and bottom caps. Then $2\pi r H_\phi = 2\pi r J_r = I_r$ is the total radial current flowing across a circle of radius r on each of the caps. Using this relation the second integral becomes

$$2 \int_0^R |E_r||I_r| \cos \psi \, dr$$

The total surface integral may then be written

$$\text{Re} \int_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{a} = \int_0^L |E_z||I_z| \cos \psi \, dz + 2 \int_0^R |E_r||I_r| \cos \psi \, dr$$

Thus the Poynting vector, when integrated over the surface of the antenna, yields the induced-emf integral.

It is evident that the contribution to the radiated power from the end surfaces of the antenna must be very small, since the current there is very small and in such directions that the various current elements produce radiation fields which cancel one another. Therefore, the second term of the above integral is usually dropped, and the power is obtained from

$$\int_0^L |E_z||I_z| \cos \psi \, dz$$

11.08 Uniform Cylindrical Waves and the Infinitely Long Wire. In foregoing sections the impedance of finite-length antennas have been computed by the induced-emf method, using an assumed sinusoidal current distribution. A simpler problem is that of determining the impedance per unit length of an infinitely long wire, which is assumed to carry a uniform, in-phase current $I\, e^{j\omega t}$. Although this may appear to be a rather unreal situation, it can be approximated in practice by a very long wire that is excited by a parallel electric field of constant value. This particular problem has the definite advantage that its solution is simple enough to permit of easy interpretation. Before solving it a brief discussion of uniform cylindrical waves will be in order.

For a homogeneous medium having the constants μ, ϵ, and σ, Maxwell's equations in cylindrical co-ordinates are

$$\left.\begin{aligned}\frac{\partial H_z}{\rho\,\partial\phi} - \frac{\partial H_\phi}{\partial z} &= (\sigma + j\omega\epsilon)E_\rho & \frac{\partial E_z}{\rho\,\partial\phi} - \frac{\partial E_\phi}{\partial z} &= -j\omega\mu H_\rho \\ \frac{\partial H_\rho}{\partial z} - \frac{\partial H_z}{\partial \rho} &= (\sigma + j\omega\epsilon)E_\phi & \frac{\partial E_\rho}{\partial z} - \frac{\partial E_z}{\partial \rho} &= -j\omega\mu H_\phi \\ \frac{\partial(\rho H_\phi)}{\rho\,\partial\rho} - \frac{\partial H_\rho}{\rho\,\partial\phi} &= (\sigma + j\omega\epsilon)E_z & \frac{\partial(\rho E_\phi)}{\rho\,\partial\rho} - \frac{\partial E_\rho}{\rho\,\partial\phi} &= -j\omega\mu H_z\end{aligned}\right\} \quad (11\text{-}27)$$

For fields that have no variation with ϕ or z, such as would be generated for example by an infinitely long wire carrying a uniform current $I\,e^{j\omega t}$, eqs. (27) reduce to

$$\left.\begin{aligned}\frac{\partial H_z}{\partial\rho} &= -(\sigma + j\omega\epsilon)E_\phi & \frac{\partial E_z}{\partial\rho} &= j\omega\mu H_\phi \\ \frac{\partial(\rho H_\phi)}{\rho\,\partial\rho} &= (\sigma + j\omega\epsilon)E_z & \frac{\partial(\rho E_\phi)}{\rho\,\partial\rho} &= -j\omega\mu H_z\end{aligned}\right\} \quad (11\text{-}28)$$

The waves obtained with these fields are *uniform cylindrical* waves, having no variation of amplitude or phase over any cylindrical surface represented by $\rho = \rho_0$. It is evident that uniform cylindrical waves are transversely electromagnetic, and that they may be divided into two types, viz., (a) those having E_z and H_ϕ components, and (b) those having E_ϕ and H_z components. The former would be generated by the infinitely long wire mentioned, whereas the latter would be produced by an infinitely long line of closely spaced coaxial loops carrying equal and uniform currents that are everywhere in phase.

Considering the first of these types, the two relations

$$\frac{\partial(\rho H_\phi)}{\rho\,\partial\rho} = (\sigma + j\omega\epsilon)E_z \qquad \frac{\partial E_z}{\partial\rho} = j\omega\mu H_\phi \qquad (11\text{-}29)$$

can be combined to yield a wave equation in cylindrical co-ordinates.

$$\frac{\partial^2 E_z}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial E_z}{\partial\rho} - \gamma^2 E_z = 0 \qquad (11\text{-}30)$$

where as usual $\gamma^2 = j\omega\mu(\sigma + j\omega\epsilon)$

For the special case of wave propagation in a nonconducting medium, $\sigma = 0$ and $\gamma^2 = -\omega^2\mu\epsilon$, so that the wave equation becomes

$$\frac{\partial^2 E_z}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial E_z}{\partial\rho} + \beta^2 E_z = 0 \qquad (11\text{-}31)$$

where $\beta = \omega\sqrt{\mu\epsilon}$ is a real number. Dividing through by β^2 in (31) shows it to be an ordinary Bessel equation of order zero with the independent variable $(\beta\rho)$:

$$\frac{\partial^2 E_z}{\partial(\beta\rho)^2} + \frac{1}{(\beta\rho)}\frac{\partial E_z}{\partial(\beta\rho)} + E_z = 0 \tag{11-32}$$

As in sec. (9.05) the general solution may be written in terms of zero-order Bessel functions of the first and second kinds.

$$E_z = AJ_0(\beta\rho) + BN_0(\beta\rho) \tag{11-33}$$

In this form the solution represents standing waves. An alternative solution may be written in terms of linear combinations of J_0 and N_0.

$$E_z = A_1H_0^{(1)}(\beta\rho) + B_1H_0^{(2)}(\beta\rho) \tag{11-34}$$

where

$$\left.\begin{aligned} H_0^{(1)}(\beta\rho) &= J_0(\beta\rho) + jN_0(\beta\rho) \\ H_0^{(2)}(\beta\rho) &= J_0(\beta\rho) - jN_0(\beta\rho) \end{aligned}\right\} \tag{11-35}$$

$H_0^{(1)}$ and $H_0^{(2)}$ are called *Hankel functions* of zero order, first and second kinds, respectively. When appropriately combined with the time factor $e^{j\omega t}$, these functions represent inward- and outward-*traveling* waves respectively. That this is so, is evident from the asymptotic expressions for large values of $(\beta\rho)$. These expressions are:

$$\left.\begin{aligned} H_0^{(1)}(\beta\rho) &\to \sqrt{\frac{2}{\pi\beta\rho}}\, e^{j\left(\beta\rho - \frac{\pi}{4}\right)} \\ H_0^{(2)}(\beta\rho) &\to \sqrt{\frac{2}{\pi\beta\rho}}\, e^{-j\left(\beta\rho - \frac{\pi}{4}\right)} \end{aligned}\right\} \text{for } \beta\rho \to \infty \tag{11-36}$$

which should be compared with the corresponding asymptotic expressions for J_0 and N_0.

$$\left.\begin{aligned} J_0(\beta\rho) &\to \sqrt{\frac{2}{\pi\beta\rho}}\cos\left(\beta\rho - \frac{\pi}{4}\right) \\ \mathrm{N}_0(\beta\rho) &\to \sqrt{\frac{2}{\pi\beta\rho}}\sin\left((\beta\rho - \frac{\pi}{4}\right) \end{aligned}\right\} \text{for } \beta\rho \to \infty \tag{11-37}$$

It is also apparent from (35) that the Hankel functions bear a relation to the Bessel functions similar to the relation between the exponential functions (with imaginary exponents) and the trigonometric (sine and cosine) functions.

For propagation in a conducting medium solutions to eq. (30) will be required. Dividing through by γ^2 in (30) shows it to be a *modified Bessel equation* of order zero in the variable $(\gamma\rho)$

$$\frac{\partial^2 E_z}{\partial(\gamma\rho)^2} + \frac{1}{\gamma\rho}\frac{\partial E_z}{\partial(\gamma\rho)} - E_z = 0 \tag{11-38}$$

Solutions to this modified Bessel equation are called *modified Bessel functions* and are denoted by $I_0(\gamma\rho)$ and $K_0(\gamma\rho)$ (for the zero order). Expressions for the I and K functions are given in the appendix. For *small* values of $(\gamma\rho)$,

$$\left.\begin{aligned} I_0(\gamma\rho) &\to 1 \\ K_0(\gamma\rho) &\to -[\ln(\gamma\rho) + C - \ln 2] \end{aligned}\right\} \text{ for } \gamma\rho \to 0 \tag{11-39}$$

Since eq. (30) reduces to (31) when γ is a pure imaginary, it is not surprising to find that the modified and ordinary Bessel functions are related to each other. The relations are

$$\begin{aligned} I_0(jz) &= J_0(z) \\ K_0(jz) &= \frac{\pi}{2j}[J_0(z) - jN_0(z)] \\ &= -\frac{\pi}{2}[N_0(z) + jJ_0(z)] \end{aligned} \tag{11-40}$$

The modified functions I and K are most suitable for propagation in a dissipative medium. For a lossless medium, for which γ is a pure imaginary, the corresponding Bessel or Hankel functions are usually more convenient.

Field about an Infinitely Long Wire. Consider now the electromagnetic field about a long wire carrying a current $I\,e^{j\omega t}$. In the region external to the wire the Hankel function solutions of eq. (31) will be appropriate, and the expression for **E** can be written as

$$E_z = A_1 H_0^{(1)}(\beta\rho) + B_1 H_0^{(2)}(\beta\rho) \tag{11-41}$$

Only the zero-order functions appear, because there is no variation of the field in the ϕ direction. If the region is assumed to extend to infinity, there is no reason for retaining the first term of (41), which represents an inward-traveling wave, and so the solution is given by

$$E_z = B_1 H_0^{(2)}(\beta\rho)$$

which represents an outward-traveling wave. Using eq. (29), the expression for magnetic intensity will be

$$H_\phi = \frac{1}{j\omega\mu}\frac{\partial E_z}{\partial \rho} = -\frac{\beta B_1}{j\omega\mu} H_1{}^{(2)}(\beta\rho) = -\frac{B_1}{j\eta} H_1{}^{(2)}(\beta\rho)$$

since $$\frac{d}{du}[H_0{}^{(2)}(u)] = -H_1{}^{(2)}(u)$$

At $$\rho = a, \qquad H_\phi = \frac{I}{2\pi a}$$

Therefore, $$B_1 = -\frac{j\eta I}{2\pi a H_1{}^{(2)}(\beta a)}$$

$$\approx -\frac{\beta\eta I}{4} \qquad \text{(for } \beta a \ll 1\text{)}$$

Then

$$\left.\begin{aligned} E_z &= -\frac{\beta\eta I}{4} H_0{}^{(2)}(\beta\rho) \xrightarrow[\rho\to\infty]{} -\frac{\eta I}{2\sqrt{\rho\lambda}} e^{-j\left(\beta\rho - \frac{\pi}{4}\right)} \\ H_\phi &= -\frac{j\beta I}{4} H_1{}^{(2)}(\beta\rho) \xrightarrow[\rho\to\infty]{} \frac{I}{2\sqrt{\rho\lambda}} e^{-j\left(\beta\rho - \frac{\pi}{4}\right)} \end{aligned}\right\} \qquad (11\text{-}42)$$

At large distances from the wire the fields decrease in amplitude as $1/\sqrt{\rho}$. Also at large distances the fields are periodic in 2π radians (this is not true close to the source) and appear to have originated at an "effective" source, which is one-eighth of a wavelength out from the center of the wire.

The outward radial impedance is

$$Z_\rho{}^+ = -\frac{E_z}{H_\phi} = j\eta \frac{H_0{}^{(2)}(\beta\rho)}{H_1{}^{(2)}(\beta\rho)} \qquad (11\text{-}43)$$

at large distances, where the asymptotic expressions for the Hankel functions can be used, the radial impedance becomes a pure resistance

$$Z_\rho{}^+ \approx \eta = 377 \text{ ohms} \qquad \text{(for } \beta\rho \gg 1\text{)}$$

The *impedance of the wire* can be obtained from a consideration of the field intensities at its surface. Assuming first a perfectly conducting wire of radius a, the total tangential electric intensity at its surface, $E(a)$, must be zero. Then

$$E(a) = E_a + E_z(a) = 0$$

or $$E_a = -E_z(a)$$

where E_a is the applied electric intensity that causes the current I to flow, and $E_z(a)$ is the self-induced electric intensity (due to the current I) evaluated at the surface of the wire, $P = a$. (In this problem the "applied" field E_a might be the incident field from a distant transmitter). Then the *external impedance* of the wire per unit length will be

$$Z_{ext} = \frac{E_a}{I} = -\frac{E_z(a)}{2\pi a H_\phi(a)}$$

Therefore,

$$Z_{ext} = \frac{j\eta}{2\pi a}\frac{H_0^{(2)}(\beta a)}{H_1^{(2)}(\beta a)} \tag{11-44}$$

For $\beta a \ll 1$, as would normally be the case, (44) reduces to

$$Z_{ext} = \frac{j\eta}{2\pi a}\left[\frac{J_0(\beta a) - jN_0(\beta a)}{J_1(\beta a) - jN_1(\beta a)}\right]$$

$$\approx \frac{j\eta}{2\pi a}\left[\frac{1 - j\frac{2}{\pi}(\ln \beta a + C - \ln 2)}{\frac{\beta a}{2} + j\frac{2}{\pi\beta a}}\right]$$

$$\approx \frac{60\pi^2}{\lambda} + \frac{j\eta}{\lambda}\ln\frac{\lambda}{2\pi a} \tag{11-45}$$

$$= \pi\omega \times 10^{-7} + j\omega\left(\frac{\mu}{2\pi}\ln\frac{\lambda}{2\pi a}\right) \tag{11-45a}$$

The real part of this external impedance is the radiation resistance per unit length, and the imaginary part is the external inductive reactance per unit length of the wire. The former is independent of wire diameter, whereas the latter becomes logarithmically infinite as the wire diameter approaches zero. The quantity $\mu/2\pi \ln \lambda/2\pi a$ is the *high-frequency external inductance* of the wire (per meter length).

If the assumption of perfect conductivity is not made, the total tangential intensity $E(a)$ at the surface will not be zero, but it will have the (small) value required to drive the current I against the *internal impedance* of the wire. This total or resultant field at the surface of the wire is as before

$$E(a) = E_a + E_z(a)$$

so that

$$E_a = E(a) - E_z(a) \tag{11-46}$$

Dividing by the current I gives the impedance per unit length of the wire.

$$Z = \frac{E_a}{I} = \frac{E(a)}{I} - \frac{E_z(a)}{I} = Z_{\text{int}} + Z_{\text{ext}} \qquad (11\text{-}47)$$

For the fields *within* the wire it is the appropriate solutions of (38), which must be used. Therefore within the wire

$$E_z\,(\text{int}) = AI_0(\gamma\rho) + BK_0(\gamma\rho)$$

The second of these functions becomes infinite at $\rho = 0$. Since E_z must always remain finite this requires that $B = 0$, so

$$E_z\,(\text{int}) = AI_0(\gamma\rho)$$

and from (29), remembering* that $I_0' = I_1$

$$H_\phi = \frac{\gamma A}{j\omega\mu} I_1(\gamma\rho)$$

At the surface of the wire, E_z (int) must equal the total or resultant electric intensity $E(a)$, and $2\pi a H_\phi = I$.

Therefore

$$E(a) = AI_0(\gamma a) \qquad I = 2\pi a \frac{\gamma A}{j\omega\mu} I_1(\gamma a)$$

and

$$Z_{\text{int}} = \frac{E(a)}{I} = \frac{j\omega\mu}{2\pi a\gamma}\frac{I_0(\gamma a)}{I_1(\gamma a)}$$

$$= \frac{\eta_m}{2\pi a}\frac{I_0(\gamma a)}{I_1(\gamma a)} \qquad (11\text{-}48)$$

where $\eta_m = \sqrt{j\omega\mu/(\sigma + j\omega\epsilon)}$ is the intrinsic impedance of the metal.

Equation (48) gives the exact expression for the internal impedance of the wire. The evaluation of this expression is simplified by recalling that for all metallic conductors at frequencies less than optical, $\sigma \gg \omega\epsilon$ and $\gamma \approx \sqrt{j\omega\mu\sigma} = \sqrt{\omega\mu\sigma}\,\sqrt{j} = \sqrt{\omega\mu\sigma}\ \underline{/45°}$. To assist in obtaining numerical values for expressions such as (48), the following auxiliary functions have been defined and tabulated

$$I_0(v\sqrt{j}) = \text{ber}\, v + j\, \text{bei}\, v$$

* Recurrence formulas for the I and K functions differ from the other Bessel functions. These formulas are listed in the Appendix.

Tables of these "farmyard functions" may be found in the references noted.* Curves showing the internal impedance of wires as given by (48) may be found in several texts.† Two special cases of this general expression are of importance practically and will be considered.

The first of these special cases occurs for thin wires at low (power) frequencies, where the wire radius is small compared with the depth of penetration. For this case, $|\gamma a| \ll 1$, and only the first two terms of the power series expansion for $I_0(\gamma a)$ and $I_1(\gamma a)$ need be used. Then

$$I_0(\gamma a) \approx 1 + \frac{(\gamma a)^2}{4}, \qquad I_1(\gamma a) \approx \frac{\gamma a}{2} + \frac{(\gamma a)^3}{16}$$

$$Z_{\text{int}} \text{ (low freq)} \approx \frac{1}{\pi a^2(\sigma + j\omega\epsilon)} + \frac{j\omega\mu}{8\pi}$$

$$\approx \frac{1}{\pi a^2 \sigma} + j\omega \frac{\mu}{8\pi} \tag{11-49}$$

These terms represent respectively the *low-frequency resistance* and *internal inductive reactance* of the wire, per unit length. The low-frequency internal inductance of the wire is $\mu/8\pi$ henry/m.

The second special case of practical importance occurs for frequencies sufficiently high that the depth of penetration is small compared with the radius of the wire. This makes $|\gamma a| \gg 1$. Except for quite thin wires, this case covers all radio frequencies. Using the asymptotic expansions for I_0 and I_1, the internal impedance becomes

$$Z_{\text{int}} \text{ (high freq)} \approx \frac{\eta_m}{2\pi a} = \frac{Z_s}{2\pi a} = \frac{R_s}{2\pi a} + \frac{jX_s}{2\pi a} \tag{11-50}$$

As would be expected, when the depth of penetration is small compared with the radius, the internal impedance per unit length of the wire is equal to the surface impedance of a thick plane sheet of the metal 1 meter long and $2\pi a$ meters wide. Evaluating (50) in terms of the constants of the metal shows that

$$Z_{\text{int}} \text{ (high freq)} = \frac{1}{2\pi a}\sqrt{\frac{\omega\mu}{2\sigma}} + \frac{j}{2\pi a}\sqrt{\frac{\omega\mu}{2\sigma}} \tag{11-51}$$

* McLachlan, *Bessel Functions for Engineers*, Oxford Press, London, 1934; Dwight, *Tables of Integrals*, Macmillan, New York, 1934.

† For example, Ramo and Whinnery, *Fields and Waves in Modern Radio* John Wiley and Sons, New York, 1944, chap. 6.

The first term is high-frequency ohmic resistance of the wire per unit length, and the second term is the high-frequency internal inductive reactance of the wire per unit length.

PART II

Circuit Relations and Field Theory

In the first part of this chapter the relations of field theory have been used to develop expressions for the *impedance* of a straight wire in two rather special cases. In the first case the wire was of finite length and was assumed to carry a sinusoidally distributed current. The impedance was calculated at the terminals. In the second case the wire was assumed infinitely long with a uniform current distribution, and the *impedance per unit length* was calculated. It is apparent that it should be possible, in a somewhat similar manner, to derive an expression for the impedance at the terminals of a wire circuit of any configuration. This is indeed the case, and it will be shown that the so-called *circuit relations*, by means of which the engineer solves for the current in a circuit in terms of the applied voltage and the circuit impedances, are derivable from field theory as special and approximate cases. Before carrying through such a derivation, it is desirable to re-examine circuit concepts for a simple closed circuit, to see how these concepts follow directly from the integral statement of Maxwell's equations.

11.09 Circuit Relations and Maxwell's Equations in the Integral Form. Consider the Maxwell emf equation (Faraday's law) applied to the simple circuit of Fig. 11-14 consisting of a loop of wire with terminals *a-b*.

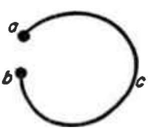


Fig. 11-14

$$\int_{(bca)} E_s\,ds + \int_{(ab)} E_s\,ds = -j\omega\Phi \qquad (11\text{-}52)$$

where E_s is the component of $\mathbf{E}$ parallel to the wire and Φ is the magnetic flux through the loop.

The first integral is taken along the wire, and the second integral is along a straight line joining the terminals. The E_s along the path of integration is the self-induced electric intensity, produced by the charges and current in the circuit. The voltage V°, which must be applied or impressed at the terminals of the circuit to transfer the charges against this self-induced field, will be equal and opposite to the second integral. That is,

$$V^\circ = -\int_{(ab)} E_s\,ds$$

Then eq. (52) can be rewritten as

$$V^\circ = -\int_{(ab)} E_s\,ds = \int_{(bca)} E_s\,ds + j\omega\Phi \qquad (11\text{-}53)$$

Dividing through by the current I, an impedance equation is obtained.

$$Z = \frac{V^\circ}{I} = \frac{\int_{(bca)} E_s\,ds}{I} + \frac{j\omega\Phi}{I} \qquad (11\text{-}54)$$

The first term on the right-hand side is the internal impedance Z_i of the wire, and the second term represents the external reactance. If the external inductance L_s is defined by Φ/I, then eq. (54) becomes

$$Z = Z_i + j\omega L_s \qquad (11\text{-}55)$$

In the d-c case ($\omega = 0$), the external reactance is zero and the internal impedance is the resistance of the wire. In the alternating case, the internal impedance Z_i is complex, and consists of a resistance R_i and an internal reactance $X_i = j\omega L_i$. If a perfectly conducting wire is assumed, the first integral on the right-hand side of (53) is zero, and the applied voltage is equal to the external reactive voltage drop $j\omega\Phi = j\omega L_e I$. For an actual conductor the total inductance L is the sum of the external and internal inductances, that is $L = L_s + L_i$. In practice $L_i \ll L_e$, so that the total inductance L is very nearly equal to the external inductance L_e. The inductance L can be increased by winding the wire in the form of a coil. In this manner the magnetic flux per ampere is increased, and the same magnetic flux is caused to link several turns. If the inductive reactance of the coil is large enough so that the inductive reactance of the rest of the circuit may be neglected, the inductance is said to be "lumped."

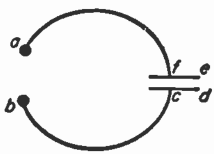


FIG. 11-15

If a condenser is connected in series with the loop as in Fig. 11-15, the emf equation becomes

$$\int_{(bcd)} E_s\,ds + \int_{(de)} E_s\,ds + \int_{(efa)} E_s\,ds + \int_{(ab)} E_s\,ds = -j\omega\Phi \quad (11\text{-}56)$$

The first and third terms are due to the internal impedance of the wire and condenser plates. The second term is the voltage between the condenser plates. This is proportional to the charge Q on the plates, and the ratio

$$\frac{Q}{\int_{(de)} E_s \, ds}$$

is the capacitance C of the condenser. Q is related to the current I by

$$Q = \int I \, dt = \frac{I}{j\omega}$$

so the second term of eq. (56) may be written

$$\int_{(de)} E_s \, ds = \frac{1}{j\omega C} I$$

The applied voltage $V°$ is equal to the negative of the fourth integral, so eq. (56) becomes

$$V° = Z_i I + j\omega\Phi + \frac{1}{j\omega C} I \tag{11-57}$$

If the small internal reactance of the wire is lumped with the external inductive reactance, (57) may be written as

$$V° = I\left(R + j\omega L + \frac{1}{j\omega C}\right) \tag{11-58}$$

where L is now the total or effective inductance of the circuit. This is the usual form of the circuit equation.

There are several approximations and assumptions involved in writing eq. (58). Some of these will be evident from the manner of its derivation from eq. (52), but others are more obscure. They will be listed here and discussed in greater detail later in this section.

1. The current I has been assumed to have the same magnitude in all parts of the circuit. This means that "distributed capacitance" effects, or displacement currents from one conductor to another have been neglected.

2. An inductance L has been defined for low frequencies (actually at $\omega = 0$) and has then been used in (58) as though it were independent of frequency.

3. Retardation effects, e.g. radiation, have been neglected.

At power frequencies, the approximations are excellent and the neglected quantities are indeed negligible. However, at radio frequencies and more especially at ultrahigh frequencies, some of the neglected factors become important, and the circuit approach breaks down unless appropriate steps are taken to make circuit concepts carry over, for example, by generalizing definitions. Generalized definitions for circuit constants can be obtained by considering the circuit as a problem in field theory. The direct derivation of **E** and **H** from Maxwell's equations in the integral form was easily done for the *closed* or *quasi-closed* circuits of Figs. 11-14 and 11-15. However, for open circuits such as antennas, where radiation is important, it is generally simpler to obtain **E** and **H** indirectly through the retarded potentials **A** and V. It is instructive to use this more general field method to derive the simple circuit relations already considered. Such a derivation points up the approximations involved in the latter relations and indicates the extent of the errors incurred when ordinary circuit theory is used at high frequencies. In addition, generalized definitions can be obtained for the circuit "constants," by means of which it becomes possible to extend the use of the circuit approach to the ultrahigh frequencies.

11.10 Derivation of Circuit Relations from Field Theory. The electric circuit laws of Ohm, Faraday, and Kirchhoff were based on experimental observations and antedated the electromagnetic theory of Maxwell and Lorentz. Indeed, the theory was developed as a generalization from these simpler and more restricted laws. It is interesting, but not surprising, then, to find that the circuit relations are just special cases of the more general field relations, and that they may be developed from the latter when suitable approximations are made. Nevertheless, the importance of the simple (and approximate) circuit relations should not be underestimated. With these beautifully simple relations the electrical engineer has been enabled to design and construct electrical systems and circuits of amazing intricacy. Without the simplifying assumptions of circuit theory the vast power and communication networks of today would not have been possible, for the exact field solutions to many of the problems would have been of overwhelming complexity.

In this section the circuit relations dealing with voltages and currents will be derived as special cases of electromagnetic field theory, the theory which treats with charge and current densities and their associated fields.

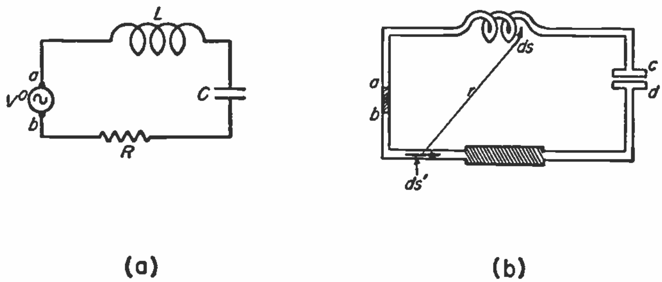


FIG. 11-16. (a) A simple *RLC* circuit. (b) A representation suitable for the application of field theory.

Consider again the simple series circuit of Fig. 11-16 for which can be written the circuit equation

$$V^\circ = IR + \frac{I}{j\omega C} + j\omega LI \tag{11-59}$$

The applied or impressed voltage V° is assumed to be independent of the resultant current I. Equation (59) can be rewritten in the form

$$V^\circ + V' = V_R \tag{11-60}$$

where
$$V' = -j\omega LI - \frac{1}{j\omega C} I \tag{11-61}$$

is the sum of the reactive voltages across the circuit elements and

$$V_R = IR \tag{11-62}$$

is the net voltage left to drive the current I through the resistance R after the reactive voltage drops have been subtracted from the applied voltage.

In field theory relations similar to (60), (61), and (62) may be written for electric fields and conduction current densities. Thus at the surface of, or within, a conductor, the conduction current density is given by Ohm's law.

$$\frac{i}{\sigma} = \mathbf{E} \tag{11-63}$$

where σ is the conductivity of the conductor and $\mathbf{E}$ is the *total* electric intensity tangential to the surface. In general, this total $\mathbf{E}$ is the sum of an *applied* or *impressed* electric intensity $\mathbf{E}^\circ$ and a "self-induced" or back electric intensity $\mathbf{E}'$ that is due to the charges and currents in the system.

That is
$$\mathbf{E} = \mathbf{E}^\circ + \mathbf{E}' \tag{11-64}$$

The impressed intensity $\mathbf{E}^\circ$ is assumed to be independent of the charges and currents in the system under consideration. This would be the case for example if $\mathbf{E}^\circ$ were the electric field of a *distant* antenna. In this circumstance, to use circuit terminology, the coupling between the two systems is sufficiently loose that the charges and currents in the second system do not affect (to any significant extent) the current flowing in the distant antenna.

The "self-induced" electric intensity $\mathbf{E}'$ that is due to the charges and currents in the system under consideration may be determined from Maxwell's equations, either directly or through the scalar and vector potentials. In terms of the potentials,

$$\mathbf{E}' = -\operatorname{grad} V - \omega\mu j\mathbf{A} \tag{11-65}$$

where V and $\mathbf{A}$ are related to the charge and current densities of the system through

$$V = \frac{1}{4\pi\epsilon}\int_{\text{vol}} \frac{\rho\left(t - \frac{r}{c}\right)}{r}\,dV \qquad \mathbf{A} = \frac{1}{4\pi}\int_{\text{vol}} \frac{i\left(t - \frac{r}{c}\right)}{r}\,dV \tag{11-66}$$

Then, rewriting (64) and using (63) and (65), the field relations at the surface of a conductor may be written as

$$\mathbf{E}^\circ = \frac{i}{\sigma} + \operatorname{grad} V + j\omega\mu\mathbf{A} \tag{11-67}$$

Integrating along the *conducting* portion of a circuit (Fig. 11-16), the general circuital relation is obtained, viz.:

$$\int_c^d E_s^\circ\,ds = \int_c^d \frac{i_s}{\sigma}\,ds + \int_c^d \frac{\partial V}{\partial s}\,ds + \int_c^d j\omega\mu A_s\,ds \tag{11-68}$$

Making suitable assumptions and approximations, the general relation (68) can be reduced to the simple circuit equation (59). The steps required to derive (59) from (68) show clearly the approximations involved in the simple circuit equation.

In eq. (68) the path of integration is taken along the surface of the conductor parallel to its axis. The expression (67) cannot be integrated across the condenser gap, because there both i and σ are zero and the first expression on the right-hand side is indeterminate. If there is no series condenser in the circuit, the points c and d are coincident, and the integration is performed around a *completely closed* conducting path. For this case of a completely closed path the circuital relation corresponding to (68) would be

$$\oint E_s^\circ\, ds = \oint \frac{i_s}{\sigma}\, ds + \oint j\omega\mu A_s\, ds \tag{11-69}$$

The second term on the right-hand side of (68) has dropped out because the gradient of a scalar potential integrated around a *closed* path is always zero. That is

$$\oint \text{grad } V \cdot \mathbf{ds} = \oint \frac{\partial V}{\partial s}\, ds \equiv 0$$

The various terms of eq. (68) will now be considered one at a time. The term on the left-hand side of (68) evidently corresponds to the applied voltage V°. In circuit work V° is supplied by an electric generator, which is usually a complicated circuit in itself. However, for purposes of solving for voltages and currents in the circuit under consideration (the driven circuit), V° is assumed to be supplied across a pair of terminals by a zero-impedance generator, or by a zero-impedance generator connected in series with a lumped impedance equal to the generator impedance. Similarly in considering the field relations, the impressed or applied field $\mathbf{E}^\circ$ usually exists along a complicated configuration of conductors (in the generator winding) and may extend over an appreciable portion of the circuit under consideration. However, for purposes of analysis the impressed field $\mathbf{E}^\circ$ is often assumed to exist only along a section of conductor of very short length; that is, a "point" or "slice" generator is assumed to exist between the points a and b, (Fig. 11-16a), so that the applied voltage is

$$\int_c^d E_s^\circ\, ds = \int_a^b E_s^\circ\, ds = V^\circ \tag{11-70}$$

Now consider the first term on the right-hand side of eq. (68). For the direct-current case, the interpretation of this term would be very simple. The current density i_s would be uniform throughout the conductor cross section and would be given by

$$i_s = \frac{I}{A} \tag{11-71}$$

where A is the cross sectional area of the conductor. The conductivity is the reciprocal of the resistivity ρ, and so

$$\frac{1}{A\sigma} = \frac{\rho}{A} = R'$$

where R' is the resistance per unit length of the conductor. Then

$$\frac{i_s}{\sigma} = \frac{i_s}{A} \cdot \frac{A}{\sigma} = IR' \tag{11-72}$$

is just the voltage drop (due to resistance) per unit length, and the first term on the right-hand side of (68) becomes

$$\int_c^d \frac{i_s}{\sigma}\, ds = \int_c^d IR'\, ds = IR \tag{11.73}$$

which is the total IR drop around the circuit.

From (72) and (63) it is seen that for the direct-current case, the ratio of total tangential electric intensity E to total current I is the resistance per unit length, that is

$$\frac{E}{I} = R'$$

For alternating currents, especially at high frequencies, the current density is no longer uniform throughout the cross-section of the conductor. Instead it varies—both in magnitude and phase through the cross-section—so that the total current I, in general, differs in phase from the current density at the surface of the conductor. The ratio E/I is now complex and defines z_i, the "internal impedance" per unit length of the conductor. Now the first term on the right-hand side of eq. (68) may be written

$$\int_c^d \frac{i_s}{\sigma}\, ds = \int_c^d E_s\, ds = \int_c^d Iz_i\, ds \tag{11-74}$$

When the current I is uniform around the circuit, as it is in the low-frequency case, (74) becomes

$$\int_c^d \frac{i_s}{\sigma} ds = I \int_c^d z_i \, ds = IZ_i \tag{11-75}$$

$Z_i = R_i + jX_i$ is the so-called internal impedance of the conductors of the circuit. For direct-current it reduces to the circuit resistance. Even in the case of high-frequency alternating currents, the internal reactance X_i is very small compared with the "external reactance" of the circuit obtained from the second and third terms on the right-hand side of (68) and may usually be neglected. In any event, in circuit work the internal reactance of the conductors is usually lumped with the external reactance to give the total circuit reactance, and the resistive or in-phase component of the first term of (68) is shown explicitly as IR.

Consider next the second term on the right-hand side of eq. (68). When integrated around a closed path, as in the case of a circuit containing only resistance and inductance, this term is zero. However, for a circuit with a condenser (Fig. 11-16b), where the integration is carried from one plate c to the other plate d, there results.

$$\int_c^d \frac{\partial V}{\partial s} ds = V_d - V_c \tag{11-76}$$

This is the potential difference between the plates of the condenser. If these plates are considered to be very close together, and if the charge distributed along the wire is small compared with the charge concentrated on the condenser plates (that is, if stray capacitance is negligible compared with the capacitance of the condenser), the potential difference $(V_d - V_c)$ will be proportional to the charge on the condenser plates. That is

$$V_d - V_c = \frac{Q}{C} = \frac{I}{j\omega C} \tag{11-77}$$

The proportionality factor C is just the capacitance of the condenser as defined for the static case. Thus, for the second term of eq. (68), it is possible to write

$$\int_c^d \frac{\partial V}{\partial s} ds = \frac{I}{j\omega C} \tag{11-78}$$

Finally consider the third term on the right-hand side of eq. (68). This could be written

$$\int_c^d j\omega\mu A_s\, ds = j\omega L'I' \tag{11-79}$$

where

$$L' = \frac{1}{I'}\int_c^d \mu A_s\, ds \tag{11-80}$$

is a "generalized inductance" of the circuit. This generalized inductance depends both on the circuit geometry and on the current distribution, and, as defined by (80), it also depends upon where in the circuit the current I' is measured. For low frequencies, where the current amplitude is constant around the circuit, this generalized definition reduces to a well-known formula for low-frequency inductance [eq. (87)]. To see how the "inductance" of the circuit changes as the frequency increases, it is necessary to examine more closely the integral expression of equation (80).

For a current flowing in a thin wire, the expression for the vector potential at any point in space, due to an elemental length, is

$$dA = \frac{1}{4\pi}\frac{I\, ds'\, e^{-j\beta r}}{r} \tag{11-81}$$

where I is the integrated value of current density over the cross section of the wire, and r is the distance from an element of length ds' along the center of the wire to the point at which **A** is evaluated. The total vector potential due to current flow in the entire circuit will be

$$\mathbf{A} = \frac{1}{4\pi}\oint \frac{I\, e^{-j\beta r}}{r}\, ds' \tag{11-82}$$

In the third term of eq. (68) the component of **A** parallel to the axis of the wire is evaluated at the surface of the wire, and integrated around the conducting part of the circuit from c to d. This term can then be written

$$j\omega \int_c^d \mu A_s\, ds = j\omega \int_c^d \oint \frac{\mu I\, e^{-j\beta r}}{4\pi r}\, d\mathbf{s}' \cdot d\mathbf{s} \tag{11-83}$$

In the general case I varies with the position of ds' and must be retained under the integral sign. The usual low-frequency approximations are to assume that I is constant around the circuit with no change of phase, and also to neglect the phase shift factor $e^{-j\beta r}$.

At very high frequencies where the circuit dimensions become appreciable fractions of a wavelength, both of these approximations lead to error. However, the effect of neglecting the phase shift factor is much the more important because it is responsible for radiation from the circuit. For circuits that are not too large in wavelengths, say less than one-tenth wavelength around, the current distribution usually departs a surprisingly small amount from the low-frequency, constant-amplitude, constant-phase condition. (The reason for this can be seen by considering the current amplitude and phase variations along a short-circuited low-loss transmission line, the length of which is less than one-twentieth of a wavelength.) Because of these facts it is often permissible to neglect variations of current amplitude and phase around the circuit while still accounting for the phase-shift factor $e^{-j\beta r}$. Under such conditions (83) becomes

$$j\omega I \int_c^d \oint \frac{\mu\, e^{-j\beta r}}{4\pi r}\, d\mathbf{s}' \cdot d\mathbf{s} \tag{11-84}$$

Comparison with $j\omega LI$ shows that

$$\int_c^d \oint \frac{\mu\, e^{-j\beta r}}{4\pi r}\, d\mathbf{s}' \cdot d\mathbf{s} \tag{11-85}$$

is the factor that, at low frequencies, is identified as the inductance of the circuit. At frequencies sufficiently low that the phase shift is negligible (that is, for which $\beta r <<< 1$)

$$e^{-j\beta r} \approx 1$$

and the low-frequency inductance is

$$L_{LF} = \int_c^d \oint \frac{\mu\, d\mathbf{s}' \cdot d\mathbf{s}}{4\pi r} \tag{11-86}$$

If the circuit is *closed*, *c* and *d* coincide and

$$L_{LF} = \oint \oint \frac{\mu\, d\mathbf{s}' \cdot d\mathbf{s}}{4\pi r} \tag{11-87}$$

which is known as Neumann's formula for the external inductance of a circuit.

At higher frequencies, where it is no longer permissible to neglect the factor $e^{-j\beta r}$, its effect can be determined by expanding it in series form and using the first few terms:

$$e^{-j\beta r} = 1 - j\beta r - \frac{\beta^2 r^2}{2!} + \frac{j\beta^3 r^3}{3!} + \frac{\beta^4 r^4}{4!} - \cdots$$
$$= \left(1 - \frac{\beta^2 r^2}{2!} + \cdots\right) - j\left(\beta r - \frac{\beta^3 r^3}{3!} + \cdots\right)$$

It is seen that the expression (85) for "inductance" now has both real and imaginary parts. The real part represents the high-frequency external inductance of the circuit. The imaginary part is the so-called radiation resistance of the circuit. From expression (84) it is evident that this imaginary part combines with the factor $j\omega I$ to yield a voltage *in-phase* with I. The power required to drive I against this in-phase component of voltage is radiated from the circuit.

The value of the radiated power is given by

$$W_{\text{rad}} = I^2 \int_c^d \oint \frac{j\omega\mu}{4\pi r}\left(-j\beta r + j\frac{\beta^3 r^3}{3!} - \cdots\right) ds' \cdot ds = I^2 R_{\text{rad}} \tag{11-88}$$

where R_{rad} is the radiation resistance of the circuit and is given by

$$R_{\text{rad}} = \int_c^d \oint \mu\left(\frac{\omega^2}{4\pi c} - \frac{\omega^4 r^2}{24\pi c^3} - \cdots\right) ds' \cdot ds \tag{11-89}$$

When integrated around a closed path, the first term drops out, leaving

$$R_{\text{rad}} = 10^{-7} \int_c^d \oint \left(-\frac{\omega^4 r^2}{3!c^3} + \frac{\omega^6 r^4}{5!c^5} - \cdots\right) ds' \cdot ds \tag{11-90}$$

Consideration of the real part of expression (85) shows how the inductance depends upon the phase factor $e^{-j\beta r}$.

$$L = \int_c^d \oint \frac{\mu}{4\pi r}\left(1 - \frac{\beta^2 r^2}{2!} + \cdots\right) ds' \cdot ds \tag{11-91}$$
$$= 10^{-7} \int_c^d \oint \frac{1}{r}\left(1 - \frac{\omega^2 r^2}{2c^2} + \cdots\right) ds' \cdot ds$$

Using eqs. (75), (76), (78), and (79), it is seen that at low frequencies eq. (68) reduces directly to eq. (59). Comparison of eqs. (68) and (59) shows clearly the approximations involved in the simple circuit relations, and makes it possible to determine the magnitude of the neglected factors. With this knowledge circuit concepts may be extended to much higher frequencies.

The extension of circuit concepts to higher frequencies is accomplished in practice by the addition of appropriately located lumped-circuit constants. For example "distributed" inductance and capacitance effect are accounted for by suitably located series inductors and shunt capacitors, and radiation effects by the inclusion of a "radiation resistance." An outstanding example in electrical engineering of the extension of circuit concepts to systems not necessarily small in wavelengths is the ordinary transmission line. Here, by suitably representing the distributed constants of the line by lumped constants, a circuit results that can be solved by ordinary circuit methods. Although the circuit is complicated, the solution is relatively simple in the important practical case of a uniform transmission line. In this manner it is possible in some problems to extend circuit concepts even to the microwave range.

ADDITIONAL PROBLEMS

9. A resonant-length dipole ($L = 2H$, slightly less than $\lambda/2$) has a free-space input impedance of $73 + j0$ ohms. What is its input impedance when placed parallel to, and a quarter-wavelength from, a large perfectly conducting screen.

10. A parasitic (unfed and short-circuited) dipole has a length of 108 cm and a radius of 0.5 cm. Determine the magnitude and phase of the current in it when placed parallel to and 0.1 wavelength from a half-wave dipole carrying 1 amp. Frequency = 150 mc. From curves, find $Z_{12} = 68 + j10$ approximately. What is the input impedance of the driven dipole (antenna 1) if it is assumed that $Z_{11} \approx 73$ ohms.

BIBLIOGRAPHY

See bibliography for chap. 10.

CHAPTER 12

DIRECTIONAL CHARACTERISTICS OF ANTENNAS

12.01 Introduction. Radio antennas have a twofold function. The first of these functions is to "radiate" the radio frequency energy that is generated in the transmitter and guided to the antenna by the transmission line. In this capacity the antenna acts as an impedance-matching device to match the impedance of the transmission line to that of free space. The other function of the antenna is to direct the energy into desired directions, and what is often more important, to suppress the radiation in other directions where it is not wanted. This second function of the antenna will be considered first under the general heading of directional characteristics.

A completely nondirectional or omnidirectional radiator radiates uniformly in all directions and is known as an *isotropic* radiator or a *unipole*. A point source of sound is an example of an isotropic radiator in acoustics. There is no such thing as an isotropic radiator of electromagnetic energy, since all radio antennas have some directivity. However, the notion of a completely nondirectional source is useful, especially for gain comparison purposes.

The *radiation pattern* of an antenna is a graphical representation of the radiation of the antenna as a function of direction. When the radiation is expressed as field strength, E volts per meter, the radiation pattern is a *field strength pattern.* If the radiation in a given direction is expressed in terms of power per unit solid angle, the resulting pattern is a *power pattern.* A power pattern is proportional to the square of the field strength pattern. Unless otherwise specified, the radiation patterns referred to in this book will be field strength patterns.

The co-ordinate system generally used in the specification of antenna radiation patterns is the spherical co-ordinate system (r, θ, ϕ), shown in Fig. 12-1. The antenna is located at or near the origin of this system, and the field intensity is specified at points on the spherical surface of radius r (or on a semispherical surface in the case of ground-based antennas). The shape of the radiation pattern is independent of r, as long as r is chosen sufficiently large (r must be very much greater than the wavelength and very much greater than the largest dimension of the antenna system.) When

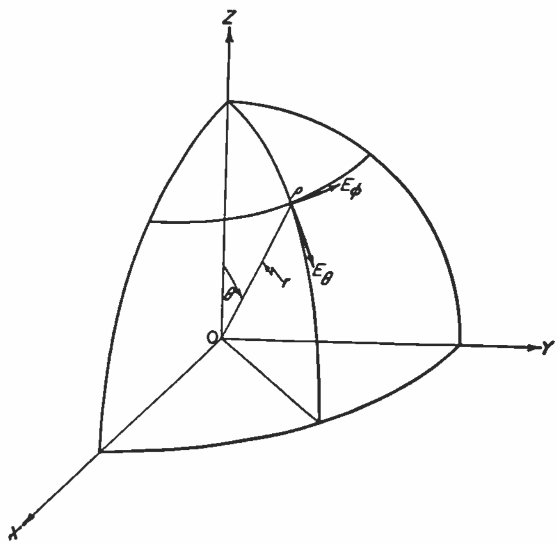


FIG. 12-1. Spherical co-ordinate system.

this is true, the magnitude of the field strength in any direction varies inversely with r, and so needs to be stated for only one value of r. For example, in broadcast antenna work it is customary to state the field strength at a radius of 1 mile. Often only the *relative* radiation pattern is used. This gives the relative field strengths in various directions, usually referred to unity in the direction of maximum radiation.

For the radiation field, the direction of **E** is always tangential to the spherical surface. For a vertical dipole **E** is in the θ direction; whereas for a horizontal loop **E** is in the ϕ direction. In general, the radiation field intensity may have both E_θ and E_ϕ components, which may or may not be in time phase. The radiation characteristics are then shown by separate patterns for the theta and phi polarizations. The terms, *theta polarization* and *phi polarization* are synonymous with and replace the older terms *vertical* polarization and *horizontal* polarization, respectively. The older terms were confusing in that a theta or vertically polarized signal is not always vertical (however it is always in the vertical plane through the radius vector), although a phi or horizontally polarized signal is always horizontal.

A complete radiation pattern gives the radiation for all angles of ϕ and θ and really requires three-dimensional presentation. This is overcome by showing cross sections of the pattern in planes of interest. Cross sections in which the radiation patterns are most frequently given are the horizontal ($\theta = 90°$) and vertical (ϕ = constant) planes. These are called the *horizontal pattern* and *vertical patterns*, respectively.

12.02 Directional Properties of Dipole Antennas. The magnitude of the radiation term for the field strength due to an elementary dipole $I\,dl$ is

$$E = E_\theta = \frac{60\pi I\,dl}{r\lambda} \sin\theta \qquad \text{volt/m} \qquad (12\text{-}1)$$

where θ is the angle between the axis of the dipole and the radius vector to the point where the field strength is measured. When the dipole is vertical, the horizontal radiation pattern is a circle (Fig. 12-2a) because in this plane ($\theta = 90°$) the radiation is uniform. In any vertical plane through the axis the field strength varies as $\sin\theta$ and the vertical patterns are all the same, having the figure-eight shape shown in Fig. 12-2b. When the dipole is horizontal, the horizontal pattern has the figure-eight shape, but the vertical pattern depends upon the angle which the vertical plane makes with the horizontal axis. The two vertical planes of chief interest are those perpendicular and parallel to the axis of the dipole. The vertical radiation pattern is a circle for the former and a figure-eight for the latter. These two vertical patterns and the horizontal pattern are known as the *principal plane* patterns.

As the length of a dipole is increased beyond the point where it may be considered short in terms of a wavelength, the radiation pattern in the planes through the axis changes as indicated in Fig.

12-2. Figure 12-2c shows the vertical radiation pattern of a center-fed half-wave vertical dipole, and Fig. 12-2d shows the same pattern when the dipole is one wavelength long. The expression

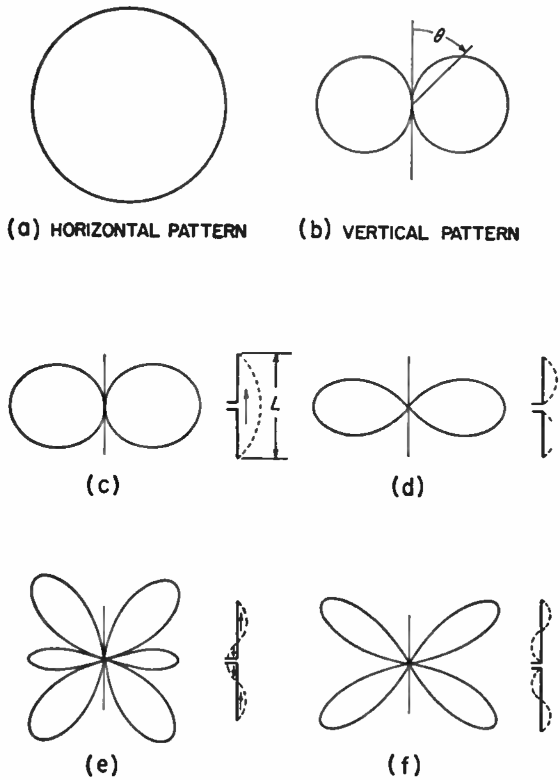


FIG. 12-2. Radiation patterns of center-fed vertical dipoles: (a) horizontal pattern; (b) vertical pattern for a short dipole. Vertical patterns for dipole lengths: (c) one-half wavelength; (d) one wavelength; (e) one-and-a-half wavelengths; (f) two wavelengths. The assumed current distribution for each case is shown dashed.

for the magnitude of the radiation field intensity due to a half-wave dipole is

$$E = \frac{60I}{r}\left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\right] \quad \text{volt/m} \qquad (12\text{-}2)$$

The more general expression for a dipole of any length $L = 2H$ is

$$E = \frac{60I}{r}\left[\frac{\cos \beta H - \cos (\beta H \cos \theta)}{\sin \theta}\right] \text{volt/m} \qquad (12\text{-}3)$$

Expressions (1) and (2) were derived in chap. 10 and expression (3) was obtained in a problem in the same chapter. The radiation patterns, as given by (3), are shown in Fig. 12-2e and 12-2f for antenna lengths of 1½ and 2 wavelengths.

The vertical radiation patterns of Fig. 12-2 also apply to the corresponding grounded vertical antennas when mounted on a perfectly conducting ground plane. The length of the grounded vertical antenna is just one-half the length of the corresponding dipole (the image forms the other half), and of course only the top half of the pattern applies.

12.03 Traveling-wave Antennas and Effect of the Point of Feed on Standing-wave Antennas. The patterns of Fig. 12-2 are for unterminated antennas that are assumed to have a standing-wave distribution of current. Sometimes antennas are terminated to make them *aperiodic* or *nonresonant* and, in this case, they have a traveling-wave distribution. The directional pattern of a traveling-wave antenna having a length of 6 wavelengths is shown in Fig. 12-3. Assuming negligible attenuation of the wave along the antenna, the expression for the pattern of a traveling-wave antenna (see problem 20, chap. 10) is

$$E = \frac{30I_m \sin \theta}{r(1 - \cos \theta)} \{2 - 2 \cos [\beta L(1 - \cos \theta)]\}^{½} \qquad (12\text{-}4)$$

It is seen that with a traveling current wave the pattern is no longer symmetrical about the $\theta = 90$ degrees plane, but instead the radiation tends to "lean" in the direction of the current wave. The angle θ between the axis of the antenna and the direction of maximum radiation becomes smaller as the antenna becomes longer.

In the case of an unterminated antenna the actual current distribution in general is a combination of standing wave and traveling wave. However, except for very long antennas, the standing wave is predominant and the traveling-wave component of current is usually neglected in pattern calculations. For the standing-wave current distribution, the pattern is always symmetrical about the $\theta = 90$ degrees plane. For *center-fed* antennas the pattern of the traveling-wave component of current is also symmetrical about $\theta = 90$ degrees, so that the effect of this latter component of current tends to be obscured. This is especially true when the angles of maximum radiation for the two current distributions nearly coincide, as is often the case. However, when an antenna is unsymmetrically fed, the pattern due to the traveling-wave current distribution is no longer symmetrical about $\theta = 90$ degrees, and its effect on the resultant pattern becomes more pronounced. This is evident in Fig. 12-4, which shows some experimental patterns of wire antennas having different locations for the feed points. The

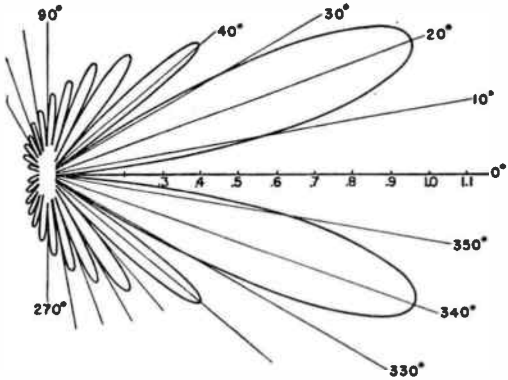


FIG. 12-3. Radiation pattern of a traveling-wave antenna that is 6 wavelengths long.

asymmetry due to the traveling-wave current shows up when the feed point is moved away from center.

Changing the location of the feed point has another, even more important, effect when the antenna is longer than a half-wavelength. This is the effect on the standing-wave current distribution, which may be quite different for different locations of the feed point as is also illustrated in Fig. 12-4 for full wave and 1½ wavelength antennas. The effects of these different current distributions is clearly evident in the measured patterns.

12.04 Two-element Array. When greater directivity is required than can be obtained by a single antenna, antenna arrays are used. An *antenna array* is a system of similar antennas, similarly oriented. Antenna arrays make use of wave interference phenomena that occur between the radiations from the different elements of the

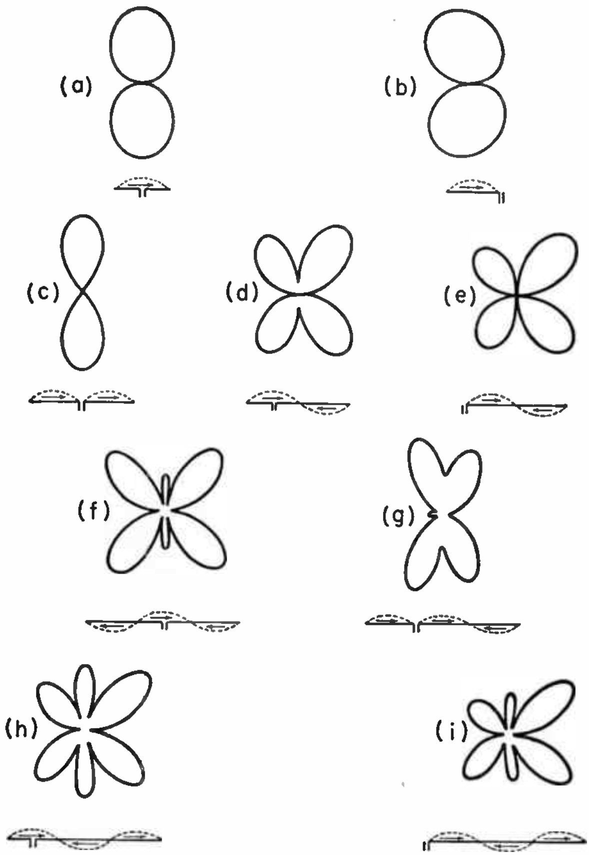


FIG. 12-4. Experimental patterns of wire antennas having different locations for feed points: (a) half-wave center-fed; (b) half-wave end-fed; (c) full-wave center-fed; (d) full-wave fed one-quarter wavelength from one end; (e) full-wave end-fed; (f), (g), (h), and (i) one-and-one-half wave, fed as indicated. (*Courtesy Electronics.*)

array. Consider the two-element array of Fig. 12-5 in which the antennas 0 and 1 are nondirectional radiators in the plane under consideration. (For example, they could be vertical radiators when the horizontal pattern is being considered.) When the point P is sufficiently remote from the antenna system, the radius vectors to the point can be considered parallel, and it is possible to write

$$r_1 = r_0 - d \cos \phi$$

in the phase factor of the fields, and

$$\frac{1}{r_1} = \frac{1}{r_0}$$

as far as the magnitudes of the fields are concerned. The phase difference between the radiations from the two antennas will be

$$\psi = \beta d \cos \phi + \alpha$$

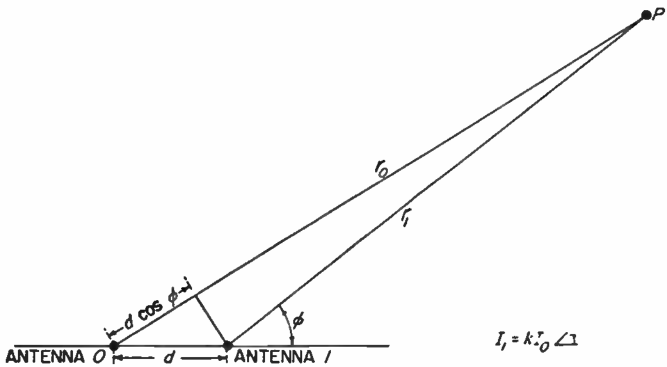


FIG. 12-5. A two-element array of nondirectional radiators.

where $\beta d = (2\pi/\lambda)d$ is the path difference in radians and α is the phase angle by which the current I_1 leads I_0. The vector sum of the fields will be

$$E = E_0(1 + k\, e^{j\psi})$$

where E_0 is the field intensity due to antenna 0 alone, and where k is the ratio of the magnitudes of I_1 and I_0. The magnitude of the total field intensity is given by

$$\begin{aligned} E_T &= |E_0(1 + k\, e^{j\psi})| \\ &= |E_0(1 + k \cos \psi + jk \sin \psi)| \\ &= E_0 \sqrt{(1 + k \cos \psi)^2 + k^2 \sin^2 \psi} \end{aligned}$$

In the particular but important case where the antenna currents have equal magnitudes, this becomes (see Fig. 12-6b)

$$\begin{aligned} E_T &= 2E_0 \cos \frac{\psi}{2} \\ &= 2E_0 \cos \left(\frac{\pi d \cos \phi}{\lambda} + \frac{\alpha}{2} \right) \end{aligned} \qquad (12\text{-}5)$$

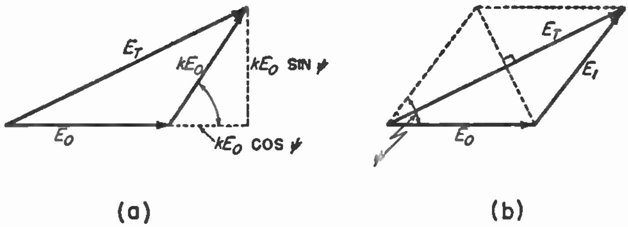


FIG. 12-6. Phasor addition of fields.

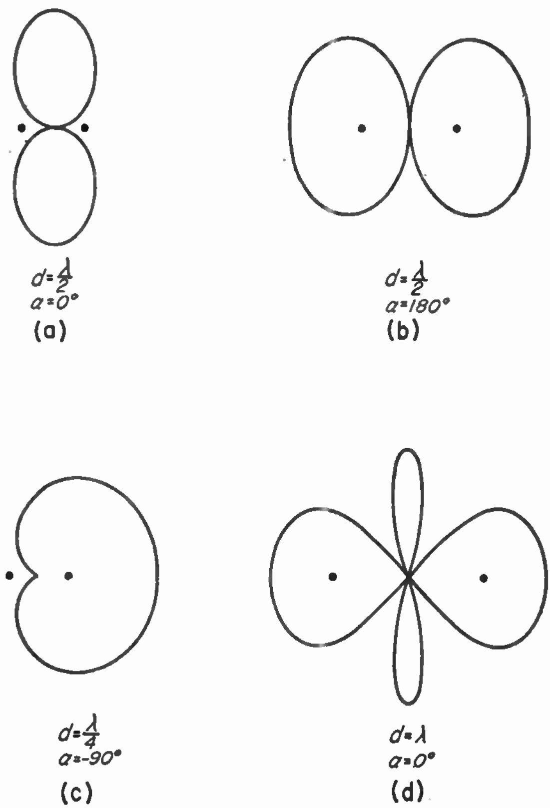


FIG. 12-7. Radiation patterns of two nondirectional radiators when fed with equal currents at the phasings shown.

The radiation patterns resulting from the expression for commonly used spacings and phasings are sketched in Fig. 12-7. These patterns are for the case where each of the antennas, when radiating alone, is a nondirectional or *point source* radiator; that is, it has a circle for its radiation pattern in the plane under consideration. This is quite evidently true for vertical antennas when the horizontal pattern is being considered.

12.05 Horizontal Patterns in Broadcast Arrays. Two element arrays are limited in the type and variety of radiation patterns that they can produce, and for broadcast arrays three or more antennas

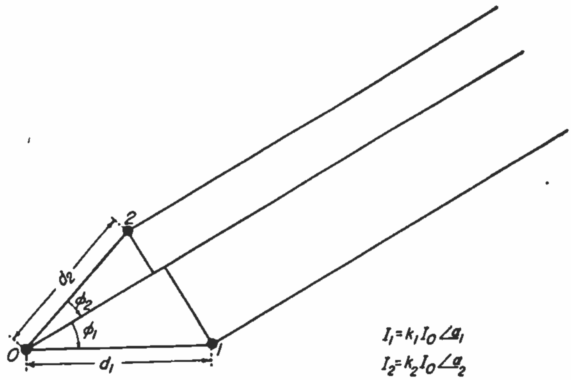


FIG. 12-8. A three-element array.

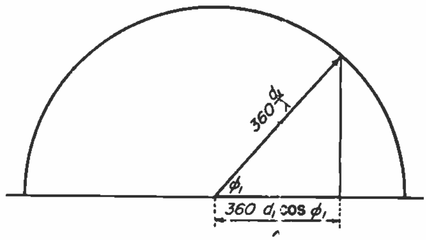
are often used. With a two-antenna array the pattern must always be symmetrical about the plane through the antennas, and the position of only two nulls can be specified. A three-element array, in which antenna configurations and spacing as well as current magnitudes and phases are all variables under the control of the designer, permits a larger number of different antenna pattern types. For a three-element array as in Fig. 12-8 the resultant horizontal intensity pattern is given by

$$E_T = |E_0(1 + k_1 e^{j\psi_1} + k_2 e^{j\psi_2})| \tag{12-6}$$

where

$$\psi_1 = (\beta d_1) \cos \phi_1 + \alpha_1$$
$$\psi_2 = (\beta d_2) \cos \phi_2 + \alpha_2$$

The evaluation of expression (6) is straightforward but rather time-consuming when a large number of points must be plotted. A graphical method which is sometimes used for evaluating an expression such as (6) is shown in Fig. 12-9. The antenna spacing is expressed in degrees and a semicircle is drawn with this spacing as radius. For each angle ϕ_1 the value of $\beta d \cos \phi$ is read off directly in degrees, and ψ_1 is obtained by adding the angle α. ψ_2 is obtained in a like manner and the vector addition of Fig. 12-9b gives the resultant E_T.



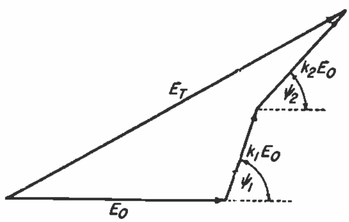


FIG. 12-9. Graphical method for obtaining antenna patterns.

The design of an array to produce a desired pattern is usually done on a cut-and-try basis. That is, certain spacings and currents are assumed and the corresponding pattern computed. Modifications are then made in the assumed conditions and the pattern is recomputed. This process is continued until a pattern close enough to the desired pattern has been obtained. Since the computation involved is simply that of vector addition, various "pattern calculators," both mechanical and electronic, have been devised to perform the computation. A book* on directional antennas shows some

* Carl E. Smith, *Directional Antennas*, Cleveland Institute of Radio Electronics, Cleveland, Ohio.

15,000 computed patterns obtained with the aid of a mechanical plotter.

12.06 Linear Arrays. For point-to-point communication at the higher frequencies the desired radiation pattern is a single narrow lobe or beam. To obtain such a characteristic (at least approximately) a multielement linear array is usually used. An array is *linear* when the elements of the array are spaced equally along a straight line (Fig. 12-10). In a *uniform linear array* the elements are fed with currents of equal magnitude and having a uniform progressive phase shift along the line. The pattern of such an array can be obtained as before by adding vectorially the field intensities due to each of the elements. For a uniform array of nondirectional elements the field intensity would be

$$E_T = E_0|1 + e^{j\psi} + e^{j2\psi} + e^{j3\psi} + \cdots + e^{j(n-1)\psi}| \qquad (12\text{-}7)$$

where

$$\psi = \beta d \cos \phi + \alpha$$

and α is the progressive phase shift between elements. (α is the angle by which the current in any element *leads* the current in the preceding element.)

For the purpose of computing the pattern of the linear array, eq. (7) may be written as

$$\frac{E_T}{E_0} = \left|\frac{1 - e^{jn\psi}}{1 - e^{j\psi}}\right| = \left|\frac{\sin \frac{n\psi}{2}}{\sin \frac{\psi}{2}}\right| \qquad (12\text{-}8)$$

The maximum value of this expression is n and occurs when $\psi = 0$. This is the *principal maximum* of the array.* Since $\psi = \beta d \cos \phi + \alpha$ the principal maximum occurs when

$$\cos \phi = -\frac{\alpha}{\beta d}$$

For a *broadside* array the maximum radiation occurs perpendicular to the line of the array at $\phi = 90$ degrees, so $\alpha = 0$ degrees. For an

* If the spacing d is equal to or greater than λ, there may be more than one principal maximum.

end-fire array the maximum radiation is along the line of the array at $\phi = 0$, so $\alpha = -\beta d$ for this case.

The expression (8) is zero when

$$\frac{n\psi}{2} = \pm k\pi \qquad k = 1, 2, 3, \cdots$$

These are the *nulls* of the pattern. Secondary maxima occur approximately midway between the nulls, when the numerator of expression (8) is a maximum, that is when

$$\frac{n\psi}{2} = \pm(2m + 1)\frac{\pi}{2} \qquad m = 1, 2, 3, \cdots$$

The first secondary maximum occurs when

$$\frac{\psi}{2} = \frac{+3\pi}{2n}$$

(note that $\psi/2 = \pi/2n$ does not give a maximum). The amplitude of the first secondary lobe is

$$\left|\frac{1}{\sin(\psi/2)}\right| = \left|\frac{1}{\sin(3\pi/2n)}\right| \approx \frac{2n}{3\pi} \qquad \text{for large } n$$

The amplitude of the principal maximum was n so the amplitude ratio of first secondary maximum to principal maximum is $2/3\pi = 0.212$. This means that the first secondary maximum is about 13.5 db below the principal maximum, and this ratio is *independent* of the number of elements in the uniform array, as long as the number is large.

The width of the principal lobe, measured between the first nulls, is twice the angle between the principal maximum and first null. This latter angle is given by

$$\frac{n\psi_1}{2} = \pi \qquad \text{or} \qquad \psi_1 = \frac{2\pi}{n}$$

For a broadside array $\cos\phi = \psi/\beta d$, and the principal maximum occurs at $\phi = \pi/2$. The first null occurs at an angle $[(\pi/2) + \Delta\phi]$

where

$$\cos\left(\frac{\pi}{2} + \Delta\phi\right) = \frac{\psi_1}{\beta d} = \frac{\lambda}{dn}$$

If $\Delta\phi$ is small, it is given approximately by

$$\Delta\phi = \frac{\lambda}{nd}$$

and the width of the principal lobe is

$$2\Delta\phi = \frac{2\lambda}{nd} \tag{12-9a}$$

For a *uniform* broadside array the width of the principal lobe (in radians) is approximately twice the reciprocal of the array length in wavelengths.

For the end-fire array $\psi = \beta d(\cos\phi - 1)$. The principal maximum is at $\phi = 0$, and the first null is at $\phi_1 = \Delta\phi$ where

$$\psi_1 = \beta d(\cos\phi_1 - 1) = -\frac{2\pi}{n}$$

or

$$(\cos\Delta\phi) - 1 = -\frac{\lambda}{nd}$$

For $\Delta\phi$ small, there results approximately

$$\frac{(\Delta\phi)^2}{2} = \frac{\lambda}{nd}$$

$$2\Delta\phi = 2\sqrt{\frac{2\lambda}{nd}} \tag{12-9b}$$

The width of the principal lobe of a uniform end fire array, as given approximately by expression (9b) is greater than that for a uniform broadside array of the same length.

12.07 Multiplication of Patterns. The methods of the preceding section provide straightforward means for determining the radiation patterns of uniform linear arrays. However, for such arrays there is also available another method for obtaining these same patterns. This second method, when it can be used, has the great advantage that it makes it possible to sketch rapidly, almost by inspection, the patterns of complicated arrays. Because of this fact, the method is a useful tool in the design of arrays.

Consider a four-element array of antennas in Fig. 12-10, in which the spacing between units is $\lambda/2$ and the currents are in phase ($\alpha = 0$). The pattern can be obtained directly by adding vectorially the four electric fields due to the four antennas. However the same radiation pattern can be obtained from the following considerations. The pattern of antennas 1 and 2 operating as a unit, that is two antennas spaced $\lambda/2$ and fed in phase, is already known and is that of Fig. 12-7a. Also antennas 3 and 4 may be considered as another similar unit with the same pattern of Fig. 12-7a. As far as the resultant radiation pattern is concerned

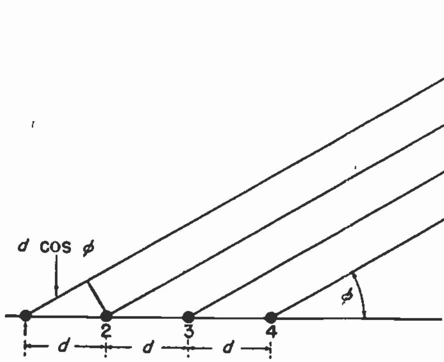


FIG. 12-10. A four-element linear array of nondirectional radiators.

antennas 1 and 2 could be replaced by a single antenna located at a point midway between them and having as its directional characteristic the "figure eight" of Fig. 12-7a. Antennas 3 and 4 could similarly be replaced by a single antenna having the figure eight pattern. The problem is then reduced to that of determining the radiation pattern of two similar antennas that are spaced a wavelength apart and each of which has a figure eight directional pattern. Now the pattern of two nondirectional radiators spaced 1λ and fed in phase is already known and is that of Fig. 12-7d. For the case of Fig. 12-7d each of the antennas alone radiates equally in all directions in the plane being considered. When these antennas are replaced by radiators that radiate different amounts in different directions, the pattern of Fig. 12-7d must be modified accordingly. The resultant pattern for the original four element array is obtained as the *product* of the pattern of Fig. 12-7d by the pattern of the unit, Fig. 12-7a.

This multiplication of patterns is illustrated below.

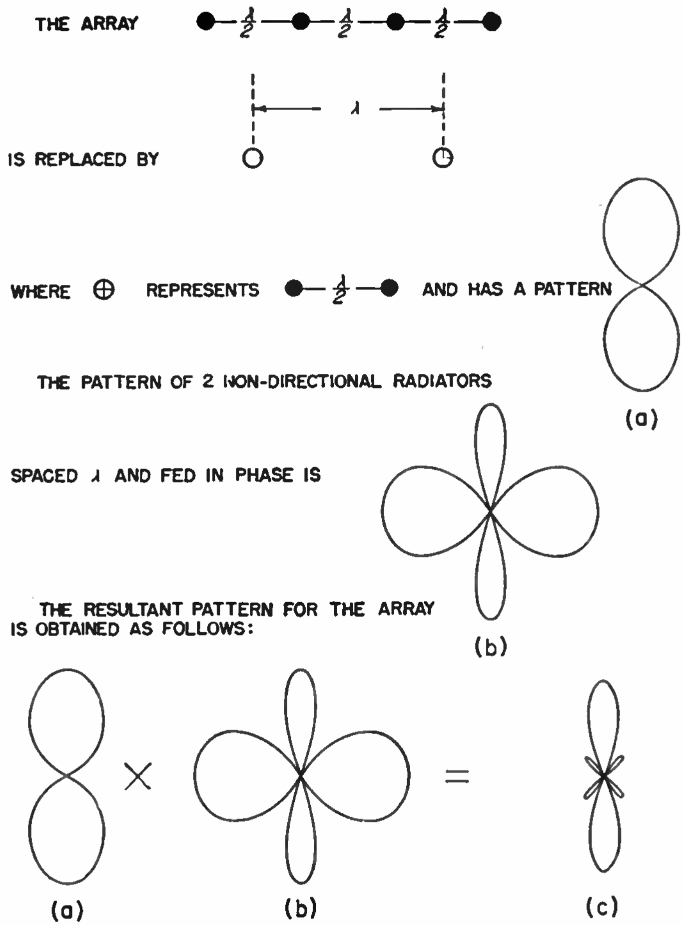


Fig. 12-11. Multiplication of patterns.

The application of this principle to more complicated arrays follows quite readily. For example the pattern of a broadside array of eight elements spaced one-half wavelength and fed in phase would be obtained by considering four elements as a unit and finding the pattern of two such units spaced a distance of two wavelengths. This is shown in Fig. 12-12. The resultant pattern is the product of the unit pattern for four elements (already obtained in Fig. 12-11) by the pattern for two nondirectional radiators spaced two wavelengths apart (calculated from eq. 7).

This procedure provides a means for rapidly determining what the resultant pattern of a complicated array will look like without making lengthy computations, since the approximate pattern can be arrived at by inspection. The width of the principal lobe (between nulls) is the same as the width of the corresponding lobe of the group pattern. The number of secondary lobes can be determined

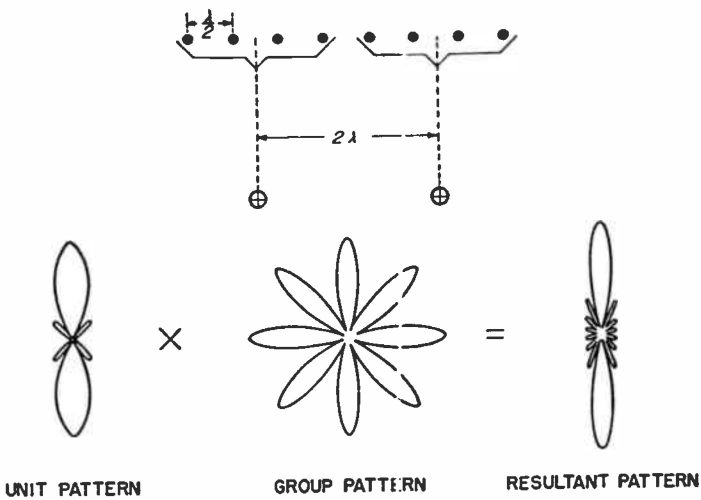


FIG. 12-12. Pattern for an eight-element uniform array obtained by principle of multiplication of patterns.

from the number of nulls in the resultant pattern, which is just the sum of the nulls in the unit and group patterns (assuming none of the nulls are coincident). Although the chief usefulness of the method is in being able to obtain an approximate idea of the pattern of a complicated array by inspection, the method itself is exact, and a point by point multiplication of patterns yields the exact pattern for the resultant.

Patterns in Other Planes. Figure 12-7a is the pattern, *in the plane normal to the axes of the antennas*, of two antennas spaced one-half wavelength apart and fed *in phase*. In this plane the antennas are nondirectional or uniform radiators. If the pattern in the plane containing the antennas is desired (in which plane the antennas are *directional*), it is necessary to multiply the pattern of Fig. 12-7a by the directional pattern of the antenna in the plane being considered. For half-wave dipole antennas this latter pattern will be the "figure-eight" pattern of Fig. 12-2c. This is shown as the unit pattern in Fig. 12-13.

The resultant pattern is then obtained as a multiplication of the group pattern by the unit pattern (Fig. 12-13A). If the antennas

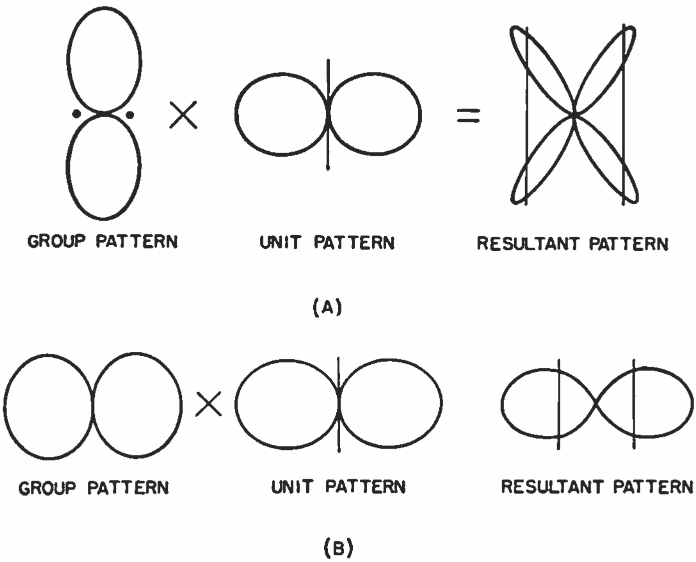


FIG. 12-13. Radiation pattern (in the plane containing the axes of the antennas) of two-element array of half-wave dipoles: (A) fed in phase; (B) fed 180 degrees out of phase.

are fed 180 degrees out of phase (end-fire array) the directions of maxima of group and unit patterns coincide and the desirable directional characteristic of Fig. 12-13B results.

12.08 Effect of the Earth on Vertical Patterns. The radiation patterns shown so far have been obtained on the assumption that the antenna or antenna array was situated in free space far removed from any other conducting bodies or reflecting surfaces. In practice, antennas are nearly always erected either right at, or within a few wavelengths of, the surface of the earth, or some other reflecting surface. Under these conditions currents flow in the reflecting surface, and the radiation pattern is modified accordingly. The magnitudes and phases of these induced currents will of course be dependent to some extent upon the surface impedance of the reflector (that is, upon σ, ϵ, ω). However, for practical purposes it is adequate to compute the resultant fields on the assumption that the surfaces are perfectly conducting. This is true, for example, for the earth at low and medium frequencies, and for metallic reflectors at any radio frequency.

In Fig. 12-14 are shown horizontal and vertical antennas located above the earth (assumed perfectly conducting). The boundary conditions to be satisfied at the surface of the perfectly conducting

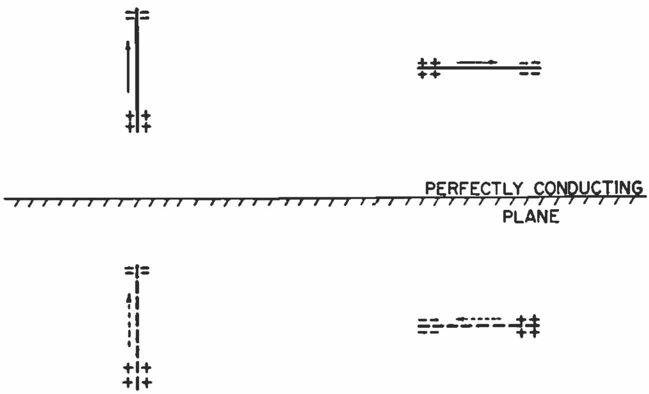


FIG. 12-14. "Image" charges and currents replace the charges and currents induced in the conducting plane.

plane are that the tangential component of **E** and the normal component of **H** must vanish. That is, at the surface **E** is normal and **H** tangential. Charges will distribute themselves and currents will flow on the conducting surface in such a manner that these boundary conditions are satisfied. The total electric and magnetic fields will be due not only to the charges and currents on the antenna, but also to these "induced" charges and currents. As far as the electric and magnetic fields in the region above the conducting plane are concerned, the same results can be obtained with the conducting plane removed and replaced with suitable located "image" charges and currents, as shown in Fig. 12-14. The image charges will be "mirror images" of the actual charges, but will have opposite sign. The currents in actual and image antennas will have the same directions for vertical antennas, but opposite directions for horizontal antennas.

For perfectly conducting planes these same results are also given in terms of simple ray theory as pictured in Fig. 12-15. The resultant field is considered as made up of direct and reflected waves, the image antenna being the virtual source of the reflected wave. The vertical component of electric field for the incident wave is

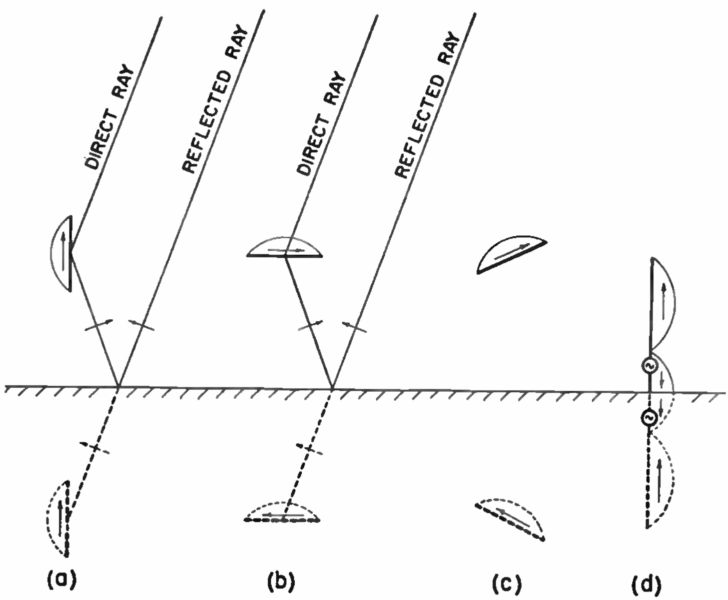


FIG. 12-15. Image antennas act as virtual sources for the reflected waves.

reflected without phase reversal, whereas the horizontal component has a 180 degrees phase reversal. It is seen that the phase delay due to path length differences (that is, the effect of retardation) is automatically taken care of.

The use of the image principle makes it a simple matter to take into account the effect of the presence of the earth on the radiation patterns. The earth is replaced by an image antenna, located a distance $2h$ below the actual antenna, where h is the height above the ground of the actual antenna. The field of this image antenna is added to that of the actual antenna to yield the resultant field.

The *relative* horizontal pattern will remain unchanged (its absolute value changes), but the vertical pattern is affected greatly.

For simple arrays above a reflecting surface the principle of multiplication of patterns can be used to obtain the resultant vertical patterns. The vertical pattern of the antenna (or array) is multiplied by the vertical pattern of two nondirectional or point-source radiators having equal amplitudes and spaced one above the

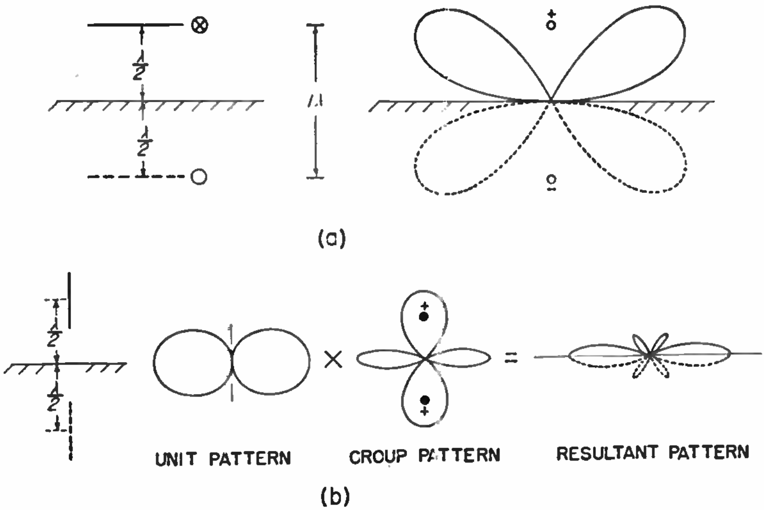


FIG. 12-16. (a) Vertical pattern of a horizontal antenna above the earth, obtained by considering the pattern of the antenna and its negative image. (b) Vertical pattern of a vertical antenna above the earth, obtained by using the principle of images and the principle of multiplication of patterns.

other a distance $2h$ apart. For vertical antennas the nondirectional radiators would be considered to have the same phase, whereas for horizontal antennas the nondirectional radiators would have opposite phases. Examples of this method are shown in Fig. 12-16. Of course, only the upper half of the resultant pattern actually exists. When the antenna is sloping as in Fig. 12-15c, the pattern cannot be obtained by this multiplication process but the image principle can still be used to obtain the resultant field. This same statement also applies for vertical antennas mounted at the surface of the earth when the antenna is not a multiple of one-half wavelength long (Fig. 12-15d).

When the finite conductivity of the earth must be considered, the idea of images is still valid, but the simple ray theory used here is no longer adequate. It is then necessary to return to field theory for accurate answers. The effect of an imperfect earth on the radiated fields is considered in chap. 16 on ground-wave propagation.

12.09 Binomial Array. An example of the usefulness of the principle of multiplication of patterns is given in the derivation of the so-called binomial array. With a uniform linear array it is found that, as the array length is increased in order to increase the

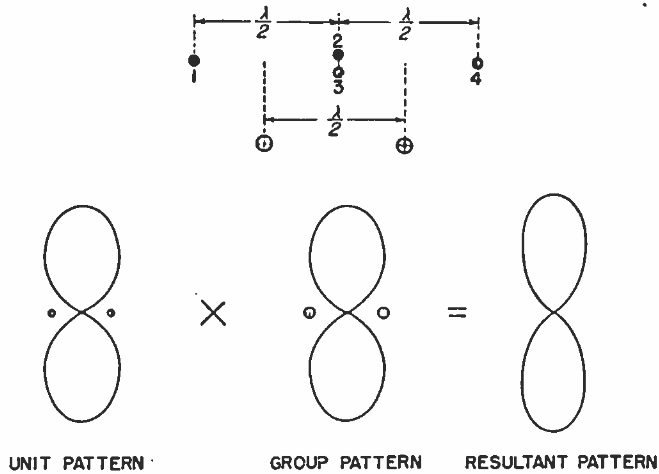


Fig. 12-17. An array that produces a pattern without secondary lobes.

directivity, secondary or minor lobes always appear in the pattern. For some applications a single narrow lobe without minor lobes is desired. A study of the uniform array, using the principle of multiplication of patterns, shows that secondary lobes appear in the resultant pattern whenever the elements that produce the unit pattern or the elements that produce the group or space pattern have a spacing greater than one-half wavelength. Thus in the uniform four-element array of Fig. 12-11, the secondary lobes appear in the resultant because the group pattern has four lobes. The group pattern has four lobes because the effective sources producing the group pattern are spaced a full wavelength. Reduction of the spacing of the elements of the group to one-half wavelength results in a two-lobed figure-eight pattern for the group pattern, and a resultant pattern that has only primary lobes. The antenna arrangement that will result in half-wavelength separation of the elements of the group, is shown in Fig. 12-17 along with the resultant patterns. In this case antennas 2 and 3 coincide so they would be replaced with a single antenna carrying double the current in the other elements. That is, a three-element array results, that has the current ratios 1:2:1 and the pattern shown as the resultant in Fig. 12-17. Since this pattern is the product of two figure-eight patterns, it can be called a "figure-eight squared" pattern.

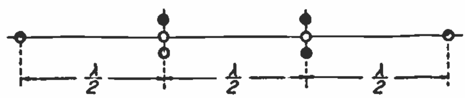


FIG. 12-18. A four-element array with a "figure-eight cubed" pattern.

Using this three-element array as a unit with a second similar unit spaced one-half wavelength from it results in the four-element array shown in Fig. 12-18. The current ratios of this array are

$$1:3:3:1$$

and the pattern is the "figure-eight squared" pattern of the unit times a figure-eight group pattern that results in a "figure-eight cubed" pattern. This process may be continued to obtain a pattern having any desired degree of directivity and no secondary lobes. The numbers that give the current ratios will be recognized as the binomial coefficients. For an array n half-wavelengths long the relative current in the rth element from one end is given by:

$$\frac{n!}{r!(n-r)!} \qquad (12\text{-}10)$$

where $r = 0, 1, 2, 3, \cdots$

12.10 Antenna Gain. The *gain* g of an antenna in a *given direction* is defined as 4π times the ratio of the radiation intensity in that direction to the total power W. When W is taken as total power delivered to the antenna, the gain is called *power gain*. When W is taken as the total power radiated from the antenna the gain is called *directive gain*. The ratio of directive gain to power gain is independent of direction and is equal to the ratio of the total power delivered to the total power radiated from the antenna. For an "efficient" antenna, directive gain and power gain are very nearly equal.

The *radiation intensity* Φ in a given direction is the power per unit solid angle radiated in that direction. For an *isotropic* radiator, for which Φ is the same in all directions, the total power radiated is

$$W_0 = 4\pi\Phi$$

Thus the gain of an antenna

$$g = \frac{4\pi\Phi}{W} = \frac{W_0}{W} \tag{12-11}$$

is just the ratio of the power radiated by an isotropic antenna to the power radiated by the actual antenna when both are producing the same radiation intensity in the direction for which the gain is specified. When the gain is expressed, in decibels, it is denoted by G, where

$$G = 10 \log_{10} g \tag{12-12}$$

The *directivity* or *maximum directive gain* of an antenna is the ratio of the maximum to the average radiation intensity. The *directivity* is obtained from

$$g_{\text{max}} = \frac{4\pi\Phi_{\text{max}}}{\int \Phi \, d\Omega} \tag{12-13}$$

where $d\Omega$ is an element of solid angle. $\int \Phi \, d\Omega$ is, of course, just W, the total power radiated. Although these definitions* have been framed by considering a transmitting antenna, they are applied to the antenna regardless of its particular function. That is, the gain of an antenna when used for receiving is the same as its gain when used for transmitting. Of course, the gain thus defined can be realized on a receiving antenna only when it is in the presence of a properly polarized field.

The gain or directivity of an antenna is easily computed when its effective length and radiation resistance are known. For example, for a current element $I\,dl$, the distant field intensity in the direction of maximum radiation is

$$E = \frac{60\pi}{r} I \left(\frac{dl}{\lambda}\right)$$

* These definitions are those given in the IRE Antenna Standards (1948).

The corresponding power flow per square meter is

$$P = \frac{E^2}{\eta_v}$$

and the radiation intensity (power per unit solid angle) is

$$\Phi = \frac{E^2 r^2}{\eta_v} \tag{12-14}$$

The radiation resistance of the current element is $80\pi^2(dl/\lambda)^2$ ohms, so that the power radiated for an effective current I is

$$W = 80\pi^2 \left(\frac{dl}{\lambda}\right)^2 I^2 \qquad \text{watts}$$

The current required to radiate 1 watt is

$$I = \frac{\lambda}{\sqrt{80}\pi\, dl} \qquad \text{amp}$$

with a corresponding field intensity, in the direction of maximum radiation, of

$$E = \frac{60}{r\sqrt{30}} \qquad \text{volt/m}$$

and a radiation intensity

$$\Phi = \frac{60^2}{80 \times 120\pi} = \frac{3}{8\pi}$$

For the same 1 watt radiated, the radiation intensity produced by an isotropic radiator would be

$$\Phi_0 = \frac{1}{4\pi}$$

so that the directivity or maximum directive gain of the current element is

$$g_{\max} = \frac{\Phi}{\Phi_0} = 1.5 \tag{12-15a}$$

or

$$G_{\max} = 10 \log_{10} 1.5 = 1.76 \text{ db} \tag{12-15b}$$

For a half-wave dipole the computed gain is 1.64 or 2.15 db. Thus the maximum directive gain of a half-wave dipole is only 0.39 db greater than for a current element (or for a very short dipole).

12.11 Effective Area of an Antenna. A term which has considerable significance, especially for receiving antennas, is effective area. The *effective area A* of an antenna is defined in terms of the gain of the antenna through the relation

$$A = \frac{\lambda^2}{4\pi} g \qquad (12\text{-}16)$$

Using this relation it can be shown that the effective area is the ratio of power available at the terminals of the antenna to the power per unit area of the appropriately polarized incident wave. That is, the received power is equal to the power flow through an area equal to the effective area of the antenna. This would be written as

$$W_R = PA \qquad (12\text{-}17)$$

where W_R is the received power and P is the power flow per square meter for the incident wave. That relation (17) holds for the current element receiving antenna can be demonstrated quite simply.

For an effective field intensity E, the power per square meter in the linearly polarized received wave is

$$P = \frac{E^2}{\eta_v} = \frac{E^2}{120\pi} \qquad \text{watts/sq m}$$

The power absorbed by a properly matched load connected to the receiving antenna would be

$$W_R = \frac{V^2_{oc}}{4R_{rad}} = \frac{E^2\, dl^2}{4R_{rad}}$$

The radiation resistance of the current element is

$$R_{rad} = 80\pi^2 \left(\frac{dl}{\lambda}\right)^2$$

so that

$$W_R = \frac{E^2\lambda^2}{320\pi^2}$$

and

$$A = \frac{W_R}{P} = 1.5\,\frac{\lambda^2}{4\pi}$$

$$= \frac{\lambda^2 g}{4\pi}$$

which agrees with definition (16).

Relations between g, l_{eff}, and R_{rad}. Only two of the three quantities, gain, effective length, and radiation resistance are required to specify the radiation characteristics of an antenna that emits linearly polarized waves. When two of these quantities are known, the third can be derived.

The radiation field of an antenna at a distance r can be expressed in terms of the effective length and the current at the terminals by the relation

$$E = \frac{60\pi}{r}\left(\frac{l_{eff}}{\lambda}\right) I$$

The radiation intensity Φ is expressible in terms of the field intensity by

$$\Phi = \frac{E^2 r^2}{\eta}$$
$$= 30\pi \left(\frac{l_{eff}}{\lambda}\right)^2 I^2$$

If $W = I^2 R_{rad}$ is the total power radiated, the gain g is

$$g = \frac{4\pi\Phi}{W}$$
$$= \frac{120\pi^2}{W}\left(\frac{l_{eff}}{\lambda}\right)^2 I^2$$
$$= \frac{120\pi^2}{R_{rad}}\left(\frac{l_{eff}}{\lambda}\right)^2 \qquad (12\text{-}18)$$

This equation relates g, l_{eff}, and R_{rad}.

Using the reciprocity theorem, it has already been shown that l_{eff} is the same for receiving as for transmitting. The power received by an antenna of effective length l_{eff} in the presence of a linearly polarized field intensity E is

$$W_R = \frac{V^2_{oc}}{4R_{rad}} = \frac{E^2 l^2_{eff}}{4R_{rad}}$$

Using (18), this becomes

$$W_R = g\,\frac{E^2\lambda^2}{4 \times 120\pi^2}$$
$$= \left(\frac{g\lambda^2}{4\pi}\right)\left(\frac{E^2}{120\pi}\right)$$
$$= \left(\frac{g\lambda^2}{4\pi}\right) P \qquad (12\text{-}19)$$

Equation (19) shows that for an antenna of any length the effective area defined by (16) gives the ratio of power received to power flow per unit area in the incident wave.

Problem 1. Knowing the effective length of a half-wave dipole, compute its effective area.

12.12 Elliptical Polarization. In the examples used in the preceding section only linearly polarized waves were considered. Such waves are emitted by the simple antenna types such as dipoles in free space, or straight vertical antennas at the surface of a perfectly-conducting flat earth. However, for antennas which have both vertical and horizontal elements or antennas that are mounted on, or backed up by, curved surfaces, the radiated fields will, in general, be elliptically polarized. Under these circumstances the "ellipticity" of the polarization as well as the gain or effective area of the antenna should be stated in order that the directional characteristics be completely specified.

Consider a transmitting antenna T of arbitrary shape located at the origin of the spherical co-ordinate system (r, θ, ϕ) of Fig. 12-1. The electric field intensity at a distant point p will, in general, have two components, E_θ and E_ϕ. These components will differ in time phase by some angle α, which will vary with direction. That is, α is a function of θ and ϕ. There will also be two components of magnetic intensity in space quadrature, but in time phase with the components of **E** (assuming a nondissipative medium).

For any particular value of r the field vectors can be represented as below.

$$
\begin{aligned}
E_\theta(t) &= E_{\theta_1}\, e^{j\omega t} \qquad & H_\phi(t) &= \frac{E_{\theta 1}}{\eta_v}\, e^{j\omega t} \\
E_\phi(t) &= E_{\phi_1}\, e^{j(\omega t+\alpha)} \qquad & H_\theta(t) &= \frac{-E_{\phi_1}}{\eta_v}\, e^{j(\omega t+\alpha)}
\end{aligned}
\tag{12-20}
$$

The real power flow per square meter in the r direction will be given by

$$
\begin{aligned}
P_r &= \tfrac{1}{2}\,\mathrm{Re}\,(E_{\theta_1}H_\phi{}^* - E_{\phi_1}H_\theta{}^*) \\
&= \frac{1}{2\eta_v}\,(E_{\theta_1}{}^2 + E_{\phi_1}{}^2) \\
&= \frac{1}{\eta_v}\,(E_\theta{}^2 + E_\phi{}^2)
\end{aligned}
\tag{12-21}
$$

where E_θ and E_ϕ are effective values. Then the radiation intensity or power per unit solid angle will be

$$\Phi = \frac{r^2}{\eta_v}(E_\theta^2 + E_\phi^2)$$
$$= \Phi_\theta + \Phi_\phi$$
$$\Phi_\theta = \frac{r^2}{\eta_v} E_\theta^2$$

is the power per unit solid angle of the θ component of the field and

$$\Phi_\phi = \frac{r^2}{\eta_v} E_\phi^2$$

is the power per unit solid angle of the ϕ component of the field. The "total" gain of the antenna would be

$$g = \frac{4\pi\Phi}{W} = \frac{4\pi}{W}(\Phi_\theta + \Phi_\phi)$$
$$= g_\theta + g_\phi$$

Correspondingly, the "total" effective area of the antenna is

$$A = \frac{\lambda^2}{4\pi}(g_\theta + g_\phi)$$
$$= A_\theta + A_\phi$$

As a transmitting antenna this "total" gain or effective area is realized only when the receiving antenna is designed for the particular polarization being radiated. Similarly, when used as a receiving antenna, these same gain and effective area figures are realized only when the received wave has the correct polarization. It is evident that in order for these "total" values to have significance, it is necessary to specify the polarization of the wave radiated by the antenna. That is, the relative magnitudes E_θ and E_ϕ and the time phase angle between them must be known. In practice the radiation patterns for the θ component and ϕ component are measured separately, but the time phase angle between these components is hardly ever obtained because of the additional complexity of the measuring technique required. Under these circumstances the gain and effective area are specified separately for the θ component and the ϕ component of the field, and the "total" values are not used. The θ-polarization gain of an antenna used for transmitting is 4π times the ratio of the radiation intensity of the

θ-polarized field to the total power. When this same antenna is used for receiving its effective area for θ-polarization as defined by

$$A_\theta = \frac{\lambda^2}{4\pi} g_\theta \tag{12-22}$$

is equal to the ratio of the power received to the power per square meter of an incident wave that is linearly polarized in the θ direction. A similar relation holds for the ϕ polarization.

12.13 Antenna Gain from Pattern Measurements. The radiation pattern of an antenna or array as obtained from full-scale or model measurements is usually a relative pattern only. That is, the relative field intensity in various directions about the antenna system is known, but its absolute value in "volts per meter at 1 mile" or some other convenient units is unknown. If the complete three-dimensional relative pattern of the antenna array has been obtained, the power radiated and hence the absolute field intensity pattern, can be determined by graphical integration of the Poynting vector over a closed surface about the system. The particular method used for performing the integration depends upon the manner in which the experimental data are presented. The following method assumes that the information is available as a set of horizontal patterns (ϕ variable, θ constant) obtained for various values of the parameter θ.

Let the components of the relative field intensity at any point on the spherical surface enclosing the antenna system (Fig. 12-1) be E_{ϕ_r} and E_{θ_r}. If k is the factor by which these relative values must be multiplied to convert them to a basis of volts per meter, then from eq. (21) the Poynting vector is

$$P_r = \frac{k^2}{\eta_v} (E_{\phi_r}{}^2 + E_{\theta_r}{}^2) \tag{12-23}$$

The total power radiated is

$$W = \int_s \mathbf{P} \cdot d\mathbf{a} = \frac{k^2}{\eta_v} \int_0^{2\pi} \int_0^{\pi} (E_{\phi_r}{}^2 + E_{\theta_r}{}^2) r^2 \sin\theta \, d\theta \, d\phi$$

Then

$$W = \frac{k^2 r^2}{\eta_v} (A + B) \tag{12-24}$$

where

$$A = \int_0^{2\pi} \int_0^{\pi} E_{\phi_r}{}^2 \sin\theta \, d\theta \, d\phi$$

$$B = \int_0^{2\pi} \int_0^{\pi} E_{\theta_r}{}^2 \sin\theta \, d\theta \, d\phi$$

From eq. (24)

$$k = \sqrt{\frac{\eta_v W}{r^2(A + B)}}$$

If the field intensity is desired in terms of volt/m at 1 mile for 1 watt radiated, then $W = 1$ watt and $r = 1609$ meters, so that

$$k = \sqrt{\frac{120\pi}{(1609)^2(A + B)}}$$

$$= \frac{.01208}{\sqrt{A + B}} \qquad (12\text{-}25)$$

Evaluation of the integrals A and B will give the desired value of the conversion factor k.

Consider the integral A and write it in the following form:

$$A = 2 \int_0^\pi \left(\sin\theta \int_0^{2\pi} \tfrac{1}{2} E_{\phi_r}{}^2 \, d\phi\right) d\theta$$

Now the integral with respect to ϕ,

$$\int_0^{2\pi} \tfrac{1}{2} E_{\phi_r}{}^2 \, d\phi \qquad (12\text{-}26)$$

is given by the area of the E_{ϕ_r} pattern plotted with respect to ϕ on polar co-ordinate graph paper (Fig. 12-19). This integration may be performed with a planimeter (remembering to convert the planimeter reading from square centimeters to square units of graph paper). Let the result of this integration be designated by A_θ. Since there will be a value of A_θ for each value of θ, A_θ is a function of θ. Then

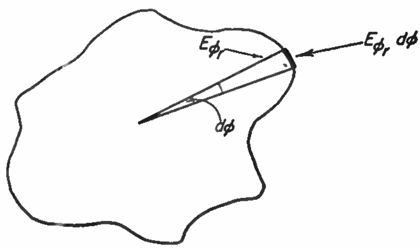


FIG. 12-19. Graphical method for obtaining antenna patterns.

$$A = 2 \int_0^\pi A_\theta \sin\theta \, d\theta \qquad (12\text{-}27)$$

There are several ways of evaluating the integral (27). About the simplest way is to approximate it with the finite sum.

$$A = 2 \sum_{\theta=0}^{\theta=180} A_\theta \sin\theta \frac{2\pi}{360} \Delta\theta$$

$$= \frac{\Delta\theta}{90} \sum_{\theta=0}^{\theta=180} A_\theta \sin\theta$$

The integral B is evaluated in the same manner. Substituting the values for A and B in eq. (25), the value of k is obtained.

With the conversion factor k known, the radiation patterns may be labeled in absolute values of volts per meter (at 1 mile for 1 watt radiated) through the relations

$$E_\phi \text{ (absolute)} = kE_{\phi_r}$$
$$E_\theta \text{ (absolute)} = kE_{\theta_r}$$

The Poynting vector is given by eq. (23). The radiation intensity for the ϕ polarization is

$$\Phi_\phi = \frac{k^2r^2}{\eta_v} E_{\phi_r}{}^2 = \frac{r^2}{\eta_v} E_\phi{}^2$$

and for the θ polarization

$$\Phi_\theta = \frac{k^2r^2}{\eta_v} E_{\theta_r}{}^2 = \frac{r^2}{\eta_v} E_\theta{}^2$$

The radiation intensity due to an isotropic antenna radiating 1 watt of power is

$$\Phi_0 = \frac{1}{4\pi}$$

The antenna gain for the ϕ polarization is

$$g_\phi = \frac{\Phi_\phi}{\Phi_0} = \frac{4\pi r^2}{\eta_v} E_\phi{}^2 = 86{,}400 E_\phi{}^2 \qquad (12\text{-}28)$$

where E_ϕ is measured in volts per meter at 1 mile for 1 watt radiated. For the θ polarization the antenna gain is

$$g_\theta = 86{,}400 E_\theta{}^2 \qquad (12\text{-}29)$$

12.14 The Mathematics of Linear Arrays. The binomial array of section 9 is but one example of a large class of linear arrays, having special current distributions by means of which the radiation patterns can be made to have almost any prescribed shape. Schelkunoff has shown* that linear arrays can be represented as polynomials and that this representation becomes a very useful tool in the analysis and synthesis of antenna arrays.

For a general linear array of equally spaced elements (Fig. 12-20), the relative amplitude of the radiated field intensity is given by

$$E = |a_0 e^{j\alpha_0} + a_1 e^{j\psi + j\alpha_1} + a_2 e^{j2\psi + j\alpha_2} + \cdots + a_{n-2} e^{j(n-2)\psi + j\alpha_{n-2}} + e^{j(n-1)\psi}| \quad (12\text{-}30)$$

where $\psi = \beta d \cos\phi + \alpha, \qquad \beta = \frac{2\pi}{\lambda}$

In this expression d is the spacing between elements. The coefficients a_0, a_1, a_2, etc., are proportional to the current amplitudes in

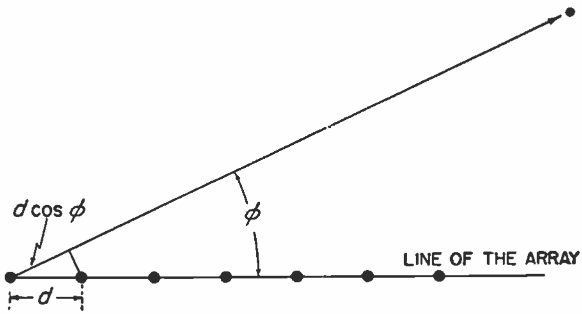


Fig. 12-20. A linear array.

the respective elements. α is the progressive phase shift (lead) from left to right; α_1, α_2, etc., are the deviations from this progressive phase shift. Expression (30) may be written

$$E = |A_0 + A_1 z + A_2 z^2 + \cdots + A_{n-2} z^{n-2} + z^{n-1}| \quad (12\text{-}31)$$

where $z = e^{j\psi}, \qquad A_m = a_m e^{j\alpha_m}$

The coefficients A_1, A_2, etc., are now complex and indicate the amplitude of current in each element and the phase deviation of that current from the progressive phase shift of the array. If any of the coefficients are zero, the corresponding element of the array will be missing, and the actual separation between adjacent elements can be greater than the "apparent separation" d. The apparent separation is the greatest common measure of the actual separations.

* S. A. Schelkunoff, "A Mathematical Theory of Linear Arrays," *BSTJ* **22**, 1, 80–107 (1943).

The following fundamental theorems are due to Schelkunoff, and lay the foundations for the method:

THEOREM I: "Every linear array with commensurable separations between the elements can be represented by a polynomial, and every polynomial can be interpreted as a linear array."
Since the product of two polynomials is a polynomial, a corollary to Theorem I is

THEOREM II: "There exists a linear array with a space factor equal to the product of the space factors of two linear arrays."

THEOREM III: "The space factor of a linear array of n apparent elements is the product of $(n - 1)$ virtual couplets with their null points at the zeros of E (eq. 31)."

The *space factor* of an array is defined as the radiation pattern of a similar array of *nondirective* or isotropic elements. The degree of the polynomial which represents an array is always one less than the apparent number of elements. The actual number of elements is at most equal to the apparent number. The total length of the array is the product of the apparent separation and the degree of the polynomial.

Consider a simple two element array in which the currents in the elements are equal in magnitude. The radiation field intensity is represented by

$$E = |1 + z| \tag{12-32}$$

where

$$z = e^{j(\beta d \cos \phi + \alpha)}$$

Making use of Theorem II, a second array can be constructed which will have a radiation pattern that is the square of that given by (32), that is,

$$|E| = |1 + z|^2 = |1 + 2z + z^2|$$

It is seen that the array that will produce this pattern is a three element array having the current ratios

1:2:1

The current in the center element will lead the left-hand element by α, and the current in the right-hand element will lead that in the left-hand element by 2α.

If the polynomial of (32) is raised to the mth power, there results the general binomial array already discussed. When the element spacing d is not greater than $\lambda/2$, such an array produces a pattern with no secondary lobes. However, the principal lobe is considerably broader than that produced by a uniform array having the same number of elements. An array having a narrower principal lobe than that given by the binomial distribution and smaller secondary lobes than that given by the uniform distribution can be obtained by raising the polynomial of the uniform array of n elements (where $n > 2$) to any desired power.

For an n-element uniform array

$$|E| = |1 + z + z^2 + \cdots + z^{n-1}| \tag{12-33}$$

It has already been shown that when n, the number of elements, is large, the ratio of the principal maximum to the first secondary maximum is approximately independent of n and is 13.5 db for the uniform array. If an array is formed to produce a pattern that is the *square* of that given by (33), the ratio of the principal to first secondary maximum will be 27 db. This second array is given by

$$\begin{aligned}|E| &= |1 + z + z^2 + \cdots + z^{n-1}|^2 \\ &= |1 + 2z + 3z^2 + \cdots + nz^{n-1} + (n-1)z^n + \cdots \\ &\qquad + 2z^{2n-3} + z^{2n-2}|\end{aligned} \tag{12-34}$$

The current ratios for this array have the triangular distribution

$$1, 2, 3, \cdots (n-1), n, (n-1), \cdots 3, 2, 1$$

Raising the uniform array to a still higher power would, of course, increase still further the ratio of principal to secondary lobes. The respective patterns for the uniform, binomial, and triangular distribution are shown in Fig. 12-21.

The significance of Theorem III, the decomposition theorem, can be understood by studying the variable z.

where $\qquad z = e^{j\psi} \qquad \psi = \beta d \cos \phi + \alpha$

Since ψ is real, $j\psi$ is a pure imaginary, and the absolute value of z is always unity. Plotted in the complex plane, z is always on the circumference of the unit circle (Fig. 12-22).

As ϕ increases from zero to 180 degrees ψ decreases from $\beta d + \alpha$ to $-\beta d + \alpha$ and z moves in a clockwise direction. Thus the range

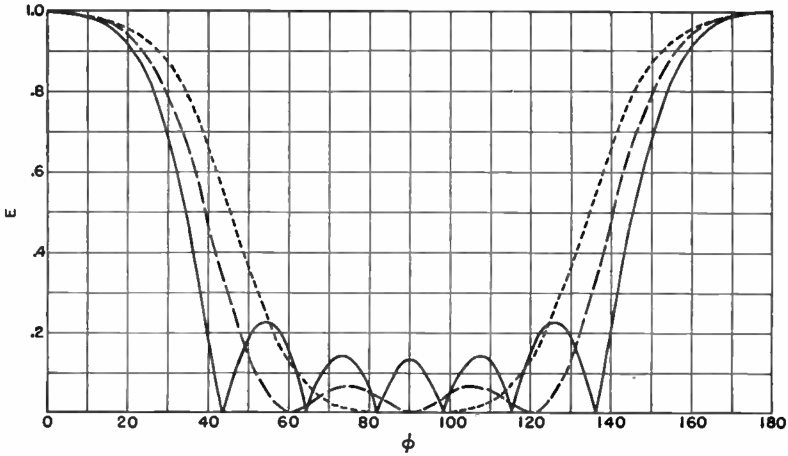


FIG. 12-21. Radiation patterns for uniform (solid), triangular (long dashed), and binomial (short dashed) amplitude distributions. (*Courtesy Bell System Technical Journal.*)

of ψ described by z is $\psi = 2\beta d$ radians. For example, for a separation between elements of $\lambda/4$, ψ varies through π radians as ϕ goes from zero to 180 degrees, and z describes a semicircle. (z retraces its path to the starting point as ϕ goes from 180 degrees to 360 degrees, and the pattern is symmetrical about the 0–180-degree line). For $d = \lambda/2$ the range of ϕ is 2π radians and z describes a complete circle as ϕ varies from zero to 180 degrees. If d is greater than $\lambda/2$, the range of ψ is greater than 2π, and z will overlap itself. The geometrical representation of Fig. 12-23 makes it a simple matter to observe the radiation characteristics as z moves around the circle within its range of operation. For example, for the simple two-element uniform array given by (32), the field intensity is the *sum* $|z + 1|$, which may be written as the difference $|z - (-1)|$.

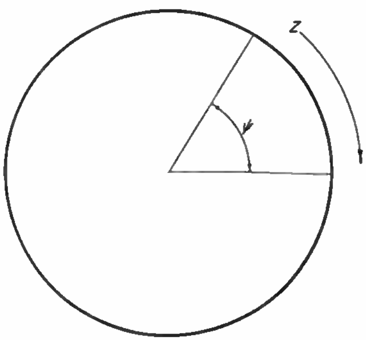
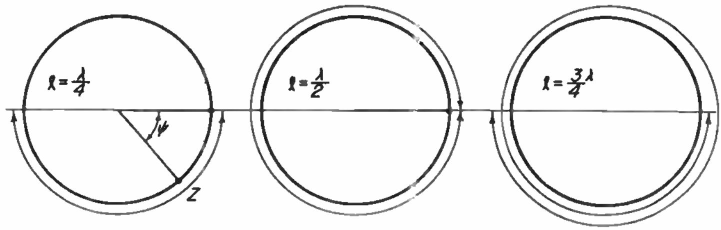


FIG. 12-22. As ϕ increases from 0 degrees to 180 degrees, z moves in a clockwise direction on the unit circle.

This value is given geometrically by the distance between z and the point -1 (Fig. 12-24). For the more general case of unequal



(a) (b) (c)

FIG. 12-23. Active range of z (shown for $\alpha = \beta l$) for a separation between elements of (a) $\lambda/4$, (b) $\lambda/2$, (c) $3\lambda/4$.

amplitudes, where the source intensities are proportional to 1 and $-t$, the radiated field intensity pattern is given by $|z - t|$ which geometrically is the distance between the points z and t. Since z is always on the unit circle, the pattern will have a zero only when t is also on the unit circle, and when t is within the range of z.

By the fundamental theorem of algebra, a polynomial of the $(n - 1)$th degree has $(n - 1)$ zeros (some of which may be multiple zeros) and can be factored into $(n - 1)$ binomials. Thus

$$|E| = |(z - t_1)(z - t_2) \cdots (z - t_{n-1})| \quad (12\text{-}35)$$

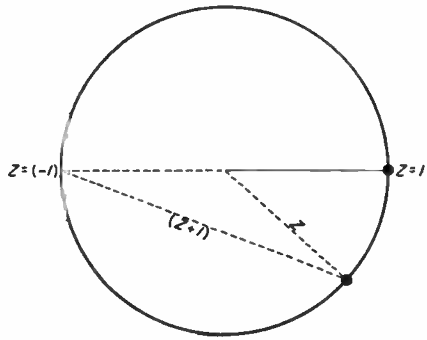


FIG. 12-24

from which Theorem III follows directly.

It is evident that the radiation intensity in any direction is given by the products of the distances from z (corresponding to the chosen direction) to the null points of the array.

EXAMPLE 1: *Uniform Array.* Consider the case of the uniform array that is represented by

$$|E| = |1 + z + z^2 + \cdots + z^{n-1}| = \left|\frac{1 - z^n}{1 - z}\right| = \left|\frac{z^n - 1}{z - 1}\right| \quad (12\text{-}36)$$

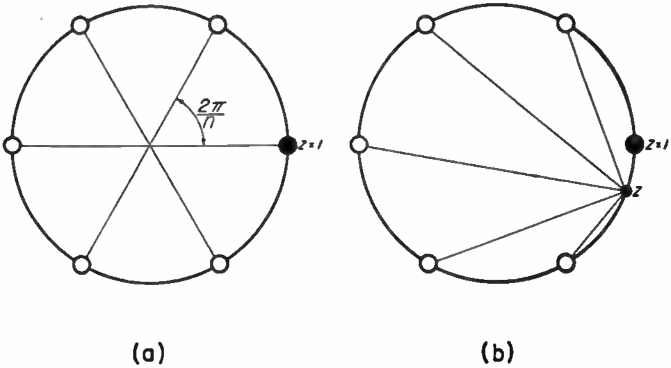


FIG. 12-25

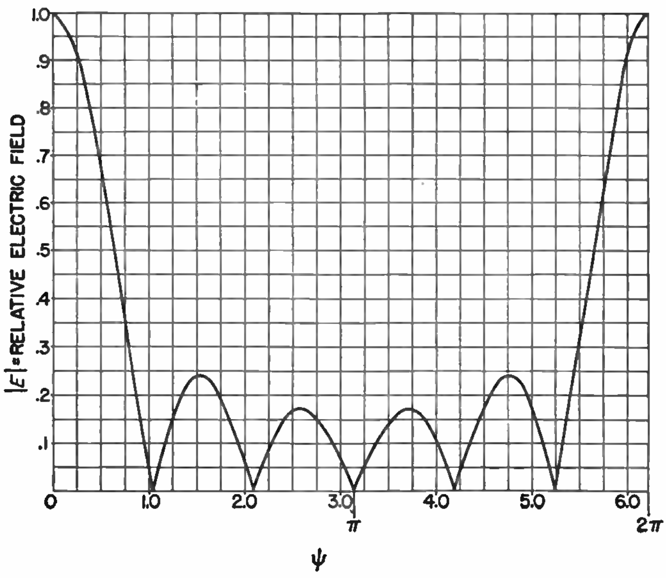


FIG. 12-26. Relative field intensity E as a function of ψ.

The null points of such an array, given by the roots of (36), are in this case the nth roots of unity (excluding $z = 1$, which is the principal maximum). In the complex plane the roots of unity all lie on the unit circle, and divide the circle into n equal parts (Fig. 12-25). The roots are

$$z = e^{-j(2\pi/n)}, e^{-j2(2\pi/n)}, \cdots e^{-jm(2\pi/n)}$$

It is seen that the null points of the array are given by $\psi_m = -m\,2\pi/n$

where $m = 1, 2, 3, \cdots (n-1)$. Since $\psi = \beta d \cos \phi + \alpha$, the null points of the radiation pattern are given in terms of the angle ϕ by

$$\cos \phi_m = -\frac{\alpha}{\beta d} - \frac{2\pi m}{n\beta d}$$

When $z = 1$, E has a principal maximum. Other maxima occur approximately midway between the nulls. As z moves around the circle the

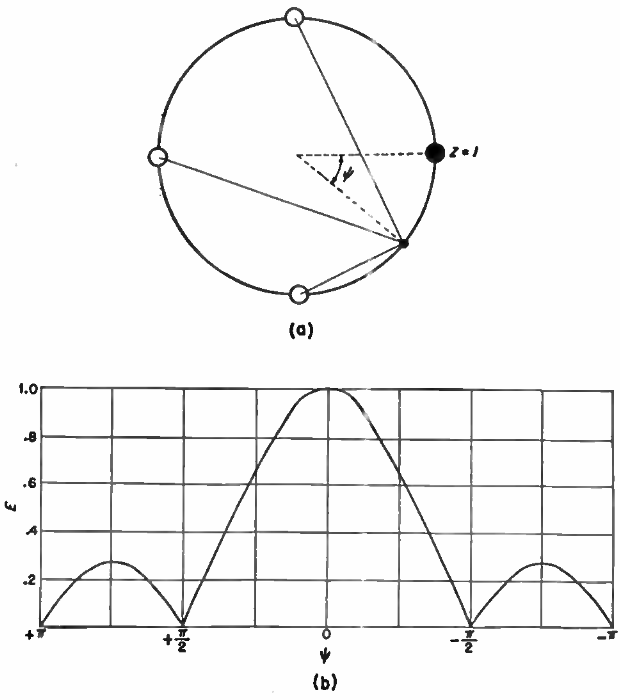


FIG. 12-27

radiation pattern is given by the product of the lines connecting the null points to z. A plot of E as a function of ψ is shown in Fig. 12-26. Using

$$\phi = \cos^{-1}\left(\frac{\psi - \alpha}{\beta d}\right)$$

E can be drawn as a function of ϕ.

EXAMPLE 2: *Four-element Broadside.* A simple array, which has already been considered is the four-element broadside having half-wave spacing between elements and equal currents fed in phase. For this case

$$\beta d = \pi \qquad \alpha = 0 \qquad \psi = \pi \cos \phi$$

The range of z is $\psi = 2\beta d = 2\pi$.

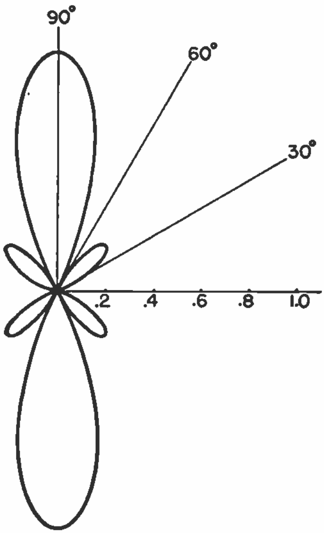


FIG. 12-28. Relative field intensity as a function of ϕ.

The relative field intensity pattern is given by

$$\begin{aligned} E &= |1 + z + z^2 + z^3| \\ &= \left|\frac{z^4 - 1}{z - 1}\right| \\ &= |(z - e^{-j(\pi/2)})(z - e^{-j\pi})(z - e^{-j(3\pi/2)})| \end{aligned} \tag{12-37}$$

The nulls are spaced equally on the unit circle as shown in Fig. 12-27a. As ϕ increases from 0 to 180 degrees, ψ decreases from π through zero to $-\pi$ and the curve of Fig. 12-27b results. This is plotted in polar co-ordinates as a function of ϕ in Fig. 12-28.

EXAMPLE 3: *Four-element End Fire.* Consider a uniform four-element end-fire array having an element spacing of one-quarter wavelength and a progressive phase shift of $-\pi/2$ radians. For this array

$$\beta d = \frac{\pi}{2} \qquad \alpha = -\frac{\pi}{2}$$

and

$$\psi = \beta d \cos \phi + \alpha = \frac{\pi}{2}(\cos \phi - 1)$$

The range of ψ is π radians.

As before, the expression for $|E|$ is given by eq. (37) and the three nulls occur at $\psi = -\pi/2, -\pi, -3\pi/2$. However, in this case the range of ψ is only from $\psi = 0$ to $\psi = -\pi$ (Fig. 12-29), so the null at $-3\pi/2$ obviously has very little effect on the pattern. An improved pattern (that is, one with a narrower principal lobe and smaller secondary lobes) can be obtained with the same number of elements by spacing the nulls equally in the range of ψ. This gives rise to the array that has the circle diagram of Fig. 12-30 and the pattern given by

$$\begin{aligned} E &= |(z - e^{-j(\pi/3)})(z - e^{-j(2\pi/3)})(z - e^{-j\pi})| \\ &= |z^3 + (1 - e^{-j(\pi/3)} - e^{-j(2\pi/3)})z^2 + (-1 - e^{-j(\pi/3)} - e^{-j(2\pi/3)})z - 1| \\ &= |1 + 2\,e^{-j(\pi/3)}z + 2\,e^{-j(2\pi/3)}z^2 + e^{-j\pi}z^3| \end{aligned} \tag{12-38}$$

The current amplitudes of the array are

$$1:2:2:1$$

and the progressive phase shift between elements will be $-\pi/2 - \pi/3 = -5\pi/6$ radians. The resultant pattern as a function of ϕ is shown as curve B in Fig. 12-31.

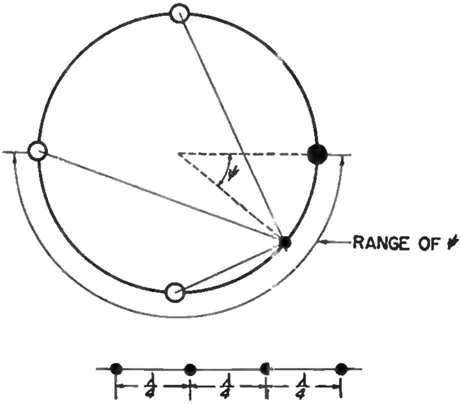


Fig. 12-29

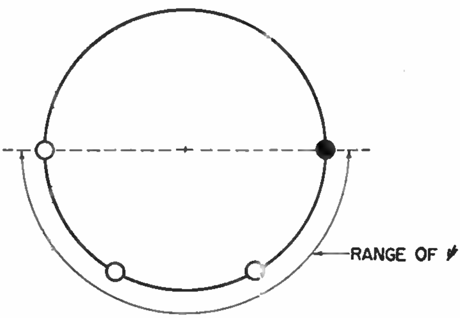


Fig. 12-30. Circle diagrams for a four-element array having nulls equispaced in the range of ψ. For an element spacing of one-quarter wavelength the range of ψ is π radians.

If the over-all length of the array is maintained constant, but the number of elements is increased, it is possible to improve the directivity still further if the nulls are properly spaced in the range of operation. Curve C of Fig. 12-31 shows the pattern that results when the number of elements is increased to seven with the spacing reduced to one-eighth wavelength, so that the over-all length is still $\frac{3}{4}\lambda$. To obtain this result the nulls were equispaced in the range $\psi = 2\beta d = \pi/2$. Curve D shows the pattern obtained for 13 elements at $\lambda/16$ spacing, with the nulls again equispaced in the range of ψ.

For the uniform array it was found that the maximum directivity and gain obtainable were directly related to the length of the array. In contrast to this, when the current ratios and phasings

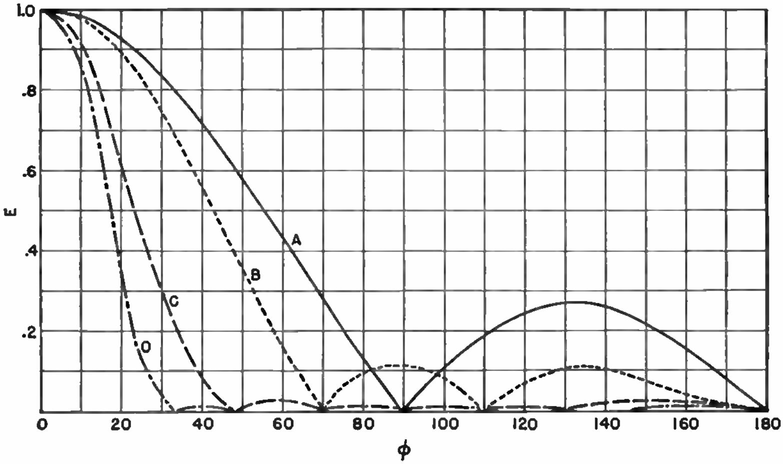


FIG. 12-31. Radiation patterns for several end-fire arrays, all having an array length of $3\lambda/4$. Shown at A is four-element uniform array ($d = \lambda/4$); B, four elements with nulls equispaced in the range of ψ ($d = \lambda/4$); C, seven elements with nulls equispaced in the range of ψ ($d = \lambda/8$); D, thirteen elements with nulls equispaced in the range of ψ ($d = \lambda/16$). (*Courtesy Bell System Technical Journal.*)

are properly chosen, it appears possible to obtain arbitrarily sharp directivity with an array of fixed length by using a sufficiently large number of elements. However, it will be found also that with the phase relations and close spacings between the elements required to obtain this result, the radiation resistance is reduced to extremely low values. That is, extremely large currents are required to produce fields of appreciable intensity. With actual antennas that have a finite ohmic resistance, the antenna efficiency enters the picture to limit the directivity and gain that can be obtained from an array of given length.

Problem 2(a). Draw the circle diagram and sketch the pattern of a three-element, uniform, end fire array, using $d = \lambda/2$. (b) Using the same number of elements and same spacing, redesign the array to have nulls at $\phi = 90$ degrees and $\phi = 60$ degrees.

12.15 Antenna Synthesis. It is a simple and straight forward job to compute the radiation pattern of an array having specified configuration and antenna currents. A somewhat more difficult problem is the design of an array to produce a prescribed radiation pattern. Making use of Fourier analysis, the methods of the preceding sections may be extended to accomplish this result.

It is convenient to consider an array having an odd number of elements with a certain symmetry of current distribution about the center element. The polynomial for an array with $n = 2m + 1$ elements is

$$|E| = |A_0 + A_1z + A_2z^2 + A_mz^m + A_{m+1}z^{m+1} + A_{2m}z^{2m}| \quad (12\text{-}39)$$

Now the absolute value of z is always unity, so equation (39) can be divided by z^m without changing the value of $|E|$. That is,

$$|E| = |A_0z^{-m} + A_1z^{-m+1} + \cdots + A_{m-1}z^{-1} + A_m + A_{m+1}z + \cdots + A_{2m}z^m| \quad (12\text{-}40)$$

It is now specified that the currents in corresponding elements on either side of the center element be equal in magnitude, but that the phase of the left-side element shall lag that of the center element by the same amount that the corresponding right-side element leads the center element (or vice versa). That is, the coefficients of corresponding elements are made complex conjugates with

$$A_m = a_0 \qquad A_{m-k} = a_k - jb_k \qquad A_{m+k} = a_k + jb_k$$

Then the sum of terms of two corresponding elements may be written

$$\begin{aligned} A_{m-k}z^{-k} + A_{m+k}z^k &= a_k(z^{+k} + z^{-k}) + jb_k(z^k - z^{-k}) \\ &= 2a_k \cos k\psi - 2b_k \sin k\psi \end{aligned}$$

since

$$z^k = e^{jk\psi}$$

The expression for $|E|$ is now

$$\begin{aligned} |E| &= 2[\tfrac{1}{2}a_0 + a_1 \cos \psi + \cdots + a_m \cos m\psi \\ &\qquad - (+b_1 \sin \psi + \cdots + b_m \sin m\psi)] \\ &= 2\left\{\frac{a_0}{2} + \sum_{k=1}^{k=m} [a_k \cos k\psi + (-b_k) \sin k\psi]\right\} \end{aligned} \quad (12\text{-}41)$$

These are the first $2m + 1$ terms of a Fourier series in which the coefficients of the cosine terms are the a_k's, and the coefficients of the sine terms are $(-b_k$'s). Now any radiation pattern specified as a function $f(\psi)$ may be expanded as a Fourier series with an infinite number of terms. Such a pattern may be approximated to any desired accuracy by means of the finite series (41). When this is done the required current distribution of the array can be written down directly.

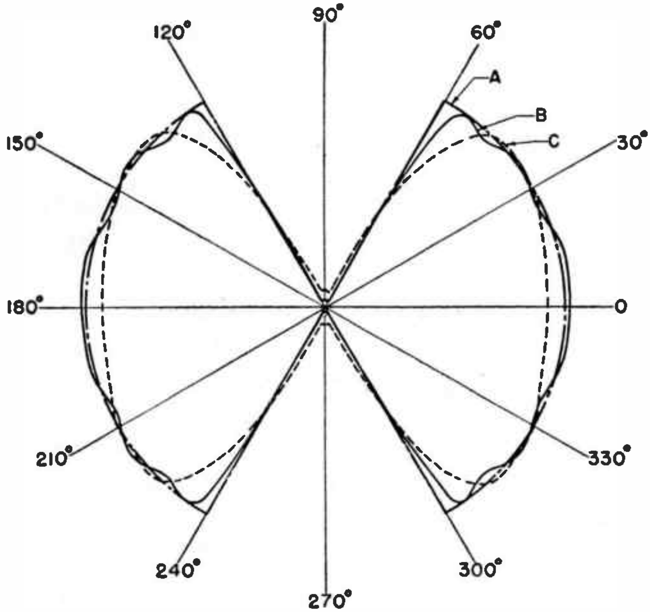


FIG. 12-32. A prescribed pattern, A, and approximations to it, obtained with an eleven-element array, B, and a five-element array, C.

EXAMPLE 4: *Synthesized Bidirectional Array.* Let it be required to design an array that will produce approximately a pattern of Fig. 12-32. This pattern is defined by

$$f(\phi) = 1 \qquad 0 < \phi < \frac{\pi}{3}$$

$$f(\phi) = 0 \qquad \frac{\pi}{3} < \phi < \frac{2\pi}{3}$$

$$f(\phi) = 1 \qquad \frac{2\pi}{3} < \phi < \pi$$

It will, of course, be symmetrical about the line of the array, $\phi = 0$. If the spacing is chosen to be $\lambda/2$, then $\psi = \pi \cos \phi + \alpha$. The corresponding ψ function is

$$F(\psi) = 1 \qquad \pi + \alpha > \psi > \frac{\pi}{2} + \alpha$$

$$F(\psi) = 0 \qquad \frac{\pi}{2} + \alpha > \psi > -\frac{\pi}{2} + \alpha$$

$$F(\psi) = 1 \qquad -\frac{\pi}{2} + \alpha > \psi > -\pi + \alpha$$

Choosing $\alpha = -\pi$ for an end fire array results in the function shown in Fig. 12-33.

The Fourier series expansion for this function is

$$F(\psi) = \left(\frac{1}{2} + \frac{2}{\pi} \sum_{k=1}^{k=\infty} \frac{1}{k} \sin \frac{k\pi}{2} \cos k\psi\right)$$

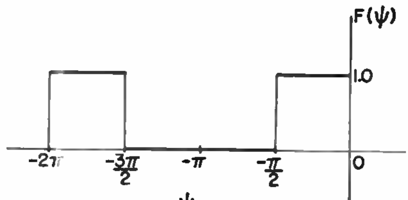


Fig. 12-33

Comparison with (41) determines the coefficients

$$a_0 = \frac{1}{2}$$

$$a_k = \frac{1}{k\pi} \sin \frac{k\pi}{2}$$

$$b_k = 0 \qquad\qquad k \neq 0$$

The pattern obtained using the value of $m = 4$, is given from eq. (40) as

$$|E| = \frac{1}{\pi}\left|-\frac{1}{3}z^{-3} + z^{-1} + \frac{\pi}{2} + z - \frac{1}{3}z^3\right| \qquad (12\text{-}42)$$

This is a five-element array having the current ratios indicated and an over-all length of three wavelengths (the apparent spacing between elements is one-half wavelength, but four of the elements are missing). The pattern produced by this array is shown in Fig. 12-32. Also shown in this figure is the pattern obtained with an 11-element array formed using $m = 9$ in the series.

In the above example the apparent element spacing was arbitrarily chosen as one-half wavelength, which made the range of ψ equal to 2π radians. If the element spacing is less than $\lambda/2$, the range of ψ will be less than 2π radians. This means that although the radiation pattern as a function of ϕ, that is $f(\phi)$, is completely specified for the whole range of ϕ, the corresponding $f(\psi)$ is specified only over its range, which is less than the interval of 2π radians required for the Fourier expansion. It is possible then to complete the interval with any function that satisfies Dirichlet's conditions. Naturally, the function chosen would be one which would simplify the series as much as possible or make it converge rapidly. It is evident that when the apparent spacing is less than $\lambda/2$ there is an unlimited number of solutions that will satisfy the conditions of the problem. If the apparent spacing of the elements is greater than $\lambda/2$, the range of ψ is more than 2π radians. Except for some special cases* it is then not possible to obtain the prescribed directional pattern by this method.

When an apparent spacing less than $\lambda/2$ is used, $f(\psi)$ is specified over only a portion of the required 2π radians, and the function used to fill in the remainder of the interval can be chosen at will by the designer. A judicious choice of "fill-in" function will produce a desirable pattern with a minimum number of elements, and conversely a poor choice of function may result in a poor pattern. An example will illustrate this point.

Example 5: *Synthesized Unidirectional Array.* Let it be required to design an end fire array that will have an approximately semicircular pattern given by

$$f(\phi) = 1 \qquad 0 < \phi < \frac{\pi}{2}$$

$$f(\phi) = 0 \qquad \frac{\pi}{2} < \phi < \pi$$

The apparent spacing is to be $\lambda/4$.

Then, for this problem,

$$\psi = \frac{\pi}{2}\cos\phi + \alpha$$

and the range of ψ is π radians. By choosing different values of α, the range of ψ which is used can be shifted anywhere in the interval of 2π radians, which is required for the Fourier expansion. This is shown in

* Examples of cases where a larger spacing is permissable are given in the following article: Irving Wolff, "Determination of the Radiating System Which will Produce a Specified Directional Characteristic," *Proc. IRE*, **25**, 5, 630 (1937).

Fig. 12-34 for three values of α. In this example there is a finite discontinuity within the range of ψ, so the coefficients of the series will decrease at a rate that is of the order of $1/n$. (They will be less in absolute magnitude than c/n, where c is some positive constant.*) If the functions were

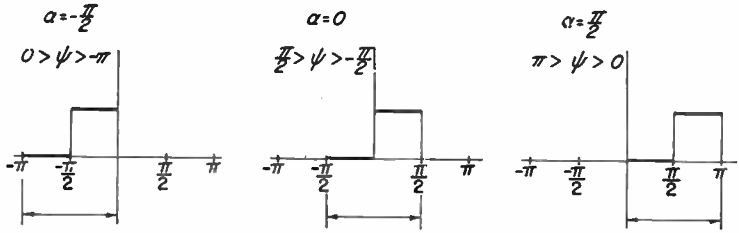


FIG. 12-34. Range of ψ for $0 < \phi < \pi$, for three different values of α. (Range used is indicated by double arrows.)

continuous in the range, the series would converge at a rate that would be at least of the order of $1/n^2$. Because no choice of fill-in function can remove this discontinuity within the range of ψ, it is anticipated that, in this case, the fill-in function may not have much effect on the number of terms required. However, it is interesting to examine some actual cases.

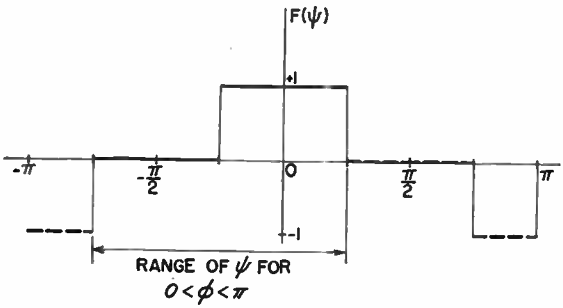


FIG. 12-35. A possible choice of fill-in function (shown dashed) for $\alpha = -\pi/4$.

CASE 1: A possible choice for α and for the fill-in function is illustrated in Fig. 12-35. This choice would appear to be good, because it results in the following conditions:

(1) $f(\pi + \psi) = -f(\psi)$. Therefore, only odd harmonics will be present.

(2) $f(\psi)$ is an even function, so the coefficients of the sine terms will be zero.

(3) $f(\psi)$ has an average value of zero, so the d-c term (or center element) is eliminated.

* Doherty and Keller, *Mathematics of Modern Engineering*, John Wiley and Sons, New York, 1936, Vol. 1, p. 89.

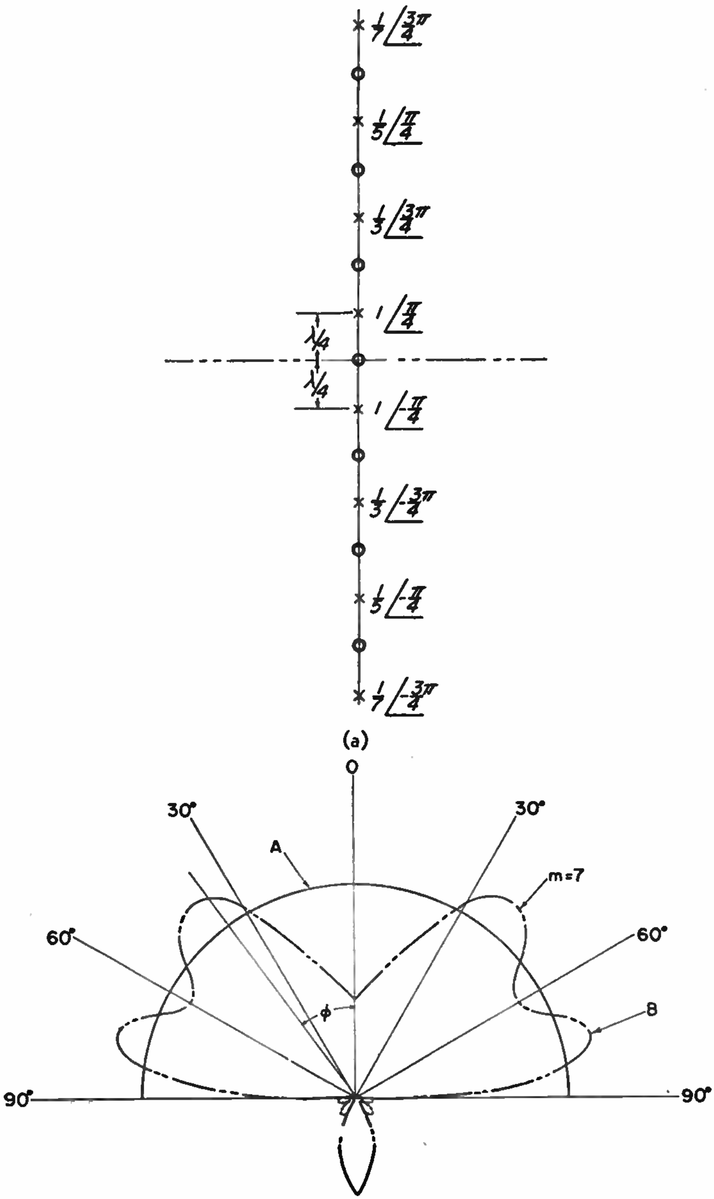


Fig. 12-36. (a) Array and (b) pattern corresponding to Fig. 12-35. (Circles in (a) indicate elements which drop out because of zero current.)

The antenna array resulting from the choice used in Fig. 12-35 is shown in Fig. 12-36a for $m = 7$ (eight elements), and the corresponding pattern is shown in Fig. 12-36b. This pattern has two serious defects. It approaches a relative value of 0.5 at $\phi = 0$ where it should be unity, and it also approaches 0.5 at $\phi = \pi$ where it should be zero. Using more terms of the series will not remedy these defects, which are inherent in the particular function used in Fig. 12-35. This function is discontinuous at the values

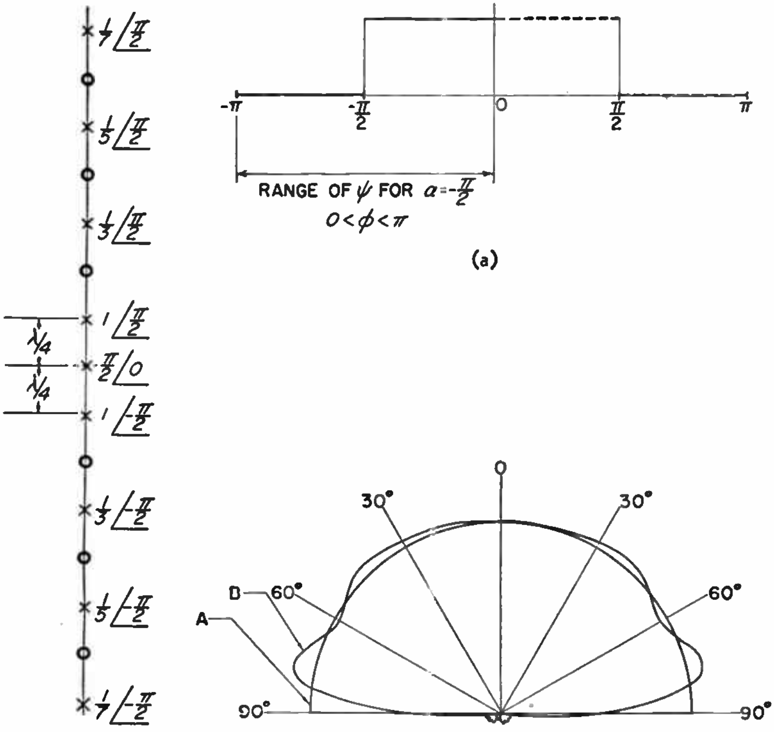


Fig. 12-37. A better design (a) results in the array (b) and the pattern (c).

of ψ corresponding to $\phi = 0$ and $\phi = \pi$, and the series converges to the average of the values taken by the function on the two sides of the discontinuity. Therefore, the function of Fig. 12-35 is an unsuitable choice.

Case 2: The discontinuities at values of ψ corresponding to $\phi = 0$ and $\phi = \pi$ can be eliminated by a different choice of α, and a suitable fill-in function. A possible function is that shown in Fig. 12-37(a).
The corresponding antenna array and resulting pattern are also shown, and it is seen that this function is a suitable one.

Whereas many other types of fill-in functions are possible, it is found for this example, where the apparent spacing is fixed at $\lambda/4$, that none of them results in appreciable improvement over the design of Fig. 12-37. However, if the apparent spacing d is permitted to have other values, this puts one more variable under the control of the designer. For a given number of antenna elements it is in general possible, with this additional control, to improve the pattern obtained. In the present example for a given number of elements, an apparent spacing of $3\lambda/8$ instead of $\lambda/4$ results in a closer approach to the ideal pattern in the critical regions, $\phi = \pm\pi/2$.

12.16 The Tchebyscheff Distribution. A particular, but very important, problem in antenna synthesis is the following: For a given linear antenna array, determine the current ratios that will result in the narrowest main lobe, for a specified side-lobe level; or, in other words, determine the current ratios that will result in the smallest side-lobe level for a given beam-width of the principal lobe. The current distribution that produces such a pattern will be considered as being the optimum.

From the material of section 14, it will be recalled that a desirable pattern (but not necessarily the optimum) can be obtained by equispacing the nulls on the appropriate arc of the unit circle. An examination of Fig. 12-26, which shows a pattern obtained by equispacing the nulls, indicates how a better pattern can be obtained. For a given width of principal lobe, the first secondary lobe can be decreased by moving the second null closer to the first. Of course, this increases the second side lobe, but that is permissible as long as it does not exceed the first. It is evident that the *optimum* pattern is obtained when all the side lobes have the same level. The problem is simply that of finding the spacing of nulls which makes this true. The answer is given in terms of the Tchebyscheff polynomials.

The *Tchebyscheff polynomials* occur quite frequently in design and synthesis problems. They are defined* by

$$T_m(x) = \cos\,(m \cos^{-1} x) \qquad -1 < x < +1$$
$$T_m(x) = \cosh\,(m \cosh^{-1} x) \qquad |x| > 1$$

The general shape of $T_m(x)$ is shown in Fig. 12-38 for both m even and m odd.

* Courant-Hilbert, *Methoden der Mathematischen Physik*, Julius Springer, Berlin, 1931. Vol. 1, p. 75.

By inspection, $T_0(x) = 1, \quad T_1(x) = x$

The higher order polynomials can be derived as follows:

$$T_2(x) = \cos (2 \cos^{-1} x) = \cos 2\delta$$

where $\delta = \cos^{-1} x \quad \text{or} \quad x = \cos \delta$

Now since
$$\cos 2\delta = 2 \cos^2 \delta - 1$$
$$T_2(x) = 2x^2 - 1$$

Similarly, it can be shown that

$$T_{m+1}(x) = 2T_m(x)T_1(x) - T_{m-1}(x)$$

so that
$$T_3(x) = 4x^3 - 3x$$
$$T_4(x) = 8x^4 - 8x^2 + 1$$

and so on.

The important characteristic of the Tchebyscheff polynomials, as far as antenna pattern synthesis is concerned, is evident from Fig. 12-38. As x is allowed to vary from some point c up to a value x_0 and then back to its starting point, the function $T_m(x)$ traces out a pattern consisting of several small side lobes and one major lobe. The secondary lobes will all be of equal amplitude (unity) and will be down from the main lobe by the ratio $1/b$. This ratio can be chosen at will by suitable choice of x_0. Such a pattern will be called the optimum or Tchebyscheff pattern.* Since a technique for obtaining the pattern is available once the positions of the nulls on the unit circle are known (section 14), all that is required from the Tchebyscheff polynomials is information on the proper distribution of the nulls. This information can be obtained by causing x to trace out the desired portion of the Tchebyscheff polynomial (of correct degree) as the variable ψ moves over its range on the unit circle. This is accomplished as follows:

Consider the Tchebyscheff polynomial of mth degree

$$T_m(x) = \cos (m \cos^{-1} x) = \cos (m\delta)$$

where
$$\cos \delta = x$$

The nulls of the pattern are given by the roots

$$\cos (m\delta) = 0$$

* C. L. Dolph, "A Current Distribution for Broadside Arrays which Optimizes the Relationship between Beam Width and Side-lobe Level," *Proc. IRE*, **34**, 6, 335 (1946); *also* H. J. Riblet, *Proc. IRE*, **35**, 5, 489 (1947).

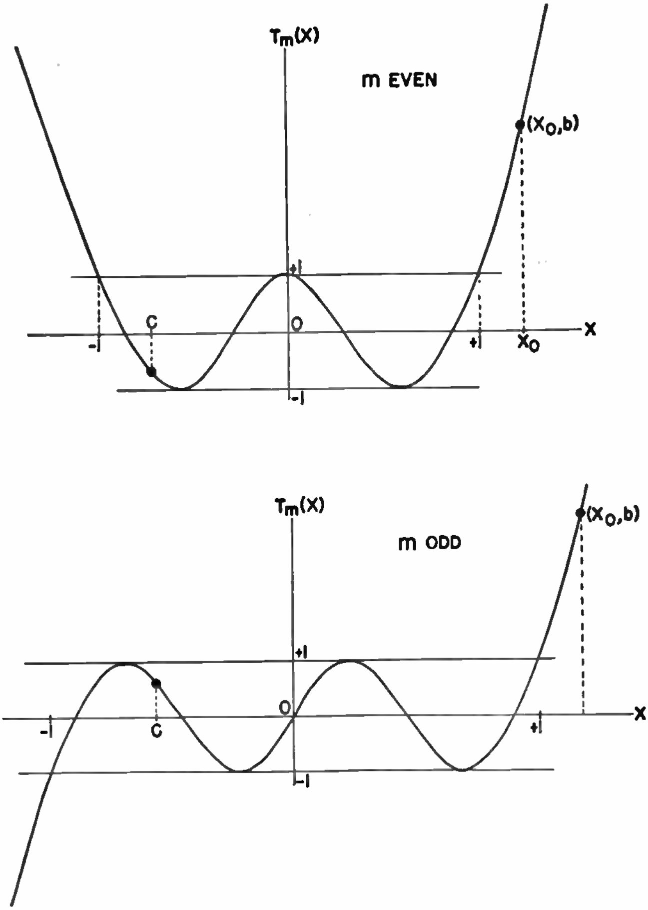


FIG. 12-38. Tchebyscheff polynomials, $Tm(x)$, for m even and m odd.

that is, by

$$\delta_k^0 = \frac{(2k - 1)\pi}{2m} \qquad k = 1, 2, \cdots m$$

Next consider the function ψ. For a broadside array for which $\alpha = 0$,

$$\psi = \beta d \cos \phi$$

as ϕ varies from 0 to $\pi/2$ to π, ψ goes from βd to 0 to $-\beta d$ and the range of ψ is $2\beta d$.

Now let $x = x_0 \cos \psi/2$. Then, as ϕ varies from 0 through $\pi/2$ to π, ψ varies from βd through zero to $-\beta d$, and x will vary from $x_0 \cos \pi d/\lambda$ to x_0 back to $x_0 \cos(-\pi d/\lambda) = x_0 \cos \pi d/\lambda$. For example, if $d = \lambda/2$, ψ will range from π through zero to $-\pi$, and x will range from 0 to $+x_0$ and back to zero. Again, if $d = \lambda$, ψ will range twice around the circle from 2π through 0 to -2π, (two major lobes) and x will range from $-x_0$ to x_0 and back to $-x_0$. This is the correspondence desired.

The nulls in the Tchebyscheff pattern occur at values of x given by

$$x_k{}^0 = \cos \delta_k{}^0$$

so the corresponding position for the nulls on the unit circle will be given by

$$x_k{}^0 = x_0 \cos \frac{\psi_k{}^0}{2}$$

or

$$\psi_k{}^0 = 2 \cos^{-1}\left[\frac{x_k{}^0}{x_0}\right]$$

$$= 2 \cos^{-1}\left[\frac{\cos \delta_k{}^0}{x_0}\right] \tag{12-43}$$

where

$$\delta_k{}^0 = \frac{(2k-1)\pi}{2m} \qquad k = 1, 2, \cdots m$$

Equation (43) gives the required spacing of the nulls on the unit circle for a pattern whose side lobes are all equal. The degree m of the polynomial used will be equal to the number of nulls on the unit circle, and this will be one less than n, the apparent number of elements. The value of x_0 is determined by the desired ratio b of principal to side lobe amplitudes. The value of x_0 is given in terms of b by

$$T_m(x_0) = b$$

It can be obtained more conveniently* from

$$x_0 = \tfrac{1}{2}[(b + \sqrt{b^2 - 1})^{1/m} + (b - \sqrt{b^2 - 1})^{1/m}]$$

The graphical-analytical method for obtaining the pattern of the array from the location of the nulls yields a detailed and accurate

* C. L. Dolph, *loc. cit.* (discussion), p. 492.

plot of relative field strength versus the defined angle ψ. For many purposes a rough sketch of the pattern may be adequate, and this can be obtained directly from the known properties of the Tchebyscheff polynomial. Thus, knowing the location of the nulls in the pattern and the amplitude of all the side lobes relative to the principal lobe, the pattern as a function of ψ may be sketched in with good accuracy. The pattern as a function of the azimuthal angle ϕ is then determined using the transform $\phi = \cos^{-1} \psi/\beta d$. The binomial expansion method of calculating the required current distribution from the location of the nulls proves satisfactory for small arrays, but tends to become unwieldy for larger arrays. For large multielement arrays, an alternative procedure outlined in the article by Dolph will be found to require less labor.

EXAMPLE 6: Design on a four-element broadside array having a spacing $d = \lambda/2$ between elements. The pattern is to be optimum with a side lobe level, which is 19.1 db down ($b = 9.0$).

For $d = \lambda/2$, the range of operation is $2\beta d = 2\pi$. Since there will be 3 nulls, use $T_3(x) = 4x^3 - 3x$. Then

$$T_3(x_0) = 4x_0^3 - 3x_0 = 9$$

Solving for x_0,

$$\begin{aligned} x_0 &= \tfrac{1}{2}[(b + \sqrt{b^2 - 1})^{1/m} + (b - \sqrt{b^2 - 1})^{1/m}] \\ &= \tfrac{1}{2}[(9 + \sqrt{80})^{1/3} + (9 - \sqrt{80})^{1/3}] \\ &= 1.5 \end{aligned}$$

The nulls are given by $\cos (m\delta) = \cos (3\delta) = 0$. Therefore

$$\delta_k{}^0 = \frac{(2k-1)\pi}{2m} = \frac{(2k-1)\pi}{6} \qquad k = 1, 2, 3, \cdots$$

Then $\qquad \delta_1{}^0 = \frac{\pi}{6} \qquad \delta_2{}^0 = \frac{3\pi}{6} \qquad \delta_3{}^0 = \frac{5\pi}{6}$

k	$\delta_k{}^0$	$x_k{}^0 = \cos \delta_k{}^0$	$x_k{}^0/x_0$	$\psi_k{}^0 = 2 \cos^{-1} \frac{x_k{}^0}{x_0}$	$\psi_k{}^0$ (radians)
1	$\pi/6$	0.866	0.577	109.5°	1.910
2	$3\pi/6$	0	0	180°	π
3	$5\pi/6$	−0.866	−0.577	250.5°	4.37

The polynomial representing the array is

$$|E| = |(z - e^{j1.91})(z - e^{j\pi})(z - e^{j4.37})|$$
$$= |z^3 + 1.667z^2 + 1.667z + 1|$$

The required relative currents in the elements are

$$1:1.667:1.667:1$$

12.17 Supergain Arrays. In the case of end fire arrays, it was found that the technique of equispacing the nulls in the range of ψ yielded desirable radiation patterns, even when the spacing between elements became small in wavelengths. Indeed, if the number of elements in the array was increased as the spacing was decreased, so that the overall length of the array remained fixed, this technique led to the design of arrays having arbitrarily sharp directivity. Such arrays, which are capable (theoretically) of attaining very high gains with reasonably small dimensions, have become known as *supergain* arrays.

When this same technique of equispacing the nulls is applied to broadside arrays having small spacings, it is found that, as the spacings are made smaller, the patterns become progressively poorer. However, if the nulls are distributed in the range of ψ according to the Tchebyscheff distribution, desirable patterns having small side lobes and arbitrarily sharp principal lobes result. The design procedure for supergain broadside arrays is indicated below.

Referring to Fig. 12-38, the range of the Tchebyscheff polynomial, which is used, lies between the points c and x_0. The position of the starting point c depends upon the element spacing and is given by

$$c = x_0 \cos \frac{\pi d}{\lambda}$$

For a spacing d equal to $\lambda/2$, the point c is at the origin. For $d > \lambda/2$, c is negative as shown in Fig. 12-38, whereas for spacings less than $\lambda/2$, c is positive, approaching x_0 as the spacing approaches zero. Since the radiation pattern is determined entirely by that portion of the Tchebyscheff curve lying between c and x_0, it is seen that for small spacings (c near x_0) full use is not being made of the pattern control available. This failing can be remedied by compressing the desired range of Tchebyscheff curve into the region that will be used. This can be accomplished by a simple change of scale on the abscissa. An example will illustrate the method.

EXAMPLE 7: Design a five-element broadside array having a total array length of $\lambda/4$ (spacing between elements $d = \lambda/16$). The side lobe level is to be 25.8 db down, or 1/19.5 times the main lobe level.

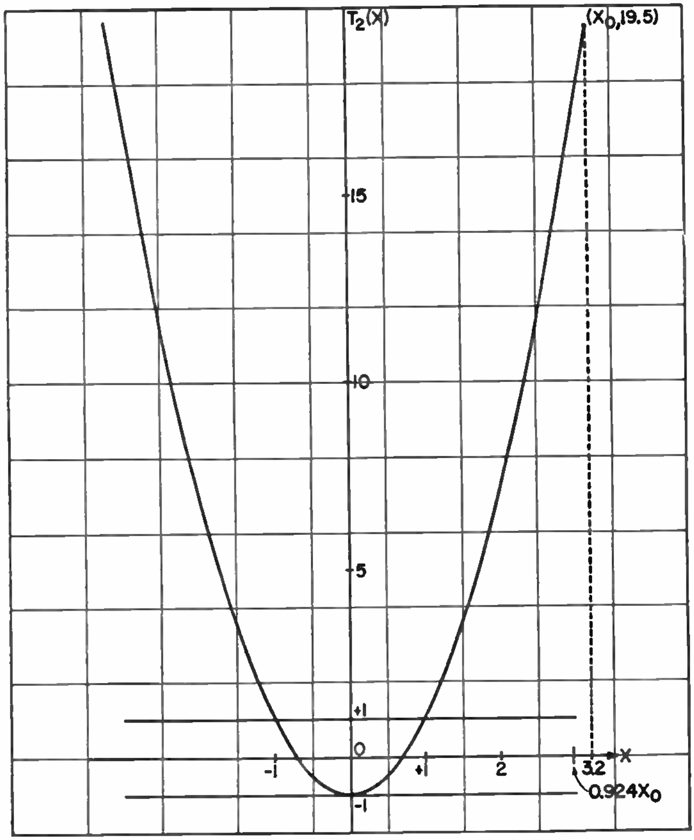


FIG. 12-39a. Second-degree Tchebyscheff polynomial.

For a five-element array there should be four nulls on the unit circle within the range of ψ, and four nulls in the range of the Tchebyscheff pattern that is used. Thus it would be possible to select $T_4(x)$ and use the range from 0 to x_0 and back (four nulls), or alternatively it is possible to select $T_2(x)$ and use the whole range from -1 to x_0 and back (again four nulls). Since it is considerably simpler to work with the lower degree polynomials, $T_2(x)$ will be used.

A sketch of $T_2(x)$ is shown in Fig. 12-39a. Let $x = x_0 \cos \psi$, where for this case of $d = \lambda/16$,

$$\psi = \frac{\pi}{8} \cos \phi$$

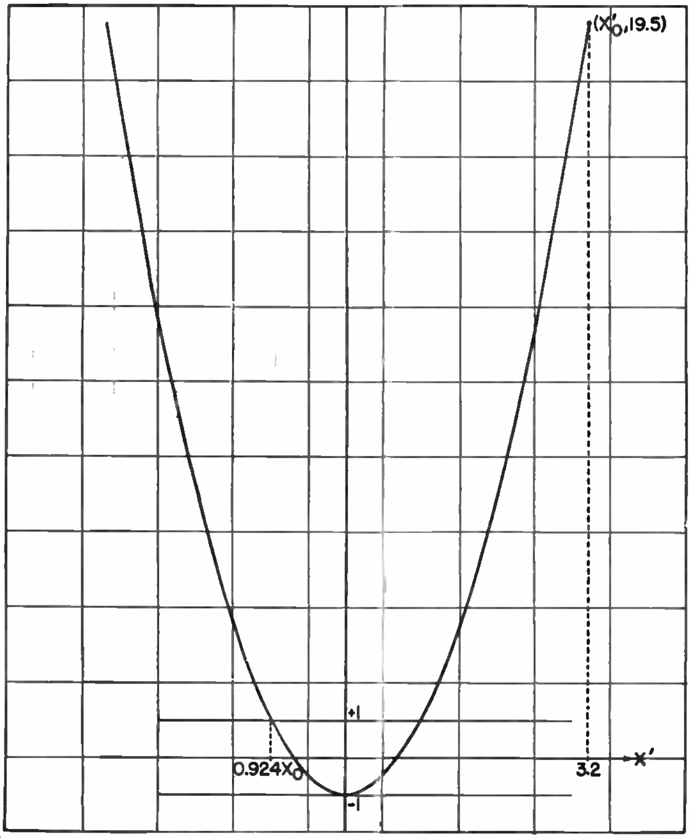


Fig. 12-39b

Then as ϕ ranges from 0 through $\pi/2$ to π, ψ will range from $\pi/8$ to $-\pi/8$, and since $\cos \pi/8 = 0.92388$, x will vary from $0.92388x_0$ to x_0 and back to $0.92388x_0$. By a simple translation and change of scale of the abscissa, the portion of the curve between -1 and x_0 can be compressed within the range $0.92388x_0$ to x_0. Thus let

$$x = ax' + b$$

so that
$$3.20156 = a(3.20156) + b$$
and
$$-1.0 = a(0.92388 \times 3.20156) + b$$
Solving
$$a = 17.236 \qquad b = -51.982$$

The curve now appears as in Fig. 12.39b, where

$$x' = \frac{x - b}{a} = \frac{x + 51.982}{17.236}$$

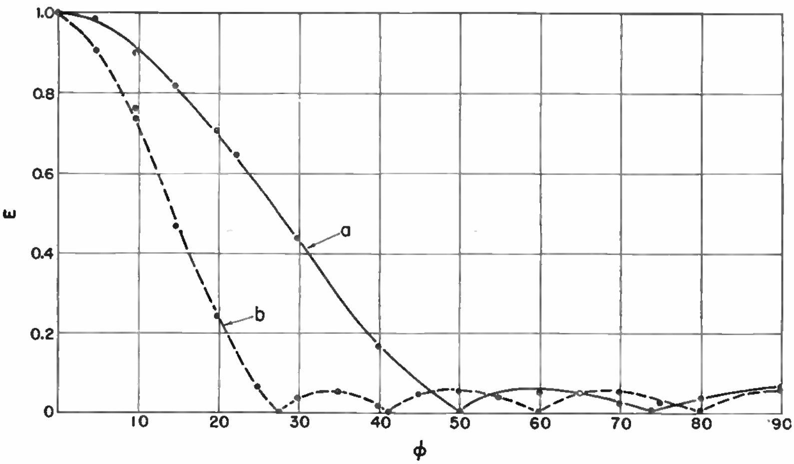


FIG. 12-40. Broadside "super-gain" patterns for arrays that have an overall length of one-quarter wavelength: *a*, five-element array; *b*, nine-element array.

The nulls in Fig. 12.39a occur at $x^0 = \pm 0.707$, so the nulls in (b) will occur at

$$x^{0\prime} = \frac{x^0 + 51.982}{17.236} = 3.0569 \qquad \text{or} \qquad 2.9749$$

Correspondingly, the nulls on the unit circle will be place at

$$\begin{aligned} \psi^0 &= \cos^{-1} \frac{x^{0\prime}}{x_0} \\ &= \cos^{-1} 0.95480 \qquad \text{or} \qquad \cos^{-1} 0.92919 \\ &= \pm 17^\circ\ 17\tfrac{5}{9}' \qquad \text{or} \qquad \pm 21^\circ\ 41\tfrac{5}{11}' \end{aligned}$$

The resulting pattern as a function of ϕ is shown in *a*, Fig. 12-40. The required relative currents are obtained in the usual manner from the coefficients of the various powers of z in the polynomial

$$|E| = (z - e^{j\psi_1^0})(z - e^{-j\psi_1^0})(z - e^{j\psi_2^0})(z - e^{-j\psi_2^0})$$

where $\psi_1^0 = 17° \, 17\frac{5}{9}'$ and $\psi_2^0 = 21° \, 41\frac{5}{11}'$

Multiplying out, the current ratios are found to be

$$1:-3.7680:5.5488:-3.7680:1$$

In the above example it will be noted that to obtain the pattern by the semigraphical method of multiplying together the distances from z to the null points of the array, as indicated in eq. (35), three-figure accuracy is adequate. However, if it is desired to obtain the pattern from the simple phasor addition of the fields due to the individual elements, as in eq. (30), it is necessary to compute the current ratios with great accuracy if a reasonably accurate pattern is to be obtained.

The reason for the extreme accuracy required in this case becomes evident when the fields are added to determine the resultant field in the direction of the maximum, broadside to the array. In this direction, there is no phase difference due to difference in path lengths and so the field trength is proportional to the simple arithmetic sum of all the currents. Adding these currents with due regard to sign,

$$1.0000 - 3,7680 + 5.5488 - 3,7680 + 1.0000 = 0.0128$$

it is seen that the "effective current" radiating in the direction of the maximum is only about one-fifth of 1 per cent of the current in the center element. This low value of radiation results from the fact that the array derives its "supergain" or high directivity properties by virtue of the addition of the radiation from elements carrying large, almost equal and opposite currents. Furthermore a slight error of the order of 1 per cent in the setting of any one of the currents would change the resultant by several hundred per cent, and so completely destroy the supergain pattern.

These effects are demonstrated even more clearly in the next example which is the case of a nine-element array having the same over-all length (one-quarter wavelength) as the array of the previous example. The pattern of this array is shown in b, Fig. 12-40. For this array the nulls were spaced corresponding to the distribution of nulls in $T_4(x)$. The calculated current ratios are given in the following table:

I_1	260,840.2268
I_2	− 2,062,922.9994
I_3	7,161,483.1266
I_4	−14,253,059.7032
I_5	17,787,318.7374
I_6	−14,253,059.7032
I_7	7,161,483.1266
I_8	− 2,062,922.9994
I_9	260,840.2268
Total	00,000,000.0390

It is seen that, if a current of the order of 17 million amperes is fed to the center element, with corresponding currents in the other elements, the total effective current radiating broadside to the array (the direction of the maximum) is equivalent to a current of 39 milliamperes in a single antenna. It is concluded that, although supergain arrays are possible in theory, they are quite impractical.

ADDITIONAL PROBLEMS

3. Derive an expression for the radiation pattern of an antenna of length L which has a traveling wave current distribution represented by $I = I_0 e^{-(\alpha+j\beta)l}$. The phase shift factor β is equal to $2\pi/\lambda$ where λ is assumed to be equal to the free-space wavelength.

4. Using the principle of multiplication of patterns, sketch the following radiation patterns:

(a) The horizontal pattern of four vertical antennas spaced one-half wavelength apart and fed with equal currents, but with 180° phasing between adjacent elements.

(b) Same as part (a), but for eight elements.

(c) The horizontal pattern of four vertical radiators spaced one-quarter wavelength and having a progressive phase shift of 90° between elements.

(d) The free-space vertical patterns (obtained for the array remote from the earth) of each of the arrays of parts (a), (b) and (c):

(1) In the plane of the array

(2) In the plane perpendicular to the plane of the array.

5. An elevated antenna is one wavelength long and is fed a quarter wavelength from one end. Assuming a sinusoidal current distribution (*not* the distribution of Fig. 12-2d) calculate its free-space radiation pattern and its radiation resistance.

6. Using the known relative radiation pattern for a half-wave dipole in free space, determine the absolute value of the pattern (in mv/m at 1 mile for 1 watt radiated) by the method of section 12.13. Hence determine the radiation resistance of the antenna.

7. Calculate and plot the vertical pattern of a 190 degree vertical monopole ($\beta H = 190\pi/180$). Determine: (a) the field intensity at the surface of the earth at a distance of one mile, in mv/m for one kw radiated, assuming negligible attenuation of the field due to earth losses; (b) the field intensity of the secondary lobe.

8. Design an end fire array that will produce approximately the pattern described by

$$f(\phi) = 1 \qquad 0 < \phi < \frac{\pi}{3}$$

$$f(\phi) = 0 \qquad \frac{\pi}{3} < \phi < \pi$$

Use an element spacing of one-quarter wavelength.

9. Design a six element broadside array having a spacing $d = \lambda/2$ between adjacent elements. The pattern is to be optimum, with the side lobe level 20 db down.

BIBLIOGRAPHY

Schelkunoff, S. A., "A Mathematical Theory of Linear Arrays," *B.S.T.J.*, XXII, 1, 80 (1943).

LaPorte, E. A., "Directional Antenna Design," *Electronics*, **9**, 22–25, 48 (1936).

Andrew, A., "A Discussion of Methods Employed in Calculations of Electromagnetic Fields of Radiating Conductors," *Electrical Communication*, **15**, 70 (1936).

Wolff, Irving, "Determination of the Radiating System which will produce a Specified Directional Characteristic," *Proc. IRE*, **25**, 5, 630 (1937).

Friis, Harald T., "A Note on a Simple Transmission Formula," *Proc. IRE*, **34**, 5, 254 (1946).

Brillouin, Leon, "Antennae for Ultra High Frequencies," *Electrical Communication*, **21**, 4 (1944); **22**, 1 (1944).

Friis, H. T. and Lewis, W. D. "Radar Antennas," *B.S.T.J.*, **26**, 2, 220–317 (1947).

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand, New York, 1943.

CHAPTER 13

IMPEDANCE CHARACTERISTICS OF ANTENNAS

To the communication engineer interested in the over-all design of a radio communication system, the antenna is but one link in the complicated chain that leads from the microphone to the loudspeaker. It is natural for him to consider the antenna simply as another circuit element that must be properly matched to the rest of the network for efficient power transfer. From this point of view the input or terminal impedance of the antenna is of primary concern. The input impedance of an antenna is a complicated function of frequency, which cannot be described in any simple analytical form. Nevertheless, at a single frequency, the antenna terminal impedance may be accurately represented by a resistance in series with a reactance. Over a small band of frequencies such representation can still be used, but it is now only approximate. If, as is often the case, the band of frequencies is centered about the "resonant frequency" of the antenna, a better approximation is obtained by representing the antenna as a series R, L, C circuit. If the range of operation extends over a wider band of frequencies, this representation is no longer adequate. It can be improved by adding elements to the "equivalent" network, but the number of elements required for reasonably good representation becomes very large as the frequency range is extended. Under these circumstances it is possible to replace the equivalent lumped-constant network with a distributed-constant network, such as an open-circuited transmission line, the input impedance of which will represent reasonably well the input impedance of the antenna over a wide range of frequencies.

13.01 Lumped-constant Representation of Antenna Input Impedance. For an antenna whose half-length H is shorter than a quarter-wavelength, the input impedance can be represented over a narrow band of frequencies by a resistance R in series with a capacitive reactance X. The resistance R is the radiation resistance (and loss resistance, if any), of the antenna, referred to the terminal or base current. It is given in Fig. 11-12, or it can be obtained from Fig. 11-2 by dividing the values of radiation resistance referred to loop current by $\sin^2 \beta H$. Approximate values of the capacitive reactance X are also given in Fig. 11-12. For antennas which have a half-length greater than about a quarter-wavelength (actually nearer 0.23λ, the exact point depending on the thickness of the antenna), the input impedance becomes inductive and would be represented by a resistance in series with an inductive reactance. For both transmitting and receiving, an antenna is often operated at its resonant frequency, that is, at the center frequency of the narrow band of operation where the antenna input impedance is a pure resistance. Below this center frequency the antenna reactance is capacitive, and above this frequency the reactance is inductive. The input impedance can then be represented approximately by a series R, L, C circuit. Quantities of interest are the required values of the equivalent R_a, L_a, and C_a, and the Q of the antenna.

Figure 13-1 illustrates this representation as a simple R, L, C circuit and shows the variation of impedance and admittance in the vicinity of resonance ($X = 0$) for such a circuit.

The general expression for the impedance is

$$Z_a = R_a + j\left(\omega L_a - \frac{1}{\omega C_a}\right) \tag{13-1}$$

at the resonant frequency $f = f_r$,

$$\omega_r L_a = \frac{1}{\omega_r C_a} \qquad Z_a = Z_r = R_a \tag{13-2}$$

For a small angular frequency increment, $\delta\omega$, from the resonance frequency, the impedance increment is

$$\delta Z_a = j\left(\delta\omega L_a + \frac{\delta\omega}{\omega^2 C_a}\right) \tag{13-3}$$

Therefore the per unit increase in impedance is

$$\frac{\delta Z_a}{Z_r} = \frac{\delta Z_a}{R_a} = j\left(\frac{\delta\omega L_a}{R_a} + \frac{\delta\omega}{R_a\omega^2 C_a}\right)$$
$$= j\left(\frac{Q}{\omega_r}\,\delta\omega + \frac{Q}{\omega_r}\,\delta\omega\right)$$
$$= j2Q\,\frac{\delta\omega}{\omega_r} \qquad (13\text{-}4)$$

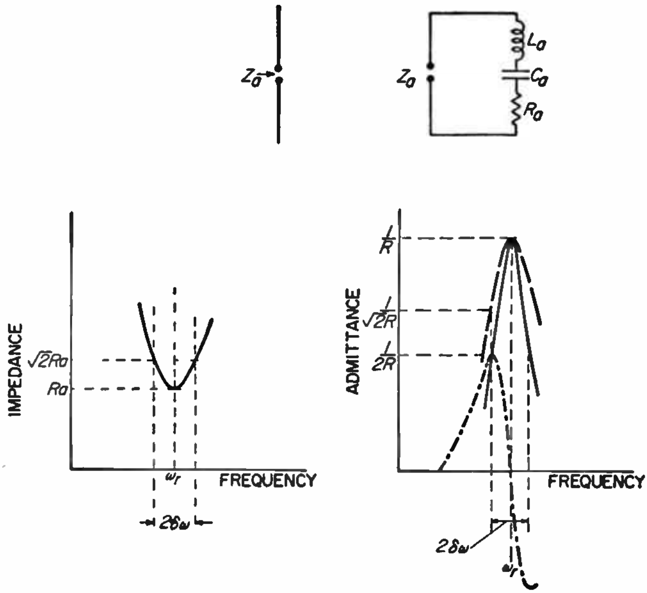


FIG. 13-1. Approximate representation of antenna input impedance by a simple R, L, C circuit, with corresponding impedance and admittance curves.

where
$$Q = \frac{\omega_r L_a}{R_a} = \frac{1}{\omega_r C_a R_a} \qquad (13\text{-}5)$$

From eq. (4)
$$\left|\frac{\delta Z_a}{R_a}\right| = 2Q\left|\frac{\delta\omega}{\omega_r}\right|$$

or
$$\frac{\delta\omega}{\omega_r} = \frac{1}{2Q}\left|\frac{\delta Z_a}{R_a}\right| \qquad (13\text{-}6)$$

When $\delta Z_a = R_a$, the current has dropped to $1/\sqrt{2}$ times its value at resonance and the power has dropped to one-half. The angular frequency difference between half-power points is

$$\Delta\omega = 2\delta\omega = \frac{\omega_r}{Q} \tag{13-7}$$

The frequency difference between half-power points is the band width of the circuit, and the *relative band width* is

$$\frac{\Delta\omega}{\omega_r} = \frac{1}{Q} \tag{13-8}$$

To the extent that the antenna impedance may be represented by the simple circuit of Fig. 13-1, this can be considered to be the band width of the antenna (unloaded). A more general definition for antenna band width is given in sec. 3.

When precise impedance or admittance measurements are made on an actual antenna in the neighborhood of the resonant frequency, it is found that the curves have a somewhat different shape from those of the simple R, L, C circuit shown in Fig. 13-1. This difference is due to the fact that the equivalent R_a, L_a, and C_a of the antenna are really functions of frequency, and not constants as in Fig. 13-1. A much better representation of the antenna impedance may be obtained with the series R, L, C, circuit by assuming L_a and C_a to be constant as before, but taking into account the variation of R_a with frequency. For this purpose, the variation of R_a with frequency can be assumed to be linear over the frequency range of interest (see Fig. 11-12), and so the resistance may be written as

$$R_a = R_r\left(1 + \rho\frac{\delta\omega}{\omega_r}\right) \tag{13-9}$$

where R_r is the resistance at resonance and ρ is a positive constant. The expression for impedance for this case, illustrated in Fig. 13-2, is (for $\omega = \omega_r + \delta\omega$)

$$\begin{aligned} Z_a &= R_a + j\left(\omega L_a - \frac{1}{\omega C_a}\right) \\ &= R_r\left(1 + \rho\frac{\delta\omega}{\omega_r}\right) + j\left(\omega_r L_a + \delta\omega L_a - \frac{1}{\omega_r C_a} + \frac{\delta\omega}{\omega_r^2 C_a}\right) \end{aligned} \tag{13-10}$$

Then, remembering that $\omega_r L_a = \dfrac{1}{\omega_r C_a}$,

$$\frac{Z_a}{R_r} = 1 + \frac{\delta\omega}{\omega_r}\left(\rho + j2\,\frac{\omega_r L_a}{R_r}\right) \tag{13-11}$$

$$= 1 + \frac{\delta\omega}{\omega_r}(\rho + j2Q) \tag{13-12}$$

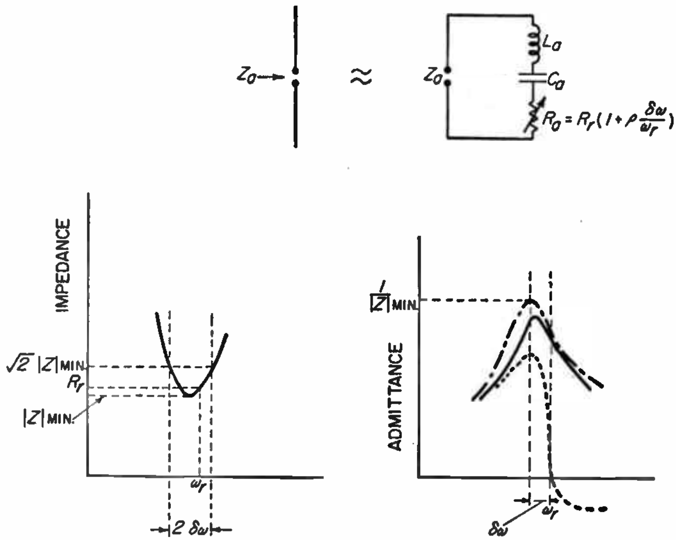


Fig. 13-2. Impedance and admittance curves about resonance when the variation of radiation resistance with frequency is considered.

The impedance and admittance curves for this case are shown in Fig. 13-2. It is seen that the impedance minimum or admittance maximum no longer occurs at resonance, but rather at some frequency below resonance. The frequency at which the impedance is a minimum can be found by minimizing the absolute value of expression (11) (or its square) with respect to $\delta\omega$.

Putting

$$\frac{d}{d(\delta\omega)}\left|\frac{Z_a}{R_r}\right|^2 = 0$$

gives

$$\frac{\delta\omega}{\omega_r} = \frac{-\rho}{\rho^2 + 4Q^2} \qquad \text{(for the minimum)}$$

The impedance at the minimum will be given by

$$Z_{\min} = R_r \left[1 - \frac{\rho}{\rho^2 + 4Q^2} (\rho + j2Q)\right]$$
$$= R_r \left[1 - \frac{\rho^2}{\rho^2 + 4Q^2} - j\frac{2\rho Q}{\rho^2 + 4Q^2}\right] \qquad (13\text{-}13)$$

In order to make use of the equivalent circuit for antenna input impedance in computations, it is necessary to know, or to be able to obtain, values of the equivalent L_a, C_a, and R_a in terms of antenna dimensions. When curves such as those of Fig. 11-12 are available in the range of interest, values of these quantities may be determined from the curves. Otherwise, L_a, C_a, and Q may be calculated in terms of a quantity called the average characteristic impedance of the antenna.

Characteristic Impedance of Antennas. A quantity that has considerable usefulness in connection with antennas is the *average characteristic impedance* of the antenna. The significance of this term as applied to cylindrical antennas can be understood by first considering a biconical transmission line or a biconical antenna. In chap. 8 the characteristic impedance of a transmission line was defined as the voltage-current ratio existing on the line when the line was infinitely long. For a uniform transmission line, this ratio is constant along the line. It is easily shown that the transmission line, formed by two infinitely long coaxial conical conductors having a common apex (Fig. 13-3), is a uniform line, and that the ratio of voltage to current along the line will remain constant, that is, independent of r. (The voltage V is applied across an infinitesimal gap at the apex and the current I flows out of one cone and into the other.) In secs. 13.06 and 13.07 the general solution for a finite length of such a transmission line (or antenna) will be obtained. It will be found that such a structure can support the TEM wave, as well as higher order TM waves. For the outgoing TEM wave (which alone is

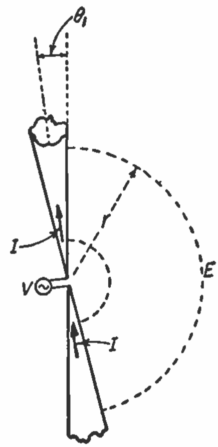


FIG. 13-3. Input to a conical transmission line (or a bi-conical antenna).

excited on the infinitely long line) the expressions for the fields are

$$\left.\begin{aligned} H_\phi &= \frac{A}{r \sin \theta} e^{-j\beta r} \\ E_\theta &= \frac{A\eta_v}{r \sin \theta} e^{-j\beta r} \\ E_r &= E_\phi = H_\theta = H_r = 0 \end{aligned}\right\} \tag{13-14}$$

Maxwell's equations in spherical co-ordinates for the case of no variation in the ϕ direction are

$$\left.\begin{aligned} \frac{1}{r}\frac{\partial(rE_\theta)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \theta} &= -j\omega\mu H_\phi \\ \frac{1}{r \sin \theta}\frac{\partial}{\partial \theta}(\sin \theta H_\phi) &= j\omega\epsilon E_r \\ -\frac{1}{r}\frac{\partial(rH_\phi)}{\partial r} &= j\omega\epsilon E_\theta \end{aligned}\right\} \tag{13-15}$$

Direct substitution of (14) into (15) shows that Maxwells equations are satisfied. In addition, because $E_r = 0$, the boundary conditions are automatically satisfied.

It will be noted that the electric field distribution is just that corresponding to the static case (problem 9, chap. 2) and that the magnitude of the voltage between the cones at any distance r from the apex is constant. That is

$$\begin{aligned} V &= \int_{\theta_1}^{\pi - \theta_1} E_\theta r \, d\theta = \eta_v A \, e^{-j\beta r} \int_{\theta_1}^{\pi - \theta_1} \frac{1}{\sin \theta} d\theta \\ &= 2\eta_v A \ln \cot \frac{\theta_1}{2} \end{aligned} \tag{13-16}$$

Also the amplitude of the current flowing in the cones is constant along the line and is given by

$$\begin{aligned} I &= 2\pi r \sin \theta_1 H_\phi \\ &= 2\pi A \end{aligned} \tag{13-17}$$

Therefore, the characteristic impedance for a biconical transmission line or biconical antenna is constant and is

$$\begin{aligned} Z_0 &= \frac{V}{I} = \frac{\eta_v}{\pi} \ln \cot \frac{\theta_1}{2} \\ &= 120 \ln \cot \frac{\theta_1}{2} \end{aligned} \tag{13-18}$$

Because the electric and magnetic field configurations are the same as for the stationary-fields case, this same result could have been obtained directly by using the *static* capacitance per unit length as calculated in chap. 2. Thus

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{1}{cC}$$

where

$$C = \frac{\pi\epsilon}{\ln \cot \frac{\theta_1}{2}}$$

and

$$\frac{1}{\sqrt{LC}} = c = \frac{1}{\sqrt{\mu_v \epsilon_v}} = 3 \times 10^8 \qquad \text{meter/sec}$$

Then

$$Z_0 = \frac{\eta_v}{\pi} \ln \cot \frac{\theta_1}{2} = 120 \ln \cot \frac{\theta_1}{2}$$

For *thin* antennas, that is when θ_1 is small, the characteristic impedance is given approximately by

$$Z_0 = 120 \ln \left(\frac{2}{\theta_1}\right) = 120 \ln \left(\frac{2r}{a}\right) \qquad (13\text{-}19)$$

where a is the cone radius at a distance r from the apex.

When a cylindrical antenna is treated in a like manner, it is evident that the corresponding "bi-cylindrical" transmission line will be nonuniform, with a capacitance per unit length and characteristic impedance that vary along the line. However, for *thin* antennas the elements dr can be considered as elements of a *biconical* line which has a cone angle $\theta_1 = a/r$, where a is the radius of the cylinder and r is the distance from the origin to the element dr (Fig. 13-4). Then the characteristic impedance at a distance r will be

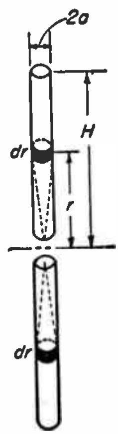


Fig. 13-4

$$Z_0(r) = 120 \ln \left(\frac{2}{\theta_1}\right) = 120 \ln \left(\frac{2r}{a}\right) \qquad (13\text{-}20)$$

It is seen that the characteristic impedance of a cylindrical antenna varies along the antenna, being larger near the ends. For a center-fed cylindrical antenna of half-length H, an "average" characteristic impedance can be defined by

$$Z_0\ (\text{av.}) = \frac{1}{H}\int_0^H Z_0(r)\,dr \tag{13-21}$$

$$= \frac{1}{H}\int_0^H 120\ln\left(\frac{2r}{a}\right)dr$$

$$= 120\left[\ln\left(\frac{2H}{a}\right) - 1\right] \tag{13-22}$$

Equivalent L_a, C_a, and Q in terms of Z_0 (av). Use of the average characteristic impedance of an antenna makes it possible to obtain values for the equivalent R, L, C circuit of the antenna in terms of the antenna dimensions. This is done by first finding a Q for the antenna by comparison with ordinary transmission line theory, and then determining L_a and C_a in terms of this Q and the known R_a.

From transmission-line theory for low-loss lines (see chap. 8), the Q of a transmission line is

$$Q = \frac{\omega L}{R} = \frac{\omega Z_0}{Rv} = \frac{2\pi Z_0}{\lambda R} \tag{13-23}$$

where R, L, and C are the resistance, inductance, and capacitance per unit length of line and

$$Z_0 \approx \sqrt{\frac{L}{C}} \qquad v \approx \frac{1}{\sqrt{LC}}$$

for low-loss lines. For a resonant length ($l = \lambda/4$) of open-circuited line, it is easily shown that the input impedance is a pure resistance of value

$$R_{\text{in}} = \frac{Rl}{2} = \frac{R\lambda}{8} \tag{13-24}$$

and this must be equal to R_a in the equivalent lumped-circuit representation of antenna input resistance (Fig. 13-1). That is,

$$\frac{R\lambda}{8} = R_a = R_{\text{rad}}$$

For the equivalent lumped-constant circuit

$$Q = \frac{\omega_r L_a}{R_a}$$

Therefore $$\omega_r L_a = R_a Q = \frac{R\lambda}{8} \cdot \frac{2\pi Z_0}{\lambda R} = \frac{\pi Z_0}{4}$$

$$\omega_r C_a = \frac{4}{\pi Z_0}$$

and

$$\left.\begin{aligned} L_a &= \frac{Z_0}{8f_r} \qquad & C_a &= \frac{2}{\pi^2 f_r Z_0} \\ Q &= \frac{\pi Z_0}{4R_a} \qquad & R_a &= R_{\text{rad}} \end{aligned}\right\} \tag{13-25}$$

where the Z_0 in the case of a cylindrical antenna, will be the average characteristic impedance Z_0 (av) defined in the preceding section. The Q of the antenna as given by eq. (25) will be the *unloaded* Q. When the antenna circuit is loaded by a properly matched generator impedance or load impedance, the total Q of the circuit is one-half the Q of the antenna above. That is

$$Q_{\text{loaded}} = \tfrac{1}{2} Q_{\text{unloaded}}$$

It is this Q_{loaded}, which is of chief concern in band width considerations.

Problem 1. (a) Using the curves of Fig. 11-12, determine an approximate value for the Q of a half-wave dipole antenna constructed with No. 12 copper wire, at a frequency of 100 megacycles; (b) compute Q for the same antenna using eq. (13-25).

Problem 2. From the point of view of band width, discuss the suitability of each of the following antennas for (a) an F-M receiving antenna, (b) a television receiving antenna: a half-wave dipole constructed of (1) No. 12 wire, (2) 1-cm. diameter rods, (3) 1-in. diameter pipes, (4) a biconical cage arrangement with a total cone angle of 10°.

NOTE: The F-M band covers from 88–108 mc. The low-frequency television channels are 6 mc. wide and, at present, are 54–60, 60–66, 66–72, 76–82, 82–88. In both cases it is desired, if practical, that a single antenna should cover the entire band with a decrease of received power of less than 3 db.

13.02 The Antenna as an Opened-out Transmission Line. Using an assumed sinusoidal current distribution, the power radiated from an antenna can be calculated, and approximate values for input resistance and reactance can be obtained for very thin antennas. However, for thicker antennas, such as tower antennas for broadcast use, especially when fed near a current node, the sinusoidal-current assumption fails to give sufficiently good answers. Faced with the necessity of having to feed such antennas, and in the absence of a rigorous solution to the antenna problem, it was natural that engineers should look for a better assumption than the sinusoidal for the current distribution. Because the input impedance of an antenna goes through variations somewhat similar to the input impedance of an open-circuited transmission line, it was also natural to attempt to treat the antenna as an opened-out transmission line. This attack had the advantage of using transmission line theory with which the engineer was already familiar, and it provided him with a simple expression for current distribution and input impedance which, although only approximate, could always be "adjusted" in the light of measured values. Although in more recent years the cylindrical antenna problem has been solved to a fairly good approximation and accurate values for input impedance are now available in the form of curves or tables, it is often still very convenient to have available simple analytic expressions, which will indicate the correct order of magnitude of input impedance and current distribution. For this reason, and because it is instructive to compare the antenna with the transmission line, one of the methods developed for treating the antenna as a transmission line will be outlined.

The most extensive study of the transmission line representation of an antenna has been made in a series of papers* by Siegel and Labus, and their results have been used in the broadcast antenna field for many years. The method treats the antenna as an opened-out transmission line that is open circuited at the end. An antenna differs from a transmission line in two important respects. An antenna radiates power, whereas transmission line theory assumes negligible radiation of power. Ordinary transmission line theory deals with *uniform* lines for which L, C, and Z_0 are constant along the line (except very close to the end). For the nonuniform line representing the antenna, L, C, and Z_0 all vary along the line, and indeed it becomes necessary to define what is meant by these quantities under such conditions. Siegel and Labus assume that

* J. Labus, "Mathematical Calculation of the Impedance of Antennas," *Hochf. und Elek.*, **41**, 17 (1933); E. Siegel and J. Labus, "Apparent Resistance of Antennas," *Hochf. und Elek.*, **43**, 166 (1934); E. Siegel and J. Labus, "Transmitting Antennas," *Hochf. und Elek.*, **49**, 87 (1937).

the radiated power can be accounted for by introducing an equal amount of ohmic loss distributed along the transmission line. Knowing this loss, an attenuation factor can be calculated, and this factor can be used to give a better approximation for the current distribution. In addition the variable characteristic impedance of the antenna is replaced by an average value. Because the value used for this average characteristic impedance determines very largely what the input impedance will be, considerable effort was expended in obtaining a truly significant expression for Z_0. The essentials of the Siegel and Labus method (with slight modifications) are outlined below.

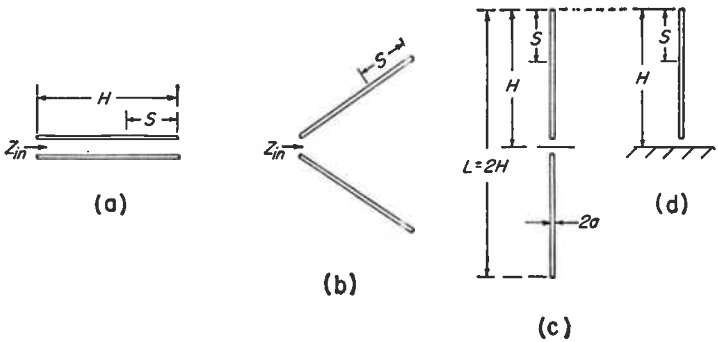


FIG. 13-5. The antenna as an opened-out transmission line.

Input Impedance by Transmission Line Analogy. Figure 13-5 illustrates the representation of a center-fed dipole antenna as an opened-out transmission line. The "equivalent" transmission line has a length equal to the half-length H of the dipole antenna. The diameter of the antenna or transmission-line conductors is $2a$. The problem of a ground-based antenna of height H, erected on a perfect reflecting plane (Fig. 13-5d), is the same as that of the dipole antenna of half-length H (Fig. 13-5c), except that values of characteristic and input impedances will be just one-half those obtained for the corresponding dipole antenna.

The first step is to obtain an expression for Z_c (av), the average characteristic impedance of the antenna. In sec. (13.01) a simple expression for average characteristic impedance was developed from elementary considerations. Siegel and Labus have developed a somewhat different expression in quite another way. The scalar potential for an antenna carrying a sinusoidal current distribution was set up and compared with the usual expression for the scalar potential along a uniform parallel-wire line. Comparison of these expressions yielded an expression for the characteristic impedance $Z_0(s)$ at each point s along the antenna. From this, an average characteristic impedance for the total length was defined by

$$Z_c\ (\text{av}) = \frac{1}{H}\int_0^H Z_0(s)\,ds \tag{13-26}$$

The final expressions obtained for Z_c (av) were

(a) For the center-fed dipole antenna of half-length H (Fig. 13-5c)

$$Z_c\ (\text{av}) = 120\left(\ln\frac{H}{a} - 1 - \frac{1}{2}\ln\frac{2H}{\lambda}\right) \quad \text{ohm} \tag{13-27}$$

(b) For the antenna of height H, fed against a perfect reflecting plane (Fig. 13-5d),

$$Z_c\ (\text{av}) = 60\left(\ln\frac{H}{a} - 1 - \frac{1}{2}\ln\frac{2H}{\lambda}\right) \quad \text{ohm} \tag{13-28}$$

The notation Z_c (av) has been used to distinguish this average characteristic impedance from the Z_0 (av) given by eq. (22). For the special case of $H = \lambda/4$, (half-wave dipole or quarter-wave ground-based antenna), the expressions for Z_c (av) become

(c) Half-wave dipole,

$$Z_c\ (\text{av}) = 120\left(\ln\frac{H}{a} - 0.65\right) \quad \text{ohm} \tag{13-29}$$

(d) Quarter-wave ground-based antenna,

$$Z_c\ (\text{av}) = 60\left(\ln\frac{H}{a} - 0.65\right) \quad \text{ohm} \tag{13-30}$$

Having a value for Z_c (av), the expressions for the voltage and current distributions and input impedance may be written down from transmission line theory. For the open-circuited line the expressions are

$$V_s = V_R \cosh \gamma s \tag{13-31}$$

$$I_s = \frac{V_R}{Z_c\ (\text{av})}\sinh \gamma s \tag{13-32}$$

$$Z_s = \frac{V_s}{I_s} = Z_c\,(\text{av})\coth\gamma s \tag{13-33}$$

$$Z_{\text{in}} = Z_c\,(\text{av})\coth\gamma H \tag{13-34}$$

In these expressions V_s and I_s are the voltage and current, respectively, at a distance s from the open end of the line. $\gamma = \alpha + j\beta$ is the complex propagation constant for the line. Its imaginary part $\beta = 2\pi/\lambda$ is the phase shift constant, and its real part α is the attenuation constant, which is still to be determined.

The expression for current distribution [eq. (32)] may be written in terms of the current at the loop or maximum, I_m, as follows: Expanding sinh γs, and taking the absolute magnitude, it is seen that, if the attenuation is not too great, the maximum amplitude of current (the loop current) will occur approximately at $s = \lambda/4$, and will be given by

$$\begin{aligned} I_m = I_{s=\lambda/4} &= j\frac{V_R}{Z_c\,(\text{av})}\cosh\alpha\frac{\lambda}{4}\sin\beta\frac{\lambda}{4} \\ &\approx j\frac{V_R}{Z_c\,(\text{av})} \end{aligned}$$

Then eq. (32) may be written as

$$\begin{aligned} I_s &= (-j)I_m\sinh\gamma s \\ &= I_m(\cosh\alpha s\sin\beta s - j\sinh\alpha s\cos\beta s) \end{aligned} \tag{13-35}$$

The next step is to determine the attenuation factor α for the "equivalent" transmission line. On the basis of a sinusoidal current distribution, the power radiated by the antenna can be computed by the Poynting vector method or the induced-emf method. This gives a value for R_{rad}, the radiation resistance, referred to the loop current. By definition,

$$\text{Power radiated} = |I_m|^2 R_{\text{rad}} \tag{13-36}$$

R_{rad} for ground-based antennas on a perfect reflecting plane is plotted as a function of antenna height in Fig. 11-2. Alternatively, for H greater than 0.2λ, R_{rad} may be obtained with good accuracy from the approximate formula

$$R_{\text{rad}} = 15\left[-\frac{\pi}{2}\sin\frac{4\pi H}{\lambda} + \left(\ln\frac{2H}{\lambda} + 1.722\right)\cos\frac{4\pi H}{\lambda} + 4.83 + 2\ln\frac{2H}{\lambda}\right] \tag{13-37}$$

For the corresponding center-fed dipole ($L = 2H$) in free space, the values of R_{rad} given by Fig. 11-2 or eq. (37) must be doubled.

In the transmission line representation the radiated power is replaced by an equal amount of power dissipated as ohmic loss along the line. This power loss may be assumed to be due to a series resistance r ohms per unit length, shunt conductance g mhos per unit length, or both. In their analysis, Siegel and Labus assumed a series resistance for the line, of such value that the total I^2R loss was equal to the radiated power. It turns out that if the power loss is considered to be due to both series resistance and shunt conductance of such values that the I^2r loss per unit length at a current loop is equal to the V^2g loss per unit length at a voltage loop, similar expressions result. There is the added advantage that input impedances calculated from these simpler expressions seem to be in better agreement with values calculated by other means. This is especially true for short antennas where the series resistance assumption, used by Siegel and Labus, leads to values of input resistance that are consistently too high, whereas the assumption used here leads to correct values.

Assuming both series resistance r per unit length and shunt conductance g per unit length, the total power loss along the line is

$$\begin{aligned} W &= \int_0^H (|I|^2 r + |V|^2 g)\, ds \\ &= \int_0^H (I_m{}^2 r|\sinh \gamma s|^2 + V_m{}^2 g|\cosh \gamma s|^2)\, ds \end{aligned} \tag{13-38}$$

The values of r and g are so chosen that

$$I_m{}^2 r = V_m{}^2 g \tag{13-39}$$

Then the expression (13) for power dissipated becomes

$$W = I^2_{m_1} r \int_0^H (|\sinh \alpha s \cos \beta s + j \cosh \alpha s \sin \beta s|^2 + |\cosh \alpha s \cos \beta s + j \sinh \alpha s \sin \beta s|^2)\, ds$$

which reduces to

$$\begin{aligned} W &= I^2_{m_1} r \int_0^H \cosh 2\alpha s\, ds \\ &= I^2_{m_1} r H \frac{\sinh 2\alpha H}{2\alpha H} \end{aligned} \tag{13-40}$$

For small values of $2\alpha H$, this becomes approximately

$$W \approx I^2_{m_1} r H$$

That is, the power loss per unit length is approximately constant and is equal to

$$I^2_{m_1} r = V^2_{m_1} g \tag{13-41}$$

The total power dissipated must equal the power that is actually radiated, so

$$I^2_{m_1} r H = I^2_{m_1} R_{\text{rad}}$$

Therefore
$$r = \frac{R_{\text{rad}}}{H}$$

Also, using (41),

$$g = \frac{I^2_{m_1} r}{V^2_{m_1}} = \frac{r}{Z^2_c \text{ (av)}} = \frac{R_{\text{rad}}}{HZ^2_c \text{ (av)}}$$

For any low-loss transmission line the attenuation factor is given by

$$\alpha = \frac{1}{2}\left(\frac{r}{Z_0} + gZ_0\right)$$

so for this "equivalent" line

$$\alpha = \frac{1}{2}\left(\frac{r}{Z_c \text{ (av)}} + \frac{R_{\text{rad}}}{HZ_c \text{ (av)}}\right) = \frac{R_{\text{rad}}}{HZ_c \text{ (av)}} \tag{13-42}$$

The total attenuation for the length H is

$$\alpha H = \frac{R_{\text{rad}}}{Z_c \text{ (av)}} \tag{13-43}$$

The input impedance can now be obtained from eq. (34). For purposes of computation eq. (34) can be expanded into more suitable forms:

$$Z_{\text{in}} = Z_c \text{ (av)} \left[\frac{\cosh (\alpha H + j\beta H)}{\sinh (\alpha H + j\beta H)}\right]$$

$$= \frac{Z_c \text{ (av)}}{2}\left(\frac{\sinh 2\alpha H - j \sin 2\beta H}{\cosh^2 \alpha H - \cos^2 \beta H}\right) \tag{13-44}$$

$$= Z_c \text{ (av)}\left(\frac{\sinh 2\alpha H - j \sin 2\beta H}{\cosh 2\alpha H - \cos 2\beta H}\right) \tag{13-45}$$

The input resistance and input reactance are, respectively,

$$R_{\text{in}} = \frac{Z_c\,(\text{av})}{2}\left(\frac{\sinh 2\alpha H}{\cosh^2 \alpha H - \cos^2 \beta H}\right) \tag{13-46}$$

$$X_{\text{in}} = \frac{Z_c\,(\text{av})}{2}\left(\frac{-\sin 2\beta H}{\cosh^2 \alpha H - \cos^2 \beta H}\right) \tag{13-47}$$

The current distribution is given by eq. (35). When only the magnitude of the current is of interest, this may be obtained from

$$\begin{aligned} I_s &= I_m|\sinh \gamma s| \\ &= I_m \sqrt{\sinh^2 \alpha s + \sin^2 \beta s} \end{aligned} \tag{13-48}$$

Correction for the End-effect. Equation (47) for the input reactance for an antenna indicates that the reactance goes through zero for lengths of line that are integral multiples of a quarter-wavelength. It is known from experiment that this is not the case, and that, in fact, the reactance zeros occur for physical lengths of antennas that are somewhat less than multiples of $\lambda/4$. This effect, which also occurs on open-ended transmission lines, is known as *end-effect.* It is due to a decrease in L and an increased C near the end of the line. This results in a decrease in Z_0 and an increase in current near the end of the line over that given by the sinusoidal distribution. With transmission lines the magnitude of the effect depends upon the line spacing in wavelengths. The region in which the change in the line "constants" occurs is known as the *terminal zone.* In the case of the transmission line, the terminal zone extends back a distance approximately equal to the line spacing. In the case of antennas, the end-effect produces an apparent lengthening of the antenna, the amount of which depends in a rather complicated manner on the characteristic impedance, the length, and the configuration of the antenna. The effect is somewhat greater for antennas of low characteristic impedance (large cross section) than it is for thin wire antennas. Siegel has investigated* the end effect on both transmission lines and antennas and has computed the following table, which shows numerical values for the amount by which the apparent electrical length of the antenna exceeds its physical length measured in wavelengths.

* Ernest M. Siegel, *Wavelength of Oscillations Along Transmission Lines and Antennas*, University of Texas Publication No. 4031, Aug. 15, 1940.

TABLE I
INCREASE IN APPARENT LENGTH OF ANTENNAS DUE TO END EFFECT

H_0/λ	% Increase	
	Tower antenna Z_c (av) = 220 ohms	Wire antenna Z_c (av) = 500 ohms
¼	5.4	4.5
⅜	5.3	4.3
½	5.2	2.2
0.59	5.1	1.9

It is seen that the rule of thumb often used in the field, by which the physical length is made 5 per cent less than the desired electrical length, is a rather good approximation for tower antennas. In computing input impedance by this method it is the apparent electrical length that should be used for H.

"Modified" Impedance. It is natural to ask what sort of agreement may be expected between measured impedances and those calculated from this simplified representation. Siegel and Labus have compared measured and calculated input impedances for elevated horizontal dipoles, and, in general, have obtained good agreement. However, when measurements are made on ground-based tower antennas, it is found that considerable difference may exist between measured and calculated values. In general, the values calculated by this method are high compared with measured values. Morrison and Smith* have made an extensive set of measurements on a 400-ft uniform cross section tower antenna and have compared the results with values calculated by this method. It was found that, if the antenna terminals were considered to be shunted by a 200-$\mu\mu$f capacitance and fed through a 6.8-μh inductance (presumably to account for base capacitance and finite ground conductivity), the resultant "modified base impedance" showed excellent agreement with measured values over a three-to-one frequency range. Although the theoretical justification for this procedure is questionable, it does produce useful answers. More-

* J. F. Morrison and P. H. Smith, "The Shunt-Excited Antenna," *Proc. I.R.E.*, vol. 25, no. 6, pp. 673–696, June, 1937.

over, it must be remembered that, regardless of the accuracy of theoretical results, some modification will always be required, because of the differences that exist between the *ideal* configuration that is calculated and the *actual* configuration that is measured. However, it is significant that answers given by the methods of Hallén and Schelkunoff (sections 13.04 and 13.08) show reasonably good agreement with measured values when corrected only for "visible" factors, such as the known base capacitance.

Example 1: Determine an approximate value for the input impedance at antiresonance of a full-wave ($L = 2H \approx \lambda$) cylindrical dipole antenna having a diameter of 2 cm. The frequency is 150 mc.

Physical half-length

$$H_0 \approx \frac{2 \times 0.95}{2} = 0.95 \text{ meter}$$

$$\frac{H_0}{a} = 95$$

$$Z_c\,(\text{av}) = 120(\ln 95 - 1 - \tfrac{1}{2} \ln 0.95) = 430 \text{ ohms}$$

$$\alpha H = \frac{R_{\text{rad}}}{Z_c\,(\text{av})} = \frac{210}{430} = 0.489$$

$$\beta H \approx \pi$$

$$Z_{\text{in}} = R_{\text{in}} = 430 \left(\frac{\sinh 0.978}{\cosh 0.978 - 1} \right) = 948 \text{ ohms}$$

Example 2: A uniform cross section tower antenna is 400 ft high and 7 ft square. Calculate the base impedance at a frequency of 1300 kc.

(a) Characteristic impedance

Equivalent* radius $a = 0.5902 \times 7 \quad = 4.14$ ft

Physical height $H_0 \quad = 400$ ft

Then $H_0/a = 96.6 \qquad 2H_0/\lambda = 1.06$

$$Z_c\,(\text{av}) = 60\left(\ln \frac{H_0}{a} - 1 - \frac{1}{2} \ln \frac{2H_0}{\lambda} \right) = 210 \text{ ohms}$$

* E. Hallén has shown that the correct value for the equivalent radius of a noncircular cylinder is obtained by finding the radius of the infinitely long circular cylinder that has the same capacitance per unit length as does an infinite length of the noncircular cylinder. For the square cross section of side length d he obtains for the equivalent radius, $a = 0.5902d$. "Theoretical Investigations into the Transmitting and Receiving Qualities of Antennae," *Nova Acta Upsal* 4, 11, 7 (1938). *Also* "Admittance Diagrams for Antennas and the Relation between Antenna Theories," Cruft Laboratory Report No. 46, June 1938.

(b) Attenuation factor

$$\text{Electrical height } \beta H = \frac{2\pi \times 1.05 \times 400}{231 \times 3.281} = 3.48 \text{ radians}$$

$$\alpha H = \frac{R_{r\,d}}{Z_0} = \frac{81}{210} = 0.386 \text{ nepers}$$

(c) Theoretical base impedance

$$R = \frac{210}{2}\left(\frac{0.851}{1.155 - 0.883}\right) = 329 \text{ ohms}$$

$$X = -\frac{210}{2}\left(\frac{0.643}{0.272}\right) = -248 \text{ ohms}$$

$$Z_{\text{base}} = 329 - j248 \text{ ohms}$$

(d) Modified base impedance. From measurements on other structures, it has been determined that the results given by this method, when

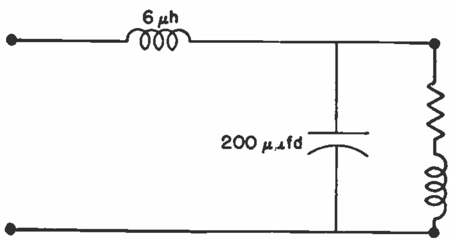


FIG. 13-6. Modified base impedance.

used on ground-based tower antennas, should be modified by the addition of a shunt capacitance of about 200 μμfd and a series inductance of about 6 μh (Fig. 13-6). Then the expected input impedance is obtainable from the circuit of Fig. 13-6:

$$\text{Modified base impedance } Z_b \text{ (mod)} = j49 + 144 - j233$$
$$= 144 - j184 \text{ ohms}$$

(For the WWJ tower which is 400 ft high and 6½ ft square, Morrison and Smith show a measured value of 140 − j200 ohms at this frequency.)

EXAMPLE 3: The antenna of example 2 is to be used as antenna (1) in a two-element directional array, with a quarter-wave tower as antenna (2). The loop current of antenna (2) is equal to, but leads, the loop current of antenna (1) by 90 degrees, that is,

$$I_2 \text{ (loop)} = I_1 \text{ (loop)}\underline{/90°}$$

Assume Z_{11} is nearly equal to the self-impedance of antenna (1) and use

$$Z_{12} = 20\underline{/-25°} \qquad \text{(referred to current loops)}$$
$$Z_{22} = 36 + j20$$

Determine the driving-point impedances:

(a) Refer the mutual impedance to the base of antenna (1). The mutual impedance Z_{21} referred to the *base* current $I_1(0)$ of antenna (1) is

$$Z_{21}\text{ (base)} = \frac{V_{21}}{I_1(0)} = \frac{I_1\text{ (loop)}}{I_1(0)} \frac{V_{21}}{I_1\text{ (loop)}} = \frac{I_1\text{ (loop)}}{I_1(0)} Z_{21}\text{ (loop)}$$

where V_{21} is the open-circuit voltage at the terminals of (2) due to the current flow in (1).

From example 2,

$$\frac{I_1\text{ (loop)}}{I_1\text{ (base)}} = \frac{\sinh \gamma\lambda/4}{\sinh \gamma H} = j \frac{\cosh \alpha\lambda/4}{\sinh 0.386 \cos 200° + j \cosh 0.386 \sin 200°}$$
$$\approx \frac{1\underline{/90°}}{0.523\underline{/224.8°}} = 1.91\underline{/-134.8°}$$

then $\quad Z_{21}\text{ (base)} = 20\underline{/-25°} \times 1.91\underline{/-134.8°} = 38.2\underline{/-159.8°}$

(b) Driving-point impedances

$$Z_1' = Z_{11} + \frac{I_2}{I_1} Z_{12} \qquad \text{(all referred to base)}$$
$$\frac{I_2\text{ (base)}}{I_1\text{ (base)}} = \frac{I_2\text{ (loop)}}{I_1\text{ (loop)}} \times \frac{I_1\text{ (loop)}}{I_1\text{ (base)}}$$
$$= 1\underline{/90°} \times 1.91\underline{/-134.8°} = 1.91\underline{/-44.8°}$$
$$Z_1' = 329 - j248 + 1.91\underline{/-44.8°} \times 38.2\underline{/-159.8°}$$
$$= 263 - j218$$

For antenna (2), base current equals loop current.

$$Z_2' = Z_{22} + \frac{I_1}{I_2} Z_{12} \qquad \text{(loop or base)}$$
$$= 36 + j20 + \frac{38.2\underline{/-159.8}}{1.91\underline{/-44.8}}$$
$$= 36 + j20 + 20\underline{/-115.0}$$
$$= 27.5 + j1.9$$

(c) Modified driving-point impedances. Treating the driving-point impedances in the same manner as the base impedances, the modified driving-point impedances are obtained.

$$Z_1' \text{ (modified)} = 130 - j154$$
$$Z_2' \text{ (modified)} = 28 + j50$$

13.03 Wide-band Impedance Matching. In general, it is desirable to match a transmitting antenna to the transmission line or *r-f* output circuit that feeds it. At a single frequency, or over a relatively narrow band of frequencies as in broadcast work, the impedance-matching problem is very simple, and one with which the reader is assumed to be familiar.* If the antenna to be matched is one of an array, then it is necessary to design the matching network to control the phase as well as magnitude of the current in the antenna, but this is still a straightforward circuit problem, an example of which is given in section 14.02. However, in communication work it is often necessary to use a single antenna through a wide band of frequencies extending over a range of 1.5 to 1 or more. In this case, wide-band impedance-matching circuits are required. When the antenna used is nonresonant or aperiodic (such as a rhombic, for example) so that its input impedance remains relatively constant over the frequency band of interest, the design of the wide-band matching network can be accomplished through the use of standard band-pass filter theory. However, more often than not, the antenna impedance that is to be matched varies between wide limits, and the design problem is quite complicated. When the analytical approach becomes too cumbersome, a very effective attack that may be used is a combination of graphical and analytical methods.†

In this approach the antenna impedance is plotted in the complex plane as a function of frequency. Figures 13-7a and 13-7b show the usual impedance and admittance curves for a typical antenna of fairly broad band width, and (c) and (d) show the corresponding plots in the complex plane. It will be recalled from chap. 8 that, in the complex plane, the locus of impedances that produce constant standing-wave ratios (on the transmission line which they terminate) is a circle. The wide-band matching problem is usually stated in terms of keeping the standing-wave ratio, ρ, below some stated value; in this case, below $\rho = 2$. Thus, in

* E.g., W. L. Everitt, *Communication Engineering*, chaps. VII and VIII, McGraw-Hill, 1937.

† F. D. Bennett, P. D. Coleman, and A. S. Meier, "The Design of Broadband Aircraft Antenna Systems," *Proc. IRE*, **33**, 10, pp. 671–700 (1945).

Figs. 13-7c and 13-7d, the problem is that of warping the impedance or admittance characteristic within the $\rho = 2$ circle over the required frequency range. The range of frequencies lying within

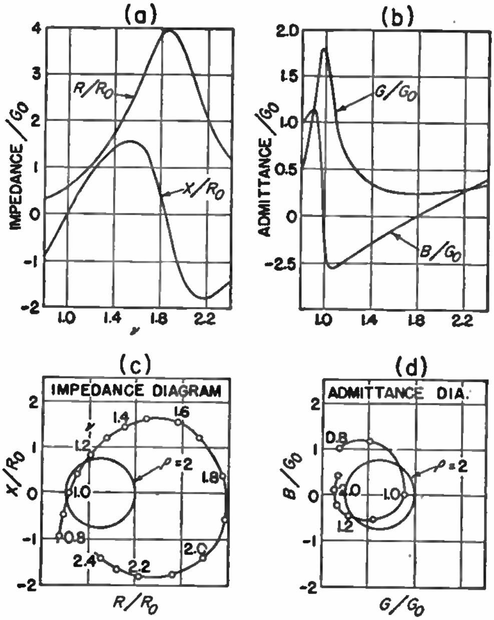


FIG. 13-7. Typical measured impedance and admittance curves for a broadband H-F or V-H-F antenna: (a) Impedance curve plotted against relative frequency. (b) Admittance curve. (c) Impedance diagram plotted in the complex plane. (d) Admittance diagram.

the $\rho = 2$ circle will be called the *band width** of the antenna, and, in this instance, per unit band width will be defined as

$$\text{band width} = \frac{f_2 - f_1}{f_1} = \frac{\Delta f}{f_1} \tag{13-49}$$

where f_1 and f_2 are the lower and upper frequencies, respectively, at which the impedance curve intersects the $\rho = 2$ circle.

* The required performance of an antenna varies greatly with the type of service, so the terms band width and percentage band width have not been standardized, but must be specified for each application.

It is convenient to show the frequency scale relative to the resonant frequency of the antenna, and this is done by use of a relative frequency, $\nu = f/f_0$. Then $\nu = 1$ at the first resonant frequency, f_0, of the antenna, and percentage band width will be given by

$$\frac{\nu_2 - \nu_1}{\nu_1} = \frac{\Delta\nu}{\nu_1} \tag{13-50}$$

Inspection of the impedance and admittance curves of Figs. 13-7a and 13-7b reveals that the susceptance curve has a negative slope at resonance (the first resonant point), and the reactance curve has a negative slope at antiresonance (usually called the second resonant point). This means that it should be possible to cancel the susceptance of the antenna in the region of first resonance by means of a suitable parallel positive susceptance. In the region of second resonance, the reactance of the antenna can be cancelled by a suitable series positive reactance. These facts form the basis of a technique for increasing the band width of an antenna. Figure 13-8 shows the effect on the impedance and admittance diagrams of adding a series capacitance (negative reactance) or series inductance (positive reactance). The effect of a series capacitance is to add a negative reactance that moves the impedance characteristic downward; on the other hand, a series inductance adds a positive reactance that moves the impedance curve upwards. On the admittance diagrams the series elements affect both conductance and susceptance and a series capacitance tends to rotate the curve counterclockwise, whereas a series inductance rotates it clockwise. In the example of Figs. 13-8a and 13-8c the series capacitance has been added to bring the antenna impedance curve down into the $\rho = 2$ circle, increasing the band width from zero to about 17 per cent [since $(1.23 - 1.05)/1.05 = 0.17$]. In Figs. 13-8b and 13-8d the series inductance has raised the impedance curve of a different antenna into the circle to produce a band width of about 46 per cent. It will be noticed that maximum band width is obtained by raising the curve slightly beyond the diameter into the upper left-hand portion of the circle. This result follows from the fact that, using the definition given above, the band width can be increased both by increasing $\Delta\nu$ and by decreasing ν_1.

The inductance or capacitance to be added can be obtained by use of lumped-constant circuits (coil or condenser) or distributed-constant circuits (sections of transmission line). In the high-frequency range, coils and condensers would usually be used, but in the V-H-F range, sections of transmission line are very convenient. Figure 13-9 indicates how such line sections can be built right into the antenna or the feed line. The lengths of line required to yield a given effective inductance or capacitance, can be calculated with the aid of the appropriate formulas from chap. 8.

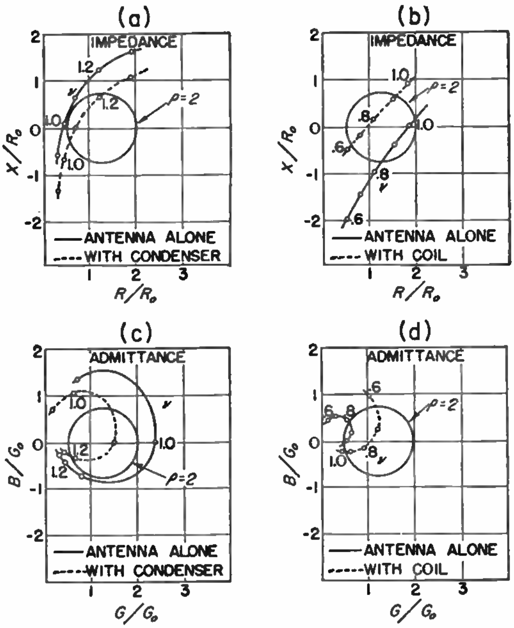


FIG. 13-8. Effect on impedance and admittance curves of the addition of a series capacitance or inductance.

The impedance-matching properties of quarter-wave and half-wave sections of transmission line at a single frequency are well known. Figure 13-10 shows in striking fashion the broad-banding properties of such sections when used in series with the antenna.

The effect of elements added in parallel with the antenna is best shown on admittance diagrams. A very important practical case is that of a parallel-resonant circuit or shorted quarter-wave stub placed in parallel with an antenna. The effects of this are shown in the admittance diagrams of Fig. 13-11. As indicated previously the susceptance curve of an antenna has a negative slope about the resonant frequency. It is evident from Fig. 13-11a that an optimum design for a parallel matching stub can be achieved by just cancelling the antenna susceptance at the points at which the

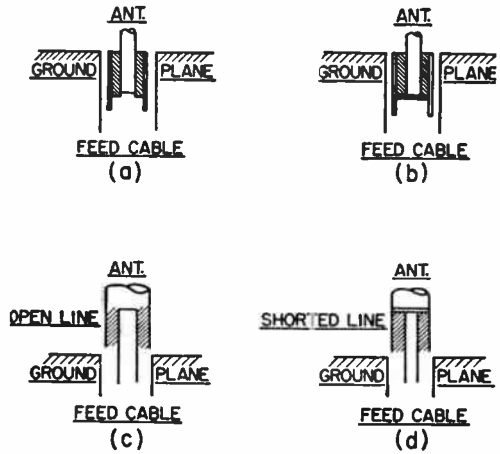


FIG. 13-9. Possible methods for building series reactance into feed line or antenna.

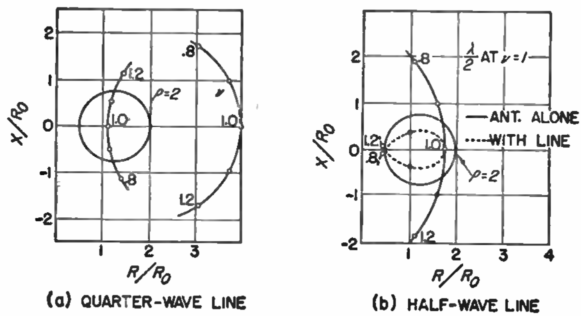


FIG. 13-10. Effect of series quarter-wave and half-wave lines on impedance characteristic.

conductance is 0.5. This will cause the admittance curve of Fig. 13.11b to tie on the $\rho = 2$ circle, and so result in maximum bandwidth for a single element. The most favorable antenna admittance curve for this purpose is one that has a conductance of just less than 2 at resonance, as shown in the figure. The required stub dimensions can then be determined by the following procedure. The points A and D, which have the negative of the antenna susceptance at the $G = 0.5$ points, are marked on the diagram. The susceptance curve of the matching stub must pass through these points. As a first approximation the susceptance variation of the stub is assumed to be linear over this frequency range and a straight

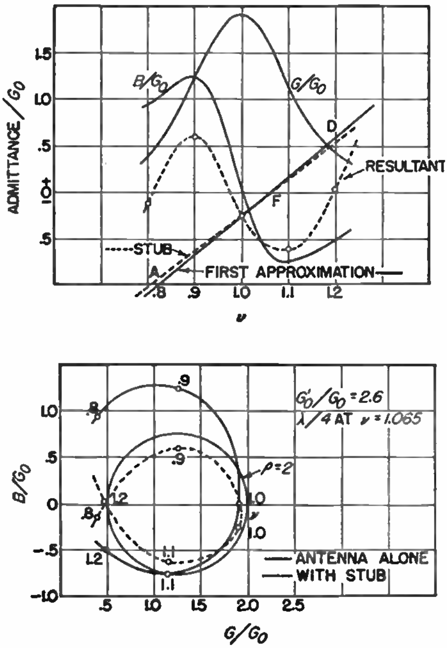


FIG. 13-11. Matching a resonant antenna over a range of frequencies by means of a parallel quarter-wave stub: admittance curves and admittance diagram.

line is drawn through the points A and D. The intersection of this line with the $Y = 0$ line gives the frequency for which the stub should be a quarter-wavelength long. The slope of the line determines the required characteristic impedance for the stub. The actual susceptance characteristic of the shorted stub can then be drawn in, as has been done in Fig. 13-11, and the resultant susceptance curve may be plotted.

The parallel broad-banding stub just considered is important practically because of the ease of incorporating it into actual antenna systems. It appears automatically in certain types of antennas, typified by the folded dipole or folded monopole. Although the primary purpose of folding is usually that of obtaining high input impedances, the resultant structure is such that the antenna impedance is effectively shunted by the input impedance of a shorted quarter-wave stub (or two stubs in series in the case of the folded dipole). The resultant wide band width has been pointed out by Carter.* In addition, it is often necessary to transform from an unbalanced (coaxial) feed to a balanced antenna or vice versa, and this is accomplished by use of "baluns" or "bazookas," some of which are described in chap. 14. Since these balance-to-unbalance transformers are formed using resonant line sections, it is often possible to design them to perform the functions of both balun and wide-band impedance-matching stub.

The use of two (or even more) elements for broad-band matching greatly increases the designer's control over the impedance characteristic and enables him to design for greater band widths. The procedure is to use the first element to "set up" the antenna curve into the ideal position for the second element to warp it into the $\rho = 2$ circle. Many interesting examples of both one and two element broad-band matching sections will be found in the original article,† from which the examples used here were taken.

13.04 The Cylindrical Antenna Problem. The methods considered earlier in this chapter for representing the impedance of an antenna by lumped-constant or distributed-constant circuits prove useful in the analytical design of suitable matching networks for the antenna. However these circuit representations are not *solutions* of the antenna problem, and the impedances or current distributions, which they approximate over a certain range of frequencies, are assumed to be known, that is, are obtainable from experiment or calculation. It is the purpose of this present section to indicate three important methods of solution that have been applied to the antenna problem. In a later section one of these methods will be considered in some detail.

Any solution of the antenna problem will of course have Maxwell's equations as a starting point. In fact, the problem is just one of solving Maxwell's equations subject to the boundary con-

* P. S. Carter, "Simple Television Antennas," *RCA Rev.*, 4, 168 (1939).

† F. D. Bennett, P. D. Coleman, and A. S. Meier, *Loc. cit.*

ditions imposed by the antenna and the source. For the simple center-fed cylindrical antenna this turns out to be a surprisingly difficult problem. Three general methods of attack have been used. The first of these methods (historically) treats the problem as a boundary-value problem. The second method sets up the problem as that of finding the solution of an integral equation for the current. The third method treats the antenna as an open-ended waveguide or electromagnetic horn.

(*a*) *As a Boundary-value Problem.* For certain symmetrical antenna shapes (e.g., the ellipsoid or prolate spheroid) it is possible to solve for the *free* oscillations or natural modes, so determining the proper frequencies and corresponding damping factors. This problem was worked out many years ago by Abraham* for very thin ellipsoids and later by Brillouin† for prolate spheroids of any eccentricity. When the antenna is excited or fed, the solution is given in terms of an infinite series of the free-oscillation modes with coefficients chosen so as to satisfy the force function. Page and Adams,‡ Ryder,§ Stratton‖ and Chu¶ are among those who have worked on this problem. This method has the advantage of yielding very reliable results, but is restricted to a relatively few shapes, among which, unfortunately, the cylinder is not included. There are two main disadvantages of the method. First, although the method is useful near resonance, for lengths considerably different from the resonant length the series converge very slowly so that an excessive amount of labor is involved in obtaining numerical answers. Second, actual antennas generally are not prolate spheroids, but have various shapes, the circular cylinder being most common. About the best that can be done in obtaining a solution for the actual antenna by this method is to assume that the solution for an "equivalent" thin prolate spheroid will hold approximately for the cylindrical antenna. The troublesome question of just what size of prolate spheroid is "equivalent" to a cylinder prevents this method from being so useful as it might otherwise be. An excellent summary and comparison of the work of different writ-

* Max Abraham, *Ann. Physik*, **66**, 435 (1898): *Math. Ann.*, **52**, 81 (1899).
† L. Brillouin, *Propagation de l'Electricite*, Hermann, Paris, 1904, Vol. 1.
‡ L. Page and N. I. Adams, *Phys. Rev.*, **53**, 819 (1938).
§ Robert M. Ryder, *J. Applied Phys.*, **13**, 327 (1942).
‖ J. A. Stratton, *Proc. Nat. Acad. Science*, **21**, 51 (317) 1935.
¶ J. A. Stratton and L. J. Chu, *J. App. Phys.*, **19**, 236 (1941).

ers using this and other methods is contained in an article by Brillouin.*

(*b*) *Integral equation solution.* Hallén† has used a different approach to the antenna problem. Starting with an arbitrary current distribution, general expressions for the field are obtained by the use of retarded potentials. Application of the boundary conditions at the surface of the antenna then leads to an integral equation for the current. Thus, instead of a set of partial differential equations, it is now an integral equation that must be solved. The method is general and applicable to antennas of different shapes, but the accurate evaluation of the resulting expressions is very difficult. However, quite recently‡ Hallén has succeeded in reducing the integrals involved to ordinary sine and cosine integrals, and iterated sine and cosine integrals, whose values he has tabulated for arguments from 0 to 7. From these tables he has constructed admittance and impedance diagrams for cylindrical antennas§ for a wide range of antenna dimensions. The impedance diagrams, showing antenna resistance and reactance separately, are reproduced in Figs. 13-12 and 13-13. In these curves, antenna resistance and reactance are shown as a function of antenna length (in radians or in wavelengths) for various ratios of half-length to radius. The single diagram of Fig. 13-14 displays this same information in a different form. Antenna conductance and susceptance are plotted in an admittance diagram for all antenna lengths in the range covered, and for six ratios of H/a. The intersecting lines mark off fixed values of βH (antenna length in radians). This method of plotting has the advantage that it makes interpolation easier. In addition the effect of base capacitance, which is always present in any actual antenna set-up, is easily allowed for in an admittance

* Leon Brillouin, "Antennae for Ultrahigh Frequencies," *Elec. Comm.*, **21**, 4, 257 (1944); and **22**, 1, 11 (1944).

† Erik Hallén, "Theoretical Investigations into the Transmitting and Receiving Qualities of Antennae," *Nova Acta Upsal* 4, **11** (1938); "Further Investigations into the Receiving Qualities of Antennae: the Absorption of Transient Unperiodic Radiation," *Arsskrift, Upsala*, 4 (1939).

‡ E. Hallén, "Iterated Sine and Cosine Integrals," *Trans. Royal Inst. Technology*, Stockholm, 12 (1947); "On Antenna Impedances," *Trans. Royal Inst. Technology*, Stockholm, 13 (1947).

§ E. Hallén, *Admittance Diagrams for Antennas and the Relation between Antenna Theories*, Cruft Laboratory, Harvard University, Technical Report 46 June, 1948.

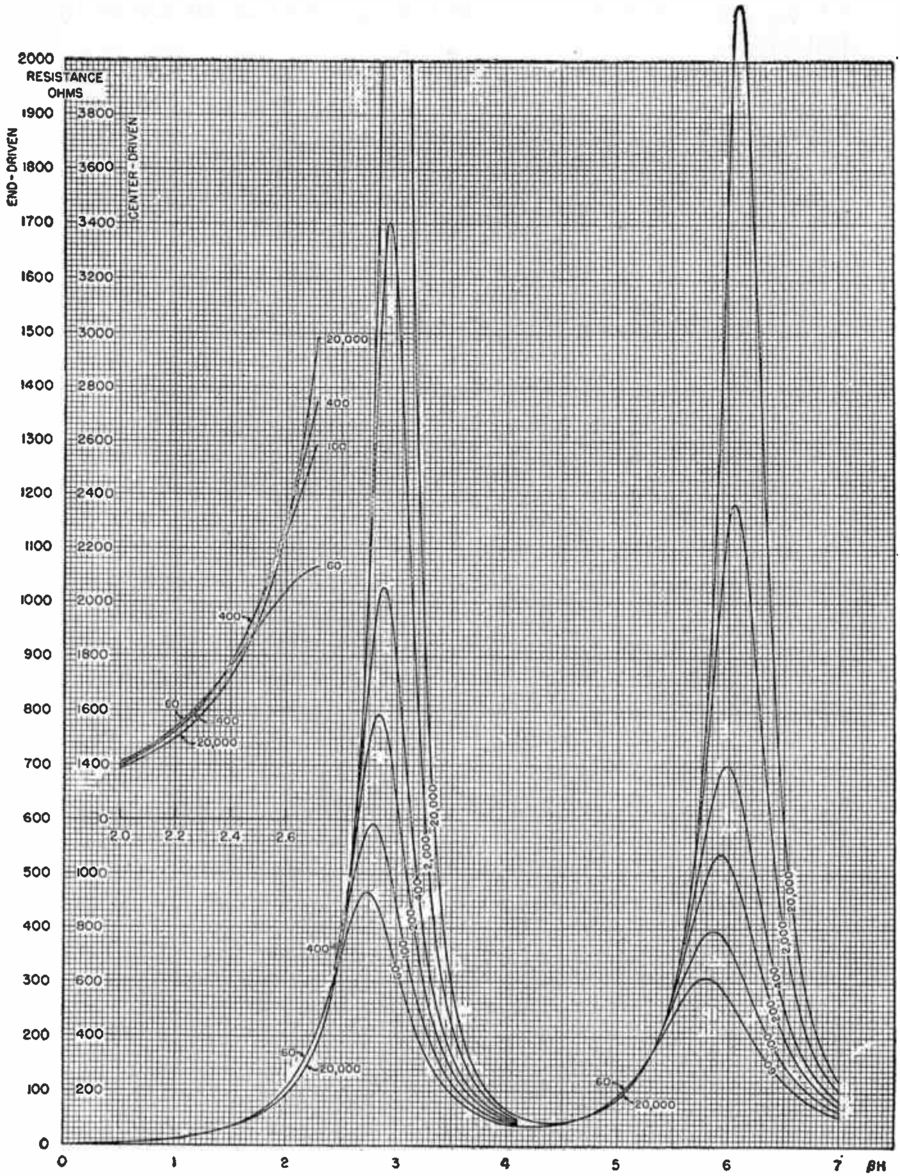


FIG. 13-12. Antenna resistance according to Hallén. The resistance of center-fed dipoles is plotted as a function of $2\pi H/\lambda$, the antenna half-length in radians, for various ratios of H/a, half-length to radius. For monopoles of length H, the ordinates should be divided by two.

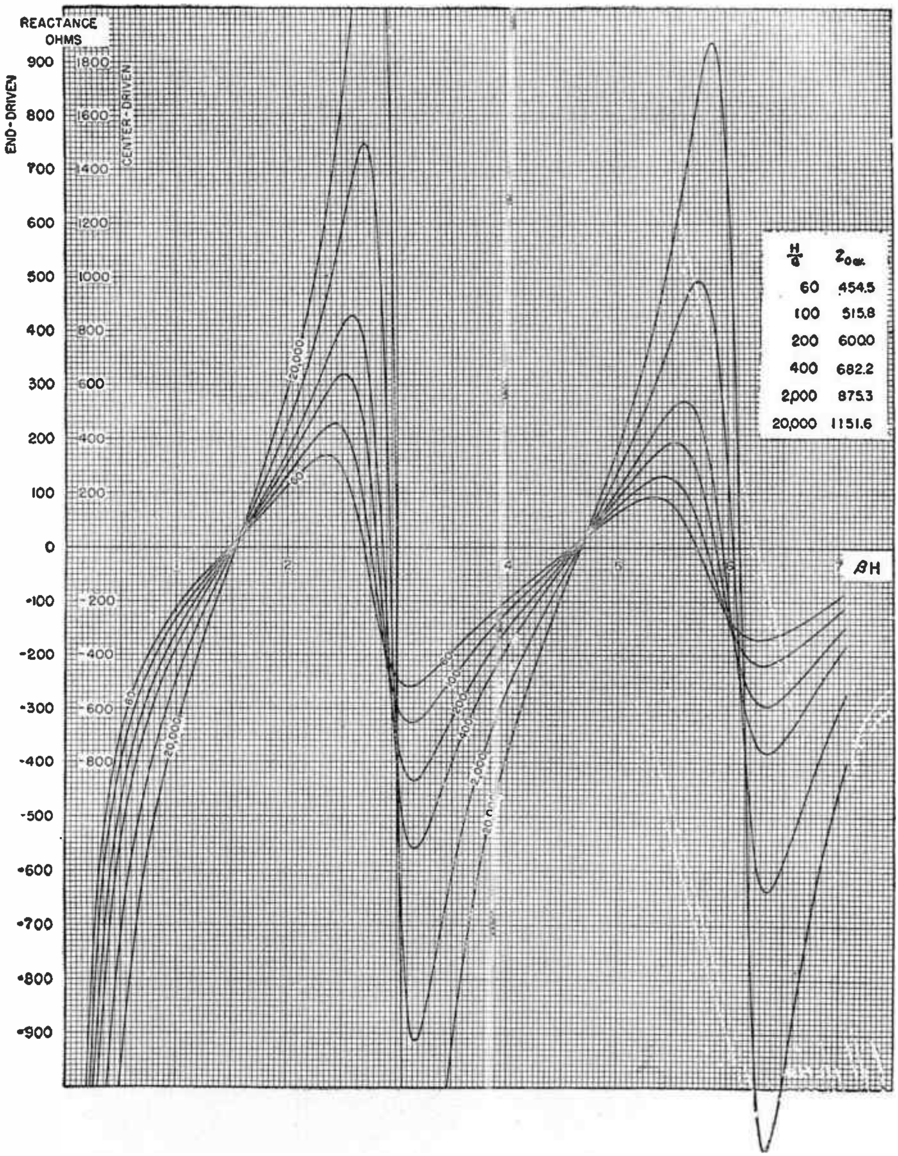


FIG. 13-13. Antenna reactance according to Hallén (see legend for Fig. 13-12).

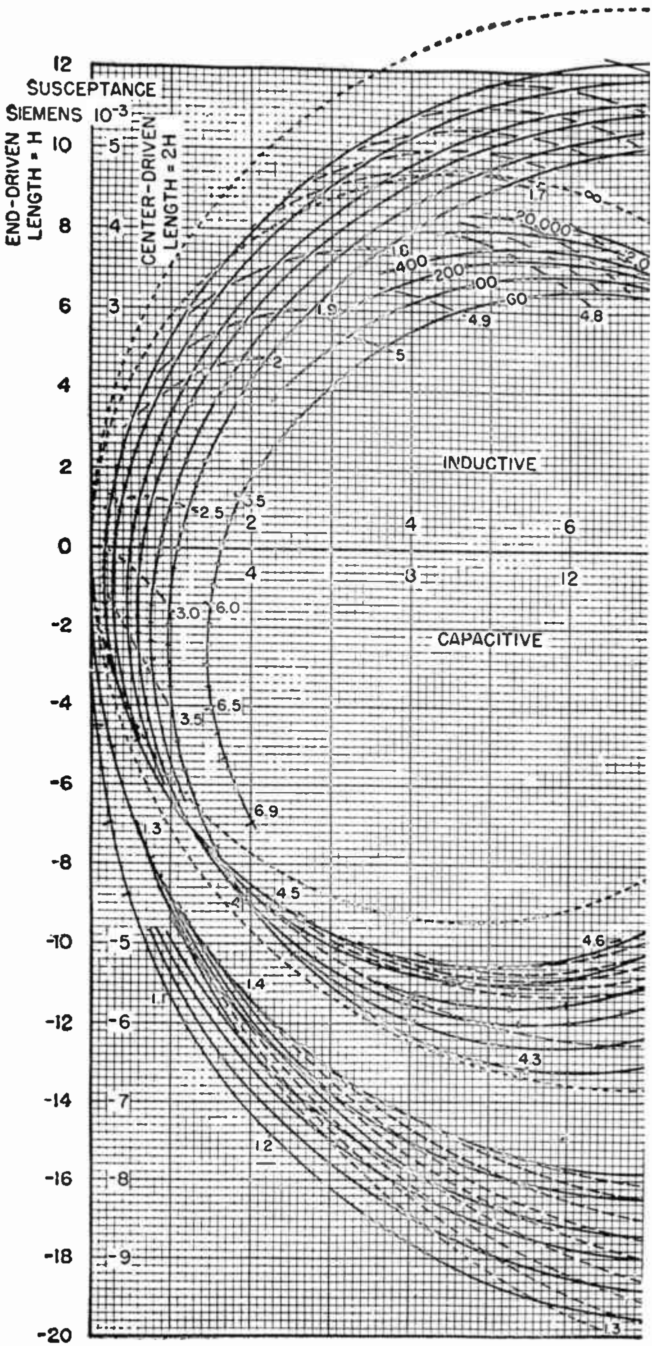
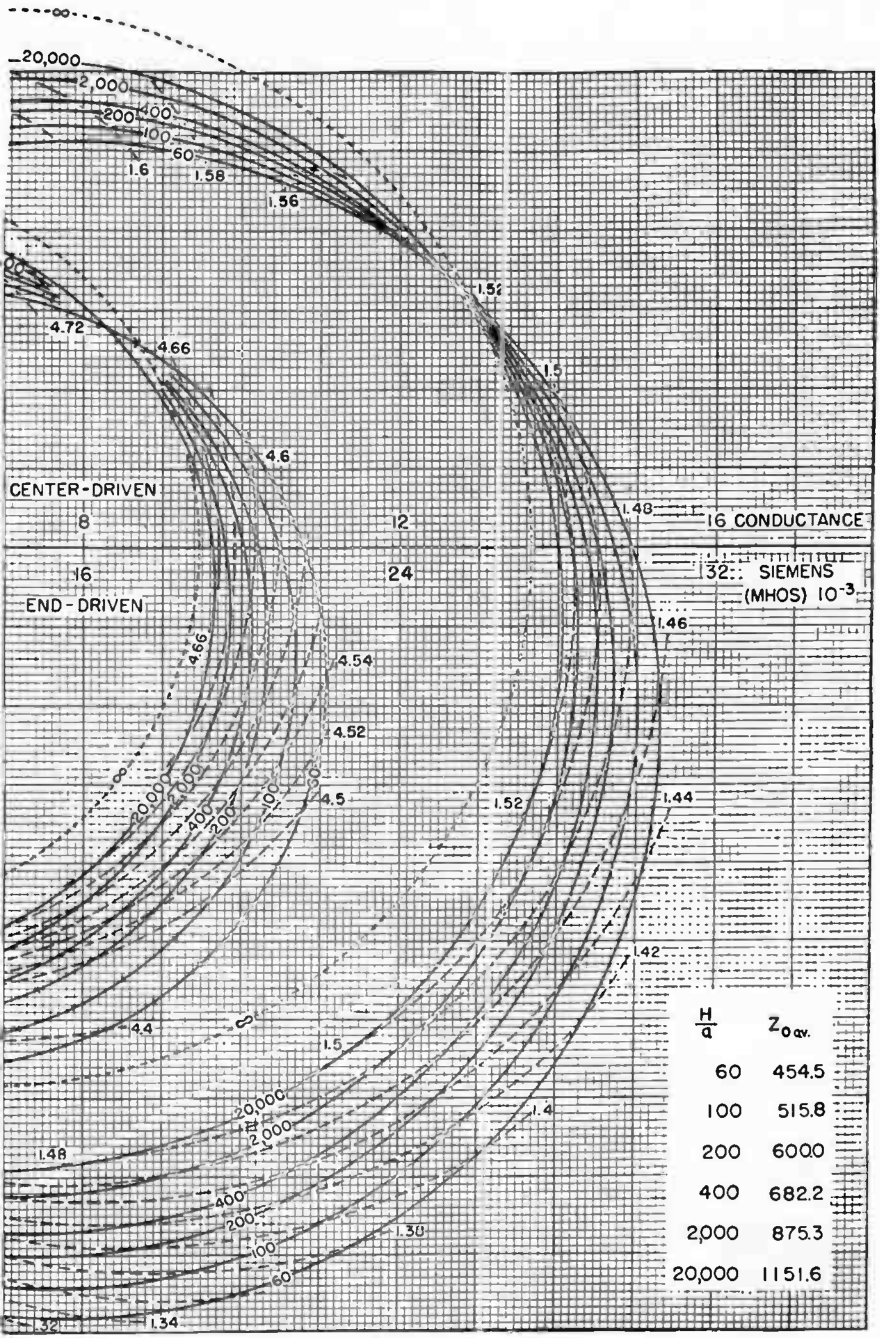


Fig. 13-14. Antenna admit-



$\frac{H}{a}$	$Z_{0\,av.}$
60	454.5
100	515.8
200	600.0
400	682.2
2,000	875.3
20,000	1151.6

tance diagram (Hallén).

diagram. Because G is unaffected by the addition of shunt susceptance, the only effect of the base capacitance, C_b, is to raise the whole admittance diagram by the amount ωC_b. In an impedance diagram, on the other hand, both R and X are affected by the shunting base capacitance.

Unfortunately Hallén's solution is too involved mathematically to be treated adequately in an engineering text of this scope. A good introductory discussion of the integral equation method can be found in the book by Aharoni.* The advanced student, who is interested in obtaining a more complete knowledge of this powerful method, should refer to the original papers by Hallén.

(c) *The antenna as a waveguide or electric horn.* An entirely different attack on the antenna problem has been made by Schelkunoff, who treats the antenna as an open-ended waveguide or electric horn. In contrast to the usual approach used in boundary value problems, where a solution is sought in terms of the natural modes or oscillations of the system, Schelkunoff solves the problem in terms of waves transmitted along the antenna. This corresponds to the engineering solution of the transmission-line problem in terms of initial and reflected waves, as against the alternative method of solution in terms of natural oscillations on a section of line. The method uses familiar transmission-line and wave-guide theories, and is an approach which the engineer finds quite satisfying. Because it represents an important application of concepts developed in earlier chapters, this method will be considered in detail. First, however, it will be necessary to give some consideration to spherical waves.

13.05 Spherical Waves. For propagation in a homogeneous medium having constants μ, ϵ, and σ the scalar wave equation is

$$\nabla^2 V = \gamma^2 V \tag{13-51}$$

where $$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$$

In spherical co-ordinates (51) becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} = \gamma^2 V \tag{13-51a}$$

* J. Aharoni, *Antennae, An Introduction to their Theory,* Clarendon Press, Oxford, 1946.

Although in general the propagation constant γ may be complex, for dissipationless media $\sigma = 0$ and γ^2 reduces to $-\omega^2\mu\epsilon$. Separating eq. (51a) by letting

$$V = R(r)P(\theta)\Phi(\phi)$$

results in the three equations

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = (b^2 + \gamma^2 r^2)R \qquad (13\text{-}52)$$

$$\frac{d^2\Phi}{d\phi^2} = -m^2\Phi \qquad (13\text{-}53)$$

$$\frac{d^2P}{d\theta^2} + \cot\theta\frac{dP}{d\theta} + \left(b^2 - \frac{m^2}{\sin^2\theta}\right)P = 0 \qquad (13\text{-}54)$$

where the constants b^2 and $-m^2$ may be real or complex. When m is an integer, Φ is periodic with a period 2π.

Equation (54) is the *associated Lengendre equation.* When b^2 is real and has the form $b^2 = n(n + 1)$, eq. (54) may be written

$$(1 - x^2)\frac{d^2p}{dx^2} - 2x\frac{dP}{dx} + \left[n(n + 1) - \frac{m^2}{1 - x^2}\right]P = 0 \quad (13\text{-}55)$$

where

$$x = \cos\theta, \qquad 1 - x^2 = \sin^2\theta, \qquad \frac{d}{d\theta} = -\sin\theta\frac{d}{dx}$$

For those problems in which there is no variation with ϕ, the constant m in eq. (53) is zero, and for these cases eq. (55) becomes

$$(1 - x^2)\frac{d^2P}{dx^2} - 2x\frac{dP}{dx} + n(n + 1)P = 0 \qquad (13\text{-}56)$$

which is the *ordinary Legendre equation.*

For nonintegral values of n the solutions of (56) are given by the functions $P_n(\cos\theta)$ and $P_n(-\cos\theta)$, where

$$P_n(\cos\theta) = \sum_{q=0}^{\infty}\frac{(-1)^q(n + q)!}{(n - q)!(q!)^2}\sin^{2q}\left(\frac{\theta}{2}\right) \qquad (13\text{-}57)$$

For integral values of n, expression (57) reduces to the Legendre polynomials and $P_n(\cos\theta)$ and $P_n(-\cos\theta)$ are no longer linearly independent. Under these circumstances it is convenient to use

$P_n(\cos\theta)$ (where n is a positive integer) for one solution and $Q_n(\cos\theta)$ for the other, where the Q_n functions are defined by

$$Q_n(\cos\theta) = P_n(\cos\theta)\log\cot\frac{\theta}{2} - \sum_{s=1}^{s=n}\frac{P_{n-s}P_{s-1}}{s} \qquad (13\text{-}58)$$

The Q functions become infinite at $\theta = 0$ and $\theta = \pi$, and so can be used to represent physically realizable fields only when the $0 - \pi$ axis is excluded from the region being considered.

When m is not equal to zero, the associated Lengendre equation must be considered. For integral values of n its solutions are $P_n^m(\cos\theta)$ and $Q_n^m(\cos\theta)$ where

$$P_n^m(\cos\theta) = (-1)^m \sin^m\theta\,\frac{d^m[P_n(\cos\theta)]}{d(\cos\theta)^m} \qquad (13\text{-}59)$$

$$Q_n^m(\cos\theta) = (-1)^m \sin^m\theta\,\frac{d^m[Q_n(\cos\theta)]}{d(\cos\theta)^m} \qquad (13\text{-}60)$$

For the first few values of n, the *Legendre* and *associated Lengendre polynomials* represented by (7) and (9) are,

$$P_0(\cos\theta) = 1$$
$$P_1(\cos\theta) = \cos\theta$$
$$P_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1)$$
$$P_3(\cos\theta) = \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta)$$
$$P_4(\cos\theta) = \tfrac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3)$$

$$P_1^1(\cos\theta) = -\sin\theta$$
$$P_2^1(\cos\theta) = -3\sin\theta\cos\theta$$
$$P_2^2(\cos\theta) = 3\sin^2\theta$$
$$P_3^1(\cos\theta) = -\tfrac{3}{2}\sin\theta(5\cos^2\theta - 1)$$
$$P_3^2(\cos\theta) = 15\sin^2\theta\cos\theta$$

Considering now the solutions to eq. (52) for R, this equation may be written as

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} - (\gamma^2r^2 + b^2)R = 0 \qquad (13\text{-}61)$$

This equation is slightly different from the ordinary Bessel equation (62) or the modified Bessel equation (63) with which it should be compared.

Ordinary Bessel $z^2 \frac{d^2w}{dz^2} + z\frac{dw}{dz} + (z^2 - \nu^2)w = 0$ (13-62)

Modified Bessel $z^2 \frac{d^2w}{dz^2} + z\frac{dw}{dz} - (z^2 + \nu^2)w = 0$ (13-63)

Equation (61) can be reduced to a standard form by suitable change of variable. Let

$$w = rR \qquad \text{or} \qquad R = \frac{w}{r}$$

Then $\frac{dR}{dr} = \frac{1}{r}\frac{dw}{dr} - \frac{w}{r^2}, \qquad \frac{d^2R}{dr^2} = \frac{1}{r}\frac{d^2w}{dr^2} - \frac{2}{r^2}\frac{dw}{dr} + \frac{2w}{r^3}$

and eq. (61) becomes

$$\frac{d^2w}{dr^2} - \left(\gamma^2 + \frac{b^2}{r^2}\right)w = 0 \tag{13-64}$$

Now put $b^2 = n(n+1)$ as in the Legendre equations and let $z = \gamma r$. Then (64) becomes

$$\frac{d^2w}{dz^2} - \left[1 + \frac{n(n+1)}{z^2}\right]w = 0 \tag{13-65}$$

Solutions to this equation are denoted by $\hat{K}_n(z)$ and $\hat{I}_n(z)$ where

$$\hat{K}_n(z) = e^{-z}\sum_{p=0}^{n}\frac{(n+p)!}{p!(n-p)!(2z)^p} \tag{13-66}$$

$$\hat{I}_n(z) = \frac{1}{2}\left[e^z\sum_{p=0}^{n}\frac{(-1)^p(n+p)!}{p!(n-p)!(2z)^p} + (-1)^{n+1}e^{-z}\sum_{p=0}^{n}\frac{(n+p)!}{p!(n-p)!(2z)^p}\right] \tag{13-67}$$

For the first few values of n these are

$$\hat{K}_0(z) = e^{-z} \qquad \hat{I}_0(z) = \sinh z$$

$$\hat{K}_1(z) = e^{-z}\left(1 + \frac{1}{z}\right) \qquad \hat{I}_1(z) = \cosh z - \frac{\sinh z}{z}$$

$$\hat{K}_2(z) = e^{-z}\left(1 + \frac{3}{z} + \frac{3}{z^2}\right) \qquad \hat{I}_2(z) = \left(1 + \frac{3}{z^2}\right)\sinh z - \frac{3}{z}\cosh z$$

In general the propagation constant γ of eq. (61) is complex. For the particular, but important, practical case where the attenuation factor α is zero, γ is a pure imaginary equal to $j\beta$, and eq. (61) becomes

$$r^2 \frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + (\beta^2r^2 - b^2)R = 0 \tag{13-68}$$

which should be compared with the ordinary Bessel equation (62). Reducing (68) in the same manner that (61) was reduced, but letting $z = \beta r$, there results

$$\frac{d^2w}{dz^2} + \left[1 - \frac{n(n+1)}{z^2}\right] w = 0 \tag{13-69}$$

instead of (65).

Solutions of this equation are denoted by $\hat{J}_n(z)$ and $\hat{N}_n(z)$, where for the first few values of n, these functions* are

$$\hat{J}_0(z) = \sin z \qquad \hat{N}_0(z) = -\cos z$$

$$\hat{J}_1(z) = \frac{\sin z}{z} - \cos z \qquad \hat{N}_1(z) = -\sin z - \frac{\cos z}{z}$$

$$\hat{J}_2(z) = \left(\frac{3}{z^2} - 1\right)\sin z - \frac{3}{z}\cos z \qquad \hat{N}_2(z) = \left(1 - \frac{3}{z^2}\right)\cos z - \frac{3}{z}\sin z$$

The $\hat{I}$ and $\hat{K}$ functions are simply related to the $\hat{J}$ and $\hat{N}$ functions by

$$\begin{aligned} \hat{I}_n(jz) &= j^{n+1}\hat{J}_n(z) \\ \hat{K}_n(jz) &= j^{-n-1}[\hat{J}_n(z) - j\hat{N}_n(z)] \end{aligned} \tag{13-70}$$

In addition the functions $\hat{J}$, $\hat{N}$, $\hat{I}$, $\hat{K}$, are related to the ordinary and modified Bessel functions, J, N, I, K. Indeed they are just the half-integral orders of the corresponding ordinary and modified Bessel functions. The relations are:

$$\left.\begin{aligned} \hat{J}_n(z) &= \sqrt{\frac{\pi z}{2}}\, J_{(n+\frac{1}{2})}(z) \qquad \hat{N}_n(z) = \sqrt{\frac{\pi z}{2}}\, N_{(n+\frac{1}{2})}(z) \\ \hat{I}_n(z) &= \sqrt{\frac{\pi z}{2}}\, I_{(n+\frac{1}{2})}(z) \qquad \hat{K}_n(z) = \sqrt{\frac{2z}{\pi}}\, K_{(n+\frac{1}{2})}(z) \end{aligned}\right\} \tag{13-71}$$

* The $\hat{J}$ and $\hat{N}$ functions used here are as defined by Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand, New York, 1943, p. 51. They are just z times the *Spherical Bessel Functions* as defined by Morse, *Vibration and Sound*, McGraw-Hill, New York, 1936, p. 246.

For propagation in a lossless medium, the propagation constant γ will be a pure imaginary equal to $j\beta$. Under these conditions (imaginary arguments), the $\hat{I}$ and $\hat{K}$ functions reduce to half-order Bessel and Hankel functions as follows:

$$
\begin{aligned}
\hat{I}_n(j\beta r) &= j^{n+1}\hat{J}_n(\beta r) \\
&= j^{n+1}\sqrt{\frac{\pi\beta r}{2}}\,J_{(n+\frac{1}{2})}(\beta r) \\
\hat{K}_n(j\beta r) &= j^{-n-1}[\hat{J}_n(\beta r) - j\hat{N}_n(\beta r)] \\
&= j^{-n-1}\sqrt{\frac{\pi\beta r}{2}}\,[J_{(n+\frac{1}{2})}(\beta r) - jN_{(n+\frac{1}{2})}(\beta r)] \\
&= j^{-n-1}\sqrt{\frac{\pi\beta r}{2}}\,H^{(2)}{}_{(n+\frac{1}{2})}(\beta r) \qquad (13\text{-}72)
\end{aligned}
$$

The first of these functions represents a spherical standing wave and is suitable for regions that include the origin. The second function represents an outward-traveling spherical wave, and is appropriate for regions that may extend to infinity, but do not include the origin. Applications of these functions* will be made in the following sections.

13.06 Spherical Waves and the Biconical Antenna. Schelkunoff has obtained a solution to the antenna problem by treating the antenna as an open-ended wave guide, or electro-magnetic horn. To accomplish this, he has started with a biconical antenna as a prototype for which a solution can be obtained from Maxwell's equations. In the process, it is demonstrated that for the biconical antenna the input impedance depends only on the principal wave. Therefore, for biconical antennas, the input impedance can be represented *exactly* as the input impedance of a uniform transmission line, terminated in an appropriate terminal impedance. Two methods for calculating the terminal impedance are given. Then, using the solution for the biconical antenna as a guide, the solution for cylindrical antennas is obtained by analogy. Whereas this approach necessarily involves making some approximations, the approxima-

* Of necessity, discussion of these functions in this section has been very brief. For a more thorough treatment, reference should be made to a mathematics text. An excellent treatment is given in S. A. Schelkunoff, *Applied Mathematics for Engineers and Scientists*, D. Van Nostrand, New York, 1948.

tions are justifiable on the basis of the physical picture gained from the biconical antenna theory.

Before investigating radiation from biconical antennas, it is desirable to give consideration to some of the general properties of spherical waves. It will be recalled that for *plane* waves it was found possible to divide the waves into transverse magnetic (TM), transverse electric (TE), and transverse electromagnetic (TEM) waves. For TM waves traveling in the z direction, $H_z = 0$, and the divergence equation for H is

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} = 0 \tag{13-73}$$

It follows that it should be possible to derive H from a *stream function* Π_m through the relations

$$H_x = \frac{\partial \Pi_m}{\partial y}, \qquad H_y = -\frac{\partial \Pi_m}{\partial x} \tag{13-74}$$

which relations satisfy (73). Since $H_z = 0$, Π_m may be regarded as the magnitude of a vector $\mathbf{A}'$ that is parallel to the z axis. Then eqs. (74) are given by

$$\mathbf{H} = \text{curl } \mathbf{A}'$$

where $$\mathbf{A}' = \mathbf{k}A_z, \qquad A_z = \Pi_m, \quad A_x = A_y = 0$$

Similarly for TE waves traveling in the z direction, $E_z = 0$, and the divergence equation for E (in a charge-free region) is

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = 0$$

so that, in this case, it is possible to obtain E from a stream function Π_e through the relations

$$E_x = \frac{\partial \Pi_e}{\partial y} \qquad E_y = -\frac{\partial \Pi_e}{\partial x}$$

Since $E_z = 0$, Π_e may be regarded as the scalar magnitude of a vector $\mathbf{F}$ that is parallel to the z axis. Then

$$\mathbf{E} = \text{curl } \mathbf{F}$$

where $$\mathbf{F} = \mathbf{k}F_z, \quad F_z = \Pi_e, \quad F_x = F_y = 0$$

Since TEM waves may be considered a special case of either TM or TE waves, it follows that the most general plane wave field traveling in the z direction can be expressed in terms of two scalar stream functions A_z and F_z.

The theory of spherical waves is similar to that of plane waves. There are TM spherical waves for which $H_r = 0$, TE spherical waves for which $E_r = 0$, and TEM spherical waves for which both E_r and H_r are zero. For TM spherical waves, the divergence equation for **H** reduces to

$$\frac{\partial}{\partial\theta}(\sin\theta H_\theta) + \frac{\partial H_\phi}{\partial\phi} = 0$$

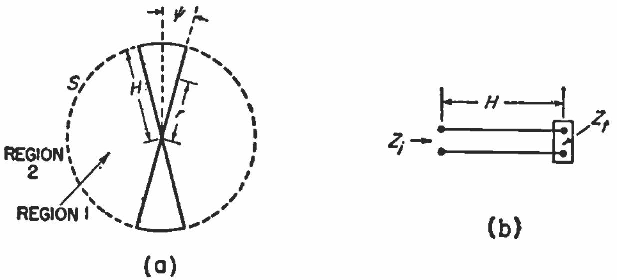


FIG. 13-15. (a) Biconical antenna and (b) its equivalent circuit (insofar as impedance is concerned).

Therefore it should be possible to obtain **H** from a stream function Π_m through the relations

$$H_\theta = \frac{1}{r\sin\theta}\frac{\partial\Pi_m}{\partial\phi} \qquad H_\phi = -\frac{1}{r}\frac{\partial\Pi_m}{\partial\theta}$$

Since $H_z = 0$, Π_m may be regarded as the magnitude of a vector **A′** which* at every point is in the direction of the r co-ordinate. Then **H** is obtained from

$$\mathbf{H} = \text{curl } \mathbf{A}'$$

where $\mathbf{A}' = \mathbf{u}_r A_r, \quad A_r = \Pi_m. \quad A_\theta = A_\phi = 0$

In a similar manner the electric field of a spherical TE wave is found to be expressible in terms of a stream function F_r. However, in dealing with biconical and cylindrical antennas, only TM (and TEM) spherical waves are encountered.

Figure 13-15 shows a biconical antenna that is assumed to be excited by a voltage applied across an infinitesimal gap at the apices.

* In general, for spherical waves, the vector $\mathbf{A}' = \mathbf{u}_r A_r$ will not be the same as the magnetic vector potential **A**.

The spherical surface S divides the space about the antenna into two regions: region I is the antenna region and region II is the outside or free-space region. The conducting cones and dielectric in region I can be considered as a wave guide that is "terminated" in a second wave guide, consisting of region II. Because of circular symmetry, currents along the cones will be radial (except at the end surfaces) and magnetic lines will be circular about the axis of the cones. That is, only transverse magnetic waves will be present.

With $H_r = 0$, the electromagnetic field about the cones can be completely specified in terms of a radial vector $\mathbf{A}' = \mathbf{u}_r A_r$. Taking the curl of $\mathbf{A}'_r$ and remembering that $\partial/\partial\phi = 0$, the magnetic intensity is given by

$$H_\phi = -\frac{1}{r}\frac{\partial A_r}{\partial \theta} \tag{13-75}$$

Then, assuming a nondissipative dielectric ($\sigma = 0$), Maxwell's equations become

$$\left.\begin{aligned} \frac{\partial}{\partial \theta}(\sin\theta H_\phi) &= j\omega\epsilon r \sin\theta E_r \\ -\frac{\partial}{\partial r}(rH_\phi) &= j\omega\epsilon r E_\theta \\ \frac{\partial}{\partial r}(rE_\theta) - \frac{\partial E_r}{\partial \theta} &= -j\omega\mu r H_\phi \end{aligned}\right\} \tag{13-76}$$

with $$H_r = H_\theta = E_\phi = 0$$

Since $E_\phi = 0$, the lines of electric intensity lie in axial planes. Also, since there is no radial magnetic current ($H_r = 0$), $\text{curl}_r\, \mathbf{E} = 0$ and the transverse electric intensity can be expressed as the gradient of a scalar potential V. That is

$$E_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta} \tag{13-77}$$

The stream function or potential A_r, from which the fields are to be obtained through (75) and (76), can be determined from the following considerations. For this problem, where there is no variation with ϕ, the separation of the wave equation in spherical co-ordinates leads to the Legendre equation (56), sec. 13.05. In region II, the free-space region, where the axis ($\theta = 0, \pi$) is included, n will be integral and the θ function solution will be given in terms of the Legendre polynomials $P_n(\cos\theta)$. The Q functions cannot be used in this region because they become infinite at the axis. Because this region extends to infinity, the appropriate radial functions will be those that represent outward-traveling waves, that is, the K functions or the spherical Hankel functions. Therefore, the expression for vector potential in this region must be of the form

$$A_r(r, \theta) = \hat{K}_n(j\beta r)P_n(\cos\theta) \tag{13-78}$$

From (78), (75), and (76) it follows that

$$rH_\phi = -\hat{K}_n(j\beta r)P_n^1(\cos\theta) \tag{13-79}$$
$$rE_\theta = \eta\hat{K}_n'(j\beta r)P_n^1(\cos\theta) \tag{13-80}$$

Expressions for E_r can be obtained from the three parts of eq. (76). Equating these expressions and using (75) the following equation in A_r is obtained,

$$r^2\frac{\partial^2 A_r}{\partial r^2} + \omega^2\mu\epsilon r^2 A_r + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial A_r}{\partial\theta}\right) = 0 \tag{13-81}$$

This equation is sometimes called a wave equation, although it is different from (13-51a) which is the wave equation in spherical co-ordinates. Separating, and letting the separation constant be $n(n + 1)$ as before, results in

$$\frac{\partial^2 A_r}{\partial r^2} = \left[\frac{n(n+1)}{r^2} - \omega^2\mu\epsilon\right]A_r \tag{13-82}$$
$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial A_r}{\partial\theta}\right) = -n(n+1)A_r \tag{13-83}$$

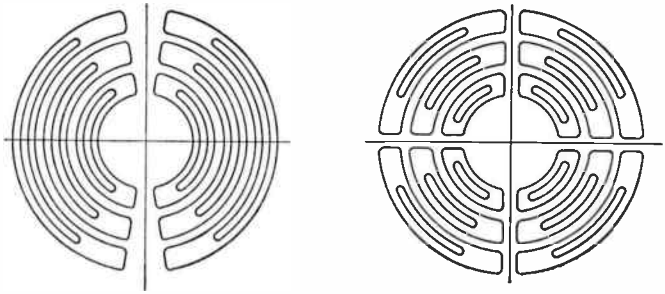
Combining (83), (76), and (75) gives an expression for E_r directly in terms of A_r:

$$\begin{aligned} j\omega\epsilon r^2 E_r &= n(n+1)A_r \\ &= n(n+1)\hat{K}_n(j\beta r)P_n(\cos\theta) \end{aligned} \tag{13-84}$$

In this region (region II), when $n = 0$ all the fields vanish, so the lowest order or principal wave is given by $n = 1$, for which

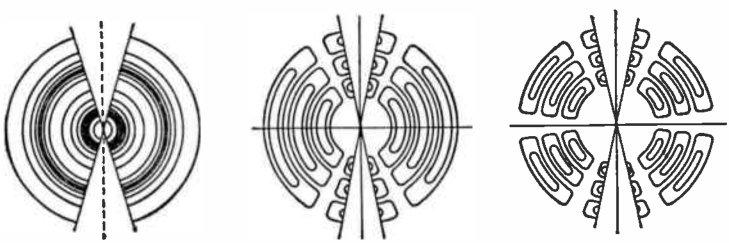
$$\left.\begin{aligned} rH_\phi &= e^{-j\beta r}\left(1 + \frac{1}{j\beta r}\right)\sin\theta \\ rE_\theta &= \eta\, e^{-j\beta r}\left(1 + \frac{1}{j\beta r} - \frac{1}{\beta^2 r^2}\right)\sin\theta \\ r^2 E_r &= 2\eta\, e^{-j\beta r}\left(\frac{1}{j\beta} - \frac{1}{\beta^2 r}\right)\cos\theta \end{aligned}\right\} \tag{13-85}$$

When multiplied by the factor $j\beta I\ dl/4\pi$, expressions (85) are exactly the expressions obtained in chap. 10 for the fields due to a current element. Thus it is seen that the simple current element generates fields that are representable by the lowest order transverse magnetic spherical waves.



(a) (b)

FIG. 13-16. Electric-field lines for first- and second-order transverse magnetic spherical waves in free space.



(a) (b) (c)

FIG. 13-17. Electric-field lines for zero-, first-, and second-order transverse magnetic spherical waves between coaxial cones. The zero-order wave is the TEM wave.

A sketch of the first- and second-order transverse magnetic spherical waves is shown in Fig. 13-16. These are for the waves in free space.

In the presence of two coaxial conductors (that is, in the antenna region I), the first- and second-order TM waves appear as shown in Figs. 13-17b and 13-17c. However, in this latter region the zero-order TM wave, that is the TEM wave, can and does exist. Its electric field lines are shown in Fig. 13-17a.

The radial impedance for outgoing waves is defined by

$$Z_r^+ = \frac{E_\theta}{H_\phi}$$

For the first-order TM wave of eqs. (85), it is

$$Z_r^+ = \frac{\eta\beta^2r^2}{1+\beta^2r^2} - \frac{j\eta}{\beta r(1+\beta^2r^2)} \tag{13-86}$$

It is noted, in passing, that at large distances this impedance approaches η, the intrinsic impedance. On the other hand, for small values of r, the radial impedance becomes a small resistance in series with a large capacitive reactance of value $-j/\omega\epsilon r$.

Within the antenna region there will exist a TEM wave as well as the higher-order waves. In general, to meet the boundary conditions, n will be nonintegral in this region and the θ function solution will be given in terms of P functions in the form

$$AP_n(-\cos\theta) + BP_n(\cos\theta)$$

An exception to this occurs for $n = 0$, which gives the TEM wave. For $n = 0$, $P_n(-\cos\theta)$ and $P_n(\cos\theta)$ are not independent solutions, and the Q function solution must be added. (The Q function is permissible in this region because the axis is excluded.) Then, for $n = 0$, the solution for A_r will have the form

$$\begin{aligned} A_r &= \hat{K}_0(j\beta r)[aP_0(\cos\theta) + bQ_0(\cos\theta)] \\ &= e^{-j\beta r}\left[a + b\ln\cot\frac{\theta}{2}\right] \end{aligned} \tag{13-87}$$

Then

$$\left.\begin{aligned} rE_\theta &= \frac{b\eta}{\sin\theta}e^{-j\beta r} \\ rH_\phi &= +\frac{b}{\sin\theta}e^{-j\beta r} \\ E_r &= 0 \end{aligned}\right\} \tag{13-88}$$

The electric field distribution for the TEM wave is seen to be the same as that obtained as a solution to Laplace's equation in the static case.

For the higher order waves between the cones, the solution will be of the form

$$A_r(r,\theta) = [a\hat{J}_n(\beta r) + b\hat{N}_n(\beta r)][a_2P_n(-\cos\theta) + b_2P_n(\cos\theta)] \tag{13-89}$$

where in general n will have nonintegral values. Now recalling from (84) that E_r is proportional to A_r, and applying the boundary conditions $E_r = 0$ at $\theta = \psi$ and at $\theta = \pi - \psi$, it follows that the second bracketed term of (89), containing the θ function, must be zero at $\theta = \psi$ and at $\theta = \pi - \psi$. Applying these conditions, it is found that

$$b_2 = -a_2$$

and
$$P_n(\cos\psi) = P_n(-\cos\psi) \tag{13-90}$$

Equation (90) may be solved for n. When this is done there results,* for small cone angles, ψ,

$$n \approx (2m+1) + \frac{1}{\ln\frac{2}{\psi}}$$
$$\approx (2m+1) + \frac{120}{Z_0} = (2m+1) + \Delta \tag{13-91}$$

where m is an integer, $\Delta = 120/Z_0$, and $Z_0 \approx 120 \ln 2/\psi$ is the characteristic impedance of the biconical antenna. As Z_0 approaches infinity (that is, as the cone angle approaches zero) n approaches an integral value, and the transmission modes approach the corresponding free-space modes.

The appropriate radial functions in the antenna region are the spherical Bessel function $\hat{J}_n$ and $\hat{N}_n$, which represent standing waves. However. the $\hat{N}_n$ cannot be used as they become infinite at the origin, which is not excluded. Except for the zero order, the $\hat{K}_n$ functions are ruled out for the same reason. It follows that the higher order waves ($n > 0$) in the antenna region will be given by

$$\left.\begin{aligned} A_r &= a\hat{J}_n(\beta r)T(\theta) \\ rH_\phi &= -a\hat{J}_n(\beta r)\frac{dT(\theta)}{d\theta} \\ rE_\theta &\doteq -ja\eta\hat{J}_n'(\beta r)\frac{dT(\theta)}{d\theta} \\ j\omega\epsilon r^2E_r &= n(n+1)a\hat{J}_n(\beta r)T(\theta) \end{aligned}\right\} \tag{13-92}$$

where
$$T(\theta) = [P_n(\cos\theta) - P_n(-\cos\theta)] \tag{13-93}$$

The current in the cones is proportional to rH_ϕ, evaluated at the surface, so the current associated with the higher order waves can

* S. A. Schelkunoff, *Electromagnetic Waves* p. 446.

be obtained from the second of eqs. (92). Now for $n > 0$, and for $r \to 0$, $J_n(r)$ varies as r^n which, of course, goes to zero as $r \to 0$. Therefore the current *at the input* $I_n(0)$, associated with the higher order waves is zero. Also the voltage caused by the higher order waves and taken along *any* meridian between the cones, can also be shown to be zero, for

$$\begin{aligned} V_n(r) &= \int_{\psi}^{\pi-\psi} rE_\theta \, d\theta \\ &= -j\eta a\hat{J}_n'(\beta r)[T(\pi - \psi) - T(\psi)] = 0 \end{aligned} \qquad (13\text{-}94)$$

for $n > 0$. Therefore the input voltage and current, and hence the input impedance, *depend only on the principal or TEM wave.* This is a very important result, because it makes it possible, without approximation, to treat the input impedance of the biconical antenna as the input impedance of a transmission line that is terminated in an appropriate impedance.

13.07 Equivalent Transmission Line and Terminal Impedance. Considering the biconical antenna as a transmission line, the voltage and current at a distance r from the origin or input terminals will be

$$\begin{aligned} V(r) &= V_0(r) \\ I(r) &= I_0(r) + \bar{I}(r) \end{aligned} \qquad (13\text{-}95)$$

where V_0 and I_0 are the principal mode ($n = 0$) values, and $\bar{I}$ is the "complementary" current due to all the higher order waves. As has already been noted, $\bar{I}(0) = 0$. Then, in terms of principal mode values, the lossless transmission line equations may be written (Fig. 13-15)

$$\begin{aligned} V_0(r) &= V_0(H) \cos \beta(H - r) + jZ_0 I_0(H) \sin \beta(H - r) \\ I_0(r) &= I_0(H) \cos \beta(H - r) + j \frac{V_0(H)}{Z_0} \sin \beta(H - r) \end{aligned}$$

where $Z_0 = 120 \ln \cot \psi/2$ is the characteristic impedance of the coaxial cones. The input impedance will be

$$Z_i = \frac{V_0(0)}{I_0(0)} = Z_0 \left[\frac{V_0(H) \cos \beta H + jZ_0 I_0(H) \sin \beta H}{Z_0 I_0(H) \cos \beta H + jV_0(H) \sin \beta H} \right] \qquad (13\text{-}96)$$

The equivalent terminal impedance Z_t will be

$$Z_t = \frac{V_0(H)}{I_0(H)} = \frac{V(H)}{I(H) - \bar{I}(H)} \qquad (13\text{-}97)$$

The equivalent terminal admittance is

$$Y_t = \frac{1}{Z_t} = \frac{I(H)}{V(H)} - \frac{\bar{I}(H)}{V(H)} = Y_{\text{caps}} + Y' \qquad (13\text{-}98)$$

The current $I(H)$ is the total current on the antenna at $r = H$ and is, therefore, just the current flow out of and into the spherical caps that are assumed to close the ends of the antenna. That is, $I(H)$ is the current flow through the capacitance between the caps, and $I(H)/V(H)$ is the admittance between the two caps. For thin

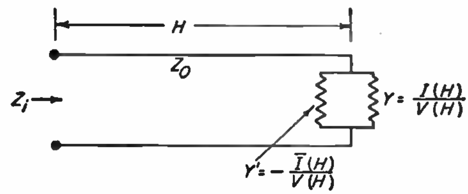


FIG. 13-18. The terminal admittance Y_t.
($Y_t = Y_1 - Y$ and $Z_t = 1/Y_t$)

antennas, the capacitance between caps is very small and $I(H)$ is approximately zero. Then

$$I_0(H) + \bar{I}(H) = I(H) \approx 0$$

or

$$I_0(H) \approx -\bar{I}(H) \qquad (13\text{-}99)$$

and

$$Y_t = \frac{I_0(H)}{V(H)} \approx -\frac{\bar{I}(H)}{V(H)}$$

also

$$Z_t = \frac{V(H)}{I_0(H)} \approx -\frac{V(H)}{\bar{I}(H)}$$

In general, the terminal admittance consists of two admittances in parallel as diagrammed in Fig. 13-18. The admittance, Y_{caps}, between caps of small radius is approximately just the capacitive susceptance $j\omega C$, where C is the electrostatic capacitance between the caps. This may be obtained by calculating the capacitance between the outside surfaces of two thin disks having radii very much smaller than their separation. The capacitance of an isolated thin circular disc treated as a very flat spheroid is found* to be

* J. H. Jeans, *Electricity and Magnetism*, Cambridge Press, London, 1946, p. 249.

$20a/9\pi\ \mu\mu$fd where a is the radius in centimeters, so the admittance between caps will be

$$Y_{\text{caps}} = \frac{ja}{30\lambda} \qquad \text{mhos}$$

where a is now the radius of the circular disks in meters.

The other part of the terminal admittance, $Y' = -\bar{I}(H)/V(H)$, is calculated from the higher order current waves at $r = H$. When the terminal impedance Z_t has been determined the input impedance will be given by

$$Z_i = Z_0\left(\frac{Z_t \cos \beta H + jZ_0 \sin \beta H}{Z_0 \cos \beta H + jZ_t \sin \beta H}\right) \tag{13-100}$$

Schelkunoff has carried out the evaluation of Y', and hence Z_t, in the following manner. Since the detailed calculations are lengthy, only an outline of the method is given here.

First, expressions for E_r in the antenna region and in the outside or free-space region are written and compared. For the antenna region, the resultant field due to the higher order waves can be expressed in the form

$$2\pi j\omega\epsilon r^2 E_r = \sum_n a_n \frac{\hat{J}_n(\beta r)}{\hat{J}_n(\beta H)} T_n(\theta) \tag{13-101}$$

where $T_n(\theta)$ is defined by (93) and where n is nonintegral, being defined by (91). Making use of (84) and (75), the corresponding expression for the complementary current will be

$$\begin{aligned} \bar{I}(r) &= 2\pi r \sin \psi H_\phi|_{\theta=\psi} \\ &= -\sum_n \frac{a_n \hat{J}_n(\beta r)}{n(n+1)\hat{J}_n(\beta H)} \sin \psi \frac{dT_n(\psi)}{d\psi} \end{aligned} \tag{13-102}$$

As $\psi \to 0$, $Z_0 \to \infty$, and $n \to 2m + 1 + \Delta$,

so that
$$\frac{dT_n(\psi)}{d\psi} \approx \frac{\Delta}{\psi} = \frac{120}{Z_0 \psi}$$

Then, for thin antennas,

$$\bar{I}(r) = -\frac{120}{Z_0} \sum_n \frac{a_n \hat{J}_n(\beta r)}{n(n+1)\hat{J}(\beta H)} \tag{13-103}$$

In the outside region E_r, can be expressed in spherical Hankel functions by

$$2\pi j\omega r^2 E_r = \sum_{\bar{n}=1}^{\infty} \frac{b_{\bar{n}} \hat{K}_{\bar{n}}(j\beta r) P_{\bar{n}}(\cos\theta)}{\hat{K}_{\bar{n}}(j\beta H)} \tag{13-104}$$

where $\bar{n}$ is integral. Equating the expressions (104) and (101) for E_r at the boundary surface $r = H$ results in

$$\sum_n a_n T_n(\theta) = \sum_{\bar{n}=1}^{\infty} b_{\bar{n}} P_{\bar{n}}(\cos\theta)$$

Now, as $\psi \to 0$ and $Z_c \to \infty$, then $n \to 2m + 1$, and $T_n(\theta) \to P_{2m+1}(\cos\theta)$. Therefore, in the limit, for infinitely thin antennas, $a_n = b_{2m+1} = \bar{n}$. Then for thin antennas, it is permissible to use the b_{2m+1} terms as first approximations for the a_n terms. The expression for the complementary current on thin antennas is then given by

$$\bar{I}(r) \approx -\frac{60}{Z_c} \sum_{m=0}^{\infty} \frac{b_{2m+1} \hat{J}_{2m+1}(\beta r)}{(2m+1)(m+1)\hat{J}_{2m+1}(\beta H)} \tag{13-105}$$

The b_{2m+1} terms can be evaluated by again considering the limiting case as $\psi \to 0$ and $Z_0 \to \infty$. For very thin antennas the current distribution approaches the sinusoidal distribution of the principal wave

$$I(r) = I_0 \sin\beta(H - r)$$

with

$$I_0 = \frac{jV_0(H)}{Z_0} \tag{13-106}$$

For this distribution the fields have been calculated in chap. 10. By expanding in terms of Legendre polynomials the distant field expression for E_r obtained from chap. 10 and comparing it with eq. (104), the b coefficients can be evaluated.

The result is:

$$b_{2m+1} = -jI_0(4m+3)\hat{J}_{2m+1}(\beta H)[\hat{J}_{2m+1}(\beta H) - j\hat{N}_{2m+1}(\beta H)]$$

Inserting this in (105) and combining with (104) gives for the complementary current

$$\bar{I}(r) = -\frac{60 V_0(H)}{Z_0^2} \sum_{m=0}^{\infty} \frac{4m+3}{(m+1)(2m+1)} [\hat{J}_{2m+1}(\beta H) - j\hat{N}_{2m+1}(\beta H)]\hat{J}_{2m+1}(\beta r) \tag{13-107}$$

Then the terminal admittance is

$$Y_t \approx -\frac{\bar{I}(H)}{\bar{V}(H)} = \frac{Z_a(\beta H)}{Z_0^2} = \frac{R_a(\beta H) + jX_a(\beta H)}{Z_0^2} \quad (13\text{-}108)$$

where $$R_a(\beta H) = 60 \sum_{m=0}^{\infty} \frac{4m+3}{(m+1)(2m+1)} \hat{J}^2_{2m+1}(\beta H) \quad (13\text{-}109)$$

$$X_a(\beta H) = -60 \sum_{m=0}^{\infty} \frac{4m+3}{(m+1)(2m+1)} \hat{J}_{2m+1}(\beta H)\hat{N}_{2m+1}(\beta H)$$

The terminal impedance Z_t is

$$Z_t = \frac{Z_0^2}{Z_a(\beta H)} \quad (13\text{-}108\text{a})$$

$Z_a(\beta H)$ is the *inverse* of the terminal impedance. The input impedance of a quarter-wave section of lossless line having a characteristic impedance Z_0 and terminated in $Z_a(\beta H)$ is Z_t.

Although it is possible to calculate Z_a directly from expressions (109), the series converge slowly and are not useful for computations except when βH is small. Schelkunoff has circumvented this difficulty by providing an ingenious alternative method for calculating Z_a. Using (100), the input impedance can be expressed in terms of Z_a by

$$Z_i = Z_0 \frac{Z_a \sin \beta H - jZ_0 \cos \beta H}{Z_0 \sin \beta H - jZ_a \cos \beta H} \quad (13\text{-}110)$$

$$= \frac{Z_a - jZ_0 \cot \beta H}{1 - jZ_a/Z_0 \cot \beta H}$$

As $Z_0 \to \infty$,

$$Z_i \approx (Z_a - jZ_0 \cot \beta H)\left(1 + j\frac{Z_a}{Z_0} \cot \beta H\right)$$

$$Z_i \to \frac{Z_a}{\sin^2 \beta H} - jZ_0 \cot \beta H \quad (13\text{-}110\text{a})$$

and, since the input current approaches $I_0 \sin \beta H$, the input power (complex) becomes

$$\tfrac{1}{2}Z_iI_0^2 \sin^2 \beta H = \tfrac{1}{2}[Z_a - jZ_0 \sin \beta H \cos \beta H]I_0^2$$

However, as $Z_0 \to \infty$, the current distribution approaches the sinusoid, and the complex input power can be obtained by the induced-emf method of chap. 11. The real part, which gives the radiation resistance, will be *independent* of the antenna shape (for thin antennas) and will, therefore, be the same as that already calculated for the infinitely thin cylindrical antenna. However, the reactive part, which determines the reactance, will be a function of shape even for thin antennas, and so must be calculated for conical

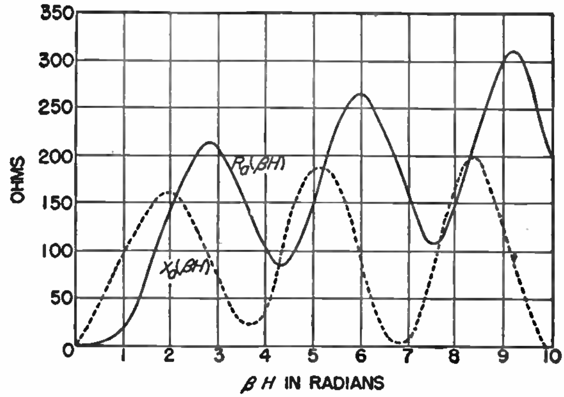


FIG. 13-19. Resistive and reactive components of the inverse terminal impedance $Z_a = R_a + jX_a$.

antennas (with $\psi \to 0$). Using this approach, Schelkunoff obtains the following results

$$\begin{aligned} R_a(\beta H) = {} & 60(\gamma + \ln 2\beta H - \text{Ci } 2\beta H) \\ & + 30(\gamma + \ln \beta H - 2\text{Ci } 2\beta H + \text{Ci } 4\beta H) \cos 2\beta H \\ & + 30(\text{Si } 4\beta H - 2\text{Si } 2\beta H) \sin 2\beta H \\ X_a(\beta H) = {} & 60\text{Si } 2\beta H + 30(\text{Ci } 4\beta H - \ln \beta H - \gamma) \sin 2\beta H \\ & - 30\text{Si } 4\beta H \cos 2\beta H \end{aligned}$$

where $\gamma = 0.5772$ (Euler's constant)

These expressions are plotted in Fig. 13-19. The input impedance is then obtained from

$$Z_i = Z_0 \frac{Z_a \sin \beta H - jZ_0 \cos \beta H}{Z_0 \sin \beta H - jZ_a \cos \beta H} \tag{13-110}$$

It is important to note that, although the approximate relation (110a) was used in calculating Z_a, it is necessary to use the exact expression (110) for calculating Z_i. Use of the approximate expression here would lead to the same answer as is given by the induced-emf method.

The final result of this attack on the problem is seen to be a surprisingly simple one. The input impedance of the conical antenna is calculated as the input impedance of a lossless transmission line which is terminated by an impedance Z_t. This terminal impedance is just the inverse of an impedance Z_a, which can be calculated by the induced-emf method.

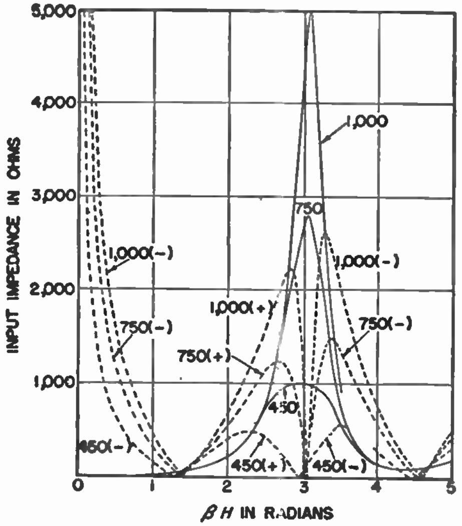


FIG. 13-20. The input impedance of hollow conical antennas for various values of Z_0 as given by Schelkunoff.

The input impedance of hollow conical antennas is shown in Fig. 13-20 for various values of Z_0. The term hollow refers to the fact that the cap capacitance has not been taken into account. For thin antennas, the cap capacitance has negligible effect, but for thicker antennas it must be accounted for. This can be done by adding the admittance Y_{caps} to the calculated value of $Y_t \approx Y'$. With this correction (108) becomes

$$Y_t = \frac{R_a(\beta H)}{Z_0^2} + j\left[\frac{X_a(\beta H)}{Z_0^2} + \omega C\right] \qquad (13\text{-}111)$$

Two effects of considerable practical importance can be observed in the curves of Fig. 13-20. The first of these is that the fatter antennas (lower Z_0) have very much smaller impedance variations with frequency, so that a fat cone is inherently a wide-band antenna. The second effect is the shortening of the resonant length for the thicker antennas. For very thin antennas resonance occurs for lengths just slightly shorter than multiples of $\lambda/4$, but for thicker antennas the shortening effect becomes quite large, especially for first resonance. The cap capacitance acts to decrease the resonant lengths still further.

13.08 Impedance of Cylindrical Antennas. The analysis for conical antennas can be extended to cover antennas of other shapes in the following manner. If the transverse dimensions of the antenna are small, the waves along it will be nearly spherical, whatever its shape. Then such antennas can be treated as nonuniform transmission lines whose inductances and capacitances per unit length and characteristic impedance vary along the line. The terminating impedance will be as calculated from (108a), except that an "average" characteristic impedance must be used for Z_0. From the theory of nonuniform transmission lines, Schelkunoff has obtained for the input impedance of antennas

$$Z_i = Z_0\,(\text{av}) \left[\frac{R_a \sin \beta H + j[(X_a - N) \sin \beta H - (Z_0\,(\text{av}) - M) \cos \beta H]}{[(Z_0\,(\text{av}) + M) \sin \beta H + (X_a + N) \cos \beta H] - jR_a \cos \beta H}\right] \tag{13-112}$$

where, for cylindrical* dipoles of radius a and half-length H,

$$M = 60(\ln 2\beta H - \text{Ci}\, 2\beta H + \gamma - 1 + \cos 2\beta H)$$
$$N = 60(\text{Si}\, 2\beta H - \sin 2\beta H)$$
$$Z_0\,(\text{av}) = 120\left(\ln \frac{2H}{a} - 1\right)$$

In Fig. 13-21 are shown curves for the input resistance and reactance of hollow cylindrical antennas for various values of Z_0 (av). Z_0 (av) for cylindrical antennas of half-length H is plotted in Fig. 13-22 as a function of the ratio H/a. For a monopole antenna Z_0 (av) has just one-half the value it has for the corresponding dipole. The

* For the M and N functions for antennas of other shapes the reader should refer to S. A. Schelkunoff, *Electromagnetic Waves*, D. Van Nostrand, New York, 1943, p. 461.

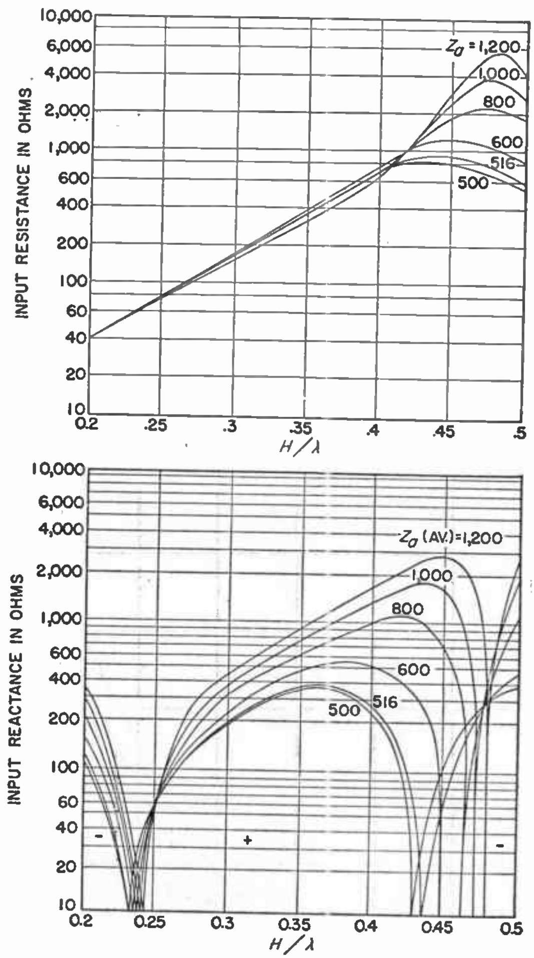


FIG. 13-21. (a) Input resistance and (b) input reactance of hollow cylindrical center-fed dipole antennas as given by Schelkunoff. For monopole antennas of height H, divide the ordinates and the value shown for Z_0 (av) by two.

input resistance and reactance of a monopole antenna are just one-half those of the corresponding dipole antenna that has the same H/a. Therefore, the input impedance of monopole antennas can be obtained from Fig. 13-21 by dividing the ordinates and Z_0 (av) by 2. Figure 13-23 shows the resonant impedance of hollow cylindrical dipole antennas at the first and second resonance points.

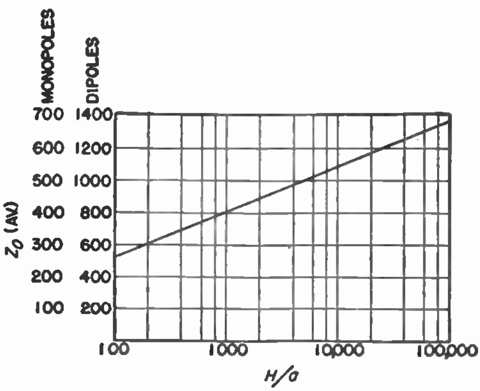


FIG. 13-22. Average characteristic impedance, Z_0 (av), for cylindrical antennas.

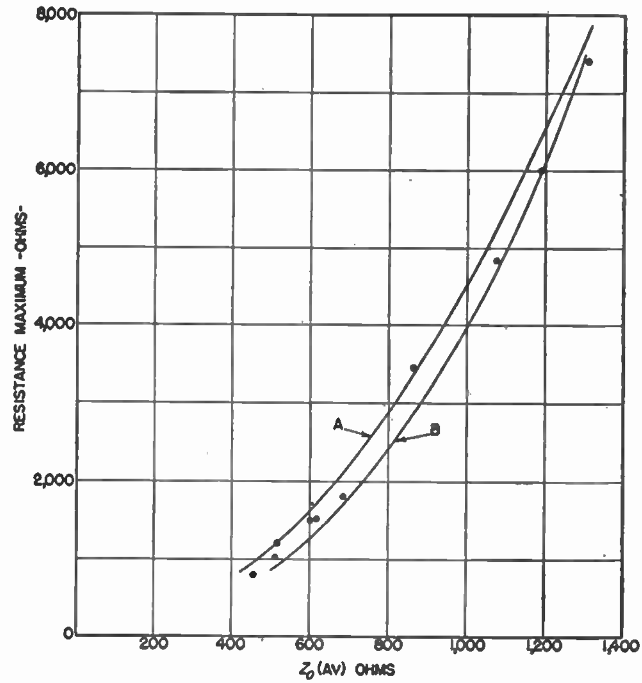


FIG. 13-23. Resonant impedance of hollow cylindrical antennas as a function of Z_0 (av) at second resonance (anti-resonance): *A*, Hallén, *Nova Acta Upsal*, 1938, Formula 39; *B*, Schelkunoff. Points are measured values from various sources.

Also shown for the second resonance point are some experimental values (circles) obtained from various sources, as well as a curve of theoretical values as given by Hallén. The agreement between theories and with experimental results at this critical point is seen to be quite good. Agreement at other points will, in general, be found to be even closer.

Schelkunoff's antenna theory is important for two reasons. First, it has provided reasonably accurate numerical answers over a fairly wide range of antenna dimensions. Second, the method itself is an excellent example of how and when to make approximations. In engineering, most problems are not amenable to exact solutions. Therefore the ability to make approximations can spell the difference between success and failure in the solution of the problem.

ADDITIONAL PROBLEMS

3. Using Schelkunoff's method, calculate the input impedance of a uniform cross section tower antenna at 1300 kc. The tower is 400 ft high and 6½ ft square. The base-insulator capacitance is 30 $\mu\mu$f.

4. The antenna for a portable test transmitter consists of a tubular steel mast 2 in. in diameter and 50 ft high. The base insulator has an effective shunting capacitance of 15 $\mu\mu$f. (a) If a test survey is to be made at 650 kc, determine (1) the radiation resistance, (2) the antenna reactance, (3) the input impedance, including the effect of the base capacitance. (b) for 1 amp through an ammeter in the lead to the antenna, what is the current in the mast near the base? (c) for an ammeter reading of 1 amp, what is the field intensity at 1 mile and how much power is radiated?

BIBLIOGRAPHY

Brillouin, L., "Antennae for Ultra-high Frequencies," *Electrical Communication*, **21**, 4, 257 (1944); **22**, 1, 11 (1944).

Schelkunoff, S. A., "Theory of Antennas of Arbitrary Size and Shape," *Proc. IRE*, **29**, 439 (1941).

Schelkunoff, S. A., "Transmission Theory of Spherical Waves," *Trans. AIEE*, **57**, 744–750 (1938).

Fubini, E. G., and Sutro, P. J. "A Wide-Band Transformer from an Unbalanced to a Balanced Line," *Proc. IRE*, **35**, 10, 1153 (1947).

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand, New York, 1943.

Aharoni, A., *Antennae, An Introduction to their Theory*, Oxford Press, London, England, 1946.

CHAPTER 14

ANTENNA PRACTICE AND DESIGN

The theory of antennas is the same at 1000 mc as it is at 100 kc, but the form that practical antennas and their matching networks take is very much a function of frequency. At low and medium frequencies, where a wavelength is long, practical antennas are usually short in wavelengths and the problem is chiefly that of efficiency. At high and very high frequencies, where a half-wave antenna has reasonable dimensions, good efficiency is more easily obtained, and the problem is usually that of obtaining directivity or gain. At ultrahigh and superhigh frequencies, the problems of "beam-shaping" become important. In addition to these general considerations, specific applications often entail particular requirements that must be met in the antenna design. In this chapter a few typical antenna systems will be considered in just enough detail to illustrate the problems and methods of antenna practice.

14.01 Low-frequency Practice—(Electrically Short Antennas). At frequencies below the broadcast band the difficulties of constructing antennas that have appreciable electrical length become very great, and so electrically short antennas must be considered. Moreover, because of the large attenuation of the horizontal component of the surface wave (see sec. 16.03), and the cancellation effect of the negative image of horizontal antennas (sec. 12.08), only the vertical portions of an antenna will be effective* in signal pickup or radiation at these frequencies. Therefore the problem becomes that of designing a vertical ground-based antenna having an effective half-length (or effective "height" in the old usage) that is as large as possible. This design is accomplished by making the

* An exception to this general statement occurs in the case of a Beverage wave antenna, which is responsive to the horizontal component of a forward-tilted, "vertically polarized" wave.

physical height as great as possible, and by *top loading*. Top loading, which consists of adding some form of "capacity hat" or a horizontal portion to the structure as in the familiar *L*- and *T*-type antennas, serves two purposes. It increases the effective half-length by a factor of two (at most), and it decreases the large capacitive reactance of the short antenna. For both transmitting and receiving antennas a large capacitive reactance means low efficiency because of the losses in the high-reactance tuning coil that is required. In addition, in the case of short transmitting antennas, the amount of power that can be fed to the antenna without voltage breakdown is dependent on the antenna reactance. Thus the voltage required to establish the antenna current is approximately $V = IX_a$, and the power radiated is $I^2R_a \approx V^2R_a/X_a^2$, so that for short antennas having low values of R_a and high values of X_a, extremely large driving voltages are required to radiate moderate amounts of power. In the case of short receiving antennas, the presence of unavoidable shunt capacitances at the receiver input reduces the signal available to the receiver when the antenna capacitive reactance is large.

Impedance and Efficiency of Short Antennas. For electrically short antennas the resistance and reactance can be expressed approximately by the following simple and convenient relations. For short vertical ground-based antennas without top loading the radiation resistance referred to the base is given by the expression developed in chap. 10, viz.,

$$R_{\text{rad}} \approx 10(\beta H)^2 \approx 400\left(\frac{H}{\lambda}\right)^2 \qquad (14\text{-}1)$$

For top-loaded antennas having a total electrical length $(H + b)$ less than one-tenth wavelength, the radiation resistance can be obtained from

$$R_{\text{rad}} = 40(\beta H)^2\left[1 - \frac{H}{H + b} + \frac{1}{4}\left(\frac{H}{H + b}\right)^2\right] \qquad (14\text{-}2)$$

In this expression, H is the height of the vertical portion of the antenna, and b is the equivalent additional length of vertical portion that would draw the same current as does the capacitive loading (Fig. 14-1). For *T*- and *L*-type antennas b is approximately equal to the length of the horizontal portion.

The approximate reactance of short unloaded antennas can be calculated from the expression

$$X_a = -Z_0' \cot \beta H \qquad (\text{for } H < 0.2\lambda) \tag{14-3}$$

where

$$Z_0' = 60\left[\ln \frac{H}{a} - 1\right] \tag{14-4}$$

is an "adjusted" characteristic impedance that is picked to give the closest fit to experimentally measured values. It is found that, if either the value of Z_0 given by expression (13-22) or that given by Siegel and Labus (13-27) is used in eq. (3), the values of reactance so calculated are consistently higher than measured

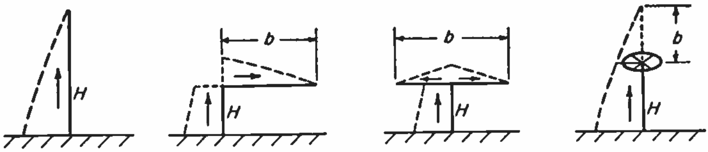


FIG. 14-1. Top-loaded antennas.

values. The use of expression (4) in eq. (3) yields calculated reactances that are in better agreement with measured reactances. There appears to be no theoretical basis for expression (4), and it must be regarded simply as an empirical formula which, when used with (3), gives a reasonably close check with values obtained by measurement.

The efficiency of electrically short antennas tends to be low because the radiation resistance is small and the loss resistances of the antenna system may be comparable with, or even considerably greater than, the radiation resistance. For short antennas the chief losses occur in the ground system and the loading coil. Ordinarily, ohmic losses in the antenna itself will be small. Ground system losses will be considered under broadcast frequency antennas where this subject has received much attention. For short antennas, where the antenna reactance is capacitive, it becomes necessary to "tune out" this capacitive reactance either by means of a series loading coil, or by use of an L, T, or π matching network. Regardless of the matching circuit used, there will always be some loss in it that will be at least as large as the loss that occurs in a simple loading coil, so the efficiency to be expected can be estimated on the basis of the latter connection. In the actual and equivalent antenna coupling circuits indicated in Fig. 14-2, the power radiated will be

$$P_r = I_a^2 R_a = \frac{I_a^2 X_a}{Q_a}$$

The power lost in the coil will be approximately

$$P_L \approx I_a^2 R_L \approx \frac{I_a^2 X_L}{Q_L}$$

$Q_a = X_a/R_a$ and $Q_L = X_L/R_L$ are the Q's of the antenna and loading coil respectively. Remembering that in the matched condition

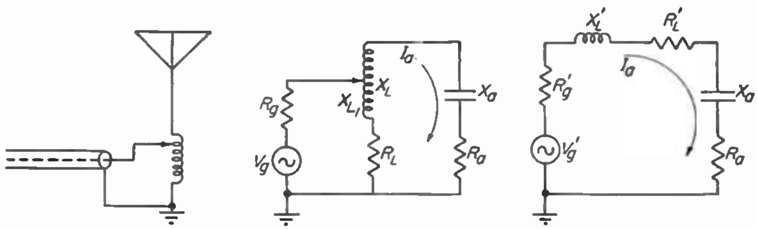


FIG. 14-2. An actual antenna coupling circuit and its equivalent circuits.

X_L and X_a are nearly equal, it is seen that the efficiency is given by

$$\text{Efficiency} = \frac{Q_L}{Q_a + Q_L}$$

For *very* short antennas, $Q_a \gg Q_L$ and $X_a \approx Z_0'/\beta H$ so that

$$\text{Efficiency} \approx \frac{Q_L}{Q_a} \approx \frac{10Q_L}{Z_0'} (\beta H)^3 \qquad (14\text{-}5)$$

Equation (5) shows that for very short vertical antennas without top loading the efficiency varies approximately as the *cube* of the antenna length.

Considerations similar to the above also apply in the case of short receiving antennas. The effective length, and hence the induced voltage, of a short receiving antenna is proportional to its physical half-length H, and, as was shown previously, its radiation resistance is proportional to the square of H. Therefore, in theory, the maximum power which can be absorbed by a matched load, viz.,

$$W = \frac{V^2_{\text{ind}}}{4R_{\text{rad}}}$$

is *independent* of the half-length H, for small values of H. Actually, when coupling circuit losses are taken into account it can be shown (see problem 6) that the efficiency, and, therefore, the maximum useful received power tends to vary as the cube of the length for very short antennas.

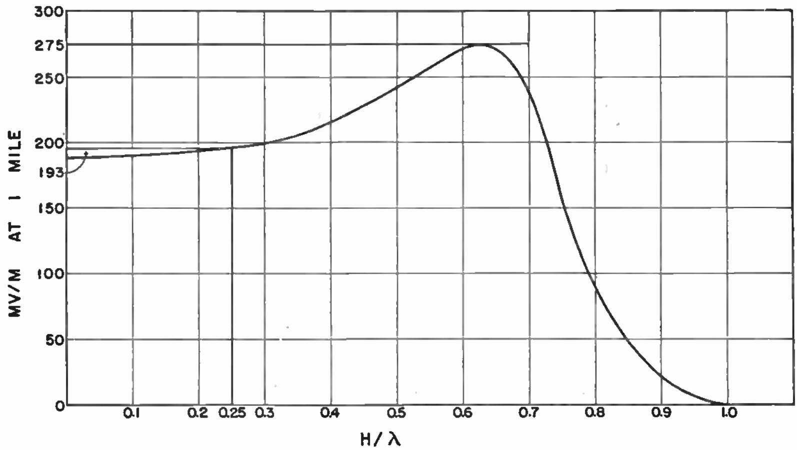


FIG. 14-3. Field intensity at one mile at $\theta = 90$ degrees for one kw radiated from a ground-based vertical monopole of height H.

14.02 Broadcast Antennas. Because of their economic importance broadcast antennas have received a great deal of attention in the literature. The factors of most concern are the height and current distribution (which determine the vertical pattern), the driving-point impedance, losses, efficiency, and, in the case of an array, the horizontal pattern.

As the height H of a ground-based vertical antenna is increased from a very short height, the field intensity on the horizon ($\theta = 90$ degrees) for a given power input first increases, and then decreases as shown* in Fig. 14-3. This dependence of the field intensity upon

* Stuart Ballantine, "On the Optimum Transmitting Wavelength for a Vertical Antenna over a Perfect Earth," *Proc. IRE*, 12, 833 (1924).

height of the antenna is a result of the change in vertical pattern with height of antenna. This effect is illustrated in Fig. 12-2 for the equivalent dipoles of length $L = 2H$. For H very small, the vertical pattern is given by sin θ, and the shape of this pattern changes quite slowly with height to $H = \lambda/4$, where the pattern is given by Fig. 12-2b. (Of course, only the top half of this pattern applies for ground-based monopoles.) Because the high-angle (small θ) radiation has been decreased, the field intensity at $\theta = 90$ degrees for a given power radiated will be greater, but at $H = \lambda/4$ the increase in field intensity over a very short monopole is only about 7 per cent. Above $H = \lambda/4$ the field intensity at $\theta = 90$ degrees continues to increase, reaching a maximum at $H = 0.64\ \lambda$, and then decreasing sharply to zero at $H = \lambda$. The decrease above $H = 0.64\ \lambda$ is due to the fact that the secondary, high-angle lobe, which begins to appear above $H = 0.5\ \lambda$, is now quite large, and hence more power is being radiated at these high angles. In broadcast work where the desired coverage is obtained by means of the surface wave alone (the wave radiated at $\theta = 90$ degrees), this high-angle radiation is deleterious for two reasons. First, it takes power which otherwise could be used to increase the field intensity of the main lobe. Second, at night, instead of being absorbed in the ionosphere as they are during the day, the waves radiated at these high angles are reflected back to earth, giving strong signals hundreds and even thousands of miles from the transmitter. In the early days of broadcasting this was a desirable result, but with the present large number of broadcast stations requiring approximately 20 stations per channel, this sky wave transmission causes severe interference with the local coverage of transmitters on the same frequency or adjacent channels, and so must be reduced to a minimum. For this reason, although the theoretical "optimum" height of 0.64 λ gives a maximum value of low-angle radiation, a somewhat smaller height gives a better ratio of low-angle to high-angle radiation. A height of 0.59 λ is one that is often used in present-day practice.

The vertical pattern of a broadcast antenna is also dependent to a small extent upon the change in cross section with height. A uniform cross section antenna has a current distribution that differs only slightly from a sinusoidal distribution. However when a tower is tapered from a large cross section at the bottom to a small cross section at the top, as it often is for mechanical reasons, the current distribution departs from the sinusoidal. The chief effect is that the current near the top is less than that which would result if the tower were uniform. For diamond-shaped antennas, the effect is more complicated. The current amplitude tends to be increased above the sinusoid, where the cross section is large, and decreased below it, where the cross section is small. Tapered towers are often self-supporting, but uniform and diamond-shaped antennas must be guyed. In general, the presence of these steel guy wires will affect both the pattern and impedance of the antenna, but these effects can be kept small by sectionalizing the guy wires with insulators, so that the current flowing in each short high-reactance section is kept small.

Losses and Efficiency. Losses in an antenna system may be divided into four classifications: (1) ohmic losses in the conductor, (2) dielectric losses in the base insulator, (3) losses in the coupling coil or matching network, (4) losses in the ground system. Ohmic losses in the conductor are negligibly small provided that the antenna cross section is sufficiently large (a condition always met in tower antennas), and that the tower members are thoroughly bonded. The power loss in the insulator(s) is generally quite small and almost always less than 1½ per cent of the total power input for broadcast antennas. However, for electrically short antennas that require large driving voltages, the percentage power loss in the insulator may become large, especially in wet weather. Losses in the matching network can be kept small by proper design, and by use of low-loss inductors. In general, losses occurring in the ground adjacent to the antenna will be quite large unless an adequate ground system is used. These ground losses may be divided into V^2G or "dielectric" losses (which are proportional to the base voltage) and I^2R or "ohmic" losses (which are proportional to the antenna current). The "dielectric" loss occurs near the base in the layer of earth above the ground wires. It is important only when the base voltage is high, as for example in the case of electrically short antennas. This type of loss may be reduced by use of a ground screen placed above the earth in the immediate vicinity of the antenna. The "ohmic" loss is due to the ground "return" current flowing through the finite impedance of the earth. This loss can be calculated (approximately) by computing the magnetic intensity at the surface of the earth in the neighborhood of the antenna on the basis of no loss (perfectly conducting earth). Then, assuming that the fields, and hence the ground currents, do not change appreciably when the earth has loss, the radial surface current density J_ρ is known, and therefore the loss per square meter $J_\rho^2 R_s$ can be calculated. The percentage of power lost in the ground depends upon the antenna height and the conductivity of the earth. It varies from a few per cent for a half-wave monopole on an earth of good conductivity up to about 75 per cent for a short monopole on a poorly conducting earth. However, in all cases these losses can be reduced almost to zero by using a suitable ground system. A grounding system composed of about 120 radial wires, each about half a wavelength long and buried a few inches beneath the surface of the earth, has proven to give almost total reduction of ground losses. Fewer wires can be used, but the loss reduction is then not so complete.

Broadcast Antenna Arrays. More often than not it is desirable to modify the horizontal pattern of the field intensity radiated by a broadcast transmitter, and this requires an array of two or more antennas. One reason for changing the horizontal radiation characteristic from a circular pattern exists when the potential audience lies on one side of the station rather than all around it. A more compelling reason exists when a station desires a frequency and power that will give more than local coverage. The Federal Communications Commission then requires that the station install a directional array that will "protect" stations on the same or adjacent channels, that lie within the radius of several hundred miles. Such protection consists of reducing the field intensities below certain specified levels in those particular directions, and this, in turn, requires a directional pattern which has a certain definite number and location of nulls or minima. In general, the more stations that must be protected, the greater will be the number of antennas required for the array. Arrays consisting of from two to five elements are common. The calculation of the patterns of such arrays has been considered in chap. 12. However after the correct current ratios have been computed there remains the problem of determining the driving point impedances, and designing the matching networks to feed the elements with currents that have the correct magnitudes and phases. The design of matching networks that will produce a specified phase shift is a circuit problem that has been solved* to give the required reactances for the elements of a T or π section in terms of the impedances to be matched and the phase shift desired. A simple example of a broadcast array design problem will be given.

The equivalent T or π section of a network of impedances can be expressed in terms of the image impedances Z_{i_1} and Z_{i_2}, and the image transfer constant Θ by the following relations.†

For a T section

$$\left.\begin{aligned} Z_1 &= \frac{Z_{i_1}\cosh\Theta - \sqrt{Z_{i_1}Z_{i_2}}}{\sinh\Theta} \\ Z_2 &= \frac{Z_{i_2}\cosh\Theta - \sqrt{Z_{i_1}Z_{i_2}}}{\sinh\Theta} \\ Z_3 &= \frac{\sqrt{Z_{i_1}Z_{i_2}}}{\sinh\Theta} \end{aligned}\right\} \tag{14-6}$$

For a π section,

$$\left.\begin{aligned} Z_A &= \frac{Z_{i_1}Z_{i_2}\sinh\Theta}{Z_{i_2}\cosh\Theta - \sqrt{Z_{i_1}Z_{i_2}}} \\ Z_B &= \sqrt{Z_{i_1}Z_{i_2}}\sinh\Theta \\ Z_C &= \frac{Z_{i_1}Z_{i_2}\sinh\Theta}{Z_{i_1}\cosh\Theta - \sqrt{Z_{i_1}Z_{i_2}}} \end{aligned}\right\} \tag{14-7}$$

For matching networks of pure reactances the complex image transfer constant Θ will reduce to the pure imaginary jB, and, considering those cases where the image impedances are pure resistances ($Z_{i_1} = R_1$, $Z_{i_2} = R_2$), the appropriate expressions for T and π reactance networks are

* A complete treatment of the method, including design curves for determining the values of reactances for the matching networks, is given in the article by W. L. Everitt, "Coupling Networks," *Communications*, **18**, 12, Sept. (1938): *also* **18**, 12, Oct. 1938. The design curves are also shown in Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943.

† E.g., W. L. Everitt, *Communication Engineering*, McGraw-Hill, New York, 1937, p. 278.

T section

$$\left.\begin{aligned} Z_1 &= -j\,\frac{R_1 \cos B - \sqrt{R_1R_2}}{\sin B} \\ Z_2 &= -j\,\frac{R_2 \cos B - \sqrt{R_1R_2}}{\sin B} \\ Z_3 &= -j\,\frac{\sqrt{R_1R_2}}{\sin B} \end{aligned}\right\} \qquad (14\text{-}8)$$

π section

$$\left.\begin{aligned} Z_A &= j\,\frac{R_1R_2 \sin B}{R_2 \cos B - \sqrt{R_1R_2}} \\ Z_B &= j\,\sqrt{R_1R_2}\,\sin B \\ Z_C &= j\,\frac{R_1R_2 \sin B}{R_1 \cos B - \sqrt{R_1R_2}} \end{aligned}\right\} \qquad (14\text{-}9)$$

The angle B is the phase shift (lag) introduced by the network. If the impedances to be matched have reactance as well as resistance, a T section can be designed to match the resistances, and then a sufficient reactance can be added to each of the values calculated for the series arms to cancel the reactances of the terminating impedances. If a π network is desired, the resulting T can then be transformed to the equivalent π.

An L network can also be used to match two resistances. However, the phase shift then cannot be chosen at will, but is determined by the image impedances. Formulas for an L network are obtained by putting one series arm, say Z_2 of a T section equal to zero. Then for an L reactance network ($Z_2 = 0$),

$$\begin{aligned} \cos B &= \sqrt{\frac{R_1}{R_2}} \qquad Z_1 = j\,\sqrt{R_1(R_1 - R_2)} \\ Z_3 &= -jR_2\sqrt{\frac{R_1}{R_2 - R_1}} \end{aligned} \qquad (14\text{-}10)$$

(For this L network R_2 must be greater than R_1, but of course the network can always be turned around.)

EXAMPLE 1:

Two quarter-wavelength tower antennas spaced one quarter-wavelength apart are to be fed with equal currents, but with the current I_B lagging I_A

by 90 degrees. Design the appropriate matching networks. The antennas are to be supplied by 70-ohm concentric cable, and the transmitter is designed to feed a 70-ohm line.

An appropriate feeding arrangement is shown in Fig. 14-4. The first step is to find the driving-point impedances of the antennas. From Fig.

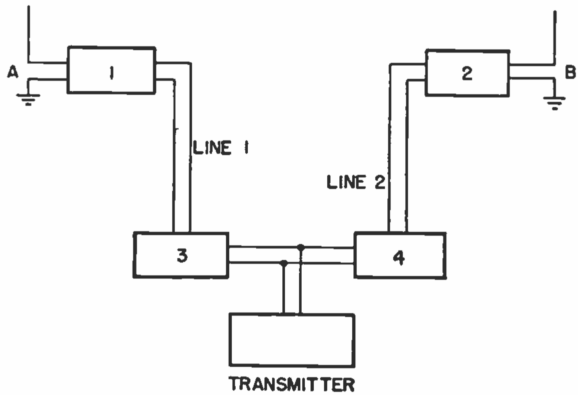


FIG. 14-4. Feeding arrangement for a two-element broadcast antenna array.

11-6 the mutual impedance between two quarter-wave monopoles at $\lambda/4$ spacing is

$$Z_{12} = 25\underline{/-35^\circ} = 21 - j14$$

For the mesh impedances Z_{11} and Z_{22}, the self-impedance of a quarter-wave monopole can be used. That is,

$$Z_{11} = Z_{22} \approx 37 + j22$$

For an array that has already been erected Z_{11}, Z_{22}, and Z_{12} can be obtained by measurement. However, ordinarily, preliminary calculations at least must be made before construction starts. From eq. (11-9), and remembering that $I_A = jI_B$, the driving-point impedances will be

$$Z_A' = \frac{V_A}{I_A} = 37 + j22 - j(21 - j14) = 23 + j1$$

$$Z_B' = \frac{V_B}{I_B} = (37 + j22) + j(21 - j14) = 51 + j43$$

Since a single three-element network can give the desired phase shift, the other three networks can be designed for impedance matching only. L networks are the simplest and most efficient for this purpose.

Network 1 must match the impedance $23 + j1$ to a resistance of 70 ohms. Then $R_1 = 23$ and $R_2 = 70$ and from eqs. (10) the required reactance to match these resistances are

$$Z_1 = j32.9, \qquad Z_3 = -j49.0$$

with $\cos B_1 = 0.70$ or $B_1 = +55$ degrees

If a reactance of $-j1$ is added to Z_1, it will cancel the positive 1-ohm reactance of the load impedance.

Therefore

$$Z_1 = j32.9 - j1 = j31.9$$

In a similar manner for network 2 it is found that for this network

$$Z_1 = -j11.9 \qquad Z_2 = -j114.8 \qquad B_2 = 31 \text{ degrees}$$

Networks 3 and 4 must be designed to satisfy the following requirements.

(1) The input resistances of the two networks must be such that each will absorb the proper amount of power for its respective antenna. Therefore the input resistance must have the ratio $R_4/R_3 = P_A/P_B = {}^{23}\!/\!_{51}$.

(2) The two networks in parallel must present a resistance of 70 ohms to the transmitter.

(3) The phase shifts must be such that the total phase shift between generator and antenna B will be 90 degrees greater than the phase shift between generator and antenna A. To satisfy simultaneously conditions (1) and (2), the required input resistances are

$$R_3 = 225 \text{ ohms} \qquad R_4 = 101.2 \text{ ohms}$$

Designing network 3 as an L network to match a resistance 225 ohms (call this R_2) to 70 ohms (call this R_1) gives for this network

$$Z_1 = j104.2 \qquad Z_2 = -j151.2 \qquad B = 56 \text{ degrees}$$

In general the line lengths to the two antennas will not be equal. Assume line 2 is 70 electrical degrees larger than line 1. Then

$$B_1 + B_3 + 90^\circ = B_2 + B_4 + 70^\circ$$
$$B_4 = 100^\circ$$

Thus, for network 4,

$$B = 100^\circ \qquad R_1 = 70 \qquad R_2 = 101.2$$

Using a π network, the required values for the elements are

$$Z_A = -j68.6 \qquad Z_B = j82.6 \qquad Z_C = -j72.8$$

The complete system will appear as shown in Fig. 14-5.

14.03 High-frequency (Short-wave) Antennas. The high-frequency band nominally covers the range of frequencies from 3 to 30 mc, but for the purposes of this section it will be considered to include frequencies down to about 2 mc. As the frequency is increased above the broadcast band, attenuation of the surface wave becomes very large and propagation by such means is restricted to local coverage within a few miles of the transmitter. Longer distance transmission at the high frequencies makes use of sky wave or ionospheric propagation. The subject of ionospheric propagation

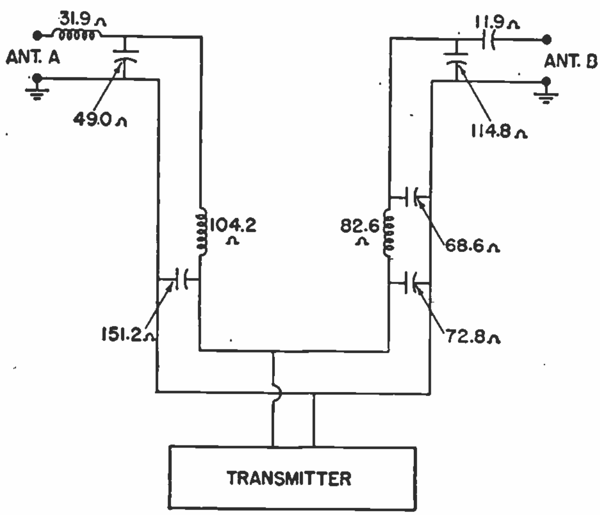


FIG. 14-5

is dealt with in chap. 17, and here it will be sufficient to point out that the higher frequencies in this band are useful for long-distance communication up to several thousand miles, whereas the lower frequencies are suitable for communication over distances less than about 500 miles. The higher frequencies suffer less attenuation, but are not reflected back to earth at the high angles of incidence required for short-distance communication. An indication of frequencies most suitable for transmission over different distances under average conditions is given in Table I. The frequencies listed in Table I are meant to serve as a rough guide only. The actual optimum frequency for any distance varies daily, seasonally, and also with the 11-year sun spot cycle, and the methods of chap. 17 should be used for a more precise determination. In general, for short distances the lower frequencies and high-angle radiation will be required; whereas for long distances higher frequencies and low-angle radiation will be needed.

In the high-frequency band it becomes practical and desirable to use elevated antennas. At these frequencies, ground system losses of ground-based vertical antennas become quite large. The use of elevated antennas gives a certain amount of control over the vertical pattern, which is very desirable. Usually, but not always,

TABLE I
USABLE FREQUENCIES (MEGACYCLES)

Distance (miles)	Summer day	Summer night	Winter day	Winter night
5000 and up	13	8–14	18	5–8
3000	18	6–14	14–18	4–8
2000	16–18	5–12	12–18	4–6
1000	12–16	4–8	10–12	3–6
500	8–12	2–6	6–10	2–5
300	6–8	2–4	5–6	2–4
150	5	2–4	4	2–4
50	2–4	2–4	2–5	2–5

the elevated antenna is horizontal. A horizontal receiving antenna is less responsive to local generated (man-made) noise which is propagated as a predominantly vertically polarized surface wave. In general, an elevated horizontal antenna is more convenient to excite, bearing in mind the desirability of keeping the antenna and feed line at right angles to each other. One of the most common high-frequency antennas is the simple center-fed half-wave dipole.

The Practical Half-wave Dipole. The directional and impedance characteristics of the theoretical infinitely thin dipole having a sinusoidal current distribution have been considered in previous chapters. An actual half-wave dipole has a finite diameter and a current distribution that is not sinusoidal. These differences affect both its directional characteristics and its impedance. However, the difference between the theoretical and actual radiation patterns of the center-fed half-wave dipole is negligible, and for all practical purposes the theoretical pattern can be used. The effect on impedance characteristics is much more pronounced. Reference to the impedance curves of Figs. 13-12, 13-13 and 13-14 shows that for finite-diameter antennas the radiation resistance at a physical half-wavelength is greater than 73 ohms and the reactance is inductive. However, in practice a "half-wave dipole" refers to a *resonant-length* dipole (for which the reactance is zero). For finite-diameter antennas the resonant length is less than a physical half-wavelength by an amount that depends upon the thickness of the antenna, but that is of the order of 5 to 10 per cent. Because the radiation resistance varies rapidly with length in this region, the radiation resistance of a finite-diameter dipole at the resonant length may be of the order of only 65 to 72 ohms—somewhat less than the theoretical 73 ohms.

In addition to the above effects, an actual dipole is always located above the ground or other supporting and reflecting surface, so that the theoretical free-space conditions do not apply. The effect of the presence of a perfectly-conducting ground on the input impedance of the antenna can most readily be accounted for by replacing the ground by an appropriately-located image antenna carrying an equal current in proper phase, and then computing the driving-point impedance under those conditions by the methods of chap. 11. Figure 14-6 shows how the input impedance of a (theoretical) half-wave dipole varies with distance above a ground plane that is assumed perfectly conducting. In general the input impedance now has a reactive component as well as resistance, and the magnitude of the resistance oscillates about the free-space value of 73 ohms. For a practical dipole a similar effect could be expected, with the input resistance oscillating about the actual free-space value. The effect of a *finitely* conducting ground could be determined in a similar manner, the only difference being that the image antenna would carry a current, the magnitude and phase of which would depend upon the actual reflection coefficients of the earth (see chap. 16). The final results would be similar to those of Fig. 14-6, except that the amplitude of oscillation about the free-space impedance would be slightly less, with a slight shift in the actual heights above ground at which the maximum and minimum impedances occurred. Because of irregularities of the ground itself, and unknown reflections from buildings, trees, and other surrounding objects, some deviation from the theoretical values must always be expected.

The effect of the presence of the ground on the vertical radiation pattern can also be obtained by use of the image principle. In Figs. 16-7 through 16-11 are shown the vertical patterns of short vertical and horizontal dipoles above earths of various conductivi-

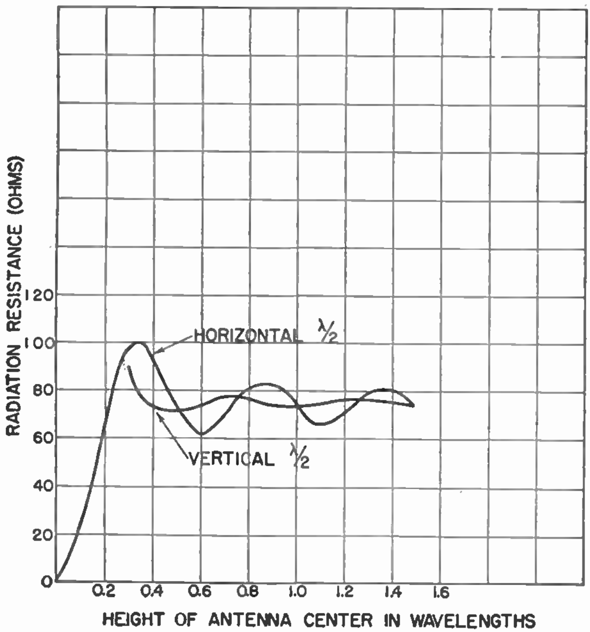


FIG. 14-6. Variation of the radiation resistance of a theoretical half-wave dipole with height above a perfectly conducting earth.

ties. For greater heights above the earth the vertical pattern becomes multilobed. The approximate location of the maxima and minima of the pattern can be obtained by considering the perfect ground case and using the principle of multiplication of patterns as in chap. 12. For horizontal antennas in the plane perpendicular to the axis of the antenna the factor by which the free-space pattern must be multiplied to account for the effect of the ground is

$$2 \sin \left(\frac{2\pi h}{\lambda} \sin \psi \right) \tag{14-11}$$

where h is the height of the center of the antenna above ground and ψ is the angle of elevation above the horizontal. The first maximum in this pattern occurs at an angle ψ_{ml} which is given by

$$\sin \psi_{ml} = \frac{\lambda}{4h} \qquad \left(h > \frac{\lambda}{4}\right)$$

For a vertical antenna the corresponding ground-effect factor for a perfectly conducting earth is

$$2 \cos \left[\frac{2\pi h}{\lambda} \sin \psi\right] \tag{14-12}$$

For finitely conducting earth, expression (11) should give a good approximation to the actual multiplying factor for all angles of ψ because, for horizontal polarization, the reflection factor of an imperfect earth is always nearly equal to minus one (Fig. 16-3). However for vertical antennas, expression (12) will give reasonably good results only for large angles of ψ. As will be shown in chap. 16 (Fig. 16-4), for angles of ψ less than about 15 degrees (the "pseudo-Brewster angle"), the phase of the reflection factor is nearer to 180 degrees than it is to zero, and the use of (12) for low angles of ψ would lead to erroneous results.

In the vertical plane *parallel* to the axis of a horizontal antenna, the radiation is *vertically* polarized, and it is the reflection factor for vertical polarization that must be used in determining the effect of the ground. This means that for large angles of ψ, the reflection factor will be approximately plus one, but for small angles (below about 15 degrees) it will more nearly approximate minus one. This result has an effect of some practical importance in connection with low-angle radiation or reception off the end of a horizontal antenna. Because the vertical components of the direct and reflected waves are oppositely directed as they leave the horizontal antenna, they will have the same direction (or phase) after reflection of the reflected wave by the imperfect ground. Therefore, at low angles, direct and reflected waves will tend to add instead of cancel, resulting in a relatively strong vertically polarized signal off the end of a horizontal antenna.

Methods of Excitation. Several common methods of feeding half-wave antennas are illustrated in Fig. 14-7. Figure 14-7a shows the balanced-line type of center feed. Because of the mismatch between the high characteristic impedance of open-wire lines and the low input resistance of a resonant dipole, this manner of excitation results in standing waves on the feed line as indicated in the figure. However with solid-dielectric, low-impedance lines this mismatch can be almost completely eliminated, but there is now

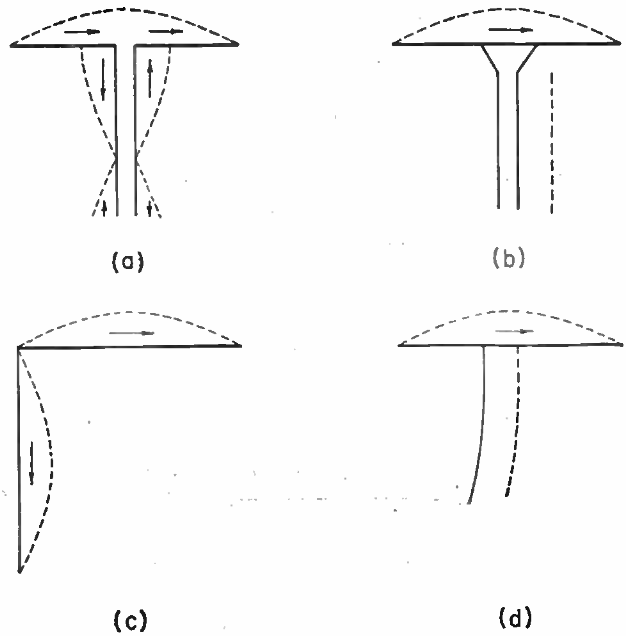


FIG. 14-7. Common methods of exciting high-frequency antennas.

some loss in the dielectric. The "delta-match" or "shunt-feed" arrangement of Fig. 14-7b can result in a good impedance match and low standing waves on the feed-line of the various dimensions are properly chosen.

Probably the simplest of all possible methods of excitation is the single-wire line "end-fed" arrangement of Fig. 14-7c. In this case the vertical "transmission line" also radiates energy, a result that may or may not be desired. By tapping on the vertical wire at a lower impedance point along the horizontal antenna as in d, a better impedance match and lower standing-wave-ratio on the feed line can be obtained.* This results in smaller radiation from the vertical wire, which now carries a traveling-wave current distribution. Optimum dimensions for the types of feed shown in b and d are dependent on the height of the antenna above ground and upon the conductivity of the ground. They may be determined by trial in each case.

Long-wire Antennas. The single-wire feeds of Figs. 14-7c and d are also suited to the excitation of horizontal long-wire antennas. When such antennas are unterminated (that is, open at the far end), the current distribution is chiefly that of a standing wave, and the antenna should preferably be cut to a resonant length, so that the input impedance is resistive. However, with the resonant-line feed of Fig. 14-7c it is usually possible to tune-out a certain amount of reactance at the point of coupling between transmitter and feed line. The patterns of *end-fed* resonant long-wire antennas are multilobed patterns given by the expressions

$$E = \frac{60I}{r}\left[\frac{\cos(\pi L/\lambda \cos\theta)}{\sin\theta}\right] \qquad (14\text{-}13a)$$

for wires that are an odd number of half-waves long, and

$$E = \frac{60I}{r}\left[\frac{\sin(\pi L/\lambda \cos\theta)}{\sin\theta}\right] \qquad (14\text{-}13b)$$

for wires that are an even number of half-waves long. Qualitative patterns for these antennas can be obtained by inspection through use of the principle of multiplication of patterns. The theoretical current distribution and calculated pattern for a two-wavelength end-fed long-wire antenna are shown in Figs. 14-8a and b. Since the actual current distribution consists of a traveling wave as well as a standing wave (because of loss due to radiation), the actual patterns will differ from the theoretical as was pointed out in section 12-03. The chief effect of this difference is to tilt the lobes towards the unfed end. This difference between actual patterns and theoretical patterns (based on standing waves only) is much less in the case of center-fed antennas.

* W. L. Everitt and J. F. Byrne, "Single-wire Transmission Lines for Short-wave Radio Antennas," *Proc. IRE*, **17**, 1840 (1929).

If a long-wire is terminated in a resistance equal to its characteristic impedance the current distribution along it is essentially that of a traveling wave, and its pattern will have the general shape of that shown in Fig. 12-3. The longer the wire, the smaller will be the angle between the wire and the first or main lobe. The important difference between the patterns of terminated and unterminated wire antennas is the absence of large rear lobes in the former.

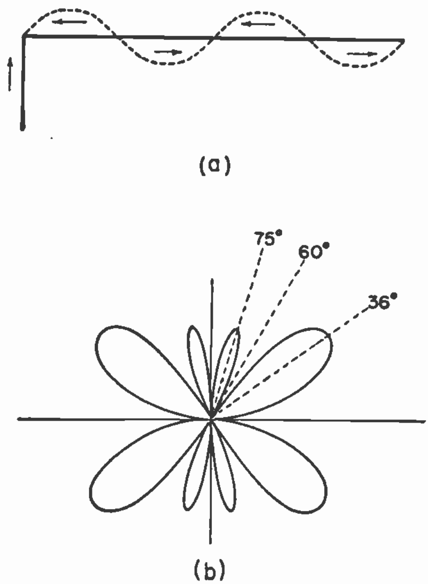


FIG. 14-8. (a) Theoretical current distribution and (b) radiation pattern of a two-wavelength end-fed antenna.

Rhombic Antennas. One of the most useful of the terminated-wire type of antennas is the rhombic antenna (Bruce antenna). This antenna, shown in Figs. 14-9a and 14-9b, consists essentially of a set of four long-wire antennas arranged in such a manner as to have reinforcement of the main beam lobes in the forward direction. Because of the importance of this antenna (rhombics or arrays of rhombics are used on most long-distance commercial circuits) complete design equations have been worked out, and design data may be found in books and handbooks. By terminating one rhombic in a second rhombic, a two-element rhombic array having improved efficiency and greater directivity results.

High-frequency Antenna Arrays. In order to obtain increased directivity for point-to-point communication, commercial companies have built many different kinds of antenna arrays. Besides the broadside "curtain array" these include antennas and arrays having such descriptive names as V, double V (or W), stacked V's, fishbone antennas, and many others. Details on these various types

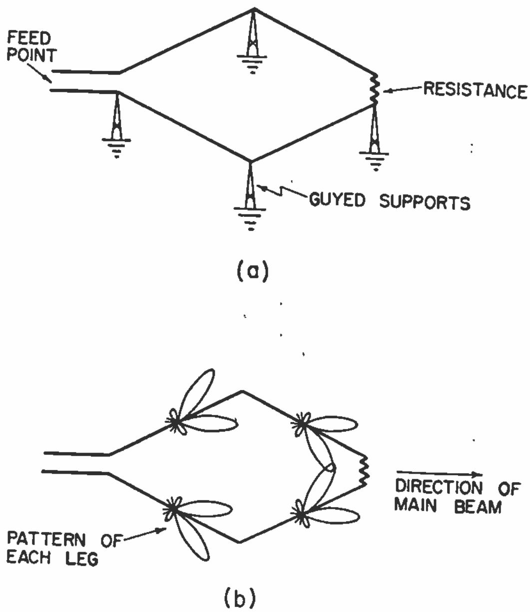


FIG. 14-9. Rhombic antennas.

may be found in the Proceedings of the IRE and other technical journals.

Parasitic Antenna Arrays. If a short-circuited antenna element is placed near an antenna carrying current there is induced in the shorted element a voltage, the magnitude and phase of which depend on the mutual impedance between the elements. The current that flows in the shorted element as a result of this induced voltage depends upon the impedance of the element. By proper control of the phase of this current a certain amount of desirable directivity can be obtained. One, two, or more such "parasitically excited" elements are often used with a driven element to form a *parasitic* *array*. The close spacing required between elements results in a compact array which can be mounted on a light, rotatable frame. The steerable directivity of this *rotary-beam* antenna, as it is called, has led to its extensive use at the upper end of the high-frequency band, and also in the very high frequency band for FM and television reception.

The magnitude and phase of the current in a parasitic element can be varied by adjusting its length and its spacing from the driven element. Fairly close spacings, between 0.1 λ and 0.25 λ are required in order to obtain a sufficiently large current, but the actual spacing is not too critical as long as the length is correctly adjusted for each spacing. When the phasing of the current in a parasitic element is such that the main lobe is on the same side of the driven element as the parasitic, the parasitic element is called a *director*. When the main lobe is on the opposite side, the term *reflector* is used. For a single director element used with a driven element, the optimum spacing* is about 0.1 λ. For a reflector alone with a driven element the optimum spacing is about 0.15 λ. When both director and reflector are used together, these spacings are not necessarily still optimum, but they are often used. When the spacings have been selected, the mutual impedances between elements can be obtained from curves such as those of Figs. 11-6 through 11-10. Then for any selected lengths (and hence impedances) of parasitic elements, the currents which will result can be calculated by solving the mesh equations.†

$$V_1 = I_1Z_{11} + I_2Z_{12} + I_3Z_{13}$$
$$0 = I_1Z_{21} + I_2Z_{22} + I_3Z_{23}$$
$$0 = I_1Z_{31} + I_2Z_{32} + I_3Z_{33}$$

A sample computation will be carried through to indicate the various factors involved.

EXAMPLE 2: Design a horizontal three-element parasitic array for 29 mc. The reflector will be spaced 0.15 λ and the director 0.10 λ from the driven

* G. H. Brown, "Directional Antennas," *Proc. IRE*, 25, 1, 78–145. (1937).

† For the solution of these equations the values of the mesh impedances Z_{11}, Z_{22}, and Z_{33} are required. In general these are not known, and it is customary to use in their stead the self-impedances, Z_{s1}, Z_{s2}, and Z_{s3} as explained in section 11-03. Although less justifiable for these close-spaced arrays, the same procedure will be followed here. It is expected that the results obtained will be at least of the correct order of magnitude.

element. The diameter of the elements will be 1½ inches. Investigation by a cut-and-try method has shown that a good front-to-back ratio is obtained when the reactances of the director (antenna 2) and reflector (antenna 3) are chosen to be

$$X_{22} = -j30 \qquad X_{33} = +j65$$

Then

$$f = 29 \text{ mc} \qquad \frac{r_0}{\lambda} = \frac{0.750}{34 \times 12} = 0.00184$$

$$\lambda = 34 \text{ ft}$$

For this problem the curves of Fig. 11-12 prove most convenient. From them it is found that

$$X_{22} = -j30, \quad R_{22} = 55.6, \quad Z_{22} = 63.2\underline{/-28.4^\circ}, \quad L_2 = 0.455\,\lambda$$

$$X_{33} = +j65, \quad R_{33} = 80, \quad Z_{33} = 103\underline{/39.1^\circ}, \quad L_3 = 0.515\,\lambda$$

Solving the mesh equations gives

$$\frac{I_2}{I_1} = \frac{Z_{31}Z_{23} - Z_{21}Z_{33}}{Z_{22}Z_{33} - (Z_{23})^2}$$

$$\frac{I_3}{I_1} = \frac{Z_{21}Z_{23} - Z_{31}Z_{22}}{Z_{22}Z_{33} - (Z_{23})^2}$$

From mutual-impedance curves the mutual impedances involved are found to be

$$Z_{31} = 60.8\underline{/-6.7^\circ}$$

$$Z_{21} = 67.9\underline{/+6.4^\circ}$$

$$Z_{23} = 49.7\underline{/-34.8^\circ}$$

Using these values,

$$\frac{I_2}{I_1} = 1.135\underline{/-143.1^\circ} \qquad \frac{I_3}{I_1} = 0.0834\underline{/67.1^\circ}$$

The horizontal pattern of the array will be given (Fig. 14-10) by

$$E = k\,[I_1 + I_2\underline{/\beta d_2 \cos\phi} + I_3\underline{/-\beta d_3 \cos\phi}]\left[\frac{\cos\left(\frac{\pi}{2}\sin\phi\right)}{\cos\phi}\right]$$

$$E = kI_1\left[1 + \left|\frac{I_2}{I_1}\right|\underline{/\alpha_2 + \beta d_2\cos\phi} + \left|\frac{I_3}{I_1}\right|\underline{/\alpha_3 - \beta d_3\cos\phi}\right]\left[\frac{\cos\left(\frac{\pi}{2}\sin\phi\right)}{\cos\phi}\right]$$

Inserting the values for I_2/I_1 and I_3/I_1, it is found that:

$$\frac{E_{0^\circ}}{E_{180^\circ}} = \frac{1.30}{0.2} = 6.5$$

which is the front to back ratio. Solving for V_1/I_1 in the mesh equations, the input impedance of the array is found to be

$$Z_1 = Z_{11} + \frac{I_2}{I_1} Z_{12} + \frac{I_3}{I_1} Z_{13}$$
$$Z_1 = Z_{11} - 53.5 - j48.4$$
$$R_1 = R_{11} - 53.5$$
$$X_1 = X_{11} - 48.4$$

Since the array should present a resistive load to the transmission line, we assign a value of $X_{11} = +j48.4$

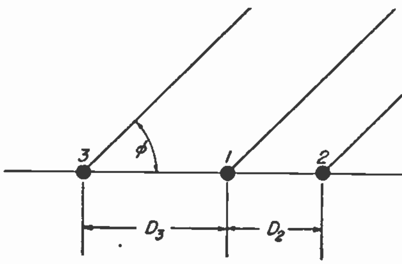


FIG. 14-10

From the curves of figure 11-12

$$X_{11} = +j48.4 \qquad R_{11} = 75 \qquad L_1 = 0.505\,\lambda$$

Then the input impedance will be

$$Z_1 = R_1 = 75 - 53.5 = 21.5 \qquad \text{ohms}$$

The low input resistance is typical of these close-spaced arrays. It can be increased to match commercially available transmission line by using a folded dipole for the driven element. A problem on this is given at the end of section 14-04.

Sec. 14.04 Very High Frequency Antennas. In the v-h-f band (30–300 mc) a half-wave antenna is a convenient physical size, and transmission line matching sections are easy to construct, so that optimum design of an antenna system becomes easier to achieve. Besides aircraft and point-to-point communication applications, this range includes frequency modulation and television. For the latter services band width becomes a consideration, so that wide-band matching circuits are of importance.

Stub-matched and Folded Dipoles. A simple dipole has an input impedance that is too low for direct connection to an ordinary open-wire transmission line, and some sort of impedance matching arrangement is required if the desirable condition of no standing waves on the line is to be attained. An easy way of obtaining this match is by means of the *stub-matched* dipole shown in Fig. 14-11. By making the length $L = 2H$ somewhat less than a half-wavelength, the input impedance will be a capacitive reactance in series with the radiation resistance. For a length s of stub line less than $\lambda/4$, the input impedance of the shorted transmission line will be

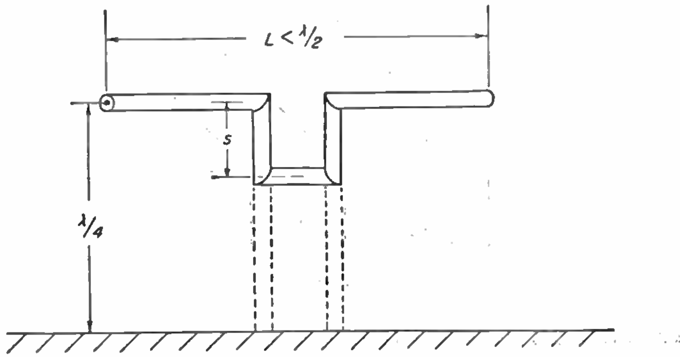


FIG. 14-11. Stub-matched dipole.

an inductive reactance, the magnitude of which can be adjusted to tune out the capacitance of the antenna. The resulting impedance at the terminals *a-b* will be a pure resistance R_{ar}, the resistance of the parallel-resonant circuit. By proper choice of L and s this resistance can be adjusted to almost any value desired. The arrangement is good mechanically, because by extending the lines of the stub back beyond the shorting bar, the antenna can be mounted on, and a quarter wavelength in front of, a reflecting ground screen, without the use of insulators.

An alternative way of obtaining a high-impedance input is by means of the *folded dipole* described by Carter* and shown in Fig. 14-12. This method has the added advantage that it also increases

* P. S. Carter, "Simple Television Antennas," *RCA Rev.*, **4**, 168, October 1939; W. Van B. Roberts, "Input Impedance of a Folded Dipole," *RCA Rev.*, **8**, 289 (1947).

the bandwidth of the antenna, an important consideration in FM and television applications. The folded half-wave dipole consists essentially of two half-wave radiators very close to each other and connected together at top and bottom. As far as the antenna currents or radiating currents are concerned, the two elements are in parallel, and if their diameters are the same, the currents in the elements will be equal and in the same direction. If 1 amp flows in each element (at the center) the total effective current is 2 amp, and the power radiated will be $(2I_1)^2R_{rad} \approx 4 \times 73I_1^2$ or 4 times that radiated by a single element carrying 1 ampere. However, the

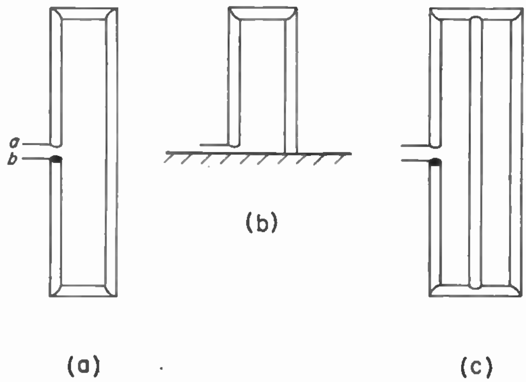


FIG. 14-12. (a) Folded dipole. (b) Folded monopole. (c) Multi-element folded dipole.

current that is required to be delivered by the generator at the terminals *a-b* is only 1 ampere, so that the input resistance is seen to be 4 times that of a simple dipole. If there are three elements of equal diameters connected together as in Fig. 14-11c, the input resistance will approximately be 9 times that of a simple dipole. If the elements are of unequal diameters, the currents will divide unequally between the elements. If it is assumed that the currents divide inversely as the characteristic impedances Z_{01} and Z_{02}, so that

$$\frac{I_2}{I_1} = \frac{Z_{01}}{Z_{02}}$$

then if element 1 is the driven element, the input resistance will be

$$R_{in} = \frac{(I_1 + I_2)^2R_{rad}}{I_1^2} = R_{rad}\left(1 + \frac{I_2}{I_1}\right)^2 = R_{rad}\left(1 + \frac{Z_{01}}{Z_{02}}\right)^2 \quad (14\text{-}14)$$

As before R_{rad} is the radiation resistance of a simple half-wave dipole (theoretically $R_{rad} = 73$ ohms, actually $R_{rad} < 73$ ohms).

In order to understand the increased band width that results with the folded dipole arrangement, consider the simple half-wave (resonant-length) dipole of Fig. 14-13*a* which is connected in parallel at its terminals with a shorted quarter-wavelength line. At the resonant frequency the dipole resistance is in parallel with the input impedance of the transmission line, which is a resistance of very high value. Below resonance, the antenna impedance becomes capacitive, but the transmission line impedance becomes inductive,

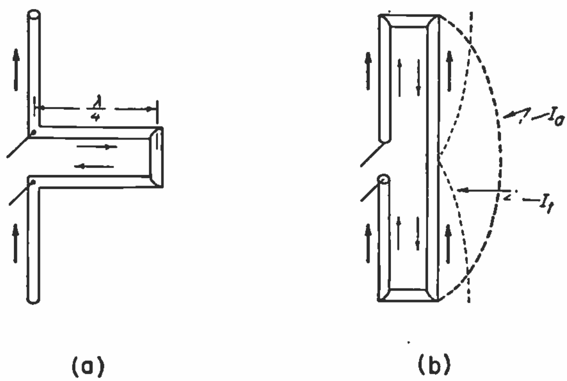


FIG. 14-13

and the parallel combination tends to remain nearer unity power factor than does the antenna alone. Conversely, above resonance the antenna impedance becomes inductive and the line impedance becomes capacitive so that compensation is again obtained. Although compensation is far from perfect, because the susceptances are not equal and opposite, if the frequency is shifted far enough in either direction from the resonant frequency of the dipole, a point of perfect susceptance compensation (where the input impedance is a pure resistance) is again obtained. Below resonance this occurs for the same conditions that led to the stub-matched dipole of Fig. 14-11. Above resonance the point of perfect susceptance compensation occurs when the capacitive susceptance of the stub is just sufficient to tune out the inductive susceptance of the antenna. Of course, the input resistance at these stub-matched or anti-resonant points will be considerably higher than at the resonant frequency, but for some purposes the resulting standing-wave ratio is good enough over the range that the effective band width may be said to extend from one anti-resonant point to the other. This represents almost a two to one frequency range.

The above considerations apply directly to the folded dipole of Fig. 14-13, which has the stub line (actually two stub lines in series) as a built-in feature. The elements of the folded dipole then carry both the antenna currents, which are in the same direction in the two elements, and the transmission line currents, which are in opposite directions in the two elements of the dipole (Fig. 14-13b). At the resonant frequency the antenna currents are relatively large $[I_a = (V/R_{in})(V/4R_{rad})$ in each element, at the center], whereas the transmission line currents are zero at the center, but have the value $I_t = V/2Z_{ot}$ at the ends, where Z_{ot} is the characteristic impedance of each of the two shorted quarter-wave transmission line sections.

For FM broadcast reception a common type of antenna is a folded dipole made of flexible solid dielectric "twin-lead." For a transmission line made of such cable the phase velocity, and hence the length of a wavelength, is only about 80 per cent of the free-space values. Therefore an electrical quarter-wavelength section is only 0.8 times $\lambda/4$ physically, and the physical line must be made shorter than would be the case with a free-space dielectric. On the other hand, the thin dielectric covering on the cable has almost negligible effect* on the apparent phase velocity and wavelength of the antenna currents, so that for resonance the physical length of the antenna should still be approximately $L \approx 0.95\lambda/2$ (that is $H \approx 0.95\lambda/4$). The method for satisfying these two conditions simultaneously is indicated in Fig. 14-14. The two elements are

* The reason for the difference in the two cases is as follows: As a transmission line the return displacement currents flow from one wire to the adjacent wire, mostly through the solid dielectric that is in parallel with the surrounding air dielectric. As an antenna the "return" displacement currents flow from both wires in parallel, through the solid dielectric and air dielectric in series to the opposite arm of the dipole (or to ground in the case of the monopole). In this latter case the length of path through the solid dielectric is short compared with the remainder of the path through the air dielectric so that the solid dielectric has negligible effect.

cut to 0.95 times λ/2, but the shorting connections are spaced only 0.8 times λ/2 apart.

An added insight into the operation of the folded dipole (and also certain of the "baluns" described in the next part of this section) is provided by the superposition principle. Consider the operation of the folded monopole of Fig. 14-12b, which operation is identical with that of the corresponding folded dipole of Fig. 14-12a. The single zero-impedance generator may be replaced by three equivalent generators having equal voltages, zero internal impedances, and connections and polarity as shown in Fig. 14-15b. If $V_1 = V_2 = V_3 = V/2$, then a quick check shows that as far as the currents in the two elements are concerned, the operation of 14-15b is identical with that of 14-15a. In b generator V_3 causes equal antenna currents to flow in the same direction in elements 1 and 2. Generators V_1 and V_2 in series cause equal and opposite transmission line currents to flow in elements 1 and 2. Because

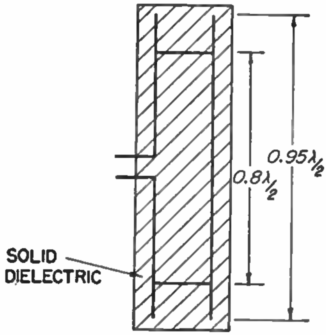


FIG. 14-14. Common type of "built-in" antenna for FM reception.

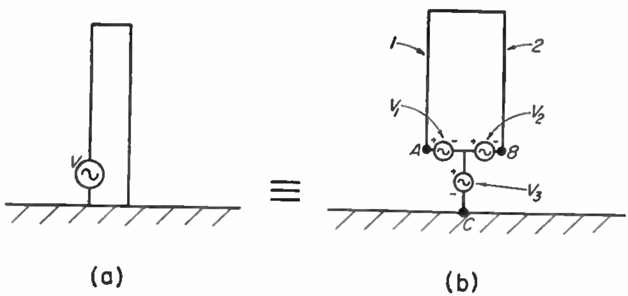


FIG. 14-15

points B and C have the same voltage at all times they could be joined together without affecting the operation (when all generators are generating) and the circuit of Fig. 14-15a would result. Now consider the superposition principle applied to the equivalent circuit b. By this principle the total currents flowing in any branch with all generators generating is just the sum of the individual currents produced by each of the generators alone, the other generators being replaced by their internal impedances. With V_1 and V_2 not generating, V_3 sends equal antenna currents into the two elements 1 and 2 (assuming these elements to have equal diameters). With V_3 not generating, V_1 and V_2 send equal and opposite transmission line currents into the two elements. The total currents that actually flow are the sum of the two sets of currents. The impedance relations given previously follow directly. At resonance the transmission line currents at the input produced by V_1 and V_2 are approximately zero. The total antenna current is V_3/R_{rad}, where R_{rad} for a $\lambda/4$ monopole is approximately 36.5 ohms. The antenna current in element 1 is one-half the total antenna current, or $I_{a1} = V_3/2R_{\text{rad}}$. The actual applied voltage $V = 2V_3$, so the input impedance at resonance is $R_{\text{in}} = V/I_{a1} = 4R_{\text{rad}}$.

Baluns. An ordinary dipole is a *balanced* load in the sense that for equal currents in the two arms, the arms should have the same impedance to ground. Such a load should be fed by a transmission line such as a two-wire line, which itself is "balanced to ground." However at very high and ultrahigh frequencies unbalanced coaxial lines are nearly always used, so the problem is encountered of transforming from an unbalanced to a balanced system or vice versa. The device that accomplishes this balance-to-unbalance transformation is called a *balun.* There are many different types baluns and four of the most common are shown in Figs. 14-16 and 14-17. In Fig. 14-16a, a balanced dipole antenna is shown connected directly to the end of an unbalanced (coaxial) line. The currents I_1 and I_2 must be equal and opposite. At the junction A, current I_2 divides into I_3, which flows down the outside of the outer conductor of the line and $I_2 - I_3$ which flows on the second arm of the dipole. The current I_3 depends upon the effective "impedance to ground" Z_g, provided by this path along the outside of the conductor. This impedance can be made very high, thus making I_3 very small, by the addition of a quarter-wave skirt around the outer conductor as in Fig. 14-16b. With the skirt shorted to the conductor at the bottom, the impedance between the points A and B, and therefore between A and the effective ground part wherever it may be, is extremely large, being limited only by the Q of the shorted quarter-wave section. In the arrangement of 14-16c the dipole feed is balanced by making the impedance of the shunt paths from A and B to the common point C equal. When the stub length is approximately a quarter of a wavelength, these equal shorting impedances are very large, and, in addition, the arrange-

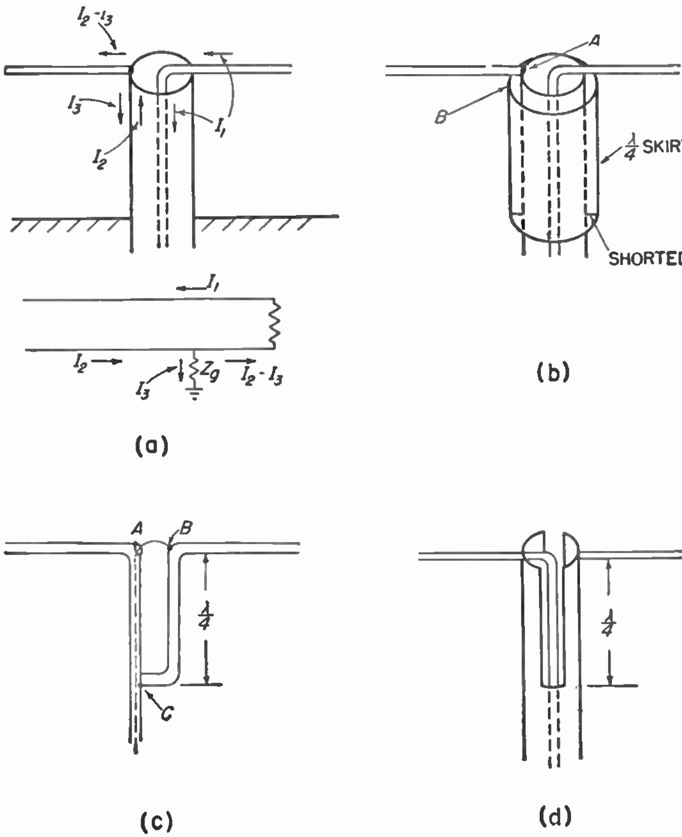


Fig. 14-16. Baluns.

ment exhibits the desirable broad-band characteristics discussed in connection with Fig. 14-13a, to which it is exactly equivalent. Figure 14-16d is a more practical version of the arrangement shown in c. Similar broad-band characteristics are obtained with the balun of Fig. 14-17a for which the equivalent circuit of Fig. 14-17b may be drawn. The impedances Z_g, which shunt each side of the balanced load, are given by

$$Z_g = jZ_0 \tan \beta s,$$

where Z_0 is the characteristic impedance of each of the parallel stubs.

Frequency Modulation and Television Antennas. Folded dipoles, either alone or with parasitic elements, are frequently-used antennas for frequency modulation and television reception. The parasitic elements give an additional directivity, which is often very desirable, but they also increase the frequency sensitivity of the antenna. This comes about because the presence of the parasitic elements tends to decrease the actual band width, and, in addition, the radiation pattern changes rapidly with frequency so that the useful frequency range is further restricted. The decreased band width is rather important in television because of the wide-band requirements of this service. These requirements are of the order of 10 per cent for single-channel reception. If a single antenna is to be used for all channels, an antenna capable of covering approximately a four to one frequency range is required. For television transmitting antennas, band width requirements are very stringent indeed, because an almost perfect match is required over the channel band width (6 mc) if "ghosts" owing to multiple reflections on the line between transmitter and antenna are to be avoided.

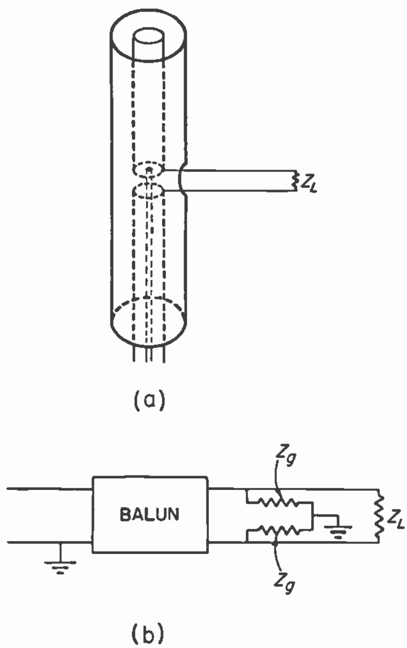


FIG. 14-17. Balun and "equivalent circuit."

In FM transmitting, the antenna problem is chiefly that of obtaining a circular pattern with horizontal (or circular) polariza-

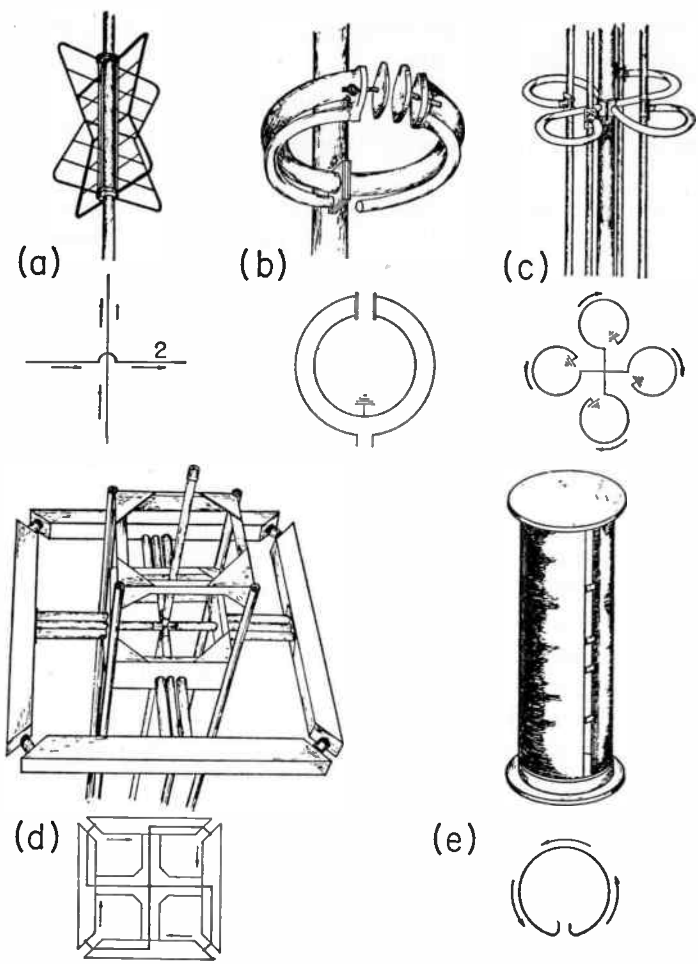


FIG. 14-18. Some commercial models of FM and television antennas: (a) *RCA* superturnstile or batwing; (b) *GE* circular; (c) *WE* clover-leaf; (d) *Federal* square loop; (e) *RCA* pylon.

tion. Power gain is obtained from directivity in the vertical plane. Figure 14-18 illustrates five commercial types of FM and television transmitting antennas that achieve these ends by different means.

In Fig. 14-18a is sketched the superturnstile or batwing type of antenna, which obtains its almost circular horizontal pattern by having the two sets of wings fed in phase quadrature. The ordinary turnstile arrangement consists of two crossed horizontal half-wave dipoles fed in quadrature. By replacing the dipoles with the wing-like structure that is fed in phase along its length, wide band width as well as some desirable vertical directivity is obtained. This antenna is suitable for television transmission. In (b) a circular antenna has been formed by rolling a folded dipole almost into a circle, and then applying capacitive loading between the ends to improve the uniformity of current around the loop. The clover-leaf antenna of (c) consists of four small loops fed in phase to give the effect of radiation from a single larger loop. One end of each of the four loops is connected to the inner conductor of a transmission line and the other end of each loop is connected to one of the four corners of a square lattice-work tower that supports the whole structure and that also forms the outer conductor of the transmission line. The square loop antenna of (d) is an adaptation for square towers of the circular V-H-F loop, which is in common use at ultra-high frequencies. This antenna makes use of the principle of the shielded loop in which the radiating element is also the outer conductor of the coaxial transmission line that feeds the loop. In addition, the coaxial feed line sections are used to obtain a suitable impedance match. The pylon or slotted-cylinder antenna shown in (e) consists of a longitudinal slot cut in a cylinder that is of the order of one wavelength long. The slot is shorted at both ends and fed at the middle by means of a coaxial transmission line that runs up the inside of the cylinder. Methods of computing the radiation patterns of slotted cylinder antennas are given in chap. 15.

All of the antennas shown in Fig. 14-18 produce a nearly circular horizontal pattern with horizontal polarization. The desired vertical directivity is obtained by stacking vertically several of the individual units.

V-H-F Antenna Arrays. For point-to-point communication and applications such as radar the gain or directivity that can be achieved by the use of arrays is usually desirable and sometimes necessary. At these frequencies, line arrays and rectangular arrays become practical. A common array is the "mattress" antenna (Fig. 14-19) which consists of a rectangular array of coplanar elements, mounted a quarter wavelength in front of a reflecting screen. The patterns of such arrays are easily obtainable by the methods of chap. 12, but methods of feeding the arrays to obtain the desired currents in the elements have not yet been considered. Figures 14-20 and 14-21 illustrate two methods for feeding a number of elements with equal currents, or with any specified currents, as

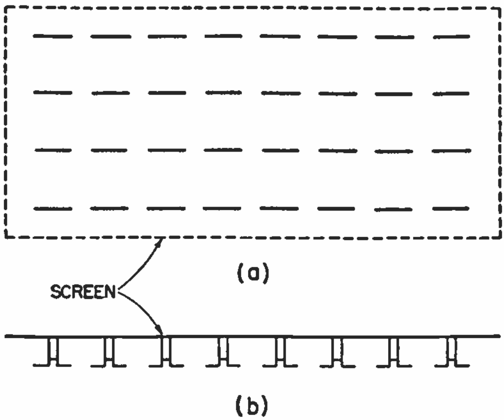


FIG. 14-19. Typical rectangular array of co-polar elements.

required, for example in the binomial array of section 12. In Fig. 14-20 the points *A*, *B*, *C*, and *D* are spaced half a wavelength apart on the transmission line, so that the voltages at these points are always equal in amplitude. By feeding the antennas from these points through quarter-wave sections, the current amplitudes

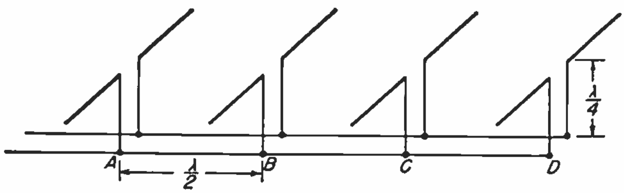


FIG. 14-20. A method of feeding antennas with specified currents.

depend only upon these equal voltages and the characteristic impedances of the sections, and will be *independent* of the antenna driving-point impedances. For similar quarter-wave sections the antenna currents will be equal, although they can be made to have almost any ratio by suitable choice of characteristic impedances for the quarter-wave sections. In the arrangement of Fig. 14-21, it is easy to show that the amplitudes of the driving voltages at the antenna terminals will be related by

$$\frac{V_2}{V_1} = \frac{Z_{02}}{Z_{01}} \qquad \frac{V_3}{V_2} = \frac{Z_{03}}{Z_{01}} \tag{14-15}$$

In practice the quarter-wave sections are made by slipping copper tubing of the correct diameter over the main feed line Z_{01}, and soldering the tubing to the inner line at both ends.

V-H-F and S-H-F Antennas. The antennas and arrays used in the v-h-f band, for the most part can be and are used in the u-h-f range also. However, at the upper end of the ultrahigh-frequency band, and especially at superhigh frequencies (3000–30,000 mc), the

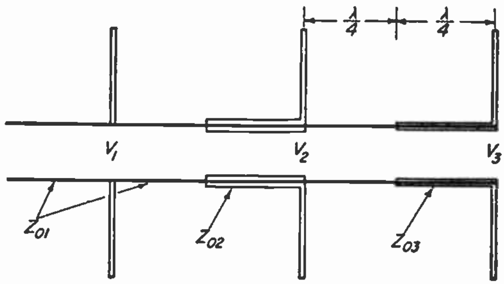


FIG. 14-21. A second method of feeding the elements of an array with specified currents.

size of the elements becomes impractically small. It is then convenient to use "current-sheet" radiators such as parabaloids, horns, and slot antennas. The radiation from such sheet radiators can be computed from the fields of the individual current elements, exactly as is done for ordinary linear radiators, providing that the current distribution on the conducting sheets is known or can be estimated. However in most cases, the current distribution is neither known nor readily estimated, so that other methods of determining the radiation must be sought. A quite powerful method consists of determining the radiation from the antenna in terms of the fields that exist across the "aperture" of the antenna. Radiation from aperture antennas is the subject of chap. 15.

Problem 1. Verify the relation given in eq. (15).

Problem 2. Using eq. (14), determine the wire size required for the excited arm of a folded dipole, in order to transform the input resistance of 21.5 ohms, obtained in example 1, to 150 ohms. The second arm of the dipole should be of the same diameter as the other elements (1½ in.), and a spacing between arms of 3 in., center-to-center, is suggested.

14.05 Receiving Antennas. In earlier chapters no special consideration has been given to receiving antennas because use of the reciprocity theorem shows that the important characteristics of an antenna, viz., impedance and pattern, are the same for receiving as for transmitting. Nevertheless, there are differences between other properties of receiving antenna and transmitting antennas, and so some mention will be made in this section of those characteristics of an antenna that are peculiar to its use as a receiving antenna. One of the important characteristics that is different for reception

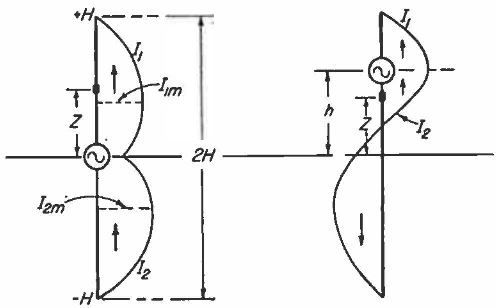


Fig. 14-22

than for transmission is the current distribution on the antenna. It has been seen that for thin transmitting antennas the current distribution may be represented approximately by a sinusoidal distribution, the approximation improving as the antenna is made thinner. On the other hand, for reception, the current distribution is a function of the length of the antenna, the direction of arrival of the received wave, and the load impedance, and except for the special case of the resonant-length antenna ($H \approx \lambda/4$), the current distribution is not even approximately sinusoidal even for very thin antennas. Once again, however, it is possible to invoke the reciprocity theorem, and by this means determine the approximate current distribution on a receiving antenna from the known approximately sinusoidal distribution that holds for the transmitting case. The application of the theorem to obtain this result is made as follows: Figure 14-22a shows a thin center-fed transmitting dipole of length $2H$ that has the symmetrical sinusoidal current distribution represented by

$$I_1 = I_{1m} \sin \beta(H - z) \qquad (0 < z < H) \qquad (14\text{-}16)$$

$$I_2 = I_{2m} \sin \beta(H + z) \qquad (-H < z < 0) \qquad (14\text{-}17)$$

In Fig. 14-22b, with the antenna fed off-center at $z = h$ (by an isolated generator that has no connection to ground), the sinusoidal current distribution is still given by eqs. (16) and (17), except that for I_1 the range of z is from h to H, and for I_2 it is from h to $-H$. At the generator ($z = h$), the currents I_1 and I_2 must be equal to each other and to the generator current I_b, and so the following relation between the maximum or loop values of currents is obtained:

$$I_b = I_{1m} \sin \beta(H - h) = I_{2m} \sin \beta(H + h) \qquad (14\text{-}18)$$

In either a or b the current at the feed point will be

$$I_b = \frac{V}{Z_a}$$

where Z_a is the antenna impedance. For thin antennas and for lengths not too near the resonant or antiresonant lengths ($H = \lambda/4$ and $H = \lambda/2$), the input impedance for the symmetrical, center-fed antenna may be represented approximately as a reactance, the value of which is given by

$$Z_a \approx -jZ_0 \cot \beta H \qquad (14\text{-}19)$$

where

$$Z_0 = 120\left[\ln \frac{2H}{a} - 1\right]$$

At the resonant length where eq. (19) indicates zero impedance, the actual impedance is a small resistance. At the antiresonant length, where (19) yields an infinite reactance, the actual impedance is a resistance of very high value.

In the unsymmetrical case of Fig. 14-22b, the impedance is not known, but a very rough approximation to it can be obtained by considering it made up of the sum of the "impedances" of the two half-sections of the antenna, one of length $(H - h)$ and the other of length $(H + h)$. That is, it will be assumed that

$$Z_a = Z_{a1} + Z_{a2}$$

where

$$Z_{a1} = -\tfrac{1}{2}jZ_{01} \cot \beta(H - h)$$

$$Z_{a2} = -\tfrac{1}{2}jZ_{02} \cot \beta(H + h)$$

Z_{01} and Z_{02} are the characteristic impedances of the two sections. In general Z_{01} and Z_{02} will be different, the first getting larger as the second gets smaller. For present purposes, it will be further assumed that

$$Z_{01} \approx Z_{02} \approx Z_0 = 120\left(\ln \frac{2H}{a} - 1\right)$$

Then

$$Z_a \approx -\tfrac{1}{2}jZ_0 [\cot \beta(H - h) + \cot \beta(H + h)] \qquad (14\text{-}20)$$

It is now possible to apply the reciprocity theorem. For a given voltage V, applied at the feed-point, $z = h$, in Fig. 14-22b, the current at any point z along the antenna will be [from eqs. (16) and (17)]

$$I_{z1} = \frac{I_b \sin \beta(H - z)}{\sin \beta(H - h)} = \frac{V \sin \beta(H - z)}{Z_a \sin \beta(H - h)} \qquad (h < z < H)$$

or $$I_{z2} = \frac{V \sin \beta(H + z)}{Z_a \sin \beta(H + h)} \qquad (-H < z < h)$$

Then by the reciprocity theorem, when the antenna is used as a receiving antenna, a voltage $E_z\, dz$ induced in an elemental length dz at z will produce a short-circuit current at $z = h$, which is given by

$$dI'_{sc} = \frac{E_z\, dz\, I_{z1}}{V} \qquad (\text{for } h < z < H)$$

or $$dI''_{sc} = \frac{E_z\, dz\, I_{z2}}{V} \qquad (\text{for } -H < z < h)$$

For a uniform plane wave incident normally ($\theta = 90$ degrees) on the receiving antenna, the induced voltage $E_z\, dz$ will be constant (same magnitude and phase) along the length of the antenna, so the total short-circuit current, I_h, at the point $z = h$ will be

$$\begin{aligned} I_h &= \frac{E_z}{V} \int_{z=h}^{z=H} I_{z1}\, dz + \frac{E_z}{V} \int_{z=-H}^{z=h} I_{z2}\, dz \\ &= \frac{E_z}{Z_a} \int_h^H \frac{\sin \beta(H - z)}{\sin \beta(H - h)}\, dz + \int_{-H}^{h} \frac{\sin \beta(H + z)}{\sin \beta(H + h)}\, dz \\ &= \frac{E_z}{\beta Z_a}\left[\frac{1 - \cos \beta(H - h)}{\sin \beta(H - h)} + \frac{1 - \cos \beta(H + h)}{\sin \beta(H + h)}\right] \qquad (14\text{-}21) \end{aligned}$$

Inserting the expression (20) for Z_a and performing the necessary trigonometric manipulations, (21) can be reduced to

$$I_h = -\frac{jE_z\lambda}{\pi Z_0}\left(\frac{\cos\beta H - \cos\beta h}{\cos\beta H}\right) \tag{14-22}$$

By choosing h at different points along the antenna the current distribution on a receiving antenna in a uniform field parallel to the antenna is obtained.

Expression (22) fails for H equal to an odd multiple of a quarter-wavelength. For these resonant lengths, expression (22) indicates that the current amplitudes would be infinite. The reason for this result is that the radiation resistance was neglected in writing the expression for the antenna impedance. Although such an approximation is permissible for nonresonant lengths where the antenna reactance may be much larger than the resistance, at resonance the input reactance is zero and the current is limited only by the radiation resistance of the antenna. Putting Z_a equal to the input resistance, that is $Z_a = R_{\text{loop}}/\sin^2\beta(H - h)$, for the resonant condition, the reciprocity theorem can be used to compute the receiving current distribution as before. For a resonant-length half-wave receiving dipole, the current distribution is found to be sinusoidal and, to a first approximation, the current amplitude is independent of the thickness of the antenna.

In Fig. 14-23, the current distributions obtained by this method have been plotted for dipole lengths $L = 2H = \lambda/4$, $3\lambda/4$, $5\lambda/4$, and for the resonant length $L = \lambda/2$, for antenna radii of 1 cm and a frequency of 100 mc.

Although the reciprocity theorem itself is exact, it was necessary in the above analysis to make some questionable assumptions regarding antenna impedances in order to obtain quantitative answers. In view of these assumptions the method would be expected to yield answers that are rough approximations, at best. It is interesting to find that expression (22) for current distribution on a receiving antenna agrees exactly with results that can be obtained by other, apparently more rigorous, methods.

Other considerations. Certain other differences between receiving and transmitting antennas result from the different requirements for the two conditions of operation. Although the directional patterns of an antenna are the same for transmitting and receiving, the optimum pattern for transmitting may not be optimum for receiving. For transmitting, the optimum pattern is often that which puts most signal into a given direction. For reception, however, the optimum condition is not maximum received power, but rather maximum signal-to-noise ratio. Although the pattern that gives the first condition may also lead to the second, such is not

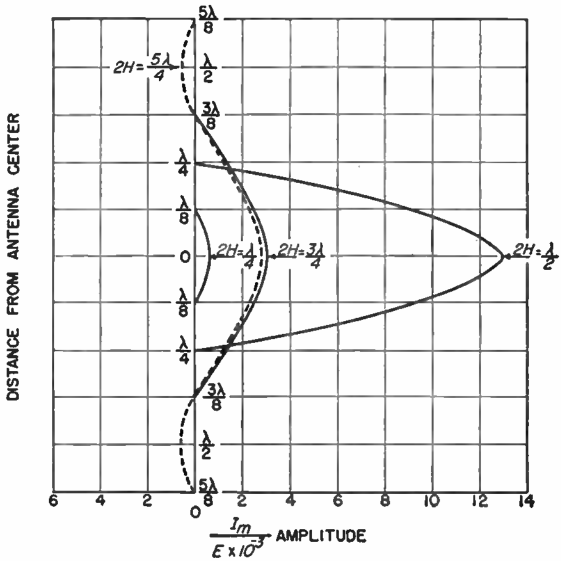


FIG. 14-23. Current distributions in a short-circuited receiving antenna where the received wave is incident normally to the antenna axis. The curves show current magnitudes on a one centimeter diameter rod at 100 mc.

necessarily the case. For example, a minor lobe in the pattern of a receiving antenna may bring in a large amount of noise if it happens to be pointed towards the noise source, and so result in a low signal-to-noise ratio. On the other hand, as a transmitting antenna, the presence of the lobe may have no ill effect, other than the loss of the small amount of power that it radiates. Increasing the directivity of a transmitting antenna will always increase the signal-to-noise ratio at the receiver (assuming the receiver to be in the correct direction). Increasing the directivity of a receiving antenna may, or may not, improve the signal-to-noise ratio. If the noise is coming equally from all directions, the improvement in signal-to-noise ratio will be exactly the same as was obtained in the transmitting case. If less noise is coming from the direction of the received signal than other directions, then the improvement can be greater than in the transmitting case. On the other hand, if all the noise is coming from the same direction as the signal, there will be no improvement in signal-to-noise ratio obtained by making the receiving antenna directive. (These conclusions are valid only if the set noise is a negligible part of the total noise.) Similar differences between transmission and reception occur in coupling the antenna to transmitter or receiver. In the transmitting case, the coupling is adjusted to feed the antenna with a specified amount of power (and not for maximum power transfer, which condition would usually overload the transmitting tubes). In receiving, the best adjustment is that which yields maximum signal-to-noise ratio. Under those conditions where maximum signal-to-noise ratio is obtained simultaneously with maximum received signal, the correct adjustment would be that which gives maximum power transfer. However, there are many cases where the optimum coupling is not that which gives maximum power transfer. In medium-frequency broadcast reception, where receivers ordinarily have ample gain, and where receiver set noise is negligible compared with atmospheric noise, there is no advantage to be gained in increasing the amount of power delivered by the antenna to the receiver beyond that required to deliver sufficient audio power output, because the noise power is increased in the same ratio as the signal power. Hence the coupling used is usually very loose, so that the receiver will "track" properly over the entire band, regardless of the length of the antenna that may be used.

Again at frequencies above about 30 mc, atmospheric noise is negligible and set noise may be the limiting factor in determining signal-to-noise ratio. Under these conditions the adjustment should be that which delivers maximum signal to the set, as long as this adjustment does not increase the set noise more than it does the signal. Set noise occurs mostly in the first stage and is dependent upon the effective resistance coupled into the grid-circuit of the first tube. Hence it depends upon the coupling to the antenna impedance and increases with this coupling. The coupling that yields maximum signal-to-noise ratio is often just slightly less than that which produces maximum power transfer.

Another difference between transmission and reception appears when the operation of an antenna of given electrical length (say a half-wave dipole) is considered as a function of frequency. Neglecting the effect of losses, a half-wave transmitting dipole fed with 100 watts of power produces the same field intensity in a given direction, regardless of whether the frequency is 100 mc or 1000 mc. When used for reception, however, a half-wave dipole delivers to a matched load an amount of power that is proportional to the square of the wavelength [from eq. (12-16) its effective area is $0.13\lambda^2$]. This means, for example, that at 1000 mc, the half-wave dipole delivers to its load only $\frac{1}{100}$ times the power it would deliver at 100 mc, for the same received field intensity. This result has important bearing in the choice of frequency used for certain v-h-f and u-h-f applications that require omnidirectional antennas and which, therefore, cannot make use of the greater directivity that is usually achieved at these frequencies.

There are also other points of difference between receiving antennas and the corresponding transmitting antennas. Power-handling capacity is often a problem with transmitting antennas; seldom with receiving antennas. Economic factors enter the engineering picture. For point-to-point communication involving a single transmitter and a single receiver, one would expect to expend the same amount of money and effort improving the receiving antenna as the transmitting antenna. On the other hand in a broadcast service, where one transmitter may serve 100,000 receivers, it is justifiable economically to spend far more on the transmitting antenna than on any one receiving antenna to obtain a given improvement.

ADDITIONAL PROBLEMS

3. Derive eq. (22) from eq. (21).

4. Using the reciprocity theorem and an assumed sinusoidal current distribution for the transmitting case, show that the current distribution on a resonant-length half-wave dipole must be sinusoidal.

5. Derive an expression for the current distribution along a receiving antenna which is terminated in an arbitrary load impedance Z_L at its center. (HINT: Use the compensation theorem.)

6. For very short receiving antennas ($H \ll \lambda$) show that the efficiency, and therefore the useful received power, varies approximately as the cube of the antenna length.

BIBLIOGRAPHY

McPherson, W. L., "Electrical Properties of Aerials for Medium and Long Wave Broadcasting," *Elec. Comm.* **16**, 306 (1938).

Cutting, Fulton, "A Simple Method of Calculating Radiation Resistance," *Proc. IRE*, **10**, 129 (1922).

Brown, G. H., "A Critical Study of Broadcast Antennas as Affected by Antenna Current Distribution," *Proc. IRE*, **24**, 48 (1936).

Alford, A., "A Discussion of Methods Employed in Calculations of Electromagnetic Fields of Radiating Conductors," *Elec. Comm.*, **15**, 70 (1936).

Brown, G. H., "Directional Antennas," *Proc. IRE*, **25**, 78 (1937).

Gihring, H. E., and G. H. Brown, "General Considerations of Tower Antennas for Broadcast Use," *Proc. IRE*, **23**, 311 (1935).

Brown, G. H., "The Phase and Magnitude of Earth Currents Near Radio Transmitting Antennas," *Proc. IRE*, **23**, 168 (1935).

Brown, G. H., R. F. Lewis, and J. Epstein, "Ground Systems as a Factor in Antenna Efficiency," *Proc. IRE*, **25**, 753 (1937).

Everitt, W. L., "Coupling Networks," *Communications*, **18**, 12 (1938).

Carter, P. S., "Simple Television Antennas," *RCA Rev.*, **4**, 168 (1939).

Van Roberts, W., "Input Impedance of a Folded Dipole," *RCA Review*, **8**, 289 (1947).

Bruce, E., "Developments in Short-wave Directive Antennas," *Proc. IRE*, **19**, 1406 (1931).

Bruce, E., A. C. Beck, and L. R. Lowry, "Horizontal Rhombic Antennas," *Proc. IRE*, **23**, 24 (1935).

Foster, Donald, "Radiation from Rhombic Antennas," *Proc. IRE*, **25**, 1327 (1937).

Carter, P. S., C. W. Hansell, and N. E. Lindenblad, "Development of Directive Transmitting Antennas by RCA Communications, Inc.," *Proc. IRE*, **19**, 1773 (1931).

Brown, George H., "The Turnstile Antenna," *Electronics*, **9**, 19 (1936).

Alford, A., and A. G. Kandoian, "Ultrahigh-frequency Loop Antennas," *Trans. AIEE*, **59**, 843 (1940).

Beverage, H. H., C. W. Rice, and E. W. Kellogg, "The Wave Antenna—A New Type of Highly Directive Antenna," *Trans. AIEE*, **42**, 215 (1923).

Friis, H. T., and C. B. Feldman, "A Multiple Unit Steerable Antenna for Short Wave Reception," *Proc. IRE*, **25**, 841 (1937).

Burgess, R. E., "The Screen Loop Aerial," *Wireless Engineer*, **16**, 492 (1939).

Everitt, W. L., *Communication Engineering*, McGraw-Hill, New York, 1937.

Terman, F. E., *Radio Engineer's Handbook*, McGraw-Hill, New York, 1943.

Grammer, George, and Byron Goodman, *The A.R.R.L. Antenna Book*, American Radio Relay League, West Hartford, Conn.

Smith, Woodrow, *Antenna Manual*, Editors and Engineers Ltd., Santa Barbara, Calif., 1948.

Radio Research Laboratory Staff, *Very High-frequency Techniques*, McGraw-Hill, New York, 1947, Vol. I.

CHAPTER 15

SECONDARY SOURCES AND APERTURE ANTENNAS

15.01 Magnetic Currents. In writing Maxwell's equations

$$\text{curl } \mathbf{H} = \dot{\mathbf{D}} + i \qquad \text{curl } \mathbf{E} = -\dot{\mathbf{B}}$$

the quantities $\dot{\mathbf{D}}$, i, and $\dot{\mathbf{B}}$ are interpreted as the densities of electric displacement current, electric conduction current, and magnetic displacement current, respectively. The absence of a magnetic quantity corresponding to i, that is, to a magnetic conduction current, is explained by the fact that so far as is yet known, there are no isolated magnetic charges. As a result, it has been found possible to set up the solution of electromagnetic problems in terms of electric currents and charges alone through the relations

$$\mathbf{H} = \text{curl } \mathbf{A} \qquad \mathbf{E} = -\text{ grad } V - \mu\dot{\mathbf{A}} \tag{15-1}$$

where

$$\mathbf{A} = \frac{1}{4\pi}\int_{\text{vol}} \frac{i\left(t - \frac{r}{v}\right)}{r}\, dV \qquad V = \frac{1}{4\pi}\int_{\text{vol}} \frac{\rho\left(t - \frac{r}{v}\right)}{\epsilon r}\, dV \tag{15-2}$$

Although the relations (1) and (2) have proven adequate for the solution of problems considered up to the present, there are many other problems where the use of fictitious magnetic currents and charges is very helpful. In such problems the fields, which are actually produced by a certain distribution of electric current and charge, can be more easily computed from an "equivalent" distribution of fictitious magnetic currents and charges. An example of such "equivalent distributions" is the case of the electric current loop and the magnetic dipole. The electromagnetic field produced by a small horizontal electric current loop is identical with that produced by a vertical magnetic dipole. Conversely, the fields produced by a magnetic current loop and electric dipole are also identical. It will be found in many problems involving radiation from "aperture antennas," that the notion of magnetic currents and charges will prove an invaluable aid in arriving at solutions. Therefore the expressions for the fields due to such magnetic charges and currents will be developed. It should be emphasized that, although the magnetic charges and currents used in this procedure are fictitious, the fields calculated from them are physical fields that are actually produced by an equivalent distribution of electric charges and currents.

Written to include magnetic as well as electric conduction current, Maxwell's equations would be

$$\text{curl}\ \mathbf{H} = \dot{\mathbf{D}} + i \qquad \text{curl}\ \mathbf{E} = -\dot{\mathbf{B}} - i_m \tag{15-3}$$

or in the integral form

$$\oint \mathbf{H} \cdot d\mathbf{s} = \dot{\Psi} + I \qquad \oint \mathbf{E} \cdot d\mathbf{s} = -\dot{\Phi} - K \tag{15-4}$$

In these equations K is a magnetic conduction current and i_m is a magnetic conduction current density. K has the dimensions of volts and i_m has the dimensions of volts per square meter. For *surface magnetic current density* (corresponding to **J** for the electric case) the symbol **M** (volt/m) will be used. It is apparent from (3) and (4) that (except for a matter of sign) complete symmetry now exists in Maxwell's equations.

The positive sign in the first equations of (3) and (4) indicates that directions of magnetomotive force and electric current are related by the right-hand rule, whereas the negative sign in the second equation indicates that the directions of electromotive force and magnetic current are related through the left-hand rule.

In general it will be desired to solve problems having both electric and magnetic distributions. However, for the purpose of developing expressions due to magnetic currents and charges, consider first the case where the fields are due to these alone. Equations (3), written for magnetic currents, and in the absence of electric currents, are

$$\text{curl}\ \mathbf{H}^m = \epsilon\dot{\mathbf{E}}^m \qquad \text{curl}\ \mathbf{E}^m = -\mu\dot{\mathbf{H}}^m - i_m \tag{15-5}$$

These should be compared with the familiar relations written for electric currents alone (without magnetic currents)

$$\text{curl } \mathbf{E}^e = -\mu\dot{\mathbf{H}}^e \qquad \text{curl } \mathbf{H}^e = \epsilon\dot{\mathbf{E}}^e + i \tag{15-6}$$

(The superscripts e refer to fields due to electric currents and superscripts m refer to fields due to magnetic currents.)

Comparison of (5) and (6) shows them to be identical sets (except for sign) if electric and magnetic quantities are interchanged, that is, if $\mathbf{H}^m$, $\mathbf{E}^m$, μ, and ϵ replace $\mathbf{E}^e$, $\mathbf{H}^e$, ϵ, and μ respectively. Therefore, the procedures used in chap. 10 for electric currents can be followed to set up potentials due to magnetic currents, and the fields can be obtained from these potentials by differentiation. Corresponding to the magnetic vector potential $\mathbf{A}$ that yields the magnetic intensity through $\mathbf{H}^e = \text{curl } \mathbf{A}$, there will be an electric vector potential $\mathbf{F}$ that will yield the electric field (due to magnetic currents) through $\mathbf{E}^m = -\text{curl } \mathbf{F}$. Similarly corresponding to the scalar electric potential V that is set up in terms of the electric charges, there will be a scalar magnetic potential $\mathfrak{F}$ that is set up in terms of the magnetic charges. Rewriting eqs. (1) and (2) with suitable superscripts for fields due to electric currents and charges results in

$$\mathbf{H}^e = \text{curl } \mathbf{A} \qquad \mathbf{E}^e = -\text{grad } V - \mu\dot{\mathbf{A}} \tag{15-7}$$

$$\mathbf{A} = \frac{1}{4\pi}\int_{\text{vol}} \frac{i\left(t - \frac{r}{v}\right)}{r}\, dV \qquad V = \frac{1}{4\pi}\int_{\text{vol}} \frac{\rho\left(t - \frac{r}{v}\right)}{\epsilon r}\, dV \tag{15-8}$$

The corresponding relations for fields due to magnetic currents and charges are

$$\mathbf{E}^m = -\text{curl } \mathbf{F} \qquad \mathbf{H}^m = -\text{grad } \mathfrak{F} - \epsilon\dot{\mathbf{F}} \tag{15-9}$$

$$\mathbf{F} = \frac{1}{4\pi}\int_{\text{vol}} \frac{i_m\left(t - \frac{r}{v}\right)}{r}\, dV \qquad \mathfrak{F} = \frac{1}{4\pi}\int_{\text{vol}} \frac{\rho_m\left(t - \frac{r}{v}\right)}{\mu r}\, dV \tag{15-10}$$

The fields due to a magnetic dipole can now be developed from (9) and (10), or they may be written down directly from comparison with the fields due to an electric dipole. For the distant field of a magnetic current element, placed at the center and lying along the polar axis of a spherical co-ordinate system, the lines of $\mathbf{E}$ will lie along circles of latitude and lines of **H** will lie along the meridians in the directions* indicated in Fig. 15-1.

For problems where *both* electric and magnetic current and charge distributions are involved, the *total* electric and magnetic intensities (indicated by no subscript) will be the sum of the intensities produced by the distributions separately. Writing

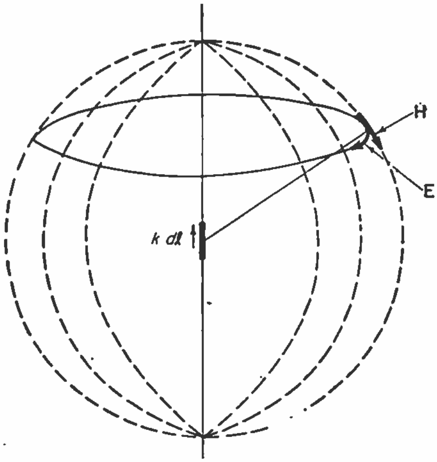


FIG. 15-1. Lines of **E** (solid) and **H** (dashes) on a large spherical surface centered on a magnetic current element $K\,dl$.

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$$

and

$$\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$$

the fields will be given by

$$\mathbf{E} = -\operatorname{grad} V - \mu\dot{\mathbf{A}} - \operatorname{curl}\mathbf{F}$$
$$\mathbf{H} = -\operatorname{grad} \mathfrak{F} - \epsilon\dot{\mathbf{F}} + \operatorname{curl}\mathbf{A}$$

For the usual case where time variations are written as $e^{j\omega t}$, these expressions become†

$$\mathbf{E} = -\operatorname{grad} V - j\omega\mu\mathbf{A} - \operatorname{curl}\mathbf{F} \qquad (15\text{-}11)$$
$$\mathbf{H} = -\operatorname{grad} \mathfrak{F} - j\omega\epsilon\mathbf{F} + \operatorname{curl}\mathbf{A} \qquad (15\text{-}12)$$

There is one last relation connected with magnetic currents that must be considered. It was found that tangential H was discontinuous across an electric-current sheet (though tangential E

* For later use it is noted in passing that, since the electric field is perpendicular to any vertical plane containing the element, a perfectly conducting sheet may be placed in such a plane without disturbing the field. Lines of H adjacent to this conducting plane will be tangential to it, and electric currents will flow on its surface, their magnitude and direction being given by

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}$$

† For dissipative media it is necessary to write the second term of (12) as $-(\sigma + j\omega\epsilon)\bar{F}$ instead of just $-j\omega\epsilon\bar{F}$ to include the effects of electric conduction currents due to magnetic sources in addition to electric displacement currents due to magnetic sources. There is, of course, no corresponding term in (11) for magnetic conduction currents due to electric sources. In this book there will be no occasion to deal with the effects of magnetic sources in dissipative media, and eqs. (11) and (12) will suffice.

remained continuous). The discontinuity in tangential H is equal to the linear current density J as shown by the relation

$$H_{1\,\text{tan}} - H_{2\,\text{tan}} = J \qquad \text{(amp/m)}$$

This result was obtained directly from an application of Maxwell's mmf equation I; the continuity of tangential E followed from the emf equation II (section 4.05). The electric-current density J and tangential H are mutually perpendicular, and this fact is indicated by the vector relation

$$\mathbf{J} = \mathbf{n} \times (\mathbf{H}_1 - \mathbf{H}_2) \qquad \text{(amp/m)} \qquad (15\text{-}13)$$

where $\mathbf{n}$, the unit vector normal to the current sheet, is regarded as positive when pointing to the side that contains H_1. In exactly the same way it is found from equations (4) that tangential E is discontinuous across a magnetic-current sheet, whereas tangential H remains continuous. For a linear magnetic current density M (volt/m) the relation corresponding to (13) is

$$\mathbf{M} = -\mathbf{n} \times (\mathbf{E}_1 - \mathbf{E}_2) \qquad \text{(volt/m)} \qquad (15\text{-}14)$$

The minus sign results from the minus sign in the second of eqs. (4). Equation (14) states that the tangential electric intensity is discontinuous across a magnetic-current sheet by an amount equal to the linear magnetic-current density.

Examples of the use of magnetic currents will appear in the sections that follow.

15.02 The Induction and Equivalence Theorems. As was pointed out in the previous section, it is always possible, at least in theory, to determine the electromagnetic field of a system from its electric currents and charges alone. In practice there are many problems whose solutions by this method are prohibitively difficult, and yet some of these may be solved without too much labor by other means. Two examples of problems that are difficult to solve in terms of the currents of the system are illustrated in Fig. 15-2. The first of these concerns the radiation from the open end of a semi-infinite coaxial line. In this case, assuming that the transverse dimensions of the line are very small in wavelengths, the current distribution is known fairly accurately, but the problem is made difficult by the fact that all currents throughout the infinite length of the line must be considered in the integration to determine the radiated field. The second problem involves the radiation from the open end of a wave guide or from an electromagnetic horn. Here the currents are known only approximately, especially around the mouth of the horn. But even though only an approximate solution is required, the integration to obtain the fields from the known or guessed at currents is extremely difficult because all currents, including those on the probe antenna and in the coaxial feed line, must be included.

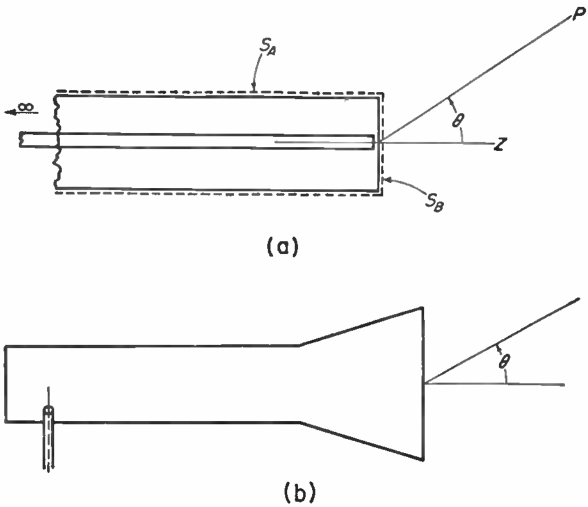


FIG. 15-2. Radiation from (a) the open end of a coaxial line (b) an electromagnetic horn.

In both of these problems it seems evident that there should be a simpler way of obtaining the radiated field. In particular, although all currents of the system are involved in determining the radiation, it appears reasonable that for systems such as those of Fig. 15-2, the currents can only affect the radiated field through some change that they make in the fields that appear across the open end of the coaxial line or wave guide horn. These latter fields are known (case *a*), or can be guessed at (case *b*), to the same order of approximation as the currents of the system. Therefore a method for computing radiated fields in terms of known fields across an aperture will be sought.

The means for accomplishing the result is suggested by *Huygens' principle.* This principle states that "each particle in any wave front acts as a new source of disturbance, sending out secondary waves, and these secondary waves combine to form a new wave front." Huygen's principle has long been used in optics to obtain qualitative answers to diffraction problems. It can be used to give quantitative results when suitably combined with two other theorems, the induction and equivalence theorems, which are due to Schelkunoff.

In Fig. 15-3 the closed surface S separates two homogeneous media, one containing a system of sources s_1, and the other being source-free. In general the field in region (2) will be different from

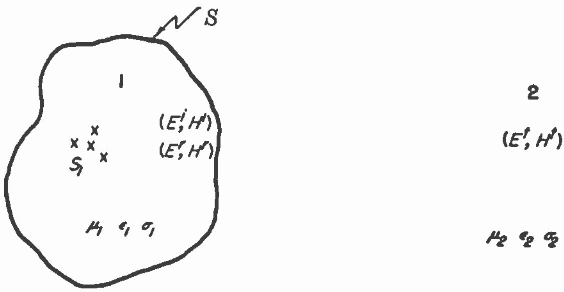


FIG. 15-3. A closed surface S divides a region (1) containing sources from a source-free region (2).

the value that it would have if media (1) and (2) were the same. The actual field in (2) can be determined by treating the problem as a reflection problem in which an incident field (E^i, H^i) sets up at the boundary surface S a reflected field (E^r, H^r) and a transmitted field (E^t, H^t). The incident field (E^i, H^i) is the field that would exist if there were no reflecting surface, that is, if the entire region were homogeneous. The actual field in region (1) is $(E^i + E^r, H^i + H^r)$; the actual field in region (2) is (E^t, H^t). At any actual boundary surface S, the tangential components of these fields are continuous. That is,

$$E_t^i + E_t^r = E_t^t \qquad H_t^i + H_t^r = H_t^t \qquad (15\text{-}15)$$

where the subscript t indicates the components tangential to the surface. Equation (15) can be rewritten in the form

$$(E_t^t - E_t^r) = E_t^i \qquad (H_t^t - H_t^r) = H_t^i \qquad (15\text{-}15a)$$

Now let attention be concentrated on the "induced" or "reradiated" fields (E^r, H^r) and (E^t, H^t). The reflected fields (E^r, H^r) satisfy Maxwell's equations in the homogeneous medium (1), and the transmitted or refracted fields satisfy Maxwell's equations in the homogeneous region (2). Together these fields constitute an electromagnetic field in the entire space. This field is source-free everywhere except on S, and the distribution of sources on S is calculable from the incident field, and therefore from the given sources s_1. In section 15.01 it was shown that the discontinuities in E and H across the surface S could be produced by current sheets on S of densities

$$\begin{aligned} \mathbf{J} &= \mathbf{n} \times (\mathbf{H}_t^t - \mathbf{H}_t^r) = \mathbf{n} \times \mathbf{H}^i \\ \mathbf{M} &= -\mathbf{n} \times (\mathbf{E}_t^t - \mathbf{E}_t^r) = -\mathbf{n} \times \mathbf{E}^i \end{aligned} \qquad (15\text{-}16)$$

Thus as far as the "induced" or reradiated field is concerned, it could be produced by electric- and magnetic-current sheets over the surface S, the densities of these sheets being given by (16). This is the *induction theorem.*

A second theorem follows directly from the induction theorem for the particular case where region (2) has the same constants as region (1), that is where the entire region is homogeneous. In this case the reflected field is zero and the transmitted field is the actual field in the homogeneous region due to the sources of s_1. But this transmitted field can also be calculated from a suitable distribution of electric- and magnetic-current sheets over the surface S. The required surface current densities of these sheets will be

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}_t^t = \mathbf{n} \times \mathbf{H}_t^i \qquad \mathbf{M} = -\mathbf{n} \times \mathbf{E}_t^t = -\mathbf{n} \times \mathbf{E}_t^i \qquad (15\text{-}17)$$

Since the vector product of $\mathbf{n}$ and the normal component of the field is zero, the t subscripts can be dropped and eqs. (17) written as

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}^i \qquad \mathbf{M} = -\mathbf{n} \times \mathbf{E}^i \qquad (15\text{-}18)$$

The vector $\mathbf{n}$ is in the direction of the transmitted wave. Thus in a source-free region bounded by a surface S, in order to compute the electromagnetic field, the source distribution s_1 (outside of S) can be replaced by a distribution of electric and magnetic currents over the surface S, where the densities of this "equivalent" source distribution are given by (18). This is the *equivalence theorem.*

These theorems prove to be powerful tools in solving many electromagnetic problems that involve radiation from apertures.

15.03 Field of a Secondary or Huygen's Source. It will be of interest to determine first the radiation field of a Huygen's source, or an element of area of an advancing wavefront in free space. In Fig. 15-4 is shown an element of area $dx\,dy$ on the wavefront of a uniform plane (TEM) wave, which is advancing in the z direction. By the theorems of the previous section this element of wavefront having electric intensity E_x^0 and magnetic intensity $H_y^0 = E_x^0/\eta_v$ can

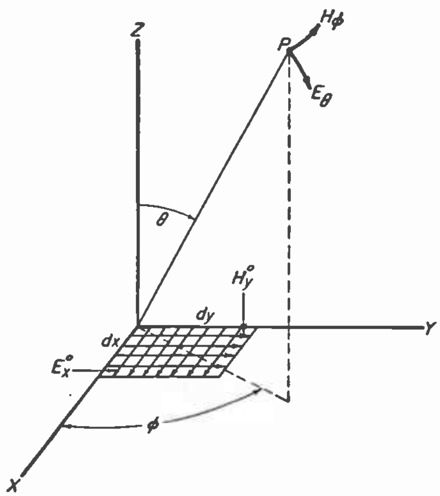


FIG. 15-4. Radiation from a Huygen's source.

be treated as a secondary source and can be replaced by electric and magnetic sheets. Using eqs. (18), the directions and densities of these current sheets will be

$$J_x = -H_y^0 = -\frac{E_x^0}{\eta_v} \qquad M_y = -E_x^0$$

The element of area $dx\,dy$ of electric surface current density J_x constitutes an electric current element $(J_x\,dy)\,dx$, and similarly the element of area of magnetic surface current density M_y constitutes a magnetic current element $(M_y\,dx)\,dy$. The problem is simply one of determining the radiation fields of these current elements. Because only the radiation fields are required, only the vector potentials need be considered. (The scalar potentials contribute nothing to the radiation or inverse-distance fields.) At large distances, the electric field of the electric-current element will be

$$\mathbf{E}^e = -j\omega\mu\mathbf{A}; \qquad \mathbf{A} = \mathbf{i}A_x = \mathbf{i}\left[\frac{(J_x\, dy)\, dx\, e^{-j\beta r}}{4\pi r}\right]$$

The θ and ϕ components of the radiation field of the electric-current element will be

$$E_\theta{}^e = -j\omega\mu A_\theta \qquad H_\phi{}^e = \frac{E_\theta{}^e}{\eta_v}$$

$$E_\phi{}^e = -j\omega\mu A_\phi \qquad H_\theta{}^e = -\frac{E_\phi{}^e}{\eta_v}$$

where $\quad A_\theta = A_x \cos\phi \cos\theta \qquad A_\phi = -A_x \sin\phi$

Similarly for the magnetic-current element the radiation fields are

$$\mathbf{H}^m = -j\omega\epsilon\mathbf{F}; \qquad \mathbf{F} = \mathbf{j}F_y = \mathbf{j}\left[\frac{(M_y\, dx)\, dy}{4\pi r}\, e^{-j\beta r}\right]$$

$$H_\phi{}^m = -j\omega\epsilon F_\phi \qquad E_\theta{}^m = \eta_v H_\phi{}^m$$

$$H_\theta{}^m = -j\omega\epsilon F_\theta \qquad E_\phi{}^m = -\eta_v H_\theta{}^m$$

$$F_\phi = F_y \cos\phi \qquad F_\theta = F_y \sin\phi \cos\theta$$

Expressing all the fields in terms of E_x^0, the radiation field of the Huygens source is found to be

$$E_\theta = E_\theta{}^e + E_\theta{}^m = \frac{jE_x^0\, dx\, dy\, e^{-j\beta r}}{2\lambda r}(\cos\phi\cos\theta + \cos\phi) \tag{15-19}$$

$$E_\phi = E_\phi{}^e + E_\phi{}^m = \frac{-jE_x^0\, dx\, dy\, e^{-j\beta r}}{2\lambda r}(\sin\phi + \sin\phi\cos\theta) \tag{15-20}$$

$$H_\phi = \frac{E_\theta}{\eta_v} \qquad H_\theta = \frac{-E_\phi}{\eta_v} \tag{15-21}$$

In the plane $\phi = 0$, the magnitude of the electric intensity is

$$|E_\theta| = \frac{E_x^0\, dx\, dy}{2\lambda r}[1 + \cos\theta] \qquad |E_\phi| = 0 \tag{15-22}$$

In the plane $\phi = 90$ degrees, the magnitude of the electric field is

$$|E_\theta| = 0 \qquad |E_\phi| = \frac{E_x^0\, dx\, dy}{2\lambda r}(1 + \cos\theta) \tag{15-23}$$

In the principal planes, which contain the axis of propagation, the radiation patterns of an element of wave-front have a heart-shaped or unidirectional pattern. The radiation is a maximum in the forward direction ($\theta = 0$); it is zero in the backward direction ($\theta = 180$ degrees). In one sense, this "explains" why an electromagnetic wave, once launched, continues to propagate in the forward direction. An electric-current sheet alone radiates equally on both sides; similarly a magnetic-current sheet alone radiates equally on both sides; but crossed electric- and magnetic-current sheets of proper relative magnitude and phase can be made to radiate on one side only. A large square surface of a plane wavefront constitutes a rectangular array of Huygen's sources, all fed in phase. The "radiation pattern" of the array is obtained by multiplying the unit pattern of the element (a cardioid pattern) by the group pattern or array factor, which in this case is a bidirectional pencil beam. The resultant pattern is a unidirectional pencil beam, the cone angle of which becomes very small as the area of the wavefront becomes large.

15.04 Radiation from the Open End of a Coaxial Line. By application of the new approaches outlined in previous sections, the problem of radiation from the open end of a coaxial cable of small cross-sectional dimensions can now be solved quite easily. If the surface S, separating the source-free region from the region containing sources, is taken to be the surface shown dotted in Fig. 15-2, it is only necessary to specify the equivalent electric- and magnetic-current sheets over this surface. The surface can be divided into two parts: S_a is the cylindrical surface that encloses the outer wall of the coaxial line; S_b is the flat circular surface which caps the end of the line. Over S_a, tangential E is tangential to the metallic wall and has zero value. Therefore the equivalent magnetic-current sheet has zero density over S_a. Also the magnetic intensity at the outer surface of the outer conductor must be zero, because for any circumferential path enclosing both conductors, the total current enclosed is zero. (This assumes that the transverse dimensions are small, so that, except right near the ends, only the TEM wave exists.) With tangential H over S_a equal to zero, the equivalent electric-current sheet must also be zero over this part of the surface, and there remains only the contribution from S_b. Over S_b the electric field is radial and the magnetic intensity is circumferential;

but whereas the electric field is relatively strong, to a first approximation the magnetic intensity is zero (the open end is a current node). Therefore the radiation from the open end can be computed by use of an equivalent magnetic-current sheet only over the surface, S_b. Having determined the radiation fields, and thence the power radiated, it is possible to compute from consideration of power flow through the open end, the small value of magnetic intensity that actually must exist there. If desired, an equivalent electric-current

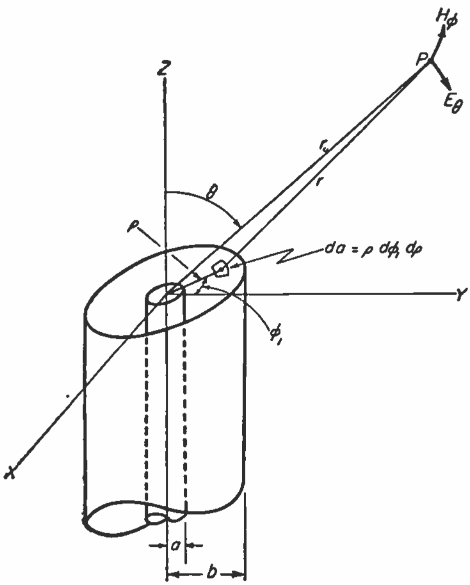


FIG. 15-5. Geometry for calculation of power radiated by open end of a coaxial cable.

sheet could then be set up for this magnetic intensity, and the small radiation field of the electric-current sheet could be calculated. In practice this second approximation rarely needs to be made, because it produces only a very small correction to the radiation fields and power radiated.

Figure 15-5 shows the geometry appropriate for calculation of the radiation from the open end of a coaxial cable. Between the inner radius a and the outer radius b, the radial electric intensity E_ρ will have a value

$$E_\rho = \frac{k}{\rho} \qquad \text{where} \qquad k = \frac{V}{\ln b/a}$$

V is the voltage between inner and outer conductors at the open end. Using (18), E_ρ can be replaced by a magnetic-current sheet

$$M_\phi = -E_\rho = -\frac{k}{\rho}$$

The electric vector potential **F** will be in the ϕ direction and will have a value

$$F_\phi = \frac{1}{4\pi}\int_A \frac{M_\phi(t - r/v)\, da}{r}$$
$$= \frac{1}{4\pi}\int_0^{2\pi}\int_a^b \frac{M_\phi\, e^{-j\beta r}}{r}\,\rho\, d\rho\, d\phi_1 \qquad (15\text{-}24)$$

where the integration is over the area A between inner and outer inductors, and where time variations as $e^{j\omega t}$ have been assumed. Without loss of generality the point P may be taken in the y-z plane, for which case only the x components, $M_\phi \cos \phi_1$, of magnetic current contribute to the potential, the y components cancelling out. Because only distant fields are being considered, the r in the denominator of (24) can be put equal to r_0, and the r in the phase factor in the numerator may be replaced by

$$r \approx r_0 - \rho \sin\theta \cos\phi_1$$

Making these substitutions, and remembering that for small values of δ the exponential $e^{j\delta}$ can be replaced by the first two terms of its power series expansion, viz.,

$$e^{j\delta} \approx 1 + j\delta$$

the integration indicated by (24) can be carried out. The result is

$$F_\phi = -\frac{j\beta k \sin\theta\, e^{-j\beta r_0}}{8r_0}(b^2 - a^2) \qquad (15\text{-}25)$$

The distant magnetic intensity is obtained from

$$H_\phi = -j\omega\epsilon F_\phi$$

so that

$$H_\phi = \frac{-\beta\omega\epsilon k \sin\theta}{8r_0}(b^2 - a^2)\, e^{-j\beta r_0} \qquad (15\text{-}26)$$

The intensity of the distant electric field will be

$$E_\theta = \eta_v H_\phi$$

By integrating the Poynting vector over a large spherical surface, the radiated power is found to be

$$W = \frac{\pi^2 V^2}{360}\left(\frac{A}{\lambda^2 \ln b/a}\right)^2 \qquad \text{watts} \qquad (15\text{-}27)$$

where $A = \pi(b^2 - a^2)$ is the area of the opening between inner and outer conductors.

15.05 Radiation through an Aperture in an Absorbing Screen. Another example of radiation through an aperture occurs in the problem of the transmission of electromagnetic energy through a rectangular aperture in a perfectly absorbing screen. Although admittedly not a very practical problem, because of the difficulties

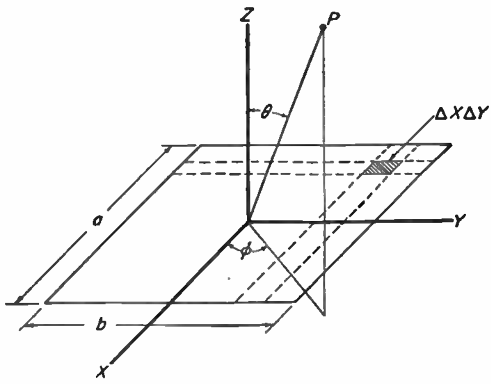


FIG. 15-6. An element of area on an advancing wavefront.

of obtaining a screen which is both infinitely thin and perfectly absorbing, the solution to this problem is required in obtaining answers to other, more practical problems.

In Fig. 15-6 the rectangle ab represents an aperture in a perfectly absorbing screen of infinite extent which occupies the $z = 0$ plane. A uniform plane electromagnetic wave traveling in the z direction is assumed to be incident upon the bottom side of the screen and aperture, and the problem is that of determining the radiation through the aperture in the positive z direction. Under the assumed conditions of the problem, the incident wave is completely absorbed at the surface of the screen. Over the aperture the field intensity will be just that of the incident wave. By dividing up the aperture into a large number of Huygen's sources of area $da = \Delta x\, \Delta y$, the aperture may be treated as a rectangular array of such sources, all fed in phase.

For a line array of nondirective sources in the x direction, having a uniform spacing Δx, the radiation pattern or space factor is (from section 12.06).

$$S_x = |1 + e^{j\psi} + e^{2j\psi} + \cdots + e^{(m-1)j\psi}|$$
$$= \left|\sum_0^{m-1} e^{j\beta(\Delta x)\sin\theta\cos\phi}\right|$$
$$= \left|\frac{\sin[\frac{1}{2}m\beta(\Delta x)\sin\theta\cos\phi]}{\sin[\frac{1}{2}\beta\,\Delta x\sin\theta\cos\phi]}\right|$$

Similarly, for a line array in the y direction with a uniform spacing Δy, the space factor is

$$S_y = \left|\frac{\sin[\frac{1}{2}n\beta(\Delta y)\sin\theta\sin\phi]}{\sin[\frac{1}{2}\beta(\Delta y)\sin\theta\sin\phi]}\right|$$

The total space factor for the rectangular array of isotropic sources is then

$$S_{x,y} = S_xS_y$$

If Δx and Δy are now allowed to become small, but m and n are made large in such manner that

$$(m-1)\,\Delta x = a \qquad (n-1)\,\Delta y = b$$

the space factor $S_{x,y}$ may be written

$$S_{x,y} = \frac{ab}{\Delta x\,\Delta y}\left|\frac{\sin(\frac{1}{2}\beta a\sin\theta\cos\phi)}{(\frac{1}{2}\beta a\sin\theta\cos\phi)}\cdot\frac{\sin(\frac{1}{2}\beta b\sin\theta\sin\phi)}{(\frac{1}{2}\beta b\sin\theta\sin\phi)}\right| \quad (15\text{-}28)$$

multiplying this space factor or group pattern by the radiation from the unit Huygen's source gives the total radiation from the aperture. Then, using (19) and (20),

$$E_\theta = \frac{jE_x^0abe^{-j\beta r}}{2\lambda r}[(1+\cos\theta)\cos\phi]\left|\frac{\sin u_1}{u_1}\cdot\frac{\sin v_1}{v_1}\right| \quad (15\text{-}29)$$

$$E_\phi = \frac{-jE_x^0abe^{-j\beta r}}{2\lambda r}[(1+\cos\theta)\sin\phi]\left|\frac{\sin u_1}{u_1}\cdot\frac{\sin v_1}{v_1}\right| \quad (15\text{-}30)$$

where

$$u_1 = \tfrac{1}{2}\beta a\sin\theta\cos\phi \qquad \text{and} \qquad v_1 = \tfrac{1}{2}\beta b\sin\theta\sin\phi$$

This is known as the *diffracted* field and this problem is an example of *Fraunhofer diffraction*. In the principal x-z plane, $\phi = 0$ or π,

$E_\phi = 0$, and the E_θ field is given by

$$E_\theta = \frac{jE_z^0 abe^{-j\beta r}}{2\lambda r}(1 + \cos\theta)\left|\frac{\sin u}{u}\right| \tag{15-31}$$

where $$u = [\tfrac{1}{2}\beta a \sin\theta] = \frac{\pi a}{\lambda}\sin\theta$$

The square bracketed expression, $(\sin u)/u$, occurs in many radiation and diffraction problems. It is plotted in Fig. 15-7.

It will be noted that the first null in this general pattern occurs at $u = \pi$. Thus, for an aperture width $a = 1\lambda$, the first null of the pattern occurs at $\theta = 90$ degrees. For aperture widths smaller

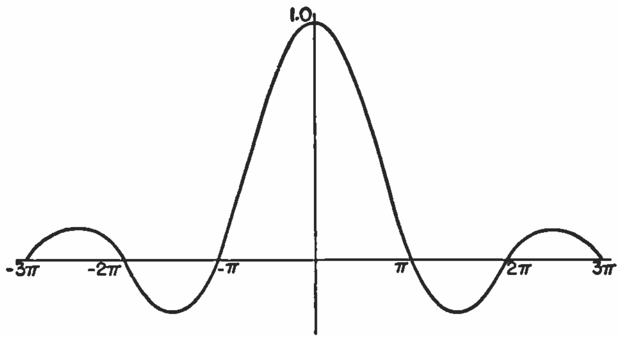


FIG. 15-7. Plot of a sin u/u.

than 1λ, there is no null in the pattern. For very large apertures, the "beam" is quite narrow, and small angles of θ are of most interest. For small values of θ, $\sin\theta \approx \theta$, and

$$u \approx \frac{\pi a\theta}{\lambda}$$

For $a = 10\lambda$, the first null occurs at $\theta_0 = 0.1$ radian or 5.7 degrees.

Application to Open-ended Waveguides. If it can be assumed that the currents on the outside walls of an open-ended waveguide have negligible effect on the radiation from the guide, the problem of diffraction through an aperture has direct application to this second, more practical, problem. Since experimentally measured radiation patterns are found to agree roughly with patterns computed by neglecting these outside currents, such calculations may be used if only approximate answers are sufficient.

Referring to Fig. 15-8, and assuming that the fields at the open end are approximately the same as they would be if the guide didn't terminate there, but continued on to infinity, the tangential fields at the end surface will be

$$E_y^0 = E^0 \sin \frac{\pi x}{a} \qquad H_x^0 = -\frac{1}{\eta}\sqrt{1 - \left(\frac{\lambda^2}{\lambda_c}\right)}\, E_y^0$$

The dominant $TE_{1,0}$ mode has been assumed, and for this mode $\lambda_c = 2a$. For a very wide guide ($a \gg \lambda/2$) carrying the dominant mode, $|E_y^0/H_x^0| \approx \eta$, and the problem is the same as that of the rectangular aperture, except for the variation of E and H in the x

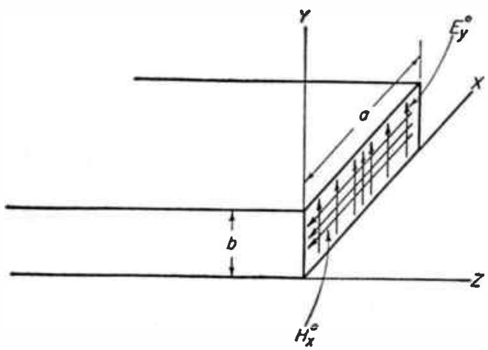


FIG. 15-8. Field at the open end of a waveguide.

direction across the mouth of the guide. For narrower guides, with operation closer to the cut-off frequency [i.e., as $\lambda \to 2a$, and $\sqrt{1 - (\lambda/\lambda_c)^2} \to 0$], H_x^0 becomes very small and there are two important effects. First, the characteristic impedance of the guide becomes very great, so that there is now a large mismatch between the impedance of the guide and the effective terminating impedance. This means that more of the energy is reflected back from the open end, and less is radiated for a given value of E_y^0. Second, the radiation pattern approaches more closely that which would be calculated from a magnetic-current sheet alone, rather than from crossed electric- and magnetic-current sheets. Experimental and calculated radiation patterns of waveguides and horns may be found in the literature.*

* Radio Research Laboratory, Staff, *Very High Frequency Techniques*, McGraw-Hill, New York, 1947, Vol. 1, Sec. 6-4.

15.06 Fraunhofer and Fresnel Diffraction. For many problems involving radiation from apertures, and also for solving certain propagation problems, some knowledge of classical diffraction theory is required. In the previous section an example of *Fraunhofer* diffraction was encountered, whereas in the problem of radiation from an electromagnetic horn, *Fresnel* diffraction will be of interest. The difference between these is illustrated in Fig. 15-9. In the case of Fraunhofer diffraction both the source and receiving point are so remote from the aperture or screen that the rays may be considered as being essentially parallel. In Fig. 15-9, this means that rays arriving from the secondary source (the aperture D), may

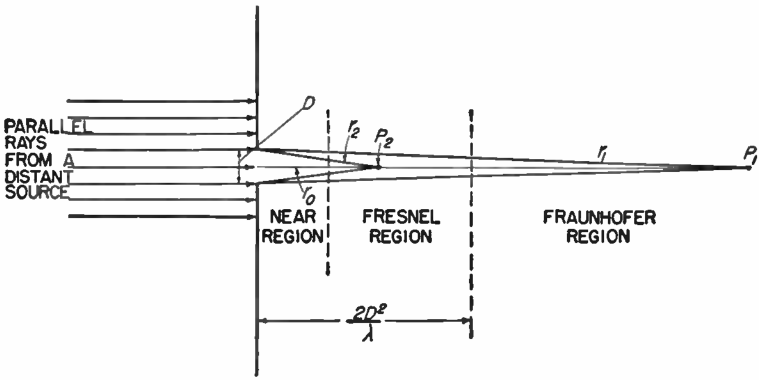


FIG. 15-9. Illustration of Fresnel and Fraunhofer regions in diffraction theory.

be considered to arrive in-phase at a point P_1 which is on a line drawn normal to the screen through the aperture. On the other hand, if the distance r to the receiving point P_2 is sufficiently large that the amplitude factor $1/r$ may still be considered constant, but is not so large that the phase difference of contributions from the various Huygen's sources over the aperture may be neglected, the point P_2 is in the region of Fresnel diffraction. The region so close to the aperture that both the amplitude and phase factors are variable with the position of the receiving point is sometimes called the *near region*. The dividing line between Fresnel and Fraunhofer diffraction depends upon the accuracy required; however, the distance to the dividing line is often taken* as $r = 2D^2/\lambda$. If the

* *IRE Standards on Antennas*, 1948.

distance to the receiving point is very great, but the source is so close to the screen that the phase of the field varies over the aperture (with the source on the normal to the screen through the aperture), Fresnel diffraction theory is required.

Fresnel Diffraction at a Straightedge. Figure 15-10 illustrates a simple example of Fresnel diffraction. An obstacle, such as a straightedge (considered to be perfectly absorbing), is inserted between a transmitting source T and a receiving location R. To keep the problem two-dimensional, the source T is assumed to be a

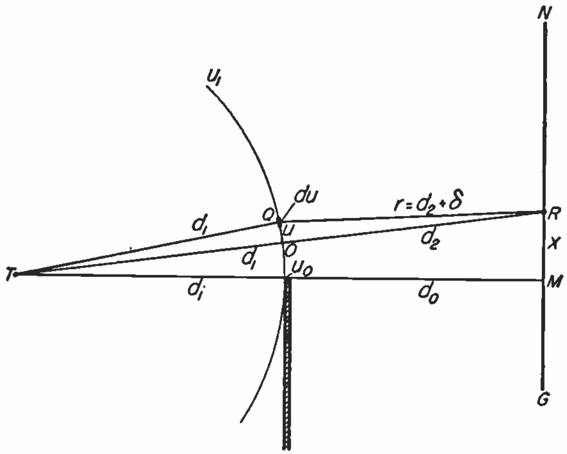


FIG. 15-10. Diffraction at a straightedge.

very long line source parallel to the long straightedge. The problem is to determine the intensity at the receiving point R, as R is moved along the line GMN. It is assumed that the distances d_0 and d_1 are sufficiently large that the approximations inherent in Fresnel diffraction theory are valid, but not large enough to permit the approximations used in Fraunhofer diffraction.

Assume that each elemental strip du of the wave front produces an effect at R given by

$$dE = \frac{k_1\, du\, e^{-j\beta r}}{f(r)} \tag{15-32}$$

where $\beta = 2\pi/\lambda$, $f(r)$ is a function of r, and k_1 is a constant. For Fresnel diffraction, the r in the denominator of (32) can be considered constant but the variation of r in the phase shift factor must be accounted for. By geometry,

$$(QR)^2 = r^2 = (d_1 + d_2)^2 + d_1^2 - 2d_1(d_1 + d_2) \cos \frac{u}{d_1}$$

$$\approx (d_1 + d_2)^2 + d_1^2 - 2\, d_1\, (d_1 + d_2) \left(1 - \frac{u^2}{2d_1^2}\right)$$

then

$$r^2 = (d_2 + \delta)^2 \approx d_2^2 + u^2 \frac{d_1 + d_2}{d_1}$$

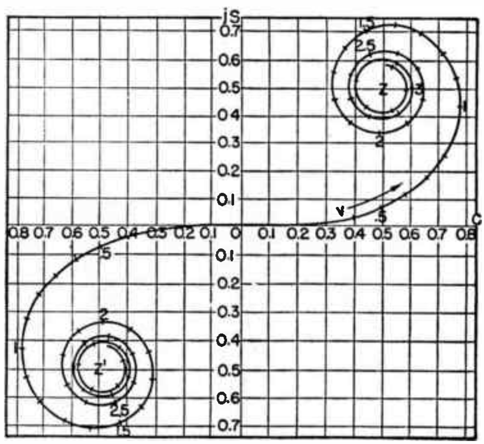


FIG. 15-11(a). Cornu's spiral.

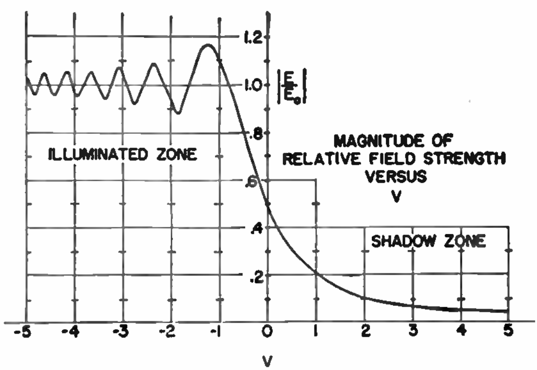


FIG. 15-11(b). Diffracted field obtained by use of (a).

Neglecting δ^2, this gives

$$\delta = u^2 \frac{d_1 + d_2}{2d_1d_2} \approx u^2 \frac{d_1 + d_0}{2d_1d_0}$$

The total effect at R, due to the portion of the wavefront between u_0 and u_1, will be

$$E = \frac{k_1}{f(r)} \int_{u_0}^{u_1} e^{-j\beta r}\, du = \frac{k_1\, e^{-j\beta d_2}}{f(d_2)} \int_{u_0}^{u_1} e^{-j\beta\delta}\, du \tag{15-33}$$

$$= \frac{k_1\, e^{-j\beta d_2}}{f(d_2)} \left(\int_{u_0}^{u_1} \cos \beta\delta\, du - j \int_{u_0}^{u_1} \sin \beta\delta\, du \right) \tag{15-33a}$$

where

$$\beta\delta = \frac{\pi}{\lambda} \left(\frac{d_1 + d_2}{d_1d_2} \right) u^2$$

The square of the magnitude of the field intensity at R is given by

$$|E|^2 = \frac{k_1^2}{f^2(d_2)} \left(\int_{u_0}^{u_1} \cos \beta\delta\, du \right)^2 + \left(\int_{u_0}^{u_1} \sin \beta\delta\, du \right)^2 \tag{15-34}$$

To evaluate and interpret this result consider the following integral

$$C(v) - jS(v) = \int_0^v e^{-j(\pi/2)v^2}\, dv \tag{15-35}$$

which is a standard form of the *Fresnel integrals*. Plotting this integral in the complex plane, with C as the abscissa and S as the ordinate, results in a curve known as *Cornu's spiral* (Fig. 15-11a). In this figure, positive values of v appear in the first quadrant and negative values of v in the third quadrant. The spiral has some interesting and important properties:

$$C = \int_0^v \cos \frac{\pi v^2}{2}\, dv \qquad S = \int_0^v \sin \frac{\pi v^2}{2}\, dv \tag{15-36}$$

$$\delta s = \sqrt{(\delta C)^2 + (\delta S)^2} = \delta v; \qquad v = s$$

$$\tan \phi = \frac{\delta S}{\delta C} = \tan \frac{\pi v^2}{2}; \qquad \phi = \frac{\pi v^2}{2} = \frac{\pi s^2}{2}$$

$$\frac{d\phi}{ds} = \pi s; \qquad \text{radius of curvature} = \frac{ds}{d\phi} = \frac{1}{\pi s}$$

$$C(\pm\infty) = \pm\tfrac{1}{2}; \qquad S(\pm\infty) = \pm\tfrac{1}{2} \tag{15-37}$$

The following properties follow from the above relations:

(1) A vector drawn from the origin to any point on the curve represents in both magnitude and phase the value of the integral (35). (The phase of the vector is the negative of the phase of the integral.)

(2) The length s of arc along the spiral, measured from the origin, is equal to v. As v approaches plus or minus infinity, the spiral winds an infinity of times about the points $(\frac{1}{2}, \frac{1}{2})$ or $(-\frac{1}{2}, -\frac{1}{2})$.

(3) The magnitude $\sqrt{C^2 + S^2}$ of the integral has a maximum value when $\phi = 3\pi/4$, or at $v = \sqrt{3/2} = 1.225$. Secondary maxima occur at

$$\phi = \frac{3\pi}{4} + 2n\pi \qquad \text{or} \qquad v = \sqrt{\frac{3}{2} + 4n} \qquad (n = 1, 2, 3, \cdots)$$

Minima occur at

$$v = \sqrt{7/2 + 4m} \qquad (m = 0, 1, 2, 3, \cdots)$$

Returning now to the integral of (33), it can be put in the standard form by writing

$$\beta\delta = \frac{\pi}{\lambda}\left(\frac{d_1 + d_2}{d_1 d_2}\right) u^2 = \frac{\pi}{2} v^2$$

or

$$v = u \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} = k_2 u$$

Then

$$E = k_3 \int_{v_0}^{v_1} e^{-j(\pi/2)v^2}\, dv$$

where

$$k_3 = \frac{k_1\, e^{-j\beta d_s}}{k_2 f(d_2)}$$

Using (36),

$$E = k_3\left(\int_0^{v_1} e^{-j\frac{\pi v^2}{2}}\, dv - \int_0^{v_0} e^{-j(\pi/2)v^2}\, dv\right)$$
$$= k_3[C(v_1) - C(v_0) - jS(v_1) + jS(v_0)]$$

Because v is proportional to u and inversely proportional to the square root of the wavelength (which is very small in optics), v_1 will be a very large number for large values of u_1. Therefore $C(v_1) \to C(\infty)$ as u_1 is allowed to become large. Then using (37), the field intensity will be approximately

$$E = K\{[\tfrac{1}{2} - C(v_0)] - j[\tfrac{1}{2} - S(v_0)]\}$$

The quantities $(\frac{1}{2} - C)$ and $j(\frac{1}{2} - S)$ represent the real and imaginary parts of a vector drawn from the upper point of convergence $(\frac{1}{2}, \frac{1}{2})$ to a point on the spiral. Thus the magnitude of E is proportional to the length of the vector drawn from $(\frac{1}{2}, \frac{1}{2})$ to the appropriate point on the spiral. This makes it possible to visualize the intensity variation as v_0 (and hence either u_0 or d_1 or d_2) is varied.

For u_0 equal to a large negative value, the free-space field intensity E_0 results. Therefore

$$E_0 = K\{[\tfrac{1}{2} - (-\tfrac{1}{2})] - j[\tfrac{1}{2} - (-\tfrac{1}{2})]\} = K(1 - j)$$

and therefore
$$K = \frac{E_0}{1 - j} = \frac{E_0}{2}(1 + j)$$

The received field intensity is given in terms of the free-space field by

$$E = \frac{E_0}{2}(1 + j) \int_{v_0}^{\infty} e^{-j\frac{\pi}{2}v^2}\, dv \qquad (15\text{-}38)$$

where
$$v_0 = u_0 \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}}$$

In order for this approximate treatment to be valid, the following inequalities must hold:

$$d_1, d_2 \gg u_0; \qquad d_1 d_2 \gg \lambda$$

In Fig. 15-11b is plotted the magnitude $|E/E_0|$ as taken off the spiral. The field intensity in the shadow zone decreases smoothly to zero. Above the line of sight the field intensity oscillates about its free-space value. On the line of sight the field intensity is just one-half of its free-space value.

This approximate theory of diffraction was developed for use in optics, where the approximations and assumptions made are usually quite valid. However, it is found that even at radio frequencies, and especially at ultrahigh frequencies, there are many problems where the theory is applicable. An example occurs in computing the radiation from electromagnetic horns.

15.07 Radiation from Electromagnetic Horns. In order to secure greater directivity a wave guide can be flared out to form an electromagnetic horn. A rectangular guide flared out in one plane only constitutes a sectoral horn, whereas a guide flared in both planes forms a pyramidal horn. The sectoral horn flared out in the plane of the electric field is the easiest case to treat and will serve as an example of the method of attack.

Figure 15-12 shows a horn flared out in the electric plane with a flare angle 2ψ.

For best operation the angle ψ is usually sufficiently small that the area of the wavefront is approximately equal to the area of the aperture. The total field at any distant point is obtained by summing the contributions from the Huygen's sources distributed over the wavefront. It is permissable to assume that the field distribution over the aperture is approximately the same as it would be there if the horn did not terminate, but was infinitely long. For the case considered in Fig. 15-12, the field will be constant over the

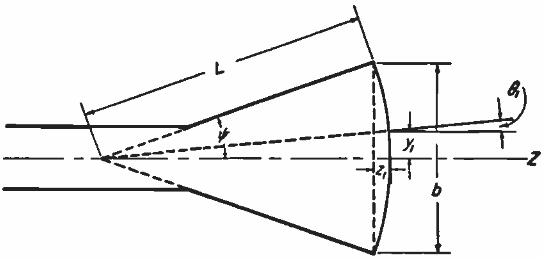


FIG. 15-12. Electromagnetic horn.

aperture in the y direction, but will vary in the x direction as $\cos \pi x_1/a$. The information of most interest will be the field intensity, and hence the gain in the forward direction, that is, along the positive z axis.

At any distant point on the z axis the field intensity due to a Huygen's source of intensity E^0H^0 will be

$$\frac{E^0\,dx\,dy}{2\lambda r}(1 + \cos\theta_1) \approx \frac{E^0\,dx\,dy}{\lambda r}$$

since $\cos\theta_1 \approx 1$ for distant points in the forward direction. The strengths of the Huygen's sources over the aperture will be given by

$$E^0 = E_y^0 \cos\frac{\pi x_1}{a}$$

The total field at a distant point on the z axis is

$$E = \frac{E_y^0}{\lambda r}\int_{-b/2}^{+b/2}\int_{-a/2}^{+a/2}\cos\frac{\pi x_1}{a}\,e^{j\beta z_1}\,dx_1\,dy_1$$

where the reference phase has been taken as that due to a source in the plane of the aperture ($z_1 = 0$). From the geometry of Fig. 15-12,

$$z_1 = L\cos\theta_1 - L\cos\psi$$
$$\approx \frac{b^2}{8L} - \frac{y_1^2}{2L}$$

Then

$$|E| = \frac{E_y^0}{\lambda r}\int_{-a/2}^{+a/2}\cos\frac{\pi x_1}{a}\,dx_1\left|\int_{-b/2}^{+b/2}e^{-j\beta\epsilon}\,dy_1\right|$$
$$= \frac{4aE_y^0}{\pi\lambda r}\left|\int_0^{b/2}e^{-j\beta\epsilon}\,dy_1\right| \qquad (15\text{-}39)$$

where
$$\epsilon = \frac{y_1^2}{2L}$$

Putting

$$\frac{\beta y_1^2}{2L} = \frac{\pi}{2}v^2$$

or
$$v = y_1\sqrt{\frac{2}{\lambda L}}, \qquad dv = dy_1\sqrt{\frac{2}{\lambda L}}$$

reduces (39) to the form of (35), and the expression for the square of the absolute magnitude of field intensity will be

$$|E|^2 = \frac{2L}{\lambda}\left(\frac{2aE_y^0}{\pi r}\right)^2\left[C^2\left(\frac{b}{\sqrt{2\lambda L}}\right) + S^2\left(\frac{b}{\sqrt{2\lambda L}}\right)\right] \qquad (15\text{-}40)$$

The effect of changes in any of the horn dimensions is made evident by using (40) and Cornu's spiral. It is seen that, if b is increased, for a given L, the forward signal will first increase to a maximum and then decrease, increasing again to secondary maximum which is smaller than the first maximum. A similar variation results if b and L are increased together keeping the horn angle constant. The explanation for this result is that as b is increased, the contributions from some of the secondary sources on the wavefronts are out of phase with others, and so tend to decrease, instead of increase, the field intensity in the forward direction.

15.08 The Infinitely Long Narrow Slit. Solutions to several problems involving radiation from or through apertures have now been obtained. In the *practical* problems, e.g., radiation from open-ended coaxial lines, wave guides, and electromagnetic horns, the solutions were approximate to the extent that the effects of any conduction currents on the outside walls were assumed to be negligible. Although in these cases the assumptions were justified by intuitive reasoning and experimental results, there are many other problems where this would not be true. In the problem of diffraction through an aperture in a thin perfectly absorbing screen, the solution was exact for the problem as stated, but the problem itself was a nonphysical one, not met with in practice. A more practical, and in general much more difficult, problem is that of diffraction

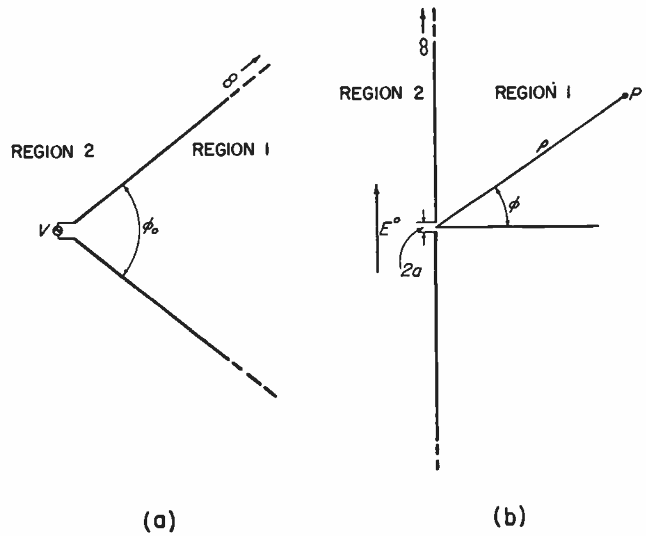


FIG. 15-13. (a) A "wedge" transmission line that supports uniform cylindrical waves. (b) Diffraction through a narrow slit in an infinite conducting plane.

through an aperture in a *conducting* screen. The simplest possible problem of this sort is the diffraction through an infinitely long and very narrow slit in a thin infinite conducting screen. Because of the simplicity of the boundary conditions and the resulting field configurations, the solution for this problem is relatively straightforward.

Figure 15-13b shows the problem to be solved. Figure 15-13a shows an apparently different problem, the solution of which applies directly to 15-13b. In the problem of 15-13b the incident wave is assumed to be directed normally to the screen with the electric vector perpendicular to the axis of the slit. In Fig. 15-13a the two semi-infinite planes form a wedge transmission line with an enclosed and ϕ_0, and with the "input edges" located a distance a from the line of intersection of the two planes. The applied voltage V is assumed to be the same at all points along the input edges, so that uniform cylindrical waves will be excited. For these waves the appropriate differential equations [taken from the set (11-28)] are

$$\frac{d(\rho E_\phi)}{\rho\, d\rho} = -j\omega\mu H_z \qquad \frac{dH_z}{d\rho} = -j\omega\epsilon E_\phi \tag{15-41}$$

where $\sigma = 0$ for the nonconducting region between the planes (region I). As in sec. (11.07) these equations combine to give a wave equation for $\mathbf{H}_z$ similar to (11-32), with solutions of the form of eq. (11-34). Retaining only the second term, which represents an outward traveling wave, the expression for $\mathbf{H}_z$ is

$$H_z = BH_0^{(2)}(\beta\rho) \tag{15-42}$$

Using the second of eqs. (41), the corresponding expression for $\mathbf{E}_\phi$ is

$$E_\phi = -j\eta BH_1^{(2)}(\beta\rho) \tag{15-43}$$

where, as usual,
$$\eta = \sqrt{\frac{u}{\epsilon}}$$

The radial wave admittance looking in the positive ρ direction in region I is

$$Y_\rho = \frac{H_z}{E_\phi} = \frac{jH_0^{(2)}(\beta\rho)}{\eta H_1^{(2)}(\beta\rho)} \tag{15-44}$$

The input voltage between the edges (at $\rho = a$) is

$$V = -\phi_0 a E_\phi(a) \tag{15-45}$$

and the input current per unit length is

$$I = J_\rho = -H_z(a) \tag{15-46}$$

where E_ϕ and H_z are evaluated at $\rho = a$. Using (43) and remembering that for $x \ll 1$,

$$H_1^{(2)}(x) \approx \frac{j2}{\pi x}$$

$$V = -\frac{2\eta\phi_0 B}{\pi\beta} \tag{15-47}$$

For the problem of Fig. 15-13b, $\phi_0 = \pi$ so that

$$B = -\frac{\beta}{2\eta} V \tag{15-48}$$

For the distant fields the expressions for **E** and **H** become

$$E_\phi = \frac{j\beta V}{2} H_1^{(2)}(\beta\rho) \underset{\rho\to\infty}{\longrightarrow} -\frac{V}{\sqrt{\rho\lambda}} e^{-j(\beta\rho-\pi/4)} \tag{15-49}$$

$$H_z = \frac{-\beta V}{2\eta} H_0^{(2)}(\beta\rho) \underset{\rho\to\infty}{\longrightarrow} -\frac{V}{\eta\sqrt{\rho\lambda}} e^{-j(\beta\rho-\pi/4)} \tag{15-50}$$

For the wedge transmission line the input admittance per meter is

$$Y_{\text{in}} = \frac{J_\rho}{V} = \frac{H_z(a)}{\phi_0 a E_\phi(a)} \tag{15-51}$$

This is the admittance presented to the generator by region I. For the slotted plane of Fig. 15-13b, $\phi_0 = \pi$, and the admittances of region I and region II are equal and in parallel. Therefore

$$Y_{\text{slit}} = 2Y_{\text{in}}$$

$$= \frac{2H_z(a)}{\pi a E_\phi(a)} = \frac{j2H_0^{(2)}(\beta a)}{\pi a \eta H_1^{(2)}(\beta a)} \tag{15-52}$$

For a very narrow slit, such as has been assumed, $\beta a \ll 1$, and (52) reduces in the same manner as did (11-44) to give

$$Y_{\text{slit}} \approx \frac{2\pi}{\eta\lambda} + \frac{j4}{\eta\lambda}\left(\ln\frac{\lambda}{\pi a} - C\right) \tag{15-53}$$

For the slit in the conducting plane the voltage V can be evaluated in terms of the incident field intensities in the following way: If there were no slit, the current per meter in the conducting plane would be $J = 2H^0$, where H^0 is the magnetic intensity of the incident wave. With the slit in the screen the *total* conduction current at the edge of the slit is zero, so the "induced" voltage V (due to charge concentrations at the edges of the slit) must be just sufficient to produce a current per meter, J_ρ, which is equal and opposite to the linear current density $J = 2H^0$. Therefore

$$V = \frac{J_\rho}{Y_{\text{slit}}} = \frac{-2H^0}{Y_{\text{slit}}} \tag{15-54}$$

Equations (49), (50), (53), and (54) constitute the solution to the problem of diffraction through a very narrow and infinitely long slit in a conducting screen.

An interesting and useful comparison can be drawn between the above problem of the long narrow slit in an infinite conducting plane and the earlier problem of the thin and infinitely long wire. Comparing eqs. (49) and (50) with those of (11-42), it is seen that except for a constant multiplying factor, the electric field diffracted by the slit has the same magnitude and direction as the *magnetic* field radiated (or reradiated) by the wire. Similarly, the magnetic field due to the slit is related in both magnitude and direction to the electric field produced by the current-carrying wire. Further, when the (external) impedance per unit length of the wire is compared with the admittance per unit length of the slit, it is observed that they are simply related by

$$Z_{\text{wire}} = \frac{\eta^2}{4} Y_{\text{slit}} \qquad (15\text{-}55)$$

or

$$Z_{\text{wire}} Z_{\text{slit}} = \frac{\eta^2}{4} = 35{,}257$$

This similarity between the properties of the wire and those of the slit are an example of a principle which is known in electromagnetics as *Babinet's principle.*

15.09 Babinet's Principle. The correspondence noted above between the magnetic and electric fields about a long narrow slit and the electric and magnetic fields about a long thin wire are just one example of a *duality principle* of electromagnetic theory. Using this principle, the solutions to certain problems can be written directly if the dual problems have been solved. In the new problems, the quantities I, V, Y, H, E correspond respectively to the quantities V, I, Z, E, and H in the dual problems. In order to see how to obtain quantitative answers using this principle, the wire and slit problems will be considered further.

If it is assumed that the long thin wire of section 11.07 is excited by an electromagnetic field E^0, H^0, which has E^0 parallel to the wire, then E^0 is the applied electric intensity E_a, and the current which flows in the wire will be

$$I = \frac{E^0}{Z_w} \qquad (15\text{-}56)$$

where Z_w is the external impedance of the wire. (A perfectly conducting wire, for which $Z_{int} = 0$, is assumed.) Using (56) and eqs. (11-42) gives for the reradiated or diffracted fields

$$E_z = \frac{-\beta\eta E^0}{4Z_w} H_0^{(2)}(\beta\rho) \tag{15-57}$$

$$H_\phi = \frac{-j\beta\eta H^0}{4Z_w} H_1^{(2)}(\beta\rho) \tag{15-58}$$

Now let the thin infinitely long wire be replaced by the narrow infinitely long slit in a conducting screen, and again let the field E^0, H^0 be incident, but with the *polarization rotated through 90 degrees* so that E^0 is *perpendicular* to the slit and H^0 is parallel to it. For this case, substituting (54) and (55) into (49) and (50) the diffracted fields of the narrow slit are found to be

$$H_z = \frac{\beta\eta H^0}{4Z_w} H_0^{(2)}(\beta\rho) \tag{15-59}$$

$$E_\phi = \frac{-j\beta\eta E^0}{4Z_w} H_1^{(2)}(\beta\rho) \tag{15-60}$$

Comparing eqs. (57) and (59) show that

$$\frac{E_z}{E^0} = -k \qquad \frac{H_z}{H^0} = +k \tag{15-61}$$

where
$$k = \frac{\beta\eta H_0^{(2)}(\beta\rho)}{4Z_w}$$

In the case of the wire, the total field E at any point is the sum of the incident field E^0 and the diffracted or reradiated field E_z; that is

$$E = E^0 + E_z$$

Then, if U_1 designates the ratio of the field intensity in the presence of the wire to the field without the wire,

$$U_1 = \frac{E}{E^0} = 1 - k \tag{15-62}$$

In the case of the infinite conducting screen with the long slit, the total field H on the right of the screen (with the incident wave approaching from the left) is just the diffracted field H_z. Denoting by U_2 the ratio of the field intensity on the right of the screen to the field without the screen,

$$U_2 = \frac{H_z}{H^0} = k \tag{15-63}$$

It is evident that the following equality holds

$$U_1 + U_2 = 1 \tag{15-64}$$

This relation, developed here for the particular case of a slit and a wire, actually is a quite general relation, true for all complementary screens and conjugate sources. *Complementary screens* are defined in the following manner: An infinite-plane conducting screen is pierced with apertures of any shape or size and the resultant screen is called S_1. Consider then the screen which is obtained by interchanging the region of metal and aperture space in S_1 and call this second screen S_2. Then screens S_1 and S_2 are said to be complementary, because, added together they result in a complete infinite metal screen. In this sense the thin wire and narrow slit may be considered as being complementary.

In deriving (64) it was necessary to interchange E and H of the incident wave. (For the plane wave considered, this was accomplished by simply rotating the plane of polarization through 90 degrees.) In general, for an arbitrary source, the effect of interchanging E and H can be obtained by replacing the distribution of electric currents and charges which constitutes the source by the corresponding distribution of magnetic currents and charges. Sources so related are called *conjugate sources.*

A generalized statement of (64) can now be stated as follows: Let a source s_1 to the left of an infinite screen S_1 produce a field on the right of S_1, and let U_1 be the ratio of this field to the field intensity that would exist there in the absence of the screen; then consider a *conjugate* source s_2 to the left of the *complementary* screen S_2, and let U_2 be the ratio of the field on the right of S_2 to the field that would exist there in the absence of the screen; then

$$U_1 + U_2 = 1 \tag{15-64}$$

Similarly, if V_1 is the ratio of field intensity at any point on the left of the screen to the field that would exist at that point if *both* screens were present (that is, if the screen were a complete infinite metal screen), and if V_2 is the ratio of the field produced by the *conjugate source* on the left of the *complementary screen* to the field that would exist there in the presence of a complete metal screen, then

$$V_1 + V_2 = 1 \tag{15-65}$$

Equations (64) and (65) along with the associated impedance relation $Z_1Z_2 = \eta^2/4$ constitute a statement of an extension* of Babinet's principle which is valid for conducting screens and radio frequencies. It should be compared with the original statement of the Babinet's principle of optics from which it derives its name. In optics, Babinet's principle relates to the transmission through apertures in *absorbing* screens, and polarization is not mentioned. The principle for optics simply states that the sum of the fields, taken separately, beyond any two complementary (absorbing) screens will add to produce the field that would exist there without any screen. It will be seen that there are important differences between this simple statement for absorbing screens in optics and the extended principle which is valid for conducting screens and polarized fields.

Application to the Half-wave Slot. One of the most straightforward applications of Babinet's principle occurs in determining the diffraction through a half-wave slot in an infinite conducting screen. For this example the complementary screen is a flat half-wave dipole of width equal to the width of the slot. For the resonant length dipole, the current induced in the dipole is independent of its cross-sectional dimensions (to a first approximation) and is given approximately by

$$I_{\infty} = \frac{V_{\infty}}{Z_a} = \frac{E^0\lambda}{\pi Z_a}$$

where $Z_a \approx 73$ ohms is the radiation resistance of the half-wave dipole. Using a spherical co-ordinate system centered at the dipole, the radiation field will be

$$E_\theta = \frac{j60I\,e^{-j\beta r}}{r}\,\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} = \frac{j60\lambda E^0\,e^{-j\beta r}}{73\pi r}\,\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$$

Therefore, by Babinet's principle, the distant diffracted field of a resonant slot in a conducting screen will be

* H. G. Booker, "Slot Aerials and Their Relation to Complementary Wire Aerials (Babinet's Principle)," *J.I.E.E.*, III A, 620–626 (1946).

$$H_\theta = -\frac{j60\lambda H^0\, e^{-j\beta r}}{73\pi r}\,\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$$

$$E\phi = -\eta H_\theta$$

In the above it has been assumed that the incident electric field was oriented parallel to the dipole, but perpendicular to the slot.

For lengths other than the resonant half-wavelength, the performance of both slot and dipole depend on the width as well as the length of the slot or (flat) dipole. However, by suitably defining a "characteristic impedance" for both of these elements, it is possible to compute at least approximately what their performance will be. It follows directly from Babinet's principle that the Q of flat dipoles and the corresponding slots in conducting planes are identical.

There may be some question about treating the thin wire and the narrow slit as complementary screens as was done earlier. The true complement of the slit in a thin conducting plane is, of course, a thin flat conducting strip, but for very narrow strips the difference between a flat strip and thin round wire becomes negligible. For slits of appreciable width, and the corresponding finite-width strips, it is necessary to define an average impedance by integrating over the width. When this integration is carried out,* it is found that the relations which constitute Babinet's principle do indeed still hold.

15.10 Slot Antennas. At very high frequencies a practical zero drag antenna for high-speed aircraft consists of a half-wave slot cut in the metal skin of the aircraft, and fed across the slot, usually at its center. Such fed slots also have interesting application in FM and television. Although the important properties of a half-wave slot in a conducting screen follow directly from Babinet's principle and the known properties of a half-wave dipole, it will be of value to consider this important case in some detail.

Figure 15-14 shows a slot in a conducting plane, and the complementary flat dipole. The slot is fed by a voltage applied between its edges at the center. The dipole is fed by a series voltage at its center.

For a first approach consider these antennas as separate bound-

* S. A. Schelkunoff, *Electromagnetic Waves*, p. 266.

ary value problems.* As such, the problem in each case is that of finding appropriate solutions to Maxwell's equations, or the derived wave equations, which will satisfy the boundary conditions. In both cases, for the free-space regions about the antennas, the wave equations to be solved are

$$\nabla^2\mathbf{E} = \mu\epsilon\ddot{\mathbf{E}} \qquad \text{or} \qquad \nabla^2\mathbf{H} = \mu\epsilon\ddot{\mathbf{H}} \tag{15-66}$$

Having obtained a solution for either **E** or **H** from one of the above equations, the other field intensity can be obtained through one of the free-space relations

$$\text{curl } \mathbf{E} = -\mu\dot{\mathbf{H}} \qquad \text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}} \tag{15-67}$$

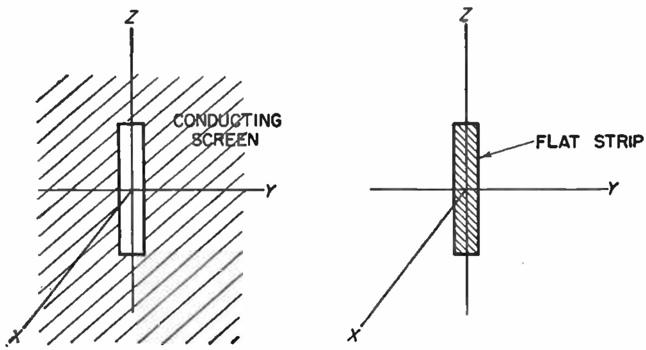


FIG. 15-14. A slot antenna in a conducting screen, and the complementary flat dipole.

For the dipole, a solution to the second of eqs. (66) would be sought subject to the following boundary conditions:

(1) In the y-z plane, and outside the perimeter of the dipole,

$$H_y = 0, \qquad H_z = 0$$

(2) In the y-z plane, and within the perimeter of the dipole,

$$H_x = 0$$

The first of these conditions follows from symmetry considerations. The second condition is a statement of the fact that the normal component of **H** must be zero at the surface of a perfect conductor.

* The treatment of the first part of this section follows a method of classroom presentation used by Professor V. H. Rumsey at the Ohio State University.

For the slot in the conducting plane the first of eqs. (66) would be solved subject to the following boundary conditions:

(1) In the y-z plane, and outside the perimeter of the slot,

$$E_y = 0, \qquad E_z = 0$$

(2) in the y-z plane, and within the perimeter of the slot

$$E_x = 0$$

The first of these conditions results from the fact that the tangential component of **E** must be zero at the surface of a perfect conductor. The second condition results from symmetry considerations.

It is apparent that, mathematically, these two problems are identical. It is necessary only to interchange **E** and **H** to pass from one problem to the other. (Of course it will have to be shown that the driving forces are also similar.) Therefore, except for a constant, the solution obtained for **E** for the slot, will be the same as the solution for **H** for the dipole, and it is possible to write for the fields at any corresponding points

$$\mathbf{E}_s = k_1\mathbf{H}_d \tag{15-68}$$

where the subscripts s and d refer to the slot and dipole respectively. Similarly, the magnetic field of the slot and the electric field of the dipole will be related by

$$\mathbf{H}_s = k_2\mathbf{E}_d \tag{15-69}$$

It follows that the pattern of **E** for the slot is the same as the pattern of **H** for the dipole, and vice versa. Also the impedance of the slot is proportional to the admittance of the dipole and vice versa. The relations between the impedance properties can be studied by use of the integral expressions for the fields near the feed points.

Strictly, for impedance to have meaning, it is necessary to consider an infinitesimal gap in each case. However, the results will apply to a finite gap if this is kept small, as indicated in Fig. 15-15. For the dipole, the impedance is given by the voltage across the gap divided by the current through the generator and into one arm of the dipole. The voltage across the gap is obtained by integrating $\mathbf{E}_d \cdot d\mathbf{s}$ along a line of **E** between two closely spaced points a and c on opposite sides of the gap.

$$V = \int_{abc} \mathbf{E}_d \cdot d\mathbf{s}$$

The current into one arm of the dipole is equal to the magnetomotive force around the (small) closed loop *efghe*

$$I = \oint \mathbf{H}_d \cdot d\mathbf{s} = 2 \int_{efg} \mathbf{H}_d \cdot d\mathbf{s}$$

The impedance of the dipole is therefore

$$Z_d = \frac{\int_{abc} \mathbf{E}_d \cdot d\mathbf{s}}{2 \int_{efg} \mathbf{H}_d \cdot d\mathbf{s}} \tag{15-70}$$

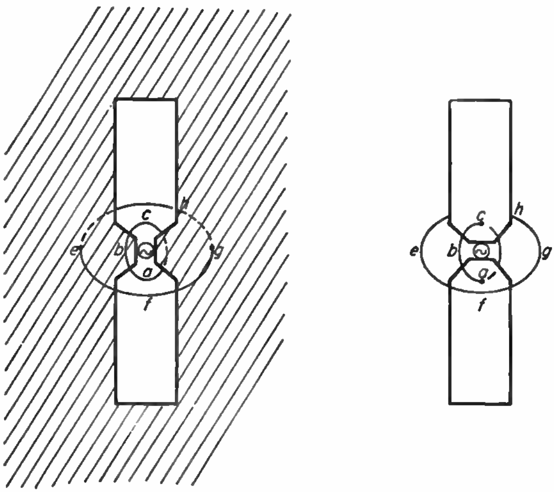


FIG. 15-15. Slot and flat-strip dipole with small gaps.

The admittance of the slot can be found by dividing the current into one edge by the voltage across the gap. The slot current is equal to the magnetomotive force around the closed path *abcda*.

$$I = \oint_{abcda} \mathbf{H}_s \cdot d\mathbf{s} = 2 \int_{abc} \mathbf{H}_s \cdot d\mathbf{s}$$

The voltage across the slot can be obtained by integrating $\mathbf{E}_s \cdot d\mathbf{s}$ along the curve *efg*.

$$V = \int_{efg} \mathbf{E}_s \cdot d\mathbf{s}$$

The slot admittance is therefore

$$Y_s = \frac{2 \int_{abc} \mathbf{H}_s \cdot d\mathbf{s}}{\int_{efg} \mathbf{E}_s \cdot d\mathbf{s}} \tag{15-71}$$

Making use of eqs. (68), (69), (70), and (71), it is seen that

$$Y_s = 4\frac{k_2}{k_1}Z_d \qquad \text{or} \qquad Z_sZ_d = \frac{k_1}{4k_2}$$

The ratio k_1/k_2 can be evaluated by considering the distant fields. At any corresponding points

$$E_s = k_1H_d \qquad H_s = k_2E_d$$

At points sufficiently distant that the fields are essentially plane wave fields

$$E_s = \eta_vH_s \qquad E_d = \eta_vH_d$$

Combining these relations shows that

$$\frac{k_1}{k_2} = \eta_v{}^2 = 377^2$$

Therefore

$$Z_sZ_d = \frac{\eta_v^2}{4} \tag{15-72}$$

For the theoretical half-wave dipole, $Z_d \approx 73 + j43$ ohms, so that for a theoretical half-wave slot,

$$Z_s \approx \frac{377^2}{4 \times (73 + j43)} \approx 418\underline{/-30.5^\circ}$$

For a resonant-length dipole, the input resistance depends upon the dipole thickness. It may be of the order of 65 ohms for practical

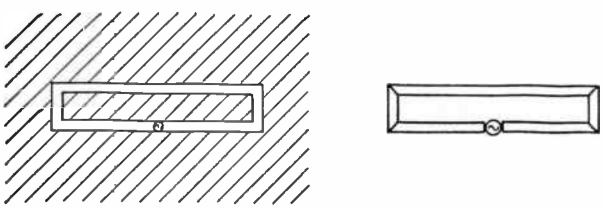


FIG. 15-16. Folded slot and folded dipole.

dipoles. The corresponding impedance for a practical resonant-length slot would be of the order of $Z_s = 550$ ohms. For a folded half-wave dipole the input impedance is roughly four times that of an ordinary dipole, so that the input impedance of the *folded half-wave slot* of Fig. 15-16 is approximately $^{550}\!/_4 = 138$ ohms.

The field about a slot in a conducting plane can also be obtained by replacing the electric field distribution that exists across the slot by its equivalent magnetic-current sheets. Elementary considerations suggest and actual measurements show that the electric intensity across a slot is distributed approximately sinusoidally along the length of the slot. The replacement of the slot in a conducting plane by magnetic currents is carried out in the following manner. Consider first conditions on one side only of the conducting plane (call this region I). A magnetic-current sheet at the surface of the plane (that now has no slot) will produce the same fields in this region as did the electric field across the slot. Since the boundary conditions at the conducting plane require zero tangential electric intensity at its surface, the magnetic-current sheet will have a positive image in the conducting plane. Because the magnetic current was assumed to be on the surface of the plane the magnetic current and its image will be almost coincident, and the only effect of the image on the field in region I is to double its value over that produced by the magnetic current alone.

The electric field across the slot also gives rise to an electromagnetic field on the back side of the plane (region II), which field could be set up by a second equivalent magnetic-current sheet at the surface of the conducting plane on the side of region II. Together with its positive image in the plane this second magnetic current will give correctly the field in region II. It should be noted that, in order to establish the correct polarity for the fields in region II with respect to those in region I (the electric field is continuous through the slot and so E has the same direction on both sides of the plane), it is necessary that the direction (or polarity) of the equivalent magnetic current in region II be opposite to that in region I. The presence of this magnetic-current sheet in region II will, of course, in no way effect the field in region I.

The use of the equivalent magnetic current to calculate the electromagnetic field due to a slot proves quite useful in determining the approximate electromagnetic field configurations about slots in conducting surfaces that are not plane. An example of this use is given in the section 15.12.

It remains to be shown that the flat-strip electric dipole of Fig. 15-15a and the slot in plane of Fig. 15-15b have corresponding methods of feed. The electric dipole is fed with an electric voltage

$V = \int_{abc} \mathbf{E}_d \cdot d\mathbf{s}$ in series with the dipole. The corresponding source for a magnetic-current sheet or magnetic dipole should be a magnetic voltage $\mathfrak{F}_1 = \int_{abc} \mathbf{H}_s \cdot d\mathbf{s}$ in series with the dipole. For the second magnetic dipole (in region II) which has its magnetic current in the opposite direction, the magnetic voltage in series with the dipole should be $\mathfrak{F}_2 = \int_{cda} \mathbf{H}_s \cdot d\mathbf{s}$. The total magnetic driving voltage required for the double-sheet magnetic dipole will be

$$\mathfrak{F}_1 + \mathfrak{F}_2 = \int_{abc} \mathbf{H}_s \cdot \mathbf{ds} + \int_{cda} \mathbf{H}_s \cdot \mathbf{ds} = \oint \mathbf{H}_s \cdot \mathbf{ds}$$

which is in fact the magnetomotive force produced by a current I through the slot generator.

15.11 The Slotted Cylinder Antennas. An antenna that has important applications at very high frequencies consists of a slot or slots cut in a conducting cylinder. For example, a longitudinal slot in a vertical cylinder produces a horizontally polarized signal suitable for FM or television. A method for obtaining the complete three-dimensional radiation pattern of a finite-length slot in a cylinder will be considered in the next section. The two-dimensional problem of an infinitely long slot in an infinite cylinder can be solved easily, and is an example that illustrates nicely a method of solution that is quite powerful for certain types of problems. The solution of this particular problem is useful because it gives the principal-plane pattern (perpendicular to the axis) of a *finite length* slot in a cylinder.

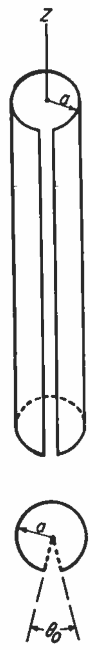


FIG. 15-17. A section of an infinitely long slotted cylinder.

Figure 15-17 shows a section of the infinitely long cylinder of radius a with a longitudinal slot of width $a\phi_0$. It is assumed that the slot is fed between its edges with a voltage $V = a\phi_0 E_0$, which is uniform in magnitude and phase along the length of the slot. This means that there will be no variations in the z direction. Also with this method of excitation there will be an E_ϕ and perhaps an E_ρ, but no E_z.

Writing Maxwell's equations in cylindrical co-ordinates for the free-space region external to the cylinder where $\sigma = 0$, and remembering that $E_z = 0$ and $\partial/\partial z = 0$, there results

$$\frac{\partial H_z}{\rho\,\partial\phi} = j\omega\epsilon E_\rho \qquad 0 = j\omega\mu H_\rho$$
$$\frac{\partial H_z}{\partial\rho} = -j\omega\epsilon E_\phi \qquad 0 = -j\omega\mu H_\phi \tag{15-73}$$
$$\frac{\partial(\rho H_\phi)}{\rho\,\partial\rho} - \frac{\partial H_\rho}{\rho\,\partial\phi} = 0 \qquad \frac{\partial(\rho E_\phi)}{\rho\,\partial\rho} - \frac{\partial E_\rho}{\rho\,\partial\phi} = -j\omega\mu H_z$$

It is seen that $H_\rho = H_\phi = 0$, and that for this problem all fields can be expressed in terms of the axial components of magnetic intensity H_z. Then

$$E_\rho = \frac{1}{j\omega\epsilon\rho}\frac{\partial H_z}{\partial\phi} \tag{15-74}$$
$$E_\phi = -\frac{1}{j\omega\epsilon}\frac{\partial H_z}{\partial\rho}$$

Substituting these expressions in the last expression of (73) gives a wave equation for H_z.

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial H_z}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 H_z}{\partial\phi^2} = -\beta^2 H_z \tag{15-75}$$

where $\beta^2 = \omega^2\mu\epsilon$

Solving in the usual manner by assuming a product solution results in

$$H_z = [A_1 H_\nu^{(1)}(\beta\rho) + B_1 H_\nu^{(2)}(\beta\rho)]e^{j\nu\phi} \tag{15-76}$$

For this problem it is apparent that ν must be an integer n. Using only the outward traveling wave for this region outside the cylinder, the general solution for this region will be

$$H_z = \sum_{n=-\infty}^{n=+\infty} b_n H_n^{(2)}(\beta\rho)\, e^{jn\phi} \tag{15-77}$$

and

$$E_\phi = \frac{j\beta}{\omega\epsilon}\sum_{n=-\infty}^{n=+\infty} b_n H_n^{(2)\prime}(\beta\rho)\, e^{jn\phi} \tag{15-78}$$

where the b_n's are coefficients which are to be evaluated by applying the boundary conditions. At $\rho = a$, the expression for E_ϕ becomes

$$E_\phi\big|_{\rho=a} = \frac{j\beta}{\omega\epsilon}\sum_{n=-\infty}^{+\infty} b_n H_n^{(2)\prime}(\beta a)\, e^{jn\phi} \tag{15-79}$$

but at $\rho = a$, the boundary conditions are

$$E_\phi = E_0 \qquad \left(-\frac{\phi_0}{2} < \phi < \frac{\phi_0}{2}\right) \tag{15-80}$$
$$E_\phi = 0 \qquad |\phi| > \frac{\phi_0}{2}$$

where the electric intensity E_0 has been assumed uniform over the gap. The field distribution at $\rho = a$, as represented by these boundary conditions is shown in Fig. 15-18.

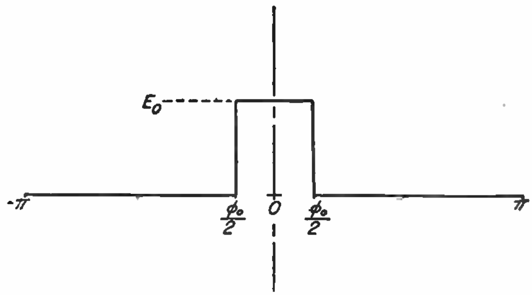


FIG. 15-18. Distribution of field intensity E_ϕ around the cylinder at $\rho = a$.

This field distribution may be resolved into a Fourier series,

$$E_\phi|_{\rho=a} = \sum_{n=-\infty}^{+\infty} C_n e^{jn\phi} \tag{15-81}$$

where

$$C_n = \frac{1}{2\pi}\int_0^{2\pi} F(\alpha)\, e^{-jn\alpha} d\alpha$$
$$= \frac{E_0}{2\pi}\int_{-\phi_0/2}^{+\phi_0/2} e^{-jn\alpha} d\alpha$$
$$= -\frac{E_0}{2jn\pi}\left(e^{-jn\phi_0/2} - e^{+jn\phi_0/2}\right)$$
$$= \frac{E_0}{n\pi}\sin\frac{n\phi_0}{2}$$

The distribution represented by (81) is the *same* as that represented by (79), so

$$\frac{E_0}{n\pi}\sin\frac{n\phi_0}{2} = \frac{j\beta}{\omega\epsilon} b_n H_n^{(2)\prime}(\beta a)$$

or

$$b_n = \frac{\omega\epsilon}{j\beta}\frac{E_0}{n\pi}\sin\frac{n\phi_0}{2}\frac{1}{H_n^{(2)\prime}(\beta a)}$$

The external field at a distance ρ is then given by (78).

$$E_\phi = \frac{E_0}{\pi} \sum_{n=-\infty}^{+\infty} \frac{\sin \frac{n\phi_0}{2} H_n^{(2)\prime}(\beta\rho)}{n H_n^{(2)\prime}(\beta a)} e^{jn\phi}$$

It is possible to evaluate this expression at large distances from the cylinder where the asymptotic expressions for the Hankel functions may be used. At large distances

$$H_n^{(2)\prime}(\beta\rho) \approx \frac{\partial}{\partial(\beta\rho)} \left[\sqrt{\frac{2}{\pi\beta\rho}}\, e^{-j\left(\beta\rho - \frac{n\pi}{2} - \frac{\pi}{4}\right)} \right]$$
$$= \left[-\frac{1}{2} \sqrt{\frac{2}{\pi(\beta\rho)^3}} - j \sqrt{\frac{2}{\pi\beta\rho}} \right] e^{-j\left(\beta\rho - \frac{n\pi}{2} - \frac{\pi}{4}\right)}$$

Neglecting the first term, at large distances

$$E_\phi \approx -\frac{jE_0}{\pi} \sqrt{\frac{2}{\pi\beta\rho}} \sum_{n=-\infty}^{+\infty} \frac{\sin \frac{n\phi_0}{2}\, e^{-j\left(\beta\rho - \frac{n\pi}{2} - \frac{\pi}{4} - n\phi\right)}}{n H_n^{(2)\prime}(\beta a)} \qquad (15\text{-}82)$$

Using only the first few terms of this expansion will usually give results of sufficient accuracy. Using less than some fixed number, say N, it is possible to write, if ϕ_0 is sufficiently small,

$$\frac{1}{n} \sin \frac{n\phi_0}{2} \approx \frac{\phi_0}{2} \qquad \text{for } |n| < N$$

Therefore, approximately,

$$E_\phi = A\, e^{-j\left(\beta\rho - \frac{\pi}{4}\right)} \sum_{n=-N}^{n=+N} \frac{e^{jn\left(\phi + \frac{\pi}{2}\right)}}{H_n^{(2)\prime}(\beta a)} \qquad (15\text{-}83)$$

where

$$A = \frac{-jE_0\phi_0}{2\pi} \sqrt{\frac{2}{\pi\beta\rho}}$$

Recalling that

$$H_{-n}^{(2)\prime}(\beta a) = (-1)^n H_n^{(2)\prime}(\beta a)$$

eq. (83) can be written

$$E_\phi = A\, e^{-j\left(\beta\rho - \frac{\pi}{4}\right)} \left[\frac{1}{H_0^{(2)\prime}(\beta a)} + 2 \sum_{n=1}^{n=N} \frac{(j)^n \cos n\phi}{H_n^{(2)\prime}(\beta a)} \right] e^{j\omega t} \qquad (15\text{-}84)$$

where the time factor has been reinserted.

This expression gives the amplitude and phase of the electric intensity at any distant point (ρ, ϕ). The relative shape of the radiation pattern is given by the absolute value of the bracketed factor.

This factor has both real and imaginary parts and may be written as $C + jD$. The field intensity pattern shape is then given by the absolute value

$$\sqrt{C^2 + D^2}$$

Figure 15-19 shows the radiation pattern calculated by this method for a $\lambda/20$-wide slot in a $5\lambda/4$-diameter cylinder. Shown for com-

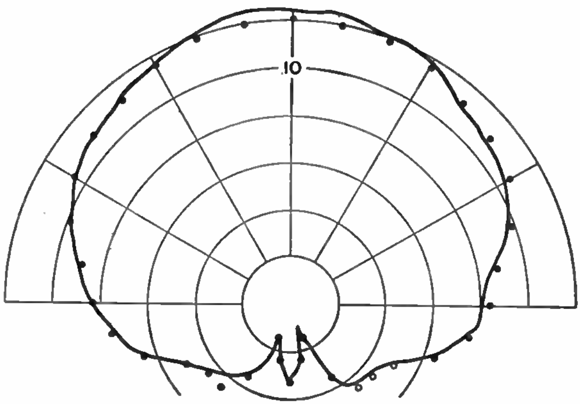


FIG. 15-19. Experimental pattern of a long, $\lambda/20$-wide, axial slot in a long cylinder of diameter $5\lambda/4$. Points are calculated for a $\lambda/20$ slot that is 1.5λ long.

parison is the measurement pattern of a $1\frac{1}{2}\lambda$-long slot of width $\lambda/20$ in a long, $5\lambda/4$-diameter cylinder. The radiation patterns of slots in cylinders of other diameters may be found in the literature.*

Transverse Slots. The same method can be used, although with less justification, to predict the radiation pattern perpendicular to the cylinder of a *transverse* slot in a cylinder. In this case the electric intensity applied across the slot will be in the z direction. For a narrow transverse slot of length $L = a\phi_0$, the electric field across the slot will be assumed to have a distribution along the

* G. Sinclair, E. C. Jordan, and E. W. Vaughan, "Measurement of Aircraft Antenna Patterns Using Models," *Proc. IRE*, **35**, 12, 1451–1462 (1947); E. C. Jordan and W. E. Miller, "Slotted-cylinder Antenna," *Electronics*, **20**, 2, 90 (1947).

length of the slot similar to that which exists on a short-circuited lossless transmission line. That is, it will be assumed that at $\rho = a$

$$E_z = E_0 \sin \beta a \left(\frac{\phi_0}{2} - \phi\right) \qquad 0 < \phi < \frac{\phi_0}{2}$$

$$E_z = E_0 \sin \beta a \left(\frac{\phi_0}{2} + \phi\right) \qquad -\frac{\phi_0}{2} < \phi < 0$$

$$E_z = 0 \qquad |\phi| > \frac{\phi_0}{2}$$

Expressing this function by the appropriate Fourier series, and equating it to an expression similar to (79) for E_z leads to the following expression for the field at any point (ρ, ϕ)

$$E_z = \frac{\beta a}{\pi} E_0 \sum_{n=-\infty}^{+\infty} \frac{\cos\left(n\frac{\phi_0}{2}\right) - \cos\left(\beta a \frac{\phi_0}{2}\right)}{H_n^{(2)}(\beta a)[(\beta a)^2 - n^2]} H_n^{(2)}(\beta\rho)\, e^{jn\phi} \qquad (15\text{-}85)$$

In Fig. 15-20 expression (85) has been evaluated for a narrow $3\lambda/4$ transverse slot in a $5\lambda/4$-diameter cylinder. Although the agreement between calculated and experimental patterns is not as close as for longitudinal slots, it is sufficiently good to predict approximate radiation patterns.

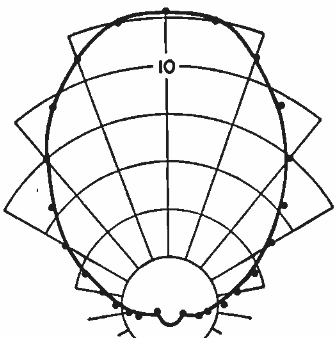


FIG. 15-20. Experimental pattern of a $3\lambda/4$-long transverse slot in a $5\lambda/4$-diameter cylinder. Points are calculated.

15.12 Dipole and Slot Arrays around Cylinders.* A class of antenna arrays of practical interest in several different fields consists of an array of vertical or horizontal dipoles or slots about a vertical conducting cylinder. In practice, the "conducting cylinder" may be an existing structure such as the spire on a tall building, or part of the superstructure on a battleship, or it may be an actual cylinder constructed as part of the antenna system. The

* The material of this section is based largely on the excellent article by P. S. Carter, "Antenna Arrays Around Cylinders," *Proc. IRE*, **31**, 12, 671–693 (1943).

mathematical problem to be solved is that of diffraction of waves of various polarizations by a conducting cylinder.

Line Source and Conducting Cylinder. Mathematically, the simplest problem of this type to solve is the two-dimensional problem of an infinitely long line source parallel to an infinitely long conducting cylinder. Consider the problem of a very long wire carrying a uniform in-phase current $I\, e^{j\omega t}$ parallel to a very long conducting cylinder (Fig. 15-21). The field due the current in the wire alone, without the cylinder, can first be obtained by integrating over the length of the wire the expression for the vector potential

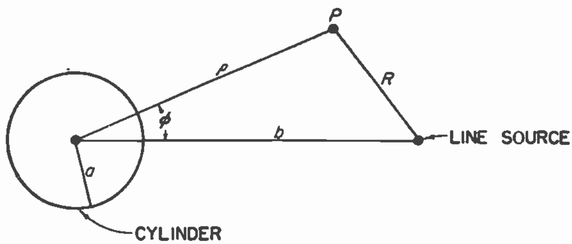


FIG. 15-21. Line source parallel to a conducting cylinder.

due to a current element. For a point P, a distance ρ from the wire (located at the origin), the vector potential will be

$$A_z = \frac{I}{4\pi} \int_{-\infty}^{+\infty} \frac{e^{-j\beta r}}{r}\, dz \tag{15-86}$$

where $r = \sqrt{\rho^2 + z^2}$ is the distance from the current element $I\, dz$ to the point P. Expression (86) can be integrated by changing the variable. Let

$$r = \rho \cosh \alpha$$

then

$$z^2 = \rho^2(\cosh^2 \alpha - 1) = \rho^2 \sinh^2 \alpha$$

$$z = \rho \sinh \alpha \qquad dz = \rho \cosh \alpha\, d\alpha = r\, d\alpha$$

$$A_z = \frac{I}{4\pi} \int_{-\infty}^{+\infty} e^{-j\beta\rho \cosh \alpha}\, d\alpha \tag{15-87}$$

This integral is a standard form* and integrates to give

$$A_z = \frac{-jI}{4} H_0^{(2)}(\beta\rho)$$

* E. Jahnke and F. Emde, *Tables of Functions*, B. G. Teubner, Leipzig, Germany, 1938, p. 218; or Dover Publications, New York, 1943, p. 150.

For the geometry of Fig. 15-21, the vector potential at P due to the line source alone will be

$$A_{z_1} = CH_0^{(2)}(\beta R)$$

where $C = -\dfrac{jI}{4}$ and $R = \sqrt{\rho^2 + b^2 - 2\rho b \cos \phi}$

The field due to the line source alone will be called the *primary wave*. The primary wave will induce currents in the conducting cylinder, and the field of these induced currents will be called the *secondary wave*. The total or resultant field at any point will be the sum of primary and secondary fields. The currents that are induced in the cylinder are of such magnitude and phase that the resultant electric intensity tangential to the (perfect) conducting cylinder is zero everywhere over the surface of the cylinder. In order to apply this boundary condition it is necessary to expand the primary wave in terms of a sum of cylindrical waves referred to the axis of the cylinder. Using the addition theorem* for Bessel functions, this expansion is given by

$$H_0^{(2)}(\beta R) = \sum_{n=0}^{\infty} \epsilon_n H_n^{(2)}(\beta b) J_n(\beta\rho) \cos n\phi \qquad \text{for } \rho < b$$

$$H_0^{(2)}(\beta R) = \sum_{n=0}^{\infty} \epsilon_n H_n^{(2)}(\beta\rho) J_n(\beta b) \cos n\phi \qquad \text{for } \rho > b$$

where ϵ_n is Neumann's number. ($\epsilon_n = 1$ for $n = 0$; $\epsilon_n = 2$ for $n \neq 0$).

The secondary waves that originate at the cylinder will be cylindrical waves, and the secondary field may be expressed as a sum of cylindrical waves originating at the axis of the cylinder. Thus for the secondary field

$$A_{z_2} = \sum_{n=0}^{\infty} \epsilon_n b_n H_n^{(2)}(\beta\rho) \cos n\phi$$

The electric intensity parallel to the surface of the cylinder is given by

$$E_z = -j\omega\mu A_z \tag{15-88}$$

* E.g., J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941, p. 372; S. A. Schelkunoff, *Electromagnetic Waves*, p. 300.

since $\partial V/\partial z = 0$ for this case. At $\rho = a$, E_z (total) and therefore A_z (total) must be zero. Therefore

$$A_z \text{ (total)} = A_{z_1} + A_{z_2}$$
$$= \sum_{n=0}^{\infty} \epsilon_n[H_n^{(2)}(\beta b)J_n(\beta a) + b_n H_n^{(2)}(\beta a)] \cos n\phi = 0$$

and therefore
$$b_n = -\frac{H_n^{(2)}(\beta b)J(\beta a)}{H_n^{(2)}(\beta a)}$$

Then at any point P, the total field will be given by

$$A_z \text{ (total)} = \sum_{n=0}^{\infty} \epsilon_n[H_n^{(2)}(\beta b)J_n(\beta \rho) + b_n H_n^{(2)}(\beta \rho)] \cos n\phi \qquad (\rho < b) \tag{15-89}$$

or

$$A_z \text{ (total)} = \sum_{n=0}^{\infty} \epsilon_n\{H_n^{(2)}(\beta \rho)[J_n(\beta b) + b_n]\} \cos n\phi \qquad (\rho > b) \tag{15-90}$$

These expressions give the field due to the infinitely long line source and conducting cylinder. The electric intensity is obtained by using (88). The magnetic intensity is given by $\mathbf{H} = \text{curl } \mathbf{A}$. The currents on the cylinder can be obtained by evaluating H at $\rho = a$.

Short Dipole near a Long Conducting Cylinder. A more practical problem than the one above is the case of a short dipole near a long conducting cylinder. If an attempt is made to solve this problem in the same manner as the preceding one, it is found that when the primary field of the dipole is expanded in cylindrical waves originating at the cylinder axis, the result is an infinite series in which each term of the series contains an infinite integral. The difficulties in evaluating such a series are very great. Fortunately, the problem can be solved by another method, described by Carter,* which makes use of the reciprocity theorem.

In Carter's method the radiation pattern of the dipole at P near a conducting cylinder (Fig. 15-22) is obtained as a receiving antenna instead of as a transmitting antenna. The wave received from a distant source will be essentially a *plane* wave that is easily expanded

* P. S. Carter, *loc. cit.*

into a sum of standing cylindrical waves. Equating the sum of primary and secondary tangential electric fields to zero at the cylinder surface gives the magnitude of the secondary or reradiated waves. The field at the dipole, and hence its open circuit voltage, V_{oc}, can then be calculated. If it is assumed that the (essentially) plane wave is produced by a current I amperes flowing in a properly oriented distant dipole at P', application of the reciprocity principle shows that a current I amperes in the dipole at P will produce an open-circuit voltage, V_{oc}, at P'. In this manner the *distant* field of the transmitting dipole near the cylinder is determined. To

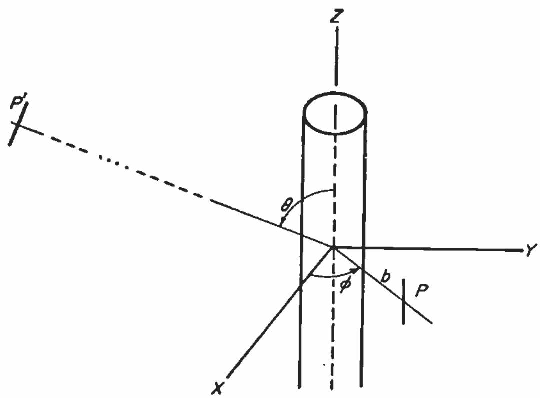


FIG. 15-22. Dipole near a conducting cylinder.

obtain the complete radiation pattern it is necessary to consider waves of both polarizations arriving at the cylinder.

Figure 15-22 shows a short vertical dipole near a very long conducting cylinder. Consider a wave, essentially plane, arriving at the cylinder from a distant dipole lying in the x-z plane, perpendicular to the radius vector, and at a polar angle θ. The magnetic field of such a wave will be horizontal and in the y direction, and can be represented by

$$\begin{aligned} H_y &= e^{j\beta(z\cos\theta + x\sin\theta)} \\ &= e^{j\beta(z\cos\theta + \rho\cos\phi\sin\theta)} \end{aligned} \tag{15-91}$$

where a wave of unit amplitude has been assumed. For this case the magnetic field will be entirely in the horizontal plane, with $H_z = 0$, so it will be possible to obtain $\mathbf{H}$ from a vector $\mathbf{A}' = \mathbf{k}A_z$

which is everywhere parallel to the z axis. The appropriate relation for H_y is

$$H_y = \text{curl}_y\, A_z = -\frac{\partial A_z}{\partial x} \tag{15-92}$$

From (92) and (91)

$$A_{z_1} = -\frac{e^{j\beta(x \sin\theta + z\cos\theta)}}{j\beta \sin\theta} = \frac{je^{j\beta(\rho\sin\theta\cos\phi + z\cos\theta)}}{\beta\sin\theta}$$

Using a standard Bessel function expansion

$$e^{j(\beta\rho\sin\theta\cos\phi)} = \sum_{n=0}^{\infty} \epsilon_n (j)^n J_n(\beta\rho\sin\theta)\cos n\phi$$

the expression for the vector potential due to the primary wave becomes

$$A_{z_1} = \frac{j\,e^{j\beta z\cos\theta}}{\beta\sin\theta}\sum_{n=0}^{\infty} \epsilon_n (j)^n J_n(\beta\rho\sin\theta)\cos n\phi \tag{15-93}$$

In this expression the primary wave has been expanded in terms of standing waves (Bessel functions of the first kind) centered at the origin. The secondary waves, produced by induced currents on the cylinder, will have the same form, but will be outward traveling waves. For such waves the J_n's in (93) will be replaced by $H_n^{(2)}$'s, so that for the secondary waves

$$A_{z_2} = \frac{j\,e^{j\beta z\cos\theta}}{\beta\sin\theta}\sum_{n=0}^{\infty} b_n\epsilon_n (j)^n H_n^{(2)}(\beta\rho\sin\theta)\cos n\phi$$

where the b_n's are arbitrary constants that must be evaluated from the boundary conditions. The total wave function is

$$A_z\ (\text{total}) = \frac{j\,c^{j\beta z\cos\theta}}{\beta\sin 0}$$

$$\sum_{n=0}^{\infty} \epsilon_n (j)^n [J_n(\beta\rho\ \text{cin}\ 0) + b_n H_n^{(2)}(\beta\rho\sin\theta)]\cos n\phi \tag{15-94}$$

Recalling that

$$\text{div}\ \mathbf{A} = -j\omega\epsilon V$$

the electric intensity can be expressed in terms of the vector potential $\mathbf{A}$ by

$$\mathbf{E} = -j\omega\mu\mathbf{A} - \text{grad } V$$
$$= -j\omega\mu\mathbf{A} - \frac{j}{\omega\epsilon}\text{grad div }\mathbf{A}$$

so that

$$E_z = -j\omega\mu A_z - \frac{j}{\omega\epsilon}\frac{\partial^2 A_z}{\partial z^2}$$
$$= \left[-j\omega\mu - \frac{j}{\omega\epsilon}(-\beta^2\cos^2\theta)\right]A_z = -j\omega\mu\sin^2\theta\, A_z$$

Since $E_z = 0$, and therefore $A_z = 0$ at $\rho = a$, it follows from (94) that

$$b_n = -\frac{J_n(\beta a \sin\theta)}{H_n^{(2)}(\beta a \sin\theta)}$$

Then the electric intensity at a vertical dipole located near the cylinder at $(b, \phi, 0)$ will be

$$E_z = \eta\sin\theta \sum_{n=0}^{\infty}\epsilon_n(j)^n\left[J_n(\beta b\sin\theta) - \frac{J_n(\beta a\sin\theta)}{H_n^{(2)}(\beta a\sin\theta)}H_n^{(2)}(\beta b\sin\theta)\right]\cos n\phi$$

By the reciprocity theorem this will also be the expression for the distant field of a vertical dipole, located a distance b from the cylinder. Therefore the relative radiation pattern for a short vertical dipole near a long vertical cylinder is

$$E_\theta = \sin\theta\sum_{n=0}^{\infty}\epsilon_n(j)^n\left[J_n(\beta b\sin\theta) - \frac{J_n(\beta a\sin\theta)}{H_n^{(2)}(\beta a\sin\theta)}H_n^{(2)}(\beta b\sin\theta)\right]\cos n\phi \qquad (15\text{-}95)$$

Expression (95) gives the radiation pattern for all values of θ. For the special case of $\theta = 90$ degrees (the horizontal pattern) it should be observed that expression (95) gives the *same* pattern as was obtained with the infinitely long wire near the cylinder. This result, obtained here for a special case, is in fact quite general. For example, the horizontal pattern of a finite-length axial slot in the vertical cylinder is independent of the length of the slot, and is the same as the pattern for an infinitely long slot.

Expression (95) also gives (exactly) the vertical pattern of a short dipole near the cylinder. This result may be used as the approximate vertical pattern of a half-wave dipole at the same location. The vertical pattern of a vertical array of dipoles may be obtained by using this pattern as the unit pattern and by applying the principle of multiplication of patterns.

The radiation patterns of horizontal and radial dipoles also may be determined by this method. In general, it is necessary to consider both polarizations for the arriving plane wave.

Application to Slots in Cylinders. By replacing the electric field distribution across a slot by its equivalent magnetic-current sheet, it is evident that the above method should have direct application in obtaining the patterns of slots in cylinders. The problem of obtaining the radiation pattern of a slot in a cylinder is now just that of determining the field patterns produced by a magnetic dipole adjacent to the cylinder. The solution carries through, just as it did for the electric dipole, except for two differences. The distant dipole in this case will be a magnetic dipole, which results in an entirely horizontal electric intensity at the cylinder, so that the fields can be expressed in terms of an electric vector potential **F** which is in the z direction. When the boundary conditions are applied at the surface of the cylinder, they cannot be applied on H_z (corresponding to the application on E_z for the electric dipole), but must be applied to E_ϕ. When the problem is worked through, keeping these facts in mind the expression obtained for H_θ due to a short axial slot in a long cylinder is

$$H_\theta = -\sin\theta \sum_{n=0}^{\infty} \epsilon_n (j)^n \left[J_n(\beta a \sin\theta) - J_n'(\beta a \sin\theta) \frac{H_n^{(2)}(\beta a \sin\theta)}{H_n^{(2)\prime}(\beta a \sin\theta)} \right] \cos n\phi \qquad (15\text{-}96)$$

which should be compared with eq. (95). In expression (96) the magnetic dipole has been allowed to approach the surface of the cylinder so that $b = a$.

Remembering that $H_n^{(2)} = J_n - jN_n$ and using the following Bessel function relation*

$$J_n N_n' - J_n' N_n = J_{n+1} N_n - J_n N_{n+1} = \frac{2}{\pi x}$$

* S. A. Schelkunoff, *Electromagnetic Waves*, p. 56, eq. 7-13.

eq. (96) reduces to

$$H_\theta = \frac{K}{\sin\theta}\sum_{n=0}^{\infty}\epsilon_n(j)^n\left[\frac{\cos n\phi}{H_n^{(2)\prime}(\beta a\sin\theta)}\right]$$

For $\theta = 90$ degrees

$$H_z = -H_\theta = -K\sum_{n=0}^{\infty}\epsilon_n(j)^n\frac{\cos n\phi}{H_n^{(2)\prime}(\beta a)} \qquad (15\text{-}97)$$

Expression (97), which is for a short axial slot in a long cylinder, gives exactly the same pattern in the horizontal plane ($\theta = 90$ degrees) as was obtained for the infinitely long slot of eq. (84).

PROBLEMS

1. A coaxial line has an inner conductor of (outer) radius $a = \frac{1}{2}$ in., and an outer conductor of (inner) radius $b = 2$ in. Determine the power radiated at 100 mc when the voltage across the open end is 1000 volts, and find the value of an equivalent resistance R that, when connected across the open end, would absorb the same amount of power as is radiated.

2. Derive the expressions (19, 20, and 21) for the electromagnetic field of a Huygen's source by direct use of equations (11 and 12).

3. Integrate the radiation fields due to all the Huygen's sources on an infinite plane to show that a plane wave results.

4. Using an "equivalent radius" $a = d/4$ for a flat-strip dipole of width d, calculate the approximate impedance of a slot in a large conducting plane at 300 mc. The slot is 35 cm long and 1 cm wide.

5. Using Carter's method, derive eq. (96) for a short slot in a cylinder, following the procedure indicated in the text.

6. Verify that eq. (96) reduces to the form shown in (97).

7. From first principles prove that the horizontal pattern of an axial slot in an infinitely long vertical cylinder is independent of the length of the slot. Hint: Use the reciprocity theorem.

8. Verify Babinet's principle (the simple version used in optics) for the case of diffraction at a straight-edge. That is, show that the vector sum of the two diffracted fields from two complementary straight-edges is equal to the free-space field.

BIBLIOGRAPHY

Schelkunoff, S. A., "Some Equivalence Theorems of Electromagnetics and their Application to Radiation Problems," *BSTJ*, **15**, 1, 92 (1936).

Booker, H. G., "Slot Aerials and Their Relation to Complementary Wire Aerials (Babinet's Principle)," *JIEE*, IIIA, 620–626 (1946).

Books

Watson, W. H., *The Physical Principles of Wave Guide Transmission and Antenna Systems*, Clarendon Press, Oxford, 1947.

Schelkunoff, S. A., *Electromagnetic Waves*, D. Van Nostrand, New York, 1943.

CHAPTER 16

GROUND WAVE PROPAGATION

The energy radiated from a transmitting antenna may reach the receiving antenna over any of several possible propagation paths, some of which are indicated in Fig. 16-1. That portion of the energy that arrives at the receiver after reflection by the ionosphere is termed the *sky wave.* Waves that are reflected at abrupt changes in the effective dielectric constant of the troposphere (that

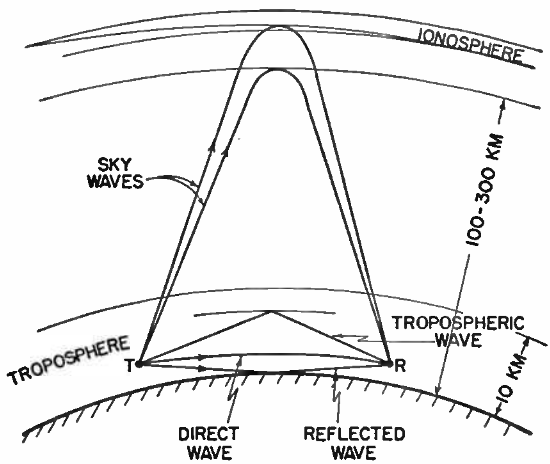


Fig. 16-1. Some possible propagation paths.

region of the atmosphere within 10 kilometers of the earth's surface) are known as *tropospheric waves.* Energy propagated over all other paths is considered to be *ground wave.* The ground wave may be divided up into a *space wave* and a *surface wave.* The space wave is made up of the *direct wave,* the signal that travels the direct path from transmitter to receiver, and the *ground-reflected wave,* which is the signal arriving at the receiver after being reflected from the surface of the earth. The space wave also includes that portion of the energy received as a result of diffraction around the earth's surface and refraction in the upper atmosphere. The surface wave is a wave that is guided along the earth's surface, much as an electromagnetic wave is guided by a transmission line. Energy is abstracted from the surface wave to supply the losses in the ground; so the attenuation of this wave is directly affected by the constants of the earth along which it travels. When both antennas are located right at the earth's surface, the direct and ground-reflected terms in the space wave cancel each other, and transmission is entirely by means of this surface wave (assuming no sky wave or tropospheric wave). The surface wave is not shown in Fig. 16-1.

The factors that affect propagation over each of these paths will be considered in detail. As a first step, the expressions for the reflection of a plane radio wave at the surface of the earth will be obtained.

16.01 Reflection at the Surface of a Finitely Conducting Plane Earth. The problem of reflection at the surface of a perfect (non-conducting) dielectric has already been solved and the reflection factors obtained for both perpendicular (horizontal) and parallel (vertical) polarizations. The earth, although not a good conductor in the sense that copper and silver are good conductors, is by no means a perfect dielectric, and its finite conductivity must be taken into account.

For a medium which has a dielectric constant ϵ and a conductivity σ, Maxwell's equation I is

$$\text{curl } \mathbf{H} = \epsilon\dot{\mathbf{E}} + \sigma\mathbf{E} \tag{16-1}$$

If the variation of **E** with time is sinusoidal, that is, if the expression for **E** at any point may be written

$$\mathbf{E} = \mathbf{E}_0\epsilon^{j\omega t} \tag{16-2}$$

Then

$$\dot{\mathbf{E}} = j\omega\mathbf{E}_0\epsilon^{j\omega t}$$
$$= j\omega\mathbf{E} \tag{16-3}$$

Putting this in eq. (1), there results

$$\text{curl } \mathbf{H} = \left(\epsilon + \frac{\sigma}{j\omega}\right)\dot{\mathbf{E}}$$
$$= \epsilon'\dot{\mathbf{E}} \tag{16-4}$$

From eq. (4) it is apparent that a partially conducting dielectric can be considered as a dielectric that has a complex dielectric constant ϵ', where

$$\epsilon' = \epsilon\left(1 + \frac{\sigma}{j\omega\epsilon}\right)$$

The wave equations and reflection coefficients derived for perfect dielectrics will apply directly to dielectrics having loss or conductance, if the dielectric constant ϵ is replaced by an equivalent complex dielectric constant

$$\epsilon' = \left(\epsilon + \frac{\sigma}{j\omega}\right)$$

Reflection Factor for Perpendicular (Horizontal) Polarization. The reflection factor R_h for a plane wave having horizontal or perpendicular polarization is obtained directly from equation (5-72). It is

$$R_h = \frac{E_r}{E_i} = \frac{\sqrt{\epsilon_v}\cos\theta - \sqrt{\left(\epsilon + \frac{\sigma}{j\omega}\right) - \epsilon_v \sin^2\theta}}{\sqrt{\epsilon_v}\cos\theta + \sqrt{\left(\epsilon + \frac{\sigma}{j\omega}\right) - \epsilon_v \sin^2\theta}} \qquad (16\text{-}5)$$

For the case of a wave incident at the surface of the earth, medium 1 is air and so ϵ_1 has been replaced by ϵ_v, the dielectric constant of free space. Also the dielectric constant ϵ_2 of the second medium has been replaced by the complex dielectric constant $[\epsilon + (\sigma/j\omega)]$. θ is the angle of incidence measured from the normal. In dealing with reflection by the earth, it is usual to express the direction of the incident wave in terms of the angle ψ which is measured from the earth's surface. That is

$$\psi = 90° - \theta$$

so that

$$\cos\theta = \sin\psi \qquad \sin\theta = \cos\psi$$

Equation (5) may then be written

$$R_h = \frac{\sin\psi - \sqrt{\left(\frac{\epsilon}{\epsilon_v} - \frac{j\sigma}{\omega\epsilon_v}\right) - \cos^2\psi}}{\sin\psi + \sqrt{\left(\frac{\epsilon}{\epsilon_v} - \frac{j}{\omega\epsilon_v}\right) - \cos^2\psi}} \qquad (16\text{-}6)$$

where (5) has been divided through by ϵ_v. It is also customary to state the earth's dielectric constant relative to that of free space by means of a relative dielectric constant ϵ_r, where

$$\epsilon_r = \frac{\epsilon}{\epsilon_v}$$

(This is the familiar dielectric constant of electrostatic units where $\epsilon_v = 1$.) The final form of the expression for the reflection factor

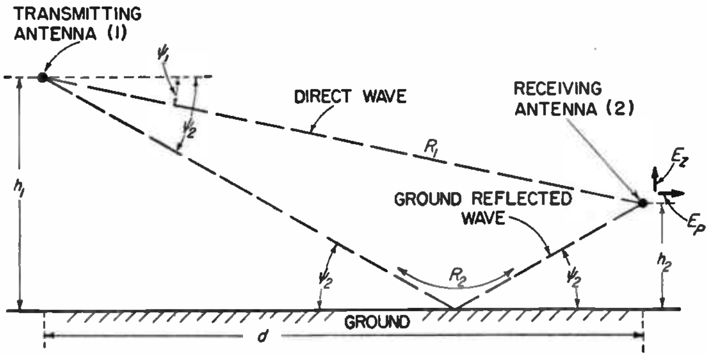


FIG. 16-2. Geometry for direct and ground-reflected waves.

for horizontal polarization is

$$R_h = \frac{\sin\psi - \sqrt{(\epsilon_r - jx) - \cos^2\psi}}{\sin\psi + \sqrt{(\epsilon_r - jx) - \cos^2\psi}} \qquad (16\text{-}7)$$

where

$$x = \frac{\sigma}{\omega\epsilon_v} = \frac{18 \times 10^9\sigma}{f} = \frac{18 \times 10^3\sigma}{f_{mc}}$$

Reflection Factor for Parallel (Vertical) Polarization. In a manner similar to the above, the reflection factor for parallel or vertical polarization is obtained from eq. (5-75). It is

$$R_v = \frac{(\epsilon_r - jx)\sin\psi - \sqrt{(\epsilon_r - jx) - \cos^2\psi}}{(\epsilon_r - jx)\sin\psi + \sqrt{(\epsilon_r - jx) - \cos^2\psi}} \qquad (16\text{-}8)$$

It is evident from eqs. (7) and (8) that the reflection factors are complex and that the reflected wave will differ both in magnitude and phase from the incident wave. The manner in which the reflection factors vary with angle of incidence is shown in Figs. 16-3 and 16-4. The various curves are for different frequencies. A study of these figures yields some interesting information. When the incident wave is horizontally polarized (Fig. 16-3), so that **E** is perpendicular to the plane of incidence and parallel to the reflecting surface, the phase of the reflected wave differs from that of the incident wave by nearly 180 degrees for all angles of incidence. For angles of incidence near grazing ($\psi = 0$), the reflected wave is equal in magnitude but 180 degrees out of phase with the incident wave

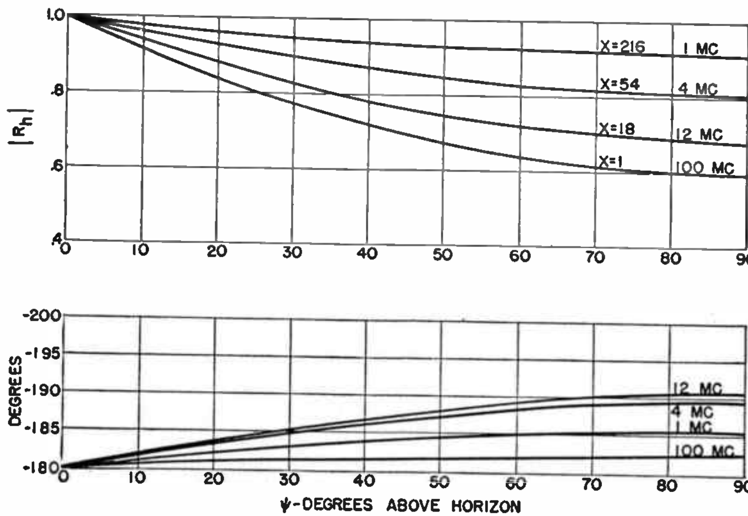


FIG. 16-3. Magnitude and phase of the plane wave reflection coefficient for horizontal polarization. The curves are for a relatively good earth ($\sigma = 12 \times 10^{-3}$, $\epsilon_r = 15$) but can be used to give approximate results for other earth conductivities and other frequencies by calculating the appropriate value of $x = 18 \times 10^3 \sigma / f_{mc}$.

for all frequencies and all ground conductivities. As the angle of incidence is increased, both the magnitude and phase of the reflection factor change, but not to any large extent. The change is greater for the higher frequencies and lower ground conductivities. The curves of Fig. 16-3 are drawn for an earth having a "good" conductivity and for a range of frequencies from 0.5 to 1000 mc. The relative dielectric constant ϵ_r varies from about 7 for a "poor" (low conductivity) earth to about 30 for a "good" (high conductivity) earth, so an average value of $\epsilon_r = 15$ has been used.

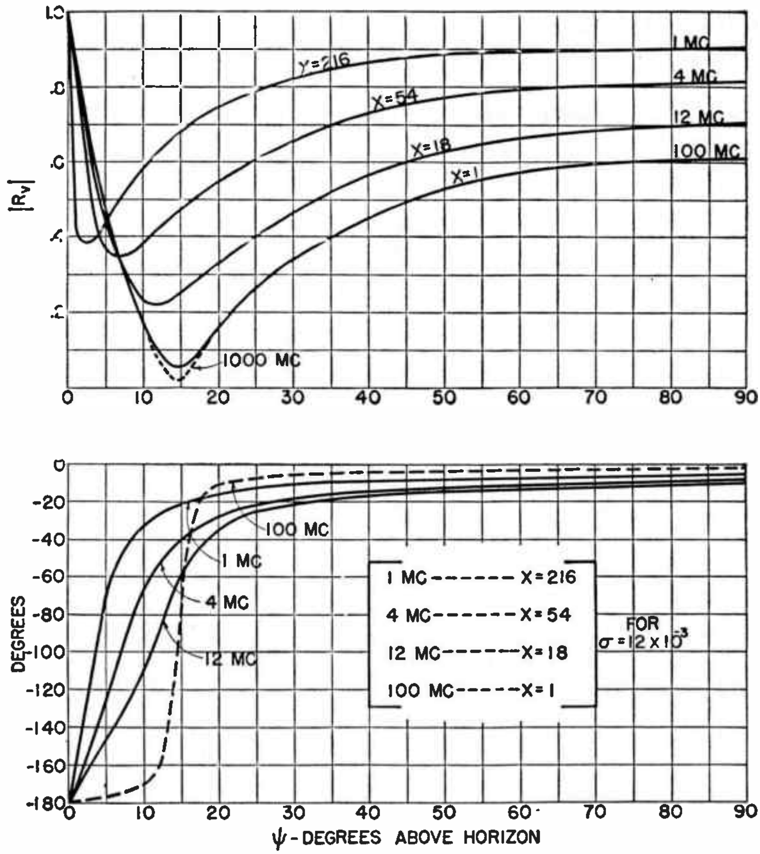
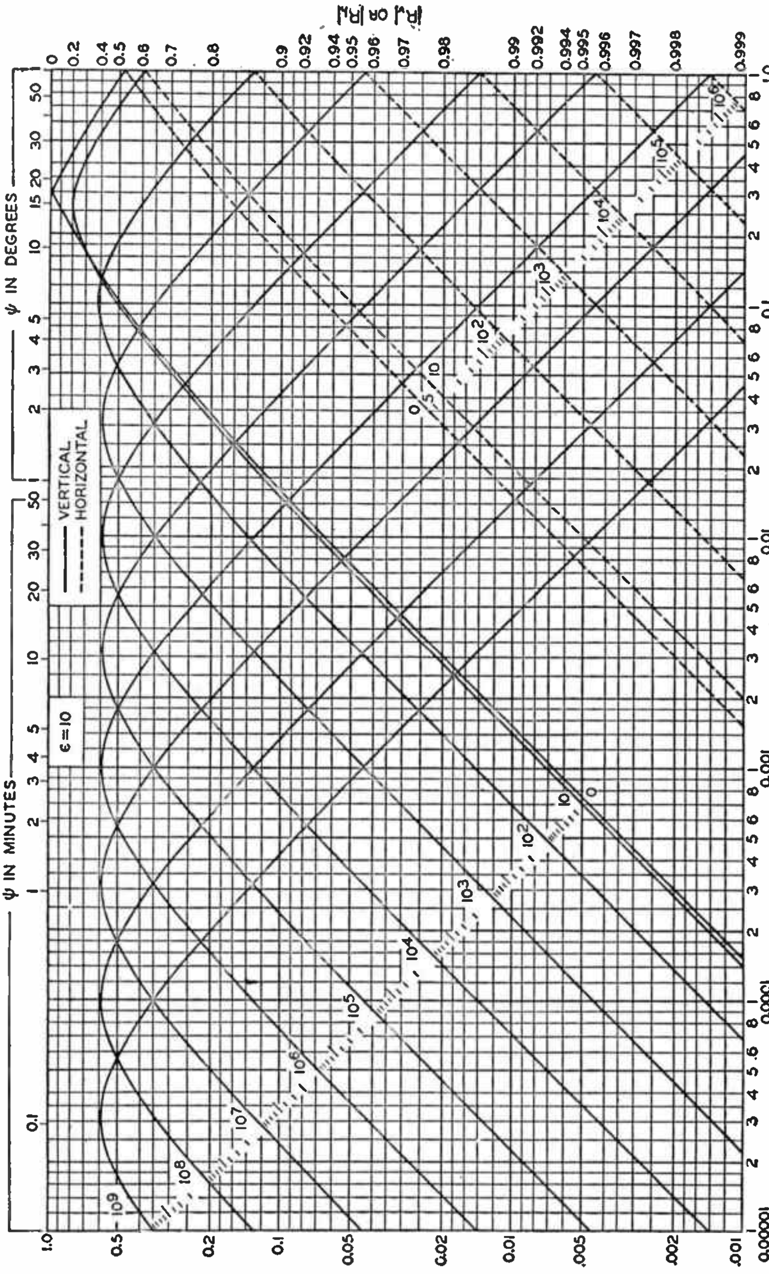


FIG. 16-4. Magnitude and phase of the plane wave reflection coefficient for vertical polarization. The curves are for a relatively good earth ($\sigma = 12 \times 10^{-3}$, $\epsilon_r = 15$) but can be used to give approximate results for other earth conductivities by calculating the appropriate value of $x = 18 \times 10^3 \sigma / f_{mc}$.

Figure 16-4 shows the manner in which the reflection factor R_v for vertical polarization varies with angle of incidence. In this case the electric vector **E** is parallel to the plane of incidence and the magnetic vector **H** is parallel to the boundary surface. The results are quite different from those obtained for horizontal polarization. As before, at grazing incidence the electric vector of the reflected wave is equal to that of the incident wave and has a


ψ IN DEGREES
ψ IN MINUTES
SIN ψ
VERTICAL
HORIZONTAL
ε = 10
|R_V| OR |R_H|
1 - |R|

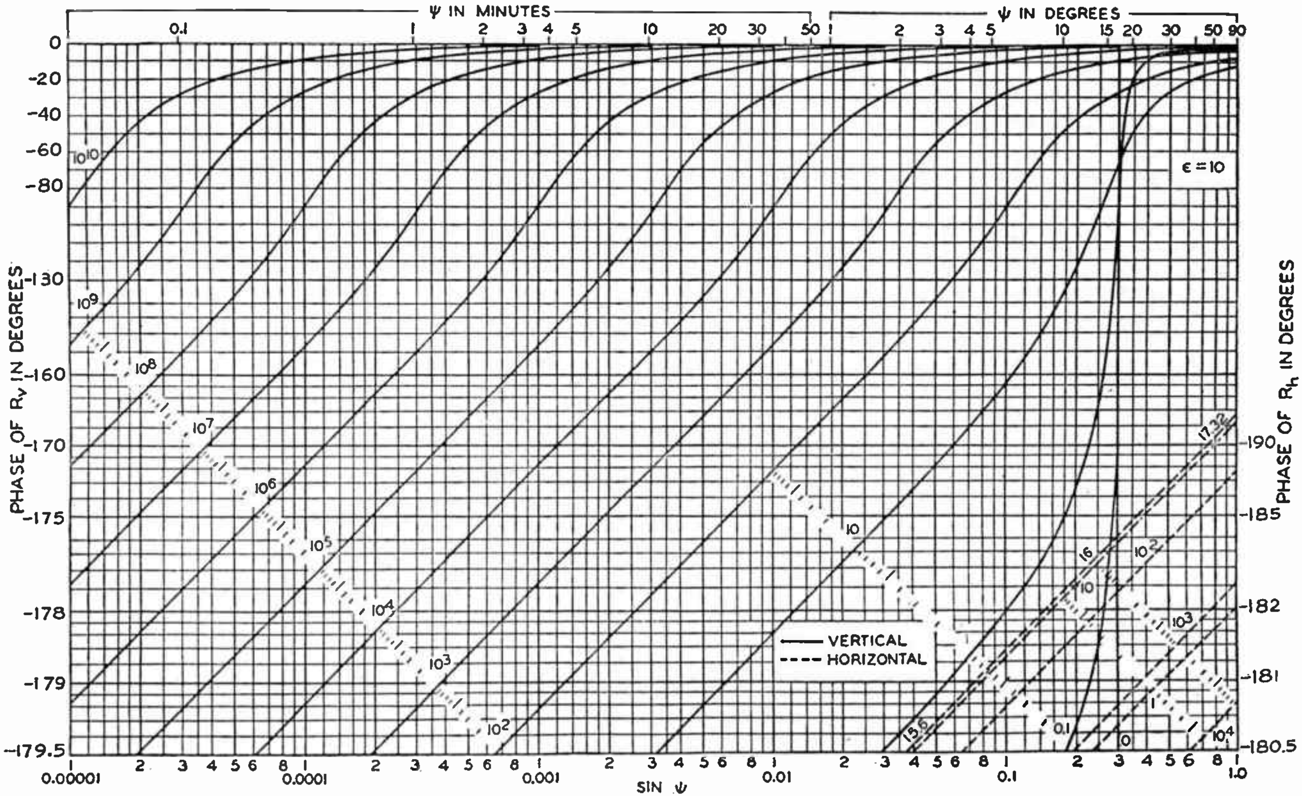


FIG. 16-5. Magnitude and phase of the reflection coefficient for $\epsilon_r = 10$. The number on each curve gives the value of the quantity $x = 18 \times 10^3\sigma/f_{mc}$. Vertical polarization is shown by solid lines, horizontal polarization by broken lines. (*Courtesy BSTJ.*)

180 degree phase reversal for all frequencies and all finite values of conductivity. However, as the angle ψ increases from zero, the magnitude and phase of the reflected wave decrease rapidly. The magnitude reaches a minimum and the phase goes through -90 degrees at an angle known as the pseudo-Brewster angle (or just Brewster angle) by analogy with the perfect dielectric case. At angles of incidence above this critical angle, the magnitude increases again and the phase approaches zero. For very high frequencies and low conductivities ($x \ll \epsilon_r$), the Brewster angle has very nearly the same value as it has for a perfect dielectric. This can be seen from eq. (8). (For $\epsilon_r = 15$, Brewster's angle occurs at $\psi = 14.5$ degrees for the perfect dielectric case.) For lower frequencies and higher conductivities the Brewster angle is less, approaching zero as x becomes much larger than ϵ_r. For a perfect conductor x is infinite and the Brewster angle occurs at $\psi = 0$ degrees.

When the incident wave is normal to the reflecting surface ($\psi = 90$ degrees), it is evident that there is no difference between horizontal and "vertical" polarization. The electric vector will be parallel to the reflecting surface in both cases and the reflection coefficients R_v and R_h should have the same values. Comparison of Figs. 16-3 and 16-4 shows that, whereas they do have the same magnitude, there is a 180 degree difference in phase. This comes about from the different definitions of positive direction for the reflected wave in the two cases and requires some explanation. For the case of reflection of a horizontally polarized wave from the surface of a perfect conductor, if the electric vector of the incident wave is in the positive x direction (Figure 5-6a), the electric vector of the reflected wave will also be in the positive x direction, but will be 180 degrees out of phase with the incident wave. This could also be interpreted as a wave in phase with the incident wave, but having its electric vector in the opposite direction. In the vertical polarization case, the positive directions for incident and reflected electric intensities are usually assumed to be as shown in Fig. 5-6b, that is, both in the positive z direction when $\psi = 0$. As ψ increases from zero, the horizontal components of both electric fields increase, but one horizontal component is positive and the other negative. At $\psi = 90$ degrees the electric intensities are wholly horizontal, but oppositely directed, one being in the positive y direction and the other in the negative y direction. From Fig. 16-4b the phase angle

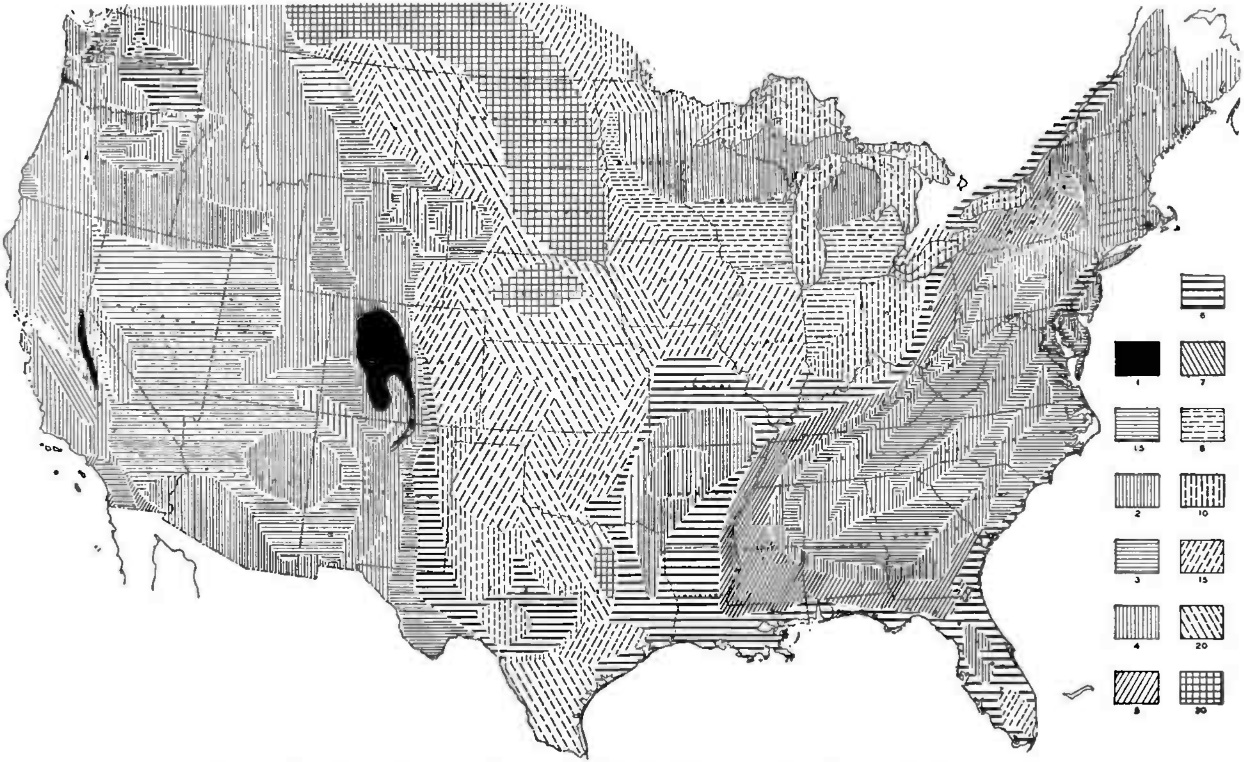


FIG. 16-6. Ground conductivity in the United States. Numbers on the legend, when multiplied by 10^{-3}, indicate ground conductivity in mhos/meter. (To obtain ground conductivity in e.m.u. multiply the numbers by 10^{-14}.) (*Map by FCC.*)

between these fields at the surface of the reflector is zero degrees (for a perfect conductor). But two vectors oppositely directed in space and having the same time phase give the same result as two vectors having the same direction and opposite phases; so this result is identical with that obtained from Fig. 16-3b at $\psi = 90$ degrees.

For angles of incidence near grazing (ψ nearly equal to zero), a more accurate plot of reflection coefficients than those given by Figs. 16-3 and 16-4 is often required. In Fig. 16-5, the magnitudes and phases of the reflection coefficients are shown* on a logarithmic scale for a relative dielectric constant $\epsilon_r = 10$.

Figure 16-6 shows how the earth's conductivity varies throughout the United States. In general, hilly or mountainous regions have low conductivity (from 10^{-3} to 5×10^{-3} mho/m) whereas the flat prairies are regions of relatively high conductivities (from 10×10^{-3} to 30×10^{-3} mho/m). The curves of Figs. 16-3 and 16-4 may be used with other conductivities than those shown on the figures by computing the appropriate values of x and interpolating between curves. For example, the curve labeled $x = 18$ corresponds to a frequency of 12 mc and a fairly good ground conductivity ($\sigma = 12 \times 10^{-3}$ mho/m). Since $x = (18 \times 10^3 \times \sigma)/f_{mc}$, it is seen that this same curve would also apply for 1 mc over an earth having a conductivity $\sigma = 1 \times 10^{-3}$ mho/m (that is a very poor earth). The curves of Fig. 16-5 are labeled directly in terms of x.

16.02 Space Wave and Surface Wave. The general problem of radiation from a vertical antenna above a plane earth having finite conductivity was originally solved by Sommerfeld† in 1909. Similar solutions have since been obtained by other writers using different attacks. All of these leave the solution in complicated forms difficult to evaluate.

* These curves are from the article by Burrows, which also shows curves for other values of ϵ_r.

C. R. Burrows, "Radio Propagation Over a Plane Earth," *Bell System Tech. J.*, **16**, 45 (1937). These curves are also shown in F. E. Terman, *Radio Engineering Handbook*, p. 700–707. In the curves of Fig. 16-5 as well as those of Figs. 16-2 and 16-3 the phase angle shown for the reflection coefficient is the angle by which the reflected wave *leads* the incident wave.

† A Sommerfeld, "The Propagation of Waves in Wireless Telegraphy," *Ann. Physik*, **28**, 665 (1909).

Norton* has reduced the complex expressions of the Sommerfeld theory to a form suitable for use in engineering work. In his original discussion, Sommerfeld stated that it was possible to divide the ground-wave field intensity into two parts, a space wave and a surface wave. The space wave predominates at large distances above the earth, whereas the surface wave is the larger near the earth's surface. As given by Norton, the expressions for the electric field of an electric dipole above the surface of a finitely conducting plane earth are in a form that clearly shows this separation into space and surface waves. At large distances from the dipole, such that the terms containing the higher orders of $1/R_1$ and $1/R_2$ may be neglected, the expressions for the vertical dipole above a finitely conducting plane earth reduce to

$$E_z = j30\beta I\,dl\left[\cos^2\psi\left(\frac{e^{-j\beta R_1}}{R_1} + R_v\frac{e^{-j\beta R_2}}{R_2}\right) + (1 - R_v)(1 - u^2 + u^4\cos^2\psi)F\,\frac{e^{-j\beta R_2}}{R_2}\right] \quad (16\text{-}9)$$

$$E_\rho = -j30\beta I\,dl\left[\sin\psi\cos\psi\left(\frac{e^{-j\beta R_1}}{R_1} + R_v\frac{e^{-j\beta R_2}}{R_2}\right) - \cos\psi(1 - R_v)u\sqrt{1 - u^2\cos^2\psi}\,F\,\frac{e^{-j\beta R_2}}{R_2}\left(1 + \sin^2\frac{\psi}{2}\right)\right] \quad (16\text{-}10)$$

In these expressions, E_z is the z component of electric field and E_ρ is the radial component (cylindrical co-ordinates, see Fig. 16-2); R_1 and R_2 are the distances from the dipole and its image, respectively, to the field point P. R_v is the plane wave reflection coefficient, the expression for which has already been developed. F is an attenuation factor that depends upon the earth's constants and upon the distance to the receiving point. It will be discussed under "surface wave." Also

$$u^2 = \frac{1}{\epsilon_r + jx}$$

where $x = \dfrac{1.8 \times 10^4\sigma \text{ mho/m}}{f_{mc}}$

σ = conductivity of the earth, mho/m

$\epsilon_r = \epsilon/\epsilon_v$ = relative dielectric constant of the earth

$\beta = 2\pi/\lambda$

* K. A. Norton, "The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere," *Proc. IRE*, **24**, 1367 (1936); *Proc. IRE*, **25**, 1203 (1937); *Proc. IRE*, **25**, 1192 (1937).

Inspection of eqs. (9) and (10) shows that the total field may be divided into two parts, a "space wave," given by the inverse-distance terms, and a "surface wave" that contains the additional attenuation factor F. Combining (9) and (10) and separating into these two types of waves, there results

$$E_{\text{total space}} = E_\psi \text{ (space)} = \sqrt{E_z^2 \text{ (space)} + E_\rho^2 \text{ (space)}}$$

$$= j30\beta I\, dl \cos\psi \left(\frac{e^{-j\beta R_1}}{R_1} + R_v \frac{e^{-j\beta R_2}}{R_2}\right) \qquad (16\text{-}11)$$

$$E_{\text{total surface}} = j30\beta I\, dl(1 - R_v)F \frac{e^{-j\beta R_2}}{R_2}$$

$$\sqrt{1 - 2u^2 + (\cos^2\psi)u^2\left(1 + \sin^2\frac{\psi}{2}\right)} \qquad (16\text{-}12)$$

In equations (11) and (12), terms involving the factor u^4 have been discarded.

The Space Wave. The expression for the space wave of a vertical dipole over a plane earth as given by eq. (11), consists of two terms. The first term $e^{-j\beta R_1}/R_1$ represents a spherical wave originating at the position of the dipole. $e^{-j\beta R_1}$ is the phase factor (the time factor $e^{j\omega t}$ has been dropped) and $1/R_1$ is the inverse-distance factor. Similarly the second term represents a spherical wave originating at the position of the image of the dipole, but in this case the magnitude and phase of the wave have been modified by the *plane* wave reflection factor R_v. Thus the space wave part of the field consists of a direct wave and a reflected wave, and the expression for the reflected wave contains the reflection factor R_v that would apply if the incident wave were plane. When the dipole is located far from the earth, the incident wave is essentially a plane wave, and, in this case, the space wave field is the total (ground wave) field. On the other hand, when the dipole is located close to the earth, the incident wave will not be plane, and the expression for the total reflected field must contain terms in addition to those given by the space wave field. These additional terms are just those which account for the surface wave.

Space Wave Patterns of a Vertical Dipole. In order to determine the effect of a finitely conducting earth upon the radiation pattern of an actual antenna, it is desirable first to investigate the radiation pattern of an elementary dipole above the earth. Expression (11) gives the space wave field of a vertical dipole located at any height above a finitely conducting earth having the reflection coefficient R_v. The expression has been evaluated and plotted as a function of frequency for a range of ground conductivities and several dipole heights (Figs. 16-7 to 16-9).

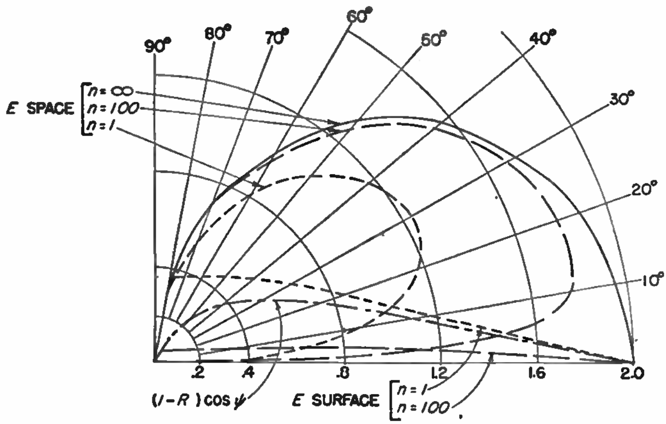


FIG. 16-7. Vertical radiation pattern of a vertical dipole at the surface of an earth having finite conductivity. The parameter $n = x/\epsilon_r$ and an average value $\epsilon_r = 15$ has been used. Both space wave and unattenuated surface wave terms are shown.

Figure 16-7 shows the vertical radiation pattern of a vertical dipole located at the surface of a finitely conducting earth. The parameter $n = x/\epsilon_r$, where as before

$$x = \frac{\sigma}{\omega\varepsilon_v} = \frac{18 \times 10^3\sigma}{f_{mc}}$$

σ is the earth conductivity in mhos per meter and f_{mc} is the frequency in megacycles. An average value of 15 has been used for ϵ_r, the relative dielectric constant of the earth. The curve $n = \infty$ represents the case of a perfectly conducting earth. $n = 100$ represents conditions at low broadcast frequencies over a good (high conductivity) earth. $n = 10$ corresponds to high broadcast frequencies over an earth of average conductivity. The curve $n = 1$ represents conditions at the medium-high frequencies. The solid curves are the space wave patterns. Shown dotted is the unattenuated surface wave curve, which will be discussed later. Figs. 16-8 and 16-9 show the vertical radiation patterns, which result when the dipole is elevated one-quarter wavelength and one-half wavelength above the earth.

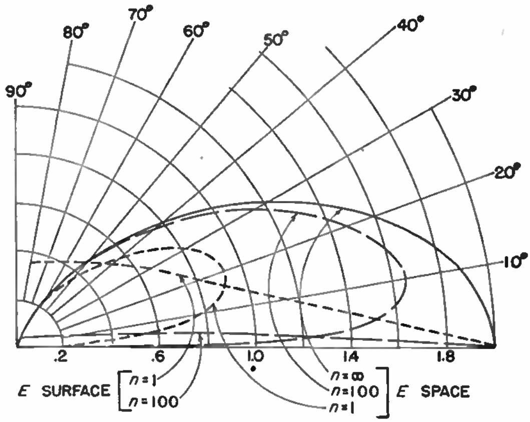


FIG. 16-8. Vertical radiation of a vertical dipole located a quarter wavelength above an earth of finite conductivity. $n = x/\epsilon_r$ and $\epsilon_r = 15$.

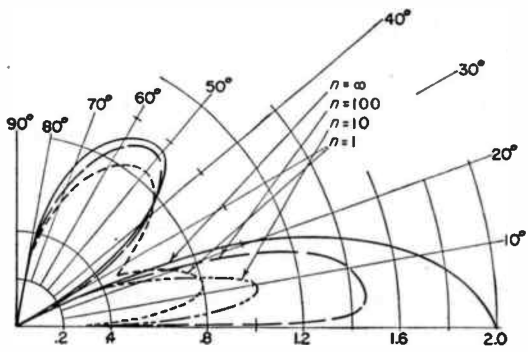


FIG. 16-9. Vertical radiation pattern of a vertical dipole one half wavelength above an earth of finite conductivity. $n = x/\epsilon_r$ and $\epsilon_r = 15$.

From these figures it is apparent that the chief effect of the finite conductivity of the earth on the vertical radiation patterns occur at the low angles where the space wave is much reduced from its value over a perfectly conducting earth. This is because of the phase of the reflection factor R_v, which changes rapidly for angles of incidence near the pseudo-Brewster angle. Above this angle the phase of R_v is nearly zero, whereas below this angle near grazing incidence the phase of R_v approaches −180 degrees. The phase of R_v is always −90 degrees at the pseudo-Brewster angle. This rapid change of phase of the reflection coefficient near the critical pseudo-Brewster angle is responsible for many of the propagation characteristics peculiar to vertical polarization.

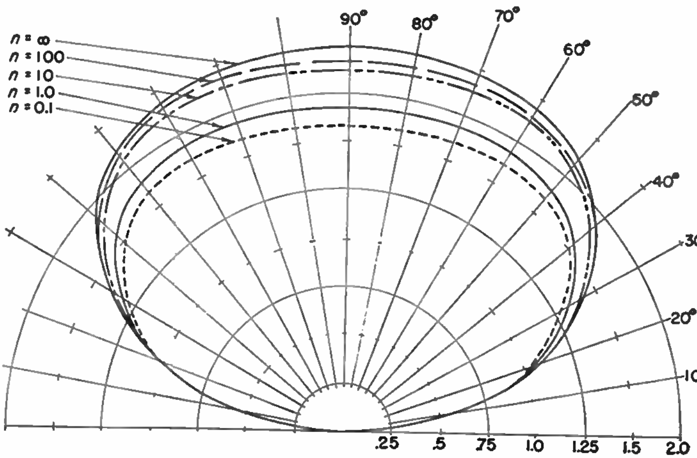


FIG. 16-10. Vertical radiation (in the plane perpendicular to the axis of the dipole) of a horizontal dipole a quarter wavelength above an earth having finite conductivity. $n = x/\epsilon_r$ and $\epsilon_r = 15$.

The patterns shown in Figs. 16-7 to 16-11 have been plotted for equal currents in the dipoles. A small radiated field, as for example in the case of $n = 1$, indicates small power radiated for a given current and, therefore, a low radiation resistance. For a given *power radiated* the dipole currents would be larger for this case ($n = 1$) and the resultant field would also be larger than shown. The relative shape of the patterns shown is the important thing; their relative size has less significance.

Space Wave Patterns for the Horizontal Dipole. The expression for the space wave field of a horizontal dipole in the plane perpendicular to the axis of the dipole is similar to that for the vertical dipole, except that R_v is replaced by R_h and the $\cos\psi$ factor is absent. It is

$$E^h_{\text{space}} = j30\beta I\, dl \left(\frac{e^{-j\beta R_1}}{R_1} + R_h \frac{e^{-j\beta R_2}}{R_2} \right)$$

The absence of the cos ψ factor is due to the fact that the horizontal dipole by itself is a uniform radiator in the plane perpendicular to its axis.

Figs. 16-10 and 16-11 show the space wave patterns of a horizontal dipole at heights of one-quarter wavelength and one-half wavelength above a finitely conducting earth. These are the

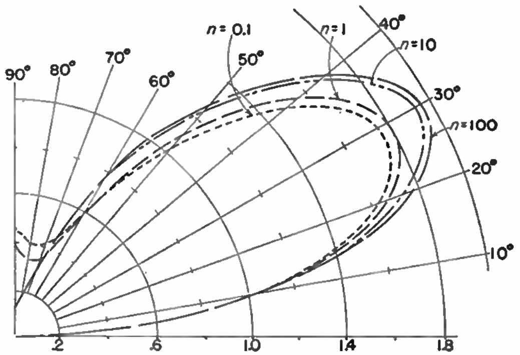


FIG. 16-11. Vertical radiation pattern (in the plane perpendicular to the axis of the dipole) of a horizontal dipole one-half wavelength above an earth having finite conductivity. $n = x/\epsilon_r$ and $\epsilon_r = 15$.

patterns in the plane perpendicular to the axis of the dipole. The effects of finite conductivity is much less marked than in the vertical dipole case because the reflection factor R_h never deviates much from the value -1, which it has for the perfect conductor case. In the plane *parallel* to the axis of the dipole, the electric field is given by the expression

$$E^h_{\text{space}} = j30\beta I\, dl \sin\psi \left(\frac{e^{-j\beta R_1}}{R_1} - R_v \frac{e^{-j\beta R_2}}{R_2} \right)$$

In this case the incident wave is polarized parallel to the plane of incidence, and the reflection factor R_v for "vertical" polarization is required. The minus sign comes about from the assumed positive directions of electric fields for the incident and reflected waves, as explained earlier in the chapter. Note that in this plane, parallel to the dipole axis, the electric field of a horizontal dipole is "vertically" polarized.

16.03 The Surface Wave. The expressions for the electric field of a vertical dipole above a finitely conducting plane earth were given in eqs. (9) and (10). When the dipole is at the surface of the earth, the expression for the surface wave part of this field reduces to

$$E_{\text{surface}} = j30\beta I\, dl(1 - R_v)F\left(\frac{e^{-j\beta R}}{R}\right)$$

$$\left[\mathbf{k}(1 - u^2) + \mathbf{r}\cos\psi\left(1 + \frac{\sin^2\psi}{2}\right)u\sqrt{1 - u^2\cos^2\psi}\right] \quad (16\text{-}13)$$

In this expression R is the distance from the dipole to the point at which the field is being considered ($R \gg \lambda$). $\mathbf{k}$ and $\mathbf{r}$ are unit vectors respectively parallel to and perpendicular to the vertical dipole. Also

$$F = [1 + j\sqrt{\pi\omega}\, e^{-\omega}\,\text{erfc}\,(-j\sqrt{\omega})]$$

$$\omega = \frac{j\beta R u^2(1 - u^2\cos^2\psi)}{2}\left[1 + \frac{\sin\psi}{u\sqrt{1 - u^2\cos^2\psi}}\right]$$

$$u^2 = \frac{1}{\epsilon_r + jx}$$

$$x = \frac{18 \times 10^3\sigma}{f_{mc}}$$

$$\text{erfc}\,(-j\sqrt{\omega}) = \frac{2}{\sqrt{\pi}}\int_{-j\sqrt{\omega}}^{\infty} e^{-v^2}\, dv$$

The function F introduces an attenuation that is dependent upon distance, frequency, and on the constants of the earth along which the wave is traveling. For distances within a few wavelengths of the dipole, F has a value of very nearly unity, and it approaches unity as the distance R approaches zero. Putting $F = 1$ in eq. (13), it is possible to evaluate and plot what is called "unattenuated surface wave." This is shown in Fig. 16-7 for two values of the parameter n. For low frequencies and good ground conductivity ($n = 100$), the unattenuated surface wave is very small, except for angles near grazing ($\psi = 0$). At $\psi = 0$, it has the value 2. At this same angle the space wave is always zero because the direct and ground-reflected waves cancel. For higher frequencies and poorer conductivity ($n = 1$), the unattenuated surface wave still has a value of 2 at $\psi = 0$, but it also has appreciable value at high angles as well. However, this wave attenuates very rapidly with distance because of the factor F.

At the surface of the earth ($\psi = 0$), the absolute value of F has been evaluated and is called the "ground wave attenuation factor." It is designated by the symbol A. That is, at $\psi = 0$

$$\begin{aligned} A &= |F| \\ &= |1 + j\sqrt{\pi\omega}\, e^{-\omega} \operatorname{erfc}(-j\sqrt{\omega})|_{\psi=0} \\ &= |1 + j\sqrt{\pi p_1}\, e^{-p_1} \operatorname{erfc}(-j\sqrt{p_1})| \end{aligned} \tag{16-14}$$

p_1 is the value of ω at the angle $\psi = 0$. In general, it is a complex quantity and may be written

$$p_1 = p\, e^{jb}$$

where p is known as the *numerical distance* and b as the *phase constant.*

Evaluating ω at $\psi = 0$ shows that

$$p = \frac{\pi R}{\lambda x} \frac{\cos^2 b''}{\cos b'} \cong \frac{\pi R}{\lambda x} \cos b$$

$$b = (2b'' - b') \cong \tan^{-1} \frac{\epsilon_r + 1}{x}$$

where
$$b'' = \tan^{-1} \frac{\epsilon_r}{x}$$

$$b' = \tan^{-1} \frac{\epsilon_r - \cos^2 \psi}{x} = \tan^{-1} \frac{\epsilon_r - 1}{x}$$

$$x = \frac{18 \times 10^3 \sigma}{f_{mc}}$$

The Surface Wave Attenuation Factor A. A plot of the ground wave attenuation factor A, as given by eq. (14), is shown in Fig. 16-12 in terms of p and b. The numerical distance p depends upon the frequency and the ground constants, as well as upon the actual distance to the transmitter. It is proportional to the distance and to the square of the frequency and varies almost inversely with the ground conductivity. The phase constant b is a measure of the power factor angle of the earth (the actual power factor angle is b''). When the earth constants and the frequency are such that $x \gg \epsilon_r$, the power factor angle will be nearly zero, and the impedance of the earth will be mainly resistive. This is the case for average or better-than-average earth at broadcast frequencies. At very high frequencies and over poor earths the condition $\epsilon_r \gg x$ may be obtained, and the earth impedance will then be reactive. It will be noticed that the same earth which acts as a conductor at very low frequencies will act as a dielectric that has a small loss at very high frequencies.

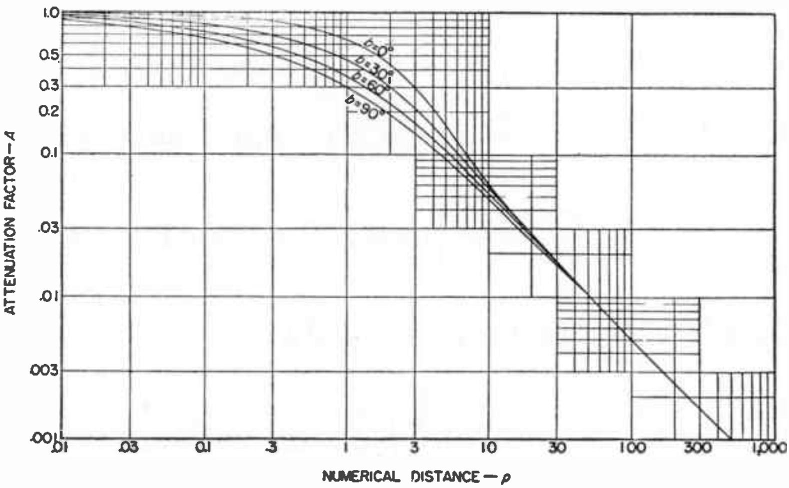


FIG. 16-12. Ground wave attenuation factor A.

The attenuation factor A can also be represented approximately by the following empirical formulas:

For $b < 5$ degrees,

$$A_1 \cong \frac{2 + 0.3p}{2 + p + 0.6p^2} \tag{16-15}$$

For all values of b,

$$A \cong A_1 - \sin b \sqrt{\frac{p}{2}}\, e^{-\frac{5}{8}p} \tag{16-16}$$

For $b < 5$ degrees and $p < 4.5$ (that is, for short numerical distances),

$$A \cong e^{-0.43p + 0.01p^2} \tag{16-17}$$

This relation shows that A varies almost exponentially with p for short numerical distance

For $b < 5$ degrees and $p \geqq 4.5$,

$$A \cong \frac{1}{2p - 3.7} \tag{16-18}$$

This relation shows that at large numerical distances A is inversely proportional to p. This means that at large numerical distances the field strength of the surfce wave will vary inversely as the *square* of the distance from the transmitter.

Surface Wave from a Horizontal Dipole. The expressions for the space and surface waves of a horizontal dipole at the surface of a finitely conducting plane earth are given by Norton as

$$E_{\text{space}} = \frac{j30\beta I\, dl\, e^{-j(\beta R - \omega t)}}{R} [\cos\phi \sin\psi(1 - R_v)\boldsymbol{\psi} + \sin\phi(1 + R_h)\boldsymbol{\phi}] \tag{16-19}$$

$$E_{\text{surface}} = \frac{j30\beta I\, dl\, e^{-j(\beta R - \omega t)}}{R} \Bigg\{ \cos\phi (u\sqrt{1 - u^2\cos^2\psi})\,(1 - R_v)F$$

$$\left[\cos\psi\left(1 + \frac{\sin^2\psi}{2}\right)\mathbf{k} + u\sqrt{1 - u^2\cos^2\psi}\right.$$

$$\left.\left(\frac{1 - \sin^4\psi - \dfrac{(1 - R_h)G}{(1 - R_v)u^2F}}{1 - u^2\cos^2}\right)\boldsymbol{\rho}\right] + \sin\phi(1 - R_h)G\boldsymbol{\phi}\Bigg\} \tag{16-20}$$

where $G = [1 + j\sqrt{\pi v}\, e^{-v} \operatorname{erfc}(-j\sqrt{v})]$

$$v = \frac{j\beta R(1 - u^2\cos^2\psi)}{2u^2}\left(1 + \frac{u\sin\psi}{\sqrt{1 - u^2\cos^2\psi}}\right)^2$$

R_h is the plane-wave reflection factor for horizontal (perpendicular) polarization. $\mathbf{k}$, $\boldsymbol{\rho}$, and $\boldsymbol{\phi}$ are unit vectors in the cylindrical co-ordinate system. The dipole lies perpendicular to $\mathbf{k}$ and in the plane $\phi = 0$. Inspection of the expressions shows that in the principle plane normal to the dipole ($\phi = 90$ degrees) the electric field is entirely in the ϕ direction, that is, it is horizontally polarized. In the direction $\phi = 0$, the electric vector lies in the plane $\phi = 0$ ("vertical" polarization). For intermediate directions the field is elliptically polarized.

The function G is an attenuation function for horizontal polarization. At large numerical distances G approaches u^4F, and since $u^2 = 1/(\epsilon_r + jx)$ is always much less than unity, it is evident that the horizontally polarized surface wave will be attenuated more rapidly than a vertically polarized wave of the same frequency.

In practical computations the attenuation of a horizontally polarized wave along the surface of the earth is determined by using the same ground wave attenuation factor A as is used for vertical polarization. However, now the numerical distance p and phase factor b are given by

$$p = \frac{\pi R}{\lambda} \frac{x}{\cos b'}$$
$$b = 180° - b'$$

where, as before,

$$x = \frac{18 \times 10^3 \sigma}{f_{mc}}$$
$$b' = \tan^{-1} \frac{\epsilon_r - 1}{x}$$

For a given actual distance R, the numerical distance p will be greater for horizontal polarization than for vertical polarization. This means greater attenuation for the horizontally polarized surface wave than for the vertically polarized wave. At low and medium frequencies, where x is large, this difference in attenuation is very great and only vertically polarized surface waves need be considered. In this frequency range the antennas used will be designed to radiate and receive vertically polarized signals. At high and very high frequencies the attenuation of the surface wave is very large for both polarizations, with the result that surface wave propogation is limited to very short distances. However, in this frequency range elevated antennas are used, and propagation paths are provided by the space wave. For this wave either vertical or horizontal polarization may be used.

16.04 Elevated Dipole Antennas above a Plane Earth. When both transmitting and receiving antennas are located at the surface of the earth, the angle ψ of the ground wave propagation path between the antennas is zero. Under these conditions the earth's reflection coefficient is -1, so the direct and ground-reflected waves cancel. Propagation is then entirely by means of the surface wave. This is the case, for example, in the daytime reception of ordinary broadcast program signals. At high and very high frequencies, however, where a wavelegth becomes sufficiently short, it is possible to elevate the antennas a quarter-wavelength or more above the ground. When the antennas are elevated, the space wave is no longer zero, and the resultant signal at the receiving antenna is the vector sum of space and surface wave.

Consider the case of two vertical antennas elevated at heights h_1 and h_2 above the surface of the earth (Fig. 16-2). From eq. (9) the vertical component of the electric field at the receiving antenna (2) due to a vertical dipole at (1) will be

$$E_z = j30\beta I\, dl \cos^2 \psi \left[\frac{e^{-j\beta R_1}}{R_1} + R_v \frac{e^{-j\beta R_2}}{R_2} + (1 - R_v)F \frac{e^{-j\beta R_2}}{R_2} \right] \quad (16\text{-}21)$$

In expression (21) u^2 and u^4 have been neglected as being small compared with unity. The first two terms of the expression constitute the space wave, and the third term is the surface wave. This expression is accurate at distances from the antenna larger than a few wavelengths. However, as it stands, it is rather involved for actual computations. Fortunately, the case of interest in practice is usually that in which the distance between antennas is very large compared with their heights above the ground, that is, for which

$$r \gg (h_1 + h_2)$$

Under these circumstances considerable simplification of the expression (21) results. The following relations will then hold approximately.

$$\cos \psi = 1$$

$$R_1 = R_2 = d \qquad \text{(for the magnitude factor in the denominator)}$$

Also for large numerical distances, the asymptotic expansion for the error function *erfc* can be used so that

$$F = 1 + j\sqrt{\pi\omega}\, e^{-\omega} \operatorname{erfc}(-j\sqrt{\omega})$$

$$= 1 - \frac{1 + \dfrac{1}{2\omega} + \dfrac{1.3}{(2\omega)^2} + \cdots}{1 + \dfrac{h_1 + h_2}{uR_2}}$$

$$\cong -\frac{1}{2\omega}$$

where $uR_2 \gg (h_1 + h_2)$ and $|\omega| > 20$.

Introducing these approximations into (21) gives

$$E_z = \frac{j30\beta I\, dl}{d}\left\{e^{-j\beta R_1} + e^{-j\beta R_2}\left[R_v - \frac{(1 - R_v)}{2\omega}\right]\right\} \tag{16-22}$$

The expression for the numerical distance ω for this case will be

$$\omega = p_1\left(1 + \frac{h_1 + h_2}{R_2 u\sqrt{1 - u^2}}\right)^2 \tag{16-23}$$

When the distance between antennas is very large compared with their height above the ground, it is evident from an inspection of expression (23) that the numerical distance ω is very nearly equal to p_1, the numerical distance along the surface of the earth. Also under the same conditions the attenuation factor F, which is approximately equal to $1/2\omega$, will not change much with height of either transmitting or receiving antenna. It will have a value approximately equal to A, the surface wave attenuation factor of Fig. 16-12. Thus, for elevated antennas, the magnitude of the surface wave will be given approximately by

$$E_{z\,\text{surface}} \approx \frac{j30\beta I\, dl}{d}\left(\frac{1 - R_v}{2p}\right) = \frac{j30\beta Id}{d}(1 - R_v)A \tag{16-24}$$

as long as the distance between antennas is very much greater than their heights above the ground.

Expression (22) has been obtained for short vertical dipoles, but it can be shown that it also holds* for elevated half-wave dipoles under the same conditions if dl is replaced by λ/π, the effective length of the half-wave dipole.

The corresponding expression for horizontal half-wave dipoles would be

$$E_\phi = \frac{j60 \sin \phi}{d}\left[e^{-j\beta R_1} + R_h\, e^{-j\beta R_2} + (1 - R_h)G\, e^{-j\beta R_2}\right] \tag{16-25}$$

At large numerical distances ($\omega > 20$), the attenuation factor G approaches u^4F, and so G is a very small quantity. The surface wave attenuation for horizontal polarization is so large that the surface wave becomes negligibly small at very short distances, and ordinarily only the space wave needs to be considered. At large

* Norton, "Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere," *Proc. IRE*, **25**. 1223 (1937).

numerical distances the factor G in expression (25) can be replaced by $A \approx 1/2p$, where p has already been defined for horizontal polarization.

EXAMPLE 1: A half-wave dipole radiator is elevated 100 ft above the ground. A receiving dipole 3 miles distant is elevated 30 ft. Determine the space and surface wave field strengths at the receiving antenna when the transmitting antenna carries a current of 1 ampere at a frequency of 50 mc. Assume an average earth having $\epsilon_r = 10$ and $\sigma = 5 \times 10^{-3}$ (a) for vertical half-wave dipoles (b) for horizontal half-wave dipoles.

CASE (a)—*Vertical half-wave dipoles.*

$$E_{sp} = \frac{j30\beta I l_{eff}}{d}(e^{-j\beta R_1} + e^{-j\beta R_2}R_v)$$

$$= \frac{j60}{d} e^{-j\beta R_1}[1 + R_v e^{-j\beta(R_2 - R_1)}]$$

$$\psi = \tan^{-1}\frac{h_1 + h_2}{r} = \tan^{-1}\frac{130}{3 \times 5280} = 0.47°$$

$$x = \frac{5 \times 10^{-3} \times 10^3 \times 18}{50} = 1.8$$

From Fig. 16-5

$$R_v = 0.94\underline{/-180°}$$

Referring to Fig. 16-2,

$$R_1 = \sqrt{d^2 + (h_1 - h_2)^2} = d\sqrt{1 + \left(\frac{h_1 - h_2}{d}\right)^2}$$

$$= d\sqrt{1.0000196} = d(1.0000098)$$

$$R_2 = \sqrt{d^2 + (h_1 + h_2)^2} = d\sqrt{1 + \left(\frac{h_1 + h_2}{d}\right)^2}$$

$$= d\sqrt{1.0000677} = d(1.0000339)$$

$$R_2 - R_1 = 3 \times 0.304 \times 5280(1.0000339 - 1.0000098)$$
$$= 0.116 \text{ meters}$$

$$\frac{360}{\lambda}(R_2 - R_1) = \frac{360}{6} \times 0.116 = 7.0°$$

$$|E_{sp}| = \frac{60}{d}|1 + 0.94\underline{/-180° - 7°}|$$

$$= \frac{60}{3 \times 1609}|1 - 0.935 + j0.113|$$

$$= \frac{60 \times 0.13}{3 \times 1609} = 1.62 \text{ mv/m}$$

$$|E_{su}| = \frac{60}{d}\left|\frac{(1 - R_v)}{2p}\right|$$

$$b \cong \tan^{-1}\frac{\epsilon_r + 1}{x} = \tan^{-1}\frac{16}{1.8} = 83.6°$$

$$p \cong \frac{\pi R}{\lambda x}\cos b = \frac{\pi \times 3 \times 1609 \times 0.112}{6 \times 1.8} = 157$$

$$|E_{su}| = \frac{60}{3 \times 1609} \times \frac{1.94}{2 \times 157} = 0.077 \text{ mv/m}$$

Case (b)—*Horizontal half-wave dipoles.* From Fig. 16-5, $R_h = 0.995 \underline{/180°}$.

$$E_{sp} = \frac{60}{d} e^{-j\beta R_1}[1 + R_h e^{-j\beta(R_2 - R_1)}]$$

$$|E_{sp}| = \frac{60}{3 \times 1609}|1 + 0.995\underline{/-180° - 7°}|$$

$$= \frac{60 \times 0.122}{3 \times 1609} = 1.52 \quad \text{mv/m}$$

$$|E_{su}| = \frac{60}{d}\left|\frac{(1 - R_h)}{2p}\right|$$

$$b' = \tan^{-1}\frac{\epsilon - 1}{x} = \tan^{-1}\frac{14}{1.8} = 82.6°$$

$$p = \frac{\pi R}{\lambda}\frac{x}{\cos b'} = \frac{\pi \times 3 \times 1609}{6} \times \frac{1.8}{0.128} = 35{,}600$$

$$E|_{su}| = \frac{60}{3 \times 1609}\frac{1.995}{2 \times 35{,}600} = 0.000349 \text{ mv/m}$$

Approximate Formula for V-H-F Propagation. The preceding example indicates that certain simplifying assumptions can be made when the elevated transmitting and receiving antennas are far apart. When these approximations are used, a quite simple formula for VHF propagation between elevated antennas results. These approximations are

(1) The surface wave can be neglected in comparison with the space wave.

(2) The angle ψ is very small so that the reflection factor R_v or $R_h \cong -1$.

Then the field at the receiving antenna due to a current I amperes in a half-wave transmitting antenna is given by

$$|E| = \frac{60I}{d}\,|1 + R_{\substack{v\\h}}\underline{/-\alpha}|$$
$$= \frac{60I}{d}\,|1 - 1\underline{/-\alpha}| \tag{16-26}$$

where α is the difference in path length between direct and reflected waves expressed in degrees. That is,

$$\alpha = \frac{2\pi}{\lambda}(R_2 - R_1) \tag{16-27}$$

Referring to Fig. 16-2,

$$R_2 \approx d\sqrt{1 + \left(\frac{h_1 + h_2}{d}\right)^2} \qquad R_1 \approx d\sqrt{1 + \left(\frac{h_1 - h_2}{d}\right)^2}$$

Using the binomial expansion, when $x \ll 1$,

$$(1 + x)^{1/2} \approx 1 + \tfrac{1}{2}x$$

Then

$$R_2 - R_1 \approx d\left[1 + \frac{1}{2}\left(\frac{h_1 + h_2}{d}\right)^2\right] - d\left[1 + \frac{1}{2}\left(\frac{h_1 - h_2}{d}\right)^2\right]$$
$$\approx \frac{2h_1h_2}{d} \tag{16-28}$$

It should be observed that, in actual computations, this *approximate* expression (28) for $R_2 - R_1$, obtained by using the first two terms of a series, will give a more accurate numerical answer than the "exact" computation, using a reasonable number of significant figures. This is because when two large and nearly equal numbers are subtracted one from the other, significant figures are lost, so that it is necessary to start with a very large number of significant figures in order to end up with only fair accuracy. In the "approximate" method one works directly on the difference between the numbers and no significant figures are lost. Then, from (27) and (28),

$$\alpha = \frac{4\pi}{\lambda}\frac{h_1h_2}{d}$$

From (26)

$$|E| = \frac{60I}{d}\,|1 - \cos\alpha + j\sin\alpha|$$

When the angle α is small, so that $\cos\alpha \cong 1$,

$$|E| \approx \frac{60I}{d}\sin\alpha \approx \frac{60I}{d}\sin\frac{4\pi h_1 h_2}{\lambda d} \tag{16-29}$$

If α is sufficiently small so that $\sin\alpha \approx \alpha$, this reduces to

$$|E| = \frac{60I\alpha}{d} = \frac{240\pi I h_1 h_2}{\lambda d^2} \qquad \text{volt/m} \tag{16-30}$$

where the approximations used are valid, the received field strength is proportional to the height of the transmitting antenna, the height of the receiving antenna, and inversely proportional to the *square* of the distance between them. In most propagation problems met with in frequency modulation and television applications, the above approximations will hold so that the simple expression of eq. (30) may be used in these important practical cases.

16.05 Wave Tilt of the Surface Wave. A vertically polarized wave at the surface of the earth will have a forward tilt, the magnitude of which depends upon the conductivity and permittivity of the earth. The slight tilt forward of the electric intensity is responsible for a small vertically downward component of the Poynting vector, sufficient to furnish the power dissipated in the earth over which the wave is passing. In general, the component of electric intensity parallel to the earth will not be in phase with the component perpendicular to it, so that the electric field just above the surface of the earth will be elliptically polarized.

In chap. 7 the problem of a wave guided along the surface of a good conductor was solved. The results obtained will apply directly to this case of a radio wave along the surface of the earth, as long as the same assumption (depth of penetration not too large a fraction of the wavelength) is valid. This will be true over most of the range of frequencies and conductivities that are of interest in surface wave propagation. Then the surface impedance of the earth is given approximately by

$$Z_s \approx \sqrt{\frac{\omega\mu}{\sqrt{\sigma^2 + \omega^2\epsilon^2}}} \Big/ \frac{1}{2}\tan^{-1}\frac{\sigma}{\omega\epsilon}$$

where σ, μ, and ϵ are respectively the conductivity, permeability, and permittivity of the earth. The horizontal component of electric intensity will be $E_h = JZ_s$ and the vertical intensity will be approximately $E_v = H\eta_v$, so that the ratio of horizontal to vertical field will be

$$\frac{E_h}{E_v} = \frac{JZ_s}{H\eta_v} = \frac{Z_s}{\eta_v}$$

$$= \frac{1}{377}\sqrt{\frac{\omega\mu}{\sqrt{\sigma^2 + \omega^2\epsilon^2}}} \Big/ \frac{1}{2}\tan^{-1}\frac{\sigma}{\omega\epsilon} \tag{16-31}$$

As an example consider a one mc radio wave at the surface of an average earth having $\sigma = 5 \times 10^{-3}$ and $\epsilon_r = 10$. For this case

$$\omega\epsilon = 5.55 \times 10^{-4}$$

$$\frac{E_h}{E_v} = 0.105\underline{/41.8^\circ}$$

The horizontal component of **E** is about one-tenth of the vertical component and leads it by an angle 41.8°. If the electric vector were plotted at various instants of time, the locus of the end point would trace out an ellipse. This elliptical polarization of the field at the earth's surface is shown in Fig. 16-13 for $\epsilon_r = 5$ and various values of x, where

$$x = \frac{\sigma}{\omega\epsilon_v} = \frac{18 \times 10^3}{f_{mc}}$$

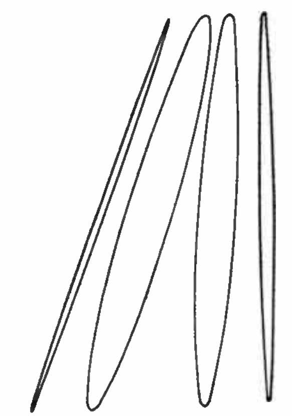


FIG. 16-13. Elliptical polarization of the electric vector at the surface of an earth for which $\epsilon_r = 5$ and for various values of $x = 18 \times 10^3/f_{mc}$.

In the example above $x = 90$.

16.06 Spherical Earth Propagation. The formulas for ground wave propagation developed in the preceding sections were obtained on the assumption of a flat or plane earth. Whereas such an assumption gives answers that are approximately correct at short distances, it cannot be expected to yield correct results at large distances. The distance up to which the curves of Fig. 16-12 can be used without serious error is given by the relation $d = 50/f^{1/3}_{mc}$ miles. Beyond this distance the actual field strength starts to deviate from that computed on the plane earth assumption. The curvature of the earth affects the propagation of the ground wave signal in several ways. First, the bulge of the earth prevents the surface wave from reaching the receiving point by a straight-line path. The surface wave, which does arrive at the receiver, reaches it by diffraction around the earth and refraction in the lower atmosphere above the earth. Secondly, for elevated antennas the space wave is affected in two different ways. The ground-reflected wave is now reflected from a curved surface, and its energy is diverged

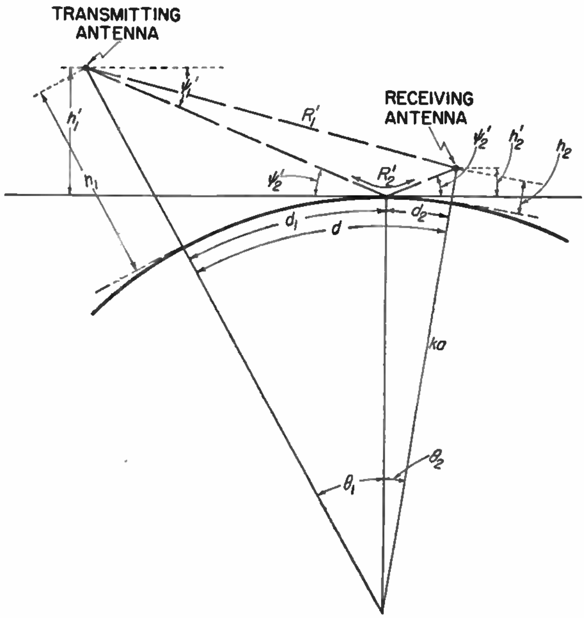


FIG. 16-14. Geometry for a spherical earth.

more than in the case when it is reflected from a flat surface. This means that the ground-reflected wave reaching the receiver will be weaker than for a flat earth by the divergence factor D, which is less than unity. Finally, for a spherical earth, the heights h_1' and h_2' of the transmitting and receiving antennas above the plane tangent to the surface of the earth at the point of reflection of the ground-reflected wave are less than the antenna heights h_1 and h_2 above the surface of the earth (Fig. 16-14).

It would seem that it should be possible to obtain an exact solution to the problem of an antenna above a spherical finitely conducting earth by solving Maxwell's equations subject to the appropriate boundary conditions. Although formal solutions to this problem have been set up, these solutions are much more involved than even the rigorous plane earth solution. For example, one such solution is in the form of an infinite series of spherical harmonics with coefficients containing twelve Bessel functions. The convergence of the series is extremely slow, the main contribution being given by those terms for which n is of the order of the ratio $2\pi R/\lambda$, where R/λ is the radius of the earth in wavelengths. For commonly used radio frequencies, this ratio is of the order of 10^3 to 10^8! It is thus apparent that a different approach must be used if numerical answers are desired. Answers of engineering accuracy can be obtained by considering separately various particular cases. The detailed analysis is complex, and in general the expressions that result are complicated. However, the results may be put in a graphical form suitable for engineering use, and this has been done by Norton.* A few of the important cases of spherical-earth ground wave propagation, which will be considered here, concern (a) the surface wave (ground-based antennas), (b) elevated antennas at medium heights, (c) optical path propagation.

(*a*) *The surface wave over a spherical earth.* Beyond the line of sight the surface wave that reaches the receiving point is due entirely to diffraction around the surface of the earth. (The effect of refraction in the troposphere is accounted for by using an effective radius for the earth that is greater than the actual radius as explained in section 16.07.) The extent of the diffraction depends upon the wavelength of the signal and the constants of the earth, and it can be determined in terms of the parameters b and k where

$$b = 2b'' - b' \qquad \text{(for vertical polarization)} \qquad (16\text{-}32)$$

$$b = 180° - b' \qquad \text{(for horizontal polarization)} \qquad (16\text{-}33)$$

$$K = \left(\frac{\lambda}{2\pi ka}\right)^{1/3}\left(\frac{x \cos b'}{\cos^2 b''}\right)^{1/3} \qquad \text{(vertical polarization)} \qquad (16\text{-}34)$$

$$K = \left(\frac{\lambda}{2\pi ka}\right)^{1/3}\left(\frac{\cos b'}{x}\right)^{1/3} \qquad \text{(horizontal polarization)} \qquad (16\text{-}35)$$

$$\tan b' = \frac{\epsilon_r - 1}{x} \qquad \tan b'' = \frac{\epsilon_r}{x} \qquad (16\text{-}36)$$

* K. A. Norton, "The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE*, **29**, 11, 623 (1941).

In the above expressions a is the radius of the earth and k is the factor by which a is multiplied to account for refraction (usually $k = 4/3$). Having determined the value of K and b in a particular problem, the curves of Figs. 16-15 and 16-16 are used to obtain the values of two other parameters β_0 and γ. Then the relative values

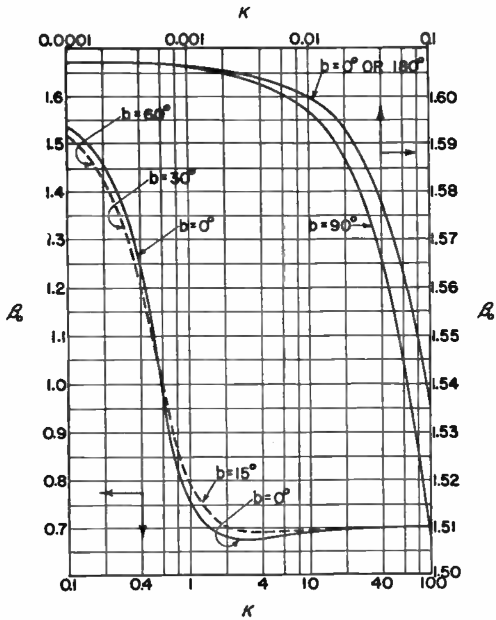


FIG. 16-15. Parameter β_0 as a function of K and b.

of field intensity at large distances over a spherical earth are given in Fig. 16-17 as a function of the parameter η', where

$$\eta' = \beta_0 \eta_0 d \tag{16-37}$$

$$\eta_0 = (k^2 a^2 \lambda)^{-1/3} \tag{16-38}$$

Figure 16-17 is used to calculate the surface wave at large distances in the following manner. First the field intensity corresponding to $\eta' = 2$ is calculated from the formula

$$E_{\text{surface }(\eta'=2)} = 2E_0 \eta_0 \gamma \tag{16-39}$$

Then the distance d corresponding to $\eta' = 2$ is obtained from

$$d_{(\eta'=2)} = \frac{2}{\beta_0 \eta_0} \tag{16-40}$$

Finally the field at any large distance is determined relative to the field at $\eta' = 2$ from Fig. 16-17.

In Fig. 16-17 it is noted that for large distances such that $\eta' > 2$, a single curve is drawn, but that for shorter distances the curve divides into two branches. The lower of these curves is valid for very large values of K (i.e., very low frequencies and good ground conductivity), whereas the upper curve applies for very small values of K (very high frequencies and poor ground conduc-

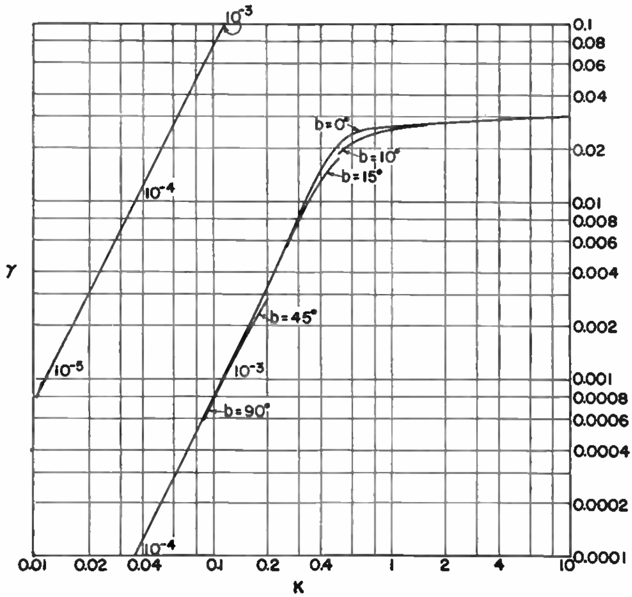


FIG. 16-16. Parameter γ as a function of K and b.

tivities). In using Fig. 16-17 it is necessary to interpolate between these curves. This is done by plotting the field at the shorter distances by means of the plane earth field intensity curves (Fig. 16-12), using the single curve of Fig. 16-17 to give the field intensities at large distances beyond $\eta' = 2$, and then joining these up with a smooth transition curve the shape of which can be estimated by keeping an eye on the curves for very large and very small values of K.

EXAMPLE 2: Determine surface wave intensity as a function of distance up to a distance of 500 miles for a 1 mc signal. Assume a fairly good earth for which $\sigma = 10 \times 10^{-3}$ and $\epsilon_r = 15$, and also assume 1 *kw* radiated from a short monopole (unattenuated field intensity at 1 mile equals 186 mv/m).

$$x = \frac{18 \times 10^3 \times \sigma}{f_{mc}} = 180 \qquad b = \tan^{-1}\frac{\epsilon_r + 1}{x} = 5.1^\circ$$

$$p = \frac{\pi R \cos b}{x\lambda} = 5.8 \times 10^{-5}R \text{ (meters)} = 0.0932R \text{ (miles)}$$

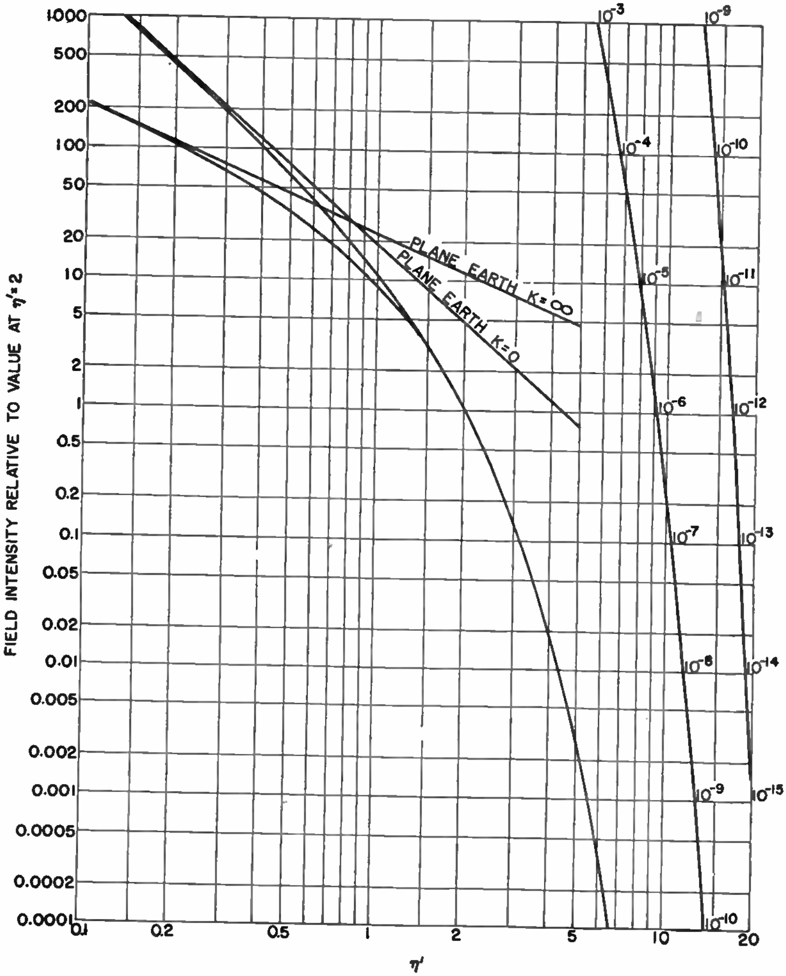


FIG. 16-17. Ground wave field intensity at large distances over a spherical earth.

(a) Plane earth calculations (using Fig. 16-12):

Miles	p	A	$E = \frac{E_0}{r} A$ (mv/m)
1	0.0932	0.93	173
10	0.932	0.63	11.7
20	1.86	0.43	4.0
100	9.32	0.065	0.12

(b) Spherical earth calculations:

$$b = 5.1°, \qquad b' = \tan^{-1} {}^{14}\!/\!_{180} = 4.46°, \qquad b'' = \tan^{-1} {}^{15}\!/\!_{180} = 4.78°$$

$$K = \left(\frac{\lambda}{2\pi ka}\right)^{\frac{1}{3}} \left(\frac{x \cos b'}{\cos^2 b''}\right)^{\frac{1}{2}}$$

$$= \left(\frac{300}{2\pi \times 5280 \times 1609}\right)^{\frac{1}{3}} \left(\frac{180 \times 0.997}{0.993}\right)^{\frac{1}{2}} = 0.239$$

From Fig. 16-15, $\beta_0 = 1.4$; from Fig. 16-16, $\gamma = .005$. Then

$$\eta_0 = [(5280 \times 1609)^2 300]^{\frac{-1}{3}} = 0.359 \times 10^{-5}$$

$$\eta' = \beta_0 \eta_0 d = 0.00808d \text{ (miles)}$$

$$E_{(\eta'=2)} = 2E_0\eta_0\gamma = 0.0107 \text{ mv/m}$$

$$d_{(\eta'=2)} = \frac{2}{\beta_0\eta_0} = 247 \text{ (miles)}$$

Problem 1. Using Fig. 16-17, sketch the field intensity *vs.* distance curve out to 500 miles for example 1.

(*b*) *Elevated antennas of medium height.* Propagation characteristics over a curved earth between elevated antennas can be obtained quite simply from the surface wave calculations as long as the antenna heights are less than $\frac{2000}{f^{2/3}_{mc}}$ ft. It is convenient to define a "numerical antenna height," q as

$$q = \frac{2\pi h}{\lambda} \left(\frac{\cos^2 b''}{x \cos b'}\right)^{\frac{1}{2}} \qquad \text{(vertical polarization)} \qquad (16\text{-}41)$$

$$q = \frac{2\pi h}{\lambda} \left(\frac{x}{\cos b'}\right)^{\frac{1}{2}} \qquad \text{(horizontal polarization)} \qquad (16\text{-}42)$$

where h/λ is the antenna height in wavelengths. Then the field intensity for antennas at numerical heights q_1 and q_2 will be given by

$$E = E_{\text{surface}} f(q_1) \cdot f(q_2) \tag{16-43}$$

where $$f(q) = \left[1 + q^2 - 2q \cos\left(\frac{\pi}{4} + \frac{b}{2}\right)\right]^{1/2}$$

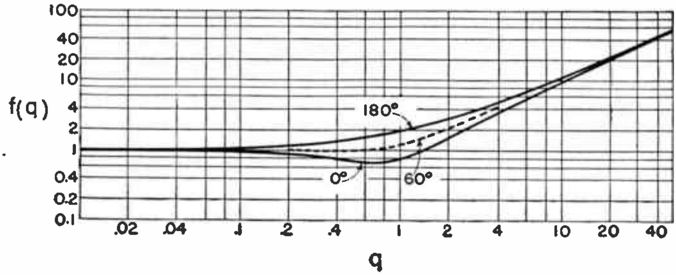


FIG. 16-18. Variation of field intensity with numerical antenna height. Within the line of sight $(q_1 + q_2) < p/100$ and $q_1q_2 < p/10$. Beyond the line of sight $q < 1/10K$.

Expression (43) can be applied whenever the antenna heights are low enough and the numerical distance p is large enough to satisfy simultaneously the following relations:

$$\left.\begin{array}{ll} h < \dfrac{2000}{f_{mc}^{2/3}}\text{ ft}; & p > 20 \\ p > 10q_1q_2; & p > 100(q_1 + q_2) \end{array}\right\} \tag{16-44}$$

This relation is particularly useful for the ultrahigh frequencies where the numerical distance p is large for distances d greater than about 1 mile. The height-gain function $f(q)$ is shown plotted in Fig. 16-18.

(c) *High antennas within the line of sight.* For high antennas well within the line of sight, the field intensity can be calculated directly from eq. (9) when this expression is suitably modified to account for the effects of a curved earth. It is easily shown by simple geometry that the distance d_0 to the horizon for an antenna of height h is given by the relation

$$d_0 \simeq \sqrt{2ah} \tag{16-45}$$

where a is the radius of the earth (3960 miles) and d, a, and h are expressed in the same units. This approximate relation holds for $h < 20{,}000$ ft. When the effect of refraction in the atmosphere is accounted for by use of an effective radius ka (see section 16.07), this formula for distance to the optical horizon becomes

$$d_0 = \sqrt{2kah} \tag{16-46}$$

Therefore the line-of-sight distance between two elevated antennas of heights h_1 and h_2 (Fig. 16-14) will be

$$d_L = \sqrt{2kah_1} + \sqrt{2kah_2}$$

For a "standard" atmosphere $k = 1.33$ and $ka = 5280$ miles, so that d_L reduces to

$$d_L = \sqrt{2h_1} + \sqrt{2h_2} \tag{16-47}$$

where h_1 and h_2 are now in feet and d_L is in miles.

For line of sight propagation the curvature of the earth will have two principal effects. First, the ground-reflected wave, being reflected from a curved surface, will have its energy diverged more than in the plane earth case. This can be accounted for by multiplying the ground-reflected wave by a divergence factor D, which is given by

$$D = \left(1 + \frac{2d_1d_2}{kad \tan \psi_2'}\right)^{1/2} \tag{16-48}$$

Second, the heights h_1' and h_2' of the antennas above the plane that is tangent to the earth at the point of reflection are less than the actual heights h_1 and h_2 above the earth. By geometry, the relations between these heights are

$$h_1' = h_1 - \frac{d_1^2}{2ka}; \qquad h_2' = h_2 - \frac{d_2^2}{2ka}$$

$$\tan \psi_2' = \frac{h_1^1 + h_2^1}{d} = \frac{h_1^1}{d_1} = \frac{h_2^1}{d_2}. \tag{16-49}$$

The modified form of expression (9) suitable for a curved earth is then

$$E_z = \frac{E_0}{d}[\cos^3 \psi_1' e^{-j\beta R_1'} + DR_{\text{refl}} \cos^3 \psi_2' e^{-j\beta R_2'} + (1 - R_{\text{refl}})$$
$$F \cos^2 \psi_2' e^{-j\beta R_2'}] \tag{16-50}$$

where R_1 has been replaced by $d/\cos\psi_1'$ and R_2 by $d/\cos\psi_2'$ in the denominator. E_0 is the free-space field intensity at unit distance and $R_{\text{refl}} = R_v$ for vertical polarization and $R_{\text{refl}} = R_h$ for horizontal polarization. The use of this formula is restricted to points well within the line of sight and to distances less than that, which makes

$$\tan\psi_2' = (\lambda/2\pi ka)^{1/3}$$

When the surface wave can be neglected, as is usually the case for high antennas, and assuming that the antennas are far enough apart that $\tan\psi_2' < 0.1$, expression (50) reduces to

$$E = \frac{E_0}{d}\left(1 + DR_{\text{refl}}\, e^{-j2\beta h_1' h_2'/d}\right) \tag{16-51}$$

where $(\lambda/2\pi ka)^{1/3} < \tan\psi_2' < 0.1$. It is seen that the received field oscillates with distance between the values

$$E = \frac{E_0}{d}\left(1 + D|R_{\text{refl}}|\right)$$

and

$$E = \frac{E_0}{d}\left(1 - D|R_{\text{refl}}|\right) \tag{16-52}$$

The above three examples of ground wave propagation above a spherical earth cover the majority, but by no means all of the cases that will be met with in practice. In general, the computations for other cases will be somewhat more complex than those above, and will involve the use of graphs and charts not shown here. For examples of such computations reference should be made to the article by Norton. Typical results for the important practical case of high antennas beyond the line of sight at FM and television frequencies are shown in the curves of Figs. 16-19 and 16-20.

16.07 Tropospheric Refraction and Reflection. For frequencies above about 50 mc the ionospheric reflection indicated in Fig. 16-1 does not occur, and propagation paths would appear to be limited to the ground wave transmission discussed in the preceding sections. As already has been seen, for these ground wave propagation paths at the very high and ultrahigh frequencies, the field intensity decreases very rapidly for distances beyond the line of sight, so that at these frequencies, useful transmission much beyond the horizon would not be expected. Nevertheless, it is found that under certain meteorological conditions useful transmission considerably beyond the line of sight does occur. Such transmission is due to refraction and reflection of the radio waves in the *troposphere*. The troposphere is that region of the earth's atmosphere immediately adjacent to the earth and extending upwards about 10 kilometers. In the troposphere the temperature decreases with height at about the rate of 6.5° C per kilometer to a value of about −50° C at its upper boundary. Above the troposphere is the stratosphere where the

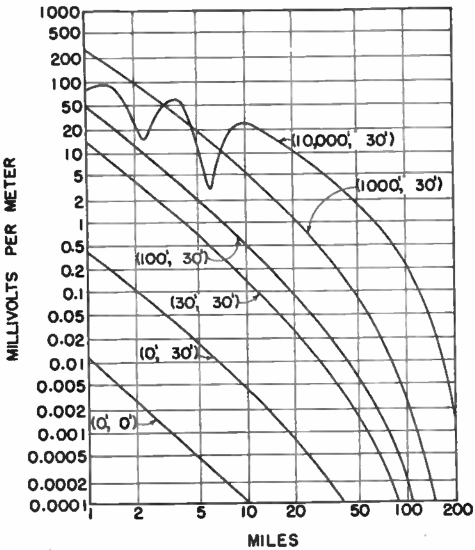


FIG. 16-19. Ground wave field intensity vs. distance for 1 kw radiated at a frequency of 50 mc. Polarization is horizontal and the labeling on the curves indicates transmitting and receiving antenna heights expressed in feet. Earth constants are $\sigma = 5 \times 10^{-3}$, $\epsilon_r = 15$.

temperature remains constant at about −50° C. Tropospheric effects on radio waves may be divided into refraction and reflection. *Tropospheric refraction* is a gradual bending of the rays that occurs because of the changing effective dielectric constant of the atmosphere through which the wave is passing. As will be shown, normal refraction effects can be accounted for by using an "effective" value for the radius of the earth that differs from the actual value. Thus the effects of tropospheric refraction are automatically included in the ground wave computations. *Tropospheric reflec-*

tions occur at a place where there are *abrupt* changes in the dielectric constant of the atmosphere. Such tropospheric reflected waves often result in useful signals at distances much greater than line of sight, a fact of considerable importance in frequency modulation and television reception.

Tropospheric Refraction. A radio wave traveling horizontally in the earth's atmosphere follows a path which has a slight downward

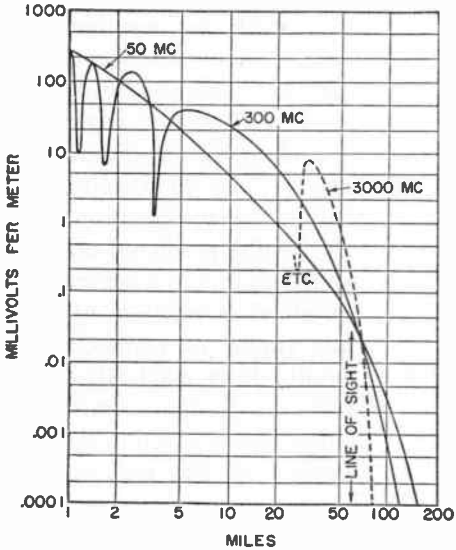


FIG. 16-20. Ground wave field intensity vs. distance for 1 kw radiated at the frequencies indicated. Polarization is horizontal and the antenna heights are 1000 feet and 30 feet. The earth constants are $\sigma = 5 \times 10^{-3}$, $\epsilon_r = 15$.

curvature due to refraction of the wave in the atmosphere. This curvature of the path tends to overcome partially the loss of signal due to curvature of the earth and permits the direct ray to reach points slightly beyond the horizon as determined by the straight-line path. In making computations the effect of refraction is accounted for by using an effective radius of curvature for the earth that is somewhat larger than the actual radius, and then assuming straight-line paths (that is, no refraction) in the atmosphere.

The refraction of a radio wave in the atmosphere occurs because the dielectric constant, and hence the refractive index of the atmosphere, varies with height above the earth. The dielectric constant of dry air is slightly greater than the value of unity that applies for a vacuum, and the presence of water vapor increases the dielectric constant still further. For this reason, the dielectric constant of the atmosphere is greater than unity near the earth's surface, but decreases to unity at great heights where the air density approaches zero. Although the dielectric constant and its variation with height are quantities that vary with the weather, the assumption is usually made that the variation of dielectric constant with height above the earth is uniform, and an atmosphere that has the assumed conditions is called a *standard atmosphere.* The justification for the use of a standard atmosphere in computations is that the results predicted on the basis of such an assumption agree fairly well with the results obtained in practice *on the average.* There are times, of course, when the observed results differ markedly from those predicted from the standard atmosphere, and some consideration is given to these nonstandard conditions at the end of this section. The relation between the radius of curvature of the path and the change of dielectric constant with height can be derived as follows.

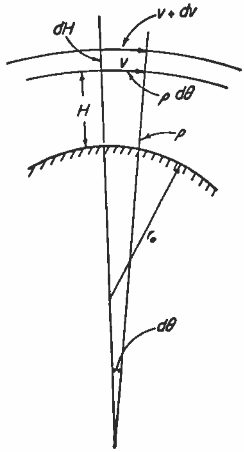


FIG. 16-21

Let ρ be the radius of curvature of the path and v the velocity of propagation at a height H above the earth. Then from Figure 16-21

$$\rho\, d\theta = v\, dt$$

or

$$\frac{d\theta}{dt} = \frac{v}{\rho} \qquad (16\text{-}53)$$

Also

$$v = \frac{1}{\sqrt{\mu_v \epsilon_r \epsilon_v}} = k_1 \epsilon_r^{-1/2} \qquad (16\text{-}54)$$

At a height $H + dH = H + d\rho$, the velocity must be

$$(v + dv) = \frac{(\rho + dH)\, d\theta}{dt}$$

Therefore
$$\frac{dv}{dH} = \frac{d\theta}{dt} = \frac{v}{\rho}$$

or
$$\rho = \frac{v}{dv/dH} = \frac{k_1\epsilon_r^{-1/2}}{-\frac{1}{2}k_1\epsilon_r^{-3/2}\frac{d\epsilon_r}{dH}} = -\frac{2\epsilon_r}{d\epsilon_r/dH} \tag{16-55}$$

Since $\epsilon_r \approx 1$,
$$\rho = -\frac{2}{d\epsilon_r/dH} \tag{16-56}$$

The radius of curvature of the path, being a function of the rate of change of the dielectric constant with height, varies from hour to

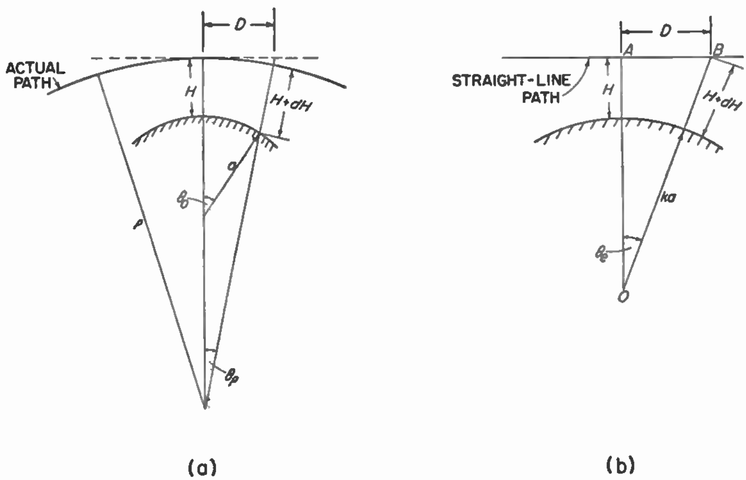


FIG. 16-22. Curved paths become straight lines when an *effective* radius *ka* is used for the earth.

hour, day to day, and season to season. However, in practice an average value of four times the radius of the earth is used for the purposes of calculations.

In working propagation problems it is often convenient to consider the ray paths as straight lines instead of being curved as they actually are, and to compensate for the curvature by using a larger value for the "effective" radius of the earth. The relations involved are shown in Fig. 16-22a and b. In Fig. 16-22a, the actual path is shown above an earth of radius a. In order for the straight-line path of Fig. 16-22b to be the equivalent of that shown in Fig. 16-22a, it is necessary that the change in height dH be the same in the two cases for the same horizontal distance D. In Fig. 16-22b,

$$dH = BO - AO = (ka + H)\left(\frac{1}{\cos\theta_e} - 1\right)$$

For small angles,

$$\frac{1}{\cos\theta_e} \approx \frac{1}{1 - \frac{\theta_e^2}{2}} \approx 1 + \frac{\theta_e^2}{2}$$

$$dH \approx \frac{ka\theta_e^2}{2}$$

when H is small compared to ka. But

$$\theta_e \approx \sin\theta_e = \frac{D}{(ka + H)} \approx \frac{D}{ka}$$

Therefore

$$dH \approx \frac{D^2}{2ka} \tag{16-57}$$

On the other hand, in Fig. 16-20a

$$dH = \frac{D^2}{2a} - \frac{D^2}{2\rho} \tag{16-58}$$

therefore

$$\frac{1}{ka} = \frac{1}{a} - \frac{1}{\rho} \tag{16-59}$$

The effective radius of the earth required is therefore

$$ka = a\left(\frac{1}{1 - \frac{a}{\rho}}\right)$$

so that

$$k = \frac{1}{1 - \frac{a}{\rho}} \tag{16-60}$$

For a radius of curvature ρ, equal four times the radius a of the earth, the effective radius of the earth is 4/3 times the actual radius. By using this effective radius instead of the actual radius in making ground wave path computations, the systematic bending of the waves in the atmosphere is accounted for, and straight-line paths may be drawn.

Tropospheric Reflection. In addition to the systematic refraction of waves that occurs in the troposphere, there is also the possibility of *reflection* occurring at places of abrupt change in the dielectric constant or its gradient. Under conditions of a standard atmosphere the line-of-sight distance is increased approximately 15 per cent by the effects of refraction (though it may be much more or much less under nonstandard conditions). In contrast to this, the effects of the reflection of waves in the troposphere may be to extend the range of reception by a hundred miles or more. In the special case of duct propagation, discussed later, the range may be extended several thousand miles.

Assuming that abrupt changes in the value of the dielectric constant do exist in the troposphere, it is easy to calculate the reflection which will occur at the surface of discontinuity. For a wave traveling in a dielectric medium having a dielectric constant ϵ_1, and incident upon a second medium of dielectric constant ϵ_2, the reflection factors have already been developed in chap. 5. Writing $\epsilon_2 = \epsilon_1 + \Delta\epsilon$, where $\Delta\epsilon$ is the change in the dielectric constant at the layer in the troposphere, eq. (5-73) for vertically polarized waves may be written

$$R_v = \frac{\left(1 + \dfrac{\Delta\epsilon}{\epsilon_1}\right)\cos\theta_1 - \sqrt{\cos^2\theta_1 + \dfrac{\Delta\epsilon}{\epsilon_1}}}{1 + \dfrac{\Delta\epsilon}{\epsilon_1}\cos\theta_1 + \sqrt{\cos^2\theta_1 + \dfrac{\Delta\epsilon}{\epsilon_1}}} \qquad (16\text{-}61)$$

It is known that the abrupt change $\Delta\epsilon$ must be very small, say of the order of 10^{-6} to 10^{-4}, and since ϵ_1 is approximately equal to unity, expression (61) can be reduced to

$$R_v \approx \frac{\Delta\epsilon}{2} - \frac{\Delta\epsilon}{4\cos^2\theta_1} \qquad (16\text{-}62)$$

Similarly for horizontally polarized waves the reflection coefficient of eq. (5-72) can be reduced to

$$R_h \approx -\frac{\Delta\epsilon}{4\cos^2\theta_1} \qquad (16\text{-}63)$$

Using these reflection coefficients and various assumed conditions for $\Delta\epsilon$ and reflecting layer height, the field intensities of tropospheric waves have been calculated.* In Fig. 16-23 the calculated

* K. A. Norton, "On a Theory of Tropospheric Wave Propagation," Report No. 40003 presented before the Federal Communications Commission, March 18, 1940. Also presented before the Broadcast Engineering Conference at the Ohio State University.

tropospheric wave field intensities have been plotted as a function of distance, for different assumed conditions. On the basis of these calculations it is apparent that there may be times when reflections from the troposphere may be expected to produce usable signals at distances considerably beyond those that result when only ground wave propagation paths are considered. Experience with fre-

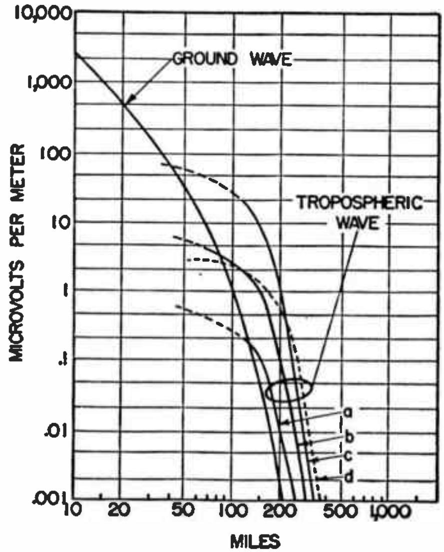


FIG. 16-23. Calculated tropospheric wave and ground wave field intensity for 1 kw radiated at 50 mc. Polarization is horizontal and antenna heights are 500 feet and 30 feet. Earth constants used are $\sigma = 5 \times 10^{-3}$, $\epsilon_r = 15$. Curves a, b, and c are for tropospheric layer height of 1.5 km and $\Delta\epsilon = 10^{-4}$, 10^{-5}, and 10^{-6} respectively. Curve d is for a layer height of 3 km and $\Delta\epsilon = 10^{-5}$. (From FCC report No. 40003 by K. A. Norton.)

quency modulation and television reception seems to bear out these predictions.

Modified Index Curves and Duct Propagation. The atmospheric condition that gives rise to the tropospheric reflection just considered, is a *nonstandard* condition. There are many different types of such nonstandard atmospheres, each of which affects wave propagation in a different way. The standard dry atmosphere has already been defined as one for which the temperature decreases at the rate oı 6.5° C per kilometer. When the temperature increases with height over a certain range of heights, it is known as *temperature inversion.* Actually the water content of the atmosphere has much more effect than temperature on its dielectric constant and on the manner in which it affects radio waves. The *moist standard atmosphere* is specified as one which has a water-vapor pressure of 10 millibars at sea level, decreasing with altitude at the rate of 1 millibar per thousand feet, up to 10,000 ft. If the temperature or water content differ from these standard conditions, *nonstandard* propagation will result. The effects to be expected can be estimated most readily by transforming the meteorological data, temperature, water content, and so on, into M curves. *M curves* are curves that show the variation of the modified index of refraction with height. (The term "modified" refers to the fact that the actual index has been modified to account for the curvature of the earth. When this is done, straight rays above a curved earth come out as curved rays (with an upward curvature) above a flat earth. This procedure, which simplifies computations when rays of different curvatures must be considered, is just the reverse of that used previously when curved rays over a curved earth were transformed to straight rays over an earth of lesser curvature.)

When M curves are available, it is possible to predict, at least roughly, the type of transmission path that can be expected. Standard propagation occurs when the modified index of refraction increases linearly with height. In this case, the M curve is a straight line with a positive slope. If the slope of the M curve decreases near the surface of the earth, *substandard* propagation results, with the rays curving upward (over the *flat* earth) more than for normal conditions. If the slope of the M curve increases near the surface of the earth, the upward curvature of the rays is less, so that greater coverage is achieved and *superstandard* conditions result. If the M curve becomes vertical (no change of modified index with height), the rays over the flat earth are straight and very great coverage can be obtained. (In this condition the actual rays have the same curvature as the curvature of the earth.)

If the modified index decreases with height (M curve slopes to left) over a portion of the range of height, the rays will be curved downward (over the flat earth) and a condition known as *trapping* or *duct* propagation can occur. Under such conditions the wave tends to be trapped or guided along the duct, much as a wave is guided by a leaky wave guide. If the lower side of the duct is at the surface of the earth, it is known as a surface duct. Sometimes when the inverted portion of the M curve is elevated above the surface of the earth, the lower side of the duct is also elevated, and the duct is called an *elevated duct.* If the receiving antenna is elevated to within the duct, the signal may be very large. However, if the receiving antenna is outside the duct, either below or above it, the received signal will be very small. Elevated ducts are due to a subsidence of large air masses and are common in Southern California and certain areas of the Pacific. They are found at elevations of 1000 to 5000 ft and may vary in thickness from a few feet to a thousand feet. In the trade wind belt over sea there appears to be a continuous surface duct about 5 ft thick. Over land areas surface ducts are produced by radiation cooling of the earth.

As with ordinary wave guide propagation, there is a certain critical frequency (which depends on the thickness of the duct) below which duct propagation will not occur. Since these non-standard refraction effects appear to be restricted to waves that make a very small angle with the horizontal, it is evident that the required thickness of the duct would have to be large in wavelengths. For this reason trapping is more likely to occur at the ultrahigh frequencies than at very high frequencies.

The tropospheric propagation considered in this section is very much a function of the weather. As more meteorological and radio transmission data become available, it is to be expected that a very close correlation between the two sets of information can be achieved. It should then prove possible to use measurements of atmospheric conditions to predict accurately tropospheric transmissions, and perhaps vice versa.

BIBLIOGRAPHY

Barfield, R. H., "The Attenuation of Wireless Waves over Land," *J. IEE*, **66**, 204 (1928).

Burrows, Charles R., "The Surface Wave in Radio Propagation over Plane Earth," *Proc. IRE*, **25**, 219 (1937).

Norton, K. A., "The Physical Reality of Space and Surface Waves in the Radiation Field of Radio Antennas," *Proc. IRE*, **25**, 1192 (1937).

Norton, K. A., "The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE*, **29**, 623 (1941).

Howe, G. W. O., "Wireless Waves at the Earth's Surface," *Wireless Engineer*, **17**, 204, 385 (1940).

Pedersen, P. O., "The Propagation of Radio Waves," *G.E.C. Gad*, Vimmelskaftet 32, Copenhagen.

Burrows, C. R., "Radio Propagation over Spherical Earth," *Proc. IRE*, **23**, 470–480 (1935).

Feldman, C. B., "Optical Behaviour of the Ground," *Proc. IRE*, **21**, 765 (1933).

van der Pol, Balth., and Bremmer, H., "The Diffraction of Electromagnetic Waves from an Electrical Point Source Round a Finitely Conducting Sphere, with Applications to Radiotelegraphy and the Theory of the Rainbow," *Phil. Mag.*, **24**, 141–176 (1937); Part II, **24**, 825–864 (1937).

Attwood, Stephen S., "The Propagation of Radio Waves through the Standard Atmosphere," NDRC Summary Technical Report, C.P. Vol. 3, 1946.

CHAPTER 17

SKY WAVE PROPAGATION

GENERAL CONSIDERATIONS

17.01 Introduction. The ground wave propagation paths considered in the previous chapter are not the only paths along which a transmitted wave may travel to reach the receiver. This was demonstrated to a surprised scientific world in 1901 by Marconi's successful transmission of radio signals across the Atlantic. Calculations had already been made to show that diffraction effects would be insufficient to permit such long-distance transmission around the curvature of the earth, and immediately other explanations were sought. The existence of a reflecting region in the earth's upper atmosphere was proposed (independently) by A. E. Kennelly and Oliver Heaviside, and the Kennelly-Heaviside layer, or *ionosphere*, as it is now known, became a much discussed part of radio propagation phenomena.

Knowledge of the characteristics of the ionosphere is based almost entirely upon its effect on radio waves, which may or may not be reflected from it back to the earth's surface. Experimentally it is found that at night signals in the broadcast frequency range are reflected back, but in the daytime the reflected signal is very weak or entirely absent. As the frequency is raised, however, these daytime reflected waves become stronger, and for frequencies between 10 and 30 mc, they may provide strong signals over distances of several thousand miles. As the frequency is increased still higher, a point is reached where the waves cease to be reflected back, but instead, penetrate the ionosphere to be lost in outer space. Thus there is a range of frequencies roughly between 3 and 30 mc where, although the surface wave is greatly attenuated, long-distance transmission may still occur because of reflections from the ionosphere. In general, these "sky wave" signals are less stable than ground wave signals, their strength depending upon the frequency, and upon the condition of the ionosphere. The state of the ionosphere is found to vary from hour to hour, day to day, and season to season in much the same way as does the weather. Also as with the weather, there may be periods of sudden storms, but many of the variations are fairly regular and may be predicted several days or even weeks ahead of time. Indeed, the art of ionosphere prediction is in much the same state as, and is very similar to, that of weather forecasting. Ionosphere stations set up in various parts of the world continuously gather and record information about the ionosphere in those regions. This information is assembled, correlated, interpreted, and issued in the form of charts that show past conditions and also make predictions for the future. Using these charts it is possible to determine in advance the optimum frequency to use for communication between any two points on the earth's surface at any given time. Thus, although long-distance ionospheric propagation does not have the stable characteristics of short-distance ground wave propagation, it does, in general, provide a predictable, and therefore usable, means of radio communication. A knowledge of some of the more important characteristics of the ionosphere will aid the engineer in an intelligent over-all design of a communications system.

17.02 The Ionosphere. The ionosphere is that region of the earth's atmosphere in which the constituent gases are ionized by radiations from outer space. This region extends from about 40 to 250 miles above the earth. The ionizing agent is chiefly ultraviolet light from the sun, which is very intense before being absorbed by the earth's atmosphere, but there is reason to believe that there are other agents as well, and cosmic rays and meteors in the high atmosphere have been suggested. Although ions and electrons are undoubtedly present to some extent throughout the whole of this region, there seem to be several layers in each of which the ionization density reaches a maximum. These layers are designated by the letter symbols D, E, and F in order of height. At times the F layer splits into separate layers called F_1 and F_2. The probable distribution of ionization density with height is indicated in Fig. 17-1, which shows conditions for a summer day. Conditions at night or on a winter day are different in that the F_1 layer merges with the F_2 layer to form the single F layer. The D layer, which does not have a permanent existence, has not been shown.

The existence of the ionosphere in the form of a layer is explained on the following basis: At great heights the ionizing radiations are very intense, but the atmosphere is rare and there are few molecules present to be ionized. Therefore in this region the ionization density (number of ions or electrons per unit volume) is very low. As height is decreased, the atmospheric pressure and ionization density increase until a height is reached where the ionization density is a maximum. Below this height the atmospheric pressure continues to increase, but the ionization density decreases because the ionizing radiation has been absorbed or used up in the process of

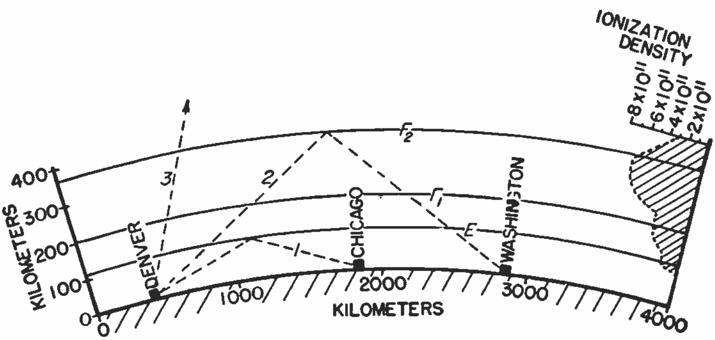


FIG. 17-1. Probable distribution of ionization density with height.

ionization. This explains in general why there should be a layer. The existence of layers within the layer is accounted for by the fact that the atmosphere is a mixture of several gases that differ in their susceptibility to the ionizing radiations, and so produce maximum ionization at different pressures.

Although the number and heights of the layers vary with time, there are two layers that have a permanent existence. These are designated as the *E* layer and the *F* layer. The *E* layer exists at a height of between 55 and 85 miles (89 to 137 kilometers) and the level of maximum ionization density remains fairly constant in height throughout the day, and from season to season. The *E* layer disappears at night.* The *F* layer exists above the *E* layer

* This statement refers to the regular day-time *E* layer that is produced by the ultraviolet light from the sun. There is evidence of a relatively low level of ionization in the *E* region that exists through the night. The cause of this night-time *E* layer has not yet been definitely established.

from about 85 to 250 miles, but the level of maximum ionization density varies between night and day and from season to season. In addition there is evidence of other layers which are present only part of the time. One of these layers is the *D* layer, which lies below the *E* layer at a height of about 40 miles. Although the *D* layer does not normally reflect back high-frequency waves, its presence decreases the intensity of signals reflected from the higher layers. Another layer that does not have a permanent existence is the so-called *sporadic E* layer. This is not so much a separate layer as it is a patch or cloud of electrons or ions having a relatively sharp boundary and existing at the height of the regular *E* layer. Reflections from sporadic *E* patches often make possible long-distance reception of waves of much higher frequency than would normally be possible.

17.03 Effective ϵ and σ of an Ionized Region. The ionosphere is a dielectric region containing free electrons and ions. In the absence of these free charges the constants of this region would be essentially those of free space, viz., $\epsilon = \epsilon_v$, $\mu = \mu_v$, and $\sigma = 0$. Under the influence of a passing electromagnetic wave the charges have imparted to them an oscillatory motion that both absorbs and reradiates some of the energy of the wave. As far as the effect on the electromagnetic wave is concerned, the ionosphere may be treated as a charge-free but imperfect dielectric having an *effective* dielectric constant ϵ and an *effective* conductivity σ, which are different from the free-space values. As will be shown, these effective values of ϵ and σ may be calculated in terms of the frequency, the ion density, and the collision frequency in the ionized region.

Consider a region in the upper atmosphere in which there is an electron or ion density of N electrons or ions per cubic meter. The electric field of the passing electromagnetic wave produces a movement of the electrons in the direction of the field, and so gives rise to a current density,

$$i = Nev \qquad \text{amp/sq m} \qquad (17\text{-}1)$$

where e is the electronic charge in coulombs and v is the average instantaneous velocity in the direction of the electric field. If E is the field strength in volt/m of the electromagnetic wave, the force on the electron is Ee. If there were no collisions between electrons and molecules of the gases of the atmosphere, the equation of motion for the electron would be

$$Ee = m \frac{dv}{dt}$$

and all of the electric force would be used in accelerating the electron. However, in the event that there are collisions between electrons and gas molecules, energy is lost in the form of heat and it may be considered that there is "frictional" or retarding force on the electron that is proportional to the velocity v. The equation of motion then becomes

$$Ee = m \frac{dv}{dt} + R_e v \tag{17-2}$$

where R_e is an effective frictional resistance. The actual average frictional force due to collisions is given by $mv\nu$, where mv is the average momentum lost on collision and ν is the frequency of collision. Thus

$$R_e = m\nu \tag{17-3}$$

Assuming that $E = E_0\epsilon^{j\omega t}$, $v = v_0\epsilon^{j\omega t}$, a solution of eq. (2) is

$$v = \frac{E_e}{R_e + j\omega m}$$

The current density is

$$\begin{aligned} i = Nev &= \frac{Ne^2E}{R_e + j\omega m} \\ &= \frac{Ne^2ER_e}{R_e^2 + \omega^2m^2} - \frac{j\omega mNe^2E}{R_e^2 + \omega^2m^2} \\ &= \frac{Ne^2\nu E}{m(\nu^2 + \omega^2)} - \frac{j\omega Ne^2E}{m(\nu^2 + \omega^2)} \end{aligned}$$

In the ionized region Maxwell's emf equation can be written

$$\begin{aligned} \text{curl } \mathbf{H} &= j\omega\epsilon_v\mathbf{E} + i \\ &= j\omega\epsilon_v\left[1 - \frac{Ne^2}{\epsilon_v m(\nu^2 + m^2)}\right]\mathbf{E} + \left[\frac{Ne^2\nu}{m(\nu^2 + \omega^2)}\right]E \\ &= (j\omega\epsilon_r\epsilon_v + \sigma)\mathbf{E} \end{aligned} \tag{17-4}$$

where

$$\epsilon_r = \left[1 - \frac{Ne^2}{\epsilon_v(\nu^2 + \omega^2)}\right] \quad \text{and} \quad \sigma = \frac{Ne^2\nu}{m(\nu^2 + \omega^2)} \tag{17-5}$$

The presence of the electrons or ions has a twofold effect. The effective dielectric constant of the region is reduced below that of free space, and the region has an effective conductivity σ which depends upon the electron density and the collision frequency. It will be observed that for a given frequency the effective conductivity in a region is a maximum when the collision frequency ν is equal to ω. The collision frequency is dependent upon the thermal agitation velocity and the gas pressure, and is therefore a function of height. It is estimated that the collision frequency varies from about 10^{12} times per second at the surface of the earth to 1 time per second at 500 miles up. Examination of eq. (4) reveals that at great heights where ν is small and $\omega \gg \nu$, the conductivity will become vanishingly small and the effective dielectric constant will be given by $\epsilon_v[1 - (Ne^2/\epsilon_v m\omega^2)]$. On the other hand, at low heights such that $\nu \gg \omega$, the conductivity again becomes small and the reduction of the dielectric constant approaches zero. These effects that occur with decreasing height are augmented by the fact that the electron density N also decreases rapidly below about 50 miles. The result is that the region of high conductivity (and therefore high absorption when the wave penetrates it) is confined to a relatively thin layer at the lower edge of the E region.

The effect of the presence of the ions and electrons in decreasing the effective dielectric constant and increasing the conductivity of the medium is just what would be expected from simple physical reasoning. In free space the electromagnetic wave results in a displacement current which leads the electric field by 90 degrees, i.e., there is a capacitive current flow. With ions and electrons present there is also a convection current flow that is in phase with the velocity of the particles. In the absence of collisions the velocity lags the electric field by 90 degrees (the acceleration is in phase with the electric force) and the convection current is an inductive current flow, in opposite phase to the displacement current. When there are collisions between the electrons and gas molecules, the velocity lags the electric intensity by an angle less than 90 degrees and there is an in-phase or power component of convection current. Thus the medium has an effective conductivity, or there is now a resistance in series with the shunting inductance. The transmission line analogy for this phenomenon will be given more detailed consideration in a later section.

17.04 Reflection and Refraction of Waves by the Ionosphere. The mechanism of reflection and refraction of radio waves by the ionosphere is very much a function of frequency. At low frequencies, say below 100 kc, the change in electron and ion density within the distance of a wavelength is so great that the layer presents virtually an abrupt discontinuity in the medium. Under these circumstances, the reflection may be treated in the same manner as the reflection of waves at the surface of a dielectric that may or may not have loss. On the other hand, at the high end of the high-frequency band, the length of a wavelength is sufficiently short that the ionization density changes only slightly in the course of a wavelength. Under such conditions the ionosphere may be treated (by methods well-known in optics) as a dielectric with a continuously variable refractive index. For in-between frequencies, not covered by these two cases, it is possible to treat the reflection region as though it consisted of several thin but discrete layers, each layer having a constant ionization density that differs from that of the adjacent layer. It follows that the incident wave will be partially refracted. The refracted wave penetrates to the second layer where it is partially reflected and partially refracted, and so on. In this case the resultant reflected signal may be considered as the sum of reflections from various parts of the ionized layer. Because they suffer greater attenuation, these in-between frequencies are of less practical interest than the others, and only the first two cases will be treated.

CASE I: *Reflection at Low Frequencies*

In this case the wavelength is considered to be sufficiently long that there is a great change in the ionization density in the course of a wavelength. The layer then may be considered a reflecting surface, for which reflection coefficients corresponding to those developed in chap. 16 may be written. These coefficients are

$$R_h = \frac{\cos\theta - \sqrt{\left(\epsilon_r + \dfrac{\sigma}{j\omega\epsilon_v}\right) - \sin^2\theta}}{\cos\theta + \sqrt{\left(\epsilon_r + \dfrac{\sigma}{j\omega\epsilon_v}\right) - \sin^2\theta}} \tag{17-6}$$

$$R_v = \frac{\left(\epsilon_r + \frac{\sigma}{j\omega\epsilon_v}\right)\cos\theta - \sqrt{\left(\epsilon_r + \frac{\sigma}{j\omega\epsilon_v}\right) - \sin^2\theta}}{\left(\epsilon_r + \frac{\sigma}{j\omega\epsilon_v}\right)\cos\theta + \sqrt{\left(\epsilon_r + \frac{\sigma}{j\omega\epsilon_v}\right) - \sin^2\theta}} \tag{17-7}$$

where the effective values of σ and ϵ_r are given by eqs. (5).

It is apparent that for this type of reflection, the reflection coefficient of the medium will depend upon the frequency, polarization, and angle of incidence of the wave. When the effective conductivity can be neglected the reflection curves will be those for reflection from a perfect dielectric that has a refractive index of less than unity. For angles of incidence greater than a certain critical angle (which depends upon the refractive index), there will be complete reflection of the signal for both polarizations of the wave. For angles less than the critical (that is, closer to the normal), the reflection coefficient will be less than unity and will depend on the angle of incidence. The effect of finite effective conductivity would be to modify these curves in much the same way as the reflection curves of Figs. 16-3, and 16-4 were modified by the conductivity. The ranges of ϵ_r and σ are, of course, here quite different from those of Figs. 16-3 and 16-4. ϵ_r is less than unity and σ is small, and their values depend upon the electron density and collision frequency at the height at which reflection occurs. These latter variables in turn will depend upon the time of day and time of year.

Case II: *Reflection (or Refraction) at High Frequencies*

(This is an important case, practically.) When the change in phase velocity within the course of a wavelength is small, the well-known methods of ray optics may be used to obtain a solution. The requirement of small change in phase velocity means, in this case, a small change in electron density, as can be seen from the following considerations: The phase velocity of a wave in a medium having negligible loss is given by

$$v_p = \frac{1}{\sqrt{\mu\epsilon}} = \frac{c}{\sqrt{\mu_r\epsilon_r}} \tag{17-8}$$

where, as usual, $c = 1/\sqrt{\mu_v\epsilon_v}$ is the velocity of light in a vacuum. Assuming the permeability of the ionosphere to be unchanged by the presence of electrons so that $\mu_r = 1$, the phase velocity will be

$$v_p = \frac{c}{\sqrt{\epsilon_r}} \tag{17-9}$$

where ϵ_r depends upon the electron density N as indicated in eq. (5). If the change in electron density in the distance of a wavelength is small the change in phase velocity will also be small.

Under the conditions for this case the wave penetrates the lower edge of the ionosphere without reflection, but within the ionosphere travels a path that is curved away from the region of greater electron density (smaller index of refraction). At any point along the path, the angle ϕ between the path and the normal (Fig. 17-2) is given by Snell's law of refraction

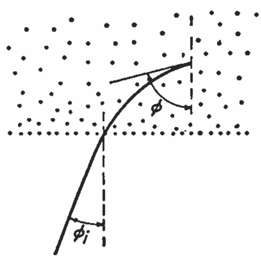


FIG. 17-2

$$\sin \phi_i = n \sin \phi$$

or

$$\sin \phi = \frac{\sin \phi_i}{n} \tag{17-10}$$

n is the index of refraction at the point where ϕ is observed, and ϕ_i is the angle of incidence (measured from the normal to the ionosphere layer). The refractive index for any medium is given by

$$n = \frac{c}{v_p} = \frac{\text{velocity of light in vacua}}{\text{phase velocity in the medium}} \tag{17-11}$$

For the lossless case, where (9) is true, eq. (11) gives for the refractive index

$$n = \sqrt{\epsilon_r} \tag{17-12}$$

In general the effective conductivity of the ionosphere cannot be neglected, but at the higher frequencies where this present analysis is applicable, reflection takes place in the F layers where the collision frequency is very small and the conductivity is correspondingly low. Therefore, for a first approximation at least, it is permissable to neglect the effects of conductivity and use the simple expression given in (12).

For $\omega^2 \gg \nu^2$, the expression for ϵ_r [from eq. (5)] is

$$\epsilon_r = \left(1 - \frac{Ne^2}{\epsilon_v m \omega^2}\right) \tag{17-13}$$

For an electron, $e = 1.59 \times 10^{-19}$ coulombs, $m = 9 \times 10^{-31}$ kg, so that (13) becomes

$$\epsilon_r = \left(1 - \frac{81N}{f^2}\right) \tag{17-14}$$

N is the number of electrons per cubic meter and f is the frequency in cycles per second. (However, if N is expressed as the number of electrons per cubic centimeter and the frequency is expressed in kc, relation (14) is still true).

From (12) the refractive index is

$$n = \sqrt{1 - \frac{81N}{f^2}} \tag{17-15}$$

The refractive index decreases as the wave penetrates into regions of greater electron density and the angle of refraction increases correspondingly. When n has decreased to the point where $n = \sin \phi_i$, the angle of refraction ϕ will be 90 degrees and the wave will be traveling horizontally. The highest point reached by the wave is therefore that point at which the electron density N satisfies the relation

$$\sqrt{1 - \frac{81N'}{f^2}} = \sin \phi_i$$

or

$$N' = \frac{f^2 \cos^2 \phi_i}{81} \qquad (17\text{-}16)$$

If the electron density at some level in a layer is sufficiently great to satisfy relation (16), the wave will be returned to earth from that level. If the maximum electron density in a layer is less than that required by (16), the wave will penetrate the layer (though it may be reflected back from a higher layer for which N is greater).

The largest electron density required for reflection occurs when the angle of incidence ϕ_i is zero, that is, for vertical incidence. For any given layer the highest frequency which will be reflected back for vertical incidence will be

$$f_c = \sqrt{81 N_{max}} \qquad (17\text{-}16a)$$

where N_{max} is the maximum ionization density (electrons per cubic meter) and f_c is the *critical frequency* for the layer.

Experimental Determination of Critical Frequencies and Virtual Heights. Ionosphere characteristics are determined experimentally by measuring the amplitude and time delay of reflected signals as a function of frequency. The commonest method is that in which the transmitted signal consists of pulses of rf energy of short duration. The receiver, which is located close to the transmitter, picks up both the direct and the reflected signal. The spacing between these signals on the time axis of a cathode ray oscilloscope gives a measurement of the height of the layer. The height so measured is the *virtual height* of the layer and is higher than the true height of the lower edge of the layer as indicated in Fig. 17-3. The virtual height h' is that height from which a wave sent up at an angle appears to be reflected. It is also the height obtained by pulse measurements, because the time delay for the actual curved path ABC is approximately the same as it would be for a wave to travel the path ADC if the ionosphere were replaced by a mirrorlike reflecting surface at the level of D. Although the path length ADC is greater than ABC, the group velocity in the ionosphere is less than in free space by just the amount required to make the time delays of the two paths equal.

As the frequency of the transmitted signal is increased, starting say at 2 mc, the measured virtual height increases slightly, indicat-

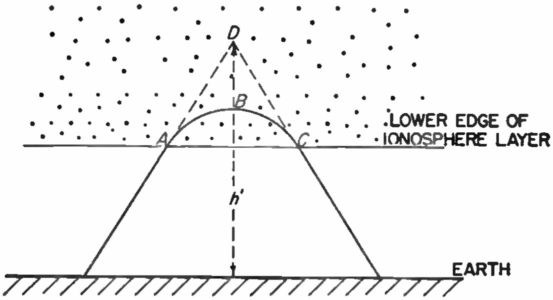


FIG. 17-3. Virtual height h' of a layer.

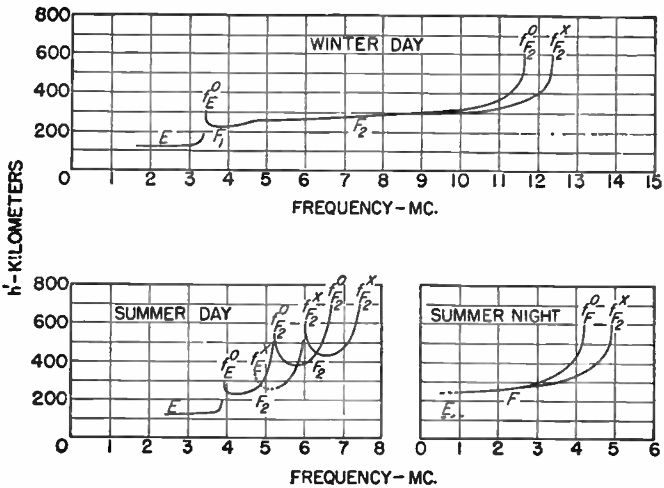


FIG. 17-4. Virtual height—frequency curves.

ing that for the higher frequencies the wave is returned back from higher levels within the layer. This continues until a critical frequency is approached near which the virtual height increases suddenly to quite high values as shown in the virtual height-frequency curve of Fig. 17-4. As the critical frequency is passed, the measured virtual height drops back to a second more or less steady value which is higher than was obtained for frequencies well below the critical frequency. The wave is now penetrating the first layer and is reflected from a second, higher layer. The critical frequency of a layer is that frequency for which a vertically incident wave just fails to be reflected back from the layer. As the frequency is increased above the critical frequency for the lower layer, a second and sometimes a third critical frequency may be reached corresponding to the critical frequencies for the higher layers. The apparent increase in the measured height of the layer in the neighborhood of the critical frequency is due to a large time delay in the ionized medium, occurring as a result of a much reduced group velocity near this frequency.

Maximum Usable and Optimum Frequencies. Although the critical frequency for any layer represents the highest frequency that will be reflected back from that layer at vertical incidence, it is not the highest frequency that can be reflected from the layer. The highest frequency that can be reflected depends also upon the angle of incidence, and hence, for a given layer height, upon the distance between the transmitting and receiving points. The maximum frequency that can be reflected back for a given distance of transmission is called the *maximum usable frequency* (*MUF*) for that distance. From eq. (16) and using (16a), it is seen that the maximum usable frequency is related to the critical frequency and the angle of incidence by the simple expression

$$\text{MUF} = f_c \sec \phi_i \tag{17-17}$$

The maximum usable frequency for a layer is greater than the critical frequency by the factor $\sec \phi_i$. Because of curvature of the earth and the ionospheric layer, the largest angle of incidence ϕ_i that can be obtained in F layer reflection is of the order of 74 degrees. This occurs for a ray that leaves the earth at the grazing angle. The geometry for this case is shown by Fig. (17-5), where ϕ_i (max) $= \sin^{-1} (r/r + h)$. The maximum usable frequency at this limiting angle is related to the critical frequency of the layer by

$$\text{MUF (max)} \approx \frac{f_c}{\cos 74°} \approx 3.6 f_c \tag{17-18}$$

When the critical frequency is known, the maximum usable frequency can be calculated for any given distance through use of eq. (17). Figure 17-6 shows a set of maximum usable frequency curves for the latitude of Washington, D. C., for a winter month

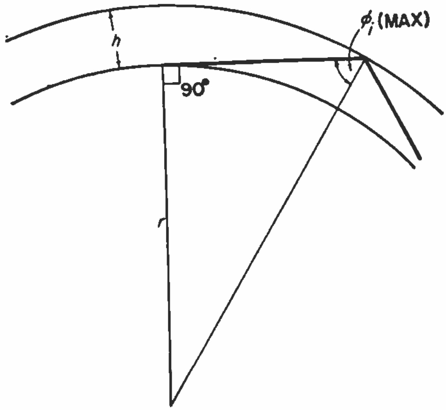


Fig. 17-5

during a period of maximum sunspot activity. It is evident that, whereas a given frequency, say 28 mc would have been satisfactory for transmitting over distances of 2000 kilometers or more near midday, the same signal would have failed to be reflected back at points less than 1500 kilometers from the transmitter at the same time. The distance within which a signal of given frequency fails to be reflected back is the *skip distance* for that frequency. The higher the frequency the greater is the skip distance.

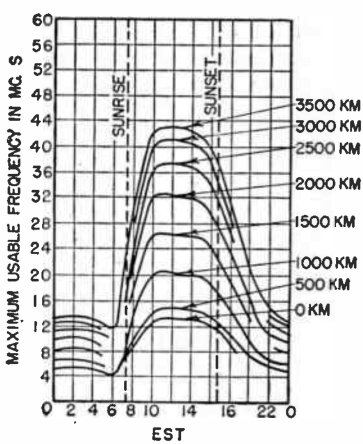


Fig. 17-6. Typical set of maximum usable frequency curves for a winter month.

Because the maximum usable frequency may show small daily variations about the monthly average of up to 15 per cent, it is customary to use a frequency somewhat lower than the predicted maximum usable frequency. Also because it is desirable to restrict the number of different frequencies required to a reasonable number, some latitude must be permitted in the choice of frequency actually used. The *optimum* frequency for transmitting between any two points is therefore selected as some frequency lying between about 50 and 85 per cent of the predicted maximum usable frequency between those points.

17.05 Regular and Irregular Variation of the Ionosphere. Conditions in the ionosphere vary throughout the day with the altitude of the sun, and in addition vary quite regularly with the season of year. A plot of critical frequencies and virtual heights as a function of time of day gives a reasonably good picture of ionospheric variations. It is customary to show monthly averages of these quantities because the day to day variations are usually quite small, except during periods of ionosphere storminess. Figure 17-7 shows some typical virtual-height and critical-frequency curves for summer and winter. The F layer, which has a virtual height of about 300 kilometers during the night, splits into two separate layers during the day. The lower of these is designated F_1 and the upper is designated F_2. The E layer exists only during the day, disappearing as soon as the sun goes down. Its virtual height remains almost constant at 110–120 kilometers from season to season and year to year.

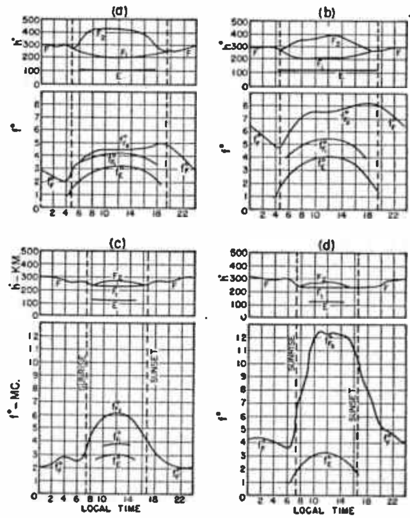


FIG. 17-7. Diurnal variation of critical frequency and virtual height of the regular ionosphere layers. (a) Summer at period of sunspot minimum. (b) Summer at sunspot maximum. (c) Winter at sunspot minimum. (d) Winter at sunspot maximum.

Besides the diurnal and seasonal variation of virtual height and critical frequency, these quantities also vary in synchronism with the 11-year sunspot cycle, as shown in Fig. 17-7. The critical frequencies are considerably higher during a year near a sunspot maximum than during a period near a sunspot minimum.

In curves such as those of Figs. 17-4 and 17-7, the critical frequencies for the different layers are designated by appropriate letter subscripts denoting the layer. Because of the presence of the earth's magnetic field, there are actually two different critical frequencies for each layer, one for the so-called *ordinary* wave and a higher one for the *extraordinary* wave. The respective critical frequency is therefore denoted by the superscript o or x; for example $f^o{}_{F_1}$ is the critical frequency of the F_1 layer for the ordinary wave. For the E layer the ordinary wave predominates and the extraordinary is usually not considered because it has negligible effect on radio reception. Ordinary and extraordinary waves will be discussed further in a later section on the effect of the earth's magnetic field.

The critical frequency of the E layer has a regular diurnal and seasonal variation. It increases with the altitude of the sun and is a maximum at noon on a summer day. For the E layer it has been found that the critical frequency is given approximately by the simple relation

$$f_E = K \sqrt[4]{\cos \psi}$$

where ψ is the zenith angle of the sun and K is a factor that depends upon the intensity of the radiation from the sun. The critical frequencies of the F layers do not obey any such simple law. For the F_2 layer the diurnal maximum lags behind the altitude of the sun and the daytime critical frequencies are higher in winter than in summer (for the northern hemisphere). Figure 17-8 shows typical curves of the distribution of the ionization density N with height. The curves shown are for day and night conditions in both summer and winter for a mid-latitude region. The values of N are obtained from sweep-frequency virtual-height measurements through the relation

$$N = \frac{\omega^2 m \epsilon_v}{e^2} \tag{17-19}$$

The maximum value of N for any layer is given by

$$N_{\max} = \frac{\omega_c{}^2 m \epsilon_v}{e^2} \tag{17-20}$$

where N is the number of electrons per cubic meter and f_c is the critical frequency in cycles for that layer. It will be observed from these curves that for daytime conditions the ionization density falls off very rapidly below about 100 kilometers.

Irregular Variation of the Ionosphere. In addition to the regular or normal variation of ionospheric characteristics indicated by Fig. 17-7, there are also irregular variations that are often unpre-

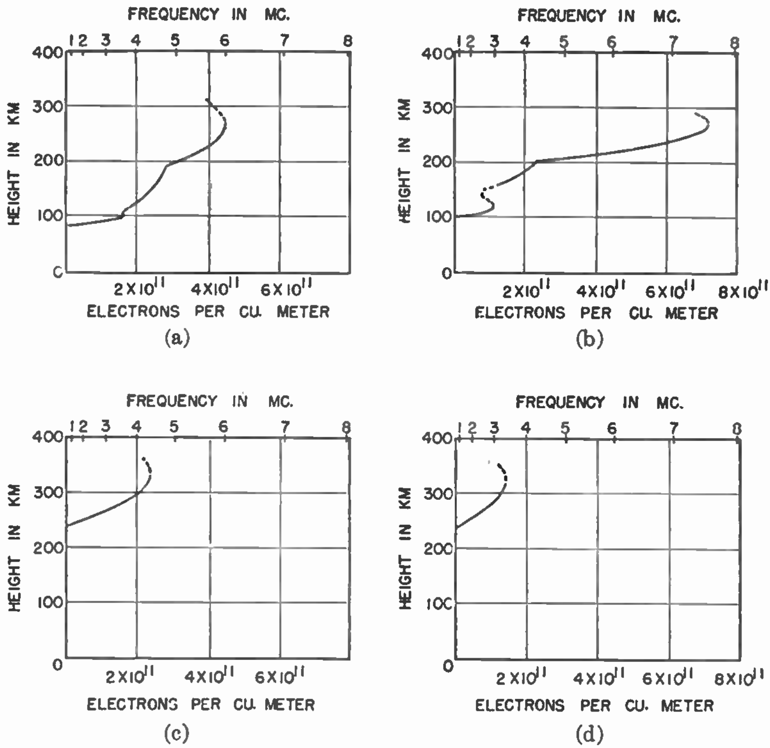


Fig. 17-8. Distribution of ionization density with height for quiet conditions. (a) Summer noon. (b) Winter noon. (c) Summer midnight. (d) Winter midnight.

dictable and that sometimes have a marked influence on radio wave propagation. One of these irregular variations is a sudden ionospheric disturbance, known as the *Dellinger effect*, which produces a complete radio "fade out" lasting from a few minutes to an hour or more. The phenomenon is caused by sudden bright eruptions on the sun that produce a large increase in the ionizing radiations that reach the D layer. The resulting increase of ionization density in this layer results in a complete absorption of all sky wave signals having a frequency greater than about 1 mc. However, for the very low frequencies that are normally reflected from this layer, the sky wave signals will increase in intensity. This sudden ionospheric disturbance is often accompanied by disturbances in terrestrial magnetism and earth currents. The effect never occurs at night.

A second type of irregular variation is somewhat similar in origin and effect to the sudden disturbance mentioned, but its beginning and ending are more gradual and it may last for several hours. Usually the absorption of radio signals is not as complete and communication may be carried on at higher frequencies.

In a third type of irregularity, known as *ionospheric storms*, the ionosphere is turbulent and loses its normal stratification. The result is that radio wave propagation becomes very erratic, and it is often necessary to lower the working frequency in order to maintain communication. The cause of the storm is thought to be the emission of a burst of electrified particles from the sun, and the fact that the storms tend to recur at 27-day intervals, the period of rotation of the sun, seems to indicate that there are active areas on the sun which produce the phenomenon. The effects of ionosphere storms may last for several days.

17.06 Attenuation Factor for Ionospheric Propagation. In section 17.03 the equivalent conductivity σ and dielectric constant ϵ of the ionosphere were obtained in terms of the ionization density N and the collision frequency ν. The attenuation factor α for wave propagation through this region will be given directly by eq. (53) of chap. 5. It is

$$\alpha = \omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1+\frac{\sigma^2}{\omega^2\epsilon^2}}-1\right)} \tag{5-53}$$

where

$$\epsilon = \epsilon_r\epsilon_v = \epsilon_v\left(1 - \frac{Ne^2}{\epsilon_v m(\nu^2+\omega^2)}\right) \tag{17-21a}$$

$$\sigma = \frac{Ne^2\nu}{m(\nu^2+\omega^2)} \tag{17-21b}$$

and

$$\mu = \mu_v.$$

Substituting (21a) and (21b) in (5-53), the expression for α may be written

$$\alpha = \frac{\omega}{c}\sqrt{\frac{\epsilon_r}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} - 1\right)}$$

$$= \frac{\omega}{c}\sqrt{\frac{1}{2}\left(\sqrt{\epsilon_r^2 + \frac{\sigma^2}{\omega^2\epsilon_v^2}} - \epsilon_r\right)}$$

$$= \frac{\omega}{c}\sqrt{\frac{1}{2}\sqrt{\epsilon_r^2 + \left[(1 - \epsilon_r)\frac{\nu}{\omega}\right]^2} - \frac{\epsilon_r}{2}} \qquad (17\text{-}22)$$

For frequencies not too near the maximum usable frequency, and for the important practical case of a section of the ionosphere where the relation $\sigma/\omega\epsilon \ll 1$ holds, eq. (5-53) for α reduces by use of the bionomial expansion to

$$\alpha = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon}} = \frac{60\pi\sigma}{\sqrt{\epsilon_r}} = \frac{60\pi Ne^2\nu}{\sqrt{\epsilon_r}\, m(\nu^2 + \omega^2)} \qquad (17\text{-}23)$$

For the frequency range where $\omega \gg \nu$, eq. (23) shows that the attenuation varies approximately as the inverse square of the frequency. Therefore it is desirable to use as high a frequency as possible without approaching too close to the maximum usable frequency. If the ionization density and collision frequency are known, the attenuation per unit length can be calculated by means of (21a) and (22). The total attenuation of the wave in the ionosphere would then be obtained by integrating α along the whole length of path through the ionosphere. In general, it is found that attenuation is negligibly small, except in the region near the lower edge of the ionosphere (the D region) and at the top of the path where the ray is being bent. The absorption that occurs in the region where the wave is bent, is called *deviative absorption*, whereas that which occurs in the D region is known as *nondeviative absorption*. For high frequencies, where reflection takes place from the F layer, deviative absorption is usually small because the collision frequency in this layer is low. Exceptions to this occur for frequencies near the maximum usable frequency where the wave is abnormally retarded and appreciable absorption of energy may take place.

Since it is known from theoretical considerations that the collision frequency ν is high near the surface of the earth but decreases very rapidly with increasing height, it can be deduced through the use of eq. (23) that the main region of nondeviative absorption will be confined to a relatively narrow range of heights lying somewhere between 60 and 100 km.

In the above analysis no account has been taken of the effect of the earth's magnetic field. This subject is treated in the next section.

17.07 Effect of the Earth's Magnetic Field. Electrons and ions in the ionosphere moving under the influence of the electric field of a passing electromagnetic wave, experience an additional force because of their velocity in the presence of the earth's magnetic field. Taking into account the effect of this steady magnetic field $\mathbf{B}_0$, but neglecting frictional forces due to collision between electrons and gas molecules, the equation of force on the particle can be written

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}_0) = m\frac{\partial \mathbf{v}}{\partial t} \tag{17-24}$$

Equation (24) is a vector equation that can be written as three scalar equations. Choosing the z axis of a rectangular co-ordinate system parallel to $\mathbf{B}_0$, so that $\mathbf{B}_0 = \mathbf{k}B_z$, and $B_x = B_y = 0$, eq. (24) becomes

$$\left.\begin{aligned} E_x + v_yB_0 &= \frac{j\omega m}{e}v_x & \text{(a)}\\ E_y - v_xB_0 &= \frac{j\omega m}{e}v_y & \text{(b)}\\ E_z &= \frac{j\omega m}{e}v_z & \text{(c)} \end{aligned}\right\} \tag{17-25}$$

In eqs. (25) all variables have been assumed periodic with a frequency $\omega/2\pi$, so the time derivative of v has been replaced by $j\omega v$. Solving (a) and (b) simultaneously for v_x and v_y gives

$$\left.\begin{aligned} v_x &= \frac{j\omega(m/e)E_x + B_0E_y}{B_0{}^2 - \omega^2(m/e)^2} & \text{(a)}\\ v_y &= \frac{j\omega(m/e)E_y - B_0E_x}{B_0{}^2 - \omega^2(m/e)^2} & \text{(b)} \end{aligned}\right\} \tag{17-26}$$

"Resonance" occurs for that frequency that makes the denominator of (a) and (b) equal to zero, that is for

$$\omega_0 = B_0\frac{e}{m} \tag{17-27}$$

An approximate value for the earth's magnetic field is 0.5 gauss which in MKS units corresponds to 0.5×10^{-4} webers/sq m.

Since $e/m \doteq 1.77 \times 10^{10}$ coulombs/kg for an electron, eq. (27) gives a resonant or *gyrofrequency* of about 1400 kc for electrons. For the hydrogen ion with a mass approximately 1800 times that of the electron, the gyro frequency is outside the radio frequency spectrum, being of the order of 800 cps. Thus, as far as the influence of the earth's magnetic field on radio-frequency wave propagation is concerned, only the electrons in the ionosphere need be considered.

The result of this resonance condition for electrons in the ionosphere is that near the gyro frequency the velocities of the electrons increase to large values for a given intensity of electric field, and the wave attenuation is increased greatly over the value existing in the absence of the earth's magnetic field. The practical result is that the absorption of radio waves in the ionosphere as a function of their frequency reaches a broad maximum in the neighborhood of 1400 kc, so that over most of the broadcast band and up to about 2 mc, the absorption of the wave is too large for daytime sky wave reception. Outside of this frequency band the attenuation is relatively much smaller, the order of magnitude being given by the expressions developed in the previous section. A general quantitative discussion of the effect of the earth's magnetic field of an electromagnetic wave of any frequency traveling through the ionosphere can become quite complicated. However, it is possible to treat certain typical particular cases,* the results of which serve to explain many of the observed phenomena.

Maxwell's first equation for a region in which there is a convection current density $N e\mathbf{v}$ is:

$$\text{Curl}\,\mathbf{H} = \epsilon_v \dot{\mathbf{E}} + Ne\mathbf{v} \tag{17-28}$$

In terms of the three scalar components and using eqs. (26), eq. (28) becomes

$$\begin{aligned}
\text{Curl}_x H &= \epsilon_v \left(1 + \frac{Ne^2}{\epsilon_v m(\omega_0^2 - \omega^2)}\right) \dot{E}_x - \frac{jNe^2\omega_0}{\omega m(\omega_0^2 - \omega^2)} \dot{E}_y \\
\text{Curl}_y H &= \epsilon_v \left(1 + \frac{Ne^2}{\epsilon_v m(\omega_0^2 - \omega^2)}\right) \dot{E}_y + \frac{jNe^2\omega_0}{\omega m(\omega_0^2 - \omega^2)} \dot{E}_x \qquad (17\text{-}29) \\
\text{Curl}_z H &= \epsilon_v \left(1 - \frac{Ne^2}{\epsilon_v m\omega^2}\right) \dot{E}_z
\end{aligned}$$

* The cases considered in this section were treated by W. Nichols and J. C. Schelleng, "Propagation of Electric Waves over the Earth," *BSTJ*, 4, 215, April, 1925.

In the above equations E_x has been replaced by $\dot{E}_x/j\omega$, etc., and $m/e = B_0/\omega_0$. Assuming for N a value approximately equal to N_{max} and using eq. (20), the quantity $Ne^2/\epsilon_v m$ is equal to ω_c^2, where ω_c is 2π times the critical frequency for the layer having that particular maximum value of N. Making this substitution, eqs. (29) can be written in the following form

$$\left.\begin{aligned} \text{Curl}_x\, H &= \epsilon'\dot{E}_x - j\epsilon''\dot{E}_y && \text{(a)} \\ \text{Curl}_y\, H &= \epsilon'\dot{E}_y + j\epsilon''\dot{E}_x && \text{(b)} \\ \text{Curl}_z\, H &= \epsilon^0\dot{E}_z && \text{(c)} \end{aligned}\right\} \quad (17\text{-}30)$$

where

$$\left.\begin{aligned} \epsilon^0 &= \epsilon_v\left(1 - \frac{\omega_c^2}{\omega^2}\right) \\ \epsilon' &= \epsilon_v\left(1 - \frac{\omega_c^2}{\omega^2 - \omega_0^2}\right) \\ \epsilon'' &= -\epsilon_v\left(\frac{\omega_c^2\omega_0}{\omega(\omega^2 - \omega_0^2)}\right) \end{aligned}\right\} \quad (17\text{-}31)$$

In addition to the above relations, another set which will be useful can be obtained by taking the curl of Maxwell's emf equation.

$$\text{curl curl } \mathbf{E} = \text{grad div } \mathbf{E} - \nabla^2\mathbf{E} = -\mu \text{ curl } \dot{\mathbf{H}} \quad (17\text{-}32)$$

From this,

$$\nabla^2\mathbf{E} - \text{grad div } \mathbf{E} = \mu \text{ curl } \dot{\mathbf{H}}$$

which can be written as the three scalar equations

$$\left.\begin{aligned} \nabla^2 E_x - \frac{\partial}{\partial x}(\text{div } \mathbf{E}) &= \mu \text{ curl}_x\, \dot{\mathbf{H}} \\ \nabla^2 E_y - \frac{\partial}{\partial y}(\text{div } \mathbf{E}) &= \mu \text{ curl}_y\, \dot{\mathbf{H}} \\ \nabla^2 E_z - \frac{\partial}{\partial z}(\text{div } \mathbf{E}) &= \mu \text{ curl}_z\, \dot{\mathbf{H}} \end{aligned}\right\} \quad (17\text{-}33)$$

It is now possible to consider separately two particular cases that are of interest.

CASE I: *Direction of Propagation Parallel to* $\mathbf{B}_0$

For this case, a uniform plane electromagnetic wave is assumed to be traveling parallel to $\mathbf{B}_0$ in the z direction. Remembering that under these conditions, $E_z = H_z = 0$, and $\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0$, eqs. (30) and (33) can be combined to yield

$$\left.\begin{aligned}\frac{\partial^2 E_x}{\partial z^2} &= \mu\epsilon'\ddot{E}_x - j\mu\epsilon''\ddot{E}_y \\ \frac{\partial^2 E_y}{\partial z^2} &= \mu\epsilon'\ddot{E}_y + j\mu\epsilon''\ddot{E}_x\end{aligned}\right\} \tag{17-34}$$

These are the relations which the electromagnetic wave traveling through the ionosphere parallel to $\mathbf{B}_0$ must satisfy. A general solution for a uniform plane wave progressing in the z direction has the form

$$\begin{aligned}E_x &= A\, e^{j\omega\left(t - \frac{z}{v}\right)} \\ E_y &= B\, e^{j\omega\left(t - \frac{z}{v}\right)}\end{aligned} \tag{17-35}$$

where A and B may be complex. Then

$$\ddot{E}_x = -\omega^2 A\, e^{j\omega\left(t - \frac{z}{v}\right)} \qquad \ddot{E}_y = -\omega^2 B\, e^{j\omega\left(t - \frac{z}{v}\right)}$$

$$\left.\begin{aligned}\frac{\partial^2 E_x}{\partial z^2} &= \frac{\omega^2}{v^2} A\, e^{j\omega\left(t - \frac{z}{v}\right)} \\ \frac{\partial^2 E_y}{\partial z^2} &= -\frac{\omega^2}{v^2} B\, e^{j\omega\left(t - \frac{z}{v}\right)}\end{aligned}\right\} \tag{17-36}$$

Substituting (36) in (34) and dividing through by

$$-\omega^2\, e^{j\omega\left(t - \frac{z}{v}\right)}$$

gives

$$\left.\begin{aligned}\left(-\frac{1}{v^2} + \mu\epsilon'\right) A &= j\mu\epsilon'' B \\ \left(-\frac{1}{v^2} + \mu\epsilon'\right) B &= -j\mu\epsilon'' A\end{aligned}\right\} \tag{17-37}$$

There are two sets of solutions to (37). They are

Solution 1:

$$\begin{aligned}B &= jA \\ v = v_1 &= \frac{1}{\sqrt{\mu(\epsilon' + \epsilon'')}}\end{aligned} \tag{17-38}$$

Solution 2:

$$\begin{aligned}B &= -jA \\ v = v_2 &= \frac{1}{\sqrt{\mu(\epsilon' - \epsilon'')}}\end{aligned} \tag{17-39}$$

Assuming A is real and using only the real parts of eqs. (35), the expressions for the electric field are:

Solution 1:

$$\left.\begin{aligned} E_x &= A \cos \omega\left(t - \frac{z}{v_1}\right) \\ E_y &= -A \sin \omega\left(t - \frac{z}{v_1}\right) \end{aligned}\right\} \tag{17-40}$$

Solution 2:

$$\left.\begin{aligned} E_x &= A \cos \omega\left(t - \frac{z}{v_2}\right) \\ E_y &= A \sin \omega\left(t - \frac{z}{v_2}\right) \end{aligned}\right\} \tag{17-41}$$

Eqs. (40) and (41) represent two circularly polarized waves, rotating in opposite directions and traveling with different velocities. From eq. (38), (39), and (31) the two velocities are

$$\left.\begin{aligned} v_1 &= \frac{c}{\sqrt{1 - \dfrac{\omega_c^2}{\omega^2 - \omega_0^2} - \dfrac{\omega_c^2\omega_0}{\omega(\omega^2 - \omega_0^2)}}} = \frac{c}{\sqrt{1 - \dfrac{\omega_c^2}{\omega(\omega - \omega_0)}}} \\ v_2 &= \frac{c}{\sqrt{1 - \dfrac{\omega_c^2}{\omega^2 - \omega_0^2} + \dfrac{\omega_c^2\omega_0}{\omega(\omega^2 - \omega_0^2)}}} = \frac{c}{\sqrt{1 - \dfrac{\omega_c^2}{\omega(\omega + \omega_0)}}} \end{aligned}\right\} \tag{17-42}$$

CASE II: *Direction of Propagation Perpendicular to* $\mathbf{B}_0$

In this case, the uniform plane electromagnetic wave will be assumed to be traveling in a direction perpendicular to the magnetic field, say in the x direction. In the general case the electric field will be arbitrarily polarized, so that there will be both E_y and E_z components. An important special case of Case II occurs when the electric vector is entirely in the z direction, that is parallel to $\mathbf{B}_0$. In this special case, only eq. (30c) applies. This corresponds to eq. (4), which was obtained previously when the earth's magnetic field was neglected. (In the present case, both σ and ν which appear in (4) are assumed to be zero). Thus, as would be expected, the earth's magnetic field $\mathbf{B}_0$ has no influence on ionospheric propagation in the particular case where $\mathbf{E}$, and therefore the electron velocity $\mathbf{v}$, is parallel to $\mathbf{B}_0$.

Now consider the more general case of propagation in the x direction (perpendicular to $\mathbf{B}_0$) when the electric field of the incident wave may have both polarizations (E_y and E_z). Remembering that for a uniform plane wave in the x direction, $\partial/\partial y = \partial/\partial z = 0$, eqs. (33) reduce to

$$\left.\begin{aligned} \frac{\partial^2 E_x}{\partial x^2} - \frac{\partial^2 E_x}{\partial x^2} &= 0 = \mu \operatorname{curl}_x \dot{\mathbf{H}} \\ \frac{\partial^2 E_y}{\partial x^2} &= \mu \operatorname{curl}_y \dot{\mathbf{H}} \\ \frac{\partial^2 E_z}{\partial x^2} &= \mu \operatorname{curl}_z \dot{\mathbf{H}} \end{aligned}\right\} \tag{17-33a}$$

Combining the first of eqs. (33a) with eq. (30a) shows that

$$\ddot{E}_x = j \frac{\epsilon''}{\epsilon'} \ddot{E}_y \tag{17-43}$$

For a uniform plane wave propagating through free space in the x direction, the wave is transverse and $E_x = 0$. However, in this case of propagation through an ionized medium in the presence of the earth's magnetic field, the electron convection current in the x-y plane results in a component (usually small) of electric intensity in the x direction, the magnitude of which can be obtained from (43).

Again from eq. (30b).

$$\nabla^2 \epsilon_y = \mu\epsilon' \ddot{E}_y + j\mu\epsilon'' \ddot{E}_x \tag{17-44}$$

and, using (43), this becomes

$$\nabla^2 E_y = \mu \left(\epsilon' - \frac{(\epsilon'')^2}{\epsilon'} \right) \ddot{E}_y \tag{17-45}$$

Also from eq. (30c),

$$\nabla^2 E_z = \mu\epsilon^0 \ddot{E}_z \tag{17-46}$$

Equations (45) and (46) indicated that for waves propagating through the ionosphere perpendicular to the earth's magnetic field, there will be two different velocities depending upon whether **E** is parallel to, or perpendicular to, the magnetic field. For the component of **E** parallel to $\mathbf{B}_0$, already considered as a special case, the velocity of wave propagation is obtained from (46) as

$$v_0 = \frac{1}{\sqrt{\mu\epsilon^0}} = \frac{1}{\sqrt{1 - \dfrac{\omega_c{}^2}{\omega^2}}} \tag{17-47}$$

This is the same velocity that was obtained when the presence of the earth's magnetic field was neglected. For E_y, the component of **E** perpendicular to $\mathbf{B}_0$, the wave velocity, obtained from (45) is

$$v_3 = \frac{1}{\sqrt{\mu \left(\epsilon' - \dfrac{(\epsilon'')^2}{\epsilon'} \right)}} = \frac{c}{\sqrt{\dfrac{1}{\epsilon_0} \left(\epsilon' - \dfrac{(\epsilon'')^2}{\epsilon'} \right)}} \tag{17-48}$$

In general, an electromagnetic wave propagating perpendicular to the earth's magnetic field will have components of **E** both parallel and perpendicular to $\mathbf{B}_0$. These components will travel with different velocities and will therefore be refracted differently in the ionosphere. Therefore two waves traveling different paths with different velocities will result. One of these has the same velocity as would be obtained if the earth's field were not present, and this is called the *ordinary wave* or *ordinary ray*. The other travels with a different velocity and is termed the *extraordinary ray*. The extraordinary ray suffers greater absorption (at high frequencies) and has a somewhat higher critical frequency than the ordinary ray.

17.08 Transmission Line Representation of the Ionosphere. When all the variables are considered, the phenomenon of ionosphere reflection is one of great complexity. In such a circumstance the engineer often finds it helpful to have available a simplified picture of the phenomenon that will indicate to him the order of magnitude (or at least the correct direction of change) of the dependent variable when one of the independent variables is varied. Figure 17-9 is a simplified equivalent-circuit representation of the ionosphere, which accounts for some of the major facts of ionosphere reflection. As with any equivalent circuit, the equivalence holds only over a restricted range of operating conditions, but within that range it may be used to give useful answers.

In Fig. 17-9a wave propagation through free space and in the ionosphere is compared to wave propagation along a transmission line having "constants" that vary along the length of the line. In the region beneath the ionosphere, the free-space constants μ_v and ϵ_v are the equivalent series inductance and shunt capacitance per unit length of the "free-space transmission line." Electron and ion convection currents flow within the ionosphere in response to the electric field of the electromagnetic wave. In the absence of collisions and neglecting the effect of the earth's magnetic field, these convection currents lag the impressed electric intensity by 90 degrees. The effect of the ionization is therefore correctly represented in the transmission line analogue as a shunt inductance L_e. In the region where collisions between electrons and gas molecules result in appreciable absorption of the wave, a resistance R_e is included in series with L_e to account for this power loss due to the convection current flow. (It could be accounted for instead by a shunt conductance G.) Shunt inductance added to a transmission line tends to neutralize or reduce the shunt capacitance of the line

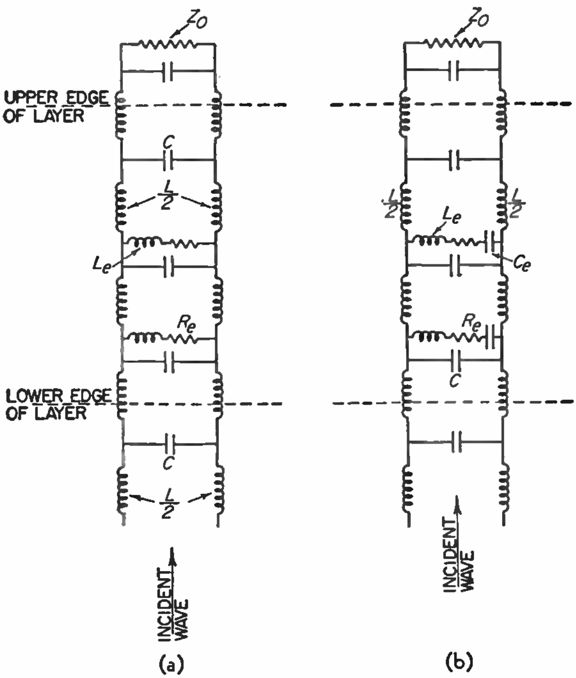


FIG. 17-9. Approximate equivalent circuit of the ionosphere, (a) neglecting, and (b) including the effect of the earth's magnetic field. $L = \mu_v$, $C = \epsilon_v$, $Z_0 = \sqrt{\mu_v/\epsilon_v}$, $L_e = \dfrac{m}{Ne^2}$, $R_e = \dfrac{m\nu}{Ne^2}$, $C_e = \dfrac{Nm}{B_0}$ (for extraordinary wave), $C_e = \infty$ (for ordinary wave).

and hence increases the phase velocity along the line as indicated by the simple relation for the lossless line

$$v = \frac{1}{\sqrt{LC_1}}$$

where for this case

$$C_1 = C - \frac{1}{\omega^2 L_e} \tag{17-49}$$

When the shunt capacitance C is completely neutralized by the shunt inductive loading, that is when $C_1 = 0$, the phase velocity becomes infinite. This means the phase shift per unit length has been reduced to zero, or in other words, wave motion has ceased.

Therefore the point where

$$L_e = \frac{1}{\omega^2_c} \tag{17-50}$$

is the farthest point to which the wave can progress, and complete reflection (in the lossless case) of the wave occurs at this point. Substitution for L_e and C indicates that the ionization density at this point must satisfy the relation

$$\omega^2 = \frac{Ne^2}{\epsilon_v m} \tag{17-51}$$

The frequency for which (51) is satisfied is by definition the critical frequency for the layer that has this value of N for its maximum ionization density.

In Fig. 17-9b the effect of the earth's magnetic field has been represented (but only approximately) as a condenser C_e in series with L_e and R_e. The capacitance of the condenser is inversely proportional to the square of the magnetic field strength and would be infinite for no magnetic field, or for propagation under circumstances where the magnetic field has no effect. At the frequency for which L_e and C_e are in series resonance,

$$\omega = \omega_0 = \frac{1}{\sqrt{L_e C_e}} = B_0 \frac{e}{m} \tag{17-52}$$

the convection current rises to a large value limited only by the collision losses; the absorption is abnormally large at and near this resonant frequency. At frequencies considerably above the resonant frequency, the absorption decreases to values just a little larger than would be obtained without the magnetic field. At frequencies below the resonant frequency the loss is somewhat less than without the magnetic field. In addition, the presence of the earth's magnetic field as represented by C_e has the effect of increasing the critical frequency for that wave polarization that is affected by it, that is, for the extraordinary wave.

In both of these representations vertical incidence has been assumed. In general the line "constants" are also functions of angle of incidence, and this fact would have to be considered if the simplified transmission line picture were used at various angles of incidence.

17.09 Sky Wave Transmission Calculations.

Maximum Usable Frequency. For ionospheric transmission to be possible, the frequency used must lie between the maximum usable frequency (*muf*) and the lowest useful high frequency (*luhf*). The

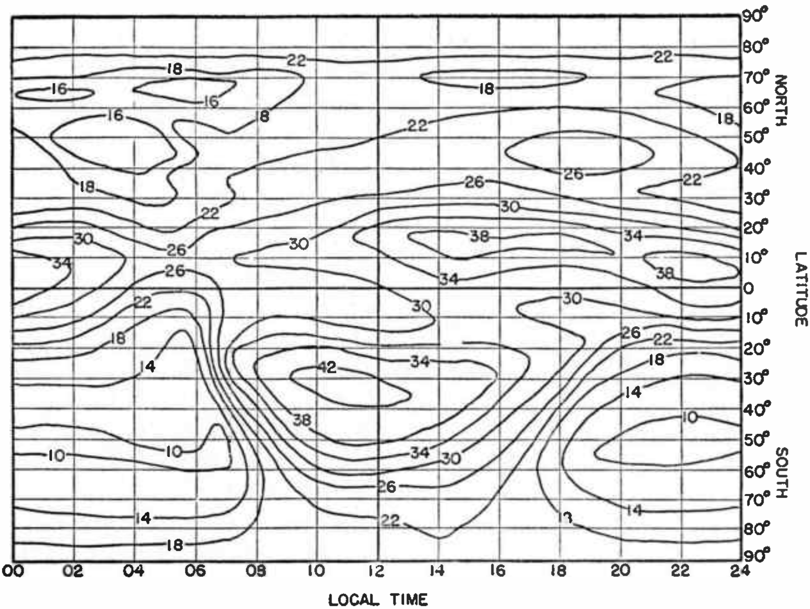


FIG. 17-10. Typical world contour *muf* chart for the F_2 layer. Chart is for distances of 4000 km and summertime conditions.

muf was defined in section 17.04, and the *luhf* will be discussed later in this section.

The *muf* for any given path at any time of day is calculated quite simply through use of convenient world contour charts obtainable from the Bureau of Standards in Washington. These charts, an example of which appears in Fig. 17-10, show the world-wide variation of *muf* with local time for all latitudes. The *muf* figures predicted on these charts are the monthly *median* values of maximum usable frequency. Thus, communication at the *muf* calculated from these charts should be effective approximately 50 per cent of the time during undisturbed periods.

To calculate the *muf* it is necessary to know the length of the transmission path and the location of certain "control" points on the path. The control points of a transmission path are points along the path, the ionospheric conditions of which seem to control transmission along the path. For paths shorter than 4000 kilometers (that is, single-hop paths) the control point is midway along the path, as would be expected. For longer paths the control points are taken as 2000 kilometers from each end for F_2 layer reflection and 1000 kilometers from each end for E layer reflection. Although the choice is empirical, it can be justified to some extent from a consideration of the probable paths in multihop transmission.

To calculate the maximum usable frequency the appropriate *muf* chart is used in conjunction with a world map and a great-circle chart of the same size. A sheet of transparent paper is placed over the world map, and on it are marked the location of the transmitting and receiving points and the equatorial line. The transparent paper is then placed on the great-circle chart. Keeping the equatorial lines on chart and paper lined up, the transparency is moved sideways until the transmitting and receiving points both lie on the same great circle line, which is then sketched in. The transparency is then placed over the *muf* chart, and the meridian whose local time is to be used for the calculation is lined up with the appropriate time meridian on the *muf* chart. Since 24 hours on the time scale of the *muf* chart is drawn to the same scale as 360 degrees of longitude on the world map, all points on the great-circle path will be lined up in their proper local time relationship. The maximum usable frequency at the control point or points can then be read off directly if the path length is the same as that for which the *muf* chart is drawn. For F_2 layer transmission over distances less than 4000 kilometers (single-hop transmission) the maximum usable frequencies are determined from zero-distance and 4000 kilometers distance *muf* charts; then the maximum usable frequency corresponding to the actual path length is obtained by interpolation with the aid of a suitable nomogram. For transmission via other layers and over greater distances the calculation procedure is slightly different. Details of these calculations, along with a complete set of sample charts and worked examples are given in a Bureau of Standards publication.* *Muf* prediction charts are also available from the same source.†

The frequencies selected by the procedures above are for undisturbed periods. During periods of "ionospheric storms" the critical frequencies are lower than usual, and it may be necessary to lower the working frequency in order to insure communication. As the frequency is lowered the absorption of the wave increases. If it is necessary to lower the operating frequency below the lowest useful high frequency, communication becomes impossible. Since the "ionospheric storm" type of disturbance is most severe in the polar regions, with less severity towards the equator, communication can often be maintained during "storm" periods by relaying through points closer to the equator. This last statement does not apply in the case of "sudden ionospheric disturbances" or "radio fadeouts" (Dellinger effect). During radio fadeouts high-frequency communication becomes impossible on all paths in the daylight side of the world. Fortunately this latter type of disturbance, which is unpredictable, rarely lasts more than two hours.

Sky-wave Absorption and Lowest Useful High Frequency. As has already been pointed out, when the operating frequency is reduced from the maximum usable frequency the absorption of the wave in the ionosphere increases, and the received signal strength becomes less. The *lowest-useful-high-frequency* (*luhf*) for a given distance and given transmitter power is defined as the lowest frequency (in the high-frequency band) that will give satisfactory reception for that distance and power. Unlike the *muf*, which depends only upon the state of the ionosphere and the distance between transmitting and receiving points, the *luhf* depends upon the following factors:

(a) The effective radiated power.

(b) The absorption characteristics of the ionosphere for the paths between transmitter and receiver.

* *Ionospheric Radio Propagation*, National Bureau of Standards Circular 462, issued June 1948.

† Series CRPL-D, *Basic Radio Propagation Predictions Three Months in Advance*. These bulletins are prepared by Central Radio Propagation Laboratory and are for sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D. C.

(c) The *required* field intensity, which in turn depends upon radio noise at the receiving location and the type of service involved. These factors will be considered in turn.

(a) *Effective Radiated Power and Unabsorbed Field Intensity.* The *effective radiated power* is the power actually radiated by the antenna, multiplied by the antenna gain in the direction of propagation. This second factor requires a knowledge of the vertical angle of radiation that is effective in producing a signal at the receiver. This angle depends upon the layer involved, the distance to the receiver, and the number of hops. The *unabsorbed field intensity* of a sky wave signal at a given distance for a transmitter is defined as the median incident field intensity that would be observed by use of an antenna of fixed linear polarization if no absorption were introduced by the ionosphere. The unabsorbed field intensity is less than that which would result from inverse-distance attenuation alone because of (1) interference and polarization fading and (2) loss of energy upon reflection at the ground between hops. In practice the unabsorbed field intensity for any distance is obtained from a graph that takes the above factors into account.

(b) *Absorption Characteristics of the Ionosphere.* Absorption in the ionosphere can be classified as deviative or nondeviative absorption. *Deviative Absorption* occurs in that region of the ionosphere where the wave is bent back to earth. Except for frequencies near the critical frequency for the reflecting layer this type of absorption is small. During daylight hours a much greater absorption of the wave occurs in the *D* region, where the collision frequency is high. This latter absorption is called *nondeviative* because it is not associated with a bending of the wave. In the *D* region recombination is rapid and the ionization density, and hence the absorption, varies almost in synchronism with the elevation of the sun. *D*-layer absorption is a maximum at noon and decreases to negligibly small values within two hours after sunset. As was pointed out in sec. 17.07, the absorption has a broad maximum in the neighborhood of the resonance frequency (1.2–1.4 mc) for the electrons in the earth's magnetic field. As the frequency is increased above the resonance frequency, the absorption decreases steadily except for frequencies close to the critical frequency of each layer.

In addition to frequency and time of day, sky-wave absorption is also dependent upon the season of year and sunspot activity. When all of these factors have been taken into account, the absorption of the wave in the ionosphere, and hence the expected incident field intensity, can be calculated. As in the case of the *muf*, such calculations are greatly facilitated by the use of numerous charts and nomograms which have been prepared, and which are available.*

(c) *Required Field Intensity.* The field intensity required for satisfactory reception for a given type of service depends among other things on the receiver sensitivity, the receiving-set noise, the radio noise level prevailing at the receiving location, and the type of modulation. Radio "noise" can be divided into *man-made noise,* that is, local electrical disturbances produced by electrical machinery, and *atmospheric noise,* or static. The latter noise depends upon the frequency, the time of day, the season of year, and location with respect to the sources of thunderstorms. Because most atmospheric noise has its origin in thunderstorms, those areas in which thunderstorms are most prevalent will have the highest atmospheric-noise level. The world can be divided into noise zones that correspond roughly to the zones showing the incidence of thunderstorms. The principal noise centers or active thunderstorm areas are located in Central Africa, Central America, and the East and West Indies. Areas of very low thunderstorm incidence are the north and south frigid zones. In general the temperate zones are areas of moderate thunderstorm incidence. The actual atmospheric noise at any location depends upon the local noise sources (thunderstorms) and also upon the sky-wave propagation characteristics between that location and the principal noise centers. It should be noted that the same factors that make for good transmission of radio signals from distant transmitters also provide good transmission of noise signals from distant noise sources.

The distribution of noise intensities throughout the world has been plotted on world maps for each month of the year. Using these noise-grade charts in conjunction with curves that show required field intensity *vs.* frequency for different times of day and different noise grades, it is a simple matter to figure the required

* Such charts are available in Bureau of Standards Circular 462 and in *Elementary Manual of Radio Propagation*, by Donald Menzel, published by Prentice-Hall, Inc., New York, 1948. The Manual gives simple, detailed procedures for the calculation of all important quantities.

field intensity for any given set of conditions.* For any frequency less than the *muf*, if the incident field intensity calculated under (b) is greater than the required field intensity calculated under (c), communication can be established. The lowest frequency for which this occurs is the lowest useful high frequency.

It is seen that the calculation of the probable received field intensity at any point is an engineering problem that can be solved when sufficient data are given. The data required are the time of day, the season of year, the transmission path, the frequency, and the effective radiated power. Although the ionosphere is a complex natural phenomenon not under the control of man, by familiarizing himself with its characteristics man has learned to predict its behaviour and so has been enabled to use it to serve his needs.

BIBLIOGRAPHY

Kirby, S. S., L. V. Berkner, D. M. Stuart, "Studies of the Ionosphere and their Application to Radio Transmission," *Proc. IRE*, **22**, 4, 481 (1934).

Darrow, Karl K., "Analysis of the Ionosphere," *BSTJ*, **19**, 455 (1940).

Dellinger, J. H., "The Role of the Ionosphere in Radio Wave Propagation," *Trans. AIEE Supplement*, **58**, 803 (1939).

Namba, Shogo, "General Theory on the Propagation of Radio Waves in the Ionized Layer of the Upper Atmosphere," *Proc. IRE*, **21**, 238 (1933).

Baker, W. G. and C. W. Rice, "Refraction of Short Waves in the Upper Atmosphere," *Trans. AIEE*, **45**, 302 (1926).

Mimno, H. R., "The Physics of the Ionosphere," *Review of Modern Physics*, Jan., 1943.

Nichols, W., and J. C. Schelleng, "Propagation of Electric Waves over the Earth," *BSTJ*, **4**, 215, 1925.

Books

Ionospheric Radio Propagation, National Bureau of Standards Circular 462, U. S. Govt. Printing Office, Washington, D. C.

Menzel, Donald H., *Elementary Manual of Radio Propagction*, Prentice-Hall, Inc., New York, 1948.

Propagation of Radio Waves, Summary Technical Report of N.D.R.C. CP Vol. 3, Columbia University Press, 1946.

Pedersen, P. O., *The Propagation of Radio Wcves*, G. E. C. Gad, Copenhagen, 1929.

* The required field intensity curves and noise-grade charts are available in the references already mentioned.

APPENDIX I

VELOCITIES OF PROPAGATION

Group Velocity and Phase Velocity. In the discussion of wave propagation between parallel planes and in wave guides two different velocities were encountered. The first of these was the phase velocity v, which represented the velocity of propagation of equiphase surfaces along the guide. The second velocity was the group velocity v_g, which in those particular cases, could be considered as the velocity of energy propagation in the direction of the axis of the guide. For wave guide propagation the phase velocity is always greater than $v_0 = 1/\sqrt{\mu\epsilon}$, whereas the group velocity is always less than v_0. The term group velocity has a more general significance than was indicated in that discussion.

In order to convey intelligence it is always necessary to modulate by some means or other the carrier frequency being transmitted. When this is done, there is a group of frequencies, usually centered about the carrier, that must be propagated along the guide or transmission line. If the phase velocity is a function of frequency, the waves of different frequencies in the group will be transmitted with slightly different velocities. The component waves combine to form a "modulation envelope," which is propagated as a wave having the group velocity v_g defined by

$$v_g = \frac{d\omega}{d\beta}$$

The frequency spread of the group is assumed to be small compared with the mean frequency of the group, and the derivative is evaluated at this mean frequency. The significance of the definition is made clear by consideration of a simple and well-known example.

Consider the case of a carrier $E_0 \cos \omega t$, amplitude-modulated by a modulation frequency $\Delta f = \Delta\omega/2\pi$. Such a signal would be represented by

$$E = E_0(1 + m \cos \Delta\omega t) \cos \omega t$$

where m is the modulation factor. This expression can be expanded in the usual manner to show the presence of the carrier and side band frequencies.

$$E = E_0 \cos \omega t + \frac{mE_0}{2} [\cos (\omega + \Delta\omega)t + \cos (\omega - \Delta\omega)t]$$

If now such a signal is propagated in the z direction under conditions where the phase velocity varies with frequency, the resultant wave would be written as

$$E = E_0 \cos (\omega t - \beta z) + \frac{mE_0}{2} \{\cos [(\omega + \Delta\omega)t - (\beta + \Delta\beta)z] + [\cos (\omega - \Delta\omega)t - (\beta - \Delta\beta)z]\}$$

This expression can be recombined to show an amplitude-modulated wave progressing in the z direction

$$E = E_0[1 + m \cos (\Delta\omega t - \Delta\beta z)] \cos (\omega t - \beta z)$$

The bracketed part of this expression represents the envelope of the wave. It is seen that the envelope progresses in the z direction with a velocity

$$v_g = \frac{\Delta\omega}{\Delta\beta}$$

If the frequency spread of the group is small enough that $\Delta\omega/\Delta\beta$ may be considered constant throughout the group, this may be written as the limit

$$v_g = \frac{d\omega}{d\beta}$$

To simplify the evaluation of v_g, this may also be written as

$$v_g = \frac{1}{d\beta/d\omega}$$

The phenomenon of phase and group velocities can be illustrated by sketching the addition of the two side band frequencies waves at a certain instant of time (the carrier frequency is omitted to simplify the discussion and the sketch). Figure 1 shows two waves of slightly different frequencies combining to form a single amplitude-modulated wave. If the component waves have the same velocity, the two crests a_1 and b_1 will move along together and the maximum of the modulation envelope will move along with them at the same velocity. Under these circumstances, phase and group velocity are the same. If it is assumed that the lower frequency wave b with the longer wavelength has a velocity slightly greater than that of a, the crests a_1 and b_1 will move apart and the crests a_2 and b_2 will come together. Therefore, at some later instant of time the maximum of the envelope will occur at the point where a_2 and b_2 are coincident, and at a still later instant where a_3 and b_3 are coincident. It is evident that the envelope is slipping backward with respect to the component waves. In other words, it is moving forward with the group velocity v_g, which is less than the phase velocit

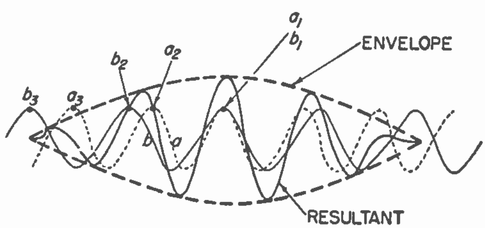


FIG. 1

of either of the component waves. Visually, as for example, in the case of water waves, it appears as though the envelope were slipping behind the component waves, or, on the other hand, as though the component waves were slipping forward through the envelope. Under those conditions, where the shorter wavelength (high frequency) wave has the greater phase velocity, the situation is reversed and the modulation envelope slips forward. The group velocity is then greater than the phase velocity of the component waves. If β is plotted as a function of ω, the phase velocity and group velocity may be determined directly from the graph. Figure 2 shows such a plot for waveguide propagation. It is observed that the slope of $\beta/\omega = 1/\bar{v}$ is always less than that of $d\beta/d\omega = 1/v_g$, so that $\bar{v}$ is always greater than v_g, but both approach v_0 as ω approaches infinity.

Figure 3 shows a typical plot of β *vs.* ω for a two-conductor transmission line having loss. For this case the slope of $1/\bar{v}$ is always greater than that of $d\beta/d\omega = 1/v_g$, so that the phase velocity i always less than the group velocity, but as the frequency is increased both velocities approach the velocity $v_0 = 1/\sqrt{LC}$, which applies in the lossless case.

Signal Velocity. It is to be noted that both phase velocity and group velocity are terms that apply only under *steady-state* conditions. If a signal be impressed suddenly at one end of a transmission line or wave guide, the time required for the disturbance to reach the other end is a measure of what is sometimes called the *signal velocity*. However, it is difficult to state just what is the

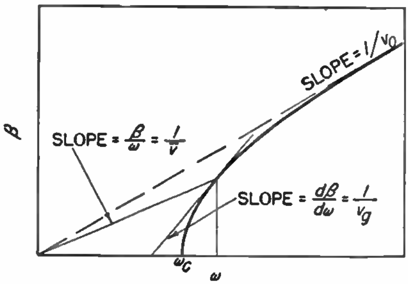


FIG. 2

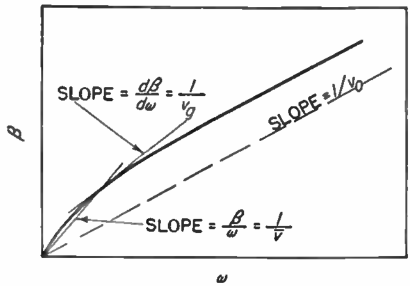


FIG. 3

value of this signal velocity, because the signal at the other end builds up more or less gradually to a steady state as the initial transient condition dies out. The first impulse always reaches the receiving end with a velocity equal to the velocity of light, with other impulses arriving at later times. However, the amplitude of the first impulse is zero and the build-up to the steady-state condition is gradual, so the time required for the signal to reach (and be indicated by) the detector is dependent on the sensitivity of the detector. A thorough discussion of this rather complex phenomena has been given by Brillouin.*

* Brillouin, Congres international d'electricite, Vol. II, Paris, 1932. *See also,* Sarbacher and Edson, *Hyper and Ultra-High Frequency Engineering,* John Wiley and Sons, New York, 1943.

APPENDIX II

BESSEL FUNCTIONS

Bessel and Hankel functions have been discussed briefly in various parts of the text (secs. 9.05, 11.08, 13.05). In working electromagnetic problems the series expansions of the functions, and a knowledge of the differentiation and recurrence formulas are often required. For convenience a few of these series and formulas are assembled together here. For a more detailed treatment of these functions, reference should be made to one of the standard texts on Bessel functions.*

Bessel Functions and Hankel Functions. Bessel functions are solutions of the following differential equation, which is known as Bessel's equation

$$z^2 \frac{d^2w}{dz^2} + z\frac{dw}{dz} + (z^2 - \nu^2)w = 0 \qquad (1)$$

The *order* of the equation is given by the value of ν. In general ν will be nonintegral. For integral values of ν the symbol n is usually used.

For nonintegral values of ν two linearly independent solutions of (1) are

$$\left.\begin{aligned} J_\nu(z) &= \sum_{m=0}^{\infty} \frac{(-1)^m z^{\nu+2m}}{m!\Gamma(m+\nu+1)2^{\nu+2m}} \\ J_{-\nu}(z) &= \sum_{m=0}^{\infty} \frac{(-1)^m z^{-\nu+2m}}{m!\Gamma(m-\nu+1)2^{-\nu+2m}} \end{aligned}\right\} \qquad (2)$$

where $J_\nu(z)$ is Bessel function of the first kind, of order ν. The function $\Gamma(m + \nu + 1) = \Gamma(p)$ is the generalized factorial function

* In addition, an excellent summary treatment of Bessel Functions is given by S. A. Schelkunoff, *Electromagnetic Waves*, Chapter III: *also Applied Mathematics for Engineers and Scientists*, Chapter XX. Where differences in definitions and notations exist among various texts, those used by Schelkunoff have been followed here.

known as the *Gamma function.* It is defined by

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} \, dx$$

When $\nu = n$ (an integer), the Gamma function becomes the factorial $\Gamma(m + n + 1) = (m + n)!$ The two solutions given by (2) are then no longer independent, but instead are related by

$$J_{-n}(z) = (-1)^n J_n(z)$$

where now

$$J_n(z) = \sum_{m=0}^{\infty} \frac{(-1)^m z^{n+2m}}{2^{n+2m} m!(m + n)!} \tag{3}$$

A second independent solution of (1) is defined by

$$N_\nu(z) = \frac{J_\nu(z) \cos \nu\pi - J_{-\nu}(z)}{\sin \nu\pi} \tag{4}$$

where $N_\nu(z)$ is known as a Neumann function or, more commonly, as a Bessel function of *second kind,* of order ν. When ν is integral, $N_n(z)$ continues to be a solution of Bessel's equation, and is still defined* by (4). For integral values of ν, a complete solution of (1) is

$$w = AJ_n(z) + BN_n(z) \tag{5}$$

Bessel functions of the second kind become infinite at $z = 0$, and so cannot be used to represent physical fields except in those problems in which the region $z = 0$ is excluded.

Solutions to equation (1) may also be written in terms of *Hankel functions.* Hankel functions are linear combinations of Bessel functions defined by

$$\begin{aligned} H_\nu^{(1)}(z) &= J_\nu(z) + jN_\nu(z) \\ H_\nu^{(2)}(z) &= J_\nu(z) - jN_\nu(z) \end{aligned} \tag{6}$$

Like the N functions, Hankel functions become infinite at $z = 0$, and so in physical problems are restricted to cases where $z = 0$ is excluded.

Bessel Functions for Small and Large Arguments. For $z \ll 1$, the J and N functions are given approximately by the expressions

$$J_\nu(z) \approx \frac{z^\nu}{2^\nu \nu!} \qquad N_\nu(z) \approx -\frac{2^\nu(\nu - 1)!}{\pi z^\nu} \tag{7}$$

* When ν is an integer, n, $N_n(z)$ as defined by (4) is indeterminate. However, it can be evaluated by usual methods.

and in particular

$$J_0(z) \approx 1 \qquad N_0(z) \approx \frac{2}{\pi}(\ln z + C - \ln 2) \tag{8}$$

On the other hand for *large* values of z the asymptotic expressions are

$$\left.\begin{aligned} J_\nu(z) &\approx \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \\ N_\nu(z) &\approx \sqrt{\frac{2}{\pi z}} \sin\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \\ H_\nu^{(1)}(z) &\approx \sqrt{\frac{2}{\pi z}}\, e^{j\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)} \\ H_\nu^{(2)}(z) &\approx \sqrt{\frac{2}{\pi z}}\, e^{-j\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)} \end{aligned}\right\} \tag{9}$$

From these expansions it is apparent that the J and N functions correspond to cosine and sine functions and as such represent standing waves when used with the time factor $e^{j\omega t}$. On the other hand, the $H^{(1)}$ and $H^{(2)}$ functions correspond to exponential functions with imaginary exponents and, when used with the time factor $e^{j\omega t}$, represent inward- and outward-traveling cylindrical waves.

Differentiation and Integration of Bessel Functions. Using the series definition for $J_0(z)$, and differentiating term by term, shows that

$$\frac{d}{dz}[J_0(z)] = -J_1(z)$$

Similarly it can be shown that the following relations are true: [In the expressions listed here $Z_\nu(z)$ may denote any of the functions $J_\nu(z)$, $N_\nu(z)$, $H_\nu^{(1)}(z)$, or $H_\nu^{(2)}(z)$; also $Z_\nu'(z)$ means $(d/dz)Z_\nu(z)$].

$$\left.\begin{aligned} Z_0'(z) &= -Z_1(z) \\ Z_1'(z) &= Z_0(z) - \frac{1}{z} Z_1(z) \\ zZ_\nu'(z) &= \nu Z_\nu(z) - zZ_{\nu+1}(z) \\ &= zZ_{\nu-1}(z) - \nu Z_\nu(z) \\ \frac{d}{dz}[z^\nu Z_\nu(z)] &= z^\nu Z_{\nu-1}(z) \\ \frac{d}{dz}[z^{-\nu} Z_\nu(z)] &= -z^{-\nu} Z_{\nu+1}(z) \end{aligned}\right\} \tag{10}$$

A useful recurrence formula is

$$2\nu Z_\nu(z) = z[Z_{\nu-1}(z) + Z_{\nu+1}(z)] \tag{11}$$

Some integration formulas are

$$\left.\begin{aligned} \int Z_1(z)\,dz &= -Z_0(z) \\ \int z^\nu Z_{\nu-1}(z)\,dz &= z^\nu Z_\nu(z) \\ \int z^{-\nu} Z_{\nu+1}(z)\,dz &= -z^{-\nu} Z_\nu(z) \end{aligned}\right\} \tag{12}$$

Modified Bessel Functions. Modified Bessel functions of zero order were encountered in sec. 11.05. The modified Bessel equation of order ν is

$$z^2 \frac{d^2w}{dz^2} + z\frac{dw}{dz} - (z^2 + \nu^2)w = 0 \tag{13}$$

For nonintegral values of ν two independent solutions are

$$I_\nu(z) \qquad \text{and} \qquad I_{-\nu}(z) \tag{14}$$

where

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{z^{\nu+2m}}{2^{\nu+2m} m!\,\Gamma(\nu + m + 1)} \tag{15}$$

As in the case of the J functions, when ν is integral, these two solutions are related by

$$I_{-n}(z) = I_n(z)$$

and it becomes necessary to seek another solution. Another solution is given by

$$K_\nu(z) = \frac{\pi}{2\sin\nu\pi}[I_{-\nu}(z) - I_\nu(z)] \tag{16}$$

For integral values of n this reduces to

$$K_n(z) = \frac{2}{\cos n\pi}\left(\frac{\partial I_{-n}}{\partial n} - \frac{\partial I_n}{\partial n}\right) \tag{17}$$

and gives a second independent solution.

For $z \ll 1$, the I and K functions are given approximately by

$$I_\nu(z) \approx \frac{z^\nu}{2^\nu \nu!} \qquad K_\nu(z) \approx \frac{2^{\nu-1}(\nu-1)!}{z^\nu} \tag{18}$$

and in particular;

$$I_0(z) \approx 1 \qquad K_0(z) \approx -[\ln z + C - \ln 2] \tag{19}$$

Series expansions for the I and K functions are given by Schelkunoff.* As noted in sec. 11.07, the I and K functions are appropriate for dissipative media. For propagation in a lossless dielectric, the arguments of these functions would be pure imaginaries, and the functions reduce to ordinary Bessel functions. For imaginary arguments the relations between modified and ordinary Bessel functions are

$$\left.\begin{aligned} I_\nu(jz) &= e^{j\nu\pi/2}J_\nu(z) \\ K_\nu(jz) &= \frac{\pi}{2}e^{-j(\nu+1)\pi/2}[J_\nu(z) - jN_\nu(z)] \\ &= \frac{\pi}{2}e^{-j(\nu+1)\pi/2}H_\nu^{(2)}(z). \end{aligned}\right\} \quad (20)$$

It is apparent that for imaginary arguments the I functions represent standing waves and the K functions represent outward traveling waves.

Recurrence Formulas for I and K Functions. In general the differentiation and recurrence formulas for the modified Bessel functions are different from those for Bessel and Hankel functions. For the modified functions the recurrence formulas corresponding to those of eqs. (10) are

$$\left.\begin{aligned} zI_\nu'(z) &= \nu I_\nu(z) + zI_{\nu+1}(z), \\ zK_\nu'(z) &= \nu K_\nu(z) - zK_{\nu+1}(z) \\ zI_\nu'(z) &= zI_{\nu-1} - \nu I_\nu(z), \\ zK_\nu'(z) &= -zK_{\nu-1}(z) - \nu K_\nu(z) \\ \frac{d}{dz}[z^\nu I_\nu(z)] &= z^\nu I_{\nu-1}(z), \\ \frac{d}{dz}[z^\nu K_\nu(z)] &= -z^\nu K_{\nu-1}(z) \\ \frac{d}{dz}[z^{-\nu} I_\nu(z)] &= z^{-\nu} I_{\nu+1}(z), \\ \frac{d}{dz}[z^{-\nu} K_\nu(z)] &= -z^{-\nu} K_{\nu+1}(z) \\ 2\nu I_\nu(z) &= z[I_{\nu-1}(z) - I_{\nu+1}(z)] \\ 2\nu K_\nu(z) &= -z[K_{\nu-1}(z) - K_{\nu+1}(z)] \end{aligned}\right\} \quad (21)$$

* *Electromagnetic Waves*, pp. 50-51: *Applied Mathematics for Engineers and Scientists*, p. 396.

LIST OF SYMBOLS

(See also pages 20–22 and Table I, page 23)

The following list contains those symbols that have been used consistently throughout the text. The symbols shown on pages 20–23 are not repeated here. Because of the large number of quantities to be represented and the undesirability of using alphabets other than English and Greek, it has been necessary to use some symbols to represent different quantities at different times and places. In every instance the symbol has been defined where introduced to avoid misinterpretation of its meaning.

Symbol	*Quantity*	*Page*
$\mathbf{A}$	Vector magnetic potential	83
c	Velocity of light in free space	19
C	Euler's constant (.5772157)	320
Ci (x)	Cosine integral	320
D	Dissipation factor	129
e	Base of natural logarithms (2.71828)	131
f	Farad	24
f	Frequency	116
$\mathbf{F}$	Vector electric potential	492
F	Attenuation factor, surface wave	625
h	Simplifying factor $h = \sqrt{\bar{\gamma}^2 + \omega^2\mu\epsilon}$	178
$H_\nu^{(1)}(x)$	Hankel function, first kind, order ν	372
$H_\nu^{(2)}(x)$	Hankel function, second kind, order ν	372
$\mathbf{i}, \mathbf{j}, \mathbf{k}$,	Unit vectors, cartesian coordinates	4
$I_\nu(x)$	Modified Bessel function, first kind, order ν	373
$\hat{I}_n(x)$	Modified Bessel function, of half order	490
Im	Imaginary part of $\cdots$	170
j	Unit imaginary number in complex plane, $\sqrt{-1}$	117
$\mathbf{J}$	Linear or surface current density	108
$J_\nu(x)$	Bessel function of first kind, order ν	271

Symbol	*Quantity*	*Page*
$\hat{J}_n(x)$	Bessel function of half order	490
k	A constant	26
K	Reflection coefficient	252
K	Magnetic conduction current	556
$K_\nu(x)$	Modified Bessel function, second kind, order ν	373
$\hat{K}_n(x)$	Modified Bessel function of half order	490
l, m, n	Direction cosines	10
m, n	An integer	179
$\mathbf{M}$	Magnetic current density	556
M, N	Antenna functions	506
$\mathbf{n}$	Unit vector normal to surface	108
$\mathbf{N}$	Unit vector normal to plane surface	143
$N_\nu(x)$	Bessel function, second kind, order ν (Neumann function)	272
$\hat{N}_n(x)$	Bessel function of half order	490
$\mathbf{P}$	Poynting vector	163
P_n, Q_n	Legendre function of order n	68, 488
Q	Quality factor	291
r, θ, ϕ	Spherical coordinate axis	14
R_s	Surface resistance	157
Re	Real part of · · ·	117
$\mathbf{s}$	Distance, $d\mathbf{s}$ element of distance	35
S	Surface, $d\mathbf{a}$ element of surface area	31, 45
$S_1(x)$	Integral related to cosine integral	320
Si (x)	Sine integral	320
$T_m(x)$	Tchebyscheff polynomial of order m	440
$\mathbf{u}_\rho, \mathbf{u}_\phi, \mathbf{u}_z$	Unit vectors, cylindrical coordinates	14
$\mathbf{u}_r, \mathbf{u}_\theta, \mathbf{u}_\phi$	Unit vectors, spherical coordinates	17
v	Wave velocity in any medium	127
v_0	Velocity, uniform plane wave in lossless dielectric	114
$\bar{v}$	Phase velocity	189
v_g	Group velocity	191
V	Volume, dV element of volume	31, 41
x, y, z	Cartesian coordinate axis	4
X_s	Surface reactance	157
Y	Admittance	150
Z	Impedance, wave impedance	150, 197
Z_0	Characteristic impedance	150

Symbol	*Quantity*	*Page*
Z_s	Surface impedance	155
α	Attenuation constant	127
$\bar{\alpha}$	General attenuation constant	176
β	Phase shift constant	119
$\bar{\beta}$	General phase shift constant	176
γ	Propagation constant	127
$\bar{\gamma}$	General propagation constant	176
Δ, δ	Small increment of a quantity	11, 27
δ	Depth of penetration	131
η	Intrinsic impedance	130
η_v	Intrinsic impedance of free space	123
θ	Polar angle in spherical coordinates	17
λ	Wavelength	119
$\bar{\lambda}$	Wavelength parallel to walls of a guide	191
μ	Prefix, micro-, denoting 10^{-6}	24
ν	General order of Bessel and Hankel functions	272
ν	Relative frequency	475
π	3.14159	19
Π	A potential stream function	492
ρ, ϕ, z	Cylindrical coordinate axis	14
ψ	An angle	81
ω	Angular frequency	116
Ω	Solid angle	30

INDEX