HYPER *and* ULTRAHIGH FREQUENCY ENGINEERING

BY

ROBERT I. SARBACHER, SC.D.
Associate Professor of Electrical Engineering
Illinois Institute of Technology

AND

WILLIAM A. EDSON, SC.D.
Assistant Professor of Electrical Engineering
Illinois Institute of Technology

FOURTH PRINTING

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

IN THE REPRINTING OF THIS BOOK, THE RECOMMENDATIONS OF THE WAR PRODUCTION BOARD HAVE BEEN OBSERVED FOR THE CONSERVATION OF PAPER AND OTHER IMPORTANT WAR MATERIALS. THE CONTENT REMAINS COMPLETE AND UNABRIDGED.

FOURTH PRINTING, SEPTEMBER, 1944

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

The genius of Hertz has been demonstrated in several far-reaching results. Beginning with the announcement of his first experiments in 1887, wide acclaim has been given to his proof of the physical reality of electromagnetic waves predicted by Maxwell's mathematical reasoning twenty-eight years earlier. But only in the last decade has great practical importance been attached to the region of the spectrum in which he worked. This region, embracing hyper frequencies or micro waves, is rapidly being opened for engineering applications and vast extensions of radio communication channels.

The development of present-day hyper-frequency technique in this country may be considered to date from the original and independent work carried out by W. L. Barrow, L. J. Chu, and others at the Massachusetts Institute of Technology and by G. C. Southworth, S. A. Schelkunoff, and others at the Bell Telephone Laboratories. Many of the extensions and applications which followed this work were carried out by L. Brillouin, E. U. Condon, W. W. Hansen, R. H. and S. F. Varian, D. L. Webster, and other leading scientists and engineers in this country and abroad.

This book is intended for use by senior students of electrical engineering, and by men with equivalent training who have had at least one course in radio engineering. Mathematical complexity has been avoided, and wherever possible each problem has been approached from fundamental physical principles. Nevertheless, a considerable amount of mathematics is necessary to permit logical development of various phases presented in the book. It may be noted that certain of the mathematical sequences are rendered formidable in appearance by the inclusion of intermediate algebraic steps which are omitted in many publications. The inclusion of these additional equations should facilitate a complete understanding of the development.

Although no definitely established frequency limits are associated with the terms hyper frequency and ultrahigh frequency as applied to radio communication systems, frequencies above about 30 megacycles per second are generally referred to as ultrahigh frequencies. Originally no upper limit was associated with this term. The term microwaves is sometimes used to identify the band of frequencies beyond those known

as ultrahigh. These wavelengths cover the range from approximately 3 centimeters to 30 centimeters. We may, therefore, consider the ultrahigh-frequency band to cover frequencies lying between 30 and 1000 megacycles per second, and the hyper-frequency band to embrace those of 1000 to 10,000 megacycles per second.

In treating problems arising in hyper- and ultrahigh-frequency engineering, the ordinary low-frequency circuit theory is inadequate, and the more general electromagnetic theory is required. Since the usual electrical curriculum does not emphasize general electromagnetic field theory, the first three chapters of this book serve to review the subject. They follow the standard plan employed in most textbooks and treatises.

Chapter 1 contains a review of electricity and magnetism. Chapter 2 presents a formulation of Maxwell's equations based on the works of Ampère, Faraday, and Maxwell. The wave equations are developed in Chapter 3, and reflection and refraction have been treated in Chapter 4. With the knowledge of the electromagnetic field equations and the laws of reflection and refraction at boundaries the student is prepared to study the propagation of waves in various types of enclosures and guides as explained in the subsequent chapters.

The authors are indebted to Lieutenant Commander S. P. Sashoff, Dr. N. Chako, and Dr. J. F. Lee for reading sections of the manuscript and for making valuable suggestions. They make grateful acknowledgment to the editors of the *Bell System Technical Journal*, *Electronics*, *Electrical Communication*, *Electrical Engineering*, *Journal of Applied Physics*, the *Proceedings of the Institute of Radio Engineers*, and *Radio*, as well as to the McGraw-Hill Publishing Company, for permission to reproduce material which originally appeared in their publications. They wish to express sincere gratitude to the many graduate students of the Illinois Institute of Technology, especially to the Messrs. J. C. Aberer, R. M. Soria, and Ensign C. E. Durkee, who worked tirelessly and painstakingly with the mimeographed notes detecting errors and obscure points.

ROBERT I. SARBACHER
WILLIAM A. EDSON

July, 1943

CONTENTS

TABLE OF SYMBOLS

Symbol	Definition	Page First Used
a	An arbitrary constant	79
a	A radius	242
A	An area	22
A	An arbitrary constant	83
A	An arbitrary angle	134
b	A radius	273
B, **B**	Magnetic induction or magnetic flux density	11
B	An arbitrary angle	134
B	Admittance	340
B_{in}	Input admittance	339
B_n	Component of magnetic induction normal to surface	51
c	Velocity of light	39
c'	Euler's constant	237
C	Capacitance	22
C	Symbolic for the intensity vectors	76
C	An arbitrary constant	125
C	Shunt capacitance of transmission line	281
C	Cathode designation	445
C	Coupling capacitance	454
C	Centigrade	505
C_c	Cathode by-pass condenser	454
C_d	Plate by-pass condenser	454
C_g	Grid-plate capacity	429
C_k	Grid-cathode capacity	429
C_p	Plate-cathode capacity	429
C_s	Screen-grid by-pass condenser	472
C_t	$C_g + C_p$	459
d	Fixed distance	22
d	Diameter of wave guide	271
D, **D**	Electric induction or electric flux density	11
D	A variable length	310
D_n	Component of electric induction normal to surface	53
e	Electronic charge	57
e	Eccentricity	288
e'	A sinusoidal voltage	459
e_a	Instantaneous voltage applied to one plate of a split-plate magnetron	563
e_b	Instantaneous voltage applied to one plate of a split-plate magnetron	563
e_g	The instantaneous grid voltage	427
e_p	The instantaneous plate voltage	427
E, **E**	Electric intensity	6
E	Designation for transverse magnetic waves	123

Symbol	Definition	Page First Used
E	A-c voltage applied to grid circuit	429
E	A-c voltage applied to grid	431
E'	Maximum value of sinusoidal variation of the electric intensity	66
E', $\mathbf{E}'$	Maximum value of electric intensity in distance and time	124
E'', $\mathbf{E}''$	Maximum value of time variation of electric intensity	121
E_1	A-c voltage applied to grid	429
E_2	A-c voltage across plate load	429
$E^{(1)}$	Component of electric intensity in $X^{(1)}$ direction	152
$E^{(2)}$	Component of electric intensity in $X^{(2)}$ direction	152
E_B	D-c plate supply voltage	454
E_{bb}	Plate polarizing potential	506
E_{cc}	Grid polarizing potential	506
E_g	A-c voltage applied to grid	433
E_g'	Maximum value of a-c voltage in grid circuit	507
E_k	Cathode voltage	463
E_0	D-c plate polarizing potential	563
E_p'	Maximum value of a-c voltage in plate circuit	507
E_r	Resonant waves	367
E_s	Standing waves	367
E_T	Component of electric intensity tangential to surface	51
E^+	Wave traveling in the positive direction	366
E^-	Wave traveling in the negative direction	366
eE_n	Waves in elliptic pipes (even)	285
oE_n	Waves in elliptic pipes (odd)	285
f	Frequency of a wave	130
f_0	Critical frequency	130
$(f_0)_n$	Critical frequency for waves in coaxial lines	278
$(f_0)_{nm}$	Critical frequency of waves in rectangular tubes	188
$(f_0)_{nm}$	Critical frequency for E waves in cylindrical tubes	249
$(f_0)_{nm}$	Critical frequency for H waves in cylindrical tubes	257
$(f_0)_{nm}$	Critical frequency of waves in a spherical cavity	391
$\mathbf{F}$	A force vector	46
F	Functional quantity	181
F	Mechanical force	438
F_e	Electrostatic force	2
F_m	Magnetostatic force	2
g	Arbitrary constant	182
g_m	The dynamic transconductance	427
G	Shunt conductance of transmission line	281
G	Grid designation	427
G	Effective input conductance	444
h	A fixed length	93
h	Arbitrary constant	182
H	Designation for transverse electric waves	122
H, $\mathbf{H}$	Magnetic intensity	6
$H^{(1)}$	Component of magnetic intensity in $X^{(1)}$ direction	148
$H^{(2)}$	Component of magnetic intensity in $X^{(2)}$ direction	148
H',$\mathbf{H}'$	Maximum value of magnetic intensity in distance and time	124
H'', $\mathbf{H}''$	Maximum value of time variation of magnetic intensity	121

Symbol	Definition	Page First Used
$\mathbf{H}$	A functional relation	37
H^+	Wave traveling in the positive direction	366
H^-	Wave traveling in the negative direction	366
H_c	Magnetic field intensity required for cut off	564
H_n	Component of magnetic intensity normal to surface	52
$H_p^{(1)}$	Hankel function of order p	238
$H_p^{(2)}$	Hankel function of order p	238
H_r	Resonant waves	367
H_s	Standing waves	367
oH_n	Waves in elliptic pipes (odd)	285
eH_n	Waves in elliptic pipes (even)	285
i	Angle of incidence	105
i	Angle of reflection	105
i	Instantaneous current	438
i	Alternating current	509
$\mathbf{i}$	Unit vector along X axis	9
i_a	Instantaneous current in one plate circuit of a split-plate magnetron	563
i_b	Instantaneous current in one plate circuit of a split-plate magnetron	563
i_g	Instantaneous grid current	427
i_p	Instantaneous plate current	427
I	Electric current	31
I'	Maximum current	137
I_g	Rms grid current	443
I_r	Rms current at receiving end of transmission line	325
I_s	Rms current at sending end of transmission line	324
$\mathbf{j}$	Unit vector along Y axis	9
j	$\sqrt{-1}$	121
j_n	Spherical Bessel function	391
j_n'	Derivative of spherical Bessel function with respect to its argument	392
J_n'	Derivative of ordinary Bessel function with respect to its argument	247
J_p	Bessel function of order p of the first kind	235
$\mathbf{k}$	Unit vector along Z axis	10
k	$\sqrt{\omega^2\mu\epsilon - \beta^2}$	287
k	A fixed distance	419
K	A constant	175
K	Cathode designation	427
l	Fixed distance	18
l	Direction cosine	77
L	Length	54
L	Inductance	55
L	Series inductance of transmission line	281
L	Plate compensating inductance	478
L_g	Effective inductance between grid and external terminals	428
L_g	Compensating inductance in video amplifier	472
L_k	Effective inductance between cathode and external terminals	428
L_p	Effective inductance between plate and external terminals	428
L_p	Compensating inductance in video amplifier	472
m	Fixed magnetic change	2

Symbol	Definition	Page First Used
m	Direction cosine	77
m	Modulation factor	133
m	An integer	184
m	Mass of electron	554
m'	Fixed magnetic change	2
m	Subscript designating rank of root of Bessel function	245
m	Subscript designating rank of root of Mathieu function	285
$\mathbf{M}$	Vector of magnetic polarization	21
M	Magnetic moment per unit area	33
M	Unit of mass	54
M	Mutual inductance	396
n	Direction cosine	77
n	An integer	125
n	$\sqrt{\kappa_e}$	271
n	Order of oscillation in magnetron oscillators	573
n	Subscript designating order of Bessel function	245
n	Subscript designating order of Mathieu function	285
N	Index of refraction	104
N	Any number	455
N	Voltage amplification ratio	466
p	Average power per unit area	89
p	An arbitrary constant	235
$\mathbf{P}$	Vector of electric polarization	19
P	Power	56
P	Plate designation	427
P	Number of pairs of plates in split-plate magnetron	573
q	Fixed electric charge	2
q'	Fixed electric charge	2
Q	Selectivity of a circuit	491
Q	Selectivity of a cavity	308
r	Variable distance	2
r	Angle of reflection	105
$\mathbf{r}$	Unit vector in direction of increasing r	231
r_{nm}	A root of rank m of J_n	245
r'_{nm}	A root of rank m of J'_n	256
r_p	The dynamic plate resistance	427
$\mathcal{R}$	Indicates real part of quantity following	90
R	Resistance	31
R	A function of r alone	234
R	Series resistance of transmission line	281
R	Effective resistance	433
R	$R_0R_g/(R_0 + R_g)$	460
R_c	Cathode resistance	454
R_d	Plate voltage dropping resistor	454
R_g	Grid leak resistance	454
R_{in}	Effective input resistance	433
R_L	Plate load resistance	454
R_0	$R_Lr_p/(R_L + r_p)$	458
R_s	Screen-grid voltage dropping resistor	472

Symbol	Definition	Page First Used
s	Area of surface	12
s	A fixed distance	76
S	Poynting vector	82
S	Periphery of elliptic pipe	285
t	Time	29
T	Period of a wave	69
T	Time	54
TE	Transverse electric waves	64
TM	Transverse magnetic waves	64
TEM	Transverse electromagnetic waves	64
u_{nm}	Root of rank m of j_n	391
u'_{nm}	Root of rank m of j'_n	391
U	Magnetostatic potential	7
v	Velocity	46
v	Output voltage	470
v_g	Group velocity	133
$(v_g)_{nm}$	Group velocity of waves in rectangular tubes	189
$(v_g)_{nm}$	Group velocity of E waves in cylindrical tubes	247
$(v_g)_{nm}$	Group velocity of H waves in cylindrical tubes	259
v_p	Phase velocity	132
$(v_p)_{nn}$	Phase velocity of waves in rectangular tubes	189
$(v_p)_{nm}$	Phase velocity of E waves in cylindrical tubes	246
$(v_p)_{nm}$	Phase velocity of H waves in cylindrical tubes	259
V	Electrostatic potential	7
V	Square wave input voltage	470
V'	Maximum voltage	137
V_a	Grid-anode d-c voltage	522
V_k	Grid-cathode d-c voltage	522
V_0	A sinusoidal voltage	474
V_r	Rms voltage at receiving end of transmission line	325
V_s	Rms voltage at sending end of transmission line	324
V_1	A-c grid voltage	459
V_2	A-c output voltage from amplifier stage	459
w	Variable width	51
W	Energy	22
W_E	Maximum value of electric energy	372
W_h	b/λ_a	406
W_M	Maximum value of magnetic energy	372
W_v	a/λ_a	403
$\mathbf{x}$	Unit vector in direction of increasing X	231
X	An angle	333
X	Reactance	337
X_{in}	Input reactance of transmission line	337
Y	A function of y alone	181
Y	Shunt admittance of transmission line	281
Y_g	Grid circuit admittance	443
Y_p	Bessel function of order p of the second kind	235
Z	A function of z alone	181
Z_0	Characteristic impedance	207

Symbol	Definition	Page First Used
Z	Series impedance of transmission line	281
Z	Plate load impedance	427
Z	$1/j\omega c_t$	460
Z_{in}	Effective input impedance	433
Z_{in}	Input impedance of transmission line	335
Z_k	Cathode impedance	463
Z_0	Characteristic impedance of transmission line	325
Z_{oc}	Input impedance when far end of transmission line is open-circuited	330
Z_{sc}	Input impedance when far end of transmission line is short-circuited	330

Greek Letter Symbols

α	An angle	76
α	Attenuation constant	128
α	Attenuation constant for waves in rectangular tubes	186
β	Phase constant	66
β	An angle	77
β_n	Phase constant for coaxial lines	278
β_{nm}	Phase constant for waves in rectangular tubes	186
β_{nm}	Phase constant for E waves in cylindrical tubes	246
β_{nm}	Phase constant for H waves in cylindrical tubes	257
γ	Propagation constant	124
γ_n	Propagation constant for waves between parallel plates	129
γ_n	Propagation constant for coaxial lines	278
γ_{nm}	Propagation constant for waves in rectangular tubes	186
γ_{nm}	Propagation constant for E waves in cylindrical tubes	244
γ_{nm}	Propagation constant for H waves in cylindrical tubes	257
Δ	A voltage employed in the separate plate circuits of a split-plate magnetron	563
δ	A small non-vanishing number	133
ϵ_0	Electric inductive capacity of empty space	2
ϵ	Electric inductive capacity	2
η	Intrinsic impedance of dielectric material	138
η	Intrinsic impedance of a conducting material	161
θ	An angle	7
θ	A phase angle	90
θ	Angular phase shift in radians	593
ι	Electric current density	27
κ_e	Specific inductive capacity or dielectric constant	2
λ	Wavelength of a wave in free space	69
λ_c	Cut-off wavelength	292
λ_k	Wavelength generated by positive-grid tube	532
λ_0	Critical wavelength	139
$(\lambda_0)_n$	Critical wavelength for waves in coaxial lines	278
$(\lambda_0)_{nm}$	Critical wavelength of waves in rectangular tubes	188
$(\lambda_0)_{nm}$	Critical wavelength for E waves in cylindrical tubes	246
$(\lambda_0)_{nm}$	Critical wavelength of H waves in cylindrical tubes	259
λ_p	Wavelength between plates or in pipe	139
$(\lambda_p)_{nm}$	Wavelength of waves in rectangular tubes	189
$(\lambda_p)_{nm}$	Wavelength of E waves in cylindrical tubes	249

Symbol	Definition	Page First Used
$(\lambda_p)_{nm}$	Wavelength of H waves in cylindrical tubes	259
μ	Magnetic inductive capacity	2
μ	The dynamic amplification factor	427
μ_0	Magnetic inductive capacity of empty space	4
μ_m	Magnetic permeability	4
π	Ratio of circumference to diameter of circle (3.14159)	2
ρ	Volume density of charge	15
ρ	Argument of Bessel functions	236
ρ	Resistivity of conducting material	396
σ	Conductivity	31
τ	Volume	83
ϕ	Magnetostatic flux	12
ϕ	A phase angle	90
ϕ	Cylindrical coordinate	228
$\boldsymbol{\phi}$	Unit vector in direction of increasing ϕ	231
Φ	A function of ϕ alone	234
ψ	Electrostatic flux	12
Ω	Solid angle	13
ω	Angular frequency	66

Other Symbols

$\log_e$	Natural logarithm	281
$\log_{10}$	Common logarithm	455
curl	Curl of	38
div	Divergence of	16
grad	Gradient of	10
∇	Symbol for grad	11
$\lvert C \rvert$	Absolute value of C	102
$\simeq$	Approximately equal to	15
$*$	Superscript indicating conjugate quantity	91
$\neq$	Not equal to	157
$\ll$	Small in comparison to	125
$\gg$	Large in comparison to	160
$\sum$	Sum of	247
∞	Infinity	8
$\equiv$	Defined as	11
$\cdot$	Dot product	15
$\times$	Cross product	38
$\rightarrow$	Approaches	15
$\overline{\ }$	Bar over letter indicates average value	27

CHAPTER 1

ELECTROSTATICS AND MAGNETOSTATICS

1·1 Introduction

The equations of electric circuit th ory in thei present form are essentially generalizations of the original work of Ohm, Faraday, Henry, Lenz, Kirchhoff, and others. Maxwell and Lorentz, in developing the electromagnetic theory, accumulated in one set of equations the work of most of the scientists who preceded them. The mathematical deductions which have resulted from these equations are of the greatest importance in hyper and ultra-high frequency engineering

Prior to the concise formulation of the electromagnetic theory by Maxwell and Lorentz, the seat of electromagnetic phenomena was taken to be within the conductor. Phenomena taking place outside the conductor were only vaguely understood. The conductor was considered to carry the electromagnetic energy much as a pipe carries water, and it was, therefore, the principal consideration. Even today many students and engineers hold to this point of view. On such a basis it is obviously difficult to account for the propagation of electromagnetic energy through free space without the existence of a conductor of some type. Many students seem to feel that the propagation of electromagnetic energy, as in radio, is something quite apart from the normal theory involving a conducting medium.

It is the purpose of the first three chapters to review the basic laws of electric and magnetic phenomena and to show how they are interrelated by the field equations of Maxwell, which we shall formulate.

1·2 Definition of the Unit of Electric Force and the Unit of Magnetic Force

The fundamental physical fact that electric charges exert forces on one another and that magnetic poles exert forces on one another is formulated into a simple law. This law was suggested, in respect to charge, by Priestley, and, in respect to magnetic poles, by Mitchell. Cavendish was probably the first man to give an accurate experimental proof of this relation. Coulomb, however, was the first to publish his results, and the law is now known by his name.

COULOMB'S LAW

Electrostatic Force

Coulomb's law for the force between two point electric charges is given by the formula

$$F_e = \frac{1}{4\pi} \cdot \frac{qq'}{\varepsilon r^2} \qquad [1\cdot1]$$

This states that the force between two point charges q and q' is directly proportional to the product of the charges and is inversely proportional to the square of the distance r between them and to the factor ε to be defined later.

Magnetostatic Force

Coulomb's law for the force between two point magnetic poles is given by the formula

$$F_m = \frac{1}{4\pi} \cdot \frac{mm'}{\mu r^2} \qquad [1\cdot2]$$

This states that the force between two point poles m and m' is directly proportional to the product of the pole strengths and is inversely proportional to the square of the distance r between them and to the factor μ to be defined later.

Magnetic pole strength is sometimes referred to as magnetic charge. The north pole of the magnet is called the positive pole or positive magnetic charge; and the south pole, the negative pole or negative magnetic charge. In either the electrostatic or magnetostatic force equations 1·1 and 1·2, if the charges are of the same sign, the force is positive; and if they are of opposite sign, the force is negative.

1·3 The Constant ε

The constant ε in equation 1·1 is defined by the relation

$$\varepsilon = \kappa_e \varepsilon_0 \qquad [1\cdot3]$$

where κ_e is a dimensionless coefficient depending upon the material, defined as unity for a vacuum, and ε_0 has a value depending upon the system of units used. In the cgs electrostatic system, ε_0 is defined as $1/4\pi$ and is tacitly assumed to be dimensionless. In the absolute rationalized* practical system discussed later, the value is given by

$$\varepsilon_0 = \frac{1}{36\pi} \times 10^{-9} \simeq 8.854 \times 10^{-12} \text{ farad/meter} \qquad [1\cdot4]$$

Equation 1·1 is expressed in coulombs of charge, meters of distance, and newtons† of force. The coulomb, defined by this relation, is chosen as the basic electrical unit in conjunction with the mechanical units of length in meters, mass in kilograms, and time in seconds.

* The term rationalized refers to a system in which the multiplier 4π is transferred into the force equation so that it does not appear in other equations. The desirability of this arrangement will become clear as we proceed.

† One newton of force will accelerate 1 kilogram at the rate of 1 meter per second per second. One newton = 10^5 dynes.

Faraday gave the name specific inductive capacity to the multiplier κ_e. More commonly the term dielectric constant is used for this symbol. The value of κ_e is defined as unity in vacuum; it is greater than unity for all material media. Two charges exert the maximum possible mutual force when they exist in vacuum. If they exist in any other homogeneous medium the force is reduced by the factor κ_e. This relation is exact only if the medium is of infinite extent and may be quite inaccurate if either charged body is not in intimate contact with the medium. Thus a flat slab of dielectric material inserted between two charged bodies in vacuum actually increases the force between them.

The dielectric constant depends upon the material and to some extent upon temperature, pressure, frequency, etc. At high frequencies a time lag is sometimes observed which is associated with a loss of power in the dielectric. This may be expressed by assigning a complex value (real and imaginary parts) to the dielectric constant. In general, the dielectric constant is a constant in the sense that it is independent of the magnitude or arrangement of charges.

TABLE 1·1

DIELECTRIC CONSTANTS

Solids

Substance	κ_e	Substance	κ_e
Bakelite	3–4	Wood, red beech, $\parallel$ fibers	4.8–2.5
Glass	5–10	Wood, red beech, $\perp$ fibers	7.7–3.6
Mica, ruby	4.2–4.6	Wood, oak, $\parallel$ fibers	4.2–2.6
Paper	2–2.5	Wood, oak, $\perp$ fibers	6.8–3.6
Paraffin	2.2		
Quartz	4.5		
Steatite	5.5–6.5		
Sulfur	4.0		
Liquids		*Gases*	
Alcohol, ethyl	26	Air	1.000588
Alcohol, methyl	31	Carbon dioxide	1.000988
Petroleum	2.2	Hydrogen	1.000264
Water	81		

The dielectric constant of a material is usually determined by measurement of the capacitance of a parallel-plate condenser in which the material appears as the dielectric. Table 1·1 gives the dielectric constants of a number of familiar substances. Most of the data presented are contained in the *Smithsonian Physical Tables.*

1·4 The Constant μ

The constant μ in equation1·2 is defined by the relation

$$\mu = \mu_m \mu_0 \qquad [1·5]$$

where μ_m is a dimensionless coefficient depending upon the material, and μ_0 has a value depending upon the system of units used. In the cgs electromagnetic system μ_0 is defined as $1/4\pi$ and is tacitly assumed to be dimensionless. In the rationalized practical system the value is given by

$$\mu_0 = 4\pi \times 10^{-7} = 1.257 \times 10^{-6} \text{ henry per meter} \qquad [1·6]$$

The term permeability is universally used for the dimensionless coefficient μ_m. For vacuum the numerical value of μ_m, like κ_e, is defined as unity. Certain materials have a permeability slightly less than unity and are called diamagnetic. Other materials have a permeability slightly greater than unity and are called paramagnetic. A few materials having permeabilities much greater than unity are called ferromagnetic.

The remarks made in regard to geometry under the dielectric constant apply directly to the present case. However, no liquids have a permeability significantly different from unity. Accordingly, with liquids, the experimental significance of μ_m is less than that of κ_e.

TABLE 1·2

PERMEABILITIES

Ferromagnetic Materials			
Substance	μ_m Maximum	μ_m at Small Magnetization	
Cobalt	60	60	
Nickel	50	50	
Cast iron	90	60	
Silicon iron	7000	3500	
Transformer iron	5500	3000	
Very pure iron	8000	4000	
Machine steel	450	300	
Paramagnetic Materials		*Diamagnetic Materials*	
Aluminum	1.00000065	Bismuth	0.99999860
Beryllium	1.00000079	Paraffin	0 99999942
$MnSO_4$	1.000100	Silver	0 99999981
$NiCl_2$	1.000040	Wood	0.99999950

In diamagnetic and paramagnetic materials the permeability is a function of temperature, pressure, and frequency but not of the magnetic intensity. In ferromagnetic materials the permeability varies markedly with the intensity of the magnetization, as well as with the above. At high frequencies the permeability may assume a complex value. Table 1·2 shows the permeability of a number of materials.

1·5 The Unit of Electric Charge and the Unit of Magnetic Charge

The relations 1·1 and 1·2 contain the definitions of the fundamental electric charge and the fundamental magnetic charge, respectively.

Definition of the Unit Electric Charge

The unit of electric charge may be defined as that charge which repels an equal charge placed 1 meter from it, in vacuum, with a force of 9×10^9 newtons. It is called the coulomb.

Definition of the Unit Magnetic Charge

The unit of magnetic charge may be defined as that charge which repels an equal charge placed 1 meter from it, in vacuum, with a force of $10^7/16\pi^2$ newtons. It is called the weber.

In this and the previous section we have spoken of the magnetic pole or magnetic charge as if the positive pole or charge could be segregated from the negative pole. With electric charges this is quite easily done. Only recently, however, have free magnetic charges been isolated.* Indeed at this time the existence of the free magnetic charge may hardly be considered a proved fact. In the past it was customary to approximate the free magnetic pole by the use of a very long slender magnet. Thus a unit magnetic pole or charge was considered to be a north pole of a long thin magnet, whose south pole was so far removed from it that its influence on the experiment was negligible.

1·6 The Electrostatic and Magnetostatic Field

The term *field* is used to identify a region in which electric or magnetic forces are acting.

The Electrostatic Field

Regions in which electric forces are acting are called electric fields. We may investigate an electric field by placing a unit electric charge in various positions in the field and observing the force acting upon it.

The Magnetostatic Field

Regions in which magnetic forces are acting are called magnetic fields. We may investigate a magnetic field by placing a unit magnetic pole in various positions in the field and observing the force acting upon it.

In either case it is necessary that the original charges creating the field be fixed in position so that they will not be disturbed by the exploring charge, and that the force observed be multiplied by the ratio of unit to actual charge in order to obtain the correct result. This result is usually expressed in the ratio form of force per unit charge.

1·7 The Electric Intensity E and the Magnetic Intensity H

The electric and magnetic fields of force are expressed quantitatively in terms of the electric and magnetic intensities **E** and **H**. These quantities are defined as follows:

* F. Ehrenhaft, "Photophoresis and Its Interpretation by Electric and Magnetic Ions," *J. Franklin Inst.*, **233**, 235, 1942.

Electric Intensity E

The force experienced by a small positive test charge when held at rest in any position or point in an electrostatic field divided by the charge is called the electric intensity **E** at that point.

Magnetic Intensity H

The force experienced by a small north magnetic pole when held at rest in any position or point in a magnetostatic field divided by the pole strength is called the magnetic intensity **H** at that point.

Both quantities **E** and **H** are vectors, which have at every point in the field both magnitude and direction.

If an electric charge q' is placed in an electric field **E**, the force which it experiences is

$$\mathbf{F}_e = q'\mathbf{E} \qquad [1\cdot7]$$

If q' is positive, $\mathbf{F}_e$ has the same direction as **E**. The equation for the electric intensity due to any electric charge q may be obtained by equating 1·1 and 1·7 and making q' unity. This gives

$$E = \frac{q}{4\pi\varepsilon r^2} \qquad [1\cdot9]$$

which is in the direction of the radius vector r joining the two points. In the rationalized practical system of units the electric intensity is expressed in volts per meter or newtons per coulomb.

If a magnetic charge m' is placed in a magnetic field **H**, the force which it experiences is

$$\mathbf{F}_m = m'\mathbf{H} \qquad [1\cdot8]$$

If m' is positive, $\mathbf{F}_m$ has the same direction as **H**. The equation for the magnetic intensity due to any magnetic charge m may be obtained by equating 1·2 and 1·8 and making m' unity. This gives

$$H = \frac{m}{4\pi\mu r^2} \qquad [1\cdot10]$$

which is in the direction of the radius vector r joining the two points. In the rationalized practical system of units the magnetic intensity is expressed in ampere-turns per meter or newtons per weber.

1·8 Electrostatic and Magnetostatic Potential

The field strength or intensity may also be defined by means of the concept of potential. This concept serves to clarify and simplify many difficult problems. It is applied quite generally to the study of field problems, such as those arising in thermodynamics, as well as to those pertaining to electricity.

In a thermal problem it is customary to define the conditions existing throughout the system in terms of temperature, a scalar quantity or function dependent upon position. The flow of heat is then expressed in terms of the variation of the temperature, the flow being greatest in regions where the temperature changes most rapidly. The conditions existing in an electrical system may similarly be defined in terms of a scalar function called potential. Here the field is strongest in regions where the potential changes most rapidly. Potential may thus be defined as a quantity whose space rate of change in any direction is the strength of the field in that direction.

Electrostatic Potential

If electrostatic potential is designated V, then from the above definition

$$\frac{-dV}{dl} = E \cos \theta \qquad [1\cdot11]$$

where θ is the angle between the direction of **E** and the path element **dl**. Hence $E \cos \theta$ is the strength of the field in the direction of dl. The negative sign indicates that in electrostatics the force between like charges is a repulsion and diminishes as the distance between them increases. Equation 1·11 may also be written

$$dV = -E \cos \theta \, dl \qquad [1\cdot13]$$

Magnetostatic Potential

If magnetostatic potential is designated U, then from the above definition

$$\frac{-dU}{dl} = H \cos \theta \qquad [1\cdot12]$$

where θ is the angle between the direction of **H** and the path element **dl**. Hence $H \cos \theta$ is the strength of the field in the direction of dl. The negative sign indicates that in magnetostatics the force between like poles is a repulsion and diminishes as the distance between them increases. Equation 1·12 may also be written

$$dU = -H \cos \theta \, dl \qquad [1\cdot14]$$

The products $Eq \cos \theta \, dl$ or $Hm \cos \theta \, dl$ represent work since $Eq \cos \theta$ or $Hm \cos \theta$ is a force, and dl is a distance in the direction of the force. If a small electric charge is carried from one point to another in an electro-

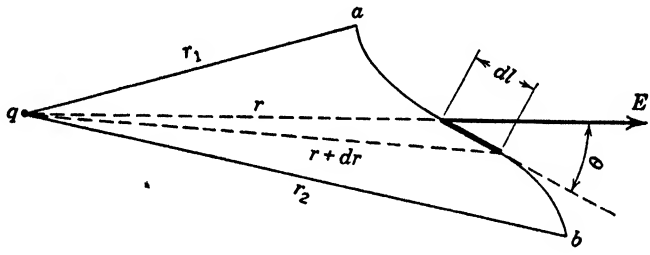


FIG. 1·1

static field, work is done. This work will be represented by the integration of equation 1·13. Consider, for example, the work done in carrying a unit charge from point b to point a, situated in the radial electrostatic field about the charge q, Fig. 1·1. Then, from equations 1·13 and 1·9

$$\int_b^a dV = -\int_b^a E \cos \theta \, dl = -\int_b^a \frac{q}{4\pi\varepsilon r^2} \cos \theta \, dl \qquad [1\cdot15]^*$$

but $\cos \theta \, dl = dr$, and hence

$$\int_b^a dV = -\int_{r_2}^{r_1} \frac{q}{4\pi\varepsilon r^2} dr = \frac{q}{4\pi\varepsilon}\left(\frac{1}{r}\right)\Bigg|_{r_2}^{r_1} \qquad [1\cdot16]$$

* Integrals of this kind are called line integrals.

Thus

$$V_a - V_b = \frac{q}{4\pi\varepsilon r_1} - \frac{q}{4\pi\varepsilon r_2} = \frac{q}{4\pi\varepsilon}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad [1\cdot17]$$

If we had carried the unit charge along some other path than that indicated, the result would have been the same. If this were not so there would be a balance of work done in carrying the charge from a to b and then along a different path, back to a. Continuous repetition of this process would constitute a power output which in a static system constitutes a violation of the principle of conservation of energy. If the work were dependent upon the path chosen, two or more potentials would necessarily be associated with each point. Since the work is independent of the path we find that for all electrostatic or magnetostatic systems there is only one value of potential at each point, and hence it is required that

$$\oint_l E \cos\theta \, dl = 0 \qquad [1\cdot18]$$

and

$$\oint_l H \cos\theta \, dl = 0 \qquad [1\cdot19]$$

where the symbols $\oint_l$ refer to the integration around any closed path. It will be shown later that this is true only in static cases but is not true in general. A system in which the potential is single-valued is called a conservative system, and the field or gradient associated is referred to as irrotational.

We see from equation 1·17 that potential is measured by its differences. If we can carry a charge between two points in a field without the expenditure of work, these points are at the same potential. Furthermore, equation 1·17 shows that all points at an infinite distance from a given charge are at the same potential, and we may arbitrarily set the potential at infinity equal to zero. Thus, by putting the point b at infinity, $r_2 = \infty$ and

$$V_a = \frac{q}{4\pi\varepsilon r_1} \qquad [1\cdot20]$$

we assign an absolute potential to the point a. Hence we may define the potential at a point in the electric field as follows:

Electrostatic Potential at a Point

The electric potential V at a point in an electrostatic system may be defined as the work done in bringing a unit (positive) charge from infinity up to the point and is equal to

$$V = \frac{q}{4\pi\varepsilon r} \qquad [1\cdot21]$$

where q represents the charge producing the field and r is the radius vector from q to the point. The electrostatic potential is measured in joules per unit charge in the practical system.

Magnetostatic Potential at a Point

The magnetic potential U at a point in a magnetostatic system may be defined as the work done in bringing a unit magnetic (north) pole from infinity up to the point and is equal to

$$U = \frac{m}{4\pi\mu r} \qquad [1\cdot22]$$

where m represents the pole producing the field and r is the radius vector from m to the point. The magnetostatic potential is measured in joules per unit charge in the practical system.

The potential is a scalar quantity or function and is especially valuable because the effect of many charges or poles is accounted for by a simple scalar addition or integration process. Vector addition or integration as required by field strength calculation is less simple.

1·9 Vectors and Scalars

Measurable quantities occurring in mathematical physics may be divided into two broad groups called vectors and scalars. Vectors represent directed magnitudes in that they require for their representation both direction and magnitude. Forces and electric and magnetic intensities are among these, as are velocity and acceleration. Scalars belong to that group of quantities which can be completely specified by their magnitude alone. Charge, mass, volume, temperature, and electrostatic or magnetostatic potential are examples of scalars.

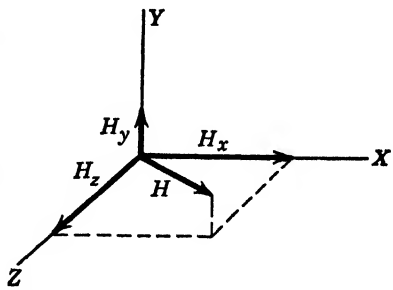


FIG. 1·2 Resolution of a vector into its components.

When dealing with vectors it is common practice to decompose them into components along some set of arbitrarily chosen coordinate axes. For example, the vector components of **H** in the direction of the x, y, z axes, Fig. 1·2, will be designated $\mathbf{H}_x$, $\mathbf{H}_y$, and $\mathbf{H}_z$. The scalar magnitude of **H** will be designated H, with the components H_x, H_y, and H_z. The usual unit vectors along the x, y, and z axes will be designated **i**, **j**,

and **k**, respectively. Thus

$$\mathbf{H} = \mathbf{H}_x + \mathbf{H}_y + \mathbf{H}_z = \mathbf{i}H_x + \mathbf{j}H_y + \mathbf{k}H_z \qquad [1\cdot23]$$

where the plus sign refers to the vector sum. See Fig. 1·2. The magnitude of **H** is given by

$$H = \sqrt{H_x^2 + H_y^2 + H_z^2}$$

When a vector has been broken down into its components, analytical transformations may conveniently be applied. After these analytical transformations have been performed, the effect on the vector may be observed by studying the effect produced on its components. This situation is sometimes rather awkward, and a method of analysis, known as vector analysis, results in extensive simplification.

1·10 Electrostatic and Magnetostatic Potential Gradient

Electrostatic Potential Gradient

In equation 1·11 we have seen that

$$E \cos\theta = -\frac{dV}{dl} \qquad [1\cdot11]$$

The various components of E along the coordinate axes are

$$\left.\begin{aligned} E_x &= -\frac{\partial V}{\partial x} \\ E_y &= -\frac{\partial V}{\partial y} \\ E_z &= -\frac{\partial V}{\partial z} \end{aligned}\right\} \qquad [1\cdot24]$$

Since $\mathbf{E} = \mathbf{i}E_x + \mathbf{j}E_y + \mathbf{k}E_z$, then

$$\mathbf{E} = -\left(\mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial y} + \mathbf{k}\frac{\partial V}{\partial z}\right) \qquad [1\cdot26]$$

or $\mathbf{E} = -$ **grad** V (joules per unit charge per meter).

Magnetostatic Potential Gradient

In equation 1·12 we have seen that

$$H \cos\theta = -\frac{dU}{dl} \qquad [1\cdot12]$$

The various components of H along the coordinate axes are

$$\left.\begin{aligned} H_x &= -\frac{\partial U}{\partial x} \\ H_y &= -\frac{\partial U}{\partial y} \\ H_z &= -\frac{\partial U}{\partial z} \end{aligned}\right\} \qquad [1\cdot25]$$

Since $\mathbf{H} = \mathbf{i}H_x + \mathbf{j}H_y + \mathbf{k}H_z$, then

$$\mathbf{H} = -\left(\mathbf{i}\frac{\partial U}{\partial x} + \mathbf{j}\frac{\partial U}{\partial y} + \mathbf{k}\frac{\partial U}{\partial z}\right) \qquad [1\cdot27]$$

or $\mathbf{H} = -$ **grad** U (joules per unit pole per meter).

Grad V (pronounced as it is spelled) is defined as

$$\mathbf{grad}\ V = \nabla V = \mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial y} + \mathbf{k}\frac{\partial V}{\partial z} \qquad [1\cdot28]$$

and is called the potential gradient at a given point. The symbol **grad** V is sometimes written ∇V, pronounced del V or nabla V, where

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

is a vector differential operator. The potential gradient is a vector, having components defined by

$$\left.\begin{aligned} \mathbf{i}E_x &= -\,\mathbf{grad}\ V_x \\ \mathbf{j}E_y &= -\,\mathbf{grad}\ V_y \\ \mathbf{k}E_z &= -\,\mathbf{grad}\ V_z \end{aligned}\right\} \quad [1\cdot29] \qquad \left.\begin{aligned} \mathbf{i}H_x &= -\,\mathbf{grad}\ U_x \\ \mathbf{j}H_y &= -\,\mathbf{grad}\ U_y \\ \mathbf{k}H_z &= -\,\mathbf{grad}\ U_z \end{aligned}\right\} \quad [1\cdot30]$$

and is directed opposite to the electric intensity **E** as indicated by the negative sign.

1·11 Electric and Magnetic Induction

Another vector convenient to employ is the induction vector, which is related to the intensity vector.

Electric Induction

The electric induction at any point is defined as the product of the electric intensity at the point by the value of the constant ϵ of the medium at the point. The electric induction is usually designated **D** with components $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$

Magnetic Induction

The magnetic induction at any point is defined as the product of the magnetic intensity at the point by the constant μ of the medium at that point. The magnetic induction is usually designated **B** with components $\mathbf{B}_x$, $\mathbf{B}_y$, and $\mathbf{B}_z$.

In this discussion we shall consider only homogeneous, isotropic media in which μ and ϵ are constant and independent of the direction in which they are measured.

Hence

$$\mathbf{D} = \epsilon\mathbf{E} \quad [1\cdot31]$$

where

$$\left.\begin{aligned} D_x &= \epsilon E_x \\ D_y &= \epsilon E_y \\ D_z &= \epsilon E_z \end{aligned}\right\} \quad [1\cdot33]$$

Hence

$$\mathbf{B} = \mu\mathbf{H} \quad [1\cdot32]$$

where

$$\left.\begin{aligned} B_x &= \mu H_x \\ B_y &= \mu H_y \\ B_z &= \mu H_z \end{aligned}\right\} \quad [1\cdot34]$$

The vectors **D** and **B** represent the electric flux density and magnetic flux density, respectively.

1·12 Electrostatic and Magnetostatic Flux

As used in mathematical physics the term flux designates either the product of an area by the component of a vector such as **D** or **B** or the

summation of such products. Such a summation is called a surface integral.

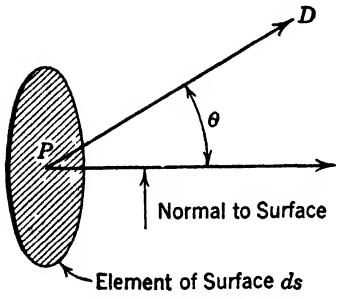


FIG. 1·3

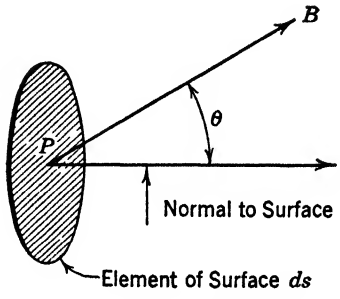


FIG. 1·4

Electrostatic Flux

Since both magnitude and direction of the electric intensity **E** are functions of the coordinates at any point P in an electrostatic field, the same is true of the electric induction **D**. Consider the point P, Fig. 1·3, where the electric induction vector **D** pierces the surface element ds at an angle θ with respect to the normal to the surface. The flux of electric induction $d\psi$ through the area ds is then defined as

$$d\psi = D \cos \theta \, ds$$

or

$$d\psi = D_n \, ds$$

where D_n is the normal component of **D** to the surface ds.

By integrating over any extended surface S we obtain the total flux ψ through the surface as

$$\psi = \int\int_s D \cos \theta \, ds \qquad [1·35]$$

or

$$\psi = \int\int_s D_n \, ds \qquad [1·37]$$

Magnetostatic Flux

Since both magnitude and direction of the magnetic intensity **H** are functions of the coordinates at any point P in a magnetostatic field, the same is true of the magnetic induction **B**. Consider the point P, Fig. 1·4, where the magnetic induction vector **B** pierces the surface element ds at an angle θ with respect to the normal to the surface. The flux of magnetic induction $d\phi$ through the area ds is then defined as

$$d\phi = B \cos \theta \, ds$$

or

$$d\phi = B_n \, ds$$

where B_n is the normal component of **B** to the surface ds.

By integrating over any extended surface S we obtain the total flux ϕ through the surface as

$$\phi = \int\int_s B \cos \theta \, ds \qquad [1·36]$$

or

$$\phi = \int\int_s B_n \, ds \qquad [1·38]$$

where $\int\int_s$ indicates the surface integral over the surface S, and ds is an element of the surface.

1·13 Gauss's Law

Gauss's law applied to the electrostatic field states that, if the surface integral [1·35] is extended over a closed surface surrounding a charge q, then the total flux over the surface is equal to q. That is, the total normal electric induction over any closed surface is equal to the total electric charge contained within the surface.

Gauss's law applied to the magnetostatic field states that, if the surface integral [1·36] is extended over a closed surface surrounding a magnetic pole m, then the total flux over the surface is equal to m. That is, the total normal magnetic induction over any closed surface is equal to the total magnetic charge contained within the surface.

Consider any point P in an electrostatic field where the induction vector is:

$$D = \varepsilon E = \frac{q}{4\pi r^2}$$

Then by equation 1·35

$$\psi = \int\int_{\text{closed surface}} \frac{q}{4\pi r^2} \cos\theta \, ds \qquad [1\cdot39]$$

Let $d\Omega$ be the solid angle subtended at q by ds. Then by the definition of the solid angle

$$d\Omega = \frac{ds_n}{r^2} = \frac{\cos\theta \, ds}{r^2}$$

where ds_n represents the component of the area ds normal to the radius vector from q.

Hence

$$\psi = \int\int_{\text{closed surface}} \frac{q}{4\pi} d\Omega = \frac{q}{4\pi} \int\int_{\text{closed surface}} d\Omega = \frac{q\Omega}{4\pi} = q \qquad [1\cdot40]^*$$

since the solid angle Ω subtended at any point within a closed surface is 4π.

From this development it is clear that one line of flux is associated with each unit charge in the rationalized system, and therefore that the flux density is $1/4\pi$ lines per square meter at a distance of 1 meter from the charge. In the familiar non-rationalized system the flux density is made equal to unity at unit distance, and therefore 4π lines are associated with each unit charge.

* If a charge is partly inside and partly outside the surface then q refers to the charge within the surface. If this is zero then $\psi = 0$ over the surface since as much flux enters it as leaves it.

1·14 Poisson's and Laplace's Equation

Consider the infinitesimal rectangular volume shown in Fig. 1·5, of dimensions Δx, Δy, Δz, situated in an electrostatic field with one corner of the volume at the point x, y, z. The volume is shown large for convenience. We shall determine the surface integral of the continuous vector **D** over the surface of the volume $\Delta\tau = \Delta x\,\Delta y\,\Delta z$.

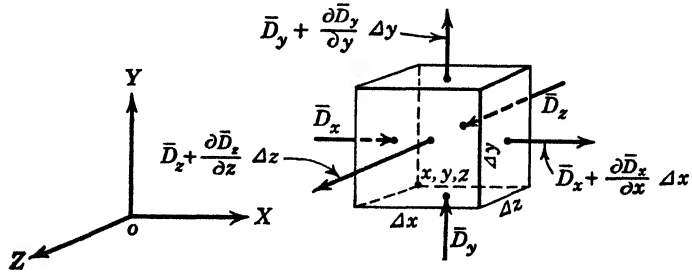


FIG. 1·5 Infinitesimal volume $\Delta\tau = \Delta x\,\Delta y\,\Delta z$ in an electric field at the point x, y, z. The bar over the electric flux density vectors indicates the average value of the vectors over the various faces of the volume.

Let $\bar{D}_x$ be the average normal component of electric induction over the area $\Delta y\,\Delta z$ which has one corner at the point x, y, z. If $\bar{D}_x$ meets the continuity requirements of Taylor's theorem* for the expansion of a function about a point, then the average normal component of **D** over the opposite face of the cube will be

$$\bar{D}_x + \frac{\partial \bar{D}_x}{\partial x}\Delta x \qquad [1\cdot41]\dagger$$

when we neglect second-order infinitesimals. Similar equations defin-

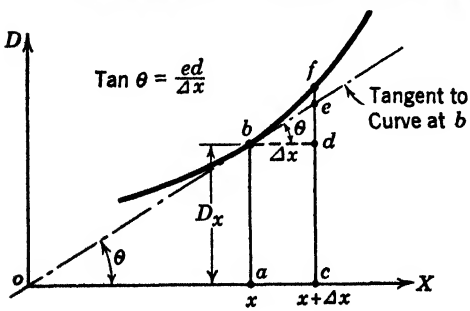


FIG. 1·5a

* For a discussion of Taylor's theorem see Burington and Torrance, *Higher Mathematics*, page 74, section 13.

† This expression for the electric induction may be clarified by the following argument. If D is a single-valued continuous function of x it may be represented graphically as shown in Fig. 1·5*a*. This curve represents the way D varies in the x direction when y and z are held fixed. If we know the ordinate to the curve at the point x and call it $D_x = ab$, our problem is to find the value of D at the point $x + \Delta x$ in terms of D_x. This value is represented by the ordinate cf where

$$cf = cd + de + ef$$

ing the average components of **D** over the remaining faces are as shown in Fig. 1·5. Hence the total surface integral of **D** in the outward direction over the volume is

$$\psi = \int\int_{\text{closed surface}} D_n\, ds = -\bar{D}_x\,\Delta y\,\Delta z + \left(\bar{D}_x + \frac{\partial \bar{D}_x}{\partial x}\Delta x\right)\Delta y\,\Delta z - \bar{D}_y\,\Delta x\,\Delta z$$
$$+ \left(\bar{D}_y + \frac{\partial \bar{D}_y}{\partial y}\Delta y\right)\Delta x\,\Delta z - \bar{D}_z\,\Delta x\,\Delta y + \left(\bar{D}_z + \frac{\partial \bar{D}_z}{\partial z}\Delta z\right)\Delta x\,\Delta y \qquad [1\cdot42]$$

giving
$$\psi = \left(\frac{\partial \bar{D}_x}{\partial x} + \frac{\partial \bar{D}_y}{\partial y} + \frac{\partial \bar{D}_z}{\partial z}\right)\Delta x\,\Delta y\,\Delta z \qquad [1\cdot43]$$

According to Gauss's theorem, ψ is also equal to the charge enclosed by the volume. If ρ is the volume density of charge, then the total charge contained in the volume element is $\rho\,\Delta x\,\Delta y\,\Delta z = \rho\,\Delta\tau$. Hence

$$\int\int_{\text{closed surface}} D_n\, ds = \left(\frac{\partial \bar{D}_x}{\partial x} + \frac{\partial \bar{D}_y}{\partial y} + \frac{\partial \bar{D}_z}{\partial z}\right)\Delta\tau = \rho\,\Delta\tau \qquad [1\cdot44]$$

Or, dividing through by $\Delta\tau$ and taking the limit as the volume approaches zero, we obtain

$$\underset{\Delta\tau\to 0}{\text{Limit}}\left\{\frac{\int\int_{\text{closed surface}} D_n\, ds}{\Delta\tau}\right\} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \qquad [1\cdot45]$$

This may be expressed in terms of potential since

$$D = \varepsilon E = -\varepsilon\frac{\partial V}{\partial l}$$

so that

$$D_x = -\varepsilon\frac{\partial V}{\partial x},\quad D_y = -\varepsilon\frac{\partial V}{\partial y},\quad D_z = -\varepsilon\frac{\partial V}{\partial z}$$

Now $cd = D_x$ and $ed = \Delta x \tan\theta$. But $\tan\theta = \dfrac{\partial D}{\partial x}$ at the point x. This we may write as $\dfrac{\partial D_x}{\partial x}$. Hence

$$ed = \Delta x\frac{\partial D_x}{\partial x}.$$

Thus to a first approximation

$$ef \simeq cd + de = D_x + \frac{\partial D_x}{\partial x}\Delta x$$

If Δx is very small, ef approaches zero and the approximation is more nearly the true value of the desired ordinate. Under such circumstances we may write for the value of D at $x + \Delta x$

$$D_x + \frac{\partial D_x}{\partial x}\Delta x$$

where x has any value.

Substituting for D_x, D_y, and D_z in equation 1·45 we have

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\varepsilon} \qquad [1·46]$$

Equation 1·46 is known as Poisson's equation. In regions in which there is no free charge, $\rho = 0$ and equation 1·46 becomes

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \qquad [1·47]^*$$

Equation 1·47 is known as Laplace's equation. Equations 1·46 and 1·47 apply to the magnetic field if V is replaced by U.

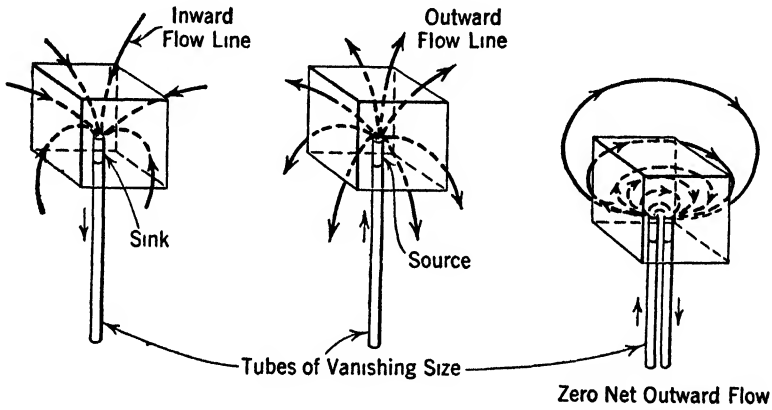


FIG. 1·6

1·15 The Divergence of a Vector

Equation 1·45 may be written more briefly by the use of a definition from vector analysis. In vector analysis we write

$$\underset{\Delta\tau \to 0}{\text{Limit}} \left\{ \frac{\iint_{\text{closed surface}} D_n \, ds}{\Delta\tau} \right\} = \text{div } \mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \qquad [1·48]$$

defining div **D**, pronounced divergence of **D**, or " div " **D**.

* In vector notation equations 1·45, 1·46, and 1·47 become, respectively:

$$\nabla \cdot \mathbf{D} = \rho$$
$$\nabla^2 V = -\rho/\epsilon$$
$$\nabla^2 V = 0$$

The divergence of a vector is a scalar quantity, since it identifies a magnitude only. It represents the outward flux of a vector per unit volume as the volume approaches zero at a given point. If this volume contains a small hole through which an incompressible liquid is flowing inward, the outward flow of the liquid through the remainder of the surface is positive and the divergence is positive. Refer to Fig. 1·6. If, on the other hand, the hole is absorbing the liquid then the divergence is negative. If the volume contains both source and sink, and the rate at which both operate is the same, the divergence is equal to zero.

The illustration just given is not rigorous, in that the volume is necessarily large in comparison to the incoming tube, which is itself of finite size. None the less the basic idea is illustrated. Moreover, the process of taking the limit as the volume approaches zero is invalid in the electrical case with which we are primarily concerned. Electric charge is not indefinitely divisible, the charge of the electron being the smallest charge which may have individual existence. Accordingly the divergence over any small closed surface will be zero or an integral multiple of the electronic charge. It is doubtful that a surface drawn to cut through the electron has physical significance.

1·16 The Divergence of the Electric and Magnetic Induction

The divergence of the electrical induction at any point in a medium where the charge density is finite is equal to the charge density ρ at the point.

$$\text{div}\,\mathbf{D} = \rho \ . \qquad [1\cdot49]$$

At points in a medium where the charge density is zero

$$\text{div}\,\mathbf{D} = 0 \qquad [1\cdot50]$$

The lines of force which represent the electrostatic field are considered to start and stop on electric charges. If free magnetic poles exist the conditions of the magnetostatic field are identical with those of the electrostatic field. If, as was formerly believed, no free magnetic poles exist we find that all lines of magnetic flux are continuous and in general

$$\text{div}\,\mathbf{B} = 0 \qquad [1\cdot51]$$

1·17 Electric and Magnetic Dipoles

If equal positive and negative electric charges occupy exactly the same point in space, the net electric field due to the charges is zero. If, however, there is a slight separation between the charges, the fields of the two no longer cancel. Such a combination is known as an electric doublet or dipole. The strength of a dipole is defined as the

product of the distance separating the charges and the strength of either charge. The value of the concept of the dipole lies in the fact that the separation of the charges is assumed small compared with the distance to other systems. Accordingly the properties of the dipole are completely defined by the strength and angular orientation of the dipole. This definition may be formulated in terms of potential.

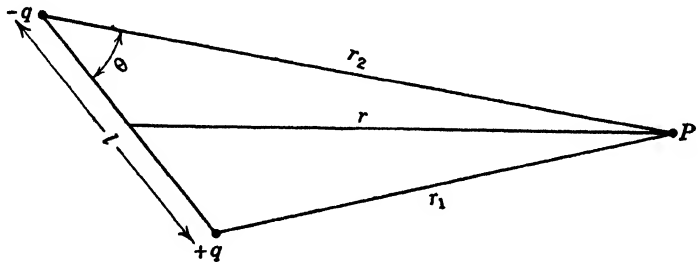


FIG. 1·7

The potential at the point P, Fig. 1·7, due to the positive charge as defined by equation 1·20 is

$$V_1 = \frac{q}{4\pi\varepsilon r_1}$$

Similarly the potential at P due to the negative charge is

$$V_2 = -\frac{q}{4\pi\varepsilon r_2}$$

The net potential is evidently the sum of these two values, or

$$\frac{q}{4\pi\varepsilon r_1} - \frac{q}{4\pi\varepsilon r_2}$$

This may be written

$$V = \frac{r_2 - r_1}{4\pi\varepsilon r_2 r_1} q$$

In so far as $r \gg l$ we may write

$$r_1 r_2 = r^2$$

and

$$r_2 - r_1 = l \cos\theta$$

With these substitutions the potential at P becomes

$$V = \frac{ql\cos\theta}{4\pi\varepsilon r^2} \qquad [1·52]$$

The line passing through the two charges is referred to as the *axis* of the dipole, the positive direction of this axis being in the direction from negative to positive charge. The product ql defines the strength or *electric moment* of the dipole. The length l is assumed to be negligible in comparison to all other dimensions of the system. Therefore we may say that the potential is zero at all points in a plane through the dipole and perpendicular to the axis, since the cosine of 90° is zero. The potential is a maximum at points which lie in the line of the axis.

The potential due to a magnetic dipole, derived in exactly the same way, is:

$$U = \frac{ml \cos \theta}{4\pi\mu r^2} \qquad [1\cdot53]$$

where ml is the strength or *magnetic moment* of the dipole.

1·18 The Polarization P and the Magnetization M

The behavior of a material dielectric in an electrostatic field may be accounted for in terms of the constant ϵ and in terms of the vector of polarization **P**. We have already considered the concept of dielectric constant in section 1·3. In section 1·11 we expressed the electric flux density quantitatively by the relation

$$\mathbf{D} = \epsilon\mathbf{E} \qquad [1\cdot31]$$

We can express the vector **D** in an alternative way. The application of the electric field may be thought of as resulting in a rearrangement of the molecular structure of the material. This may be expressed quantitatively by the relation

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \qquad [1\cdot54]$$

where the first term $\epsilon_0\mathbf{E}$ represents the induction which would occur if no dielectric were present. The second term **P** represents the contribution to the induction due to the presence of the dielectric. This equation is taken to define the vector quantity **P**. The two equations 1·31 and 1·54 may be combined to give

$$\epsilon = \epsilon_0 + \frac{\mathbf{P}}{\mathbf{E}}$$

or

$$\mathbf{P} = (\epsilon - \epsilon_0)\mathbf{E}$$

We usually think of a dielectric as comprising a random distribution of positive and negative charges in equal numbers bound together in an elastic medium. Such a random distribution can have a net external field equal to zero since the action of each charge may be canceled by

the combined actions of the adjoining charges. When an electric field is applied to such a dielectric the positive charges are, on the average, displaced a small distance in the direction of the field and the negative charges are displaced a similar small distance in the opposite direction. An elastic force tending to restore the charges to their original position accounts for the fact that the relation is linear and also, as will be shown later, that energy is stored in the dielectric.

This displaced condition may be interpreted in two distinct ways at our option. The original distribution may be expressed as coincident pairs of unlike charges, in which case the pairs are displaced by the field and a condition called polarization exists throughout the volume. The resulting physical system is indistinguishable from a uniform volume distribution of dipoles of suitable strength whose axes are in the direction of the applied electric field.

An alternative interpretation is based upon the conditions existing at the boundaries of the dielectric. If we omit for the moment a suitable number of charges near each boundary it is always possible to find a pairing of the remaining charges so that no external field is produced. A certain number of positive charges near one boundary acts with an equal number of negative charges at the other boundary to produce exactly the same effect as the volume distribution of dipoles previously discussed. The effect of these charges we have chosen to isolate is to annul part of the field in the dielectric which caused the displacement. Accordingly, the field effective in producing the displacement is less than that which would be required if the dielectric were not present.

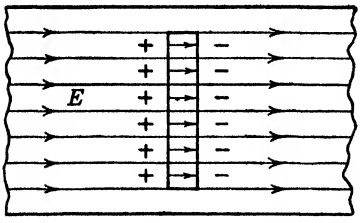


FIG. 1·8 Illustration of the increased electric intensity in a flat cavity in a dielectric material.

As an example let us consider the system of Fig. 1·8. A very narrow flat cavity, drawn enlarged for convenience, is located in a dielectric material of constant ε. In the dielectric the electric intensity is $\mathbf{E}$ and the induction is

$$\mathbf{D} = \varepsilon\mathbf{E}$$

The induction vector is continuous across the boundaries, and accordingly the electric intensity in the cavity is ε times as great as that in the dielectric. The intensity in the cavity, $\mathbf{E}'$, may also be written

$$\mathbf{E}' = \varepsilon_0\mathbf{E} + \mathbf{P}$$

The difference between $\mathbf{E}$ and $\mathbf{E}'$ is accounted for by the charges shown

at the surfaces. If σ is the density of charge at each surface and P_n the normal component of the polarization, it may be shown that $\sigma = P_n$. This follows from our use of rationalized units in which one unit charge produces one (not 4π) unit of induction. The geometry of the problem is such that the total induction passes from the positive to the negative charges directly, producing an induction or intensity numerically equal to the surface charge density.

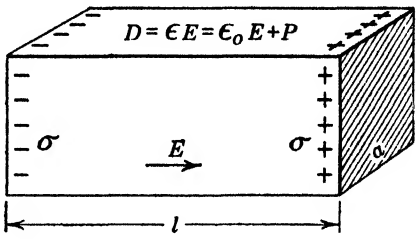


FIG. 1·9 Rectangular section of a homogeneous dielectric.

Consider the system of Fig. 1·9. The electric intensity **E** in the dielectric is uniform and produces an induction

$$\mathbf{D} = \varepsilon\mathbf{E} = \varepsilon_0\mathbf{E} + \mathbf{P}$$

The atomic displacement produced by this field results in an effective charge of density σ on the bounding surfaces. The electric moment of the entire block as a sort of enlarged dipole is equal to σal. But we have shown from the preceding work that $\sigma = P_n$. Accordingly the total moment is equal to $P_n al$. Therefore we deduce that the polarization **P** is expressible as a volume density of electric moment due to a distribution of dipoles.

The situation existing in the magnetic field is exactly analogous. We may write

$$\mathbf{B} = \mu\mathbf{H} = \mu_0\mathbf{H} + \mathbf{M} \qquad [1\cdot55]$$

where **M**, a vector called magnetization, analogous to **P**, is the volume density of magnetic moment of the equivalent dipoles.

$$\mathbf{M} = (\mu - \mu_0)\mathbf{H}, \qquad \mu = \mu_0 + \frac{\mathbf{M}}{\mathbf{H}} \qquad [1\cdot56]$$

1·19 Energy in the Electrostatic and Magnetostatic Fields

When a dielectric is inserted in an electric field, the displacements suffered in its atomic structure result in a condition very similar to that resulting from mechanical strain. As in the mechanical system, a certain definite amount of energy is stored per unit volume. This energy may be evaluated by assuming a uniform parallel field in the X direction. This condition exists in the familiar parallel-plate con-

denser. It will be recalled that the capacitance C of such a condenser is

$$C = \frac{\varepsilon A}{d} \qquad [1\cdot57]^*$$

where A is the area of the plates and d is the spacing between the plates. Also it may be shown that the energy W required to establish the charge is

$$W = \tfrac{1}{2}CV^2 = \frac{1}{2}\frac{\varepsilon A}{d}V^2 \qquad [1\cdot58]$$

where V is the potential difference between the two plates. Also $V = Ed$, since the field is uniform. Accordingly the total energy stored is

$$W = \frac{\varepsilon E^2}{2} Ad \qquad [1\cdot59]$$

Since the electric strain is equal in every part of the dielectric we are justified in dividing the total energy by the volume Ad. Hence the energy stored by the electric field per unit volume is

$$\frac{\varepsilon E^2}{2} \qquad [1\cdot60]$$

A similar development for the magnetic field gives for the energy stored per unit volume by the magnetic field

$$\frac{\mu H^2}{2} \qquad [1\cdot61]$$

Equations 1·60 and 1·61 are valuable because they permit us to evaluate the total energy stored in a system by integration of these quantities. Electrical energy is stored in all dielectrics and in vacuum but not in conductors. Magnetic energy is stored in all tangible materials, both paramagnetic and ferromagnetic, as well as in vacuum.

* The student is possibly more familiar with the relation

$$C = \frac{\varepsilon A}{4\pi d}$$

that is obtained when unrationalized units are used. With the rationalized system the 4π is included in the electric intensity.

SUMMARY OF CHAPTER 1

In the two columns below are summarized the laws and relations discussed in this chapter. The brief description of these relations that has been given is intended only as a review of electrostatic and magnetostatic theories. If any doubt remains the student should consult textbooks on these subjects.

UNITS AND EQUATIONS

Electrostatics	Magnetostatics
Coulomb's Law	
Electrostatic Units (cgs)	**Electromagnetic Units** (cgs)
$F_e = \frac{q_1 q_2}{\kappa_e r^2}$ dynes	$F_m = \frac{m_1 m_2}{\mu_m r^2}$ dynes
r in centimeters	r in centimeters
q in stat-coulombs	m in electromagnetic units of pole strength
In vacuum κ_e = unity and	In vacuum μ_m = unity and
$F_e = \frac{q_1 q_2}{r^2}$ dynes	$F_m = \frac{m_1 m_2}{r^2}$ dynes
Rationalized Practical Units	
$F_e = \frac{1}{4\pi} \cdot \frac{q_1 q_2}{\varepsilon r^2}$ newtons	$F_m = \frac{1}{4\pi} \frac{m_1 m_2}{\mu r^2}$ newtons
r in meters	r in meters
q in coulombs	m in webers
$\varepsilon = \kappa_e \varepsilon_0$	$\mu = \mu_m \mu_0$
$\varepsilon_0 = 8.854 \times 10^{-12}$ farad per meter	$\mu_0 = 1.257 \times 10^{-6}$ henry per meter
κ_e = the dielectric constant as given in standard tables	μ_m = the permeability as given in standard tables
In vacuum	In vacuum
$F_e = \frac{1}{4\pi} \frac{q_1 q_2}{\varepsilon_0 r^2}$ newtons	$F_m = \frac{1}{4\pi} \frac{m_1 m_2}{\mu_0 r^2}$ newtons
Electric Field Strength	
Electrostatic Units (cgs)	**Magnetostatic Units** (cgs)
$E = \frac{q}{\kappa_e r^2}$ stat-volts per cm	$H = \frac{m}{\mu_m r^2}$ gilberts per cm
r in centimeters	r in centimeters
q in stat-coulombs	m in emu of pole strength

In vacuum

$$E = \frac{q}{r^2} \quad \text{stat-volts per cm}$$

In vacuum

$$H = \frac{m}{r^2} \quad \text{gilberts per cm}$$

Rationalized Practical Units

$$E = \frac{1}{4\pi\varepsilon_0}\frac{q}{\kappa_e r^2} \quad \text{volts per meter}$$

r in meters
q in coulombs

$$H = \frac{1}{4\pi\mu_0}\frac{m}{\mu_m r^2} \quad \text{ampere-turns per meter}$$

r in meters
m in webers

In vacuum

$$E = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2} \quad \text{volts per meter}$$

In vacuum

$$H = \frac{1}{4\pi\mu_0}\frac{m}{r^2} \quad \text{ampere-turns per meter}$$

Potential

Electrostatic Units (cgs)

$$V = \frac{q}{\kappa_e r} \quad \text{stat-volts}$$

r in centimeters
q in stat-coulombs

Magnetostatic Units (cgs)

$$U = \frac{m}{\mu_m r} \quad \text{ergs per pole}$$

r in centimeters
m in emu of pole strength

In vacuum

$$V = \frac{q}{r} \quad \text{stat-volts}$$

In vacuum

$$U = \frac{m}{r} \quad \text{ergs per pole}$$

Rationalized Practical Units

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{\kappa_e r} \quad \text{volts}$$

r in meters
q in coulombs

$$U = \frac{1}{4\pi\mu_0}\frac{m}{\mu_m r} \quad \text{joules per pole}$$

r in meters
m in webers

In vacuum

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{r} \quad \text{volts}$$

In vacuum

$$U = \frac{1}{4\pi\mu_0}\frac{m}{r} \quad \text{joules per pole}$$

The Potential May Be Expressed As

$$V = -\int_l E \cos\theta \, dl \qquad U = -\int_l H \cos\theta \, dl$$

The Field Intensity May Be Expressed As

$$E = -\frac{dV}{dl} \qquad H = -\frac{dU}{dl}$$

The Flux Density or Induction Is Defined As

$$\mathbf{D} = \varepsilon\mathbf{E} \qquad \mathbf{B} = \mu\mathbf{H}$$

The Total Flux Is Defined As

$$\psi = \int\int_s D_n \, ds \qquad \phi = \int\int_s B_n \, ds$$

The Divergence of the Induction Vector in Unrationalized Units

$$\text{div } \mathbf{D} = 0 \quad \text{or} \quad 4\pi\rho \qquad \text{div } \mathbf{B} = 0$$

In Rationalized Units

$$\text{div } \mathbf{D} = 0 \quad \text{or} \quad \rho \qquad \text{div } \mathbf{B} = 0$$

The Energy per Unit Volume in Unrationalized Units

$$\frac{\epsilon E^2}{8\pi} \qquad \frac{\mu H^2}{8\pi}$$

In Rationalized Units

$$\frac{\epsilon E^2}{2} \qquad \frac{\mu H^2}{2}$$

PROBLEMS

When electric charges are uniformly distributed over a spherical surface, the force exerted upon a charge at a point outside the surface is the same as if the charges on the sphere were all concentrated at the center of the sphere. The net force exerted upon an exploring charge placed inside the charged sphere is zero. In the problems below, which involve spherical charged bodies, this proposition may be used because the charges in question are assumed to be uniformly distributed over the spherical surface.

1·1 Two small charged metal balls are placed 0.5 meter apart. The charge on one of them is 5×10^{-8} coulomb, and the charge on the other is 10^{-9} coulomb. Both are charged positively. Calculate the force on each ball, and express it in pounds, dynes, newtons, and grams.

1·2 Two similar metal balls each weighing 0.1 gram have equal charges of 10^{-8} coulomb. The balls are suspended from two points 5 cm apart by silk threads of equal length and negligible weight. If the force of repulsion between the charges holds them 7 cm apart, how long is the suspending thread?

1·3 Two metal balls having a diameter of 3 cm are placed 1 meter apart between centers. If the charge on each is 10^{-8} coulomb, calculate and plot the electric field intensity for points between them on the line of centers: (*a*) when they are both charged positively; (*b*) when one is positive, the other negative.

1·4 Calculate and plot the potential in volts between the two charged balls of problem 3 on their line of centers if both are charged positively.

1·5 Find the magnetic intensity and potential at a point located on a line making an angle of 30° with the axis of a small thin magnet and passing through its center. The length of the magnet is 0.1 meter, and the magnetic pole strength is m. Let the distance between the center of the magnet and the point in question be 1 meter.

1·6 Two magnets are fixed in space, both being perpendicular to a line joining their centers, which are 0.4 meter apart. One has a length between poles of 0.2

meter and pole strengths of 10^{-5} weber. The second magnet has a length between poles of 0.01 meter and pole strengths of 10^{-7} weber. Using Coulomb's law, calculate the net force on each pole of the small magnet due to the action of both poles of the large magnet.

1·7 In the arrangement of problem 6, calculate the forces by first evaluating the field at the small magnet due to the action of the larger magnet. Use equation 1·8.

1·8. Two iron bars are identical in appearance, hardness, and other easily observed characteristics, but one is a permanent magnet and the other has the magnetic properties of soft iron. How may the two be identified without the use of any additional equipment?

1·9 An extremely slender bar magnet is relatively long, and one of its poles is used to simulate an isolated north pole. How long must the bar be made if the field produced by it differs only 1 per cent from the field due to an isolated pole of strength equal to that of the magnet, at a distance 5 cm from the pole and along the axis of the bar?

1·10 A certain glass has a dielectric constant κ_e equal to 2.5. Evaluate the polarization vector **P** for this material in mks units.

1·11 A spherical cavity is cut in a polarized dielectric. Determine the electric intensity at the center of the sphere due to the surface charge on the cavity.

1·12 A particular grade of iron has a permeability μ_m equal to 1000. Evaluate the magnetization vector **M** for this material in mks units.

1·13 A narrow flat cavity is cut at right angles to the direction of the magnetic induction in a ferromagnetic solid. Determine the intensity in the cavity due to the magnetic poles which appear at the surfaces.

1·14 A particular material has a dielectric constant κ_e equal to 4.0 and will withstand a maximum electric gradient of 100,000 volts per meter. Calculate the energy in joules per cubic meter which may be stored in this dielectric when used in a condenser.

1·15 A metallic sphere has a radius equal to a and is suspended in air which will withstand a maximum electric gradient of 10^6 volts per meter. How large an electric charge may be placed on this sphere before the air breaks down and causes discharge?

CHAPTER 2

THE ELECTROMAGNETIC EQUATIONS

2·1 The Electron Theory of the Electric Current

According to the electron theory, conducting bodies include in their structure a large number of charged particles or electrons which are free to move through the conductor when acted upon by an electric force. The movement of these particles is considered to be quite erratic. This is caused by the numerous collisions which the electrons suffer when they are set in motion. They collide with one another and with the relatively stationary atoms and molecules of the conductor. There is, however, a steady net drift of the charged particles in the direction of the force. This drift constitutes the flow of electric current.

It is an unfortunate fact that the conventional direction of current is opposite to the actual electron motion, since the electron is the principal carrying agent in most forms of conduction. This convention is a result of early work in electrostatics in which the positive direction of charge and current were defined. It was found that the charge produced by friction between certain pairs of materials always had a definite polarity, and the positive sign was arbitrarily applied to the charge of one of a pair of such materials. In particular, glass and similar materials become positively electrified if they are rubbed with a cloth of cotton or silk. Other materials such as resin or amber are negatively electrified by the same treatment. The terms "vitreous" and "resinous" for the two polarities were once in common use. Electrical effects due to friction or mechanical fracture are referred to as *triboelectric* effects.

2·2 The Continuity Equation for Current

Let us imagine a small volume element $\Delta\tau$, having dimensions Δx, Δy, Δz, placed within a conductor, one corner of which is at the point x, y, z. Let us observe the motion of the free charges within this volume under the action of an electric force. The motion of the electrons through the conductor may be expressed in terms of a current density ι having components ι_x, ι_y, ι_z, at the point x, y, z, in the direction of the axes as shown in Fig. 2·1. Let the average current density over the face $\Delta y\ \Delta z$ be $\bar{\iota}_x$. The total current entering the volume from

the x direction is thus $\bar{\imath}_x \Delta y \Delta z$. The current leaving the opposite face is

$$\left(\bar{\imath}_x + \frac{\partial \bar{\imath}_x}{\partial x} \Delta x\right) \Delta y \Delta z \qquad [2\cdot1]^*$$

The difference between the current entering and the current leaving in the x direction is

$$\bar{\imath}_x \Delta y \Delta z - \left(\bar{\imath}_x + \frac{\partial \bar{\imath}_x}{\partial x} \Delta x\right) \Delta y \Delta z = - \frac{\partial \bar{\imath}_x}{\partial x} \Delta x \Delta y \Delta z \qquad [2\cdot2]$$

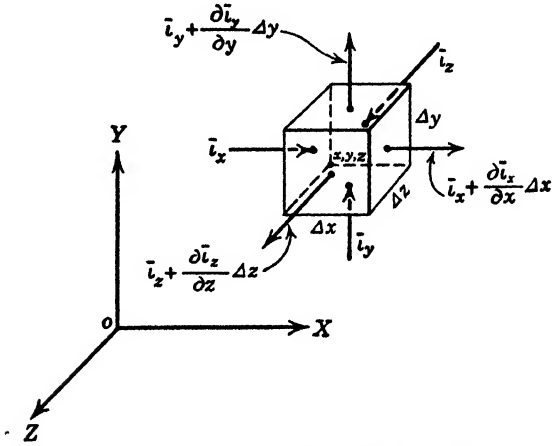


FIG. 2·1 An infinitesimal volume $\Delta\tau = \Delta x \Delta y \Delta z$ at a point x, y, z inside a conductor through which a current is flowing. The bar over the current-density vectors indicates the average value of the vectors over the various faces of the volume.

Similarly, the difference between the current entering and the current leaving the small volume taken in the y direction is

$$- \frac{\partial \bar{\imath}_y}{\partial y} \Delta x \Delta y \Delta z \qquad [2\cdot3]$$

* This relation may be obtained by application of Taylor's theorem for the expansion of a function about a point. Assuming proper continuity this gives

$$\bar{\imath}_x + \frac{\partial \bar{\imath}_x}{\partial x} \Delta x + \frac{\partial^2 \bar{\imath}_x}{\partial x^2} \frac{\overline{\Delta x^2}}{2!} + \frac{\partial^3 \bar{\imath}_x}{\partial x^3} \frac{\overline{\Delta x^3}}{3!} + \cdots$$

Neglecting second- and higher-order effects, we have

$$\bar{\imath}_x + \frac{\partial \bar{\imath}_x}{\partial x} \Delta x$$

for the current at a distance Δx from the point where the current is $\bar{\imath}_x$. See also footnote on page 14.

and taken in the z direction is

$$-\frac{\partial \bar{\iota}_z}{\partial z}\Delta x\,\Delta y\,\Delta z \qquad [2\cdot4]$$

The difference between the current entering and the current leaving the volume $\Delta\tau$ is

$$-\left(\frac{\partial \bar{\iota}_x}{\partial x}+\frac{\partial \bar{\iota}_y}{\partial y}+\frac{\partial \bar{\iota}_z}{\partial z}\right)\Delta x\,\Delta y\,\Delta z \qquad [2\cdot5]$$

which is equal to the rate of change of charge inside the volume. If the average charge density in the volume $\Delta\tau$ is $\bar{\rho}$, the charge in the volume $\Delta\tau$ is $\bar{\rho}\,\Delta x\,\Delta y\,\Delta z$. The rate of change of charge within the volume is then

$$\frac{\partial \bar{\rho}}{\partial t}\Delta x\,\Delta y\,\Delta z \qquad [2\cdot6]$$

Hence, equating 2·5 and 2·6, we obtain

$$\left(\frac{\partial \bar{\iota}_x}{\partial x}+\frac{\partial \bar{\iota}_y}{\partial y}+\frac{\partial \bar{\iota}_z}{\partial z}\right)\Delta x\,\Delta y\,\Delta z=-\frac{\partial \bar{\rho}}{\partial t}\Delta x\,\Delta y\,\Delta z \qquad [2\cdot7]$$

To reduce to a point relation, let us divide through by $\Delta x\ \Delta y\ \Delta z$ and take the limit as Δx, Δy, and Δz approach zero. The average values then approach their values at the point x, y, z, and we have

$$\frac{\partial \iota_x}{\partial x}+\frac{\partial \iota_y}{\partial y}+\frac{\partial \iota_z}{\partial z}=-\frac{\partial \rho}{\partial t} \qquad [2\cdot8]$$

This equation is correct in the practical or any other consistent set of units.

If the rate of change of charge is zero, then

$$\frac{\partial \iota_x}{\partial x}+\frac{\partial \iota_y}{\partial y}+\frac{\partial \iota_z}{\partial z}=0 \qquad [2\cdot9]$$

Equation 2·8 may be expressed as

$$\operatorname{div}\boldsymbol{\iota}=-\frac{\partial \rho}{\partial t} \qquad [2\cdot10]^*$$

* This will follow from the definition of divergence given in Chapter 1, since the right-hand side of 2 · 8 represents the limit, as $\Delta\tau$ approaches zero, of the surface integral of $\boldsymbol{\iota}$ over $\Delta\tau$, divided by $\Delta\tau$, i.e.

$$\operatorname{div}\boldsymbol{\iota}=\lim_{\Delta\tau\to 0}\left[\frac{\int\int_s \iota_n\,ds}{\Delta\tau}\right]=\frac{\partial \iota_x}{\partial x}+\frac{\partial \iota_y}{\partial y}+\frac{\partial \iota_z}{\partial z}$$

and equation 2·9 as

$$\text{div } \iota = 0 \qquad [2\cdot11]$$

This equation states that, if there is no accumulation or loss of charge, the amount of current entering a volume is exactly equal to the amount leaving the volume. This is Kirchhoff's first law. If, however, there is a change of charge, it is equal to the difference between the inflowing and outflowing current. This is equivalent to saying that electricity can be neither created nor destroyed.

2·3 The Static, Steady, and Non-Steady States

When electric or magnetic charges are held fixed in a medium, and there is no motion of any kind, the conditions are referred to as static. The results derived in Chapter 1 are based upon the static state. The act of carrying test charges about, of course, constitutes a movement. We are concerned, however, not with the conditions arising during this movement, but with the static conditions arising at each point at which the exploring charges are placed.

When a continual drift of electric charges has been taking place for some time the conditions that exist are referred to as the steady state. When a battery has been connected to a circuit for sufficient time for all transient effects to have died down, the steady state is obtained.

The non-steady state refers to the interval during which transient effects are being observed, as during the interval directly after the closing of a switch in an electric circuit, or the application of electric forces to a medium. The conditions which exist at all times in an alternating-current system are properly referred to as the non-steady state. Commonly, however, the condition in which the voltage and current are periodic functions of time is referred to as the steady state of alternating current.

2·4 Ohm's Law

By means of the electron theory we may develop an analytical formulation of the relations existing between the free electrons and the electric field intensity within a conductor. For our analysis let us take a small section of a wire conductor at right angles to its length, and observe the electrons in this section under the influence of an electric field. We shall assume that steady-state conditions prevail, and that there is an average charge density ρ moving with a velocity $\mathbf{v}$ through this section. The product $\rho\mathbf{v}$ at any point in the conductor is called the current density ι at that point. But the electrons experience continuous succession of collisions in passing through the conductor. It is to these repeated collisions that the heating effect of an electric

current is attributed. They represent a form of resistance to the flow of the current. The flow of electrons is sometimes compared with the movement of marbles through a viscous liquid. This analogy is very useful in considering the action of electrons in metals when the field intensity is varying quite rapidly, since it helps one to visualize how the velocity of the electrons may lag the field variations. Certainly at lower frequencies or with steady fields, the average velocity of the electrons is proportional to the field intensity **E**. Since ρ is the average density of charge, then $\iota = \rho \mathbf{v}$ is proportional to **E**. The proportionality constant σ is called the conductivity of the conductor and we may write

$$\iota = \sigma \mathbf{E} \qquad [2\cdot12]$$

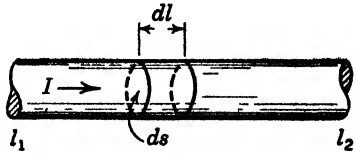


FIG 2 2 A section of a cylindrical conductor of area ds through which a current I is flowing.

Values of the conductivity σ, in mhos per meter, for common conductors and dielectrics are given in Table 2·1. Since the total current I flowing in the wire is $I = \iota\, ds$ (Fig. 2·2), we have

$$I = \iota\, ds = \sigma E\, ds \qquad [2\cdot13]$$

Since the current is the same through all sections of the wire we may multiply and divide 2·13 by dl, obtaining

$$I = \sigma E\, ds \frac{dl}{dl} = \sigma \frac{ds}{dl} E\, dl \qquad [2\cdot14]$$

Now $\sigma ds/dl$ is a constant depending on the conductivity and dimensions of the conductor. Let us set it equal to $1/dR$ so that we obtain

$$I\, dR = E\, dl \qquad [2\cdot15]$$

TABLE 2 · 1

CONDUCTIVITY OF COMMON MATERIALS AT 20° C.

Metals	mhos/meter	Dielectrics	mhos/meter
Aluminum	3.5×10^7	Bakelite	10^{-8} to 10^{-10}
Copper	5.8×10^7	Celluloid	10^{-8}
Gold	4.1×10^7	Glass	10^{-12}
Iron	1.0×10^7	Hard rubber	10^{-14} to 10^{-16}
Lead	0.5×10^7	Mica	10^{-11} to 10^{-15}
Nichrome	0.1×10^7	Paraffin	10^{-14} to 10^{-16}
Nickel	1.3×10^7	Porcelain	3×10^{-13}
Silver	6.1×10^7	Wood, paraffined	10^{-8} to 10^{-11}
Tin	0.9×10^7	Petroleum	10^{-14}
Tungsten	1.8×10^7	Distilled water	2×10^{-4}
Zinc	1.7×10^7	Sea water	3 to 5

The data in this table have been taken largely from the *Smithsonian Physical Tables*. For more complete data and for other materials see this reference.

Since I and E are constant with respect to l in a uniform conductor, we may write

$$I \int_{l_1}^{l_2} dR = E \int_{l_1}^{l_2} dl \qquad [2\cdot16]$$

or

$$IR = El_2 - El_1 \qquad [2\cdot17]$$

But $(El_2 - El_1)$ is the work done in carrying a unit charge from l_1 to l_2, and this has been defined as the potential difference between l_1 and l_2. Let this potential difference equal V. Then we obtain the familiar form of Ohm's law:

$$V = RI \qquad [2\cdot18]$$

where V is the total voltage drop, I is the total current, and R is the total resistance of the conductor.

2·5 The Magnetic Shell and the Unit Current

Oersted discovered in 1819 that a conductor carrying a current of electricity would deflect a magnetic needle placed near by. The fact that an electric current produces magnetic effects provides the link between the subjects of electricity and magnetism.

Shortly after Oersted's discovery, Ampère showed that a small loop of wire carrying a current produced a magnetic field similar to that due to a short magnet. It was found that, at distances from the loop large in comparison to the linear dimensions of the loop, the magnetic effect observed was proportional to the area of the loop and to the current flowing in it. In practice we must be able to calculate the magnetic effect at much smaller distances. The following considerations, due to Ampère, permit such calculation.

Let us consider the circuit of Fig. 2·3, which consists of a wire, having any arbitrary shape, carrying a current I. Let us imagine the circuit divided into a large number of small square loops or meshes, each of which carries the same current as that which actually exists in the boundary circuit. Now, if each of the small imaginary circuits has its sides infinitely close to its neighboring circuits, the effects of the currents will neutralize or cancel one another at all places except in the outside bounding circuit. Also we may choose these loops as small as we desire. Hence they may be chosen so as to be small in comparison with the distance to any point near the surface formed by the loops. Now the effect of each small mesh is equivalent to a magnetic doublet previously described, and the total magnetic effect of the true current in the bounding wire is equal to the vector sum of the effects of all the magnetic doublets or meshes. Taken altogether the magnetic doublets

constitute a magnetic shell, that is, a thin lamina of arbitrary shape, the boundary of which coincides with the conductor forming the boundary current.

A magnetic shell may be thought of as formed by the aggregation of many like magnetic dipoles arranged in such a way as to resemble an infinitely thin slab cut from a magnet. The magnetic moment of a dipole is defined as the product of its length by its pole strength. The moment of a magnetic shell is defined as the sum of the moments of the dipoles in a unit area or as the product of the thickness of the shell by the number of poles per unit area.

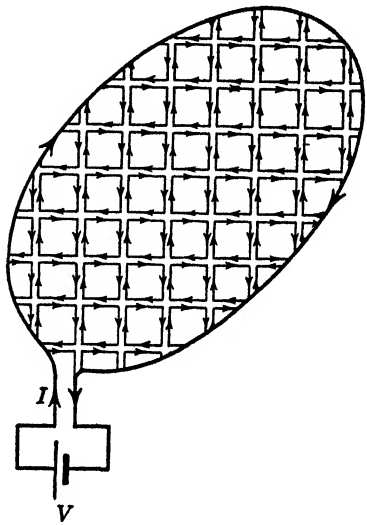


FIG. 2·3 Imaginary rectangular loop circuits which completely fill the closed area formed by the physical circuit I. The small loops may be considered of vanishing size, and infinite in number. Each carries the same current as the external peripheral circuit.

This magnetic shell is equivalent in its magnetic effect to an electrical current whose circuit is its periphery. The magnetic effect of each element of the shell is $M\ da/4\pi\mu$, and of the whole shell $MA/4\pi\mu$, where M is the magnetic moment per unit area, da is the area of each mesh, and A is the total area of the shell.* Also the magnetic effect of each mesh is $I\ da/4\pi$, or $IA/4\pi$ for the boundary circuit. Hence the unit electric current may be defined as one which will produce the same magnetic effect at external points as a magnetic shell of unit strength whose boundary coincides with the current provided that the medium is vacuum.

Consider the circuit of Fig. 2·4, in which a current of I amperes is flowing. The equivalent magnetic shell for this circuit will, by the definition of unit current, have a magnetic moment M per unit area, where $M = \mu I$. Hence the magnetic moment for a small area da about a point Q on the shell will be $M\ da/4\pi$. This small section of the shell

* The field due to a free electric or magnetic pole varies inversely with the dielectric constant or permeability of the medium surrounding the pole. The magnetic field due to an electric current is independent of the surrounding material. Accordingly the strength of the magnetic shell and of the dipoles which comprise it must be made proportional to the permeability of the surrounding medium. In this way the field and potential distribution which result are independent of the permeability.

is exactly equivalent to a small magnetic dipole. Now at a point P, a distance r from Q, the magnetic potential due to this elementary dipole is, by equation 1·53,

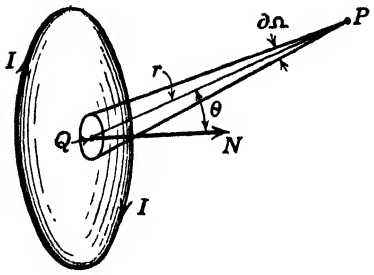


FIG. 2·4

$$dU = \frac{M\, da \cos\theta}{4\pi r^2 \mu} \qquad [2\cdot19]$$

where θ is the angle between r and the normal N to the area da. The solid angle subtended at P by the area da is $d\Omega = (da \cos\theta / r^2)$. Hence

$$dU = \frac{M\, d\Omega}{4\pi\mu} \qquad [2\cdot20]$$

Integrating over the whole shell, the potential at P is

$$U = \frac{M\Omega}{4\pi\mu} \qquad [2\cdot21]$$

where Ω is the solid angle subtended at P for the whole shell. Since the circuit forms the boundary for the shell, the solid angle is the same for circuit or shell. Therefore

$$U = \frac{I\Omega}{4\pi} \qquad [2\cdot22]$$

As the point P approaches the surface of the shell, the solid angle approaches 2π and $U = I/2$ at the surface.

2·6 Magnetomotive Force and Ampère's Law

The magnetomotive force may be defined as the work done by the magnetic field as a unit magnetic pole is moved along a path joining two points P and Q in the field. This may be expressed by the line integral

$$\text{MMF} = \int_P^Q H \cos\theta\, dl \qquad [2\cdot23]$$

where θ is the angle between the direction of H and the path element dl. In the field external to permanent magnets, a problem in the static state, this integral taken around any closed path is zero. When we consider the magnetic field due to the flow of current in a wire, a problem in the steady, not static, state, we find a different situation where the truth of this statement depends upon the path taken. Let us consider, for example, the magnetic field produced by a straight

wire carrying current. If our path of integration encloses the wire, as, for example, a circular path about the wire, then the integral is no longer zero. This is true for any circuit carrying a current, as will be shown.

Consider the circuit drawn in Fig. 2·5. Let us imagine a surface formed so that its outside boundary is the wire of the circuit. We may imagine this surface to be of the type that would be produced by dipping our wire circuit into a pail of soapy water. If we restrict the paths along which we calculate our line integral from cutting the surface, then all closed paths along which we calculate the integral will be zero. Let us calculate the value of equation 2·23 for a path which cuts the surface and includes within it the wire carrying the current I. First we may calculate the line integral for a path starting at a point P in front of the surface shown in Fig. 2·5, and proceeding around C to a point Q on the other side of the surface. The difference in magnetic potential between P and Q is

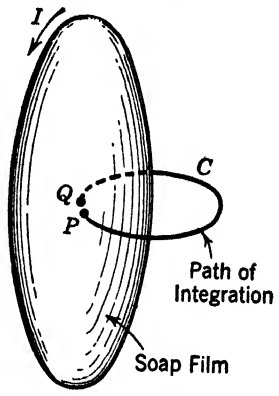


FIG. 2·5 Interpretation of a current-carrying circuit in terms of the magnetic shell.

$$U_P - U_Q = \int_P^Q H \cos\theta \, dl \qquad [2\cdot24]$$

If we let the point P approach Q the potential difference does not approach zero since then our path would cut the surface. It was shown in section 2·5, where our circuit was replaced by a magnetic shell, that, as P approaches the surface of the shell, the angle Ω subtended at P by the surface approaches 2π and U_P approaches $M/2$ or $I/2$. Similarly, as Q approaches the surface, U_Q approaches $-M/2$ or $-I/2$. Hence

$$U_P - U_Q = \pm M \quad \text{or} \quad \pm I \qquad [2\cdot25]$$

Or

$$\oint_l H \cos\theta \, dl = \pm I \qquad [2\cdot26]$$

Thus if we go around the path in such a direction that the integral is positive we have

$$\oint_l H \cos\theta \, dl = I \qquad [2\cdot27]$$

where the integration is carried once around a path enclosing the wire. If we encircle the wire n times, the work done is In.

It must not be assumed from the foregoing that the magnetomotive force is located at the surface of the equivalent magnetic shell. Such an assumption is obviously illogical, in view of the fact that the magnetic shell was arbitrarily chosen in position, and is contrary to all physical experience. Actually work is done upon a test pole throughout its motion, and the magnetomotive force is accordingly of a distributed nature.

In the practical system the mmf in a closed path linking unit current is stated simply as one ampere-turn. The strength of the unit pole is such that 1 joule of work is required to carry a unit pole once around a unit current.

2·7 Deductions from Ampère's Law

Consider an isotropic medium such as a large conducting mass of metal through which an electric current is flowing. Let us set up in this medium the coordinate axes shown in Fig. 2·6. The current may

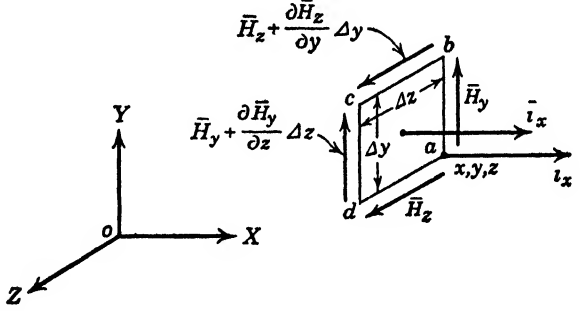


FIG. 2·6

be considered to be a vector function of position with respect to these axes and may have different values at different points in the space. Let the current density at any point x, y, z, be ι with components ι_x, ι_y, and ι_z along the three axes, respectively. Construct the small rectangular area $abcd$ so that the corner a of this rectangle is at the point x, y, z. Let the side $ab = dc = \Delta y$, and let the side $ad = bc = \Delta z$. For convenience this area is shown greatly enlarged in Fig. 2·6.

At the point x, y, z, the component of current density perpendicular to the area is ι_x. The remaining components of current density at x, y, z, that is, ι_y, and ι_z, contribute nothing to the magnetomotive force around the area. Now the components of current density normal to the area may be different for different points in the area. Let $\bar{\iota}_x$ be the average normal component of current density through the area.

The total current through the area is then

$$I = i_x \, \Delta y \, \Delta z \qquad [2\cdot28]$$

Let us consider the magnetic intensity acting along the edges of the area *abcd*. If we place a unit magnetic pole at the point x, y, z, it is acted upon by a magnetic force **H**, having components H_x, H_y, and H_z along the three axes. These components may vary for different points in the medium. This can be expressed mathematically by saying that **H** is a function of x, y, and z; that is

$$\mathbf{H} = \mathbf{H}(x, y, z) \qquad [2\cdot29]$$

Let us define the average magnetic field intensity acting along the edges *ab* and *ad* of our area as $\overline{H}_z$ and $\overline{H}_y$, respectively, having the arrow direction indicated. We wish to find the values of the magnetic field intensity acting along edges *bc* and *dc*. This we can do with the aid of Taylor's theorem if we assume that our function $\mathbf{H}(x, y, z)$ obeys the continuity requirements of the theorem. Now the value of the field intensity along the edge *dc* will be H_y plus the change in H_y as we move the distance Δz from *ab* to *dc*. Assuming proper continuity and neglecting second-order effects, Taylor's theorem tells us that the average intensity acting along *dc* is

$$\overline{H}_y + \frac{\partial \overline{H}_y}{\partial z} \Delta z \qquad [2\cdot30]$$

Similarly the average magnetic intensity along *bc* will be

$$\overline{H}_z + \frac{\partial \overline{H}_z}{\partial y} \Delta y \qquad [2\cdot31]$$

We will now determine the amount of work done by the magnetizing forces when a unit magnetic pole is carried around the edges of the area, starting at *a*, moving to *b*, then through *c* and *d*, and back to *a*. The work done by a force in displacing its point of application is the product of the force and the displacement in the direction of the force. Hence the work done is

$$\oint_l H \cos\theta \, dl =$$

$$\underbrace{\overline{H}_y \, \Delta y}_{a \text{ to } b} + \underbrace{\left[\overline{H}_z + \frac{\partial \overline{H}_z}{\partial y} \Delta y\right] \Delta z}_{b \text{ to } c} - \underbrace{\left[\overline{H}_y + \frac{\partial \overline{H}_y}{\partial z} \Delta z\right] \Delta y}_{c \text{ to } d} - \underbrace{\overline{H}_z \, \Delta z}_{d \text{ to } a} \qquad [2\cdot32]^*$$

* Since H, in this case, is everywhere along the path of integration, $\theta = 0$ or 180°. Hence $\cos\theta = 1$ or -1, accounting for the choice of sign in this equation.

According to Ampère's law the work in practical units is equal to the current flowing normally through this infinitesimal area. This current is given in equation 2·28, and thus we have, equating 2·32 and 2·28,

$$\overline{H}_y\,\Delta y + \overline{H}_z\,\Delta z + \frac{\partial \overline{H}_z}{\partial y}\Delta y\,\Delta z - \overline{H}_y\,\Delta y - \frac{\partial \overline{H}_y}{\partial z}\Delta y\,\Delta z - \overline{H}_z\,\Delta z = \bar{\iota}_x\,\Delta y\,\Delta z \qquad [2\cdot33]$$

Dividing through by $\Delta y\,\Delta z$, and taking the limit as Δy and Δz approach zero, we obtain

$$\underset{\Delta y\,\Delta z \to 0}{\text{Limit}} \oint_l \frac{H\cos\theta\,dl}{\Delta y\,\Delta z} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \iota_x \qquad [2\cdot34]$$

where the average values of the magnetic intensity and current approach their values at the point x, y, z, in the limit.

By constructing similar mathematical areas $\Delta x\,\Delta z$ and $\Delta x\,\Delta y$, and then from geometry estimating the work done, we may obtain by Ampère's law the relations for the y and z components of the current:

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \iota_y \qquad [2\cdot35]$$

and

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \iota_z \qquad [2\cdot36]$$

Equations 2·34, 2·35, and 2·36 may be expressed more briefly by the single vector equation

$$\nabla \times \mathbf{H} \equiv \mathbf{curl}\,\mathbf{H} = \boldsymbol{\iota} \qquad [2\cdot37]^*$$

* The symbol $\nabla \times \mathbf{H}$ (pronounced nabla cross H) is often used in place of the symbol **curl H**. The curl of a vector is defined by

$$\underset{\Delta s \to 0}{\text{Limit}} \left[\frac{\oint_l H\cos\theta\,dl}{\Delta s}\right] = \mathbf{curl}\,\mathbf{H}$$

If **curl H** is zero, the vector field is referred to as irrotational. In the static cases considered in Chapter 1, the fields were irrotational. When **curl H** is not equal to zero, as in this section, the field is said to be rotational. A convenient definition for curl of a vector is

$$\mathbf{curl}\,\mathbf{H} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix}$$

and the symbol $\nabla = \mathbf{i}\dfrac{\partial}{\partial x} + \mathbf{j}\dfrac{\partial}{\partial y} + \mathbf{k}\dfrac{\partial}{\partial z}$.

In the familiar unrationalized units $\mathbf{curl}\ \mathbf{H} = 4\pi\iota$. If $\mathbf{H}$ is measured in emu and ι is in esu, the equation becomes

$$\mathbf{curl}\ \mathbf{H} = \frac{4\pi\iota}{c}$$

where **curl H** is a vector having x, y, and z components defined by

$$\left.\begin{aligned} \mathrm{curl}_x H &= \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \\ \mathrm{curl}_y H &= \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \\ \mathrm{curl}_z H &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \end{aligned}\right\} \qquad [2{\cdot}38]$$

That is,

$$\mathbf{curl}\ \mathbf{H} = \mathbf{i}\ \mathrm{curl}_x H + \mathbf{j}\ \mathrm{curl}_y H + \mathbf{k}\ \mathrm{curl}_z H$$

2·8 The Electromotive Force Equation and Faraday's Law

The electromotive force may be defined as the work done by the electric field as a small electric charge is moved along a path joining two points P and Q in the field divided by the charge. This may be expressed by the line integral

$$\mathrm{EMF} = \int_P^Q E \cos\theta\, dl \qquad [2{\cdot}39]$$

where θ is the angle between the direction of E and the path element dl. In the field of static charges this integral taken around any closed path is zero.

Faraday, in 1831, discovered that, when a conducting circuit is placed in a magnetic field whose intensity is varying, an electromotive force is developed in the circuit, causing a current to flow. This electromotive force is proportional to the rate at which the magnetic flux is changing and varies with the dimensions and configuration of the circuit. Thus

$$\mathrm{EMF} = -\frac{\partial \phi}{\partial t} \qquad [2{\cdot}40]$$

The magnetic flux ϕ is the total flux linking the circuit; it is given by the relation

$$\phi = \int\int_s B_n\, ds$$

where the surface integral is taken over the area bounded by the current and ds is an element of the area. B_n is the component of the flux

density normal to the element ds. Relation 2·40 is known as Faraday's law for electromagnetic induction. From 2·39 and 2·40 we have

$$\oint_l E \cos\theta \, dl = -\frac{\partial \phi}{\partial t} = -\frac{\partial}{\partial t}\int\int_s B_n \, ds \qquad [2\cdot41]$$

in any consistent set of units. If E is measured in esu and B in emu, equation 2·41 becomes

$$\oint_l E \cos\theta \, dl = -\frac{1}{c}\frac{\partial}{\partial t}\int\int_s B_n \, ds$$

2·9 Deductions from Faraday's Law

The presence of a physical conductor in the changing magnetic field is not in any way a necessary condition for the existence of an electric field. Actually, wherever there is a magnetic field changing with time, it must necessarily be accompanied by an electric force given by relation 2·41.

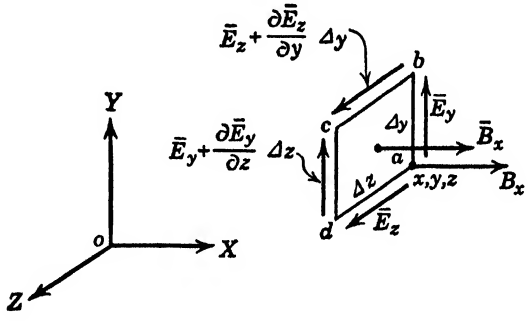


FIG. 2·7

To reduce this equation to a point relation we may proceed as in section 2·7. The relation between flux density **B** and the total flux ϕ is analogous to the relation between the current density ι and the total current I.

Let us consider the flux linking the rectangle $\Delta y \, \Delta z$ placed in a magnetic field where the coordinate axes X, Y, Z of Fig. 2·7 have been set up. Let one corner of this rectangle be at the point x, y, z. The flux ϕ linking the rectangle is

$$\phi_x = \int\int_s \bar{B}_x \, ds = \bar{B}_x \int\int_s ds = \bar{B}_x \, \Delta y \, \Delta z \qquad [2\cdot42]$$

where $\bar{B}_x$ is the average normal flux density over the area $\Delta y \, \Delta z$.

Let us now consider the line integral of the electric intensity acting

around the edges of the area *abcd*. Proceeding as in section 2·7, we find the average electric intensity to have the values shown in Fig. 2·7. The work done in carrying a unit charge around the rectangle is

$$\oint E \cos\theta \, dl =$$

$$\underbrace{\bar{E}_y \Delta y}_{a \text{ to } b} + \underbrace{\left[\bar{E}_z + \frac{\partial \bar{E}_z}{\partial y}\Delta y\right]\Delta z}_{b \text{ to } c} - \underbrace{\left[\bar{E}_y + \frac{\partial \bar{E}_y}{\partial z}\Delta z\right]\Delta y}_{c \text{ to } d} - \underbrace{\bar{E}_z \Delta z}_{d \text{ to } a} \quad [2\cdot43]^*$$

This work according to Faraday's law is equal to $-\partial\phi/\partial t$. Substituting for ϕ from equation 2·42, we obtain

$$\bar{E}_y \Delta_y + \bar{E}_z \Delta z + \frac{\partial \bar{E}_z}{\partial y}\Delta y \, \Delta z - \bar{E}_y \Delta y - \frac{\partial \bar{E}_y}{\partial z}\Delta y \, \Delta z - \bar{E}_z \Delta z =$$

$$- \frac{\partial}{\partial t}\bar{B}_x \Delta y \, \Delta_z \qquad [2\cdot44]$$

or

$$\left(\frac{\partial \bar{E}_z}{\partial y} - \frac{\partial \bar{E}_y}{\partial z}\right)\Delta y \, \Delta z = - \frac{\partial \bar{B}_x}{\partial t}\Delta y \, \Delta z \qquad [2\cdot45]$$

Dividing through by $\Delta y \, \Delta z$, and taking the limit as Δy and Δz approach zero, we have

$$\underset{\Delta y \, \Delta z \to 0}{\text{Limit}} \left\{\frac{\oint_l E \cos\theta \, dl}{\Delta y \, \Delta z}\right\} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = - \frac{\partial B_x}{\partial t} \qquad [2\cdot46]$$

By constructing similar mathematical areas $\Delta x \, \Delta z$ and $\Delta x \, \Delta y$, and deducing from geometry the work done, we obtain

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = - \frac{\partial B_y}{\partial t} \qquad [2\cdot47]$$

and

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = - \frac{\partial B_z}{\partial t} \qquad [2\cdot48]$$

These relations may be expressed more briefly in vector notation by

$$\nabla \times \mathbf{E} = \text{curl}\, \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \qquad [2\cdot49]$$

* As in the case discussed in section 2·7, E is always directed along the path of integration. From a to b and from b to c, $\theta = 0°$, so that $\cos\theta = 1$. Along the paths c to d and d to a, $\theta = 180°$, so that $\cos\theta = -1$, thus accounting for the choice of sign in this equation.

2·10 Maxwell's Displacement Current

One of Maxwell's outstanding contributions to electrical theory is his conception of the electrical displacement current and its magnetic effects. The term displacement current resulted from Maxwell's attempt to explain the phenomena which occur in a physical dielectric placed in an electric field. He conceived of separate positive and negative charges bound together in a perfectly elastic way by the molecular structure of the dielectric. When no electric field is applied these charges annul the effects of one another and no electric effect is produced by the dielectric itself. When an electric field is applied by an external source, the positive and negative charges are *displaced* in opposite directions and a state of polarization exists. In an isotropic medium the amount of this movement or displacement is proportional to the applied voltage or field.

Since this is an elastic process it is evident that no steady or continuous state will result. Maxwell, however, considered the transient or unsteady state. He showed that this motion of positive and negative charges in opposite directions constitutes a current which flows only while the electric field is changing. This concept is of the greatest importance in that it explains how magnetic fields may arise in regions where no physical conductor of electricity exists. In fact, the displacement is in no way different from the familiar conduction current except that it cannot exist as a direct current. Accordingly, it is a vector quantity, having direction and magnitude and being capable of resolution into components like any other vector.

If a battery is connected to a resistance, a current I flows in the circuit consisting of the battery, the wire, and the resistance. Such a current is called a conduction current, and the circuit is said to be closed. If a condenser is connected in series with this system, there is an accumulation of charge of opposite sign on the two plates of the condenser, no steady current flows, and the circuit is viewed as an open circuit. If, however, an alternating voltage is applied in place of the battery, an alternating current flows in the connecting wires, and charges of opposite sign appear alternately on the two plates of the condenser. Before Maxwell's time such a circuit was viewed as an open circuit, where the current existed only in the wire. Maxwell suggested the possibility of a current flowing in the dielectric between the condenser plates. Since the plates of the condenser contain at any instant an alternating electric charge, there must exist an alternating electric field in the dielectric between them. Maxwell postulated that any change in the electric intensity in a medium *is* an electric current.

Let us consider the circuit of Fig. 2·8, in which an alternating-current

generator has been connected. If C is the capacitance of the condenser and V the voltage across it, the instantaneous charge on the condenser is defined by $q = CV$. Now the current I flowing in the circuit is, by the definition of current, at every instant equal to

$$I = \frac{\partial q}{\partial t}$$

Substituting for q, we obtain

$$I = \frac{\partial q}{\partial t} = C\frac{\partial V}{\partial t} \qquad [2{\cdot}50]$$

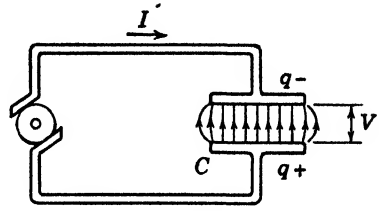


FIG 2 8

Now with a parallel-plate condenser as shown in Fig. 2·8 the capacitance is, in the practical system,

$$C = \varepsilon\frac{A}{d} \qquad [2{\cdot}51]$$

where d is the distance between the plates and A is the area of the plates. Substituting for C in equation 2·50 we have for the current density

$$\iota = \frac{I}{A} = \frac{\varepsilon}{d}\cdot\frac{\partial V}{\partial t} \qquad [2{\cdot}52]$$

Since the potential difference between the plates is the line integral of the field intensity E,

$$V = Ed \qquad [2{\cdot}53]$$

Substituting this value for V in 2·52, we have, in vector form,

$$\boldsymbol{\iota} = \varepsilon\frac{\partial \mathbf{E}}{\partial t} = \frac{\partial \mathbf{D}}{\partial t} \qquad [2{\cdot}54]$$

It is seen that $\partial\mathbf{D}/\partial t$ is proportional to a current. Maxwell suggested that we call the quantity $\partial\mathbf{D}/\partial t$ a displacement current. It is seen that this displacement current flows in the condenser whenever the current flows in the wire. Hence our total current is continuous. If, as often happens, both conduction current $\boldsymbol{\iota}_c$ and displacement current $\partial\mathbf{D}/\partial t$ take place in the same material, the two components of current may be added. Thus

$$\boldsymbol{\iota} = \boldsymbol{\iota}_c + \frac{\partial \mathbf{D}}{\partial t} \qquad [2{\cdot}55]$$

In the conductor, $\boldsymbol{\iota}_c$ is large in comparison with $\partial\mathbf{D}/\partial t$, and the current is largely conduction current. In the space between the condenser

plates, the conduction current density ι_c is small compared to the displacement current density $\partial \mathbf{D}/\partial t$. Any conduction current flowing between the condenser plates is called leakage current.

With these considerations in mind we may substitute for the current density ι, in the equations developed in section 2·7, the total current density of equation 2·55. This gives

$$\text{curl } \mathbf{H} = \left[\iota_c + \frac{\partial \mathbf{D}}{\partial t}\right] \qquad [2\cdot56]$$

It has been proved experimentally that the displacement current produces a magnetic field just as a conduction current does.

2·11 Maxwell's Equations

Collecting the general equations which we have obtained in the first two sections, we have:*

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \iota_{cx} + \frac{\partial D_x}{\partial t} \qquad (a)\\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \iota_{cy} + \frac{\partial D_y}{\partial t} \qquad (b)\\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \iota_{cz} + \frac{\partial D_z}{\partial t} \qquad (c) \end{aligned}\right\} \text{curl } \mathbf{H} = \iota_c + \frac{\partial \mathbf{D}}{\partial t} \qquad [2\cdot57]$$

* If magnetic poles can be isolated, then the form taken by Maxwell's equations is

$$\text{curl } \mathbf{H} = \iota_e + \frac{\partial \mathbf{D}}{\partial t}$$

$$-\text{curl } \mathbf{E} = \iota_m + \frac{\partial \mathbf{B}}{\partial t}$$

$$\text{div } \mathbf{D} = \rho_e$$

$$\text{div } \mathbf{B} = \rho_m$$

where ι_m, the magnetic current, is equal to $\sigma_m \mathbf{H}$, and σ_m is the magnetic conductivity. The volume density of magnetic charge, ρ_m, is analogous to the volume density of electric charge ρ_e. Under these conditions it is seen that Maxwell's equations are symmetrical. The practical significance of the change is questionable because no materials are known to contain an abundance of free magnetic charges. Hence in practical engineering problems we write

$$\iota_m = 0$$

$$\rho_m = 0$$

and the equations take the form given above.

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t} \end{aligned}\right\} \quad \text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \begin{matrix}(a)\\(b)\\(c)\end{matrix} \qquad [2\cdot58]$$

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \text{div } \mathbf{D} = \rho \qquad [2\cdot59]$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = \text{div } \mathbf{B} = 0 \qquad [2\cdot60]$$

$$\mathbf{D} = \varepsilon \mathbf{E} \qquad [2\cdot61]$$

$$\mathbf{B} = \mu \mathbf{H} \qquad [2\cdot62]$$

These equations, known as Maxwell's equations, are called point relations since with their aid the electric or magnetic field may be determined at any point when the currents and charges are specified.

To determine the forces acting on a charge or current other relations are required. The force acting on a unit charge at a point was defined in Chapter 1 as the electric intensity **E** at that point. The force on any charge is given by

$$\mathbf{F}_e = q\mathbf{E} \qquad [2\cdot63]$$

where **F** is in the direction of **E**.

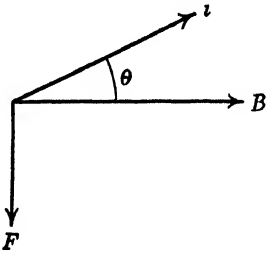


FIG. 2 9 Action of the magnetic field upon a current. Force $F_m = \iota B \sin \theta$ and is directed into the paper if F and B lie in the plane of the paper.

In the magnetic field, the force acting on a unit magnetic pole at any point is the magnetic intensity **H** at that point. From the definition of unit current in terms of the unit magnetic pole, the force acting on unit current at a point is equal to $\mu\mathbf{H}$, the magnetic induction **B**. This force, however, acts at right angles to the magnetic induction and to the current density as indicated in Fig. 2·9. That is,

$$F_m = \iota B \sin \theta \qquad [2\cdot64]$$

where θ is the angle between the direction of ι and the direction of **B**. The force $\mathbf{F}_m$ is perpendicular to the plane in which ι and **B** lie. This

is expressed in vector analysis by the vector or cross product

$$\mathbf{F}_m = \boldsymbol{\iota} \times \mathbf{B} \qquad [2\cdot65]$$

(pronounced "$\boldsymbol{\iota}$ cross $\mathbf{B}$"). When $\boldsymbol{\iota} = \rho\mathbf{v}$ this may be expressed as

$$\mathbf{F}_m = \rho(\mathbf{v} \times \mathbf{B}) \qquad [2\cdot66]$$

$\mathbf{F}_m$ is directed as indicated in the figure.

If practically all the current is constrained to flow in a wire conductor, the magnetic field in which the wire is placed acts on the current and, because of the constraint, on the wire. The force produced on the wire is, of course, the well-known motor action, and it is at right angles to the plane in which the wire and magnetic intensity vector lie.

If the current is not constrained to flow in the wire, but consists of a beam of electrons traveling through a magnetic field, the force acting on the electrons will cause them to deviate from their original direction.

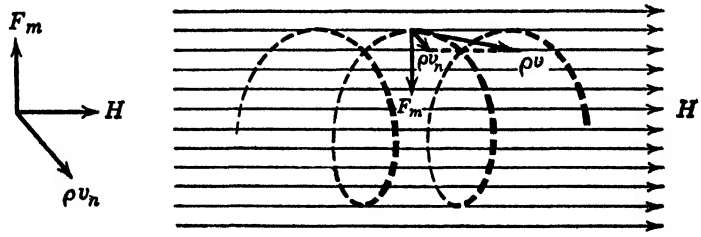


FIG. 2·10 Path followed by an electron when projected into a magnetic field with velocity $\mathbf{v}$. v_n is the component of velocity normal to the magnetic field.

This force, which is always at right angles to the direction of the field and to the instantaneous velocity, causes the electron to follow a curved path. If no electric field exists in the region, the electrons travel a helical path in passing through the magnetic field. This may be seen from a study of Fig. 2·10. Hence we may write for the force vector

$$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m \qquad [2\cdot67]$$

$$\mathbf{F} = q\,[\mathbf{E} + (\mathbf{v} \times \mathbf{B})] \qquad [2\cdot68]$$

2·12 Static State

In the static state, that is, in the electrostatic and magnetostatic cases discussed in Chapter 1, Maxwell's equations may be expressed more simply, since there is then no change with respect to time. They

are:

Electrostatic

$$\left.\begin{aligned}\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}&=0\\ \frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}&=0\\ \frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}&=0\end{aligned}\right\}\ \text{curl } \mathbf{E}=0 \qquad [2\cdot69]$$

$$\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}=\text{div } \mathbf{D}=\rho \qquad [2\cdot71]$$

where

$$\mathbf{D}=\varepsilon\mathbf{E} \qquad [2\cdot73]$$

Magnetostatic

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}&=0\\ \frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}&=0\\ \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}&=0\end{aligned}\right\}\ \text{curl } \mathbf{H}=0 \qquad [2\cdot70]$$

$$\frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}=\text{div } \mathbf{B}=0 \qquad [2\cdot72]$$

where

$$\mathbf{B}=\mu\mathbf{H} \qquad [2\cdot74]$$

As was previously stated, the systems of electrostatics and magnetostatics are completely independent of one another, and there is no current flow. The force between electric or magnetic charges is given by the relations:

$$\mathbf{F}_e=\mathbf{E}q \qquad [2\cdot75]$$

and

$$\mathbf{F}_m=\mathbf{H}m \qquad [2\cdot76]$$

2·13 Steady State

In the steady or stationary state a current exists, yet both the time derivatives of electric and magnetic induction are zero. We have, then, for the equations defining the steady state:

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}&=\iota_x\\ \frac{\partial H_x}{\partial_z}-\frac{\partial H_z}{\partial x}&=\iota_y\\ \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}&=\iota_z\end{aligned}\right\}\ \text{curl } \mathbf{H}=\boldsymbol{\iota} \qquad [2\cdot77]$$

$$\left.\begin{aligned}\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}&=0\\ \frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}&=0\\ \frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}&=0\end{aligned}\right\}\ \text{curl } \mathbf{E}=0 \qquad [2\cdot78]$$

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \text{div } \mathbf{D} = \rho \qquad [2\cdot79]$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = \text{div } \mathbf{B} = 0 \qquad [2\cdot80]$$

$$\iota = \sigma \mathbf{E} \qquad [2\cdot81]$$

Equation 2·81 expresses Ohm's law in terms of current density, electric field intensity, and specific conductivity. In this form it represents a point relation and is most suitable for use with Maxwell's equations.

2·14 Quasi-Steady State

In the non-steady state we must use the relations in their complete form. We may define, however, a state intermediate between the steady and non-steady state, which has been called the quasi-steady state. This state characterizes most of the problems of electrical engineering. The displacement current, except that in condensers, is taken as negligible in comparison with the conduction current. That is,

$$\iota_c \gg \frac{\partial \mathbf{D}}{\partial t} \qquad [2\cdot82]$$

and the field equations become

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \iota_{cx} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \iota_{cy} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \iota_{cz} \end{aligned}\right\} \text{curl } \mathbf{H} = \iota_c \qquad [2\cdot83]$$

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t} \end{aligned}\right\} \text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad [2\cdot84]$$

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \text{div } \mathbf{D} = \rho \qquad [2\cdot85]$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_y}{\partial z} = \text{div } \mathbf{B} = 0 \qquad [2\cdot86]$$

The equations as given above include the induction law, but the magnetic field both inside and outside of the conductor is calculated as though it were produced only by the conduction current. The results obtained when the displacement current is neglected are then correct only when it is negligible in comparison to the conduction current. This is true in most of the problems associated with low-frequency engineering. However, at high frequencies, the displacement current is often extremely important and we must use the complete Maxwellian equations.

2·15 Further Discussion of the Continuity Equations*

It may be shown that the continuity equation for electric current is contained in equation 2·57. Differentiating 2·57*a*, 2·57*b*, and 2·57*c* with respect to x, y, and z, respectively, we obtain:

$$\frac{\partial^2 H_z}{\partial x\,\partial y} - \frac{\partial^2 H_y}{\partial x\,\partial z} = \frac{\partial}{\partial x}\left(\iota_{cx} + \frac{\partial D_x}{\partial t}\right) \qquad [2\cdot87]$$

$$\frac{\partial^2 H_x}{\partial y\,\partial z} - \frac{\partial^2 H_z}{\partial y\,\partial x} = \frac{\partial}{\partial y}\left(\iota_{cy} + \frac{\partial D_y}{\partial t}\right) \qquad [2\cdot88]$$

$$\frac{\partial^2 H_y}{\partial z\,\partial x} - \frac{\partial^2 H_x}{\partial z\,\partial y} = \frac{\partial}{\partial z}\left(\iota_{cz} + \frac{\partial D_z}{\partial t}\right) \qquad [2\cdot89]$$

* The method of vector analysis applied to this manipulation is as follows:

Given: $\quad \mathbf{curl}\ \mathbf{H} = \left(\boldsymbol{\iota}_c + \frac{\partial \mathbf{D}}{\partial t}\right) = \boldsymbol{\iota}$

then $\quad \mathbf{div\ curl}\ \mathbf{H} = \mathbf{div}\ \boldsymbol{\iota}$

but $\quad \mathbf{div\ curl}$ of any vector is 0

therefore $\quad \mathbf{div}\ \boldsymbol{\iota} = 0$

Similarly, *given* $\quad \mathbf{curl}\ \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$

therefore $\quad \mathbf{div\ curl}\ \mathbf{E} = 0 = -\mathbf{div}\frac{\partial \mathbf{B}}{\partial t}$

and $\quad \mathbf{div}\frac{\partial \mathbf{B}}{\partial t} = 0$

or $\quad \mathbf{div}\ \mathbf{B} = 0$

Adding and rearranging

$$\left.\begin{aligned} &\frac{\partial^2 H_z}{\partial x\,\partial y} - \frac{\partial^2 H_z}{\partial y\,\partial x} \\ &+ \frac{\partial^2 H_x}{\partial y\,\partial z} - \frac{\partial^2 H_x}{\partial z\,\partial y} \\ &+ \frac{\partial^2 H_y}{\partial z\,\partial x} - \frac{\partial^2 H_y}{\partial x\,\partial z} \end{aligned}\right\} = \left\{\begin{aligned} &\frac{\partial}{\partial x}\left(\iota_{cx} + \frac{\partial D_x}{\partial t}\right) \\ &+ \frac{\partial}{\partial y}\left(\iota_{cy} + \frac{\partial D_y}{\partial t}\right) \\ &+ \frac{\partial}{\partial z}\left(\iota_{cz} + \frac{\partial D_z}{\partial t}\right) \end{aligned}\right. \qquad [2\cdot90]$$

The left member of equation 2·90 is zero if

$$\frac{\partial^2 H_z}{\partial x\,\partial y} = \frac{\partial^2 H_z}{\partial y\,\partial x}; \qquad \frac{\partial^2 H_y}{\partial x\,\partial z} = \frac{\partial^2 H_y}{\partial z\,\partial x}; \qquad \frac{\partial^2 H_x}{\partial y\,\partial z} = \frac{\partial^2 H_x}{\partial z\,\partial y}$$

These relations are equal if the order of differentiation may be changed, and this is true, with few exceptions,* if all the second-order derivatives exist.

Thus

$$\frac{\partial}{\partial x}\left(\iota_{cx} + \frac{\partial D_x}{\partial t}\right) + \frac{\partial}{\partial y}\left(\iota_{cy} + \frac{\partial D_y}{\partial t}\right) + \frac{\partial}{\partial z}\left(\iota_{cz} + \frac{\partial D_z}{\partial t}\right) = 0$$

But

$$\iota_x = \iota_{cx} + \frac{\partial D_x}{\partial t}$$

$$\iota_y = \iota_{cy} + \frac{\partial D_y}{\partial t}$$

$$\iota_z = \iota_{cz} + \frac{\partial D_z}{\partial t}$$

Hence

$$\frac{\partial \iota_x}{\partial x} + \frac{\partial \iota_y}{\partial y} + \frac{\partial \iota_z}{\partial z} = 0$$

Similarly from equation 2·58 we can show that

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$$

This equation establishes the continuity of the magnetic induction, or flux density, as it is often called. Further discussion of this point was presented in a previous section.

2·16 Continuity of the Tangential Components of E and H

In the solution of problems involving surfaces of discontinuity, it is

*See Burington and Torrance, *Higher Mathematics*, page 95, Section 17. McGraw-Hill Book Company.

necessary to prove certain relations in regard to the boundary conditions that must hold in passing from one medium into another. These boundary conditions may be established by means of deductions from Ampère's and Faraday's laws.

Let us consider, for example, the surface of discontinuity between two media shown in Fig. 2·11. This surface may represent a discontinuity in dielectric constant, permeability, or conductivity. Let us construct the small rectangle *abcd* shown in Fig. 2·11 so that one side of it is in medium 1 and the other side is in medium 2. Let the length of the rectangle be l and the width w, as indicated in the figure. Along the four sides of the rectangle let the average values of the electric intensity be $\bar{E}_{1T}$ acting in medium 1, $\bar{E}_{2T}$ acting in medium 2, $\bar{E}_3 = \bar{E}_{1N} + \bar{E}_{2N}$ and $\bar{E}_4 = \bar{E}_{1N'} + \bar{E}_{2N'}$ acting in both media 1 and 2. Also let $\bar{B}_n$ be the average value of the normal component of magnetic flux density over the area of the rectangle. From Faraday's law we have

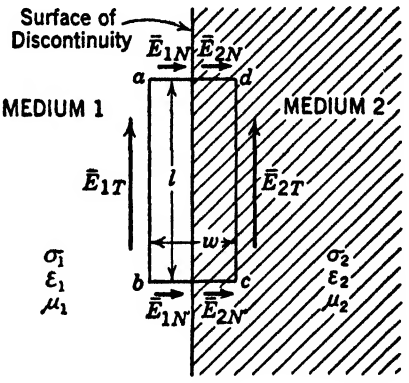


FIG. 2·11

$$\oint_l E \cos\theta \, dl = -\frac{\partial}{\partial t}\int\int B_n \, ds$$

Taking the path of integration around the rectangle

$$\oint_l E \cos\theta \, dl = -\underset{a \text{ to } b}{\bar{E}_{1T} l} + \underset{b \text{ to } c}{\bar{E}_4 w} + \underset{c \text{ to } d}{\bar{E}_{2T} l} - \underset{d \text{ to } a}{\bar{E}_3 w} \qquad [2\cdot91]$$

where $\theta = 0$ along path from b to c and c to d so that $\cos\theta = 1$, and $\theta \doteq 180°$ along path from a to b and d to a so that $\cos\theta = -1$. Also

$$\int\int_s \bar{B}_n \, ds = \bar{B}_n \int\int_s ds = \bar{B}_n wl \qquad [2\cdot92]$$

where wl is the surface area of the rectangle. Thus

$$\frac{\partial}{\partial t}\bar{B}_n wl = -\bar{E}_{1T} l + \bar{E}_4 w + \bar{E}_{2T} l - \bar{E}_3 w \qquad [2\cdot93]$$

Let us divide through by l and take the limit as w approaches zero. This gives

$$\underset{w \to 0}{\text{Limit}}\left[\frac{\partial}{\partial t}\bar{B}_n w\right] = \underset{w \to 0}{\text{Limit}}\left[\bar{E}_{2T} - \bar{E}_{1T} + \frac{w}{l}(\bar{E}_4 - \bar{E}_3)\right] \qquad [2\cdot94]$$

In the limit we have

$$0 = E_{2T} - E_{1T}$$

or

$$E_{1T} = E_{2T} \qquad [2\cdot95]$$

at the surface of discontinuity. Thus the tangential component of the electric intensity is everywhere continuous.

By use of the magnetomotive force equation, i.e.,

$$\oint_l H \cos\theta \, dl = \int\int_c \iota_n \, ds \qquad [2\cdot96]$$

we can show in a similar way the continuity of the tangential component of the magnetic intensity when the current density is everywhere continuous. Let us replace the electric intensity vectors around the area *abcd* by the magnetic intensity vectors $\overline{H}_{1T}$, $\overline{H}_{2T}$, $\overline{H}_3$, and $\overline{H}_4$. Let $\bar{\iota}_n$ be the total average current density normal to the rectangle *abcd*. Then

$$\int\int_s \bar{\iota}_n \, ds = \bar{\iota}_n \int\int_s ds = \bar{\iota}_n \, wl \qquad [2\cdot97]$$

Also

$$\oint_l H \cos\theta \, dl = -\overline{H}_{1T} l + \overline{H}_4 w + \overline{H}_{2T} l - \overline{H}_3 w \qquad [2\cdot98]$$

Equating 2·98 and 2·97, dividing through by l, and taking the limit as w approaches zero, we have

$$\underset{w\to 0}{\text{Limit}} \{\bar{\iota}_n w\} = \underset{w\to 0}{\text{Limit}} \left\{ -\overline{H}_{1T} + \overline{H}_{2T} + (\overline{H}_4 - \overline{H}_3)\frac{w}{l} \right\} \qquad [2\cdot99]$$

In the limit

$$0 = -H_{1T} + H_{2T}$$

or

$$H_{1T} = H_{2T} \qquad [2\cdot100]$$

at the surface of discontinuity. Thus the tangential component of the magnetic intensity is everywhere continuous.

2·17 Continuity of the Normal Components of D and B

Two additional relations, useful in the solution of problems in refraction and reflection, may be derived with the aid of section 2·15. The first of these is based upon the assumption that free magnetic poles do not exist, and upon the general equation 1·51, which states that

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = \text{div}\,\mathbf{B} = 0 \qquad [1\cdot51]$$

Because lines of magnetic flux may be considered continuous it is necessary that as many leave one side of any given area of the bounding surface as arrive at the opposite side. This statement defines only the normal components of the magnetic induction and permits us to write

$$B_{1N} = B_{2N} \qquad [2.101]$$

In section 2·15 it was shown that the total current is continuous. In any medium where the conduction current is negligible the displacement current is therefore continuous also. This statement is equivalent to the equation

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \text{div } \mathbf{D} = 0$$

which is valid in any region where there are no free electric charges. By means of an argument equivalent to that above it is shown that the normal component of the electric induction is also continuous. That is,

$$D_{1N} = D_{2N} \qquad [2·102]$$

2·18 Units and Dimensions

Probably no science has inherited a more bewildering array of units than electromagnetism. This oversupply is evidence of the extremely wide application and great basic importance of the phenomena. The subjects of electrostatics and magnetostatics were built up independently and each had a workable, self-consistent system of units. Then Oersted's discovery linked the two fields, and Maxwell's researches completed the union. It was found that any quantity in electricity or magnetism could be expressed numerically in terms of either electrostatic or magnetostatic definitions, and this resulted in a confusing and unfortunate situation. Moreover, it became apparent that a basic problem of dimensions exists. In either system alone it is possible to assume some parameter dimensionless, and dimensions are readily developed for all other quantities of the system. It was customary to assume the dielectric constant dimensionless in electrostatics and the permeability dimensionless in magnetostatics. Maxwell's equations permit the ready deduction of dimensions of electric quantities in the magnetic system, and vice versa. However, ε and μ cannot both be dimensionless.

The situation was still further beclouded by the commercial application of electricity. The men who built our early generators and motors were practical engineers, not physicists, and they demanded a convenient set of units for their work. The so-called practical units, a com-

promise between the two previous systems, arose from this demand. It works out rather well for problems in electric circuits of resistance and inductance, but the unit of capacitance in this system is entirely too large to be practical.

We may evaluate the electric quantities with the aid of dimensional analysis in the following manner. It is convenient to take the magnetostatic force equation as our starting point, where the force is given by the relation

$$F = \frac{m_1 m_2}{\mu r^2} = \frac{m^2}{\mu r^2} \quad \text{for} \quad m_1 = m_2 \qquad [2\cdot103]$$

We can define $m_1 = m_2 = 1$ when $F = 1$ and $r = 1$ on the assumption that μ is unity in vacuum. But force may be expressed in terms of the fundamental units of mass $[M]$, length $[L]$, and time $[T]$ thus:

$$F = [MA] = [MLT^{-2}] \qquad [2\cdot104]$$

Replacing F in 2·103 as given in 2·104, and solving for m, we obtain

$$[m] = [\sqrt{ML^3T^{-2}\mu}] = [M^{1/2}L^{3/2}T^{-1}\mu^{1/2}] \qquad [2\cdot105]$$

Using the equation $F = mH$ or $H = F/m$, we have the dimensions of magnetic field intensity

$$[H] = [M^{1/2}L^{-1/2}T^{-1}\mu^{-1/2}] \qquad [2\cdot106]$$

From the definition of current we deduce its dimensions

$$I = \int_l H \cos\theta \, ds = H[L] \qquad [2\cdot107]$$

$$[I] = [M^{1/2}L^{1/2}T^{-1}\mu^{-1/2}] \qquad [2\cdot108]$$

But electric current is the time rate of transfer of charge. $I = q/t$. Accordingly the dimensions of charge in the emu system are

$$[q] = [M^{1/2}L^{1/2}\mu^{-1/2}] \qquad [2\cdot109]$$

In order to compare the systems we now deduce the dimensions of electric charge in the cgs electrostatic system. In the electrostatic system, the definitive equation is

$$F = \frac{q_1 q_2}{\varepsilon r^2} \qquad [2\cdot110]$$

The dimensions of q which follow from this equation are

$$[q] = [M^{1/2}L^{3/2}T^{-1}\varepsilon^{1/2}] \qquad [2\cdot111]$$

Evidently electric charge is a fundamental entity and must have the

same dimension regardless of any definition used. Accordingly we set

$$[M^{1/2}L^{1/2}\mu^{-1/2}] = [M^{1/2}L^{3/2}T^{-1}\varepsilon^{1/2}] \qquad [2\cdot112]$$

or

$$[\varepsilon^{-1/2}\mu^{-1/2}] = [LT^{-1}] \qquad [2\cdot113]$$

We may not ascertain the dimensions of ε or of μ separately from this process but we do know that the quantity

$$\frac{1}{\sqrt{\varepsilon\mu}} = \frac{L}{T} \qquad [2\cdot114]$$

has the dimensions of velocity. It might be thought that suitable manipulations of equations would yield another relation permitting explicit solution of the dimensions of μ and ε. Unfortunately, this is not true. We are free to define the dimension of one of the two arbitrarily if the above relation is adhered to.

One attractive possibility is to set

$$[\varepsilon] = [M^{-1}L^{-1}T^2] \quad \text{and} \quad [\mu] = [ML^{-1}] \qquad [2\cdot115]$$

From this assumption

$$\text{Charge } [q] = [L] \quad \text{(length)} \qquad [2\cdot116]$$

$$\text{Current } [I] = [LT^{-1}] \quad \text{(velocity)} \qquad [2\cdot117]$$

$$\text{Magnetic field } [H] = [T^{-1}] \quad \text{(angular velocity)} \qquad [2\cdot118]$$

$$\text{Inductance } [L] = [M] \quad \text{(mass)} \qquad [2\cdot119]$$

$$\text{Capacitance } [C] = [M^{-1}T^2] \quad \text{(compliance of a spring)} \qquad [2\cdot120]$$

$$\text{Potential } [V] = [MLT^{-2}] \quad \text{(mechanical force)} \qquad [2\cdot121]$$

It is seen that the familiar properties of inductance, capacitance, and potential difference have exactly the same dimensions as the analogous mechanical properties. To those who employ physical reasoning this is a very real advantage. The energy stored in an inductance, $\frac{1}{2}LI^2$, has the dimensions $[ML^2T^{-2}]$ of kinetic energy. Similarly the energy stored in a condenser, $\frac{1}{2}CV^2$, has the same dimensions $[ML^2T^{-2}]$.

Another possible definition, and one which has found considerable favor, hinges upon the fact that the ohm is the most permanent, reproducible, and transportable of the electric parameters. It has therefore been suggested that the ohm be used as the basic electrical dimension $[O]$. The dimensions of ε then become

$$[\varepsilon] = [L^{-1}\,TO^{-1}]$$

and the dimensions of μ are

$$[\mu] = [L^{-1}\ TO]$$

Were the ohm dimensionless, these equations would reduce to the unreasonable assumption that ε and μ have the same dimensions. The advantages of the ohm as a basic unit of reference are not to be ignored, but it is not necessary to carry through the dimensional implications in order to obtain this practical advantage.

The practice now becoming standard is to use the electric charge as the basic electrical dimension. Dimensional relations that follow are:

Quantity	Symbol	Dimensions	Practical Units
Length	L	L	Meter
Mass	M	M	Kilogram
Time	T	T	Second
Force	F	MLT^{-2}	Newton
Energy	W	ML^2T^{-2}	Joule
Power	P	ML^2T^{-3}	Watt
Charge	q	Q	Coulomb
Current	I	$T^{-1}Q$	Ampere
Resistance	R	$ML^2T^{-1}Q^2$	Ohm
Electrical potential	V	$ML^2T^{-2}Q^{-1}$	Volt
Electric field intensity	E	$MLT^{-2}Q^{-1}$	Volt/meter
Electric displacement	D	$L^{-2}Q$	Coulomb/square meter
Dielectric constant	ε	$M^{-1}L^{-3}T^2Q^2$	Farad/meter
Capacitance	C	$M^{-1}L^{-2}T^2Q^2$	Farad
Magnetomotive force	MMF	$T^{-1}Q$	Ampere-turn
Magnetic field intensity	H	$L^{-1}T^{-1}Q$	Ampere-turn/meter
Magnetic induction	B	$MT^{-1}Q^{-1}$	Weber/square meter
Permeability	μ	MLQ^{-2}	Henry/meter
Inductance	L	ML^2Q^{-2}	Henry

It will be observed that no fractional exponents appear in the foregoing table and that the units are the familiar practical ones. There is, however, a very important basic principle that is not apparent. The mks definitions now in use reduce the practical system from an inconsistent mass of definitions in terms of standard resistors, standard cells, etc., to a consistent reproducible absolute system. That is, any electrical property may be expressed in terms of the four basic quantities, the meter, the kilogram, the second, and the coulomb. Since the coulomb may be established quantitatively from the other three units it follows that the system is absolute and that physical relations may be stated precisely by its use.

The question of rationalization does not appear in the foregoing dimensional analysis since it involves only a numerical constant. The

question involved is, fundamentally, in just what equations the factor 4π shall appear. In the esu and emu systems the force and potential equations are free from the 4π term, while the term does appear in the flux, mmf, and field equations. Since we are dealing here primarily with the latter equations, we shall find it advantageous to use the rationalized system.

Heaviside was one of the first to advocate a rationalized system. The Heaviside-Lorentz system, which he helped to formulate, decreased the unit charge of the electrostatic and magnetostatic systems by the factor $\sqrt{4\pi}$. This leads to a set of Maxwell's equations that are free from the coefficient 4π. Certain of the intervening equations, however, involve fractional powers of this multiplier, a most undesirable feature.

The rationalized practical system as here used is due to Giorgi. The charge in the electric force equation is unchanged, while the unit pole of the magnetic force equation is modified by the factor 4π. Thus the unit electric charge is 1 coulomb, and the electric current is 1 coulomb per second, or 1 ampere. With a unit pole of the unrationalized system, the mmf (work per unit pole) in a closed contour enclosing unit current involves the factor 4π. The pole of the rationalized system is $1/4\pi$ of the other pole, and therefore cancels the factor already introduced. In addition, the unit of electrostatic flux is associated with a unit charge. This leads to a divergence equation and a displacement current that are free from the factor 4π.

PROBLEMS

2·1 A circular copper wire has an area of 0.005 cm^2 and conducts a current of 2 amperes. How many electrons pass a given point in this wire per second? Assuming that there are 16.8×10^{21} free electrons per cubic centimeter in copper, compute the average velocity of the electrons. An electron has a charge of 1.600×10^{-19} coulomb.

2·2 It is known that the force exerted on a conductor at right angles to a system of lines of magnetic induction is BlI, where B is the induction, l is the length of the conductor, and I is the total current. Consider a stream of electrons having a charge e and a velocity v. By conversion of the above equation show that the force on each electron is Bev.

2·3 In the typical antenna a current is caused to flow at the center of a relatively long straight conductor by inserting a generator there. Explain how the continuity of the total current applies in this case.

2·4 Ordinary telegraphy is accomplished by keying or switching a battery circuit. Identify this operation in terms of static, steady, or non-steady state. Repeat for telephony using a carbon microphone. Repeat for a vacuum-tube oscillator system. Repeat for the electric lens system of a cathode-ray tube.

2·5 Calculate the current which will flow in a circular aluminum rod of radius 0.2 cm if the electric intensity is 0.001 volt per meter.

2·6 A current of 2 amperes flows in a circular wire circuit having a radius of 10 cm located in free space. Calculate the magnetic moment per unit area of a magnetic shell which will replace this current circuit.

2·7 An annular torus of soft iron has a major radius of 5 cm and a section radius of 0.5 cm. It is wrapped with 100 turns of wire uniformly distributed, carrying a current of 1 ampere. Calculate the average magnetic intensity, the total MMF, the induction if $\mu_m = 5000$, and the total magnetic flux.

2·8 Discuss the analogy between the electric circuit composed of batteries and resistances and the magnetic circuit composed of magnetic shells (or windings) and reluctance. In particular, mention numerical differences and certain other basic differences.

2·9 Discuss in detail why the permeability of the medium must be considered in writing the relation between the magnetic shell and the current in an electric circuit.

2·10 A straight wire 20 cm long is moved at right angles through a magnetic field having a value of induction $B = 10^6$ webers per square meter with a velocity of 5 meters per second. Calculate the voltage induced in this wire. Interpret in terms of the line integral of the electric intensity.

2·11 In one familiar form of the unipolar generator a disk of a conducting material is revolved between the poles of a permanent magnet. The electrical contacts are applied at the center of the disk and at the periphery. Consider such a device in which the magnet is greatly expanded and produces a uniform magnetic field everywhere normal to the disk. Why do not the flux lines rotate with the disk? If they did rotate with the disc what would happen?

2·12 A circular slab of dielectric is 3 cm in radius and 1 cm thick. It forms the dielectric of a parallel-plate condenser which is subjected to an rms voltage of 100 volts at a frequency of 10^8 cycles per second. Calculate the MMF (rms) due to the displacement current in this slab.

CHAPTER 3

MAXWELL'S EQUATIONS

3·1 Maxwell's Field Equations in Integral Form

The fundamental field equations of electromagnetic theory as presented and developed in the first two chapters may be summarized as follows:*

From Gauss

Classical Form	Vector Form†	
$\int\int_s D_n\, ds = \int\int\int_\tau \rho\, d\tau$	$\int\int_s \mathbf{D}\cdot\mathbf{ds} = \int\int\int_\tau \rho\, d\tau$	[3·1]
$\int\int_s B_n\, ds = 0$	$\int\int_s \mathbf{B}\cdot\mathbf{ds} = 0$	[3·2]

From Ampère and Maxwell

$$\oint_l H\cos\theta\, dl = \int\int_s \left\{\iota_c + \frac{\partial D}{\partial t}\right\}_n ds \qquad \oint \mathbf{H}\cdot\mathbf{dl} = \int\int_s \left(\iota_c + \frac{\partial \mathbf{D}}{\partial t}\right)\cdot\mathbf{ds} \quad [3\cdot3]$$

From Faraday

$$\oint_l E\cos\theta\, dl = -\frac{\partial}{\partial t}\int\int_s B_n\, ds \qquad \oint_l \mathbf{E}\cdot\mathbf{dl} = -\int\int \frac{\partial \mathbf{B}}{\partial t}\cdot\mathbf{ds} \quad [3\cdot4]$$

where $$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = \epsilon\mathbf{E} \quad [3\cdot5]$$

and $$\mathbf{B} = \mu_0\mathbf{H} + \mathbf{M} = \mu\mathbf{H} \quad [3\cdot6]$$

Equations 3·1 and 3·2 state that the surface integral of the total flux through any closed surface S is equal to the volume integral of the charge density taken throughout the volume enclosed by the surface. The element of the surface is ds, and $d\tau$ is an element of volume. The subscript n refers to the component of the flux vector normal to the surface element. In the electric field, the charge density is ρ, whereas in the magnetic field, we assume that there is no residual charge density of magnetization.

* The form in which these equations are given here is used extensively. These equations can, however, be expressed in many other ways.

† The vector ds has the direction of an outward normal to the surface and a magnitude equal to the increment of surface area.

Equation 3·3 states that the line integral of the component of magnetic intensity along the path element *dl* is equal to the surface integral of the component of total current density normal to the surface element *ds* of the surface bounded by the path along which the line integral is taken. The angle θ is the angle between the direction of **H** and the direction of the path element **dl** at any point along the path. The displacement current is $\partial \mathbf{D}/\partial t$.

Equation 3·4 states that the line integral of the component of electric intensity along the path element *dl* is equal to the negative time derivative of the surface integral of the component of the magnetic flux normal to the surface element *ds* of the surface bounded by the path along which the line integral is taken. The angle θ is the angle between the direction of **E** and the direction of the path element **dl** at any point along the path.

It is understood that the relations $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$, where ε and μ are constant, hold only in isotropic dielectrics and in isotropic paramagnetic and diamagnetic materials.

In order to determine the force acting on a charge or current in the presence of electric or magnetic fields, or both, we need in addition to these equations the relation

$$\mathbf{F} = q(\mathbf{E} + \boldsymbol{v} \times \mathbf{B}) \qquad [3\cdot7]$$

The vector cross product was explained in a previous chapter. Also the relation between the current density and the electric intensity **E** (Ohm's law) is

$$\boldsymbol{\iota} = \sigma\mathbf{E} \qquad [3\cdot8]$$

where σ is the conductivity of the medium. The right-hand side of 3·3 is often written in terms of the electric intensity. To do this replace $\boldsymbol{\iota}$ and **D** in terms of **E** according to equations 3·5 and 3·8. We then obtain

$$\oint_l H \cos\theta \, dl = \int\int_s \left(\sigma E + \varepsilon \frac{\partial E}{\partial t}\right)_n ds \qquad [3\cdot9]$$

3·2 Maxwell's Field Equations in Differential Form

The field equations as developed in differential form may be summarized as follows:

From Gauss

Classical Form	Vector Form	
$\dfrac{\partial D_x}{\partial x} + \dfrac{\partial D_y}{\partial y} + \dfrac{\partial D_z}{\partial z} = \rho$	$\operatorname{div} \mathbf{D} = \rho$	[3·10]
$\dfrac{\partial B_x}{\partial x} + \dfrac{\partial B_y}{\partial y} + \dfrac{\partial B_z}{\partial z} = 0$	$\operatorname{div} \mathbf{B} = 0$	[3·11]

From Ampère and Maxwell

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \iota_{cx} + \frac{\partial D_x}{\partial t} \qquad \operatorname{curl}_x \mathrm{H} = \iota_{cx} + \frac{\partial D_x}{\partial t} \qquad (a) \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \iota_{cy} + \frac{\partial D_y}{\partial t} \qquad \operatorname{curl}_y \mathrm{H} = \iota_{cy} + \frac{\partial D_y}{\partial t} \qquad (b) \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \iota_{cz} + \frac{\partial D_z}{\partial t} \qquad \operatorname{curl}_z \mathrm{H} = \iota_{cz} + \frac{\partial D_z}{\partial t} \qquad (c) \end{aligned}\right\} \quad [3\cdot12]$$

From Faraday

$$\left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\frac{\partial B_x}{\partial t} \qquad \operatorname{curl}_x \mathrm{E} = -\frac{\partial B_x}{\partial t} \qquad (a) \\ \frac{\partial E_x}{z} - \frac{\partial E_z}{\partial x} &= -\frac{\partial B_y}{\partial t} \qquad \operatorname{curl}_y \mathrm{E} = -\frac{\partial B_y}{\partial t} \qquad (b) \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\frac{\partial B_z}{\partial t} \qquad \operatorname{curl}_z \mathrm{E} = -\frac{\partial B_z}{\partial t} \qquad (c) \end{aligned}\right\} \quad [3\cdot13]$$

Equations 3·12 and 3·13 may be written in the abbreviated vector form:

$$\operatorname{curl} \mathbf{H} = \iota_c + \frac{\partial \mathbf{D}}{\partial t} \qquad [3\cdot14]$$

$$\operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad [3\cdot15]$$

The above equations refer to the most general medium with which we shall deal. This medium is restricted in the following sense:

1. It is homogeneous in that ε, μ, and σ, are constant over the region in question.

2. It is isotropic in that these parameters have the same properties in all directions.

The degree to which these requirements are met by different media is quite varied. Some media, such as air or free space, may be considered to meet them rather exactly. Other materials, such as wood or earth, are extremely inhomogeneous and anisotropic. A large class of materials including rubber, glass, and metals meet these requirements to an intermediate degree. Average parameters can sometimes be selected when applying the field equations in media which deviate from these requirements. The field equations then give approximately correct results depending on the degree of approximation. In many systems, it often happens that the conducting properties are negligible in comparison to the magnetic and dielectric properties, or

vice versa. In such cases the field equations are greatly simplified. The simplest of all media is free space or vacuum.

Maxwell's equations include practically all the fundamental laws and relations of electrostatics, magnetostatics, and electromagnetism except those relating to mechanical forces and to acceleration.

3·3 Plane Waves and the Wave Front

We are all familiar with the water waves set up when we drop a pebble into a still pond. The surface waves set up in this way travel outward from the disturbance and form clearly defined concentric rings. We may observe that if we drop in a number of stones at the same time the disturbance on the surface is quite complicated in the region where the stones hit the water. At some distance away, however, the waves

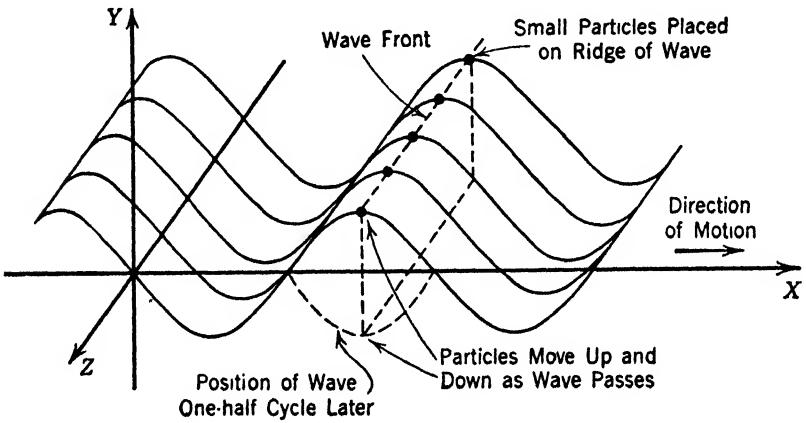


FIG. 3·1 Illustration of wave motion.

appear to be clearly defined ridges traveling with a definite velocity. At a distant point from the disturbance, let us observe closely the movement of a series of small particles all of which lie on one of the ridges as shown in Fig. 3·1. These particles will rise and fall together as the waves pass. If we imagine a line joining them, this line is parallel to the ridges and is called the " wave front." All along the wave front the phase of the wave is constant at any instant. If the pond is deep, and the water waves under consideration have small amplitude, the height of the wave will also be constant at all points of constant phase. Electric waves likewise resolve themselves at great distances from the source into clearly defined layers, thin sections of which change in intensity as the wave progresses. Like water waves, electric waves progress at right angles to the wave front.

When water waves have small amplitude, their motion may be defined by the use of two vectors. One vector represents the direction and velocity of propagation of the wave; the other, the direction and amplitude of the wave front from some arbitrary reference level. This level may conveniently be chosen as the height of the water when no wave is present. To represent an electromagnetic wave, three vectors are required. One of them represents the direction and velocity of propagation; the other two, the amplitude and direction of the electric and magnetic field intensities. If the medium is not isotropic, the wave propagation may not be at right angles to the front. A wave which has a plane front is called a plane wave; one having a spherical front, a spherical wave, etc. Thus the words plane, spherical, etc., refer to the shape of the surfaces throughout which the phase of the wave is constant at any given instant. The amplitude of the wave may or may not be constant over this equiphase surface. With plane waves these equiphase surfaces form a system of parallel planes. With spherical waves, they form a system of concentric spheres. The term "ray" is used to define a straight line which is everywhere normal to the equiphase surfaces and hence is in the direction in which the phase of the wave changes most rapidly at any given instant. The rays of a plane wave are straight lines, parallel to one another; those of a spherical wave are straight lines radiating from the center of the wave system.

3·4 Longitudinal and Transverse Waves

Water waves of small amplitude are essentially transverse waves * in that the variation in amplitude of the waves (the vibration) takes place at right angles to the direction of propagation. In the case of sound waves the displacement in the medium (the vibration) takes place along the direction of propagation, and they are called longitudinal waves. Thus, with transverse waves the vibration takes place at right angles to the rays, whereas in longitudinal waves the vibration takes place in the direction of the rays. In the electromagnetic wave the situation is complicated by the fact that both electric and magnetic vectors must be considered. Electromagnetic waves are, in general, neither longitudinal nor transverse in the sense that neither the electric vector **E** nor the magnetic vector **H** is completely longitudinal or completely transverse.

In some types of electromagnetic waves, the electric vector **E** is at

* In general, water waves are both longitudinal and transverse. Where the amplitude of the wave is small, however, they approximate very closely transverse waves, and their cross section is nearly sinusoidal.

right angles to the ray while the magnetic vector **H** is not. Waves of this type are called transverse electric waves or *TE* waves. Since the **H** vector is not normal to the ray it must have a component along the ray. In this sense such a wave is sometimes called longitudinal magnetic or an H wave. In other types of waves the magnetic vector **H** is at right angles to the ray while the electric vector **E** is not. Waves of this type are called transverse magnetic or *TM* waves. In this case, **E** must have a component in the direction of the ray, and for this reason the wave is sometimes called longitudinal electric or an *E* wave. If both **E** and **H** vectors are at right angles to the ray, the wave is called transverse electromagnetic or a *TEM* wave. No electromagnetic field exists in which both **E** and **H** are wholly in the direction of the ray.

3·5 Propagation of Plane Electromagnetic Waves in Free Space

One of the simplest examples of wave motion, which we shall deduce from Maxwell's equations, is a plane wave in free space. With such a wave, the electric and magnetic intensities are of constant value over any plane perpendicular to the direction of propagation. Such a plane is thus a surface of equal phase. Let this plane, which is called the plane of the wave, be the *YOZ* plane. Then the electric and magnetic intensities have constant values for all values of y and z. Hence, the derivatives of the **E** and **H** vectors with respect to y and z are zero.

From this we know immediately that equations 3·12 and 3·13 reduce to:

$$\left.\begin{aligned} 0 &= \varepsilon \frac{\partial E_x}{\partial t} && (a)\\ -\frac{\partial H_z}{\partial x} &= \varepsilon \frac{\partial E_y}{\partial t} && (b)\\ \frac{\partial H_y}{\partial x} &= \varepsilon \frac{\partial E_z}{\partial t} && (c) \end{aligned}\right\} \quad [3\cdot16] \qquad \left.\begin{aligned} 0 &= \mu \frac{\partial H_x}{\partial t} && (a)\\ \frac{\partial E_z}{\partial x} &= \mu \frac{\partial H_y}{\partial t} && (b)\\ \frac{\partial E_y}{\partial x} &= -\mu \frac{\partial H_z}{\partial t} && (c) \end{aligned}\right\} \quad [3\cdot17]$$

where $\iota_c = 0$ and **D** and **B** have been replaced by $\varepsilon\mathbf{E}$ and $\mu\mathbf{H}$, respectively.

From equation 3·16*a* we know that $E_x = 0$ or a constant over all space. Since constant values do not enter into wave propagation we take $E_x = 0$. Similarly, from 3·17*a*, $H_x = 0$. Hence it is evident that the vectors **E** and **H** lie entirely in the *YOZ* plane.

The choice of the direction of either the electric intensity or the magnetic intensity in the *YOZ* plane is wholly arbitrary. Let us choose the direction of the electric intensity **E** parallel to the *OY* axis, and let us further require that **E** remain in this direction for all values

of distance and time. Under these conditions $E_z = 0$ everywhere.* Hence from equations 3·16c and 3·17b

$$\frac{\partial H_y}{\partial x} = 0 \quad \text{and} \quad \mu \frac{\partial H_y}{\partial t} = 0 \qquad [3\cdot18]$$

and we see that H_y is also zero. This leaves E_y and H_z as the only remaining components of the field, and hence the **E** and **H** vectors are at right angles to each other. Equations 3·16 and 3·17 reduce to

$$-\frac{\partial H_z}{\partial x} = \varepsilon \frac{\partial E_y}{\partial t} \qquad [3\cdot19]$$

and

$$-\frac{\partial E_y}{\partial x} = \mu \frac{\partial H_z}{\partial t} \qquad [3\cdot20]$$

We can obtain from these relations an equation in E_y alone or in H_z alone by the following manipulation. Differentiate 3·19 with respect to x and 3·20 with respect to t, obtaining

$$\frac{\partial^2 H_z}{\partial x^2} + \varepsilon \frac{\partial^2 E_y}{\partial x\, \partial t} = 0 \qquad [3\cdot21]$$

and

$$\frac{\partial^2 E_y}{\partial t\, \partial x} + \mu \frac{\partial^2 H_z}{\partial t^2} = 0 \qquad [3\cdot22]$$

Subtracting 3·22 multiplied by ε from 3·21, we have

$$\varepsilon \frac{\partial^2 E_y}{\partial x\, \partial t} - \varepsilon \frac{\partial^2 E_y}{\partial t\, \partial x} - \mu\varepsilon \frac{\partial^2 H_z}{\partial t^2} + \frac{\partial^2 H_z}{\partial x^2} = 0 \qquad [3\cdot23]$$

If

$$\frac{\partial^2 E_y}{\partial x\, \partial t} = \frac{\partial^2 E_y}{\partial t\, \partial x} \dagger \qquad [3\cdot24]$$

equation 3·23 becomes

$$\frac{\partial^2 H_z}{\partial x^2} = \mu\varepsilon \frac{\partial^2 H_z}{\partial t^2} \qquad [3\cdot25]$$

By differentiating 3·19 with respect to t, and 3·20 with respect to x, and carrying through a similar manipulation, we obtain

$$\frac{\partial^2 E_y}{\partial x^2} = \mu\varepsilon \frac{\partial^2 E_y}{\partial t^2} \qquad [3\cdot26]$$

* This assumes that **E** varies in amplitude only. A more general case exists in which **E** may be chosen to vary in angle also.

† See footnote, page 50.

Equations of the form 3·25 or 3·26 are well known in mathematical physics.* Equations of this type arise whenever wave motion is studied, for example, in the study of the vibrating string or the vibrating membrane. Hence equations of the type of 3·25 or 3·26 are called plane wave equations.

In solving differential equations it is often expedient to assume a solution, substitute it in the equation, and see whether there are any conditions under which it satisfies the equation. If conditions exist for which the assumed solution satisfies the equation, then it is a solution subject to these conditions. Let us assume that

$$E_y = E_y' \cos(\omega t - \beta x) \qquad [3\cdot27]\dagger$$

is a solution to 3·26 where β and ω are constants to be determined. This may be written in the form

$$E_y = E_y' \cos \omega \left(t - \frac{x}{v}\right) \qquad [3\cdot27a]$$

where $v = \omega/\beta$. Let us find the required partial derivatives and substitute them into 3·26. We have

$$\frac{\partial}{\partial x}[E_y' \cos(\omega t - \beta x)] = \beta E_y' \sin(\omega t - \beta x)$$

$$\frac{\partial^2}{\partial x^2}[E_y' \cos(\omega t - \beta x)] = -\beta^2 E_y' \cos(\omega t - \beta x)$$

Similarly

$$\frac{\partial^2}{\partial t^2}[E_y' \cos(\omega t - \beta x)] = -\omega^2 E_y' \cos(\omega t - \beta x)$$

Substituting in 3·26 we obtain

$$\frac{1}{\mu\varepsilon} E_y' \cos(\omega t - \beta x) = \frac{\omega^2}{\beta^2} E_y' \cos(\omega t - \beta x)$$

which yields

$$v^2 = \frac{\omega^2}{\beta^2} = \frac{1}{\mu\varepsilon}$$

* Page, *Introduction to Theoretical Physics*, D. Van Nostrand, Second Printing, 1929, page 150.

† The prime as used in this equation refers to the maximum value of the sinusoidal variation of the vector. Thus E_y' represents the maximum value of E_y. Other symbols that are commonly used to designate the maximum are E_0, E_m, and E_{max}. Since it is desirable to include the component designation in addition to the designation for maximum, double subscripts were avoided by the use of the "prime" designation.

or, taking the radical as positive

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \quad \text{or} \quad v = -\frac{1}{\sqrt{\mu\varepsilon}}$$

There are evidently two solutions corresponding to the two values of v. These are

$$E_y = E'_y \cos \omega(t - \sqrt{\mu\varepsilon}\, x)$$

and

$$E_y = E'_y \cos \omega(t + \sqrt{\mu\varepsilon}\, x)$$

If we have two solutions to a given differential equation, the sum of these is also a solution. Hence the solution of 3·26 may be written

$$E_y = E'_y \cos \omega\left(t - \frac{x}{v}\right) + E'_y \cos \omega\left(t + \frac{x}{v}\right) \qquad [3\cdot28]$$

where we take

$$v = +\frac{1}{\sqrt{\mu\varepsilon}}$$

The first term on the right side of equation 3·28 represents a wave traveling in the $+x$ direction, that is, away from the origin. To show this let us consider the value of E_y as given by this term at a distance x_1 from the origin at a time t_1. Then, neglecting for the moment the second term,

$$E_y = E'_y \cos \omega\left(t_1 - \frac{x_1}{v}\right)$$

This value of E_y is the same all over a plane normal to the X direction at the distance x_1 from the origin. Since E_y is a periodic function of distance and time, it will have this same value over a plane at a distance $x_2 > x_1$ from the origin at a later time t_2 if

$$t_1 - \frac{x_1}{v} = t_2 - \frac{x_2}{v}$$

that is, if the argument $\left(t - \frac{x}{v}\right)$ is the same at both planes. This will be true provided that

$$\frac{1}{v}(x_2 - x_1) = t_2 - t_1$$

or provided that

$$\frac{x_2 - x_1}{t_2 - t_1} = v = \frac{1}{\sqrt{\mu\varepsilon}}$$

We see, therefore, that v is the velocity at which the wave is propagated in the direction of x increasing, that is, the $+x$ direction. The first term of equation 3·28 is called the incident wave. The second term, however, represents a wave moving in the opposite direction, for if we consider the argument $\left(t + \frac{x}{v}\right)$ we find that in order for $E'_y \cos \omega \left(t + \frac{x}{v}\right)$ to have the same value in both planes it is sufficient that

$$t_1 + \frac{x_1}{v} = t_2 + \frac{x_2}{v}$$

or

$$v = -\frac{x_2 - x_1}{t_2 - t_1}$$

Since in this discussion velocity and time are not negative, and since $x_2 > x_1$, it is evident that the wave must be traveling in the direction of x decreasing, that is the $-x$ direction. This term is referred to as the reflected wave.

Let us consider the wave moving in the direction of x increasing, that is, the forward-moving wave

$$E_y = E'_y \cos \omega \left(t - \frac{x}{v}\right) \qquad [3·28a]$$

At the time $t = 0$

$$E_y = E'_y \cos \left(-\frac{\omega}{v} x\right) = E'_y \cos \beta x \qquad [3·29]$$

From this equation we can determine the value of E_y at all points in space at this instant. Let us set up the axes shown in Fig. 3·2 in which we may draw the curve

$$y = E'_y \cos \beta x$$

Over any plane parallel to the YOZ plane the electric intensity has a value given by the ordinate of this curve at the value of x for which the plane is drawn. This value of E_y is the same at all points in this

plane, and is in the y direction at all points. Similarly over any other plane parallel to this one the value of E_y is the same as the value of the ordinate to the curve $E_y' \cos \beta x$ at which the plane is drawn. Hence equation 3 28a gives us the distribution of E_y throughout all space at any instant.

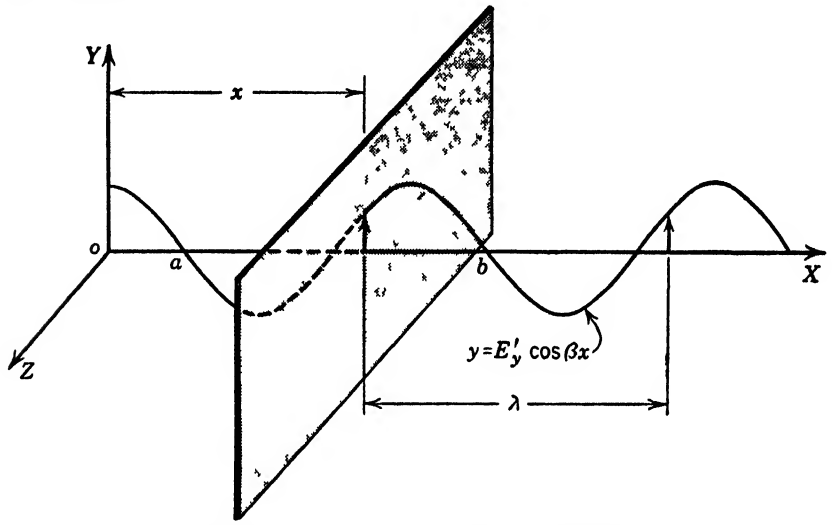


FIG 3 2 Illustration of a wave in free space

From Fig 3 2 it is clear that the value of E_y is unchanged if x is increased by some amount λ. Let us set $\beta = 2\pi/\lambda$. Then we may write

$$E_y' \cos \beta x = E_y' \cos \frac{2\pi}{\lambda} (x + \lambda)$$

since $\cos \beta x = \cos (\beta x + 2\pi)$ The constant λ is known as the wavelength; it is the distance traveled in one cycle by a periodic function. It is equal to the distance ab, Fig. 3 2. If the wave travels a distance λ in a time T

$$v = \frac{\lambda}{T}$$

The constant T is called the period of the wave. It represents the time required for the wave to pass through a complete cycle of positive and negative values. Hence the frequency f of the wave, defining the number of cycles per second, is

$$f = \frac{1}{T}$$

giving the very useful relation

$$v = f\lambda$$

The velocity of propagation is determined by the constants ϵ and μ of the medium. If the medium is air, $\epsilon = \epsilon_0$ and $\mu = \mu_0$ and

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 3 \times 10^8 \text{ meters per second}$$

It has been found experimentally that this velocity is equal to the velocity of light, within the limits of measurement. Hence Maxwell's equations predict the identity of the velocity of light and that of radio waves in free space.

The magnetic intensity H_z may be found from equation 3·19, where

$$\frac{\partial H_z}{\partial x} = -\varepsilon \frac{\partial E_y}{\partial t}$$

Since (forward-moving wave)

$$E_y = E'_y \cos(\omega t - \beta x)$$

$$\frac{\partial E_y}{\partial t} = -\omega E'_y \sin(\omega t - \beta x)$$

hence

$$\frac{\partial H_z}{\partial x} = \varepsilon\omega E'_y \sin(\omega t - \beta x)$$

Integrating, we obtain

$$H_z = \varepsilon \frac{\omega}{\beta} E'_y \cos(\omega t - \beta x)$$

$$= \varepsilon v E'_y \cos(\omega t - \beta x)$$

Since

$$v = \frac{1}{\sqrt{\varepsilon\mu}} \qquad \varepsilon v = \sqrt{\frac{\varepsilon}{\mu}}$$

We obtain

$$H_z = \sqrt{\frac{\varepsilon}{\mu}} E'_y \cos(\omega t - \beta x) \qquad [3\cdot30]$$

$$= H'_z \cos(\omega t - \beta x)$$

Hence the maximum amplitude of the magnetic intensity is

$$H'_z = \sqrt{\frac{\varepsilon}{\mu}} E'_y \qquad [3\cdot30a]$$

The relation 3·30 is plotted in Fig. 3·3 with equation 3·29 at the time $t = 0$. The remarks made in connection with the vector E_y apply also to the vector H_z.

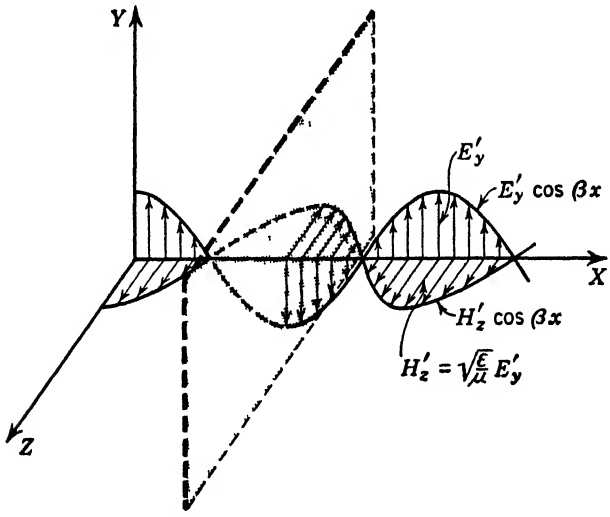


FIG. 3·3 Phase relations in a wave in free space.

3·6 General Discussion of Electromagnetic Waves

The water model illustrates in an extremely limited way the behavior of electromagnetic waves, but it is helpful in obtaining the concept of the wave front and the plane of the wave. Waves may be either guided or free in the same sense that a sound wave traveling inside a speaking tube is guided whereas one traveling in the outside air is free. Electric waves traveling in hollow pipes or along a transmission line are guided whereas waves radiated from an outside antenna are free.

It is often convenient in picturing the process occurring in wave phenomena to represent the intensity of electric and magnetic disturbances by the lines-of-force concept mentioned previously. The method is as follows. Lines are drawn in the direction of the intensity vectors. Solid lines represent the electric vector, and dotted lines the magnetic vector. Small arrows are usually drawn on the lines to indicate the direction of the force which would act on a small positive charge placed in the field. When the intensity is large the lines are drawn close together; when weak, they are drawn far apart. Thus the density of the lines indicates the strength of the field. Figure 3·4 is such a pictorial representation, where the plane waves discussed in the preceding section are reproduced. The height at which the lines stop is chosen

arbitrarily. Actually they extend in all directions through space wherever the electromagnetic wave exists. In Fig. 3·4 time is held fixed. Actually the whole picture moves in the x direction with velocity v and represents a harmonic plane wave traveling forward. The electric intensity **E** and magnetic intensity **H** are everywhere at right

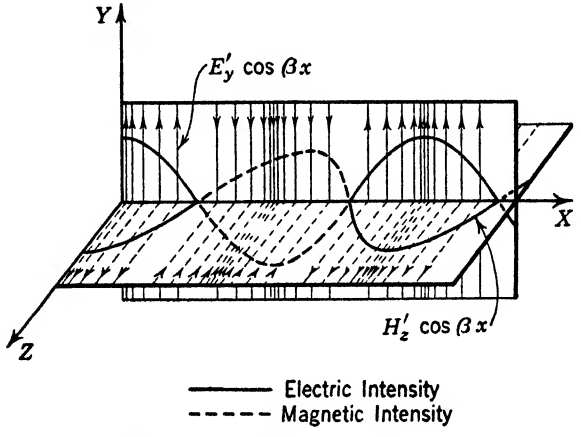


FIG. 3·4 A plane-polarized wave.

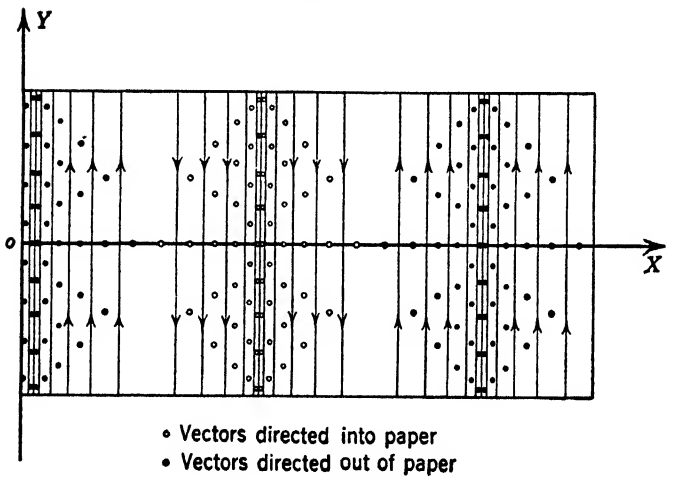


FIG. 3·5 Alternative representation of a plane-polarized wave.

angles to each other and to the direction of propagation. Such a wave as this is called polarized, in the optical sense. From the study of reflection of such waves it can be shown that the magnetic intensity lies in the plane of polarization, as the term is used in optics, and the electric intensity is therefore at right angles to the plane of polarization.

This plane, in Fig. 3·4, is the *XOZ* plane. It must be remembered that we have not considered the source of the waves at this time; we have assumed ourselves at a large distance from it. If we remove ourselves a sufficient distance from the seat of the electromagnetic disturbance setting up the waves, small sections of the wave front will be sufficiently flat to be considered plane.

When it is necessary to represent the wave configuration shown in Fig. 3·4 in a normal rather than a perspective drawing, it is convenient to use open and solid circles to represent vectors normal to the plane of the paper. The waves of Fig. 3·4 are represented in this way in Fig. 3·5. The small solid circles represent vectors directed toward the observer; the open circles, vectors directed away from the observer. Sometimes the convention is adopted of representing vectors directed toward the observer by open or solid circles and vectors directed away by small crosses. The former convention will be used throughout this book. The closeness or density of the circles is proportional to the strength of the field.

3·7 Energy in the Electromagnetic Wave

It was shown in an earlier section that the energy per unit volume associated with the electric intensity in a dielectric is $\epsilon E^2/2$, where E is measured in volts per meter and the energy is expressed in joules per cubic meter. In the plane wave, discussed in the preceding sections, the electric intensity varies sinusoidally in time and space. Moreover the magnetic field varies in the same manner and reaches its maximum at a particular point at the same instant that the electric field there is a maximum. The maximum density of energy at any instant during the cycle at a particular point is thus equal to

$$\frac{\mu H_z'^2}{2} + \frac{\epsilon E_y'^2}{2} \qquad [3\cdot31]$$

3·8 General Case of Propagation of Plane-Polarized Electromagnetic Waves in an Isotropic Medium

A more general solution to Maxwell's equations in a homogeneous isotropic medium may be obtained by an elimination process which will now be described. In this development the substitutions $\iota_c = \sigma\mathbf{E}$, $\mathbf{D} = \epsilon\mathbf{E}$, and $\mathbf{B} = \mu\mathbf{H}$ are used. Let us differentiate equation 3·12*a* with respect to time. Then

$$\frac{\partial}{\partial t}\left\{\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right\} = \frac{\partial^2 H_z}{\partial t\,\partial y} - \frac{\partial^2 H_y}{\partial t\,\partial z} = \sigma\frac{\partial E_x}{\partial t} + \epsilon\frac{\partial^2 E_x}{\partial t^2} \qquad [3\cdot32]$$

Differentiate 3·13*b* with respect to z and multiply through by $-1/\mu$,

obtaining

$$-\frac{1}{\mu}\frac{\partial}{\partial z}\left\{\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right\}=-\frac{1}{\mu}\frac{\partial^2 E_x}{\partial z^2}+\frac{1}{\mu}\frac{\partial^2 E_z}{\partial z\,\partial x}=+\frac{\partial^2 H_y}{\partial z\,\partial t} \quad [3\cdot33]$$

Differentiate 3·13c with respect to y and multiply through by $1/\mu$, obtaining

$$\frac{1}{\mu}\frac{\partial}{\partial y}\left\{\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right\}=\frac{1}{\mu}\frac{\partial^2 E_y}{\partial y\,\partial x}-\frac{1}{\mu}\frac{\partial^2 E_x}{\partial y^2}=-\frac{\partial^2 H_z}{\partial y\,\partial t} \quad [3\cdot34]$$

Add equations 3·32, 3·33, and 3·34, collecting terms involving H on the left-hand side and terms involving E on the right-hand side. This gives

$$\frac{\partial^2 H_z}{\partial t\,\partial y}-\frac{\partial^2 H_y}{\partial t\,\partial z}+\frac{\partial^2 H_y}{\partial z\,\partial t}-\frac{\partial^2 H_z}{\partial y\,\partial t}=\sigma\frac{\partial E_x}{\partial t}$$
$$+\varepsilon\frac{\partial^2 E_x}{\partial t^2}-\frac{1}{\mu}\left\{\frac{\partial^2 E_x}{dz^2}-\frac{\partial^2 E_z}{\partial z\,\partial x}-\frac{\partial^2 E_y}{\partial y\,\partial x}+\frac{\partial^2 E_x}{\partial y^2}\right\} \quad [3\cdot35]$$

If

$$\frac{\partial^2 H_z}{\partial t\,\partial y}=\frac{\partial^2 H_z}{\partial y\,\partial t} \quad \text{and} \quad \frac{\partial^2 H_y}{\partial t\,\partial z}=\frac{\partial^2 H_y}{\partial z\,\partial t}$$

then the left member of 3·35 is zero. Rearranging the right member we have

$$0=-\frac{1}{\mu}\left\{-\left(\frac{\partial^2 E_y}{\partial y\,\partial x}+\frac{\partial^2 E_z}{\partial z\,\partial x}\right)+\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}\right\}+\sigma\frac{\partial E_x}{\partial t}+\varepsilon\frac{\partial^2 E_x}{\partial t^2} \quad [3\cdot36]$$

Again, if

$$\frac{\partial^2 E_y}{\partial y\,\partial x}=\frac{\partial^2 E_y}{\partial x\,\partial y} \quad \text{and} \quad \frac{\partial^2 E_z}{\partial z\,\partial x}=\frac{\partial^2 E_z}{\partial x\,\partial z}$$

then

$$\frac{\partial^2 E_y}{\partial y\,\partial x}+\frac{\partial^2 E_z}{\partial z\,\partial x}=\frac{\partial}{\partial x}\left(\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}\right) \quad [3\cdot37]$$

From equation 3·10

$$\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=-\frac{\partial E_x}{\partial x}+\frac{\rho}{\varepsilon} \quad [3\cdot38]$$

and hence

$$\frac{\partial^2 E_y}{\partial y\,\partial x}+\frac{\partial^2 E_z}{\partial z\,\partial x}=\frac{\partial}{\partial x}\left(-\frac{\partial E_x}{\partial x}+\frac{\rho}{\varepsilon}\right)=-\frac{\partial^2 E_x}{\partial x^2}+\frac{1}{\varepsilon}\frac{\partial \rho}{\partial x} \quad [3\cdot39]$$

On substituting this relation in 3·36 we get

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} = \mu\sigma \frac{\partial E_x}{\partial t} + \mu\varepsilon \frac{\partial^2 E_x}{\partial t^2} + \frac{1}{\varepsilon}\frac{\partial \rho}{\partial x} \qquad [3\cdot40]$$

Similar equations may be obtained in E_y and E_z by eliminating the H components from equations 3·12 and 3·13. For example, to obtain the equation in E_y, first differentiate 3·12*b* with respect to t. Then differentiate 3·13*a* with respect to z so as to eliminate the terms $\dfrac{\partial^2 H_x}{\partial t\, \partial z}$ and $\dfrac{\partial^2 H_x}{\partial z\, \partial t}$. In order to eliminate the term $\dfrac{\partial^2 H_z}{\partial t\, \partial x}$ that is obtained from 3·12*b*, it is necessary to differentiate 3·13*c* with respect to x. Then, when the terms are collected as before, we obtain the desired equation in E_y.* The equations in E may be written then as

$$\left.\begin{aligned}
\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} &= \mu\sigma \frac{\partial E_x}{\partial t} + \mu\varepsilon \frac{\partial^2 E_x}{\partial t^2} + \frac{1}{\varepsilon}\frac{\partial \rho}{\partial x} \quad (a)\\
\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_y}{\partial z^2} &= \mu\sigma \frac{\partial E_y}{\partial t} + \mu\varepsilon \frac{\partial^2 E_y}{\partial t^2} + \frac{1}{\varepsilon}\frac{\partial \rho}{\partial y} \quad (b)\\
\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} &= \mu\sigma \frac{\partial E_z}{\partial t} + \mu\varepsilon \frac{\partial^2 E_z}{\partial t^2} + \frac{1}{\varepsilon}\frac{\partial \rho}{\partial z} \quad (c)
\end{aligned}\right\} \qquad [3\cdot41]$$

We can obtain a similar set of relations in H by eliminating E just as H was eliminated above. Carrying through the manipulation:

$$\left.\begin{aligned}
\frac{\partial^2 H_x}{\partial x^2} + \frac{\partial^2 H_x}{\partial y^2} + \frac{\partial^2 H_x}{\partial z^2} &= \mu\varepsilon \frac{\partial^2 H_x}{\partial t^2} + \mu\sigma \frac{\partial H_x}{\partial t} \quad (a)\\
\frac{\partial^2 H_y}{\partial x^2} + \frac{\partial^2 H_y}{\partial y^2} + \frac{\partial^2 H_y}{\partial z^2} &= \mu\varepsilon \frac{\partial^2 H_y}{\partial t^2} + \mu\sigma \frac{\partial H_y}{\partial t} \quad (b)\\
\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \frac{\partial^2 H_z}{\partial z^2} &= \mu\varepsilon \frac{\partial^2 H_z}{\partial t^2} + \mu\sigma \frac{\partial H_z}{\partial t} \quad (c)
\end{aligned}\right\} \qquad [3\cdot42]$$

Relations 3·41 and 3·42 may be more concisely written as

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu\sigma \frac{\partial \mathbf{E}}{\partial t} + \frac{1}{\varepsilon} \operatorname{\mathbf{grad}} \rho \qquad [3\cdot43]$$

$$\nabla^2 \mathbf{H} = \mu\varepsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} + \mu\sigma \frac{\partial \mathbf{H}}{\partial t} \qquad [3\cdot44]$$

* It is highly desirable for the student to carry through this manipulation in order to familiarize himself with the process.

In an uncharged insulating medium $\rho = 0$ and $\sigma = 0$, and these equations become

$$\nabla^2\mathbf{E} = \mu\varepsilon \frac{\partial^2\mathbf{E}}{\partial t^2} \qquad [3\cdot45]$$

$$\nabla^2\mathbf{H} = \mu\varepsilon \frac{\partial^2\mathbf{H}}{\partial t^2} \qquad [3\cdot46]$$

We may express the relations 3·45 and 3·46 by means of the equation

$$\nabla^2\mathbf{C} = \frac{1}{v^2} \frac{\partial^2\mathbf{C}}{\partial t^2} \qquad [3\cdot47]$$

where **C** is symbolic for E_x, E_y, E_z, H_x, H_y, or H_z, and $v^2 = 1/\mu\varepsilon$. As was mentioned before, this equation arises wherever wave motions are discussed and is therefore known as the wave equation. It is one of the most important partial differential equations of mathematical physics.

3·9 Direction Cosines and the Normal Form of the Equation of a Plane

We may find it helpful to review here several relations from analytic geometry.

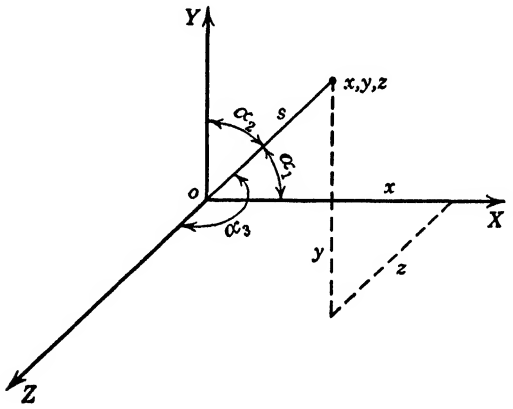


FIG. 3·6 Representation of direction cosines.

Consider the point x, y, z, Fig. 3·6, at a distance s from the origin. If α_1, α_2, and α_3 are the angles made by s with respect to the X, Y, and Z axes, we have

$$x = s \cos \alpha_1$$

$$y = s \cos \alpha_2$$

$$z = s \cos \alpha_3$$

But

$$x^2 + y^2 + z^2 = s^2 \cos^2 \alpha_1 + s^2 \cos^2 \alpha_2 + s^2 \cos^2 \alpha_3$$
$$= s^2 (\cos^2 \alpha_1 + \cos^2 \alpha_2 + \cos^2 \alpha_3)$$

Since $s^2 = x^2 + y^2 + z^2$ it follows that

$$\cos^2 \alpha_1 + \cos^2 \alpha_2 + \cos^2 \alpha_3 = 1 \qquad [3\cdot48]$$

It is often convenient to let $l = \cos \alpha_1$, $m = \cos \alpha_2$, and $n = \cos \alpha_3$, where l, m, and n are called the direction cosines of s. Then

$$l^2 + m^2 + n^2 = 1 \qquad [3\cdot49]$$

Let us now consider the angle θ between the lines AB and CD of Fig. 3·7. The projection of CD on AB is clearly the sum of the projections of CM, MK, and KD on AB. Thus

$$CD \cos \theta = CM \cos \alpha_1 + MK \cos \alpha_2 + KD \cos \alpha_3$$

where $l_1 = \cos \alpha_1$, $m_1 = \cos \alpha_2$, and $n_1 = \cos \alpha_3$ are the direction cosines of the line AB. But

$$CM = CD \cos \beta_1$$
$$MK = CD \cos \beta_2$$
$$KD = CD \cos \beta_3$$

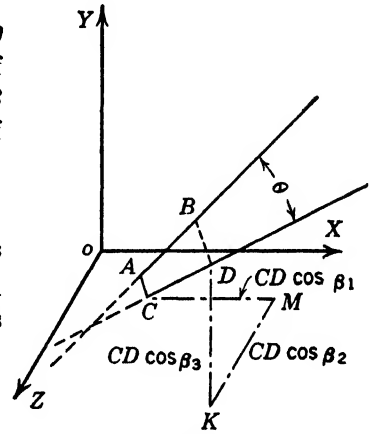


FIG. 3·7 Application of direction cosines.

where $l_2 = \cos \beta_1$, $m_2 = \cos \beta_2$, and $n_2 = \cos \beta_3$ are the direction cosines of CD. Thus we have

$$CD \cos \theta = CD \cos \alpha_1 \cos \beta_1 + CD \cos \alpha_2 \cos \beta_2 + CD \cos \alpha_3 \cos \beta_3$$

or

$$CD \cos \theta = CD(l_1 l_2 + m_1 m_2 + n_1 n_2)$$

Hence

$$\cos \theta = l_1 l_2 + m_1 m_2 + n_1 n_2 \qquad [3\cdot50]$$

giving the angle between two lines in terms of the direction cosines of each. This may also be expressed in terms of the sine of the angle θ since

$$\sin \theta = \sqrt{1 - \cos^2 \theta}$$

Therefore

$$\sin \theta = \sqrt{1 - (l_1 l_2 + m_1 m_2 + n_1 n_2)^2} \qquad [3\cdot50a]$$

Another relation which it will be useful to consider is the following: Let s be the length of a perpendicular drawn from the origin to the plane shown in Fig. 3·8. Let the point P, whose coordinates are x,

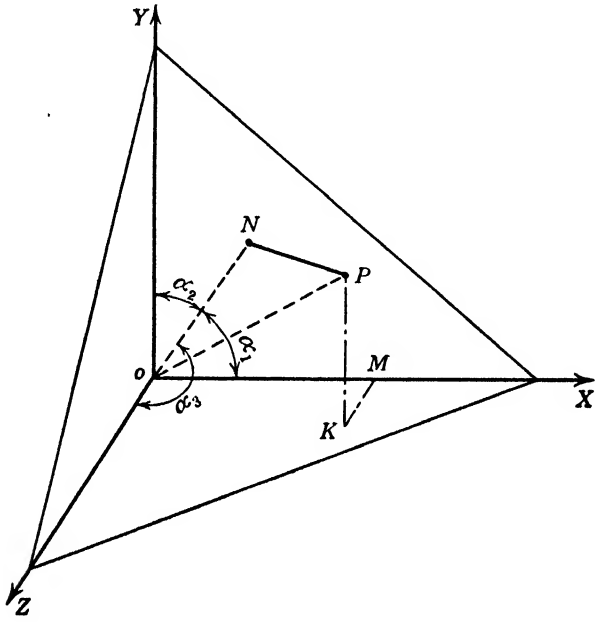


FIG. 3·8

y, z, be any point in the plane. Now the projection of OP on the normal ON to the plane is the sum of the projections of OM, MK, and KP on ON. But $OM = x$, $MK = y$, and $KP = z$. Also the projection of OM on ON is equal to $OM \cos \alpha_1$.

Similarly the projection of MK on ON is $MK \cos \alpha_3$, and of KP on ON is $KP \cos \alpha_2$. Thus

$$ON = s = OM \cos \alpha_1 + KP \cos \alpha_2 + MK \cos \alpha_3$$

giving

$$s = x \cos \alpha_1 + y \cos \alpha_2 + z \cos \alpha_3$$

or

$$s = lx + my + mz \qquad [3·51]$$

giving the normal form of the equation of a plane in rectangular coordinates, in terms of the direction cosines of the normal drawn from the origin to the plane.

3·10 · A General Solution of the Wave Equation

The wave equation 3·47 may be expressed as

$$\frac{\partial^2 C}{\partial s^2} = \frac{1}{v^2}\frac{\partial^2 C}{\partial t^2} \qquad [3\cdot52]$$

where C is symbolic for the components of E or H and is some function of s and t, i.e., $C = f(s, t)$, and s is given by

$$s = lx + my + nz$$

In the previous section it was seen that this is the equation of a plane, where l, m, n are the direction cosines of s, the perpendicular distance from the origin to the plane. The direction cosines may be expressed as

$$l = \frac{\partial s}{\partial x}, \quad m = \frac{\partial s}{\partial y}, \quad n = \frac{\partial s}{\partial z}$$

We may show that 3·52 is valid by obtaining the required partial derivatives and substituting in 3·47. Thus

$$\frac{\partial C}{\partial x} = \frac{\partial s}{\partial x}\cdot\frac{\partial C}{\partial s} = l\frac{\partial C}{\partial s}$$

$$\frac{\partial^2 C}{\partial x^2} = l\frac{\partial s}{\partial x}\cdot\frac{\partial^2 C}{\partial s^2} = l^2\frac{\partial^2 C}{\partial s^2}$$

Similarly

$$\frac{\partial^2 C}{\partial y^2} = m^2\frac{\partial^2 C}{\partial s^2} \qquad \text{and} \qquad \frac{\partial^2 C}{\partial z^2} = n^2\frac{\partial^2 C}{\partial s^2}$$

Thus

$$\nabla^2 \mathbf{C} = \frac{\partial^2 \mathbf{C}}{\partial x^2} + \frac{\partial^2 \mathbf{C}}{\partial y^2} + \frac{\partial^2 \mathbf{C}}{\partial z^2} = (l^2 + m^2 + n^2)\frac{\partial^2 \mathbf{C}}{\partial s^2} = \frac{1}{v^2}\frac{\partial^2 \mathbf{C}}{\partial t^2}$$

Since $l^2 + m^2 + n^2 = 1$ we obtain the wave equation 3·52. It is well known that any function of $s + at$ is a solution of this equation, where a is a constant to be determined. To illustrate, let

$$C = C(s + at) \qquad [3\cdot53]$$

where $C(s + at)$ is any function of $s + at$. Let us substitute this function into the partial differential equation 3·52. We may obtain

the required partial derivatives as follows:

$$\frac{\partial C}{\partial s} = \frac{\partial C}{\partial (s + at)} \cdot \frac{\partial (s + at)}{\partial s} = \frac{\partial C}{\partial (s + at)} = C'$$

$$\frac{\partial^2 C}{\partial s^2} = \frac{\partial^2 C}{\partial (s+at)^2} \cdot \frac{\partial (s + at)}{\partial s} = \frac{\partial^2 C}{\partial (s + at)^2} = C''$$

$$\frac{\partial C}{\partial t} = \frac{\partial C}{\partial (s + at)} \cdot \frac{\partial (s + at)}{\partial t} = a \frac{\partial C}{\partial (s + at)} = aC'$$

$$\frac{\partial^2 C}{\partial t^2} = \frac{\partial^2 C}{\partial (s+at)^2} \cdot \frac{\partial (s + at)}{\partial t} = a^2 \frac{\partial^2 C}{\partial (s+at)^2} = a^2 C''$$

Hence

$$C'' = \frac{a^2}{v^2} C'' \qquad [3{\cdot}54]$$

giving

$$a^2 = v^2$$

Thus $C(s + at)$ satisfies 3·52, provided that $a = \pm v$. Therefore $C_2(s + vt)$ is a solution of 3·52. Also, from 3·54, any function of $(s - vt)$ is also a solution. Let this solution be $C_1(s - vt)$. Thus the complete solution is

$$C = C_1(s - vt) + C_2(s + vt)$$

where C_1 and C_2 are arbitrary functions in which

$$v = \frac{1}{\sqrt{\mu\varepsilon}}$$

Let us consider the argument $(s - vt)$. Add to s an increment Δs and to t an increment Δt. Then $C_1(s - vt)$ will have the same value at a distance $s + \Delta s$ and time $t + \Delta t$ if

$$s + \Delta s - v(t + \Delta t) = s - vt$$

that is, provided that

$$\Delta s - v\Delta t = 0$$

or if

$$v = \frac{\Delta s}{\Delta t}$$

Thus we see that Δs is the distance that the plane of $C_1(s - vt)$ travels in the time Δt so that the velocity of propagation, v, of the plane of the

wave is

$$v = \frac{\Delta s}{\Delta t} = \frac{1}{\sqrt{\mu\varepsilon}}$$

in the direction s. In free space, $\mu = \mu_0 = 4\pi \times 10^{-7}$ henry per meter, and $\varepsilon = \varepsilon_0 = 8.85 \times 10^{-12}$ farad per meter. Thus

$$v = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = 2.998 \times 10^8 \cong 3 \times 10^8$$

meters per second. Since **C** is symbolic for the **E** and **H** vectors, or their components, Maxwell's equations tells us that the velocity at which a plane wave is propagated in free space is the velocity of light. In any other dielectric medium where ε and μ are constant quantities, the velocity of propagation of the electromagnetic waves is

$$v = \frac{1}{\sqrt{\mu\varepsilon}}$$

It can be shown in a similar manner that $C_2(s + vt)$ represents a wave traveling in the opposite direction with the same velocity. As far as the wave equations are concerned the general configuration of the waves C_1 and C_2 is wholly arbitrary. It is necessary only that **E** and **H** be mutually perpendicular and that they lie in the wave front. The form of the functions C_1 and C_2 depends upon the exciting source which produces the waves.

3·11 The Poynting Vector

Let us consider a portion of a region in which an electromagnetic field is changing. It is evident that energy must be flowing into and out of this portion across the surface surrounding it. We may determine an expression for this energy flow from equations 3·12 and 3·13 which are repeated below for convenience.

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \iota_{cx} + \varepsilon\frac{\partial E_x}{\partial t} \quad (a)\\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \iota_{cy} + \varepsilon\frac{\partial E_y}{\partial t} \quad (b)\\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \iota_{cz} + \varepsilon\frac{\partial E_z}{\partial t} \quad (c) \end{aligned}\right\} \text{[3·12]} \qquad \left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu\frac{\partial H_x}{\partial t} \quad (a)\\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu\frac{\partial H_y}{\partial t} \quad (b)\\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu\frac{\partial H_z}{\partial t} \quad (c) \end{aligned}\right\} \text{[3·13]}$$

D has been replaced by ε**E** and **B** by μ**H**. Let us multiply the three equations 3·12 by E_x, E_y, and E_z, respectively, and the three equations

3·13 by $-H_x$, $-H_y$, and $-H_z$, obtaining:

$$\left.\begin{aligned} E_x\frac{\partial H_z}{\partial y} - E_x\frac{\partial H_y}{\partial z} &= \iota_{cx}E_x + \varepsilon E_x\frac{\partial E_x}{\partial t} \\ E_y\frac{\partial H_x}{\partial z} - E_y\frac{\partial H_z}{\partial x} &= \iota_{cy}E_y + \varepsilon E_y\frac{\partial E_y}{\partial t} \\ E_z\frac{\partial H_y}{\partial x} - E_z\frac{\partial H_x}{\partial y} &= \iota_{cz}E_z + \varepsilon E_z\frac{\partial E_z}{\partial t} \end{aligned}\right\} \text{[3·55]} \qquad \left.\begin{aligned} -H_x\frac{\partial E_z}{\partial y} + H_x\frac{\partial E_y}{\partial z} &= \mu H_x\frac{\partial H_x}{\partial t} \\ -H_y\frac{\partial E_x}{\partial z} + H_y\frac{\partial E_z}{\partial x} &= \mu H_y\frac{\partial H_y}{\partial t} \\ -H_z\frac{\partial E_y}{\partial x} + H_z\frac{\partial E_x}{\partial y} &= \mu H_z\frac{\partial H_z}{\partial t} \end{aligned}\right\} \text{[3·56]}$$

Let us now add the equations 3·55 and 3·56 and at the same time collect terms in such a way that terms differentiated with respect to x, y, and z are grouped together on one side of the equation while derivatives with respect to t are grouped on the other side.

We obtain:

$$\left.\begin{aligned} -E_y\frac{\partial H_z}{\partial x} + E_z\frac{\partial H_y}{\partial x} + H_y\frac{\partial E_z}{\partial x} - H_z\frac{\partial E_y}{\partial x} \\ -E_z\frac{\partial H_x}{\partial y} + E_x\frac{\partial H_z}{\partial y} + H_z\frac{\partial E_x}{\partial y} - H_x\frac{\partial E_z}{\partial y} \\ -E_x\frac{\partial H_y}{\partial z} + E_y\frac{\partial H_x}{\partial z} + H_x\frac{\partial E_y}{\partial z} - H_y\frac{\partial E_x}{\partial z} \end{aligned}\right\} = \left\{\begin{aligned} &\iota_{cx}E_x + \iota_{cy}E_y + \iota_{cz}E_z \\ &+\varepsilon E_x\frac{\partial E_x}{\partial t} + \varepsilon E_y\frac{\partial E_y}{\partial t} + \varepsilon E_z\frac{\partial E_z}{\partial t} \\ &+\mu H_x\frac{\partial H_x}{\partial t} + \mu H_y\frac{\partial H_y}{\partial t} + \mu H_z\frac{\partial H_z}{\partial t} \end{aligned}\right. \text{[3·57]}$$

Let us consider the first four terms of equation 3·57. It can be seen that these may be represented by the negative of the partial derivative of a vector component S_x, where $S_x = (E_yH_z - E_zH_y)$. That is

$$-\frac{\partial S_x}{\partial x} = -\frac{\partial}{\partial x}(E_yH_z - E_zH_y) = -E_y\frac{\partial H_z}{\partial x} - H_z\frac{\partial E_y}{\partial x} + E_z\frac{\partial H_y}{\partial x} + H_y\frac{\partial E_z}{\partial x} \qquad \text{[3.57}a\text{]}$$

Similarly the next four terms may be represented by $-\dfrac{\partial S_y}{\partial y}$, where

$$S_y = (E_zH_x - E_xH_z) \qquad \text{[3·57}b\text{]}$$

and the following four terms by $-\dfrac{\partial S_z}{\partial z}$, where

$$S_z = (E_xH_y - E_yH_x) \qquad \text{[3·57}c\text{]}$$

Thus the left side of equation 3·57 becomes

$$-\left(\frac{\partial S_x}{\partial x} + \frac{\partial S_y}{\partial y} + \frac{\partial S_z}{\partial z}\right) = -\operatorname{div}\mathbf{S} \qquad \text{[3·58]}$$

Let us now consider the terms involving the partial time derivative on

the right-hand side of equation 3·57. Take, for example, the term $\varepsilon E_x \frac{\partial E_x}{\partial t}$. This may be expressed as $\frac{\varepsilon}{2}\frac{\partial (E_x)^2}{\partial t}$.* Thus equation 3·57 may be written

$$-\left(\frac{\partial S_x}{\partial x}+\frac{\partial S_y}{\partial y}+\frac{\partial S_z}{\partial z}\right)=\iota_{cx}E_x+\iota_{cy}E_y+\iota_{cz}E_z$$
$$+\frac{\partial}{\partial t}\left(\frac{\varepsilon E_x^2}{2}+\frac{\varepsilon E_y^2}{2}+\frac{\varepsilon E_z^2}{2}\right)+\frac{\partial}{\partial t}\left(\frac{\mu H_x^2}{2}+\frac{\mu H_y^2}{2}+\frac{\mu H_z^2}{2}\right) \quad [3\cdot59]$$

The field equations 3·10, 3·11, 3·12, and 3·13 are sometimes called point relations, in that they define conditions at any point in the electromagnetic field, subject to the restrictions which we imposed upon them during their development. Equation 3·59 has been derived from them directly; it is itself a point relation. To analyze it further let us integrate each term in it over a volume τ in the field and attempt to interpret the meaning of the various volume integrals.

Equation 3·59, integrated over a volume τ in the field, becomes

$$\iiint_\tau\left(\frac{\partial S_x}{\partial x}+\frac{\partial S_y}{\partial y}+\frac{\partial S_z}{\partial z}\right)d\tau+\iiint_\tau(\iota_{cx}E_x+\iota_{cy}E_y+\iota_{cz}E_z)\,d\tau$$
$$=-\iiint_\tau\frac{\partial}{\partial t}\left(\frac{\varepsilon E_x^2}{2}+\frac{\varepsilon E_y^2}{2}+\frac{\varepsilon E_z^2}{2}\right)d\tau$$
$$-\iiint_\tau\frac{\partial}{\partial t}\left(\frac{\mu H_x^2}{2}+\frac{\mu H_y^2}{2}+\frac{\mu H_z^2}{2}\right)d\tau \quad [3\cdot60]$$

It can be shown that the first integral of equation 3·60 represents the outward flux of the vector **S** over the surface of the volume τ. Consider the surface integral of **S** out of the volume element $\Delta\tau$ shown in

* To illustrate, let $E_x = Ae^{(x-vt)}$. Then $E_x^2 = A^2e^{2(x-vt)}$. But

$$\frac{\partial E_x}{\partial t} = -Ave^{(x-vt)} \quad \text{and} \quad \frac{\partial (E_x)^2}{\partial t} = -2A^2ve^{2(x-vt)}$$

Hence

$$E_x\frac{\partial E_x}{\partial t} = (Ae^{(x-vt)})(-Ave^{(x-vt)})$$
$$= -A^2ve^{2(x-vt)}$$

Therefore

$$E_x\frac{\partial E_x}{\partial t} = \frac{1}{2}\frac{\partial (E_x)^2}{\partial t}$$

Fig. 3·9. The net outward flux through the surfaces of the volume parallel to the YOZ plane is

$$-\bar{S}_x\,\Delta y\,\Delta z+\left(\bar{S}_x+\frac{\partial \bar{S}_x}{\partial x}\Delta x\right)\Delta y\,\Delta z=\frac{\partial \bar{S}_x}{\partial x}\Delta\tau$$

where $\bar{S}_x$ is the average of S_x over the areas $\Delta y\,\Delta z$. Similarly the outward flux through the other two pairs of surfaces is

$$\frac{\partial \bar{S}_y}{\partial y}\Delta\tau \qquad \text{and} \qquad \frac{\partial \bar{S}_z}{\partial z}\Delta\tau$$

Thus the net outward flux of S over the surface of the volume element $\Delta\tau$ is

$$\left(\frac{\partial \bar{S}_x}{\partial x}+\frac{\partial \bar{S}_y}{\partial y}+\frac{\partial \bar{S}_z}{\partial z}\right)$$

and this is equal to the surface integral of S over the surface of $\Delta\tau$.

The total amount of S flowing out of a volume τ per second is simply

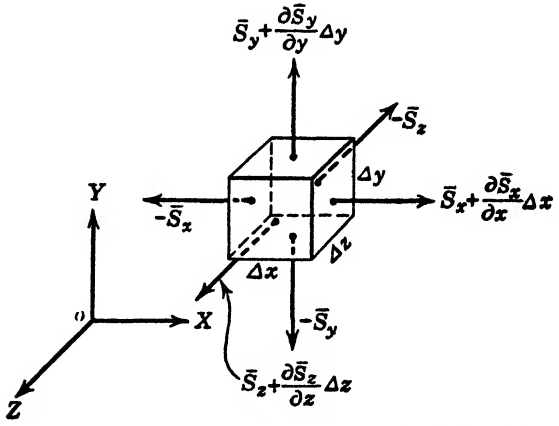


Fig. 3·9 An infinitesimal volume $\Delta\tau=\Delta x\,\Delta y\,\Delta z$ at a point x, y, z in a region in which an electromagnetic field exists. The bar over the energy vectors indicates the average value of the vectors over the various faces of the volume.

the sum of the amounts flowing from each element of the volume. This, in the limit as $\Delta\tau$ approaches zero, is

$$\int\int\int_{\tau}\left(\frac{\partial S_x}{\partial x}+\frac{\partial S_y}{\partial y}+\frac{\partial S_z}{\partial z}\right)d\tau$$

Further, all this amount flows through the surface s surrounding the volume τ, so that the rate of outward flow is $\int\int S_n\,ds$. Since these

two expressions are equal to each other we have

$$\int\int\int_{\tau}\left(\frac{\partial S_x}{\partial x}+\frac{\partial S_y}{\partial y}+\frac{\partial S_z}{\partial z}\right)d\tau=\int\int_{s}S_n\,ds$$

This relation is true for any vector function of position; it is known as Gauss's theorem.

Let us next consider the second term in equation 3·60. If the region in which our volume element is taken has a conductivity σ per unit volume, then there will be a diffusion of current through the volume.

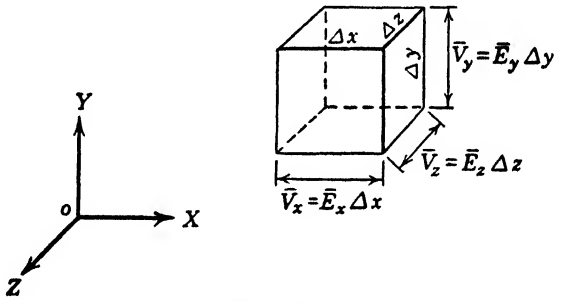


FIG. 3·10

We may study the energy relations in this volume by considering the current flow through it as related to the field intensity vectors. See Fig. 3·10.

Let the average electric intensity between the $\Delta y\,\Delta z$ faces be $\bar{E}_x$. Then the average potential difference between these faces will be $\bar{E}_x\,\Delta x$. Let the average component of current density flowing between the faces be ι_{cx}. Then the average current flowing between them will be $\iota_{cx}\,\Delta y\,\Delta z$. The product of potential difference and current is power, or energy expended per second. Hence the power expended in causing the current to flow in the x direction is

$$\Delta\bar{P}_x=\bar{E}_x\iota_{cx}\,\Delta x\,\Delta y\,\Delta z$$

Similarly the average energy expended per second by the field in setting up currents in the y and z directions is

$$\Delta\bar{P}_y=\bar{E}_y\iota_{cy}\,\Delta x\,\Delta y\,\Delta z$$

$$\Delta\bar{P}_z=\bar{E}_z\iota_{cz}\,\Delta x\,\Delta y\,\Delta z$$

The average power expended in the volume element is thus

$$\Delta\bar{P}=(\bar{E}_x\iota_{cx}+\bar{E}_y\iota_{cy}+\bar{E}_z\iota_{cz})\,\Delta\tau$$

where

$$\Delta\tau=\Delta x\,\Delta y\,\Delta z$$

By dividing through by $\Delta\tau$ and taking the limit as $\Delta\tau$ approaches zero we may obtain a point relation for the energy in the medium. Thus

$$\underset{\Delta\tau\to 0}{\text{Limit}}\ \frac{\Delta\bar{P}}{\Delta\tau} = \frac{\partial P}{\partial\tau} = E_x\iota_{cx} + E_y\iota_{cy} + E_z\iota_{cz}$$

Hence the energy expended throughout the volume τ is

$$P = \int\int\int_{\tau} (E_x\iota_{cx} + E_y\iota_{cy} + E_z\iota_{cz})\, d\tau = \int\int\int_{\tau} \iota_c E \cos\theta\, d\tau$$

where θ is the angle between the directions of ι_c and **E** at any point in the region.

Since the current density ι_c consists of the movement of a charge density ρ at an average velocity **v**, the electromagnetic field must supply this energy.

The integrands appearing on the right side of equation 3·60 may be expressed more compactly by writing

$$\frac{\varepsilon E^2}{2} = \frac{\varepsilon E_x^2}{2} + \frac{\varepsilon E_y^2}{2} + \frac{\varepsilon E_z^2}{2}$$

and

$$\frac{\mu H^2}{2} = \frac{\mu H_x^2}{2} + \frac{\mu H_y^2}{2} + \frac{\mu H_z^2}{2}$$

It was shown in Chapter 1 that the relations $\varepsilon E^2/2$ and $\mu H^2/2$ represent the energy per unit volume contained in the electric and magnetic fields, respectively. Their sum

$$\frac{\varepsilon E^2}{2} + \frac{\mu H^2}{2}$$

represents the total energy contained in the electromagnetic field. Hence the integral

$$\int\int\int_{\tau}\left(\frac{\varepsilon E^2}{2} + \frac{\mu H^2}{2}\right) d\tau$$

represents the total energy of the field in the volume τ. The time derivative represents the rate of change of this total energy. Let us rewrite the energy equation 3·60 in the light of the above. It becomes

$$\underbrace{\int\int_{s} S_n\, ds}_{\text{Outward flux of S or rate at which energy is radiated.}} + \underbrace{\int\int\int_{\tau} \iota_c E \cos\theta\, d\tau}_{\text{Rate at which energy is expended in establishing current.}} = \underbrace{-\frac{\partial}{\partial t}\int\int\int_{\tau}\left(\frac{\varepsilon E^2}{2} + \frac{\mu H^2}{2}\right) d\tau}_{\text{Rate of decrease of electromagnetic field energy.}} \qquad [3\cdot61]$$

THE ENERGY EQUATION

It is therefore seen that the rate of decrease of the electromagnetic energy in any volume τ is given by the right side of equation 3·61, and this is equal to the rate at which energy is expended in establishing the current (generation of heat in τ), and the rate at which the vector **S** flows through the surface surrounding τ. We may conclude, therefore, that the first term of equation 3·61 represents the rate at which energy is passing through the surface, or the rate at which energy is being radiated. Equation 3·61, then, is a statement of the conservation of energy of the electromagnetic field.

The vector **S**, known as the Poynting vector, represents the flow of energy per second through unit area of surface. The components of **S** as given by equation 3·57 are

$$\begin{aligned} S_x &= E_y H_z - E_z H_y \\ S_y &= E_z H_x - E_x H_z \\ S_z &= E_x H_y - E_y H_x \end{aligned} \qquad [3\cdot62]$$

We can determine both the direction and magnitude of **S** by the use of direction cosines. Let us assign to the three vectors the direction cosines as follows:

Vector	Direction Cosines		
	x	y	z
S	l	m	n
E	l_1	m_1	n_1
H	l_2	m_2	n_2

[3·63]

Then from 3·62

$$\begin{aligned} S_x &= EH(m_1 n_2 - n_1 m_2) \\ S_y &= EH(n_1 l_2 - l_1 n_2) \\ S_z &= EH(l_1 m_2 - m_1 l_2) \end{aligned} \qquad [3\cdot64]$$

Now the magnitude of **S** is given by

$$S = \sqrt{S_x^2 + S_y^2 + S_z^2} \qquad [3\cdot65]$$

Substituting for S_x, S_y, and S_z from 3·64, we will obtain, after the algebraic manipulation:

$$S = EH\sqrt{1 - (l_1 l_2 + m_1 m_2 + n_1 n_2)^2} \qquad [3\cdot66]$$

From equation 3·50*a* it is seen that the radical in 3·66 is equal to the sine of the angle between the vectors whose direction cosines are represented in it. Thus

$$S = EH \sin\theta \quad \text{watts/meter}^2 \qquad [3\cdot67]$$

where θ is the angle between **E** and **H**. This establishes the magnitude of **S**. Its direction may be found by the following process. Let us first determine the direction cosines of **S** in terms of those of **E** and **H**. Now

$$S_x = (EH \sin\theta)l = EH(m_1n_2 - n_1m_2)$$
$$S_y = (EH \sin\theta)m = EH(n_1l_2 - l_1n_2)$$
$$S_z = (EH \sin\theta)n = EH(l_1m_2 - m_1l_2)$$

as given by 3·64. Hence

$$l = \frac{m_1n_2 - n_1m_2}{\sin\theta}$$
$$m = \frac{n_1l_2 - l_1n_2}{\sin\theta} \qquad [3\cdot68]$$
$$n = \frac{l_1m_2 - m_1l_2}{\sin\theta}$$

If θ_1 is the angle between **E** and **S**, we know from equation 3·50 that

$$\cos\theta_1 = ll_1 + mm_1 + nn_1$$

and, if θ_2 is the angle between **H** and **S**, we also know that

$$\cos\theta_2 = ll_2 + mm_2 + nn_2$$

Substituting for l, m, n as given in 3·68, we find after cancellation that $\cos\theta_1 = 0$ and $\cos\theta_2 = 0$. Hence θ_1 and θ_2 must be 90°. The space orientation of the **E**, **H**, and **S** vectors is shown in Fig. 3·11. Since

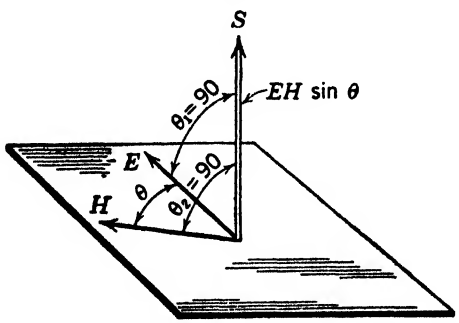


FIG. 3·11 The Poynting vector

$\theta_1 = \theta_2 = 90°$, it is evident that **S** is perpendicular to the plane in which the **E** and **H** vectors lie. The magnitude of **S** is given in equation 3·67.

The vector **S**, called Poynting's vector, satisfies the requirements

of the vector or cross product, since it is perpendicular to two other vectors and has a magnitude equal to the product of these vectors times the sine of the angle between them. In vector notation $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ or $[\mathbf{EH}]$, and is designated positive if taken in the direction indicated in Fig. 3·11. It will be observed that this is the direction a right-handed (standard) screw would move if the vector **E**, coinciding with the slot in the screw, were turned into **H**.

From this analysis we see that the direction of flow of energy is determined by the direction of the electric and magnetic intensity vectors, and its value is proportional to their product and to the sine of the angle between them. If the direction of either **E** or **H** is reversed, the direction of energy flow is reversed. If the direction of both **E** and **H** is reversed, their relative position is unchanged and the direction of energy flow is the same. Since the vectors **E** and **H** are traveling forward with a velocity **v**, the energy is streaming past any fixed point at this velocity.

The average power flowing across unit area of surface normal to the vector **S** may be computed with the aid of the Poynting vector. Since **S** represents the energy per second transferred across a unit area normal to **S**, the average power **p** flowing across the area is the integral of **S** over a period. Then

$$\mathbf{p} = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{S} d(\omega t) \qquad [3\cdot69]$$

Now

$$\mathbf{S} = \mathbf{i}S_x + \mathbf{j}S_y + \mathbf{k}S_z \qquad [3\cdot70]$$

Replacing S_x, S_y, and S_z from 3·62,

$$\mathbf{S} = \mathbf{i}(E_y H_z - E_z H_y) + \mathbf{j}(E_z H_x - E_x H_z) + \mathbf{k}(E_x H_y - E_y H_x) \qquad [3\cdot71]$$

This relation may be more concisely expressed by the determinant

$$\mathbf{S} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ E_x & E_y & E_z \\ H_x & H_y & H_z \end{vmatrix} \equiv \mathbf{E} \times \mathbf{H} \qquad [3\cdot72]$$

defining the vector or cross product of **E** and **H**. Hence equation 3·69 may be written

$$\mathbf{p} = \frac{1}{2\pi} \int_0^{2\pi} (\mathbf{E} \times \mathbf{H}) d(\omega t) \qquad [3\cdot73]$$

It is usually convenient to use the exponential form for the vectors **E** and **H** instead of the sinusoidal or cosinusoidal form when performing

mathematical manipulations with them. For example, we may express the relation

$$\mathbf{E} = \mathbf{E}' \cos(\omega t - \beta x) \quad \text{in the form} \quad \mathbf{E} = \mathbf{E}' e^{j(\omega t - \beta x)} *$$

where it is understood that, in the final solution, we will take only the real part.

This is sometimes written

$$\mathbf{E} = \text{real}\,(\mathbf{E}' e^{j(\omega t - \beta x)})$$

or

$$\mathbf{E} = \mathcal{R}(\mathbf{E}' e^{j(\omega t - \beta x)}) \qquad [3\cdot74]$$

with similar expressions for the **H** vector. The electric and magnetic vectors may in some cases be out of phase with one another, and in general we may write

$$\mathbf{E} = \mathbf{E}' \cos(\omega t + \theta_1) \qquad [3\cdot75]$$

$$\mathbf{H} = \mathbf{H}' \cos(\omega t + \theta_2) \qquad [3\cdot76]$$

where $\theta_1 = -\beta x + \phi_1$ and $\theta_2 = -\beta x + \phi_2$, ϕ_1 and ϕ_2 representing phase angles whose values are dependent upon the choice of the origin of time. Now $\cos(\omega t + \theta_1)$ and $\cos(\omega t + \theta_2)$ are scalar quantities, and hence

$$\mathbf{E} \times \mathbf{H} = \mathbf{E}' \times \mathbf{H}' \cos(\omega t + \theta_1) \cos(\omega t + \theta_2) \dagger \qquad [3\cdot77]$$

Hence equation 3·73 becomes

$$\mathbf{p} = \int_0^{2\pi} \frac{\mathbf{E}' \times \mathbf{H}'}{2\pi} \{\cos(\omega t + \theta_1) \cos(\omega t + \theta_2)\}\, d(\omega t)$$

Since $\mathbf{E}'$ and $\mathbf{H}'$ are not functions of time, this may be written

$$\mathbf{p} = \frac{\mathbf{E}' \times \mathbf{H}'}{2\pi} \int_0^{2\pi} \tfrac{1}{2}\{\cos(2\omega t + \theta_1 + \theta_2) + \cos(\theta_1 - \theta_2)\} d(\omega t) \ddagger$$

$$= \frac{\mathbf{E}' \times \mathbf{H}'}{2} \cos(\theta_1 - \theta_2) \qquad [3\cdot78]$$

* Since $\mathbf{E} = \mathbf{E}' e^{j(\omega t - \beta x)} = \mathbf{E}' [\cos(\omega t - \beta x) + j \sin(\omega t - \beta x)]$ the real part is $\mathbf{E}' \cos(\omega t - \beta x)$. Hence

$$\mathbf{E} = \mathcal{R}(\mathbf{E}' e^{j(\omega t - \beta x)}) = \mathbf{E}' \cos(\omega t - \beta x)$$

where it is understood that $\mathbf{E}'$, which represents the maximum value of the time and distance variation of **E**, is real.

† If we substitute $\mathbf{E} = a\mathbf{E}_1$ and $\mathbf{H} = b\mathbf{H}_1$, where a and b are scalars, into the determinant 3·72 defining the cross product, it may be shown that $\mathbf{E} \times \mathbf{H} = \mathbf{E}_1 \times \mathbf{H}_1\,(ab)$.

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

Now

$$\mathbf{E} = \mathcal{R}\mathbf{E}'e^{j(\omega t+\theta_1)} \qquad [3\cdot79]$$

and

$$\mathbf{H} = \mathcal{R}\mathbf{H}'e^{j(\omega t+\theta_2)} \qquad [3\cdot80]$$

We find it convenient at this point to introduce a new vector which is the conjugate of $\mathbf{H}$ and is written $\mathbf{H}^*$. It is given by

$$\mathbf{H}^* = \mathbf{H}'e^{-j(\omega t+\theta_2)} \qquad [3\cdot81]$$

By its use we have

$$\mathcal{R}(\mathbf{E}\times\mathbf{H}^*) = \mathcal{R}(\mathbf{E}'\times\mathbf{H}')e^{j(\theta_1-\theta_2)} \qquad [3\cdot82]$$

Hence 3·78 may be written

$$\mathbf{p} = \tfrac{1}{2}\mathcal{R}(\mathbf{E}\times\mathbf{H}^*) = \tfrac{1}{2}\mathcal{R}(\mathbf{E}'\times\mathbf{H}')e^{j(\theta_1-\theta_2)} \qquad [3\cdot83]$$

Thus the average power flowing through a unit area of surface normal to the direction of flow is given by the real part of the vector or cross product of the instantaneous electric intensity $\mathbf{E}$ and the conjugate of the instantaneous magnetic intensity $\mathbf{H}$. The vector $\mathbf{p}$ is known as the complex Poynting vector.

In the forms of the electromagnetic field which we shall consider, the phase angle $\theta_1 - \theta_2$ is usually zero. Under these circumstances the average power $\mathbf{p}$ reduces to

$$\mathbf{p} = \tfrac{1}{2}(\mathbf{E}'\times\mathbf{H}') = \tfrac{1}{2}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ E'_x & E'_y & E'_z \\ H'_x & H'_y & H'_z \end{vmatrix} \qquad [3\cdot84]$$

The vector $\mathbf{p}$ may be resolved into components p_x, p_y, and p_z directed along the coordinate axes so that

$$\mathbf{p} = \mathbf{i}p_x + \mathbf{j}p_y + \mathbf{k}p_z \qquad [3\cdot85]$$

$$= \mathbf{i}\tfrac{1}{2}(E'_yH'_z - E'_zH'_y) + \mathbf{j}\tfrac{1}{2}(E'_zH'_x - E'_xH'_z) + \mathbf{k}\tfrac{1}{2}(E'_xH'_y - E'_yH'_x) \qquad [3\cdot86]$$

$$\begin{aligned} p_x &= \tfrac{1}{2}(E'_yH'_z - E'_zH'_y) \\ p_y &= \tfrac{1}{2}(E'_zH'_x - E'_xH'_z) \\ p_z &= \tfrac{1}{2}(E'_xH'_y - E'_yH'_x) \end{aligned} \qquad [3\cdot87]$$

where p_x is the power flowing in the X direction across unit area parallel to the YOZ plane. Similarly p_y and p_z represent power flowing in the Y and Z directions through areas parallel to the XOZ and XOY planes, respectively. The total average power P flowing across any

given surface can now be computed by taking the surface integral of the normal component of **p** over the surface. Thus

$$P = \int\int_s p_n \, ds \qquad [3\cdot88]$$

In the problems which we shall consider, it is possible to choose the coordinate axes so that the surfaces under consideration represent constant values of the independent variables. For example, in the study of the rectangular wave guide, the sides of the guide are chosen parallel to the coordinate system. Thus setting $y = 0$ or $y = a$ defines a

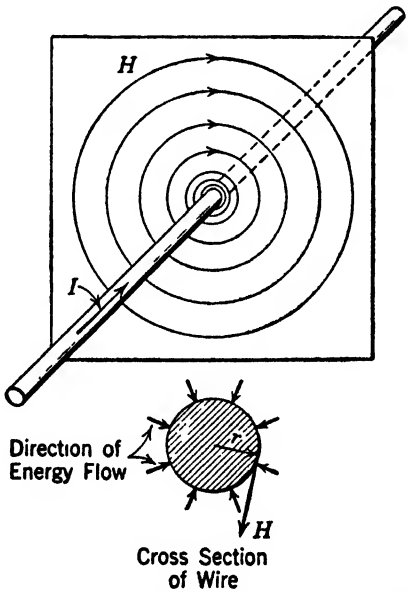


FIG. 3·12 Application of the Poynting vector.

surface which coincides with the top and bottom walls of the tube. Similarly setting $z = 0$ or b defines a surface coinciding with the side walls of the tube. When considering the flow of power into these surfaces from the electromagnetic fields within the guide, the normal component of **p** in the two cases are respectively p_y and p_z. The total power then is the product of these components and the surface areas in question.

To illustrate the operation of Poynting's vector, let us consider the following example. In Fig. 3·12 is shown a wire of finite resistance carrying a steady current I. In the dielectric surrounding the wire the lines of magnetic intensity form concentric circles around the wire

while the lines of electric intensity are in the direction of the wire. At the surface of the wire of radius r, the magnetic intensity is

$$H = \frac{I}{2\pi r} \qquad [3\cdot89]^*$$

Let the component of the electric intensity along the surface of the wire be E. Now E and H are at right angles to each other so that the rate at which energy is transferred from the dielectric into the wire is given by Poynting's vector S which, because the current is steady, equals p.

$$p = EH \sin\theta = \frac{EI}{2\pi r} \qquad [3\cdot90]$$

If we consider the energy entering the surface of the wire of length h, the surface area involved is $2\pi rh$. Thus the power through this surface is

$$P = 2\pi rh \frac{EI}{2\pi r} = EIh$$

The potential difference (IR drop) V across the section of the wire of length h is

$$V = Eh$$

* This relation may be obtained directly from equation 3·3, where

$$I = \oint_l H \cos\theta \, dl$$

Taking the line integral around a path coinciding with a line of constant magnetic intensity, H may be taken outside the integral sign and

$$I = H \oint_l dl$$

If r is the radius vector from the center of the wire to the path,

$$\oint_l dl = 2\pi r$$

Hence $I = 2\pi rH$ or $H = I/2\pi r$ in rationalized practical units. The student is probably more familiar with this expression as obtained in unrationalized emu where Maxwell's equation is

$$4\pi I = \oint_l H \cos\theta \, dl.$$

Carrying out the integration we get:

$$4\pi I = 2\pi rH \qquad \text{or} \qquad H = \frac{2I}{r}$$

Hence the flow of power into the wire is at the rate of

$$P = IV \text{ joules per second or watts} \qquad [3\cdot91]$$

This is a relation very familiar to electrical engineers. We can readily see that this result is correct only in the case of direct current or when the displacement current is small in comparison with the conduction current. Otherwise 3·89 would be

$$H = \frac{1}{2\pi r}\left(I + \pi r^2 \frac{\partial D}{\partial t}\right)$$

Under these conditions the power entering the wire per unit area is

$$p = \frac{E}{2\pi r}\left(I + \pi r^2 \varepsilon \frac{\partial E}{\partial t}\right) \qquad [3\cdot92]$$

and the power entering the surface of area $2\pi rh$ is

$$P = Eh\left\{I + \pi r^2 \varepsilon \frac{\partial E}{\partial t}\right\} = VI + \pi r^2 \varepsilon V \frac{\partial E}{\partial t} \qquad [3\cdot93]$$

The second term in equation 3·93 may be very important at high frequencies when $\partial E/\partial t$ is large and the conductivity is low.

PROBLEMS

3·1 Identify the following wave systems as to longitudinal or transverse: (*a*) motion of a violin string; (*b*) motion of a slender steel bar struck on the end; (*c*) motion of a slender steel bar struck transversely at its center; (*d*) motion of a thin circular steel plate struck at its center.

3·2 Water waves of the ordinary kind travel on the surface. Are they longitudinal or transverse? Sound waves travel through the volume of water. Are they longitudinal or transverse?

3·3 A loop antenna in air is formed by a single square turn of wire 0.5 meter on a side. The edges lie in the OX and OY directions and are acted upon by a plane electromagnetic wave moving in the $+X$ direction with its electric vector in the Y direction. The wave has a length of 3 meters and maximum electric intensity of 2 volts per meter. Calculate from the electric intensity the open-circuit voltage induced in the antenna.

3·4 Repeat problem 3·3, basing the calculation entirely on the magnetic intensity. Compare the results.

3·5 In sea water ρ is zero but σ is not negligible. Show what forms result when a solution of equation 3·43 is attempted by substituting $E = e^{j(s-\omega t)}$.

3·6 A vertical antenna (in the OY direction) 1 meter long is acted upon by the wave described in problem 3. Calculate the voltage induced in this antenna, considering first the action of the electric field, then the action of the magnetic field. Are these two voltages additive?

3·7 The loop antenna of problem 3 remains in the XOY plane but is rotated so that its diagonal lies in the OY direction. Calculate the open-circuit voltage induced.

3·8 A vector joins the origin and the point $x = 3$, $y = 4$, $z = 5$. Find the direction cosines.

3·9 Find the equation of the plane which passes through the point (3, 4, 5) and is normal to the vector which joins that point and the origin.

3·10 A vector joins the origin and the point $x = 4$, $y = 3$, $z = 5$. Find the direction cosines and the angle which this vector makes with the vector of problem 8 by means of the direction cosines.

3·11 Consider a coaxial conductor in which the radius of the inner conductor is a and the inner radius of the outer conductor is b. Show that the Poynting vector integrated over the area between conductors is equal to VI, indicating that the power is propagated between, rather than within, the conductors.

3·12 A plane electromagnetic wave in free space has a maximum electric intensity of 100 volts per meter. How much power is transmitted per square meter of wave front?

3·13 A plane electromagnetic wave travels in a medium having $\kappa_e = 2$, $\mu_m = 1.5$. The maximum value of the electric field is 5 volts per meter. Determine the intensity of the magnetic field and distribution of energy between the electric and magnetic fields.

3·14 A plane ectromagnetic wave in free space transmits power at the rate of 7 microwatts per square meter. Evaluate the intensity of the electric and magnetic fields involved.

3·15 Verify equation 3·24 for the following functions:

$$E_y = ax + by + cz$$
$$E_y = ax + by + ct$$
$$E_y = a \cos x + b + ct$$

where a, b, and c are constants.

CHAPTER 4

REFLECTION AND REFRACTION OF PLANE WAVES

It is well known that electromagnetic waves are reflected, at least in part, whenever they pass a surface of discontinuity. The most familiar example is visible light, which is reflected to a greater or lesser extent by every material surface. Polished metals reflect nearly all the incident light in a regular fashion; that is, the angle of incidence is equal to the angle of reflection. Other materials which we call white or light colored reflect a large proportion of the incident light, but in an irregular or scattered fashion. Dark-colored materials reflect only a small portion of the incident radiation. In all the above examples the radiant energy not reflected is absorbed in the medium. Glass and other transparent materials reflect only a small portion of the incident radiation and transmit the remainder with small loss so that almost all the power which enters the system as light is delivered as such. Regular reflection is characteristic of transparent materials even though some dark material is occasionally added to absorb the radiation which is not reflected.

In the following sections the familiar laws of physical optics will be developed by means of Maxwell's equations. It will be found advantageous to keep the optical problem constantly in mind in order to obtain a better understanding of the mathematical argument.

4·1 Reflection from a Perfect Conductor at Normal Incidence

It has already been shown that the tangential components of **E** and **H** are continuous at any boundary. That is

$$E_{1T} = E_{2T} \qquad [4\cdot1]$$

and

$$H_{1T} = H_{2T} \qquad [4\cdot2]$$

By a similar process we showed that the normal components of **D** and **B** are continuous at the boundary between any two media. These relations may be written

$$D_{1N} = D_{2N} \qquad [4\cdot3]$$

and

$$B_{1N} = B_{2N} \qquad [4\cdot4]$$

where the subscript 1 refers to conditions in the medium on one side of the boundary or dividing surface, and the subscript 2 refers to the

conditions in the other medium. The letter T designates the component of intensity which is at the point in question parallel or tangential to the surface. The letter N refers to the component of induction which is at right angles or normal to the surface.

Equations 4·3 and 4·4 are exact only when there is no distribution of free charge at the boundary surface. We are here concerned, however, only with wave phenomena, and the action of any fixed or static charge is not considered. Accordingly we shall use 4·3 and 4·4 as they are written, ignoring the separate problems in electrostatics and magnetostatics which may exist as a result of the action of free charges at the boundary surface.

The four relations expressed by equations 4·1 to 4·4 must be satisfied by an electromagnetic wave at any surface which separates two different media. These relations in addition to the fundamental wave equations which must be satisfied at all points permit us to calculate the basic laws of reflection and refraction.

Let us consider first the case in which an electromagnetic wave in free space impinges at right angles upon a very large flat plate of perfectly conducting metal. More general problems involving imperfect conductors, dielectric materials, and incidence at other angles will be considered later. Let the conducting surface be parallel to the YOZ plane. An electromagnetic wave similar to that shown in Figs. 4·1*a* and 4·1*b* is propagated to the right along the X axis. Such a wave may be described by the equations

$$E_y = E'_y \sin(\omega t - \beta x) \quad [4\cdot5]$$

and

$$H_z = H'_z \sin(\omega t - \beta x) \quad [4\cdot6]$$

} Incident wave

where

$$H'_z = \sqrt{\frac{\varepsilon}{\mu}}\, E'_y \qquad \text{and} \qquad \frac{\omega}{\beta} = v = \frac{1}{\sqrt{\varepsilon\mu}}$$

These equations describe a wave which is a harmonic function of time and space and is propagated in the $+X$ direction with the velocity v. The constants ε and μ refer to the space through which the wave moves and not to the metal reflecting surface.

A perfect conductor is defined by the fact that the electric intensity within the conductor is zero. From equation 4·1 it is therefore necessary that the tangential component of the electric intensity just outside the conductor is also zero. This fact is the basis of the argument by which the reflected wave is established. Since the electric intensity defined by equation 4·5 is not equal to zero at the conductor it is neces-

sary that an equal and opposite intensity be assumed which with 4·5 will add to zero. This second intensity characterizes the reflected wave. A trivial solution in which the electric intensity is everywhere zero results if the second or reflected wave is assumed to travel in the same direction as the original or incident wave. We therefore assume a wave of the same frequency and velocity traveling in the opposite direction. The reversal of the electric intensity, demanded by the

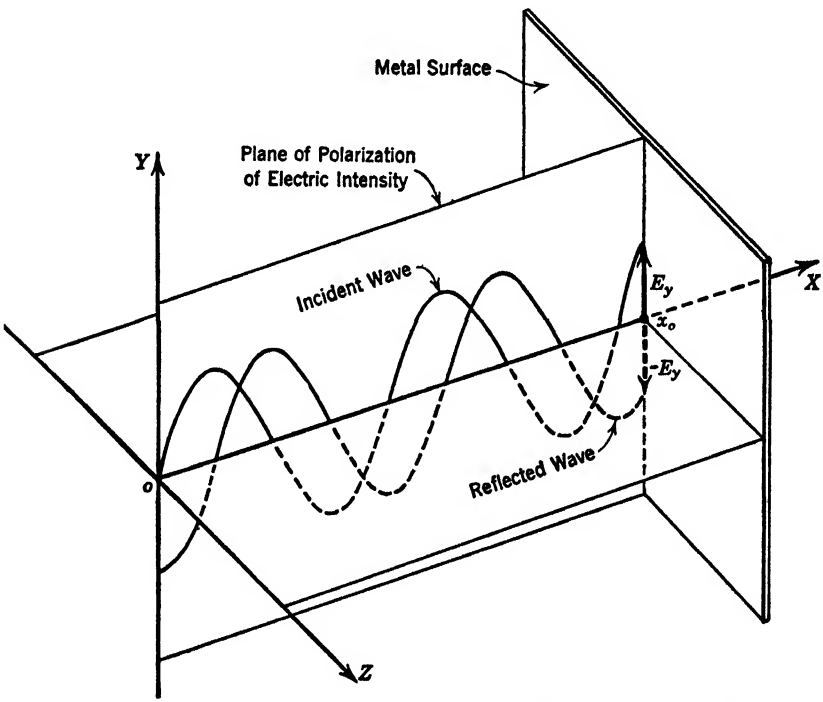


FIG. 4·1*a* Instantaneous distribution of electric intensity for reflection at normal incidence.

continuity relation at the surface of the conductor, and the reversal of the direction of propagation require that the magnetic intensity of the reflected wave be in the same direction as that of the incident wave. This conclusion is immediately apparent from a consideration of the rule expressed in Fig. 3·11. The two components of the reflected wave are characterized by the equations

$$\left.\begin{aligned} E_y &= -E'_y \sin(\omega t + \beta x) \\ H_z &= H'_z \sin(\omega t + \beta x) \end{aligned}\right\} \text{Reflected wave} \qquad \begin{matrix}[4\cdot7] \\ [4\cdot8]\end{matrix}$$

The sum of the incident and reflected waves produces standing waves, which are given for the electric intensity by the sum of equations 4·5 and 4·7. Thus

$$E_{y\,(\text{standing})} = E'_y\{\sin(\omega t - \beta x) - \sin(\omega t + \beta x)\} \qquad [4\cdot9]$$

For the magnetic intensity they are given by the sum of 4·6 and 4·8:

$$H_{z\,(\text{standing})} = H'_z\{\sin(\omega t - \beta x) + \sin(\omega t + \beta x)\} \qquad [4\cdot10]$$

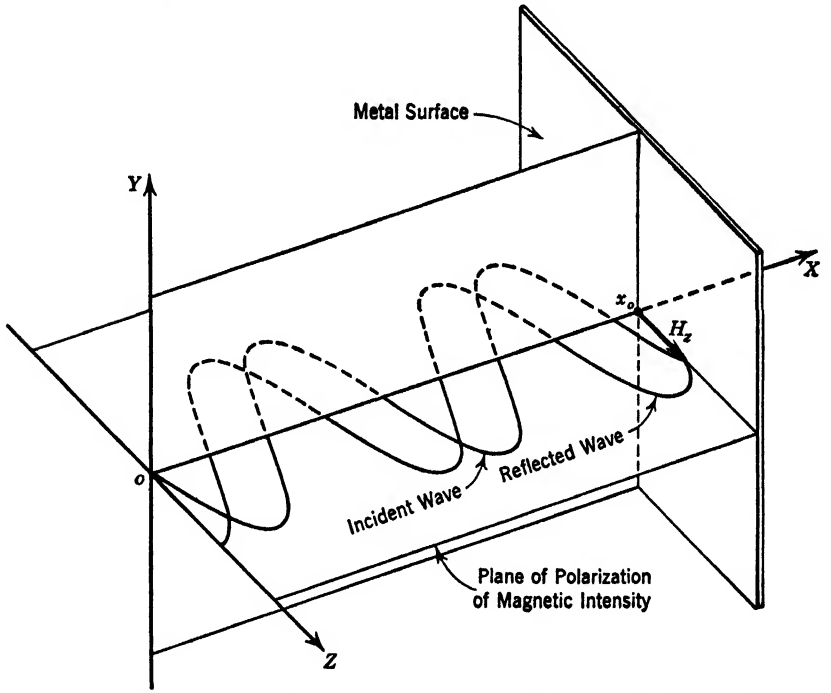


FIG. 4·1*b* Instantaneous distribution of magnetic intensity for reflection at normal incidence.

Equations 4·9 and 4·10, reduced by the trigonometric identity

$$\sin A \pm \sin B = 2\{\sin \tfrac{1}{2}(A \pm B)\cos \tfrac{1}{2}(A \mp B)\} \qquad [4\cdot11]$$

are

$$\left.\begin{aligned} E_{y\,(\text{standing})} &= -2E'_y \sin \beta x \cos \omega t \qquad [4\cdot12] \\ H_{z\,(\text{standing})} &= +2H'_z \sin \omega t \cos \beta x \qquad [4\cdot13] \end{aligned}\right\} \text{Standing wave}$$

At the plane $x = x_0$, where the perfect conductor is located, the standing waves are

$$E_{y\,(\text{standing})} = -2E'_y \sin \beta x_0 \cos \omega t \qquad [4\cdot14]$$

and

$$H_{z\,(\text{standing})} = -2H'_z \sin \omega t \cos \beta x_0 \qquad [4\cdot15]$$

Our particular choice of the phase of the wave makes it necessary to assign to x_0 such a value that $\sin \beta x_0$ is zero. Thus, E_y will be zero at the reflecting surface if x_0 is given by the relation

$$\beta x_0 = \frac{n\pi}{2} \qquad n = 0, 2, 4, 6 \cdots$$

At the planes defined by these values of x_0, the electric intensity is always zero, satisfying the initial continuity condition, and the magnetic intensity varies with time from zero to $2H'_z$.

A useful plot shown in Fig. 4·2 results if we develop, as a function of x, the greatest values of the electric and magnetic intensities which occur

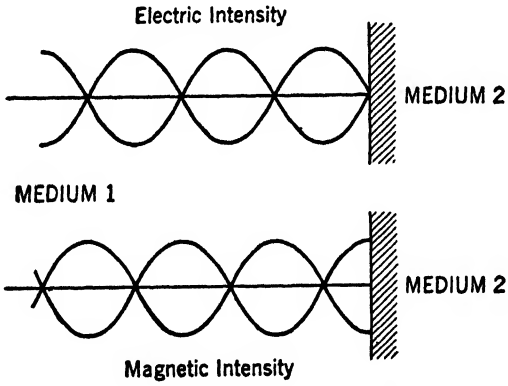


FIG. 4·2 Stationary waves.

over a cycle. By placing $\cos \omega t = \pm 1$ in equation 4·12 we obtain the maximum values of E_y over a cycle. The functional relationship is then given by

$$E_{y\,(\text{max standing})} = \pm 2E'_y \sin \beta x \qquad [4\cdot16]$$

Similarly, by placing $\sin \omega t = \pm 1$ in equation 4·13 we obtain the maximum values of H_z over a cycle. Hence

$$H_{z\,(\text{max standing})} = \pm 2H'_z \cos \beta x \qquad [4\cdot17]$$

Relations 4·16 and 4·17 are recognized as a system of standing waves similar to those existing on vibrating strings or other mechanical systems, and are shown in Fig. 4·2. Thus, although **E** and **H** of the incident wave are in time and space phase as are the **E** and **H** of the reflected wave, the total **E** and **H** vectors of the standing wave are in

time and space quadrature. The doubling of **H** at the reflecting surface may be thought of as due to finite currents induced in the reflecting metal by the action of the incident wave. In a conductor of zero resistance such currents entail neither electric field nor loss of power.

It must be understood that curves similar to those of Fig. 4·2 result if any value other than 1 is substituted for cos ωt. Thus if cos $\omega t = 0.707$ is used the resulting plot represents the space distribution of readings of rms instruments. One rather surprising aspect of this situation is that at certain points in space, called nodes, the electric intensity is always zero and at certain other nodes the magnetic intensity is always zero. At all other points the fields vary sinusoidally with time, and the values at any instant obey the sinusoidal space distribution shown.

4·2 Polarization

All the waves so far discussed have been of the type which is referred to as plane polarized. In such a wave the electric intensity is everywhere parallel to some plane and the magnetic intensity is everywhere perpendicular to that plane. Such a wave system was described in Chapter 3 and is illustrated in Fig. 3·4.

Plane-polarized waves are widely used in optics for such purposes as the testing of mechanical models, the determination of sugar content, and the elimination of glare. The waves produced by most commercial radio broadcasting stations are polarized, the electrical intensity being vertical and the magnetic intensity being horizontal, i.e., parallel to the earth's surface.

Ordinary light is composed of radiation from a large number of atoms or molecules. Each atom or molecule acts individually in creating the light waves, and accordingly the resulting radiation is of a random nature. That is, ordinary light consists of many separate components each polarized at some arbitrary angle. The overall effect of such combinations shows none of the properties of a simple polarized wave. Such light is therefore referred to as unpolarized unless it has been specially treated by some selective process which removes all components that do not lie in some particular plane. It is by this selecting or filtering action that polarized light is produced. Polarized radio waves, on the other hand, result from the nature of the source or antenna which produces them.

Elliptically polarized waves result from the superposition of two plane-polarized wave systems which are propagated in the same direction and have the same wavelength and frequency. The two waves are, in general, of unequal intensity and differ by 90° in the plane of their polarization. The electric components of the two waves shown

in Fig. 4·3 are characterized by the equations:

$$E_y = E'_y \sin(\omega t - \beta x) \qquad [4\cdot18]$$

$$E_z = E'_z \cos(\omega t - \beta x) \qquad [4\cdot19]$$

It is seen that one wave is equal to zero at any given time and place where the other wave is a maximum. These two waves may be considered as the separate components of a wave whose electric intensity $\mathbf{E} = \mathbf{j}E_y + \mathbf{k}E_z$ At any given place, as time is varied, the vector which characterizes the electric intensity describes an ellipse as shown

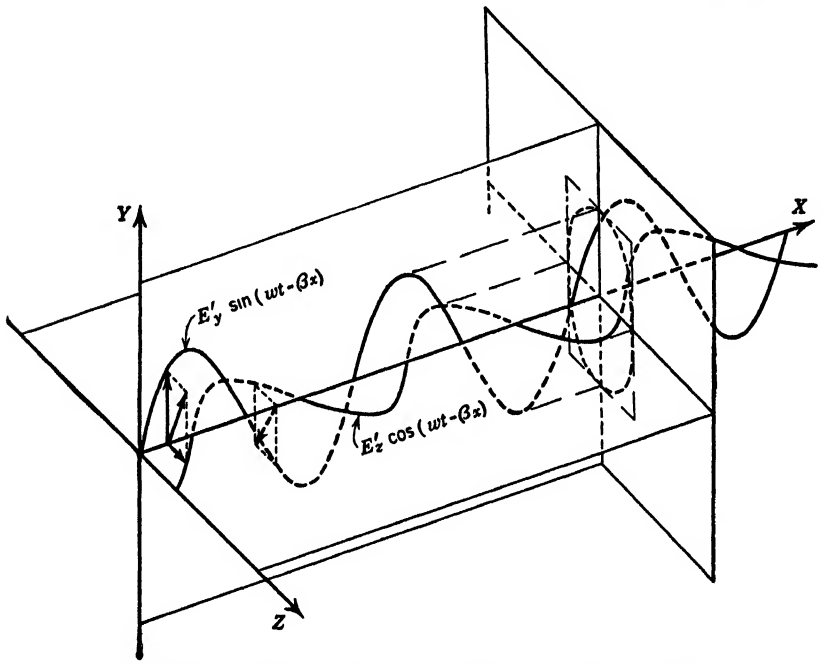


FIG. 4 3 Analysis of an elliptically polarized wave into component plane-polarized waves.

in Fig. 4·3. At any given instant of time the vector which characterizes the total electric intensity describes an elliptical helix throughout space.

The magnetic vector which is associated with the sum of the two elementary waves is everywhere at right angles to the resultant electric vector and is in time phase with it.

Circularly polarized waves are a special case of the elliptically polarized wave in which $|E'_y| = |E'_z|$.* Here the electric vector at a given

* The double bars around a letter refer to its absolute value.

place describes a circle during the course of one cycle. At a given instant the electric and magnetic vectors describe the form of regular helixes, the two being everywhere 90° apart.

4·3 Boundary between Two Dielectric Materials

Let us consider a plane surface separating two unlike media, as for instance a surface between air and glass. Since the dielectric constant of one material is different from that of the other, it is clear that the same equations cannot describe the wave on the two sides of the boundary. Accordingly, an additional expression representing a reflected wave must be added. It can be shown that three distinct wave systems exist at every boundary of media. In the familiar example of total internal reflection, characteristic of binocular prisms and lucite light guides, a wave of extremely small geometrical extent exists in the medium of lower dielectric constant. It will appear later that this is essentially a standing wave system involving no transmission of energy. The three wave systems are generally referred to as the incident, the reflected, and the refracted waves or rays. It is our problem to find relations between the amplitude and the phase of these three waves which satisfy the general boundary conditions stated above.

The simplest, and a very important, example of reflection occurs when a plane electromagnetic wave meets a boundary between two materials of different dielectric constant. It will be assumed that both media have negligible conductivity and the same permeability.

In order to simplify the equations, and without loss of generality, we choose the incident ray in the XOZ plane of Fig. 4·4 and let the XOY plane be the boundary surface. The incident wave may be characterized by the exponential form:

$$e^{j\omega\left(t-\frac{s}{v_1}\right)} = e^{j\omega\left(t-\frac{lx+my+nz}{v_1}\right)} \qquad [4\cdot20]$$

where l, m, and n are the direction cosines described previously. However, $m = 0$ by our choice of axes, and at the boundary $z = 0$. Therefore at the boundary the function reduces to

$$e^{j\omega\left(t-\frac{lx}{v_1}\right)} \qquad [4\cdot21]$$

Evidently this wave together with the reflected and refracted waves must satisfy the boundary conditions.

Let us assume first a refracted form of the wave

$$e^{j\omega_2\left(t-\frac{l_2x+m_2y+n_2z}{v_2}\right)} \qquad [4\cdot22]$$

This must have the same form as 4·21 at $z = 0$ if it is a solution to the problem. The two can be identical for all values of x, y, and t only if $\omega_2 = \omega$, $l_2/v_2 = l/v_1$, and $m_2 = 0$. These relatively simple equalities contain a wealth of information. The statement $\omega_2 = \omega$ tells us that the frequency is not changed at the boundary, a familiar and reasonable result. The fact that $m_2 = 0$ is evidence that the refracted wave lies in the same plane as the incident wave. Finally the equation $l_2/v_2 = l/v_1$ is the familiar sine law of refraction. This becomes more evident if we rearrange the equation to the form $v_1/v_2 = l/l_2$. In terms of the angles i and r of Fig. 4·4

$$l = \cos(90° - i) = \sin i$$

and

$$l_2 = \cos(90° - r) = \sin r$$

Moreover, v_1/v_2 is the index of refraction N of the upper material with respect to the lower.* We may therefore write

$$N = \frac{v_1}{v_2} = \frac{l}{l_2} = \frac{\sin i}{\sin r} \qquad [4·23]$$

Thus through the wave equations we have deduced the standard law of refraction ordinarily developed from purely optical considerations.

The wave reflected at the boundary may be treated in much the same way. The incident wave function at the boundary is again of the form

$$e^{j\omega\left(t-\frac{lx}{v_1}\right)} \qquad [4·21]$$

Since the medium is the same for the reflected as for the incident wave we know that both waves have the same velocity v_1. The reflected wave function is assumed to be of the form

$$e^{j\omega_1\left(t-\frac{l_1x+m_1y+n_1z}{v_1}\right)} \qquad [4·24]$$

which reduces to the form of 4·21 for $z = 0$ only if $\omega_1 = \omega$, $m_1 = 0$, and $l_1 = l$. That is, the direction cosine of the reflected ray is equal to that of the incident ray. This is possible under the conditions of Fig. 4·4, since the cosine of a positive angle is equal to the cosine of the equal negative angle.

We have then established from the boundary conditions the fact that the incident, refracted, and reflected rays will lie in the same plane, a plane perpendicular to the boundary. In this plane the reflected and incident rays make equal angles with the normal. The refracted ray obeys the well-known sine law.

* The index of refraction N is given here as a number greater than unity. It is the ratio of the velocities of the wave in the two media.

4·4 Reflection at Normal Incidence

The above treatment, though perfectly valid, is incomplete in that it does not deal directly with either the polarization or the intensity of the various waves. These will now be developed, first for the special condition of normal incidence, later for a more general problem. Again the treatment hinges entirely upon the boundary conditions and the wave equations, and again the results are the same as those of the optical development.

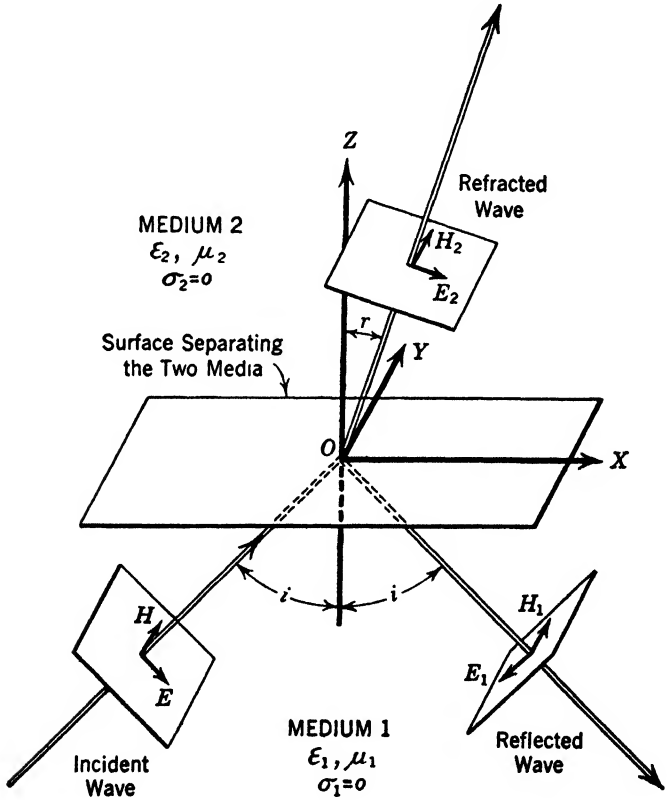


FIG. 4·4 Reflection and refraction.

Let us consider a case similar to the one drawn in Fig. 4·4 but differing in that the incident ray is normal to the XOY boundary plane. Let us choose E, the electric component of the incident wave, in the X direction, and H, the magnetic component of the incident wave, in the Y direction. Assume a refracted wave with the electric component E_2 in the X direction and the magnetic component H_2 in the Y direction. Finally assume a reflected wave with the electric component E_1

reversed or in the negative X direction and a magnetic component H_1 again in the Y direction. Reversal of a single component suffices to reverse the direction of the Poynting vector of energy flow. The permeability of both materials is taken as equal to that of free space. Accordingly, the index of refraction

$$N = \frac{v_1}{v_2} = \sqrt{\frac{\varepsilon_2\mu_2}{\varepsilon_1\mu_1}} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}} \qquad [4\cdot25]^*$$

The inversion of the last fraction results from the fact that a larger velocity is associated with a smaller dielectric constant.

It is now possible to apply the boundary conditions previously discussed. From our choice of wave direction all normal components of field are everywhere zero, and it is unnecessary to consider their continuity. Continuity of the tangential components of E and H gives us

$$E - E_1 = E_2 \qquad [4\cdot26]$$

$$H + H_1 = H_2 \qquad [4\cdot27]$$

Moreover, in any medium the numerical relation $H = \sqrt{\varepsilon/\mu}\,E$ holds. Accordingly

$$\frac{H}{E} = \sqrt{\frac{\varepsilon_1}{\mu_1}} = \frac{H_1}{E_1} \quad \text{and} \quad \frac{H_2}{E_2} = \sqrt{\frac{\varepsilon_2}{\mu_2}} \qquad [4\cdot28]$$

From 4·25, $\sqrt{\varepsilon_2} = N\sqrt{\varepsilon_1}$. Substituting this value for ε_2 in 4·28 and replacing μ_2 by its equal μ_1 we obtain

$$\frac{H_2}{E_2} = \frac{N\sqrt{\varepsilon_1}}{\sqrt{\mu_1}} = N\frac{H_1}{E_1} = N\frac{H}{E} \qquad [4\cdot29]$$

Dividing through by N we have

$$\frac{H}{E} = \frac{H_1}{E_1} = \frac{H_2}{NE_2} \qquad [4\cdot30]$$

Hence

$$H_1 = \frac{E_1}{E}H \qquad [4\cdot31]$$

and

$$H_2 = \frac{NE_2}{E}H \qquad [4\,32]$$

* When dealing with media in which $\mu_1 = \mu_2$. No dielectrics are known in which the permeability differs appreciably from that of free space.

Substituting these values of H_1 and H_2 into equation 4·27 we obtain

$$H + \frac{E_1}{E} H = \frac{NE_2}{E} H \qquad [4·33]$$

or, dividing through by H,

$$E + E_1 = NE_2 \qquad [4·34]$$

Eliminating E_2 between equations 4·34 and 4·26 we have

$$E_1 = \frac{N-1}{N+1} E \qquad \text{or} \qquad \frac{E_1}{E} = \frac{N-1}{N+1} \qquad [4·35]$$

But E_1/E is the reflection coefficient at normal incidence. The ratio of intensities or powers is equal to the square of the ratio of the electric vectors. Hence

$$\frac{P_1}{P} = \left(\frac{N-1}{N+1}\right)^2 \qquad [4·36]$$

As the index approaches unity, corresponding to no discontinuity of medium, the reflection coefficient approaches zero. If the index is very large, the reflection coefficient approaches unity but can never be exactly one.

An example may prove interesting. Consider the reflection of a wave traveling from air into glass having an index of refraction of 1.5 ($\varepsilon_1 = \varepsilon_0$, $\varepsilon_2 = 2.25\varepsilon_0$).* The power in the reflected wave is $(0.5/2.5)^2 = 0.04$ of the power in the incident wave. Conservation of energy demands that the remaining 96 per cent of the incident power be transmitted in the refracted wave. That such is the fact is readily shown by means of Poynting's vector and the values of E_2 and H_2 developed above. The proof is suggested as an exercise for the student, the algebra involved being relatively simple.

4·5 Fresnel's Equations†

We shall now consider the general case of reflection and refraction at a boundary between two dielectric materials. The equations just developed are simply special cases of this very general problem. The most significant factor that has not yet been considered is polarization. Not only is the angle between the incident ray and the normal arbitrary,

* It will be observed that the value of $\kappa_e = 2.25$ is lower than the minimum given for glass in Table 1·1, page 3. The values in Table 1·1 are derived from radio-frequency measurements.

† The name is pronounced Frĕ nĕl′ as if the *s* were absent.

but also the electric vector may have any angle with respect to the plane of the ray.

In Fig. 4·4 all the electric vectors lie in the XOZ plane, and the magnetic vectors are parallel to the Y axis. In Fig. 4·5 the magnetic vectors are shown in the XOZ plane. Consideration of these two figures shows that an incident wave having its electric vector in any

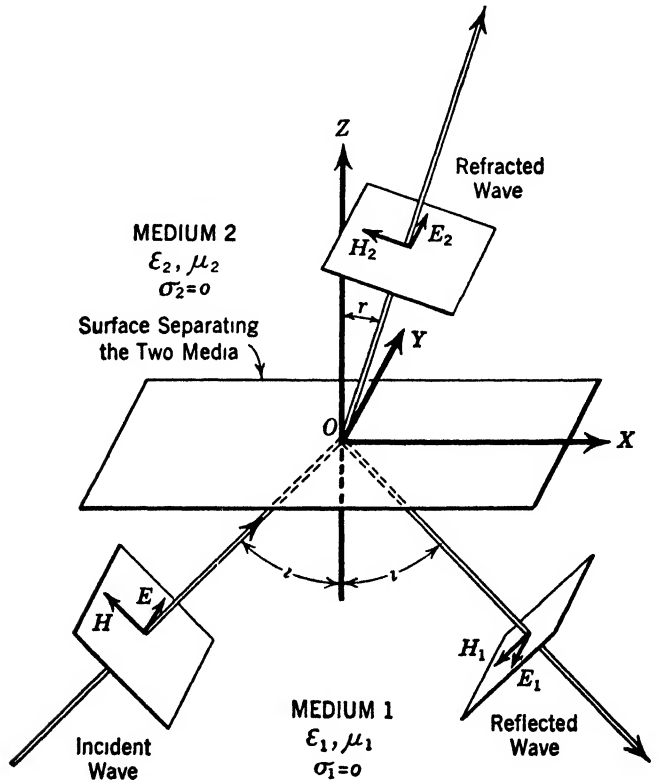


FIG 4 5 Reflection and refraction.

direction may be resolved into the two components shown. Accordingly, a solution of these two cases is a complete solution to the problem.

Let us consider first the case of Fig. 4·4. The magnetic vectors are entirely in the Y direction, and therefore H_x, H_{1x}, and H_{2x} are all zero. We may write from the basic wave equations

$$H_y = \sqrt{\frac{\varepsilon_1}{\mu_1}}\,E; \quad H_{1y} = \sqrt{\frac{\varepsilon_1}{\mu_1}}\,E_1; \quad \text{and} \quad H_{2y} = \sqrt{\frac{\varepsilon_2}{\mu_2}}\,E_2 \qquad [4\text{·}37]$$

Moreover,

$$\left.\begin{aligned} E_x &= E\cos i & E_z &= -E\sin i \\ E_{1x} &= -E_1\cos i & E_{1z} &= -E_1\sin i \\ E_{2x} &= E_2\cos r & E_{2z} &= -E_2\sin r \end{aligned}\right\} \quad [4\cdot38]$$

The continuity of the normal component of D gives us

$$D_z + D_{1z} = D_{2z} \quad [4\cdot39]$$

or

$$\varepsilon_1(E + E_1)\sin i = \varepsilon_2 E_2 \sin r \quad [4\cdot40]$$

The continuity of the tangential component of E gives us

$$E_x + E_{1x} = E_{2x} \quad [4\cdot41]$$

or

$$(E - E_1)\cos i = E_2 \cos r \quad [4\cdot42]$$

The continuity of the tangential component of H gives us

$$H_y + H_{1y} = H_{2y} \quad [4\cdot43]$$

Substituting for H_y, H_{1y}, and H_{2y} from equations 4·37

$$\sqrt{\frac{\varepsilon_1}{\mu_1}}E + \sqrt{\frac{\varepsilon_1}{\mu_1}}E_1 = \sqrt{\frac{\varepsilon_2}{\mu_2}}E_2 \quad [4\cdot44]$$

which may be written, when $\mu_2 = \mu_1$

$$E + E_1 = NE_2 \quad [4\cdot45]$$

This last equation may be combined with 4·40 to verify the sine law of refraction

$$N = \frac{\sin i}{\sin r} \quad [4\cdot46]$$

Equation 4·42 may be written

$$E - E_1 = E_2\frac{\cos r}{\cos i} \quad [4\cdot47]$$

Eliminating E_2 between 4·45 and 4·47

$$\frac{1}{N}(E + E_1) = E\frac{\cos i}{\cos r} - E_1\frac{\cos i}{\cos r} \quad [4\cdot48]$$

Separating variables,

$$E\frac{N\cos i - \cos r}{N\cos r} = E_1\frac{\cos r + N\cos i}{N\cos r} \quad [4\cdot49]$$

But

$$N = \frac{\sin i}{\sin r} \qquad [4\cdot50]$$

so that

$$\frac{E_1}{E} = \frac{N\cos i - \cos r}{N\cos i + \cos r} = \frac{\sin i\cos i - \sin r\cos r}{\sin i\cos i + \sin r\cos r} \qquad [4\cdot51]$$

With the aid of the trigonometric relations

$$\sin\theta\cos\theta = \tfrac{1}{2}\sin 2\theta \text{ and } \sin A \pm \sin B = 2\sin\tfrac{1}{2}(A \pm B)\cos\tfrac{1}{2}(A \mp B) \qquad [4\cdot52]$$

equation 4.51 may be written

$$\frac{E_1}{E} = \frac{\sin 2i - \sin 2r}{\sin 2i + \sin 2r} = \frac{\sin(i-r)\cos(i+r)}{\sin(i+r)\cos(i-r)} = \frac{\tan(i-r)}{\tan(i+r)} \qquad [4\cdot53]$$

This is one of Fresnel's equations. Reduction of this equation in terms only of the index and the angle of incidence is possible. Unfortunately the form is too complicated to be very useful.

The alternative case of Fig. 4.5 will now be considered in a parallel development. The electric vectors are all parallel to the Y axis. Accordingly we write

$$E = E_y, E_1 = -E_{1y}, E_2 = E_{2y} \qquad [4\cdot54]$$

and

$$\left.\begin{aligned} H_x &= -H\cos i & H_z &= H\sin i \\ H_{1x} &= -H_1\cos i & H_{1z} &= -H_1\sin i \\ H_{2x} &= -H_2\cos r & H_{2z} &= H_2\sin r \end{aligned}\right\} \qquad [4\cdot55]$$

The magnetic components may be written in terms of the electric components since $H = \sqrt{\varepsilon_1/\mu_1}E$, $H_1 = \sqrt{\varepsilon_1/\mu_1}E_1$, and $H_2 = \sqrt{\varepsilon_2/\mu_2}E_2$. Thus

$$\left.\begin{aligned} H_x &= -\sqrt{\frac{\varepsilon_1}{\mu_1}}E\cos i & H_z &= \sqrt{\frac{\varepsilon_1}{\mu_1}}E\sin i \\ H_{1x} &= -\sqrt{\frac{\varepsilon_1}{\mu_1}}E_1\cos i & H_{1z} &= -\sqrt{\frac{\varepsilon_1}{\mu_1}}E_1\sin i \\ H_{2x} &= -\sqrt{\frac{\varepsilon_2}{\mu_2}}E_2\cos r & H_{2z} &= \sqrt{\frac{\varepsilon_2}{\mu_2}}E_2\sin r \end{aligned}\right\} \qquad [4\cdot56]$$

The continuity of the tangential component of E gives

$$E_2 = E - E_1 \qquad [4\cdot57]$$

The continuity of the tangential component of H gives

$$H_{2x} = H_x + H_{1x} \qquad [4\cdot58]$$

Substituting for H_x, H_{1x} and H_{2x} as given in equation 4·56, we may write

$$\sqrt{\frac{\varepsilon_2}{\mu_2}}\, E_2 \cos r = \sqrt{\frac{\varepsilon_1}{\mu_1}}\, E \cos i + \sqrt{\frac{\varepsilon_1}{\mu_1}}\, E_1 \cos i \qquad [4\cdot59]$$

or, since $\mu_2 = \mu_1$,

$$\sqrt{\varepsilon_2} E_2 \cos r = \sqrt{\varepsilon_1} E \cos i + \sqrt{\varepsilon_1} E_1 \cos i \qquad [4\cdot60]$$

The continuity of the normal component of B is equivalent to continuity of the normal component of H since $\mu_1 = \mu_2$ throughout.

$$H_{2z} = H_z + H_{1z} \qquad [4\cdot61]$$

Substituting for H_z, H_{1z}, and H_{2z} as given in equation 4·56, this may be written

$$\sqrt{\frac{\varepsilon_2}{\mu_2}}\, E_2 \sin r = \sqrt{\frac{\varepsilon_1}{\mu_1}}\, E \sin i - \sqrt{\frac{\varepsilon_1}{\mu_1}}\, E_1 \sin i \qquad [4\cdot62]$$

or

$$\sqrt{\varepsilon_2} E_2 \sin r = \sqrt{\varepsilon_1} E \sin i - \sqrt{\varepsilon_1} E_1 \sin i \qquad [4\cdot63]$$

From 4·25, $N = \sqrt{\varepsilon_2/\varepsilon_1}$. Hence 4·60 and 4·63 become

$$NE_2 \cos r = E \cos i + E_1 \cos i = (E + E_1) \cos i \qquad [4\cdot64]$$

$$NE_2 \sin r = E \sin i - E_1 \sin i = (E - E_1) \sin i \qquad [4\cdot65]$$

Equations 4·57 and 4·65 again give us $N = \dfrac{\sin i}{\sin r}$. Eliminating E_2 from equations 4·64 and 4·65, we have

$$(E + E_1) \cos i \sin r = (E - E_1) \sin i \cos r \qquad [4\cdot66]$$

$$E_1 (\cos i \sin r + \sin i \cos r) = E (\sin i \cos r - \cos i \sin r) \qquad [4\cdot67]$$

$$\frac{E_1}{E} = \frac{\sin i \cos r - \cos i \sin r}{\sin i \cos r + \cos i \sin r} = \frac{\sin (i - r)}{\sin (i + r)} \qquad [4\cdot68]$$

This is another of Fresnel's equations. The expressions of 4·53 and 4·68,

$$\frac{\sin (i - r)}{\sin (i + r)} \quad \text{and} \quad \frac{\tan (i - r)}{\tan (i + r)} \qquad [4\cdot69]$$

approach each other as i and r become small, i.e., as normal incidence is approached. Moreover, $\dfrac{\sin i}{\sin r} = N$ approaches $\dfrac{i}{r} = N$ for small values

of i and r. Substituting $i = Nr$ into Fresnel's equations, we see that

$$\frac{\sin (Nr - r)}{\sin (Nr + r)} \simeq \frac{\tan (Nr - r)}{\tan (Nr + r)} \simeq \frac{Nr - r}{Nr + r} = \frac{N - 1}{N + 1} \text{ (for small } i \text{ and } r) \quad [4\cdot70]$$

the value previously deduced from equation 4·35 at normal incidence.

4·6 Polarization by Reflection

The equation for reflection applying to Fig. 4·4 was shown to be

$$\frac{E_1}{E} = \frac{\tan (i - r)}{\tan (i + r)} \quad [4\cdot53]$$

When $(i + r) = 90°$, the denominator of the right member becomes infinite and E_1 becomes zero. Hence there is no reflected wave. Inspection of the geometry indicates that such a condition is readily possible, the angles being $i = 56°$ and $r = 34°$ for ordinary glass.* The equation for reflection applying to Fig. 4·5 was shown to be

$$\frac{E_1}{E} = \frac{\sin (i - r)}{\sin (i + r)} \quad [4\cdot68]$$

This expression is finite for $i + r = 90°$, and accordingly a finite reflection obtains. It is seen, therefore, that, under one condition of polarization, reflection occurs, whereas under another it does not. Since an unpolarized wave may be expressed in terms of the two cases treated, it is evident that a method for obtaining a polarized wave from an unpolarized one is available. Thus, when an unpolarized wave impinges upon a glass surface at an angle of incidence of 56°, only the components of E and H in the directions indicated in Fig. 4·5 are reflected. This reflected wave is polarized as indicated by the reflected wave E_1, H_1 in this figure.

We may also detect the plane of polarization of a wave by means of this phenomenon. By rotating the glass, while maintaining the proper angle of incidence, it is clear that a maximum energy in the reflected wave will occur when the vectors of the polarized wave are directed as indicated in Fig. 4·5. Conversely, a null method may be employed in which the incident vectors correspond to those of Fig. 4·4. Consequently it is possible to produce completely plane-polarized waves by this means or to use such a device as a detector of polarization. The only restrictions on the dielectric material employed are that it have a low conductivity and be sufficiently thick. Also, the use of a blackened

* These values of the angles i and r are obtained from the simultaneous solution of the equations $i + r = 90°$ and $\frac{\sin i}{\sin r} = N = 1.5$ for glass.

glass having a low coefficient of transmission is desirable in order to avoid difficulty with reflections from back surfaces. This critical angle of incidence is known as Brewster's angle.

4·7 Total Internal Reflection

Total internal reflection is of interest to the electrical engineer because it forms the basis of the dielectric wave guide. The term internal reflection refers to the fact that the wave is traveling from the medium having the larger to the medium having the smaller index of refraction. See Fig. 4·6. Under these conditions $N = \sqrt{\varepsilon_1/\varepsilon_2}$ is always greater than unity. The term total refers to the fact that under suitable conditions no power is transmitted to the second medium.

In this case* we must write

$$N = \frac{v_2}{v_1} = \frac{\sin r}{\sin i} \qquad \text{or} \qquad N \sin i = \sin r \tag{4·71}$$

Since $N > 1$, $\sin r$ is always greater than $\sin i$. When the angle of incidence has such a value that $N \sin i = 1$, it is evident from equation 4·71 that $\sin r = 1$ also. This corresponds to $r = 90°$, which means that the refracted wave grazes the surface separating the two media. For values of i such that $N \sin i$ is greater than 1, equation 4·71 requires that $\sin r$ be greater by the same amount. This, of course, requires imaginary values of r. Physically this condition corresponds to complete reflection of the incident wave in medium 1. It must not be assumed, however, that no wave exists in the second, less dense, medium. Actually a rather special form of guided wave exists at the boundary and is propagated along the boundary plane. The situation may be treated by mathematical methods similar to those already employed.

Let the incident ray be characterized by the function

$$e^{j\omega\left(t-\frac{lx+my+nz}{v_1}\right)} \tag{4·72}$$

where v_1 is the velocity in medium 1, which has the higher index of refraction. From our choice of axes, Fig. 4·6, we may write $m = 0$, $l = \sin i$, and $n = \cos i$.

Accordingly the function may be written

$$e^{j\omega\left(t-\frac{x\sin i+z\cos i}{v_1}\right)} \tag{4·73}$$

Similarly the refracted wave may be written

$$e^{j\omega\left(t-\frac{x\sin r+z\cos r}{v_2}\right)} \tag{4·74}$$

* See footnote on page 104.

where, from equation 4·71,

$$v_2 = Nv_1 \qquad [4\cdot71]$$

is the velocity in the upper medium. This form is correct for any angle of incidence. From the relations

$$\cos r = \sqrt{1 - \sin^2 r} \qquad \text{and} \qquad \sin r = N \sin i$$

we have

$$\cos r = \sqrt{1 - N^2 \sin^2 i} \quad = \quad j\sqrt{N^2 \sin^2 i - 1} \qquad [4\cdot75]$$

depending upon whether $N \sin i$ is less or greater than unity. If it is greater, corresponding to total reflection, the angle is a pure imaginary.

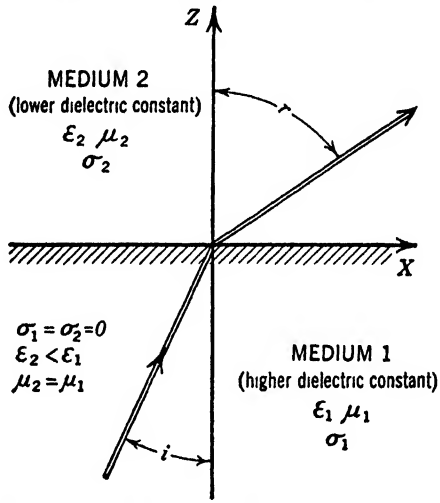


FIG. 4·6 Transmission from medium of higher dielectric constant to one of lower dielectric constant.

Now the relation 4·74 representing the refracted wave may be written

$$e^{j\omega\left(t - \frac{x \sin r}{v_2}\right)} e^{j\omega\left(\frac{z \cos r}{v_2}\right)} \qquad [4\cdot76]$$

Replacing cos r by its value given in equation 4·75 for total reflection we may write

$$e^{j\omega\left(t - \frac{x \sin r}{v_2}\right)} e^{-\frac{\omega\sqrt{N^2 \sin^2 i - 1}}{v_2} z} \qquad [4\cdot77]$$

Such a function represents a wave propagated in the $+x$ direction with a velocity $v_2/\sin r$ and attenuated in the $+z$ direction by the exponential term

$$e^{-\frac{\omega\sqrt{N^2 \sin^2 i - 1}}{v_2} z} \qquad [4\cdot78]$$

Since $N \sin i > 1$ is the fundamental assumption, the velocity in the x direction is always less than v_2. The attenuation in the z direction is high. The wave falls to $1/e$ of its value at the boundary for

$$z = \frac{v_2}{\omega \sqrt{N^2 \sin^2 i - 1}} \qquad [4{\cdot}79]$$

If we let $N \sin i = 1{\cdot}41$, a reasonable assumption for glass of value $N = 1.5$, the radical reduces to unity. Also we have as identities

$$v_2 = f\lambda_2 \qquad \text{or} \qquad 2\pi v_2 = \omega\lambda_2 \qquad [4{\cdot}80]$$

Accordingly the condition fo attenuation to $1/e$ of the value at the surface of the glass is

$$z = \frac{\lambda_2 v_2}{2\pi v_2} = \frac{\lambda_2}{2\pi} \qquad [4{\cdot}81]$$

The physical existence of such waves has been demonstrated. Less than total reflection results if any absorptive material comes within a distance comparable to λ_2 of the boundary. In fact, the trouble with dirt upon the prisms of binoculars is a familiar example. The specks of dirt absorb energy from the guided wave at the boundary and so impair the reflection.

It may be observed from equation 4·77 that the planes of constant phase are parallel to the YOZ plane (x = constant) and that the planes of constant amplitude are parallel to the boundary (z = constant). It may readily be shown that Poynting's vector for such a wave has no component normal to the surface, and therefore no energy is propagated in that direction.

An expression for the energy in the reflected wave may be developed by means of Fresnel's equations. It is only through this development that the total reflection of energy is established. In the first case treated corresponding to Fig. 4·4 we showed that

$$\frac{E_1}{E} = \frac{\sin i \cos i - \sin r \cos r}{\sin i \cos i + \sin r \cos r} = \frac{\sin i \cos i - jN \sin i \sqrt{N^2 \sin^2 i - 1}}{\sin i \cos i + jN \sin i \sqrt{N^2 \sin^2 i - 1}} \qquad [4{\cdot}82]$$

Since numerator and denominator are conjugate complex numbers, it is evident that E_1 and E are numerically equal. A phase shift exists, however, which is equal to

$$2 \tan^{-1} \left(\frac{N \sqrt{N^2 \sin^2 i - 1}}{\cos i} \right) \qquad [4{\cdot}83]$$

In the other case, corresponding to Fig. 4·5

$$\frac{E_1}{E} = \frac{\sin \imath \cos r - \cos i \sin r}{\sin i \cos r + \cos \imath \sin r} \qquad [4\cdot68]$$

$$= \frac{j \sin \imath \sqrt{N^2 \sin^2 \imath - 1} - N \cos i \sin i}{j \sin i \sqrt{N^2 \sin^2 \imath - 1} + N \cos i \sin i} \qquad [4\cdot84]$$

or

$$\frac{E_1}{E} = -\frac{N \cos i - j\sqrt{N^2 \sin^2 i - 1}}{N \cos \imath + j\sqrt{N^2 \sin^2 \imath - 1}} \qquad [4\cdot85]$$

Again numerator and denominator are conjugate, and E_1 and E are numerically equal. Thus the statement of total reflection of energy is verified in both cases. In this latter case a different phase shift equal to

$$2 \tan^{-1}\left(\frac{\sqrt{N^2 \sin^2 \imath - 1}}{N \cos i}\right) \qquad [4\cdot86]$$

is indicated.

Several important wave phenomena result from this difference of phase shift. One of the most interesting is the formation of elliptically polarized waves from plane-polarized waves. If a beam of plane-

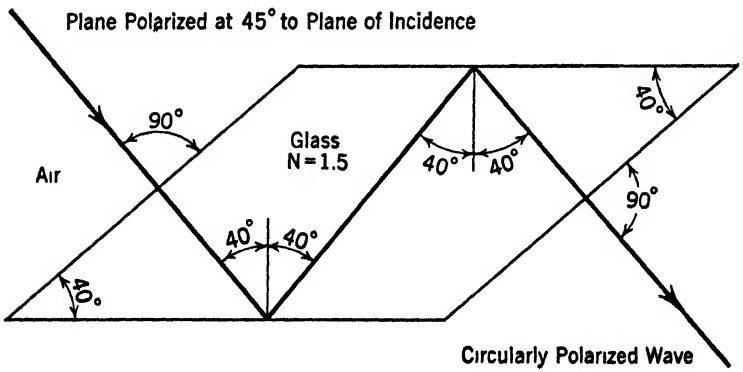


FIG. 4·7 Fresnel's rhomb.

polarized light having components corresponding to Figs. 4·4 and 4·5 is totally reflected, the two plane components will have a phase difference. In such a wave the electric vector never falls to zero at any place or at any time. Rather it acts as a rotating vector of variable amplitude. By means of two successive internal reflections it is possible to establish circularly polarized light in which the rotating vector is of constant

amplitude. Figure 4·7 shows the so-called Fresnel rhomb with constants suitable for the production of circularly polarized from plane-polarized light. Evidently the reverse function is also satisfied in that circularly polarized light is converted into plane polarized. This form of reciprocity is true in general of all such cases. That is, if elliptically polarized light is produced from plane-polarized, then plane-polarized light will result from the introduction of a suitable elliptically polarized beam.

PROBLEMS

4·1 A slab of material is to be prepared which will reflect 10 per cent of the incident wave power (light) and absorb the remainder, transmitting a negligible power from its rear face. What properties should the material have?

4·2 Show that the tangential component of H is not continuous at the surface of a conductor if a finite current flows on the surface of that conductor.

4·3 Show analytically how two plane waves traveling in the same direction in free space may be combined to form an elliptically polarized wave. Show the phase relation and relative amplitudes and polarization required.

4·4 A plane wave falls normally upon the surface of a dielectric slab. Calculate the dielectric constant and index of refraction required if the reflection coefficient on a power basis is 5, 10, 20, and 50 per cent. (Take $\mu_m = 1$.)

4·5 A material has a dielectric constant of 4 and a permeability of 1. A plane wave travels normally from free space into this material. Calculate the relative field strengths E and H and the power in the transmitted and reflected waves relative to the incident wave.

4·6 Repeat problem 4·5 for a material in which the dielectric constant is 100.

4·7 Referring to Fig. 4·4, the conductivities are both zero and both permeabilities are unity. The dielectric constant of the lower material is unity, and that of the upper material is 3. The angle of incidence i is 30°. Given E in the incident wave, calculate all other values of E and H. Also calculate the angle of refraction r and the powers in the reflected and refracted waves in terms of that in the original wave.

4·8 Repeat problem 4·7 for the orientation shown in Fig. 4·5.

4·9 Repeat problem 4·7 for a case in which the lower medium is free space and the upper medium has a permeability of 1 and a dielectric constant of 10.

4·10 Repeat problem 4·9 for the orientation of Fig. 4·5.

4·11 The electric vector of a wave has equal components in the directions indicated in Figs. 4·4 and 4·5. Repeat problem 4·7 for such a wave.

4·12 A plane wave traveling in free space makes an angle of incidence i of 45° with a material having a dielectric constant of 4 and a permeability of 1. The electric intensity E in the incident wave is rotated 30° from the position shown in Fig. 4·4 toward that shown in Fig. 4·5. Calculate all other values of E and H, the angle of refraction r, and the power in the reflected and refracted waves in terms of the incident wave.

4·13 A wave traveling in a medium of unit permeability and dielectric constant 9 strikes a plane boundary with free space. Given $i = 45°$, $\omega = 10^{10}$, and the orienta-

tion of Fig. 4·5, calculate the direction and magnitude of the electric field in free space at distances of 1, 10, and 100 cm from the bounding surface. Calculate the velocity of propagation of the guided wave in free space.

4·14 A wave traveling in a medium of unit permeability and dielectric constant 4 strikes a plane boundary with free space. Given $i = 60°$, $\omega = 10^{11}$, and the orientation of Fig. 4·4, calculate the magnitude of the electric field at distances of 0.2,2, and 20 cm from the bounding surface. Calculate the phase shift between the incident and reflected waves in the medium.

4·15 A Fresnel rhomb is to be designed for a material of unit permeability and dielectric constant 3. Choose the angles of the rhomb so that plane-polarized light is converted into circularly polarized light and so that light enters and leaves the faces normally.

4·16 Design a Fresnel rhomb for a material of permeability 1 and dielectric constant 9.

4·17 Non-polarized light is to be polarized by reflection from the surface of a material of dielectric constant 6 and permeability 1. Determine the angle (Brewster's angle) for this material.

4·18 Non-polarized light is to be polarized by reflection from the surface of a material of dielectric constant 2 and permeability 1. Determine Brewster's angle for this material.

4·19 Starting from the general form of equation 4·25, $N = \sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}}$, develop the equation which corresponds to equation 4·35.

4·20 Given $N = \sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}}$, consider Fig. 4·4, and develop the equation which corresponds to equation 4·46.

CHAPTER 5

PARALLEL PLANE WAVE GUIDES

5·1 Boundary Conditions at a Perfect Conductor

The material in this chapter is concerned with the propagation of electromagnetic waves that are guided by plane parallel conductors immersed in a homogeneous isotropic medium. Under such conditions, σ, ε, and μ are constant, $\mathbf{B} = \mu\mathbf{H}$, and $\mathbf{D} = \varepsilon\mathbf{E}$. Let us first consider that the guides are perfect conductors, that is, conductors which have zero resistance or infinite conductivity. Inside such a perfect conductor the electric intensity $\mathbf{E}$ vanishes everywhere, and electric charge can reside only on its surface. Actually, of course, no such conductor exists. However, much useful information can be obtained by making such a supposition, and, because of the simplifications which it introduces into the wave equations, it represents an excellent starting point. The conditions existing when the conductivity is finite are considered by an extension of the method.

In Chapter 2 it was shown that the tangential component of the electric intensity $\mathbf{E}$ is continuous at the boundary between two media. Therefore the existence of a component of electric intensity along the conducting surface would require the existence of electric intensity in the conducting material. But this is contrary to the definition of a perfect conductor. Hence the only electric intensity vectors that can exist at the surface of a perfect conductor must be normal to the surface.

5·2 Waves Guided between Conducting Parallel Planes

Let us consider the propagation of electromagnetic waves traveling in the X direction in a homogeneous isotropic medium which fills the space between two perfectly conducting sheets of metal of infinite extent. Let these sheets be parallel to the XOZ plane of our coordinate system. Let one of them cut the Y axis at the point $y = -y_0/2$ (Fig. 5·1), and the other at the point $y = y_0/2$, so that the distance between them is y_0.

For the plane wave in free space, the field equations show that the direction of propagation is at right angles to the plane of the wave. When we chose the YOZ plane as the plane of the wave, we found that

the direction of propagation was along the X axis, and that the **E** and **H** vectors were at right angles to each other and to the direction of propagation. In considering guided waves in this section it is our purpose to investigate propagation in the X direction. It does not follow that the wave front will be parallel to the YOZ plane, however, for in the Y direction we have definite boundary conditions to be met, and it cannot be expected that the intensity vectors will be constant in this direction at any instant of time. In the Z direction, how-

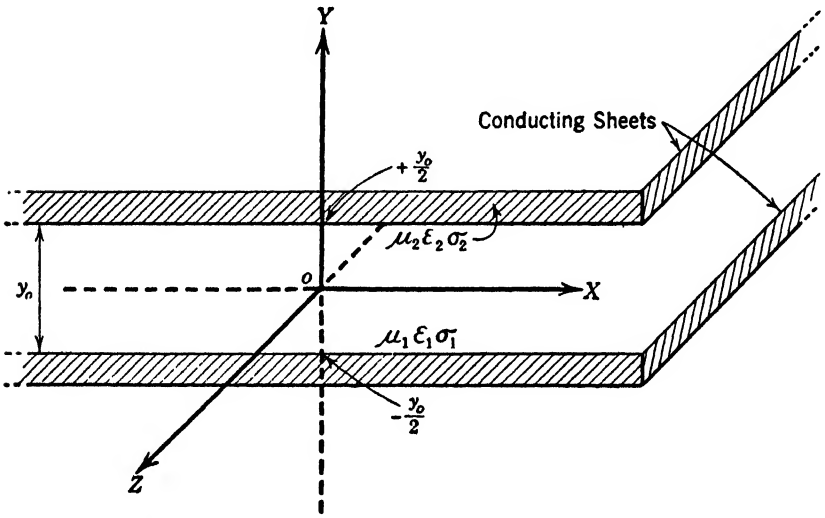


FIG. 5.1 Showing arrangement of axes between two conducting sheets of metal of infinite extent.

ever, we do not have boundaries, since our planes have been chosen of infinite extent. We can, therefore, consider that the electric and magnetic intensities at any instant of time are constant in this direction and therefore that their derivatives with respect to z are everywhere zero. With this restriction the wave equations:

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \sigma E_x + \varepsilon \frac{\partial E_x}{\partial t} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \sigma E_y + \varepsilon \frac{\partial E_y}{\partial t} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \sigma E_z + \varepsilon \frac{\partial E_z}{\partial t} \end{aligned}\right\} [5{\cdot}1] \qquad \left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu \frac{\partial H_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu \frac{\partial H_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu \frac{\partial H_z}{\partial t} \end{aligned}\right\} [5{\cdot}2]$$

become, when $\partial/\partial z = 0$,

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} &= \sigma E_x + \varepsilon \frac{\partial E_x}{\partial t} \\ -\frac{\partial H_z}{\partial x} &= \sigma E_y + \varepsilon \frac{\partial E_y}{\partial t} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \sigma E_z + \varepsilon \frac{\partial E_z}{\partial t} \end{aligned}\right\} [5\cdot3] \qquad \left.\begin{aligned} \frac{\partial E_z}{\partial y} &= -\mu \frac{\partial H_x}{\partial t} \\ -\frac{\partial E_z}{\partial x} &= -\mu \frac{\partial H_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu \frac{\partial H_z}{\partial t} \end{aligned}\right\} [5\cdot4]$$

Our problem now is to solve these partial differential equations subject to the boundary conditions imposed upon them.

Let us introduce at this point a further restriction, namely, that **E** and **H** involve time only through the factor $e^{j\omega t}$. This limitation is in no sense a narrow one, since we may take the real part or the imaginary part of the resulting solutions and obtain sinusoidal or cosinusoidal functions of time. These solutions will be particular solutions of the equations for any specified frequency $f = \omega/2\pi$. Since these equations are linear and homogeneous, the solutions for different frequencies are additive, and the complete solution is the sum of the particular solutions. Now any periodic function of time may be developed in a Fourier series of sines and cosines of harmonically related frequencies. Hence, by proper summation, our equations will apply wherever **E** and **H** are periodic functions of time, even though the requirement that time appear only in the factor $e^{j\omega t}$ is imposed. Accordingly, as far as time is concerned,

$$\mathbf{E} = \mathbf{E}'' e^{j\omega t} \qquad [5\cdot5]$$

and

$$\mathbf{H} = \mathbf{H}'' e^{j\omega t} \qquad [5\cdot6]$$

where $\mathbf{E}''$ and $\mathbf{H}''$ are the maximum values of the time variations of the intensity vectors. Substituting these values for **E** and **H** into equations 5·3 and 5·4, and dividing through by $e^{j\omega t}$, we obtain

$$\left.\begin{aligned} \frac{\partial H_z''}{\partial y} &= (\sigma + j\omega\varepsilon) E_x'' \\ -\frac{\partial H_z''}{\partial x} &= (\sigma + j\omega\varepsilon) E_y'' \\ \frac{\partial H_y''}{\partial x} - \frac{\partial H_x''}{\partial y} &= (\sigma + j\omega\varepsilon) E_z'' \end{aligned}\right\} [5\cdot7] \qquad \left.\begin{aligned} \frac{\partial E_z''}{\partial y} &= -j\omega\mu H_x'' \\ -\frac{\partial E_z''}{\partial x} &= -j\omega\mu H_y'' \\ \frac{\partial E_y''}{\partial x} - \frac{\partial E_x''}{\partial y} &= -j\omega\mu H_z'' \end{aligned}\right\} [5\cdot8]$$

We shall show that these equations represent two basically different modes of transmission. We may choose **E** in the Z direction and

obtain from these relations the components of the **H** vector, or we may choose **H** in the Z direction and obtain the components of the **E** vector.

5·3 Transmission Modes; E and H Waves

If we choose the electric intensity vector **E** in the Z direction, we obtain what is known as the transverse electric or TE mode of transmission. Waves satisfying this mode are also called H waves.* Under these conditions $E = E_z$, and $E_x = E_y = 0$. Substituting these values for E_z and E_y into equations 5·7 and 5·8, we obtain:

$$\left.\begin{aligned} \frac{\partial H_z''}{\partial y} &= 0 \qquad (a)\\ \frac{\partial H_z''}{\partial x} &= 0 \qquad (b)\\ \frac{\partial H_y''}{\partial x} - \frac{\partial H_x''}{\partial y} &= (\sigma + j\omega\epsilon)E_z'' \quad (c) \end{aligned}\right\} \quad [5\cdot9]$$

$$\left.\begin{aligned} \frac{\partial E_z''}{\partial y} &= -j\omega\mu H_x'' \qquad (a)\\ -\frac{\partial E_z''}{\partial x} &= -j\omega\mu H_y'' \qquad (b)\\ 0 &= -j\omega\mu H_z'' \qquad (c) \end{aligned}\right\} \quad [5\cdot10]$$

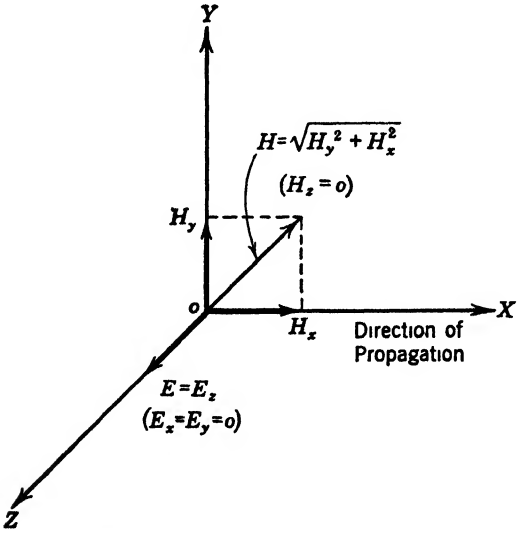


FIG. 5·2 Transverse electric or TE mode of transmission (H waves).

From 5·10c it is evident that $H_z'' = 0$. The remaining components with this type of transmission may be represented as shown in Fig. 5·2. It is evident that the **H** vector does not lie entirely in the YOZ plane.

If we choose the magnetic intensity vector **H** in the Z direction, that is, at right angles to the direction of propagation, we obtain what is known as the transverse magnetic or TM mode of transmission. Waves

* The designations TE and H are synonymous symbols.

satisfying this mode are also called E waves.* Under these conditions, $H = H_z$, and $H_x = H_y = 0$. Substituting these values for H_x and H_y into equations 5·7 and 5·8, we obtain:

$$\left.\begin{aligned} \frac{\partial H_z''}{\partial y} &= (\sigma + j\omega\varepsilon)E_x'' \quad (a)\\ -\frac{\partial H_z''}{\partial x} &= (\sigma + j\omega\varepsilon)E_y'' \quad (b)\\ 0 &= (\sigma + j\omega\varepsilon)E_z'' \quad (c) \end{aligned}\right\} \quad [5\,11]$$

$$\left.\begin{aligned} \frac{\partial E_z''}{\partial y} &= 0 \quad (a)\\ \frac{\partial E_z''}{\partial x} &= 0 \quad (b)\\ \frac{\partial E_y''}{\partial x} - \frac{\partial E_x''}{\partial y} &= -j\omega\mu H_z'' \quad (c) \end{aligned}\right\} \quad [5\cdot12]$$

From equation 5·11c it is evident that $E_z'' = 0$. The remaining components with this type of transmission may be represented as shown in Fig. 5·3. It is evident here, also, that the E vector does not lie entirely in the YOZ plane.

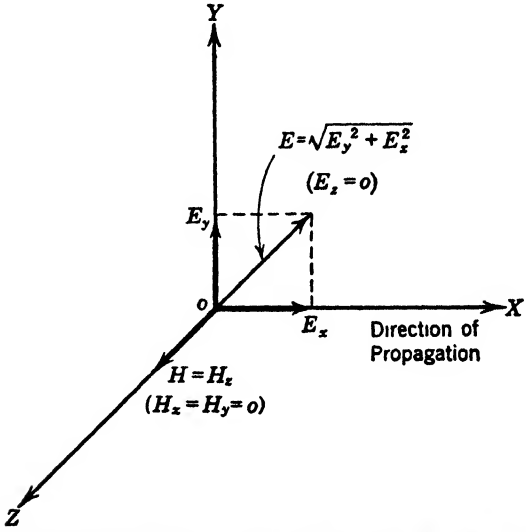


FIG. 5·3 Transverse magnetic or TM mode of transmission (E waves).

The equations defining these two modes may be collected as follows:

$$\left.\begin{aligned} \frac{\partial H_y''}{\partial x} - \frac{\partial H_x''}{\partial y} &= (\sigma + j\omega\varepsilon)E_z''\\ \frac{\partial E_z''}{\partial y} &= -j\omega\mu H_x''\\ -\frac{\partial E_z''}{\partial x} &= -j\omega\mu H_y''\\ E_z = E_y = H_z &= 0 \end{aligned}\right\} \quad \begin{matrix} TE \text{ or } H \\ \text{waves} \\ [5\cdot13] \end{matrix}$$

$$\left.\begin{aligned} \frac{\partial H_z''}{\partial y} &= (\sigma + j\omega\varepsilon)E_x''\\ -\frac{\partial H_z''}{\partial x} &= (\sigma + j\omega\varepsilon)E_y''\\ \frac{\partial E_y''}{\partial x} - \frac{\partial E_x''}{\partial y} &= -j\omega\mu H_z''\\ H_z = H_y = E_z &= 0 \end{aligned}\right\} \quad \begin{matrix} TM \text{ or } E \\ \text{waves} \\ [5\cdot14] \end{matrix}$$

* The designations TM and E are synonymous symbols and may be used interchangeably.

These equations for the two types of transmission modes are expressed in terms of the maximum values of the time variation of the **E** and **H** vectors. We will now investigate the properties of these vectors as they are propagated in the X direction. Let us assume that the functional relation between $\mathbf{E}''$ or $\mathbf{H}''$ and distance is given by

$$\mathbf{E}'' = \mathbf{E}'e^{-\gamma x} \qquad [5\cdot15]$$

and

$$\mathbf{H}'' = \mathbf{H}'e^{-\gamma x} \qquad [5\cdot16]$$

where $\mathbf{E}'$ and $\mathbf{H}'$ are the maximum values of the intensity vectors in distance and time, and γ is a constant whose value remains to be determined. This constant, called the propagation constant, will be discussed in detail later.

Thus from equations 5·15 and 5·16 the **E** and **H** vectors are now functionally related to distance and time by the expressions

$$\mathbf{E} = \mathbf{E}'e^{j\omega t}e^{-\gamma x} = \mathbf{E}'e^{-\gamma x+j\omega t} \qquad [5\cdot17]$$

and

$$\mathbf{H} = \mathbf{H}'e^{j\omega t}e^{-\gamma x} = \mathbf{H}'e^{-\gamma x+j\omega t} \qquad [5\cdot18]$$

Substituting the relations 5·15 and 5·16 into 5·13 and 5·14 and dividing through by $e^{-\gamma x}$ we obtain

$$\left.\begin{aligned} -\gamma H_y' - \frac{\partial H_x'}{\partial y} &= (\sigma + j\omega\varepsilon)E_z' \quad (a)\\ \frac{\partial E_z'}{\partial y} &= -j\omega\mu H_x' \quad (b)\\ \gamma E_z' &= -j\omega\mu H_y' \quad (c) \end{aligned}\right\} \text{TE or } H \text{ waves [5·19]}$$

$$\left.\begin{aligned} \frac{\partial H_z'}{\partial y} &= (\sigma + j\omega\varepsilon)E_x' \quad (a)\\ \gamma H_z' &= (\sigma + j\omega\varepsilon)E_y' \quad (b)\\ \gamma E_y' + \frac{\partial E_x'}{\partial y} &= j\omega\mu H_z' \quad (c) \end{aligned}\right\} TM \text{ or } E \text{ waves [5·20]}$$

5·4 Transmission Properties of *E* Waves

We may obtain from equations 5·20 an equation in H_z' alone. Differentiate 5·20a with respect to y. This gives

$$\frac{\partial^2 H_z'}{\partial y^2} = (\sigma + j\omega\varepsilon)\frac{\partial E_x'}{\partial y} \qquad [5\cdot21]$$

From 5·20c

$$\frac{\partial E_x'}{\partial y} = j\omega\mu H_z' - \gamma E_y' \qquad [5\cdot22]$$

Substituting for E_y' from 5·20b we have

$$\frac{\partial E_x'}{\partial y} = j\omega\mu H_z' - \frac{\gamma^2}{\sigma + j\omega\varepsilon}H_z' \qquad [5\cdot23]$$

Replacing $\dfrac{\partial E_x'}{\partial y}$ in equation 5·21 with its value as given in 5·23, and

rearranging terms, we obtain

$$\frac{\partial^2 H_z'}{\partial y^2} = -[\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)]H_z' \qquad [5\cdot24]$$

Let μ_1, ε_1, and σ_1 refer to the dielectric between the plates, and μ_2, ε_2, σ_2 to the conductor. Now in the dielectric $\sigma_1 \ll \omega\varepsilon_1$, so that 5·24 becomes

$$\frac{\partial^2 H_z'}{\partial y^2} = -(\gamma^2 + \omega^2\mu_1\varepsilon_1)H_z' \qquad [5\cdot25]$$

The solution of this equation is well known.* It may be written as

$$H_z' = C_1 \sin(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y + C_2 \cos(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y \qquad [5\cdot26]$$

where C_1 and C_2 are arbitrary constants depending upon the amplitude of the original excitation. We may obtain E_x' from equation 5·23. In the dielectric, neglecting σ_1 equation 5·23 may be written

$$\frac{\partial E_x'}{\partial y} = -\frac{(\gamma^2 + \omega^2\mu_1\varepsilon_1)}{j\omega\varepsilon_1} H_z' \qquad [5\cdot27]$$

Substituting for H_z' from 5.26 and integrating,

$$E_x' = \frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\varepsilon_1}\{C_1 \cos(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y - C_2 \sin(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y\} \qquad [5\cdot28]$$

where we have omitted the constant of integration since it is of no importance in wave propagation.

The E_y' component may be obtained from equation 5·20*b*. Thus, neglecting σ_1,

$$E_y' = \frac{\gamma}{j\omega\varepsilon_1} H_z' = \frac{\gamma}{j\omega\varepsilon_1}\{C_1 \sin(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y + C_2 \cos(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y\} \qquad [5\cdot29]$$

At the boundary $E_x' = 0$ and $y = \pm y_0/2$. Let us set

$$\frac{(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y_0}{2} = \frac{n\pi}{2}$$

where n is an integer, so that

$$\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = \frac{n\pi}{y_0} \qquad [5\cdot30]$$

* See, for example, Burington and Torrance, *Higher Mathematics*, McGraw-Hill Book Company, page 355, equations 1 and 2.

Then

$$E'_x = \frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\varepsilon_1}\left\{C_1 \cos\left(\frac{n\pi y}{y_0}\right) - C_2 \sin\left(\frac{n\pi y}{y_0}\right)\right\} \qquad [5\cdot31]$$

Now three distinct conditions arise, depending upon whether n is zero, odd, or even. Under these conditions C_1 and C_2 are alternately zero in order that the boundary conditions be satisfied. Thus, when $y = \pm y_0/2$ and n is even, i.e., $n = 2, 4, 6 \cdots$

$$\cos\left(\frac{n\pi}{y_0} \cdot \frac{y_0}{2}\right) = \cos\left(n\,\frac{\pi}{2}\right) = \pm 1 \qquad \text{and} \qquad \sin\left(n\,\frac{\pi}{2}\right) = 0$$

This will give

$$E'_x = \pm \frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\varepsilon_1} C_1 \qquad [5\cdot32]$$

But E'_x must be zero when $y = \pm y_0/2$. Therefore C_1 must be zero if n is even. Similarly, when n is odd, i.e., $n = 1, 3, 5, \cdots$

$$E'_x = \pm \frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\varepsilon_1} C_2 \qquad [5\cdot33]$$

when $y = \pm y_0/2$. Therefore C_2 must be zero when n is odd. Finally, when $n = 0$, E'_x is given by 5·32 at $y = \pm y_0/2$ and C_1 must be zero. Thus we have at

	$n = 0$,	$C_1 = 0$	
(E waves)	n odd,	$C_2 = 0$	[5·34]
	n even,	$C_1 = 0$	

We see, therefore, that we may have different E waves characterized by the given value of the integer n. We may refer to these as E_n waves in general or, specifically, as E_0, E_1, E_2, etc., waves, or as TM_0, TM_1, TM_2, etc., waves.

With the aid of 5·30 and subject to the restrictions of 5·34, we may write, since $\mathbf{H} = \mathbf{H}'e^{j\omega t - \gamma x}$ and $\mathbf{E} = \mathbf{E}'e^{j\omega t - \gamma x}$

$$\left.\begin{aligned}
H_z &= \left\{C_1 \sin\left(\frac{n\pi}{y_0} y\right) + C_2 \cos\left(\frac{n\pi}{y_0} y\right)\right\} e^{j\omega t - \gamma x} \\
E_x &= \frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\varepsilon_1}\left\{C_1 \cos\left(\frac{n\pi}{y_0} y\right) - C_2 \sin\left(\frac{n\pi}{y_0} y\right)\right\} e^{j\omega t - \gamma x} \\
E_y &= \frac{\gamma}{j\omega\varepsilon_1}\left\{C_1 \sin\left(\frac{n\pi}{y_0} y\right) + C_2 \cos\left(\frac{n\pi}{y_0} y\right)\right\} e^{j\omega t - \gamma x}
\end{aligned}\right\} \quad \begin{matrix} E_n \text{ or } TM_n \\ \text{waves in} \\ \text{the di-} \\ \text{electric} \\ [5\cdot35] \end{matrix}$$

5·5 Transmission Properties of H Waves

In a manner similar to that employed in section 5·4, we may obtain from equations 5·19 an equation in E_z' alone. It is

$$\frac{\partial^2 E_z'}{\partial y^2} = -(\gamma^2 + \omega^2\mu_1\varepsilon_1)E_z' \qquad [5\cdot36]$$

The solution of this equation is

$$E_z' = C_3 \sin(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y + C_4 \cos(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y \qquad [5\cdot37]$$

where C_3 and C_4 are arbitrary constants depending on the amplitude of the original excitation. To satisfy the boundary conditions for a perfect conductor E_z must be zero at $y = \pm y_0/2$. Again, let

$$(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})\frac{y_0}{2} = \frac{n\pi}{2}$$

where n is an integer, so that

$$\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = \frac{n\pi}{y_0} \qquad [5\cdot37a]$$

Then

$$E_z' = C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right) \qquad [5\cdot37b]$$

As for the E waves, three distinct conditions arise, depending on whether n is zero, odd, or even. Under these conditions, the constants C_3 and C_4 must be alternately zero to satisfy the boundary requirements. Thus, if at $y = \pm y_0/2$,

$n = 0$	$E_z = C_4$	and hence C_4 must be zero
$n = 1, 3, 5,$ etc.	$E_z = \pm C_3$	and hence C_3 must be zero
$n = 2, 4, 6,$ etc.	$E_z = \pm C_4$	and hence C_4 must be zero

These requirements may be tabulated as follows:

	$n = 0$	$C_4 = 0$
(H waves)	n odd	$C_3 = 0$
	n even	$C_4 = 0$

[5·38]

Equations for H_x' and H_y' may be obtained from equations 5·19 and 5·37.

Thus from 5·19*b*

$$H_x' = -\frac{1}{j\omega\mu_1}\frac{\partial E_z'}{\partial y} = -\frac{\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}}{j\omega\mu_1}\{C_3 \cos(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y - C_4 \sin(\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})y\} \qquad [5\cdot39]$$

and from 5·19c

$$H_y' = -\frac{\gamma}{j\omega\mu_1}E_z' = -\frac{\gamma}{j\omega\mu_1}\{C_3 \sin(\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1})y + C_4 \cos(\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1})y\} \quad [5\cdot40]$$

Since $\mathbf{H} = \mathbf{H}'e^{j\omega t-\gamma x}$ and $\mathbf{E} = \mathbf{E}'e^{j\omega t-\gamma x}$, we may write, for the equations of the transverse electric mode of transmission in the dielectric:

$$\left.\begin{aligned} E_z &= \left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j\omega t-\gamma x} \\ H_x &= -\frac{1}{j\omega\mu_1}\left(\frac{n\pi}{y_0}\right)\left\{C_3 \cos\left(\frac{n\pi}{y_0}y\right) - C_4 \sin\left(\frac{n\pi}{y_0}y\right)\right\} e^{j\omega t-\gamma x} \\ H_y &= -\frac{\gamma}{j\omega\mu_1}\left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j\omega t-\gamma x} \end{aligned}\right\} \begin{array}{l} H_n \text{ or } TE_n \\ \text{waves in the} \\ \text{dielectric} \\ [5\cdot41]* \end{array}$$

These equations are of course subject to the restrictions imposed by equation 5·38.

5·6 Conditions for Wave Propagation

The constant γ appearing in equations 5·35 and 5·41 is known as the propagation constant. It is customary to set

$$\gamma \equiv \alpha + j\beta$$

where α is the attenuation constant and β is the phase constant. Thus

$$e^{-\gamma x} = e^{-\alpha x - j\beta x} = e^{-\alpha x}e^{-j\beta x} \quad [5\cdot42]$$

The term $e^{-\alpha x}$, the attenuation factor, determines the rate at which the amplitude of the wave is reduced as it progresses in the X direction. The term $e^{-j\beta x}$ determines the phase properties of the wave.

We may determine γ from equation 5·30. Thus

$$\gamma^2 = \left(\frac{n\pi}{y_0}\right)^2 - \omega^2\mu_1\varepsilon_1 \quad [5\cdot43]$$

and

$$\gamma_n = \sqrt{\left(\frac{n\pi}{y_0}\right)^2 - \omega^2\mu_1\varepsilon_1} = \alpha_n + j\beta_n \quad [5\cdot44]\dagger$$

So long as $\omega^2\mu_1\varepsilon_1 > (n\pi/y_0)^2$, γ_n has no real part and $\alpha_n = 0$. The attenu-

* In writing these equations, $\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1}$ has been replaced by $n\pi/y_0$ according to 5·37*a*.

† We take only the positive sign of the radical since we are interested only in propagation in the X direction. Also, since γ depends on the value of n chosen, it may be written γ_n. Thus γ_1 corresponds to the value of γ when $n = 1$, γ_2 for $n = 2$, etc. This procedure will be followed in similar cases where convenient.

ation is therefore zero, and $\gamma_n = j\beta_n$, $\gamma_n^2 = -\beta_n^2$, giving

$$\beta_n = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2} \qquad [5\cdot45]$$

Equations 5·35 for the transverse magnetic mode of transmission, and equations 5·41 for the transverse electric mode of transmission, may now be expressed in the following form,* subject, of course, to the restrictions of the relations 5·34 for the TM_n mode, and to the restrictions of relations 5·38 for the TE_n mode:

$$\left.\begin{aligned} H_z &= \left\{C_1 \sin\left(\frac{n\pi}{y_0}y\right) + C_2 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \\ E_x &= -\frac{j}{\omega\varepsilon_1}\left(\frac{n\pi}{y_0}\right)\left\{C_1 \cos\left(\frac{n\pi}{y_0}y\right) - C_2 \sin\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \\ E_y &= \frac{\beta_n}{\omega\varepsilon_1}\left\{C_1 \sin\left(\frac{n\pi}{y_0}y\right) + C_2 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \end{aligned}\right\} \begin{array}{l} E_n \text{ or } TM_n \\ \text{waves in the} \\ \text{dielectric} \\ [5\cdot46] \end{array}$$

$$\left.\begin{aligned} E_z &= \left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \\ H_x &= \frac{j}{\omega\mu_1}\left(\frac{n\pi}{y_0}\right)\left\{C_3 \cos\left(\frac{n\pi}{y_0}y\right) - C_4 \sin\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \\ H_y &= -\frac{\beta_n}{\omega\mu_1}\left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} e^{j(\omega t-\beta_n x)} \end{aligned}\right\} \begin{array}{l} H_n \text{ or } TE_n \\ \text{waves in the} \\ \text{dielectric} \\ [5\cdot47] \end{array}$$

These may also be expressed in the sinusoidal or cosinusoidal form instead of the exponential by taking the real part. Thus

$$\left.\begin{aligned} H_z &= \left\{C_1 \sin\left(\frac{n\pi}{y_0}y\right) + C_2 \cos\left(\frac{n\pi}{y_0}y\right)\right\} \cos(\omega t - \beta_n x) \\ E_x &= \frac{1}{\omega\varepsilon_1}\left(\frac{n\pi}{y_0}\right)\left\{C_1 \cos\left(\frac{n\pi}{y_0}y\right) - C_2 \sin\left(\frac{n\pi}{y_0}y\right)\right\} \sin(\omega t - \beta_n x) \\ E_y &= \frac{\beta_n}{\omega\varepsilon_1}\left\{C_1 \sin\left(\frac{n\pi}{y_0}y\right) + C_2 \cos\left(\frac{n\pi}{y_0}y\right)\right\} \cos(\omega t - \beta_n x) \end{aligned}\right\} \begin{array}{l} E_n \text{ or } TM_n \\ \text{waves in the} \\ \text{dielectric} \\ [5\cdot48] \end{array}$$

$$\left.\begin{aligned} E_z &= \left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} \cos(\omega t - \beta_n x) \\ H_x &= -\frac{1}{\omega\mu_1}\left(\frac{n\pi}{y_0}\right)\left\{C_3 \cos\left(\frac{n\pi}{y_0}y\right) - C_4 \sin\left(\frac{n\pi}{y_0}y\right)\right\} \sin(\omega t - \beta_n x) \\ H_y &= -\frac{\beta_n}{\omega\mu_1}\left\{C_3 \sin\left(\frac{n\pi}{y_0}y\right) + C_4 \cos\left(\frac{n\pi}{y_0}y\right)\right\} \cos(\omega t - \beta_n x) \end{aligned}\right\} \begin{array}{l} H_n \text{ or } TE_n \\ \text{waves in the} \\ \text{dielectric} \\ [5\cdot49] \end{array}$$

* In writing these relations γ_n has been replaced by $j\beta_n$, $\sqrt{\gamma_n^2 + \omega^2\mu_1\varepsilon_1}$ by $n\pi/y_0$.

If $\omega^2\mu_1\varepsilon_1 < (n\pi/y_0)^2$, γ_n has no imaginary part, $\beta_n = 0$, and the exponential term involving γ_n reduces to the hyperbolic sine or cosine. Under these conditions, there is no propagation at all. We can, therefore, define a critical frequency $f_0 = \omega_0/2\pi$ at which

$$\omega^2\mu_1\varepsilon_1 = \left(\frac{n\pi}{y_0}\right)^2 \qquad [5\cdot50]$$

or

$$(f_0)_n = \frac{1}{\sqrt{\mu_1\varepsilon_1}}\left(\frac{n}{2y_0}\right) \qquad [5\cdot51]$$

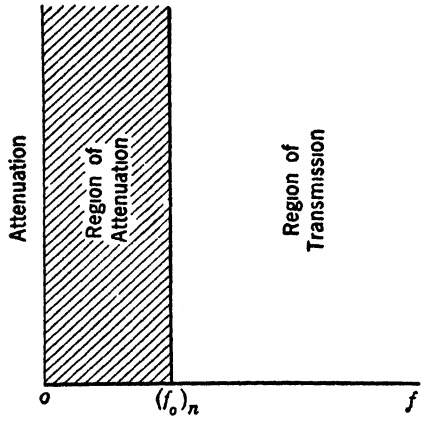


FIG 5·4 High-pass filter action of parallel plates. (Ideal case of perfect conductors)

For frequencies above this value γ_n is imaginary and propagation without attenuation takes place. For frequencies below this value, there is no propagation. Hence the parallel planes act as a high-pass wave filter having the characteristic indicated in Fig. 5·4. This is true, of course, only if the planes are perfect conductors.

Even if the conductivity of the bounding plates is infinite it must not be assumed that a physical source and receiver between the plates will not be coupled. The *TEM* mode discussed below has no cutoff frequency, and some power will be propagated in this way or by the familiar induction fields of low frequency. What is meant is that no wave in the particular *TE* or *TM* mode is propagated.

5·7 The Principal or *TEM* Transmission Mode (Perfect Conductors)

(Special case when $n = 0$)

When $n = 0$, $C_1 = 0$, and β_n reduces to

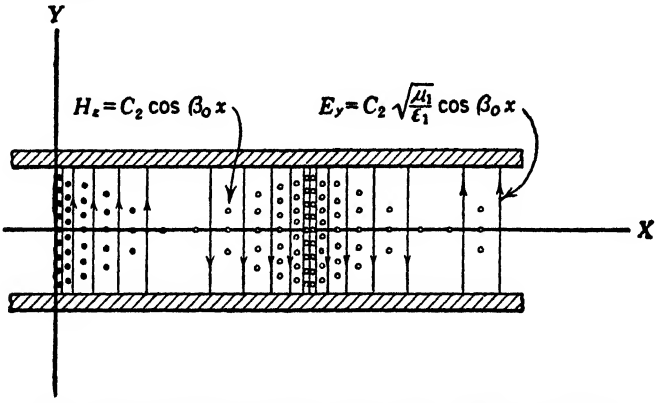
$$\beta_0 = \omega\sqrt{\mu_1\varepsilon_1}$$

Equations 5·48 then become

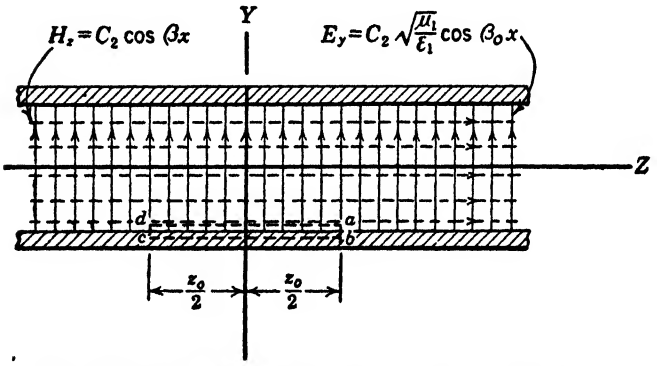
$$\left.\begin{aligned} H_z &= C_2\cos(\omega t - \beta_0 x) \\ E_x &= 0 \\ E_y &= C_2\sqrt{\frac{\mu_1}{\varepsilon_1}}\cos(\omega t - \beta_0 x) \end{aligned}\right\} \begin{matrix} TEM \\ \text{waves} \\ [5\cdot52] \end{matrix}$$

Since $E_x = 0$, both **E** and **H** vectors are at right angles to the direction of propagation. Thus both **E** and **H** are transverse, and we have

what might be called a transverse electromagnetic (*TEM*) mode of transmission. Under these conditions the critical frequency f_0 as given by equation 5·51 reduces to zero, so that there is no filter action. All frequencies will be transmitted without attenuation, and the waves will be guided between the plates without loss. The field distribution is shown in Fig. 5·5.



Side view of *TEM* wave when time is held fixed at t = zero.



End view of *TEM* wave when time is held fixed at t = zero.

FIG. 5·5

The situation may be visualized in another way. Suppose that a system of plane waves in free space impinges edgewise upon a pair of parallel plane conducting sheets in such a way that the electric vector is perpendicular to the sheets. Since the electric intensity has no component in the plane of the sheet there is no reflection and the wave continues without any change of form. Evidently part of the wave with its associated energy is propagated in the space between the sheets while the remainder continues outside them. The important point,

however, is that the wave is a true transverse electromagnetic wave, and that the guiding sheets serve only to limit the area of the wave front.

The familiar dissipationless parallel wire transmission line is closely similar to this case. The electric and magnetic lines or vectors are everywhere perpendicular to each other and to the direction of propagation. Energy of any frequency from zero upward may be transmitted, and the characteristic impedance is a pure resistance independent of frequency. If two pairs of wires are used as a transmission system, results are similar but the fields are relatively more confined. In fact, it may be shown that the plane case here discussed is merely the limit approached as more and more pairs of wires are added to the transmission system.

The velocity with which equiphase surfaces of this wave are propagated is given by the relation

$$(v_p)_n = \frac{\omega}{\beta_n} \qquad [5\cdot53]$$

Hence, for $n = 0$, equation 5·45 reduces to $\beta_0 = \omega\sqrt{\mu_1\varepsilon_1}$, and equation 5·53 may be written

$$(v_p)_0 = \frac{\omega}{\omega\sqrt{\mu_1\varepsilon_1}} = \frac{1}{\sqrt{\mu_1\varepsilon_1}} \qquad [5\cdot54]$$

giving the phase velocity v_p for the *TEM* mode. For air

$$\mu_1 = \mu_0 \qquad \text{and} \qquad \varepsilon_1 = \varepsilon_0$$

so that

$$(v_p)_0 = c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = 3 \times 10^8 \text{ meters per second} \qquad [5\cdot55]$$

5·8 Velocities of Propagation*

The term phase velocity, already used, is only one of three basically important velocities associated with wave motion. The others are group velocity and signal velocity. It is an unfortunate fact that the signal velocity, the most important of the three since it defines the effective time of signaling, is also the most obscure. Let us consider these various concepts.

The phase velocity is defined by the relation $v_p = f\lambda$, where λ is the distance between equiphase points of two successive waves in a steady-state system. Such waves we have typically represented by the function

$$f(\omega t - \beta x) \qquad [5\cdot56]$$

* For an excellent discussion of group and phase velocity see *Fundamentals of Electric Waves*, by H. H. Skilling, John Wiley & Sons, pages 164–168.

The surfaces of equal phase are characterized by the relation $f(\omega t - \beta x) = \text{constant}$ or

$$\omega t - \beta x = \text{constant} \qquad [5\cdot57]$$

These equiphase surfaces are propagated with a velocity

$$v_p = \frac{dx}{dt} = \frac{\omega}{\beta} = \frac{2\pi f}{\beta} = f \cdot \frac{2\pi}{\beta} = f\lambda \qquad [5\cdot58]$$

It will be shown later that this phase velocity may approach infinity or that it may be very low. Since energy is never propagated with a velocity exceeding that of light it is certain that such a quantity does not represent the propagation of energy or intelligence. Moreover, this definition for the phase velocity as given in 5·58 is based upon steady-state conditions and has nothing to do with the rate at which the wave system is initially built up.

Another important velocity concept is that designated group velocity. As its name implies, group velocity is associated with the velocity of more than one wave. It is defined by the relation

$$v_g = \cdot \frac{1}{\frac{\partial \beta}{\partial \omega}} \cdot \qquad [5\cdot59]$$

and is the velocity with which the envelope of an amplitude-modulated wave is propagated when the modulation frequency is very low compared to the carrier.

Let us consider an amplitude-modulated wave in which the function $\cos(\omega t - \beta x)$ describes the high-frequency or carrier wave and the function $\cos(\delta\omega t - \delta\beta x)$* describes the low modulation-frequency wave. The amplitude-modulated wave may be expressed in the form

$$E = E'[(1 + m\cos(\delta\omega t - \delta\beta x)]\cos(\omega t - \beta x) \qquad [5\cdot60]$$

where E' is the maximum value of the intensity E in distance and time and m is the modulation factor, which may have any value from zero to unity. Expanding 5·60, we obtain

$$\begin{aligned} E &= E'\{\cos(\omega t - \beta x) + m\cos(\delta\omega t - \delta\beta x)\cos(\omega t - \beta x)\} \\ &= E'\{\cos(\omega t - \beta x) + \frac{m}{2}\cos[(\omega + \delta\omega)t - (\beta + \delta\beta)x] + \\ &\qquad \frac{m}{2}\cos[(\omega - \delta\omega)t - (\beta - \delta\beta)x]\} \end{aligned} \qquad [5\cdot61]$$

* The symbol δ is used here to represent a small non-vanishing number. Thus $\delta\omega/2\pi$ represents the audio frequency, which is small in comparison to the carrier frequency $\omega/2\pi$.

It is seen that the wave of 5·61 evidently contains three component waves, including the original carrier and two others, one having a frequency slightly greater and the other a frequency slightly less than the carrier frequency.

Equation 5·61, when m is unity, may be written as

$$E = \frac{E'}{2}\Big[\{\cos(\omega t - \beta x) + \cos[(\omega + \delta\omega)t - (\beta + \delta\beta)x]\} + \{\cos(\omega t - \beta x) + \cos[(\omega - \delta\omega)t - (\beta - \delta\beta)x]\}\Big] \qquad [5\cdot62]$$

Let us consider the first term in the brackets of equation 5·62. By means of the trigonometric identity

$$\cos A + \cos B = 2\cos\tfrac{1}{2}(A - B)\cos\tfrac{1}{2}(A + B)$$

it may be expressed as

$$2\cos\tfrac{1}{2}(\delta\beta x - \delta\omega t)\cos\left[\left(\omega + \frac{\delta\omega}{2}\right)t - \left(\beta + \frac{\delta\beta}{2}\right)x\right] \qquad [5\cdot63]$$

This expression is recognizable as that of the familiar phenomenon of beats illustrated in Fig. 5·6. It must be remembered that this figure represents only a part of the transmitted wave of equation 5·62. The

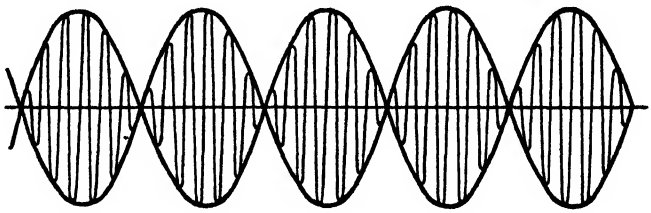


FIG. 5·6 Beats propagated with a velocity equal to the group velocity.

actual electric field oscillates at a frequency negligibly different from $\omega/2\pi$, and the amplitude varies slowly according to the envelope relation

$$\cos\tfrac{1}{2}(\delta\beta x - \delta\omega t) \qquad [5\cdot64]$$

from zero to a maximum value given by the sum of the amplitudes of the carrier and modulation frequencies. Consideration of the second term of 5·62 will yield the same result. Evidently constancy of the envelope demands

$$x\delta\beta - t\delta\omega = \text{constant} \qquad [5\cdot65]$$

Accordingly the beats or groups are propagated with a velocity,

$$v_g = \frac{x}{t} = \frac{\delta\omega}{\delta\beta} \quad \text{or in the limit} \quad \cdot\frac{1}{\dfrac{\partial\beta}{\partial\omega}}\cdot \qquad [5\cdot66]$$

From our definition of group velocity it is clear that it applies to all frequencies lying within a narrow band of some mean frequency. Evidently this is exactly the case of a wave modulated with a pure sine wave or any superimposed array of sinusoidals such as speech, provided only that the total band width of the signal is small compared to the carrier frequency.

The group velocity is of considerable practical significance in that it defines the velocity at which low-frequency signals are propagated over a high-frequency carrier system. Thus speech or television signals in ordinary carrier practice are propagated at the group velocity of the transmission system.

From the definition of phase velocity, we have

$$v_p = \frac{\omega}{\beta} \qquad \text{or} \qquad \omega = v_p \beta \tag{5·58}$$

If v_p is independent of ω we have, taking the derivative,

$$1 = v_p \frac{\partial \beta}{\partial \omega} \qquad \text{or} \qquad v_p = \cdot \frac{1}{\dfrac{\partial \beta}{\partial \omega}} \cdot = v_g \tag{5·67}$$

A medium in which v_p is independent of ω is called non-dispersive. It is seen that in such media the phase and group velocities are equal. Other media, in which v_p is dependent upon ω, are called dispersive.

The concept of signal velocity becomes necessary only in dispersive media, that is, media in which the phase and group velocities are different. Since information can be transmitted only by some form of transient disturbance and since transients often occupy wide regions of the frequency spectrum, the problem is far from simple. Let us assume a specific example in which a sinusoidal voltage is applied to an antenna in a wave-guide system. Let the frequency be somewhat higher than the cut-off value of the system but low enough so that the phase velocity is not independent of frequency. Under these conditions the phase velocity is larger than c, the velocity of light, and the group velocity is less than c.

If the antenna is energized at a time $t = 0$, we wish to determine at what time an effect is first observed at a distance d away. Evidently the phase and group velocity concepts are not directly applicable since no form of steady state exists. Accordingly we must examine the problem carefully from a fundamental viewpoint. Fortunately the problem has much in common with that of a voltage suddenly applied to a lumped electric circuit.

When an alternating voltage is suddenly applied to a lumped net-

work, the expression for the resulting current is quite complex even for simple networks. In general, however, two terms serve to express the current function. The first of these terms, the so-called steady-state term, begins at the instant that the switch is closed and continues without any change until the circuit is again modified. The second term expresses the transient current which flows only for a limited time after the switch is closed. With unimportant exceptions the transient current is always of such a form as to cancel the steady-state current at time $t = 0$. Accordingly the total current builds up at a finite rate and in a more or less complex fashion from zero to the steady-state value.

It is thus impossible to say exactly when the current commences to flow, for at any finite value of time there is some finite value of current, and mathematically the steady state is never reached. Depending upon how sensitive a device we use to record the presence or absence of current, we can make the interval as short or as long as we choose.

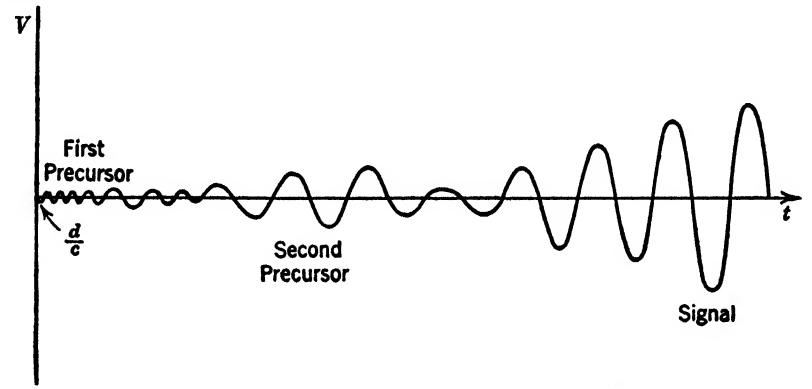


FIG. 5·7 The arrival of a transient wave.*

In wave propagation the situation is similar but even more complex. The first impulse reaches the point d with the velocity c regardless of the nature of the medium, but the amplitude of this first impulse or wave is zero. This first impulse is a wave train of very high frequency which first grows and then diminishes in amplitude with the passage of time. This first wave train or precursor, as it is called, may be followed by a second precursor of lower frequency and larger amplitude before the final wave train of the steady state arrives. The strength V of the received signal is plotted against time for a typical case in Fig. 5·7.

It is thus clear that no definite arrival time can be specified under

*An excellent account of this subject is given by Brillouin, *Congrès international d'électricité*, Vol. II, Paris, 1932.

these conditions. At all times greater than d/c some power is present at the point d. In general this power increases continuously with the passage of time, reaching a final value only after an infinite length of time. Accordingly, our determination of the velocity must depend entirely upon the sensitivity of the detector used to judge the arrival of the waves. In practice, the portion of the disturbance designated signal and associated with the steady state is often much larger than the precursors and arrives with a reasonably sharp wave front. Accordingly detectors of a considerable range of sensitivity may give essentially the same result for the signal velocity.

In coaxial cables or other wave guides which have no low-frequency cut-off, the phase and group velocities are essentially the same and are equal to the signal velocity.

5·9 Voltage, Current, and Power Relations in the *TEM* Mode

We may determine the maximum value of the voltage V' between the upper and lower plate in the OY direction, Fig. 5·5, by taking the line integral of the electric intensity between the upper and lower plates. Integrating along a line of constant electric intensity, we obtain:

$$V' = \int_l E' \cos\theta \, dl = \int_{-y_0/2}^{+y_0/2} E'_y \, dy$$

Substituting for E'_y its value as given in equation 5·52, we obtain

$$V' = \int_{-y_0/2}^{+y_0/2} C_2 \sqrt{\frac{\mu_1}{\varepsilon_1}} \, dy = C_2 \sqrt{\frac{\mu_1}{\varepsilon_1}} \, y_0 \qquad [5·68]$$

Also, the longitudinal current, per unit width, along the upper or lower plate, Fig. 5·5, may be found to be

$$I' = \oint_l H' \cos\theta \, dl = \int_d^a H'_z \, dz + \int_a^b 0 \, dy + \int_b^c 0 \, dz + \int_c^d 0 \, dy$$

Substituting for H'_z its value C_2 as given in equation 5·52 we have

$$I' = \int_d^a C_2 \, dz = C_2 z_0 \qquad [5·69]$$

where the path of integration is shown in Fig. 5·5.

By means of the complex Poynting vector, we may calculate the average power flowing across unit area in the X direction. From equations 3·87

$$p_x = \tfrac{1}{2}(E'_y H'_z - E'_z H'_y) \qquad [5·70]$$

Since E'_z and H'_y are zero

$$p_x = \tfrac{1}{2}(E'_y H'_z)$$

From equations 5·52

$$E_y' = C_2\sqrt{\frac{\mu_1}{\epsilon_1}} \qquad \text{and} \qquad H_z' = C_2$$

Hence

$$p_x = \tfrac{1}{2}C_2^2\sqrt{\frac{\mu_1}{\epsilon_1}} \tag{5·71}$$

giving the average power flowing across unit area. The average power flowing across an area z_0 centimeters in the Z direction and y_0 centimeters in the Y direction is

$$P_x = \tfrac{1}{2}C_2^2\sqrt{\frac{\mu_1}{\epsilon_1}}\, z_0 y_0 \tag{5·72}$$

We may define a characteristic impedance on a voltage or power basis thus:

On a voltage-current basis

$$Z_0 = \frac{V'}{I'} = \sqrt{\frac{\mu_1}{\epsilon_1}}\,\frac{y_0}{z_0} \tag{5·73}$$

On a power basis

$$Z_0 = \frac{P}{(I_{\text{rms}})^2} = \frac{\tfrac{1}{2}C_2^2\sqrt{\dfrac{\mu_1}{\epsilon_1}}\, z_0 y_0}{\tfrac{1}{2}C_2^2 z_0^2} = \sqrt{\frac{\mu_1}{\epsilon_1}}\,\frac{y_0}{z_0}$$

We may also define a quantity η which is characteristic of the particular wave type and is somewhat like an impedance. This quantity defined in terms of the intensity vectors is

$$\eta = \frac{E_y}{H_z} = \sqrt{\frac{\mu_1}{\epsilon_1}} \tag{5·74}$$

If the dielectric between the plates is free space, $\mu_1 = \mu_0$ and $\epsilon_1 = \epsilon_0$, then

$$\eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377\ \Omega \tag{5·75}$$

giving the so-called intrinsic impedance of free space.

The concept of the impedance of free space is not a familiar one. Nor is it particularly easy to visualize such a concept. If, however, we evaluate the ratio of E to H in the *TEM* mode in air between guides or in the plane wave in free space we always secure the number 120π =

377 ohms. If a specified electric intensity is impressed in the medium by some source, then the larger ε becomes the more power is radiated from the source. Conversely the larger μ becomes the less power is radiated. This behavior is directly analogous to that observed when a long coaxial cable is connected to a generator. A large dielectric constant is associated with a low characteristic impedance and a large power input.

5·10 The TM_n Transmission Mode (E_n Waves) $n \neq 0$

The E waves in the dielectric may be conveniently classified into two groups depending on whether n is odd or even. We have from equations 5·48, subject to restrictions imposed by 5·34,

$$\left.\begin{array}{c} n \text{ odd}, \quad C_2 = 0 \\ H_z = C_1 \sin\left(\frac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \\ E_x = \frac{1}{\omega \varepsilon_1}\left(\frac{n\pi}{y_0}\right) C_1 \cos\left(\frac{n\pi}{y_0} y\right) \\ \sin(\omega t - \beta_n x) \\ E_y = \frac{\beta_n}{\omega \varepsilon_1} C_1 \sin\left(\frac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \end{array}\right\} [5{\cdot}76] \quad \left.\begin{array}{c} n \text{ even}, \quad C_1 = 0 \\ H_z = C_2 \cos\left(\frac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \\ E_x = -\frac{1}{\omega \varepsilon_1}\left(\frac{n\pi}{y_0}\right) C_2 \sin\left(\frac{n\pi}{y_0} y\right) \\ \sin(\omega t - \beta_n x) \\ E_y = \frac{\beta_n}{\omega \varepsilon_1} C_2 \cos\left(\frac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \end{array}\right\} [5{\cdot}77]$$

where the phase constant is given by

$$\beta_n = \sqrt{\omega^2 \mu_1 \varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2} \qquad [5{\cdot}78]$$

and the critical frequency from equation 5·51 is

$$(f_0)_n = \frac{1}{\sqrt{\mu_1 \varepsilon_1}} \left(\frac{n}{2y_0}\right) = \frac{n v_1}{2y_0} \qquad [5{\cdot}79]$$

where $v_1 = 1/\sqrt{\mu_1 \varepsilon_1}$ is the velocity of a free wave in a medium of dielectric constant ε_1 and permeability μ_1.

The critical wavelength is

$$(\lambda_0)_n = \frac{v_1}{(f_0)_n} = \frac{2y_0}{n} \qquad [5{\cdot}80]$$

The wavelength between plates is

$$\lambda_n = \frac{v_p}{f} = \frac{\omega}{\beta_n f} = \frac{2\pi}{\beta_n} \qquad [5{\cdot}81]$$

The phase velocity is

$$(v_p)_n = \frac{\omega}{\beta_n} \qquad [5\cdot82]$$

The group velocity is

$$(v_g)_n = \cdot \frac{1}{\dfrac{d\beta_n}{d\omega}} \cdot = \frac{\beta_n}{\omega\mu_1\varepsilon_1} = \frac{v_1^2}{(v_p)_n} \qquad [5\cdot83]$$

It is seen that the product of the phase and group velocity is equal to the square of the free wave velocity v_1. That is,

$$v_1^2 = (v_p)_n(v_g)_n = \frac{1}{\mu_1\varepsilon_1}$$

In air

$$(v_p)_n(v_g)_n = \frac{1}{\mu_0\varepsilon_0} = c^2 \qquad [5\cdot84]$$

where c is the velocity of light.

The value of the integer n is taken to define the order of the wave. For the lowest-order wave, $n = 1$, the critical frequency is lowest; it increases as the order of the wave is increased. It also increases as the distance y_0 between the plates decreases. It is decreased as the dielectric constant ε_1 of the medium between the plates is increased.

5·11 The TM_1 Transmission Mode (E_1 Waves) ($n = 1$)

The amplitude of the components of the first-order E_n wave, i.e., the E_1 wave, may be written from equations 5·76 when $n = 1$ as

$$\left.\begin{aligned} H_z' &= C_1 \sin\left(\frac{\pi}{y_0} y\right) \\ E_x' &= \frac{\pi}{\omega\varepsilon_1 y_0} C_1 \cos\left(\frac{\pi}{y_0} y\right) \\ E_y' &= \frac{\beta_1}{\omega\varepsilon_1} C_1 \sin\left(\frac{\pi}{y_0} y\right) \end{aligned}\right\} \begin{matrix} TM_1 \text{ or } E_1 \\ \text{wave} \\ [5\cdot85] \end{matrix}$$

The propagation properties of this wave are:

$$\beta_1 = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{y_0}\right)^2} \qquad [5\cdot86]$$

$$(f_0)_1 = \frac{v_1}{2y_0} \qquad [5\cdot87]$$

$$(\lambda_0)_1 = 2y_0 \qquad [5.88]$$

$$(\lambda)_1 = \frac{2\pi}{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}} = \frac{2\pi}{\sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{y_0}\right)^2}} \qquad [5.89]$$

$$(v_p)_1 = \frac{\omega}{\beta_1} = \frac{\omega}{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}} = \frac{\omega}{\sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{y_0}\right)^2}} \qquad [5.90]$$

$$(v_g)_1 = \frac{v_1^2}{(v_p)_1} \qquad [5.91]$$

$$v_1 = \frac{1}{\sqrt{\mu_1\varepsilon_1}} \qquad [5.92]$$

The field distribution at any instant of time may be obtained by setting t = constant = 0 in equations 5·76. Then, from Fig. 5·8

$$\frac{E_y}{E_x} = \tan\theta = \frac{dy}{dx} \qquad [5.93]$$

and

$$\frac{E_y}{E_x} = \frac{\beta_1 y_0}{\pi} \tan\left(\frac{\pi}{y_0} y\right) \cot(-\beta_1 x) \qquad [5.94]$$

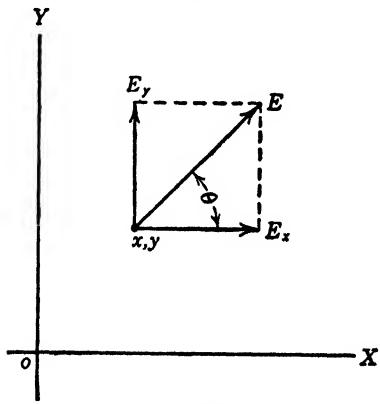


FIG. 5·8

This equation may be solved by analytical or graphical methods and leads to the field distribution shown in Fig. 5·9. The side view of Fig 5·9 is propagated in the X direction between the plates at a velocity given by $(v_p)_1$ in equation 5·90.

If we factor $(\omega/v_1)^2$ from equation 5·90 and replace $v_1/2y_0$ by $(f_0)_1$, we obtain:

$$(v_p)_1 = \frac{v_1}{\sqrt{1 - \left(\frac{v_1\pi}{\omega y_0}\right)^2}} = \frac{v_1}{\sqrt{1 - \left(\frac{(f_0)_1}{f}\right)^2}} \qquad [5\cdot95]$$

also

$$(v_g)_1 = \frac{v_1^2}{(v_p)_1} = v_1\sqrt{1 - \left(\frac{(f_0)_1}{f}\right)^2} \qquad [5\cdot96]$$

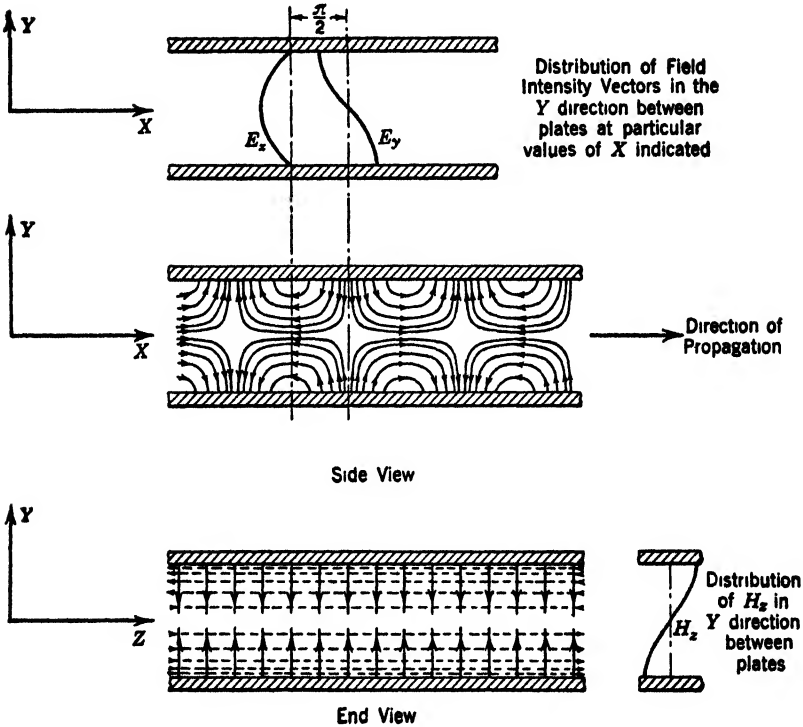


FIG. 5·9 Instantaneous field configuration for an E_1 or TM_1 wave in the dielectric medium between two parallel conducting planes of infinite extent. This field structure is propagated in the X direction between the plates.

These equations may be plotted as shown in Fig. 5·10. As the frequency f of the wave is increased without limit, its velocity v_p or v_g approaches the velocity v_1 of a free wave traveling in an infinite dielectric medium having the same properties as the medium between the plates. If the dielectric is free space, the velocity that is approached is c, the

velocity of light. As the frequency of the wave approaches the critical frequency, the phase velocity $(v_p)_1$ increases without limit. This will be explained in section 5·15 by decomposing the waves of Fig. 5·9 into two sets of ordinary plane waves that are reflected back and forth between the upper and lower plates. It will be shown that, as the critical frequency is approached, the direction of propagation of the elementary waves approaches the Y direction and there is no transmission of the waves between the plates in the X direction. Since the equiphase surfaces of the waves are infinite in extent, the equiphase surfaces of the elementary plane waves extend throughout the region between the plates when their direction of propagation is normal to the surface of the plates. Thus, as the direction of propagation approaches the Y direction, the phase velocity, i.e., the velocity of propagation of the equiphase surfaces, approaches infinity. To say that the phase velocity is infinite is simply stating that it takes no time at all for a wave to travel between two points both of which lie in an equiphase surface of the wave.

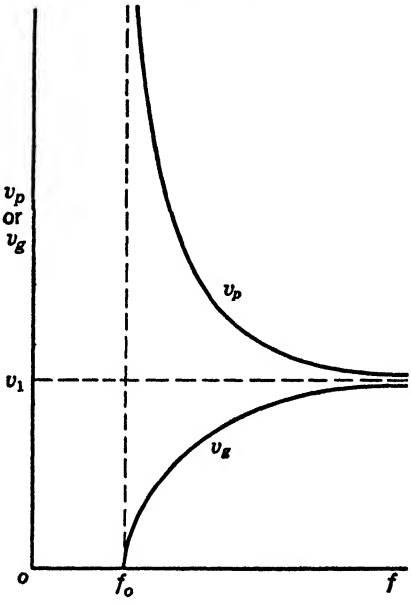


FIG. 5·10 Variation of phase and group velocity as a function of frequency.

5·12 The TM_2 Transmission Mode (E_2 Waves) $n = 2$

The amplitude of the second-order E_n wave, i.e., the E_2 wave, may be written from equation 5·77 when $n = 2$, as

$$\left.\begin{aligned} H'_z &= C_2 \cos\left(\frac{2\pi}{y_0} y\right) \\ E'_x &= -\frac{1}{\omega\varepsilon_1}\left(\frac{2\pi}{y_0}\right) C_2 \sin\left(\frac{2\pi}{y_0} y\right) \\ E'_y &= \frac{\beta_2}{\omega\varepsilon_1} C_2 \cos\left(\frac{2\pi}{y_0} y\right) \end{aligned}\right\} \quad \begin{matrix} TM_2 \text{ or } E_2 \\ \text{wave} \\ [5\text{·}97] \end{matrix}$$

The propagation properties of this wave are:

$$\beta_2 = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{2\pi}{y_0}\right)^2} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{2\pi}{y_0}\right)^2} \qquad [5\cdot98]$$

$$(f_0)_2 = \frac{v_1}{y_0} \qquad [5\cdot99]$$

$$(\lambda_0)_2 = y_0 \qquad [5\cdot100]$$

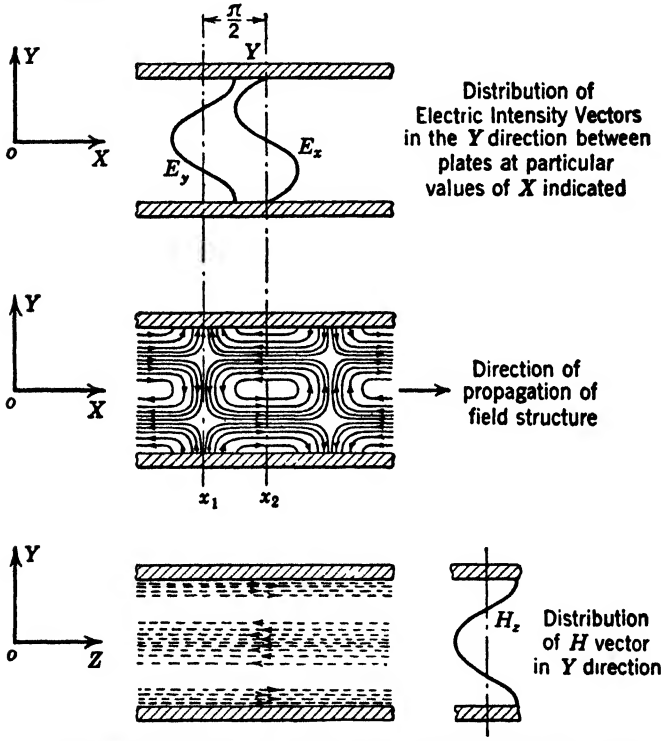


FIG. 5·11 Instantaneous field configuration for an E_2 or TM_2 wave in the dielectric medium between two parallel conducting planes of infinite extent.

$$(\lambda)_2 = \frac{2\pi}{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{2\pi}{y_0}\right)^2}} = \frac{2\pi}{\sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{2\pi}{y_0}\right)^2}} \qquad [5\cdot101]$$

$$(v_p)_2 = \frac{\omega}{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{2\pi}{y_0}\right)^2}} = \frac{\omega}{\sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{2\pi}{y_0}\right)^2}} \qquad [5\cdot102]$$

$$(v_g)_2 = \frac{v_1^2}{(v_p)_2} \qquad [5\cdot103]$$

$$v_1 = \frac{1}{\sqrt{\mu_1 \varepsilon_1}} \qquad [5\cdot104]$$

In this case we may substitute $v_1 = y_0(f_0)_2$ from 5·99 in 5·102 after multiplying numerator and denominator by v_1/ω. The expression for $(v_p)_2$ may then be written

$$(v_p)_2 = \frac{v_1}{\sqrt{1 - \left(\frac{2\pi v_1}{y_0 \omega}\right)^2}} = \frac{v_1}{\sqrt{1 - \left((f_0)_2 \frac{2\pi}{\omega}\right)^2}} = \frac{v_1}{\sqrt{1 - \left(\frac{(f_0)_2}{f}\right)^2}}$$

also

$$(v_g)_2 = \frac{v_1^2}{(v_p)_2} = v_1 \sqrt{1 - \left(\frac{(f_0)_2}{f}\right)^2}$$

The field distribution may be obtained as for the E_1 wave; it is shown in Fig. 5·11.

5·13 The TE_0 Transmission Mode (H_0 Waves) $n = 0$

When $n = 0$, it is seen from the relations 5·38 for the transverse electric transmission mode that $C_4 = 0$. When the conditions $n = 0$ and $C_4 = 0$ are applied to the general equations 5·49 for the H_n wave we have

$$E_z = 0, \quad H_x = 0, \quad H_y = 0$$

Thus no wave of this type is possible between the plane parallel conductors. More accurately, the *TEM* wave discussed in section 5.7 replaces the H_0 wave. When the order of the *TE* wave is reduced to zero the longitudinal component of H becomes zero and a *TEM* wave results.

5·14 The TE_n Transmission Mode (H Waves) $n \neq 0$

The H waves in the dielectric may also be classified into two groups. For one of these groups n is odd, and for the other n is even. In either case n is not equal to zero. From the general equation for the transverse electric mode of transmission as given in equations 5·49, we have,

subject to the restrictions imposed by the relations 5·38,

$$\left.\begin{array}{l} n \text{ odd, } C_3 = 0 \\ E_z = C_4 \cos\left(\dfrac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \\ H_x = \dfrac{1}{\omega\mu_1}\left(\dfrac{n\pi}{y_0}\right) C_4 \sin\left(\dfrac{n\pi}{y_0} y\right) \sin(\omega t - \beta_n x) \\ H_y = -\dfrac{\beta_n}{\omega\mu_1} C_4 \cos\left(\dfrac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \end{array}\right\} \quad [5\cdot105]$$

$$\left.\begin{array}{l} n \text{ even, } C_4 = 0 \\ E_z = C_3 \sin\left(\dfrac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \\ H_x = -\dfrac{1}{\omega\mu_1}\left(\dfrac{n\pi}{y_0}\right) C_3 \cos\left(\dfrac{n\pi}{y_0} y\right) \sin(\omega t - \beta_n x) \\ H_y = -\dfrac{\beta_n}{\omega\mu_1} C_3 \sin\left(\dfrac{n\pi}{y_0} y\right) \cos(\omega t - \beta_n x) \end{array}\right\} \quad [5\cdot106]$$

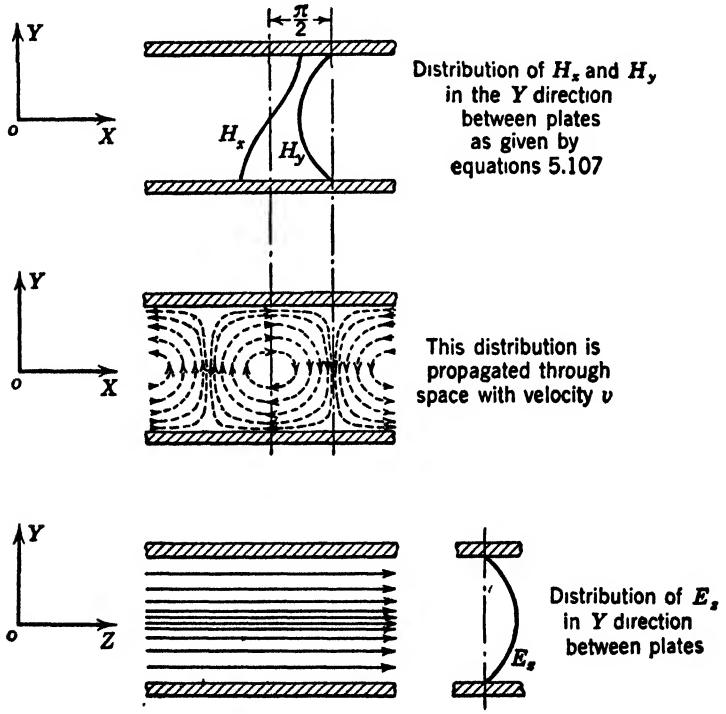


FIG. 5·12 Instantaneous field configuration for an H_1 or TE_1 wave in the dielectric medium between two parallel conducting planes of infinite extent.

The expressions for β_n, $(f_0)_n$, $(\lambda_0)_n$, $(\lambda)_n$, $(v_p)_n$, and $(v_g)_n$ for the H waves are the same as those already given for the E waves in equations 5·78 to 5·83 inclusive. The transmission properties are identical for E and H waves of the same order.

The TE_1 transmission mode (H_1 waves) $n = 1$

The amplitudes of the first-order H_n wave, i.e., the H_1 wave, may be written from equations 5·105 when $n = 1$. Thus

$$\left.\begin{aligned} E'_z &= C_4 \cos\left(\frac{\pi}{y_0} y\right) \\ H'_x &= \frac{1}{\omega\mu_1}\left(\frac{\pi}{y_0}\right) C_4 \sin\left(\frac{\pi}{y_0} y\right) \\ H'_y &= -\frac{\beta_1}{\omega\mu_1} C_4 \cos\left(\frac{\pi}{y_0} y\right) \end{aligned}\right\} \qquad [5\cdot107]$$

The transmission properties of this wave are given in equations 5·86 through 5.92. The field distribution is shown in Fig. 5·12.

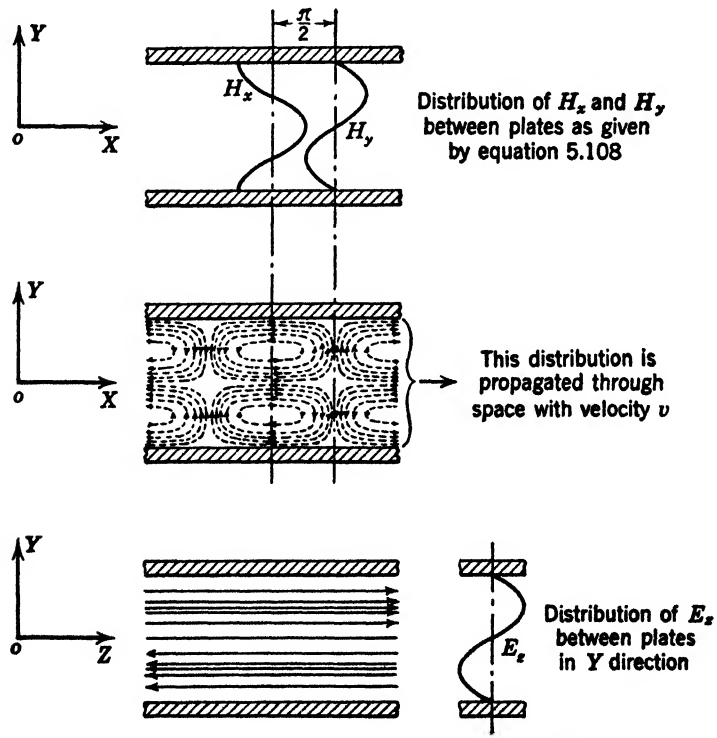


FIG. 5·13 Instantaneous field configuration for an H_2 or TE_2 wave in the dielectric medium between two parallel conducting planes of infinite extent.

The TE_2 transmission mode (H_2 wave) $n = 2$

The amplitude of the second-order H_n wave may be written from equations 5·106 when $n = 2$. Thus for the H_2 wave

$$\left.\begin{aligned} E_z' &= C_3 \sin\left(\frac{2\pi}{y_0} y\right) \\ H_x' &= -\frac{1}{\omega\mu_1}\left(\frac{2\pi}{y_0}\right) C_3 \cos\left(\frac{2\pi}{y_0} y\right) \\ H_y' &= -\frac{\beta_2}{\omega\mu_1} C_3 \sin\left(\frac{2\pi}{y_0} y\right) \end{aligned}\right\} \qquad [5\cdot108]$$

The transmission properties of the H_2 wave are given in equations 5·98 to 5·104, and the field distribution is shown in Fig. 5·13.

5·15 The E_1 Wave: Resolution into Elementary Waves*

Let us change the expressions 5·85–5·92 for the E_1 wave into an equivalent form in which this wave appears as the superposition of two ordinary plane waves. We may then deduce the direction of propagation of these waves and explain the transmission of the E_1 wave in terms of its more elementary or component parts. From equations 5·46 with $n = 1$ and $C_2 = 0$ we may write in exponential form

$$\left.\begin{aligned} H_z &= C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{j\omega t - j\beta_1 x} \\ E_x &= -\frac{j\pi}{\omega\varepsilon_1 y_0} C_1 \cos\left(\frac{\pi}{y_0} y\right) e^{j\omega t - j\beta_1 x} \\ E_y &= \frac{\beta_1}{\omega\varepsilon_1} C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{j\omega t - j\beta_1 x} \end{aligned}\right\} \begin{matrix} E_1 \text{ or } TM_1 \\ \text{wave} \\ [5\cdot109] \end{matrix}$$

Let us consider first the expression for H_z. Since

$$\sin\left(\frac{\pi}{y_0} y\right) = \frac{1}{2j}\left(e^{j\frac{\pi}{y_0}y} - e^{-j\frac{\pi}{y_0}y}\right)$$

we may write

$$H_z = C_1 \frac{1}{2j}\left\{e^{j\omega t - j\left(\beta_1 x - \frac{\pi}{y_0} y\right)} - e^{j\omega t - j\left(\beta_1 x + \frac{\pi}{y_0} y\right)}\right\} \qquad [5\cdot110]$$

Thus H_z is separated into two components $H_z^{(1)}$ and $H_z^{(2)}$ defined by

$$H_z^{(1)} = C_1 \frac{1}{2j} e^{j\omega t - j\left(\beta_1 x - \frac{\pi}{y_0} y\right)} \qquad [5\cdot111]$$

* This type of resolution was first suggested by Brillouin.

and

$$H_z^{(2)} = -C_1 \frac{1}{2j} e^{j\omega t - j\left(\beta_1 x + \frac{\pi}{y_0} y\right)} \qquad [5\cdot112]$$

where

$$H = H_z = H_z^{(1)} + H_z^{(2)} \qquad [5\cdot113]$$

These vector components may be interpreted as two component waves traveling in different directions. The direction of the $H_z^{(1)}$ wave may be found from the exponent $(\beta_1 x - (\pi/y_0)y)$ whose geometrical representation is given in Fig. 5·14.

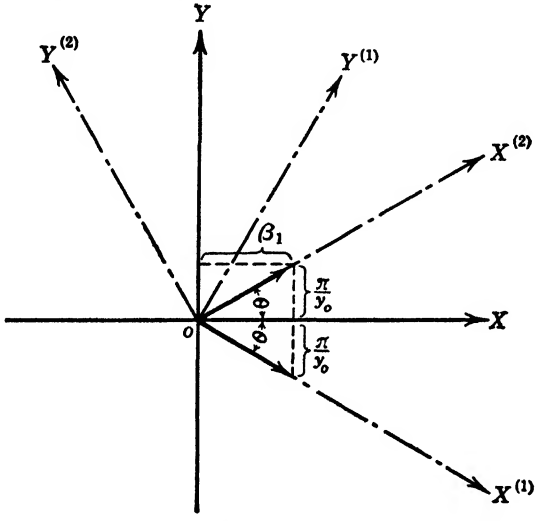


FIG. 5·14 Geometrical representation of the exponent $\left(\beta_1 x - \frac{\pi}{y_0} y\right)$.

Thus equation 5·111 represents a vector component wave traveling in the $X^{(1)}$ direction at an angle $\theta = \tan^{-1}(-\pi/\beta_1 y_0)$ with the X axis. Similarly equation 5·112 represents a wave traveling in the $X^{(2)}$ direction at an angle $\theta = \tan^{-1}(\pi/\beta_1 y_0)$ with the X axis. Hence the H_z component is expressed in terms of two component waves, $H_z^{(1)}$ and $H_z^{(2)}$, which are traveling in the $X^{(1)}$ and $X^{(2)}$ directions, respectively. To analyze these components it is convenient to express $H_z^{(1)}$ in terms of the new axes $X^{(1)}$ $Y^{(1)}$, and $H_z^{(2)}$ in terms of the new axes $X^{(2)}$ $Y^{(2)}$.

Let us consider first the $H_z^{(1)}$ component. To make the required transformation to the new axes it is necessary to find an expression for x and y in terms of $x^{(1)}$ and $y^{(1)}$. To do this set up the axes of Fig. 5·15 and consider the coordinates of any point x, y in terms of the coordinates $x^{(1)}$ $y^{(1)}$ of the new axes. From the geometry of the prob-

lem we see that the required analytic transformation is

$$y = y^{(1)} \cos\theta - x^{(1)} \sin\theta \qquad [5\cdot114]$$

$$x = x^{(1)} \cos\theta + y^{(1)} \sin\theta \qquad [5\cdot115]$$

Therefore

$$\beta_1 x - \frac{\pi}{y_0} y = \beta_1 (x^{(1)} \cos\theta + y^{(1)} \sin\theta) - \frac{\pi}{y_0} (y^{(1)} \cos\theta - x^{(1)} \sin\theta)$$

$$= \left(\beta_1 \cos\theta + \frac{\pi}{y_0} \sin\theta\right) x^{(1)} + \left(\beta_1 \sin\theta - \frac{\pi}{y_0} \cos\theta\right) y^{(1)} \qquad [5\cdot116]$$

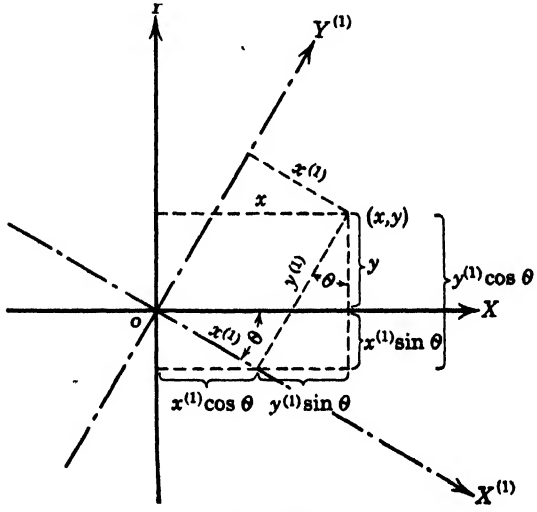


FIG. 5·15

Now from Fig. 5·14

$$\sin\theta = \frac{\pi}{y_0 \sqrt{\beta_1^2 + \left(\frac{\pi}{y_0}\right)^2}} \quad \text{and} \quad \cos\theta = \frac{\beta_1}{\sqrt{\beta_1^2 + \left(\frac{\pi}{y_0}\right)^2}} \qquad [5\cdot117]$$

Also, from equation 5·86

$$\beta_1 = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{y_0}\right)^2}$$

where v_1 is the velocity of a free wave in a medium of dielectric constant ε_1 and permeability μ_1, and $v_1 = 1/\sqrt{\mu_1 \varepsilon_1}$. Substituting for β_1 in

equations 5·117 we obtain

$$\sin\theta = \frac{\pi}{y_0}\left(\frac{v_1}{\omega}\right) \qquad \text{and} \qquad \cos\theta = \beta_1\left(\frac{v_1}{\omega}\right) \tag{5·118}$$

Replacing the sin θ and cos θ in equation 5·116 by their equivalents as given in 5·118, we have

$$\beta_1 x - \frac{\pi}{y_0} y = \left\{\beta_1^2 + \left(\frac{\pi}{y_0}\right)^2\right\}\frac{v_1}{\omega}x^{(1)} - \left\{\beta_1\frac{\pi}{y_0} - \beta_1\frac{\pi}{y_0}\right\}\frac{v_1}{\omega}y^{(1)} = \left(\frac{\omega}{v_1}\right)^2\frac{v_1}{\omega}x^{(1)}$$

or

$$\beta_1 x - \frac{\pi}{y_0} y = \frac{\omega}{v_1}x^{(1)} \tag{5·119}$$

Hence

$$H_z^{(1)} = C_1\frac{1}{2j}e^{j\omega t - j\frac{\omega}{v_1}x^{(1)}} = C_1\frac{1}{2j}e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)} \tag{5·120}$$

By carrying through a similar transformation for the $H_z^{(2)}$ wave, we obtain

$$H_z^{(2)} = -C_1\frac{1}{2j}e^{j\omega t - j\frac{\omega}{v_1}x^{(2)}} = -C_1\frac{1}{2j}e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)} \tag{5·121}$$

where $H_z^{(2)}$ is expressed in terms of the propagation along the $X^{(2)}$ axis of Fig. 5·14.

H_z is the sum of equations 5·120 and 5·121. By resolving the E_y and E_x components in a similar way, we may obtain the following expressions for the E_1 wave:

$$\begin{aligned}
H_z &= \frac{C_1}{2j}e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)} - \frac{C_1}{2j}e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)} \\
E_x &= \frac{\pi C_1}{2j\omega\varepsilon_1 y_0}e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)} + \frac{\pi C_1}{2j\omega\varepsilon y_0}e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)} \\
E_y &= \underbrace{\frac{\beta_1 C_1}{2j\omega\varepsilon_1}e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)}}_{\substack{E_1 \text{ Wave} \\ \text{Group I}}} - \underbrace{\frac{\beta_1 C_1}{2j\omega\varepsilon_1}e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)}}_{\substack{E_1 \text{ Wave} \\ \text{Group II}}}
\end{aligned} \tag{5·122}$$

The E_1 wave may therefore be divided into two groups, one of which represents a wave traveling in the $X^{(1)}$ direction and the other a wave traveling in the $X^{(2)}$ direction. In both these groups **H** is transverse to the direction of propagation, being in the Z direction, which is normal to the XOY, $X^{(1)}OY^{(1)}$, or $X^{(2)}OY^{(2)}$ planes. It can be shown that the **E** vector in each wave is normal to its direction of propagation. Let us

resolve $E_y^{(1)}$ and $E_x^{(2)}$ of the first group into components along the normal to the $X^{(1)}$ direction. The graphical resolution of these vectors is shown in Fig. 5·16. The amplitude of the component of $E_x^{(1)}$ (indicated by the notation $(E_x^{(1)})'$ along the $X^{(1)}$ direction is

$$(E_x^{(1)})' \cos\theta = \frac{\pi C_1}{2j\omega\varepsilon_1 y_0} \cos\theta$$

Substituting for $\cos\theta$ from 5·118 we have

$$(E_x^{(1)})' \cos\theta = \frac{\pi C_1}{2j\omega\varepsilon_1 y_0} \beta_1 \left(\frac{v_1}{\omega}\right)$$

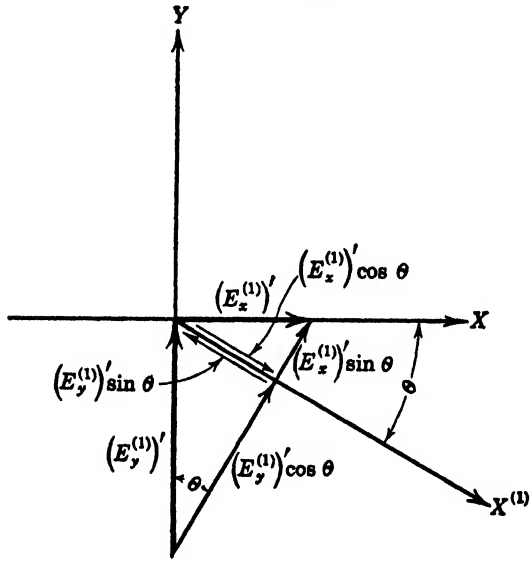


FIG. 5·16 Resolution of the elementary vector components $E_x^{(1)}$ and $E_y^{(1)}$ along and at right angles to the direction of propagation of the elementary wave of Group I. Instantaneous amplitudes are indicated.

Also the amplitude of the component of $E_y^{(1)}$ along the $X^{(1)}$ direction is

$$(E_y^{(1)})' \sin\theta$$

Substituting $\sin\theta$ from 5·118 we have

$$(E_y^{(1)})' \sin\theta = \frac{\beta_1 C_1}{2j\omega\varepsilon_1} \frac{\pi}{y_0} \left(\frac{v_1}{\omega}\right)$$

Thus it is seen that

$$(E_y^{(1)})' \sin\theta = (E_x^{(1)})' \cos\theta$$

From Fig. 5·16 it is clear that these two components are directed opposite to each other. Hence there is no $X^{(1)}$ component of electric intensity.

To obtain the component of $E^{(1)}$ normal to the direction of propagation, we see from Fig. 5·16 that

$$(E^{(1)}_{y_{(1)}})' = (E^{(1)}_x)' \sin\theta + (E^{(1)}_y)' \cos\theta \qquad [5\cdot123]$$

Thus

$$(E^{(1)}_{y_{(1)}})' = \frac{\pi C_1}{2j\omega\varepsilon_1 y_0}\left(\frac{\pi v_1}{\omega y_0}\right) + \frac{\beta_1 C_1}{2j\omega\varepsilon_1}\left(\frac{\beta_1 v_1}{\omega}\right)$$

$$= \frac{v_1 C_1}{2j\omega^2\varepsilon_1}\left\{\left(\frac{\pi}{y_0}\right)^2 + \beta_1^2\right\} \qquad [5\cdot124]$$

But

$$\beta_1^2 = \left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{y_0}\right)^2$$

from equation 5·86. Hence

$$(E^{(1)}_{y_{(1)}})' = \frac{C_1}{j2\varepsilon_1}\frac{1}{v_1} \qquad [5\cdot125]$$

From the first group we may now write the final expressions for the elementary waves as

$$\left.\begin{aligned} H^{(1)}_{z_{(1)}} &= \frac{C_1}{j2} e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)} \\ E^{(1)}_{y_{(1)}} &= \frac{C_1}{j2\varepsilon_1}\frac{1}{v_1} e^{j\omega\left(t-\frac{x^{(1)}}{v_1}\right)} \end{aligned}\right\} \begin{array}{l} E_1 \text{ elementary} \\ \text{wave in } X^{(1)} \\ \text{direction} \\ [5\cdot126] \end{array}$$

It can readily be shown by division that the ratio $E^{(1)}_{y_{(1)}}/H^{(1)}_{z_{(1)}} = \sqrt{\mu_1/\varepsilon_1}$, the value previously obtained for plane waves.

In a similar manner we may write for the second group

$$\left.\begin{aligned} H^{(2)}_{z_{(2)}} &= -\frac{C_1}{j2} e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)} \\ E^{(2)}_{y_{(2)}} &= -\frac{C_1}{j2\varepsilon}\frac{1}{v_1} e^{j\omega\left(t-\frac{x^{(2)}}{v_1}\right)} \end{aligned}\right\} \begin{array}{l} E_1 \\ \text{elementary} \\ \text{wave in } X^{(2)} \\ \text{direction} \\ [5\cdot127] \end{array}$$

Thus these elementary waves are plane waves having the electric and magnetic vectors at right angles to each other and to their direction of propagation, and having the velocity v_1 of a free wave in a medium of dielectric constant ε_1 and permeability μ_1. Figure 5·17 shows the direction which these waves take as they progress in the medium between the plates.

The equations 5·126 and 5·127 for the elementary waves are entirely

equivalent to equations 5·109 for the E_1 wave, from which they were obtained. We see, therefore, that the E_1 wave consists of the superposition of two elementary plane waves, which are reflected back and forth by the metal surfaces of the conducting bounding plates. Equations 5·126 may be considered to represent a wave striking the lower plate at $y = -y_0/2$ at an angle θ, and equation 5·127 may be considered to

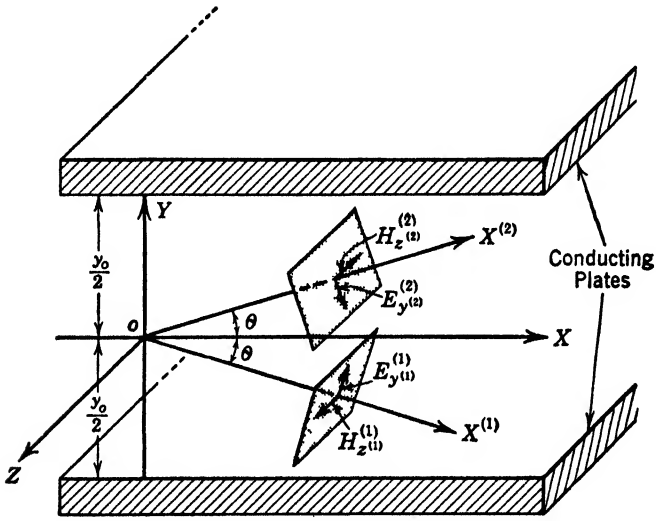


FIG. 5.17 Direction of propagation of elementary waves.

represent a wave being reflected from this lower surface at an angle θ and traveling toward the upper plate. At the upper plate, at $y = y_0/2$, the roles of the waves are interchanged. The angle θ which determines the directions $X^{(1)}$ and $X^{(2)}$ may be expressed as a function of the frequency of the waves. From Fig. 5·14,

$$\theta = \tan^{-1}\frac{\pi}{\beta_1 y_0} = \tan^{-1}\frac{1}{\sqrt{\left(\dfrac{f}{(f_0)_1}\right)^2 - 1}} \qquad [5\cdot128]^*$$

* From equation 5·86

$$\frac{\pi}{\beta_1 y_0} = \frac{1}{\dfrac{y_0}{\pi}\sqrt{\left(\dfrac{\omega}{v_1}\right)^2 - \left(\dfrac{\pi}{y_0}\right)^2}}$$

$$= \frac{1}{\sqrt{\left(\dfrac{\omega y_0}{\pi v_1}\right)^2 - 1}}$$

Thus, as the frequency f approaches the critical frequency $(f_0)_1 \tan\theta$ approaches ∞ and θ approaches 90°. At $\theta = 90°$, there is no transmission in the X direction and the waves bounce back and forth between the upper and lower plates. Thus the group velocity $(v_g)_1 = 0$ at $\theta = 90°$ while the phase velocity is infinite since the plane of the wave

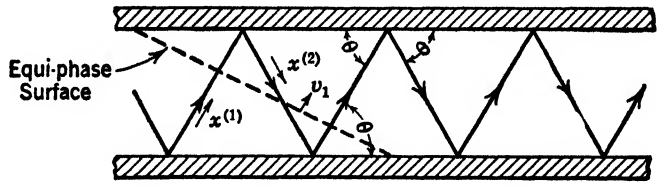


FIG. 5·18 Path of elementary waves when f is slightly larger than $(f_0)_1$.

is parallel to the surfaces of the plates, and two points an infinite distance from each other between the plates are in the equiphase surface of the waves. When f is slightly greater than $(f_0)_1$, there is a slow progression of the wave in the X direction (see Fig. 5·18). Thus the group velocity of the wave is low. Figure 5·19 shows the path of the elementary waves when f is several times larger then (f_0).

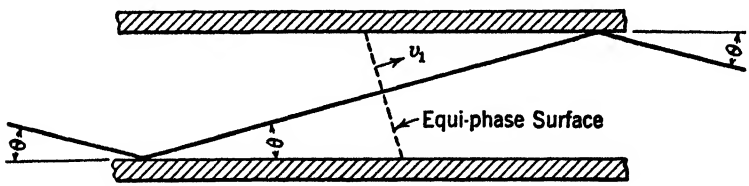


FIG. 5·19 Path of elementary waves when f is large in comparison to $(f_0)_1$.

The phase velocity, however, is quite high since equiphase surfaces of the wave progress along the plates at a much more rapid pace than the equiphase surfaces move in the $X^{(1)}$ or $X^{(2)}$ directions. This is a kind of shear action. Consider, for example, the velocity v_a of a point a on the blade of a paper-cutting machine (Fig. 5·20), compared to the velocity v_b of the point of contact b of the cutting edges. As ϕ ap-

---

We may write

$$\frac{\omega y_0}{\pi v_1} = \frac{2\pi f y_0}{\pi v_1} = \frac{f}{\dfrac{v_1}{2y_0}}$$

But $v_1/2y_0 = (f_0)_1$ (equation 5·87). Hence

$$\frac{\pi}{\beta_1 y_0} = \frac{1}{\sqrt{\left(\dfrac{f}{(f_0)_1}\right)^2 - 1}}$$

proaches 90°, the velocity v_b approaches infinity while the velocity v_a may be quite small. The velocity v_b is analogous to the phase velocity, and increases without limit as ϕ approaches 90°.

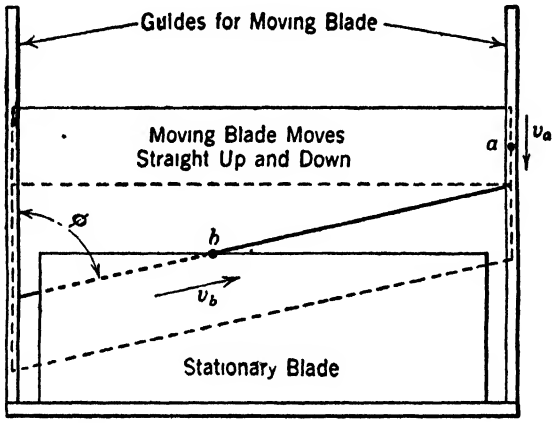


FIG. 5·20 Shear mechanism.

Returning to wave propagation between the plates, we can see that as f is increased without limit $\tan\theta$ approaches infinity and θ approaches zero. Thus the phase and group velocity approach the free wave velocity v_1.

5·16 Imperfectly Conducting Plates

As was previously stated, the conductivity of practical metals, although quite high, is never infinite. See Table 2·1, page 31, for a comparison of the conductivity of some common metals and other media. The problem of obtaining an exact solution of the wave equations in a region bounded by a metal of finite conductivity is extremely difficult, and as yet no complete solution has been found. By making certain assumptions, however, we can obtain quite satisfactory approximate values for the attenuation. It has been found, fortunately, that the results obtained by making these assumptions apply for most values of conductivity and frequency that are of practical importance.

Where the bounding plates are perfectly conducting the condition required at the surface of the plates is that the tangential component of the electric intensity vector **E** be zero there. All the waves so far discussed in this chapter meet this condition in that their derivation is based upon it. Further, when these waves are transmitted through the dielectric medium between the plates they suffer no attenuation; i.e.,

$\alpha = 0$. We can readily show that this is consistent with our concept of energy flow as obtained from the Poynting vector.

Let us recall the conditions set forth for the flow of energy by Poynting's concept. First, we know that the flow of energy in the electromagnetic field is directed at right angles to the plane in which the **E** and **H** vectors lie. Next, we know that the energy flows in the same direction that a right-hand screw would progress when we turn **E** into **H**. See Fig. 5·21. Further, if we reverse one or the other of these vectors,

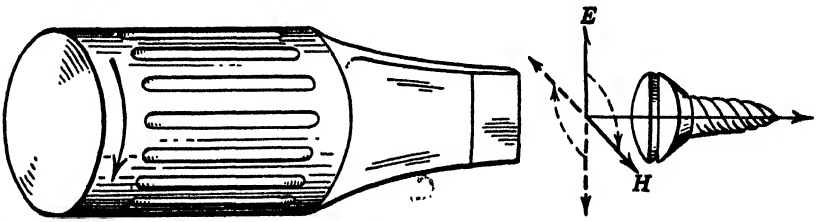


FIG. 5·21 The energy flow as given by the Poynting vector is in the same direction that a right-hand screw would progress when we turn E into H.

the direction of flow is reversed; if both are reversed, the direction of flow is the same. Finally, we know that the magnitude of the energy flowing across a unit area of surface per second is the product of the magnitude of the **E** and **H** vectors at the surface times the sine of the angle between them, this surface being parallel to the plane in which the **E** and **H** vectors lie.

Now, for the waves discussed in the previous sections, only the tangential component of the **H** vector exists at the surface of the bounding plates, i.e., $H_{tan} \neq 0$ at $y = \pm y_0/2$, while in every case $E_{tan} = 0$ at $y = \pm y_0/2$. Therefore there can be no energy flow into the boundary plates, and all the energy is transmitted through the dielectric between the plates, without loss. In practice, where the conductivity is high but finite, a small tangential component of **E** will exist at the surface of the plates where $y = \pm y_0/2$. Under these circumstances energy will flow into the bounding plates, given by the relation

$$\tfrac{1}{2}\mathcal{R}(\mathbf{E}_{tan} \times \mathbf{H}^*_{tan})$$

as shown in Chapter 3. This energy will be subtracted from the wave in the dielectric. Thus in the practical example the wave will suffer a diminution of energy as it is transmitted through the dielectric. This loss of energy may be expressed in terms of the attenuation factor

$e^{-\alpha x}$, appearing in equation 5·42, where α is given by the relation

$$\alpha = \frac{1}{2} \frac{\text{Power loss per meter of length}}{\text{Power transmitted through the dielectric}} \quad [5\cdot129]$$

*

In the previous examples we have solved the wave equations subject to the boundary conditions that **E** tangential is equal to zero at the boundary, this being a necessary condition when σ is infinite. If we attempt to solve them for the condition that **E** tangential is equal to some finite but small value at the boundary, and in this way determine α, the problem becomes extremely difficult. A method for obtaining α which we shall use here is based upon certain simplifying assumptions which are sufficiently accurate for practical purposes and which lead to a result that has been confirmed experimentally.

Let us assume that the distribution of the field vectors when the plates are imperfect conductors is not materially changed from that which is obtained when the boundary plates are perfectly conducting. The imperfect conductivity permits a finite dissipation of energy in the metal. The current which flows is accompanied by the presence of a component of electric intensity in the direction of the current. We will assume that this is the only change in the field distribution in the dielectric. The existence of this component results in a tipping forward of the electric vector in the neighborhood of the boundary and therefore in the presence of an electric vector $E_{\tan}$ at the boundary. We do not know the magnitude of this vector component, but we may estimate it in the following way. First, we know the value of the tangential component of **H**, i.e., $H_{\tan}$ at the boundary, and we know that it is continuous at the boundary. From a solution of the wave equations in a metallic medium we may obtain a relation between the **E** and **H** vectors which must hold wherever $\sigma \gg \omega\epsilon$. Since our metallic boundaries do meet this condition, and since we know $H_{\tan}$ at the surface of the metal, we may establish from this relation the value of $E_{\tan}$ at the surface of the metal. Then by means of the Poynting vector we may calculate the flow of energy across the surface formed by the dielectric and the metal, and thus deduce the attenuation constant α. Let us first solve the wave equations in an infinite metallic medium of conductivity σ and obtain a relation between **E** and **H** in this medium.

5·17 Wave Propagation in a Metal

Let us consider the propagation of a plane wave in the Y direction in a metallic medium of infinite extent having a finite conductivity σ_2. Let us choose **H** in the Z direction so that $H = H_z$ and $H_x = H_y = 0$.

* A derivation of this relation is given on page 165.

Further

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial x} = 0$$

since we are dealing with a plane wave in a medium of infinite extent. Then the wave equations

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \sigma E_x + \varepsilon \frac{\partial E_x}{\partial t} \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \sigma E_y + \varepsilon \frac{\partial E_y}{\partial t} \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \sigma E_z + \varepsilon \frac{\partial E_z}{\partial t} \end{aligned}\right\} \quad [3\cdot12] \qquad \left.\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu \frac{\partial H_x}{\partial t} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu \frac{\partial H_y}{\partial t} \\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu \frac{\partial H_z}{\partial t} \end{aligned}\right\} \quad [3\cdot13]$$

reduce to

$$\frac{\partial H_z}{\partial y} = \sigma_2 E_x + \varepsilon_2 \frac{\partial E_x}{\partial t} \qquad [5\cdot130]$$

and

$$\frac{\partial E_x}{\partial y} = \mu_2 \frac{\partial H_z}{\partial t} \qquad [5\cdot131]$$

where $E_y = E_z = 0$ and $E = -E_x$. The subscript 2 refers to the constants of the metallic medium. The geometrical configuration of this wave is shown in Fig. 5·22.

Let

$$H_z = H_z' e^{j\omega t - \gamma y} \qquad [5\cdot132]$$

$$E_x = E_x' e^{j\omega t - \gamma y} \qquad [5\cdot133]$$

where $\gamma = \alpha + j\beta$.* Substituting H_z and E_x in equations 5·130 and 5·131 and dividing by $e^{j\omega t - \gamma y}$ we obtain

$$-\gamma H_z' = (\sigma_2 + j\omega\varepsilon_2) E_x' \qquad [5\cdot134]$$

$$-\gamma E_x' = j\omega\mu_2 H_z' \qquad [5\cdot135]$$

The propagation constant γ may be obtained directly by substituting for H_z in equation 5·134 its value as obtained from equation 5·135. Thus

$$\frac{\gamma^2}{j\omega\mu_2} E_x' = (\sigma_2 + j\omega\varepsilon_2) E_x' \qquad [5\cdot136]$$

giving

$$\gamma^2 = j\omega\mu_2(\sigma_2 + j\omega\varepsilon_2) \qquad [5\cdot137]$$

* The value of α given in this equation defines the attenuation of the wave in the metal in the Y direction. It is not the same as that of equation 5·129.

If $\sigma_2 \gg \omega\epsilon_2$

$$\gamma = \sqrt{j\omega\mu_2\sigma_2} = (1+j)\sqrt{\frac{\omega\mu_2\sigma_2}{2}} = (1+j)\sqrt{\pi f\mu_2\sigma_2} \qquad [5\cdot138]$$

since

$$(1+j)^2 = 2j \qquad \text{or} \qquad \sqrt{j} = \frac{(1+j)}{\sqrt{2}}$$

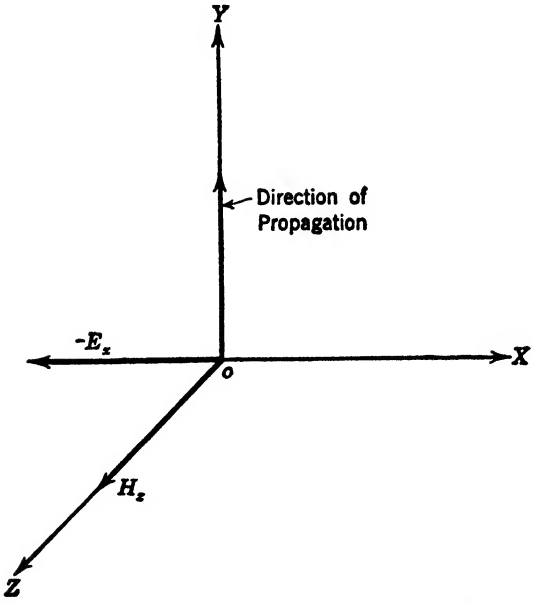


FIG. 5·22 Propagation of a wave into metal

Now

$$\gamma = \alpha + j\beta = \sqrt{\pi f\mu_2\sigma_2} + j\sqrt{\pi f\mu_2\sigma_2}$$

Hence

$$\alpha = \beta = \sqrt{\pi f\mu_2\sigma_2} \qquad [5\cdot139]$$

Thus, in the metal

$$H = H_z' e^{-(\sqrt{\pi f\mu_2\sigma_2})y}\, e^{j[\omega t-(\sqrt{\pi f\mu_2\sigma_2})y]} \qquad [5\cdot140]$$

$$E = E_x' e^{-(\sqrt{\pi f\mu_2\sigma_2})y}\, e^{j[\omega t-(\sqrt{\pi f\mu_2\sigma_2})y]} \qquad [5\cdot141]$$

and therefore the amplitude of the E and H waves is decreased at a rate determined by the attenuation factor $e^{-(\sqrt{\pi f\mu_2\sigma_2})y}$ as the wave progresses in the Y direction. This factor is plotted in Fig. 5·23

as a function of y. It is assumed in the graph that μ_2 and σ_2 are constant as the frequency is varied. This is not true in general but holds well up to frequencies of the order of 10^{10} cycles per second.

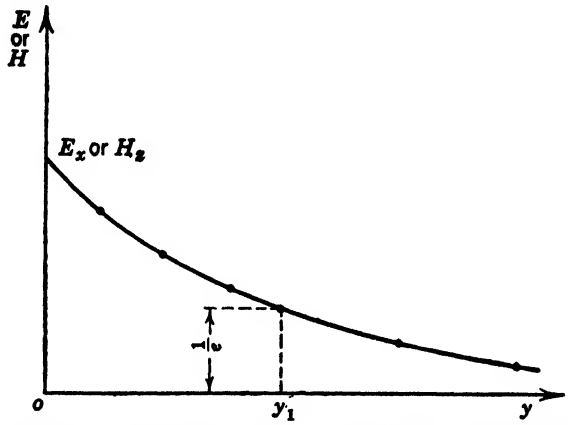


FIG. 5·23 Reduction of the amplitude of the wave as it penetrates the metal in the Y direction.

A convenient measure of the penetrating power of the wave is the value of y which reduces the exponent αy to unity. Let us call this value y_1. Then $\alpha y_1 = 1$ and $y_1 = 1/\alpha$. Thus at a distance y_1 the wave amplitude will be reduced to e^{-1} or $1/e = 0.368$ of its value at $y = 0$.

The phase velocity v_p in the metal is given as

$$v_p = \frac{\omega}{\beta} = \frac{\omega}{\sqrt{\pi f \mu_2 \sigma_2}} = \sqrt{\frac{2\omega}{\mu_2 \sigma_2}} \qquad [5\cdot142]$$

Values of α, y_1, and v_p are given as functions of frequency in Table 5·1 for copper.

TABLE 5·1

Frequency, cycles per second	$\alpha = \beta = \sqrt{\pi f \mu_2 \sigma_2}$	y_1 meters	v_p, meters per second
0	0		0
1	15.13	0.066	0.416
10^2	151.3	0.0066	4.16
10^6	15,130	0.000066	416

The intrinsic impedance η defined in terms of the intensity vectors is $\eta = \mathbf{E}/\mathbf{H}$. Hence from 5·135

$$\eta = \frac{-E_x'}{H_z'} = \frac{j\omega\mu_2}{\gamma} \qquad [5\cdot143]$$

From 5·138

$$\gamma = \sqrt{j\omega\mu_2\sigma_2}$$

when $\sigma_2 \gg \omega\varepsilon_2$. Hence

$$\eta = \frac{j\omega\mu_2}{\sqrt{j\omega\mu_2\sigma_2}} = \sqrt{\frac{j\omega\mu_2}{\sigma_2}} = \sqrt{\frac{\pi f\mu_2}{\sigma_2}} + j\sqrt{\frac{\pi f\mu_2}{\sigma_2}} \qquad [5\cdot144]$$

giving

$$-E_x' = \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, H_z' \qquad [5\cdot145]$$

Since $\mathbf{E} = \eta\mathbf{H}$, this indicates that **E** and **H** are 45° out of time and space phase in a metal in which $\sigma_2 \gg \omega\varepsilon_2$. Since this condition is met in our boundary-value problem involving the infinite metal planes separated by a dielectric, we may use this value of η in establishing the electric intensity at the surface of the plates. Let us consider first the attenuation associated with the transverse electromagnetic wave.

5·18 The Transverse Electromagnetic Mode (*TEM*) Imperfect Conductor

The equations defining the transverse electromagnetic transmission mode may be written from equations 5·35 subject to the restrictions imposed by relations 5·34; i.e., $C_1 = 0$, $n = 0$.

$$\left.\begin{aligned} H_{z1} &= C_2 e^{-\gamma x + j\omega t} = C_2 e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \\ E_{y1} &= C_2\sqrt{\frac{\mu_1}{\varepsilon_1}}\, e^{-\gamma x + j\omega t} = C_2\sqrt{\frac{\mu_1}{\varepsilon_1}}\, e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \end{aligned}\right\} \qquad [5\cdot146]^*$$

These equations have already been discussed in section 5·7 for the case of perfect conductivity ($\sigma_2 = \infty$), where $\alpha = 0$. We will now determine the attenuation constant α in the case of imperfect conductivity.

Consider the conditions represented diagrammatically in Fig. 5·24, where t is held constant. The magnetic intensity is constant in the Y direction, and therefore its value at the surface of either upper or lower plate is given in equation 5·146. If a value of $-E_x$ exists at the upper surface of the metal, so that energy will flow into the metal, then the value of $-E_x$ will be given by 5·145. Thus at the surface of the upper

* It will now be desirable to distinguish between waves in the metal and waves in the dielectric. To do this we may write H_{z1} and E_{y1} for the waves in the dielectric material. Then H_{z2}, etc., will refer to waves in the metal.

plate

$$\left.\begin{aligned} H_{z2} &= H_{z1} = C_2 e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \\ -E_{x2} &= \eta H_{z2} = C_2 \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \end{aligned}\right\} \begin{array}{l}\textit{TEM}\\ \text{wave at}\\ \text{surface of}\\ \text{upper plate}\\ [5\cdot147]\end{array}$$

And at the surface of the lower plate

$$\left.\begin{aligned} H_{z2} &= H_{z1} = C_2 e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \\ E_{x2} &= \eta H_{z2} = C_2 \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, e^{-\alpha x} e^{j(\omega t - \beta_0 x)} \end{aligned}\right\} \begin{array}{l}\textit{TEM}\\ \text{wave at}\\ \text{surface of}\\ \text{lower plate}\\ [5\cdot148]\end{array}$$

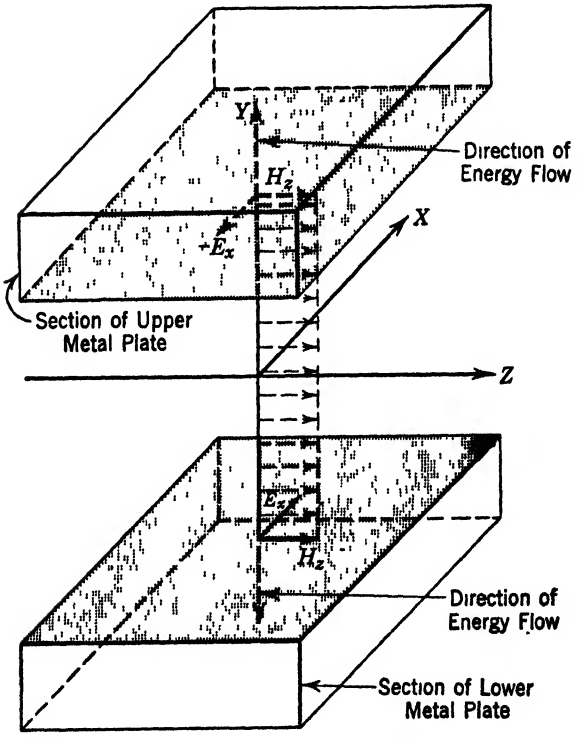


FIG. 5·24 Electric intensity vectors at the surface of a guide of imperfect conductivity. *TEM* wave.

We may now calculate the power flow associated with this wave. Let us call the power flowing into the upper plate, per second per square meter of surface, p_y, and that flowing into the lower plate p'_y. Then

$$p_y = \tfrac{1}{2}\mathcal{R}(E_{z2}H^*_{x2} - E_{x2}H^*_{z2}) = \tfrac{1}{2}\mathcal{R}(-E_{x2}H^*_{z2}) \qquad [5\cdot149]$$

Substituting for E_{x2} and H_{z2} from equations 5·147 we have

$$p_y = \tfrac{1}{2}C_2^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot150]^*$$

The power loss over an area of the metal surface 1 meter wide in the X direction and z_0 meters long in the Z direction is p_y times z_0 or

$$p_y z_0 = \frac{z_0}{2}\,C_2^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot151]$$

In a similar manner we may calculate the power loss over the same area of the lower plate. It likewise is

$$p_y' z_0 = \frac{z_0}{2}\,C_2^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot152]$$

The power loss P_y over the same area of both plates is then

$$P_y = z_0(p_y + p_y') = C_2^2 z_0\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot153]$$

This power is a function of the distance the wave has progressed between the plates in the X direction. This we would expect since the amplitude of the **E** and **H** vectors in the dielectric are being continuously reduced as the wave progresses. Hence the power which the wave delivers to the plates in its progress between them is decreased with distance.

Now let us calculate the flow of power across a surface in the dielectric normal to the plane of the wave. It is given by

$$p_x = \tfrac{1}{2}(E_{y1}'H_{z1}' - E_{z1}'H_{y1}') = \tfrac{1}{2}(E_{y1}'H_{z1}') \qquad [5\cdot154]$$

where p_x is the power flow per second per square meter of surface. Hence

$$p_x = \tfrac{1}{2}C_2^2\sqrt{\frac{\mu_1}{\varepsilon_1}}\,e^{-2\alpha x} \qquad [5\cdot155]$$

and the power passing through a surface of the dielectric z_0 meters in the Z direction and y_0 meters in the Y direction is $p_x z_0 y_0$. Thus

$$P_x = \tfrac{1}{2}C_2^2 z_0 y_0\sqrt{\frac{\mu_1}{\varepsilon_1}}\,e^{-2\alpha x} \qquad [5\cdot156]$$

Now the power loss in the metal boundary is equal to the rate of diminution of the power flowing through the dielectric. Since P_x is the power flowing through an area $z_0 y_0$ normal to the Poynting vector, $-\partial P_x/\partial x$

* Only the real part of the radical as given in equation 5·144 is used as indicated by the symbol $\mathfrak{R}$ in equation 5·149.

is the rate of decrease of the power. See Fig. 5·25. This must be equal to P_y, the power lost by the wave in the upper and lower plates. Thus

$$P_y = -\frac{\partial P_x}{\partial x} \qquad [5\cdot157]$$

Differentiating equation 5·156 with respect to x

$$\frac{\partial P_x}{\partial x} = -2\alpha\left(\tfrac{1}{2}C_2^2 y_0 z_0 \sqrt{\frac{\mu_1}{\varepsilon_1}}\, e^{-2\alpha x}\right) = -2\alpha P_x \qquad [5\cdot158]$$

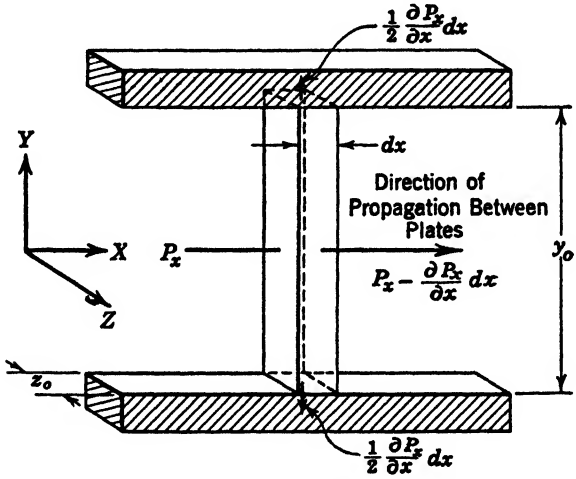


FIG. 5·25 Development of the attenuation constant α.

Hence

$$P_y = 2\alpha P_x \qquad [5\cdot159]$$

giving

$$\alpha = \tfrac{1}{2}\frac{P_y}{P_x} \qquad [5\cdot160]$$

Thus

$$\alpha = \frac{C_2^2 z_0 \sqrt{\dfrac{\pi f \mu_2}{\sigma_2}}\, e^{-2\alpha x}}{C_2^2 z_0 y_0 \sqrt{\dfrac{\mu_1}{\varepsilon_1}}\, e^{-2\alpha x}} \qquad [5\cdot161]$$

or

$$\alpha = \frac{1}{y_0}\sqrt{\frac{\pi f \mu_2 \varepsilon_1}{\sigma_2 \mu_1}} \quad \text{nepers per meter} \qquad [5\cdot162]$$

5·19 The E_1 Wave (TM_1 Mode) Imperfect Conductor

The equations for the first-order transverse magnetic mode of transmission, as obtained from the general equations 5·35 subjected to the restrictions of 5·34, i.e., $n = 1$, $C_2 = 0$, are:

$$\left.\begin{aligned} H_{z1} &= C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \\ E_{x1} &= \frac{1}{j\omega\varepsilon_1}\left(\frac{\pi}{y_0}\right) C_1 \cos\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \\ E_{y1} &= \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - (\pi/y_0)^2}}{\omega\varepsilon_1} C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \end{aligned}\right\} \begin{array}{l}\text{E_1 waves} \\ \text{TM_1 mode in the} \\ \text{dielectric} \\ \text{[5·163]}\end{array}$$

The geometric configuration of the H_{z1} vector between a section of the parallel planes is shown in Fig. 5 26, where t is held constant. The magnetic intensity vector varies sinusoidally in the Y direction and has its maximum value at the upper and lower metal surfaces. The value of E_{x2} at the surface of the metal is given by

$$-E_{x2} = \eta H_{z1} = \eta H_{z2}$$

since the tangential component of the magnetic intensity is continuous at the boundary. Hence

$$-E_{x2} = \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, H_{z1} \qquad [5\cdot164]$$

$$= \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \qquad [5\cdot165]$$

At the metal surface

$$\left.\begin{aligned} H_{z2} &= C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \\ -E_{x2} &= \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, C_1 \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \end{aligned}\right\} \begin{array}{l}\text{E_1 or TM_1} \\ \text{wave at the} \\ \text{boundary} \\ \text{[5·166]}\end{array}$$

The power flow into the metal per second per square meter of surface of the upper plate is

$$p_y = \tfrac{1}{2}\mathcal{R}(E_{x2}H_{x2}^* - E_{x2}H_{x2}^*) = \tfrac{1}{2}\mathcal{R}(-E_{x2}H_{x2}^*) \qquad [5\cdot167]$$

At the upper surface $y = y_0/2$, so that

$$\sin^2\left(\frac{\pi}{y_0} y\right) = \sin^2\left(\frac{\pi}{2}\right) = 1$$

Hence at the upper surface

$$p_y = \tfrac{1}{2}\mathcal{R}\left[C_1^2\sqrt{\frac{j\omega\mu_2}{\sigma_2}}\,e^{-2\alpha x}\right] \qquad [5\cdot168]$$

$$= \tfrac{1}{2}C_1^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot169]$$

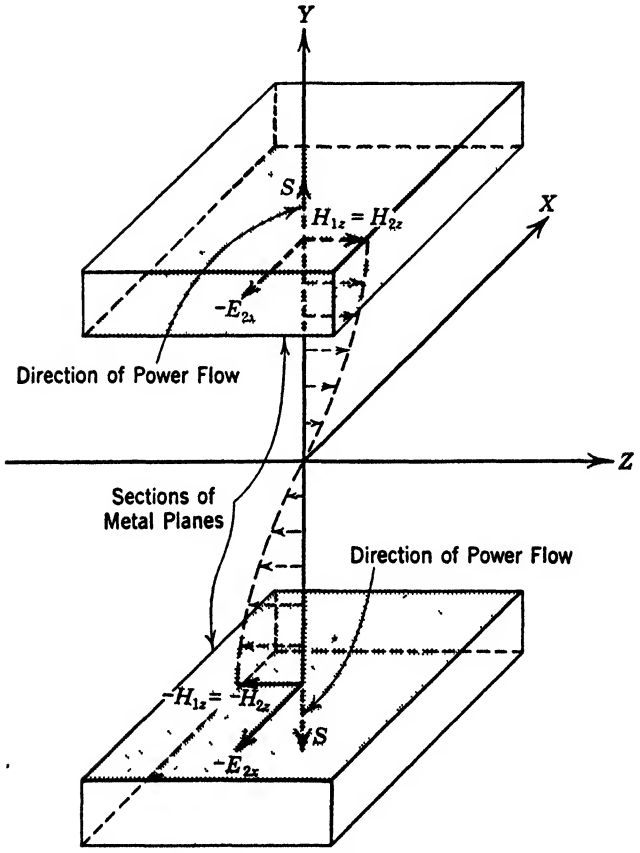


FIG. 5·26 Electric intensity vectors at the surface of a guide of imperfect conductivity. TM_1 wave.

The power flowing into an area z_0 meters long in the Z direction and 1 meter wide in the X direction is z_0 times p_y. Hence the power flowing into this area is

$$\frac{z_0}{2}\,C_1^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,e^{-2\alpha x} \qquad [5\cdot170]$$

In a similar way we may calculate the power p'_y flowing into a unit area in the lower plate. These are equal, so that the total power flowing into the upper and lower plates over a surface area 1 meter wide and z_0 meters long in each is

$$P_y = z_0(p_y + p'_y) = C_1^2 z_0 \sqrt{\frac{\pi f \mu_2}{\sigma_2}}\, e^{-2\alpha x} \qquad [5\cdot171]$$

The power flowing through a square meter of surface per second in the dielectric normal to the direction of propagation of the wave is

$$p_x = \tfrac{1}{2}(E'_{y1}H'_{z1} - E'_{z1}H'_{y1}) = \tfrac{1}{2}(E'_{y1}H'_{z1}) \qquad [5\cdot172]$$

Now

$$E'_{y1}H'_{z1} = C_1^2 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\varepsilon_1} \sin^2\left(\frac{\pi}{y_0} y\right) e^{-2\alpha x} \qquad [5\cdot173]$$

Hence

$$P_x = \frac{C_1^2}{2} \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\varepsilon_1} \sin^2\left(\frac{\pi}{y_0} y\right) e^{-2\alpha x} \qquad [5\cdot174]$$

giving the power flow per square meter per second. The power flowing across a surface z_0 meters in the Z direction and y_0 meters in the Y direction may be cxpressed as

$$P_x = \frac{C_1^2}{2}\left(\frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\varepsilon_1}\right) e^{-2\alpha x} \int_{z=-z_0/2}^{z=z_0/2} \int_{y=-y_0/2}^{y=y_0/2} \sin^2\left(\frac{\pi}{y_0} y\right) dy\, dz \qquad [5\cdot175]$$

$$= \frac{C_1^2}{4} y_0 z_0 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\varepsilon_1} e^{-2\alpha x} \qquad [5\cdot176]^*$$

This gives

$$\alpha = \frac{1}{2}\frac{P_y}{P_x} = \frac{1}{2} \frac{C_1^2 z_0 \sqrt{\frac{\pi f \mu_2}{\sigma_2}}\, e^{-2\alpha x}}{\frac{C_1^2}{4} y_0 z_0 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\varepsilon_1} e^{-2\alpha x}} \qquad [5\cdot177]$$

* $\displaystyle\int_{z=-z_0/2}^{z=z_0/2} \int_{y=-y_0/2}^{y=y_0/2} \sin^2\left(\frac{\pi}{y_0} y\right) dy\, dz = \int_{z=-z_0/2}^{z=z_0/2} \frac{y_0}{2}\, dz = \frac{y_0 z_0}{2}$

Hence for the E_1 wave

$$\alpha = 2\frac{\dfrac{1}{y_0}\sqrt{\dfrac{\pi f\mu_2}{\sigma_2}}}{\dfrac{1}{\omega\varepsilon_1}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\dfrac{\pi}{y_0}\right)^2}} \quad \text{nepers per meter} \qquad [5\cdot178]^*$$

or

$$\alpha = y_0^{-3/2}\sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}\left\{\frac{\left(\dfrac{f}{(f_0)_1}\right)^{3/2}}{\sqrt{\left(\dfrac{f}{(f_0)_1}\right)^2 - 1}}\right\} \text{ nepers per meter} \qquad [5\cdot179]$$

5·20 The H_1 Wave (TE_1 Mode) Imperfect Conductor

The equations for the first order-transverse electric mode of transmission as obtained from the general equation 5·41 subject to the restric-

* If we multiply numerator and denominator of this equation by $\sqrt{\varepsilon_1/\mu_1}$ the numerator becomes

$$\frac{1}{y_0}\sqrt{\frac{\pi f\mu_2\varepsilon_1}{\sigma_2\mu_1}} = \frac{1}{y_0}\sqrt{\frac{\pi f\mu_2(\mu_1\varepsilon_1)}{\sigma_2\mu_1^2}} = \sqrt{\frac{\pi f\mu_2}{y_0^2v_1^2\sigma_2\mu_1^2}} \qquad [5\cdot178a]$$

But the critical frequency $(f_0)_1 = v_1/2y_0$ for the E_1 wave. Hence we may write 5.178a as

$$\sqrt{\frac{\pi\mu_2}{2v_1y_0^3\sigma_2\mu_1^2}\left(\frac{f}{(f_0)_1}\right)} = y_0^{-3/2}\,\tfrac{1}{2}\sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}\left(\frac{f}{(f_0)_1}\right)^{1/2} \qquad [5\cdot178b]$$

The denominator becomes

$$\frac{1}{\omega\sqrt{\mu_1\varepsilon_1}}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2} = \sqrt{1 - \left(\frac{(f_0)_1}{f}\right)^2} = \frac{(f_0)_1}{f}\sqrt{\left(\frac{f}{(f_0)_1}\right)^2 - 1} \qquad [5\cdot178c]$$

Thus

$$\alpha = 2\frac{y_0^{-3/2}\,\tfrac{1}{2}\sqrt{\dfrac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}\left(\dfrac{f}{(f_0)_1}\right)^{1/2}}{\dfrac{(f_0)_1}{f}\sqrt{\left(\dfrac{f}{(f_0)_1}\right)^2 - 1}} \qquad [5\cdot178d]$$

$$= y_0^{-3/2}\sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}\left\{\frac{\left(\dfrac{f}{(f_0)_1}\right)^{3/2}}{\sqrt{\left(\dfrac{f}{(f_0)_1}\right)^2 - 1}}\right\} \qquad [5\cdot178e]$$

tions of 5·38, i.e., $n = 1$, $C_3 = 0$, are

$$E_{z1} = C_4 \cos\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)}$$

$$H_{x1} = C_4 \frac{1}{j\omega\mu_1} \left(\frac{\pi}{y_0}\right) \sin\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)}$$

$$H_{y1} = -C_4 \frac{\sqrt{\omega^2 \mu_1 \varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\mu_1} \cos\left(\frac{\pi}{y_0} y\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)}$$

} H_1 or TE_1 wave in the dielectric [5·180]

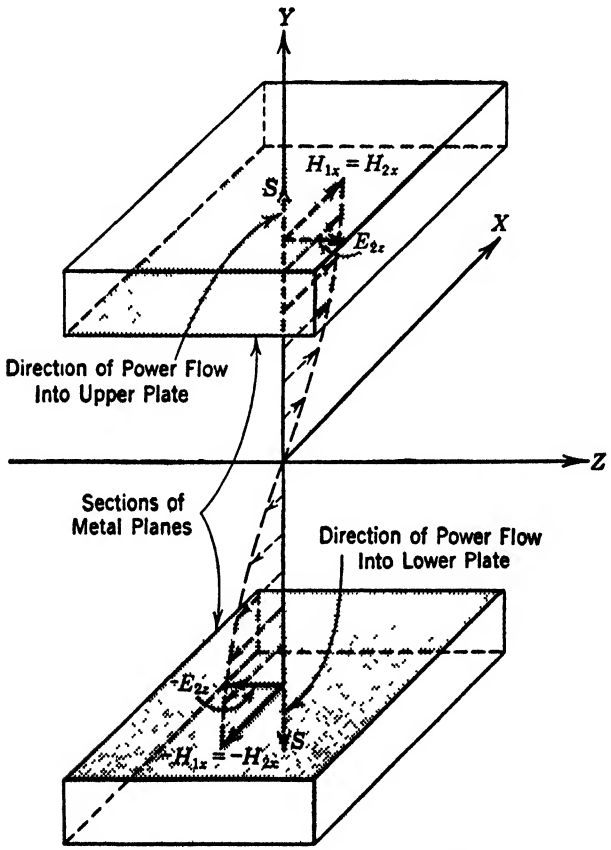


FIG. 5·27 Electric intensity vectors at the surface of a guide of imperfect conductivity. TE_1 wave.

A study of the geometric configuration of this wave between the plates, when time is held fixed, shows the existence of a component of

$\mathbf{H}_{\tan}$, i.e., H_{x1} at the boundary. See Fig. 5·27. The H_{y1} component is zero there. The value of E_{z2} that must exist at the boundary is given by

$$E_{z2} = \eta H_{x1} \tag{5·181}$$

or

$$E_{z2} = \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, H_{x1}$$

But at the boundaries $y = y_0/2$ and $\sin \pi/2 = 1$. Accordingly

$$\left.\begin{aligned} E_{z2} &= \sqrt{\frac{j\omega\mu_2}{\sigma_2}}\, C_4 \frac{1}{j\omega\mu_1}\left(\frac{\pi}{y_0}\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \\ H_{x2} &= \frac{C_4}{j\omega\mu_1}\left(\frac{\pi}{y_0}\right) e^{-\alpha x} e^{j(\omega t - \beta_1 x)} \end{aligned}\right\} \begin{array}{l} H_1 \text{ or } TE_1 \\ \text{wave at the} \\ \text{metal surface} \\ [5·182] \end{array}$$

The power flow into the metal per second per square meter of surface of the upper plate is

$$p_y = \tfrac{1}{2}\mathcal{R}(E_{z2}H_{x2}^* - E_{x2}H_{z2}^*) = \tfrac{1}{2}\mathcal{R}(E_{z2}H_{x2}^*) \tag{5·183}$$

Now

$$E_{z2}H_{x2}^* = C_4^2 \frac{1}{\omega^2\mu_1^2}\sqrt{2j}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2 e^{-2\alpha x} \tag{5·184}$$

Since $\sqrt{2j} = (1 + j)$ the real part of $E_{z2}H_{x2}^*$ is

$$C_4^2 \frac{1}{\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2 e^{-2\alpha x} \tag{5·185}$$

and

$$p_y = \tfrac{1}{2}C_4^2 \frac{1}{\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2 e^{-2\alpha x} \tag{5·186}$$

The power entering a rectangular area in the upper plate of length z_0 in the Z direction and 1 meter in the X direction is z_0 times p_y. The power p_y' entering the lower plate across an equal area directly below this one may be calculated in a similar manner. This power is also equal to $z_0 p_y$. Thus the total power loss is

$$z_0(p_y + p_y') = 2z_0 p_y$$

or

$$P_y = C_4^2 \frac{z_0}{\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2 e^{-2\alpha x} \tag{5·187}$$

The power flowing per second through a square meter of surface of the

dielectric normal to the direction of propagation of the wave is

$$p_x = \tfrac{1}{2}(-E'_{z1}H'_{y1}) \qquad [5\cdot188]$$

$$-E'_{z1}H'_{y1} = C_4^2 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\mu_1} \cos^2\left(\frac{\pi}{y_0}y\right) e^{-2\alpha x} \qquad [5\cdot189]$$

Hence

$$p_x = \tfrac{1}{2}C_4^2 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\mu_1} \cos^2\left(\frac{\pi}{y_0}y\right) e^{-2\alpha x} \qquad [5\cdot190]$$

The power flowing across an arca z_0 meters in the Z direction and y_0 meters in the Y direction may be expressed as

$$P_x = \tfrac{1}{2}C_4^2 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\mu_1} e^{-2\alpha x} \int_{z=-z_0/2}^{z=z_0/2} \int_{y=-y_0/2}^{y=y_0/2} \cos^2\left(\frac{\pi}{y_0}y\right) dy\, dz \qquad [5\cdot191]$$

$$P_x = \tfrac{1}{4}y_0 z_0 C_4^2 \frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}{\omega\mu_1} e^{-2\alpha x} \qquad [5\cdot192]$$

Since

$$\int_{z=-z_0/2}^{z=z_0/2} \int_{y=-y_0/2}^{y=y_0/2} \cos^2\left(\frac{\pi}{y_0}y\right) dy\, dz = \int_{z=-z_0/2}^{z=z_0/2} \frac{y_0}{2}\, dz = \frac{z_0 y_0}{2}$$

The attenuation factor for the H_1 wave is then

$$\alpha = \frac{1}{2}\frac{P_y}{P_x} = \frac{1}{2}\frac{C_4^2 \frac{z_0}{\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2 e^{-2\alpha x}}{\frac{1}{4}C_4^2 \frac{y_0 z_0}{\omega\mu_1}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2} e^{-2\alpha x}} \qquad [5\cdot193]$$

$$= 2\frac{\frac{1}{y_0\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{\pi}{y_0}\right)^2}{\frac{1}{\omega\mu_1}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}} \quad \text{nepers per meter} \qquad [5\cdot194]^*$$

* If we multiply numerator and denominator of equation 5·194 by μ_1/ε_1 the equation may be written

$$\alpha = 2\frac{\frac{1}{y_0}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}}{\frac{1}{\omega\varepsilon_1}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2}}\left[\left(\frac{\pi}{y_0}\right)^2 \frac{1}{\omega^2\mu_1\varepsilon_1}\right] \qquad [5\cdot194a]$$

$$\alpha = y_0^{-3/2}\sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}} \left\{\frac{\left(\frac{f}{(f_0)_1}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_1}\right)^2 - 1}}\right\} \text{ nepers per meters} \quad [5\cdot195]$$

5·21 Attenuation of the H_n Wave (TE_n Mode) Imperfect Conductor

The equations for the nth order (n odd) transverse electric mode of transmission give, for the power lost in the upper and lower plates per square meter of surface,

$$P_{y(n \text{ odd})} = \tfrac{1}{2}C_4^2\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left\{\frac{1}{\omega^2\mu_1^2}\left(\frac{n\pi}{y_0}\right)^2 e^{-2\alpha x}\right\} \quad [5\cdot196]$$

The power flowing across an area 1 meter wide in the Z direction and y_0 meters in the Y direction, in the dielectric, is

$$P_{x(n \text{ odd})} = \tfrac{1}{2}C_4^2\frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2}}{\omega\mu_1} e^{-2\alpha x}\int_{y=-y_0/2}^{y=y_0/2}\cos^2\left(\frac{n\pi}{y_0}y\right)dy \quad [5\cdot197]$$

$$= \frac{y_0}{4}C_4^2\frac{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2}}{\omega\mu_1} e^{-2\alpha x} \quad [5\cdot198]$$

Since

$$\int_{y=-y_0/2}^{y=y_0/2}\cos^2\left(\frac{n\pi}{y_0}y\right)dy = \frac{y}{2} \quad [5\cdot199]$$

Thus

$$\alpha_{(n \text{ odd})} = \frac{1}{2}\frac{P_y}{P_x} = \frac{\frac{1}{y_0\omega^2\mu_1^2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left(\frac{n\pi}{y_0}\right)^2}{\frac{1}{\omega\mu_1}\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2}} \text{ nepers per meter} \quad [5\cdot200]$$

It is seen that the first term of this relation is the same as equation 5·178 obtained for the E_1 wave. The second term may be written

$$\left(\frac{\pi}{y_0}\right)^2\frac{1}{\omega^2\mu_1\varepsilon_1} = \left(\frac{\pi}{y_0}\cdot\frac{v_1}{\omega}\right)^2 = \left(\frac{\pi}{y_0}\cdot\frac{v_1}{2\pi f}\right)^2 = \left(\frac{v_1}{2y_0}\right)^2\left(\frac{1}{f}\right)^2 \quad [5\cdot194b]$$

The critical frequency $(f_0)_1 = v_1/2y_0$ for the H_1 wave. Thus

$$\left(\frac{\pi}{y_0}\right)^2\frac{1}{\omega^2\mu_1\varepsilon_1} = \left(\frac{(f_0)_1}{f}\right)^2 \quad [5\cdot194c]$$

Hence equation 5·194 may be written as equation 5·179 multiplied by $(f/(f_0)_1)^{-2}$.

It may be shown that the attenuation for the nth order H_n wave, n even, is the same as given by equation 5·200.

After algebraic manipulation similar to that employed previously, the attenuation constant for the H_n wave may be expressed as

$$\alpha = y_0^{-3/2}\sqrt{\frac{2\pi\mu_2 n}{v_1\sigma_2\mu_1^2}}\left\{\frac{\left(\frac{f}{(f_0)_n}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_n}\right)^2-1}}\right\} \quad \text{nepers per meter} \quad [5\cdot201]$$

where the critical frequency $(f_0)_n = nv_1/2y_0$ for the nth order magnetic wave.

5·22 The Attenuation of the E_n Wave (TM_n Mode) Imperfect Conductor

The attenuation constant for the nth order transverse magnetic mode of transmission may be obtained in a manner similar to that employed in the general transverse electric case, of the previous section; it is

$$\alpha = y_0^{-3/2}\sqrt{\frac{2\pi\mu_2 n}{v_1\sigma_2\mu_1^2}}\left\{\frac{\left(\frac{f}{(f_0)_n}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_n}\right)^2-1}}\right\} \text{ nepers per meter} \quad [5\cdot202]$$

where the critical frequency $(f_0)_n = nv_1/2y_0$ for the nth order electric wave.

5·23 The Attenuation Constants

The attenuation constants given in the preceding sections are:

$$\underset{n=0}{\alpha TEM} = \frac{1}{y_0}\sqrt{\frac{\pi\mu_2\epsilon_1}{\sigma_2\mu_1}}\,(f)^{1/2} \quad \text{nepers per meter} \quad [5\cdot162]$$

$$\underset{n\neq0}{\alpha TE} = y_0^{-3/2}K\sqrt{n}\,\frac{\left(\frac{f}{(f_0)_n}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_n}\right)^2-1}} \quad \text{nepers per meter} \quad [5\cdot201]$$

$$\underset{n\neq0}{\alpha TM} = y_0^{-3/2}K\sqrt{n}\,\frac{\left(\frac{f}{(f_0)_n}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_n}\right)^2-1}} \quad \text{nepers per meter} \quad [5\cdot202]$$

where

$$K = \sqrt{\frac{2\mu_2\pi}{\sigma_2 v_1 \mu_1^2}}$$

The factor K depends on the material of which the bounding plates and dielectric are constructed, and on the velocity v_1. If the dielectric material between the plates is air, $v_1 = c$. It is seen that the attenuation is proportional to the minus three-halves power of the separation y_0 between the plates. Except where $n = 0$, the attenuation is infinite at $f = (f_0)_n$. In the case of the TM mode, the attenuation has a minimum at $f/(f_0)_n = \sqrt{3}$, and then rises as the frequency is increased. For large values of $f/(f_0)_n$ the attenuation increases as the square root of the frequency. In the case of the TE mode, the attenuation has no minimum but continues to decrease with increased frequency. For large values of $f/(f_0)_n$, the decrease in attenuation is proportional to the minus three halves power of the frequency. Figure 5·28 illustrates the variation of the attenuation as the function of frequency for the TE_1, TM_1, and TEM modes.

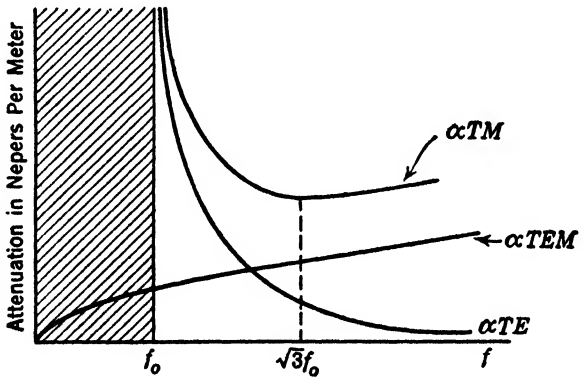


FIG. 5·28 Variation of the attenuation with frequency for various modes of transmission.

PROBLEMS

5·1 In the TEM wave in free space or between parallel plane guides the wave front is plane and perpendicular to the direction of propagation. The wave front is readily defined by the fact that both E and H are at one instant equal to zero over the entire surface. Examine critically the wave front of the TE_1 and TM_1 waves in terms of this definition. Consider E and H separately. What conclusions may be drawn?

5·2 Sketch the field distributions E and H, in the TE_3 and TM_3 modes of propagation.

5·3 Sketch the field distributions, E and H, in the TE_4 and TM_4 modes of propagation.

5·4 Indicate qualitatively how a complex traveling wave might be built up by the superposition of several simple modes. What requirements on relative velocity and phase exist? Compare with elliptically polarized light waves.

5·5 Develop equation 5·36 from equations 5·19.

5·6 Is air a dispersive or a non-dispersive medium for the propagation of sound waves? By what simple experiment may this be proved? Is the surface of a pond of water a dispersive or a non-dispersive medium to ordinary water waves? How may this statement be tested?

5·7 In electric wave filters composed of lumped inductances and capacitances the transition between the region of free transmission and that of attenuation is always smooth. Discuss Fig. 5·4 in the light of this fact. Is it likely that a curve such as that of Fig. 5·4 would be observed experimentally?

5·8 Evaluate the characteristic impedance of a wave guide to the TM_1 mode of transmission on the basis of voltage and current.

5·9 Evaluate the characteristic impedance of a wave guide to the TM_1 mode of transmission on the basis of the Poynting vector and the current.

5·10 Discuss qualitatively the performance to be expected if a section curved in the direction of propagation is introduced between two sections of plane wave guide. Consider the TEM, TE_1, and TM_1 modes of transmission.

5·11 Consider the possibility of resolving the TM_2 mode of transmission into plane waves by the method of section 5·15.

5·12 By the use of a standard reference work compare the velocity of electric and acoustic (mechanical) waves in copper at 60 cycles per second.

5·13 Examine critically the development of equation 5·160. Explain why a correct answer is obtained in spite of the fact that α appears to be assumed in the initial equation.

5·14 A wave of the TEM mode is propagated between two parallel copper sheets spaced 0.1 meter apart in air. Evaluate the attenuation at frequencies of 10^7, 10^8, and 10^9 cycles per second. Interpret these results in terms of the distance required for the wave to fall to $1/e$ of its initial value.

5·15 Repeat problem 5·14 for the TE_1 mode of propagation and frequencies of 3×10^9, 10^{10}, and 3×10^{10} cycles per second.

5·16 Repeat problem 5·15 for the TM_1 mode of propagation.

5·17 Discuss the form of equations 5·162, 5·201, and 5·202. Why should the attenuation be a function of frequency for any of these waves?

5·18 A wave guide is formed of two parallel sheets of copper spaced 20 cm in air. The frequency is 10^9 cycles per second. Determine the effect, upon the cut-off frequency, velocity of propagation, and attenuation, of substituting a loss-free insulator of dielectric constant 4 for the air.

5·19 What effect upon the attenuation of a wave guide is to be expected if the medium between the plates has a finite conductivity σ?

5·20 Discuss the practical application of plane parallel wave guides. What limitations do they suffer?

5·21 Does it seem probable that no modes of transmission other than those here described exist between parallel planes? What others may exist?

CHAPTER 6

RECTANGULAR WAVE GUIDES

This chapter is concerned with the propagation of electromagnetic waves in metallic tubes of rectangular cross section, filled with a dielectric material. The method of presentation is similar to that developed originally by Chu and Barrow.* As in Chapter 5, we shall consider first that the conductivity of the metal is infinite, and later discuss the practical case of imperfect conductivity.

6·1 Waves Guided by a Rectangular Metallic Tube

Let us consider the propagation of electromagnetic waves traveling in the X direction in a homogeneous isotropic medium which fills the interior of a rectangular metal tube of infinite length. A cross section of the tube is shown in Fig. 6·1. The inside edges of the upper and

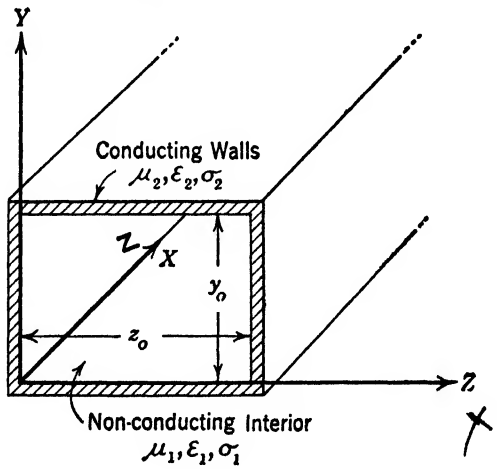


FIG. 6·1 Rectangular wave guide.

lower sides of the tube cut the Y axis at the points $y = 0$ and $y = y_0$; the inside edges of the two side walls cut the Z axis at $z = 0$ and $z = z_0$. Thus the interior dimensions of the cross section are z_0 wide and y_0 high.

As in Chapter 5 we shall introduce here the restrictions that the

* Chu and Barrow, "Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section," *IRE*, 26, 1520, 1938.

electric intensity **E** and the magnetic intensity **H** involve time only in the form $e^{j\omega t}$ and distance only in the form $e^{-\gamma x}$ so that

$$\mathbf{E} = \mathbf{E}'e^{j\omega t-\gamma x} \qquad [6\cdot1]$$

$$\mathbf{H} = \mathbf{H}'e^{j\omega t-\gamma x} \qquad [6\cdot2]$$

In the final solution we will, as before, take only the real part. Thus

$$\mathbf{E} = \mathcal{R}[\mathbf{E}'e^{j\omega t-\gamma x}]$$

and

$$\mathbf{H} = \mathcal{R}[\mathbf{H}'e^{j\omega t-\gamma x}]$$

The factor $e^{j\omega t-\gamma x}$ represents a wave traveling in the $+X$ direction. The frequency of the wave is given by $f = \omega/2\pi$, and the propagation constant is γ. As before, $\gamma = \alpha + j\beta$, where α is the attenuation constant and β the phase constant of the system. Substituting the relations 6·1 and 6·2 into the wave equations

$$\left.\begin{aligned}\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \sigma E_x + \varepsilon\frac{\partial E_x}{\partial t}\\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \sigma E_y + \varepsilon\frac{\partial E_y}{\partial t}\\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \sigma E_z + \varepsilon\frac{\partial E_z}{\partial t}\end{aligned}\right\} \quad [6\cdot3] \qquad \left.\begin{aligned}\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu\frac{\partial H_x}{\partial t}\\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu\frac{\partial H_y}{\partial t}\\ \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu\frac{\partial H_z}{\partial t}\end{aligned}\right\} \quad [6\cdot4]$$

We obtain

$$\left.\begin{aligned}\frac{\partial H'_z}{\partial y} - \frac{\partial H'_y}{\partial z} &= (\sigma + j\omega\varepsilon)E'_x\\ \frac{\partial H'_x}{\partial z} + \gamma H'_z &= (\sigma + j\omega\varepsilon)E'_y\\ -\gamma H'_y - \frac{\partial H'_x}{\partial y} &= (\sigma + j\omega\varepsilon)E'_z\end{aligned}\right\} \quad [6\cdot5] \qquad \left.\begin{aligned}\frac{\partial E'_z}{\partial y} - \frac{\partial E'_y}{\partial z} &= -j\omega\mu H'_x\\ \frac{\partial E'_x}{\partial z} + \gamma E'_z &= -j\omega\mu H'_y\\ \gamma E'_y + \frac{\partial E'_x}{\partial y} &= j\omega\mu H'_z\end{aligned}\right\} \quad [6\cdot6]$$

There are two types of waves which satisfy these equations, and which may exist independently. These waves, known as the *E* or *TM* wave and the *H* or *TE* wave, respectively, are characterized by two types of transmission modes, depending on the presence or absence of the x component of the electric or magnetic intensity.

6·2 Transmission Modes. *E* and *H* Waves

Solutions which have a component of electric intensity in the direction of propagation but no component of magnetic intensity in the direction of propagation are called *E* waves. Where propagation in

the X direction is considered, we may state this more briefly as $H_x = 0$, $E_x \neq 0$. Since $H_x = 0$, the magnetic intensity is normal to the direction of propagation. Hence this mode of transmission may also be described as a transverse magnetic mode, or *TM* mode. This wave is sometimes called longitudinal electric, since there is a component of **E** in the direction of propagation.

Solutions which have a component of magnetic intensity but no component of electric intensity in the direction of propagation are called *H* waves. More briefly, when propagation is in the X direction, *H* waves are those in which $E_x = 0$ and $H_x \neq 0$. Since $E_x = 0$, the electric intensity is normal to the direction of propagation. Hence this mode of transmission may be characterized as a transverse electric mode or *TE* mode. This wave is sometimes called longitudinal magnetic, since there is a component of **H** in the direction of propagation.

6·3 Transverse Electric (*TE* Mode) *H* Waves

With the restriction that $E_x = 0$, equations 6·5 and 6·6 become

$$\left.\begin{aligned} \frac{\partial H_z'}{\partial y} - \frac{\partial H_y'}{\partial z} &= 0 \qquad (a) \\ \frac{\partial H_x'}{\partial z} + \gamma H_z' &= (\sigma + j\omega\varepsilon)E_y' \quad (b) \\ -\gamma H_y' - \frac{\partial H_x'}{\partial y} &= (\sigma + j\omega\varepsilon)E_z' \, (c) \end{aligned}\right\} [6\cdot7] \qquad \left.\begin{aligned} \frac{\partial E_z'}{\partial y} - \frac{\partial E_y'}{\partial z} &= -j\omega\mu H_x' \, (a) \\ \gamma E_z' &= -j\omega\mu H_y' \, (b) \\ \gamma E_y' &= j\omega\mu H_z' \quad (c) \end{aligned}\right\} [6\cdot8]$$

We may obtain from equation 6·7 and 6·8 an equation in H_x' alone, as follows:

Differentiate 6·7*b* with respect to z, obtaining

$$\frac{\partial^2 H_x'}{\partial z^2} + \gamma \frac{\partial H_z'}{\partial z} = (\sigma + j\omega\varepsilon) \frac{\partial E_y'}{\partial z} \qquad [6\cdot9]$$

Differentiate 6·7*c* with respect to y, obtaining

$$-\gamma \frac{\partial H_y'}{\partial y} - \frac{\partial^2 H_x'}{\partial y^2} = (\sigma + j\omega\varepsilon) \frac{\partial E_z'}{\partial y} \qquad [6\cdot10]$$

Subtract 6·10 from 6·9 to give

$$\frac{\partial^2 H_x'}{\partial z^2} + \frac{\partial^2 H_x'}{\partial y^2} + \gamma \left\{ \frac{\partial H_z'}{\partial z} + \frac{\partial H_y'}{\partial y} \right\} = (\sigma + j\omega\varepsilon) \left\{ \frac{\partial E_y'}{\partial z} - \frac{\partial E_z'}{\partial y} \right\} \qquad [6\cdot11]$$

Next differentiate 6·8*b* with respect to y and obtain

$$\gamma \frac{\partial E_z'}{\partial y} = -j\omega\mu \frac{\partial H_y'}{\partial y}$$

Similarly differentiating 6·8*c* with respect to z gives

$$\gamma \frac{\partial E'_y}{\partial z} = j\omega\mu \frac{\partial H'_z}{\partial z}$$

The difference between these equations is

$$\gamma \left\{ \frac{\partial E'_z}{\partial y} - \frac{\partial E'_y}{\partial z} \right\} = -j\omega\mu \left\{ \frac{\partial H'_y}{\partial y} + \frac{\partial H'_z}{\partial z} \right\} \qquad [6\cdot12]$$

From 6·8*a* the term in the brackets on the left side of equation 6·12 is $-j\omega\mu H'_x$. Hence

$$\frac{\partial H'_y}{\partial y} + \frac{\partial H'_z}{\partial z} = \gamma H'_x \qquad [6\cdot13]$$

Thus using 6 8*a* equation 6·11 becomes

$$\frac{\partial^2 H'_x}{\partial z^2} + \frac{\partial^2 H'_x}{\partial y^2} + \gamma^2 H'_x = (\sigma + j\omega\varepsilon) j\omega\mu H'_x \qquad [6\cdot14]$$

or

$$\frac{\partial^2 H'_x}{\partial z^2} + \frac{\partial^2 H'_x}{\partial y^2} = -\{\gamma^2 - (\sigma + j\omega\varepsilon) j\omega\mu\} H'_x \qquad [6\cdot15]$$

In the dielectric $\sigma_1 \ll \omega\varepsilon_1$. Therefore, in the dielectric, equation 6·15 becomes

$$\frac{\partial^2 H'_x}{\partial z^2} + \frac{\partial^2 H'_x}{\partial y^2} = -(\gamma^2 + \omega^2\mu_1\varepsilon_1) H'_x \qquad [6\cdot16]$$

Equation 6·16 could have been obtained directly from the wave equation 3·46, where, in the dielectric,

$$\frac{\partial^2 H_x}{\partial x^2} + \frac{\partial^2 H_x}{\partial y^2} + \frac{\partial^2 H_x}{\partial z^2} = \mu_1\varepsilon_1 \frac{\partial^2 H_x}{\partial t^2} \qquad [6\cdot17]$$

Substituting $H_x = H'_x e^{j\omega t - \gamma x}$ into 6·17 and dividing through by $e^{j\omega t - \gamma x}$ we obtain

$$\gamma^2 H'_x + \frac{\partial^2 H'_x}{\partial y^2} + \frac{\partial^2 H'_x}{\partial z^2} = -\mu_1\varepsilon_1\omega^2 H'_x$$

or

$$\frac{\partial^2 H'_x}{\partial y^2} + \frac{\partial^2 H'_x}{\partial z^2} = -(\gamma^2 + \mu_1\varepsilon_1\omega^2) H'_x \qquad [6\cdot16]$$

6·4 Solution by the Method of Separation of Variables

Equation 6·16 may be solved by the method of separation of variables. This method is extremely useful in the solution of partial differential

equations of this type. To apply this method to the solution of 6·16 we attempt to find a solution which will consist of the product of a function of y alone and a function of z alone. Let $Y(y)$ and $Z(z)$ be these functions so that our solution is $H'_x = Y(y)Z(z)$, or, expressed more briefly

$$H'_x = YZ \qquad [6\cdot18]$$

Let us substitute this solution into 6·16. The required partial derivatives are

$$\frac{\partial}{\partial y}(YZ) = Z\frac{\partial Y}{\partial y}\,; \quad \frac{\partial^2}{\partial y^2}(YZ) = Z\frac{\partial^2 Y}{\partial y^2}$$

$$\frac{\partial}{\partial z}(YZ) = Y\frac{\partial Z}{\partial z}\,; \quad \frac{\partial^2}{\partial z^2}(YZ) = Y\frac{\partial^2 Z}{\partial z^2}$$

Thus

$$Z\frac{\partial^2 Y}{\partial y^2} + Y\frac{\partial^2 Z}{\partial z^2} = -(\gamma^2 + \mu_1\varepsilon_1\omega^2)YZ$$

Dividing through by YZ we obtain

$$\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -(\gamma^2 + \mu_1\varepsilon_1\omega^2) \qquad [6\cdot19]$$

The first term of equation 6·19 is a function of y alone, say $F(y)$; the second term is a function of z alone, say $F(z)$. Then from 6·19

$$F(y) + F(z) = C \qquad [6\cdot20]$$

where C is a constant equal to $-(\gamma^2 + \omega^2\mu_1\varepsilon_1)$. The equation therefore states that a function of y plus a function of z is equal to a constant, no matter what values y and z may have. This is surely not possible in general, for, if we keep y fixed and vary z, $F(z)$ would change while $F(y)$ remained the same. Thus the equation would not be satisfied. The only exception would, of course, occur where $F(y)$ and $F(z)$ are each constant and independent of y and z. If we impose these conditions, we have

$$\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} = -A_1 \qquad [6\cdot21]$$

and

$$\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -A_2 \qquad [6\cdot22]$$

where A_1 and A_2 are constants and

$$A_1 + A_2 = (\gamma^2 + \omega^2\mu_1\varepsilon_1) \qquad [6\cdot23]$$

We may rewrite 6·21 as

$$\frac{\partial^2 Y}{\partial y^2} + A_1 Y = 0 \qquad [6\cdot24]$$

and 6·22 as

$$\frac{\partial^2 Z}{\partial z^2} + A_2 Z = 0 \qquad [6\cdot25]$$

Equation 6·16 is thus reduced to the ordinary differential equations 6·24 and 6·25. The solution of 6·24 is well known in the theory of differential equations and may be written down directly. It is

$$Y = e^{jgy} \pm e^{-jgy} \qquad [6\cdot26]$$

$$= C_1 \sin gy \qquad \text{or} \qquad C_2 \cos gy \qquad [6\cdot27]$$

where C_1, C_2, and g are constants.

Since $C_1 \sin gy$ and $C_2 \cos gy$ are each solutions of 6·24 their sum is also a solution. Thus $Y = C_1 \sin gy + C_2 \cos gy$. The constant g may be determined by substitution of 6·27 in 6·24. Carrying out this substitution we obtain

$$g = \sqrt{A_1}$$

Hence

$$Y = C_1 \sin \sqrt{A_1}\, y + C_2 \cos \sqrt{A_1}\, y \qquad [6\cdot28]$$

Similarly, the solution of 6.25 is

$$Z = e^{jhz} \pm e^{-jhz} \qquad [6\cdot29]$$

$$= C_3 \sin hz \qquad \text{or} \qquad C_4 \cos hz \qquad [6\cdot30]$$

where

$$h = \sqrt{A_2}$$

and C_3 and C_4 are arbitrary constants.

Thus

$$Z = C_3 \sin \sqrt{A_2}\, z + C_4 \cos \sqrt{A_2}\, z \qquad [6\cdot31]$$

Since the solution of 6·16 is $H_x = YZ$, we may choose, as the final solution,

$$H_x' = A \cos \sqrt{A_1}\, y \cos \sqrt{A_2}\, z \qquad [6\cdot32]$$

where $A = C_2 C_4$.

Other solutions may be chosen from 6·28 and 6·31, as, for example, the product of the sines, or products of sines and cosines, but 6·32 is the one we want, since, as will be evident later, it alone will satisfy the boundary conditions of the problem.

Since we have now determined H_x' we may readily determine the remaining amplitudes of the components of the H wave from equations 6·7 and 6·8. For example, we may obtain an equation relating H_y' and H_x' by eliminating E_z' from equations 6·7*c* and 6·8*b*. Substituting for E_z', in equation 6·7*c*, its value in terms of H_y' as given in 6·8*b*, we have

$$-\gamma H_y' - \frac{\partial H_x'}{\partial y} = \frac{\omega^2 \mu_1 \varepsilon_1}{\gamma} H_y' \qquad [6\cdot33]$$

where σ_1 has been neglected in comparison to $\omega\varepsilon_1$ in the dielectric. Thus,

$$H_y' = -\frac{\gamma}{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} \frac{\partial H_x'}{\partial y} \qquad [6\cdot34]$$

From 6·32

$$\frac{\partial H_x'}{\partial y} = -A \sqrt{A_1} \sin \sqrt{A_1}\, y \cos \sqrt{A_2}\, z \qquad [6\cdot35]$$

Therefore

$$H_y' = A \frac{\gamma}{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} \sqrt{A_1} \sin \sqrt{A_1}\, y \cos \sqrt{A_2}\, z \qquad [6\cdot36]$$

Since equation 6·8*b* expresses E_z' in terms of H_y' we have for E_z'

$$E_z' = -\frac{j\omega\mu_1}{\gamma} H_y' = -A \frac{j\omega\mu_1}{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} \sqrt{A_1} \sin \sqrt{A_1}\, y \cos \sqrt{A_2}\, z \qquad [6\cdot37]$$

From equations 6·7*b* and 6·8*c* we may get a relation between H_z' and H_x'. Again neglecting σ in comparison to $\omega\varepsilon_1$ in the dielectric, we obtain

$$\frac{\partial H_x'}{\partial z} + \gamma H_z' = -\frac{\omega^2 \mu_1 \varepsilon_1}{\gamma} H_z' \qquad [6\cdot38]$$

so that

$$H_z' = -\frac{\gamma}{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} \frac{\partial H_x'}{\partial z} \qquad [6\cdot39]$$

From equation 6·32

$$\frac{\partial H_x'}{\partial z} = -A \sqrt{A_2} \cos \sqrt{A_1}\, y \sin \sqrt{A_2}\, z \qquad [6\cdot40]$$

Therefore

$$H_z' = A \frac{\gamma}{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} \sqrt{A_2} \cos \sqrt{A_1}\, y \sin \sqrt{A_2}\, z \qquad [6\cdot41]$$

Finally we may obtain E_y' from equation 6·8c

$$E_y' = \frac{j\omega\mu_1}{\gamma} H_z' = A \frac{j\omega\mu_1}{\gamma^2 + \omega^2\mu_1\varepsilon_1} \sqrt{A_2} \cos \sqrt{A_1}\, y \sin \sqrt{A_2}\, z \qquad [6\cdot42]$$

6·5 Boundary Conditions for *H* Waves

We have now obtained all the components of the H wave in the dielectric. Let us apply the boundary conditions of the rectangular tube to the amplitude of the E_y and E_z components.

Since the tangential component of an electric vector must vanish at the surface of a perfect conductor, E_y' must equal zero at $z = 0$ or z_0. Let $\sqrt{A_2} = m\pi/z_0$, where m is an integer. Then from 6·42

$$E_y' = A \frac{j\omega\mu_1}{\gamma^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{m\pi}{z_0}\right) \cos \sqrt{A_1}\, y \sin \left(\frac{m\pi}{z_0} z\right) \qquad [6\cdot43]$$

Now, at $z = 0$, $\sin \left(\frac{m\pi}{z_0} z\right) = 0$; and at $z = z_0$ it reduces to $\sin (m\pi)$. The latter is zero for any integral value of m. Thus $E_y' = 0$ at $z = 0$ or z_0, and the boundary conditions are satisfied in the Z direction.

Likewise E_z' must be zero at $y = 0$ or y_0. Let

$$\sqrt{A_1} = \frac{n\pi}{y_0}$$

where n is an integer. Then from 6·37

$$E_z' = -A \frac{j\omega\mu_1}{\gamma^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{n\pi}{y_0}\right) \sin \left(\frac{n\pi}{y_0} y\right) \cos \left(\frac{m\pi}{z_0} z\right) \qquad [6\cdot44]$$

and the boundary conditions are satisfied in the Y direction as well as in the Z direction. Also from 6·23

$$\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2 = \gamma^2 + \omega^2\mu_1\varepsilon_1 \qquad [6\cdot45]$$

so that

$$\gamma_{nm} = \sqrt{\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2 - \omega^2\mu_1\varepsilon_1} \qquad [6\cdot46]^*$$

* Since γ depends on the particular values of m and n chosen, it may be written γ_{nm}. Thus, when $n = 2$, $m = 3$, $\gamma_{nm} = \gamma_{23}$.

6·6 H_{nm} Waves in the Dielectric

Let us collect the various components of the H wave in the dielectric which we have obtained. They are

$$\left.\begin{aligned}
H_x &= A \cos\left(\frac{n\pi}{y_0} y\right) \cos\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\
H_y &= A \frac{\gamma_{nm}}{\gamma_{nm}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{n\pi}{y_0}\right) \sin\left(\frac{n\pi}{y_0} y\right) \cos\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\
H_z &= A \frac{\gamma_{nm}}{\gamma_{nm}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{m\pi}{z_0}\right) \cos\left(\frac{n\pi}{y_0} y\right) \sin\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\
E_x &= 0 \\
E_y &= A \frac{j\omega\mu_1}{\gamma_{nm}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{m\pi}{z_0}\right) \cos\left(\frac{n\pi}{y_0} y\right) \sin\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\
E_z &= -A \frac{j\omega\mu_1}{\gamma_{nm}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{n\pi}{y_0}\right) \sin\left(\frac{n\pi}{y_0} y\right) \cos\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x}
\end{aligned}\right\} \quad \begin{array}{l} H_{nm} \text{ or } TE_{nm} \\ \text{waves in the} \\ \text{dielectric} \\ [6\cdot47] \end{array}$$

The constant A depends only upon the original excitation creating the wave and denotes absolute magnitude only.

The use of subscripts to indicate the order of the wave in each coordinate is convenient. Thus two are necessary here, whereas in the previous study of parallel planes only one was required. The subscript n refers to the number of half sinusoids or maxima of field intensity in the Y direction from $y = 0$ to $y = y_0$. The subscript m similarly indicates the number of half sinusoids occurring in the Z direction from $z = 0$ to $z = z_0$. We can therefore refer to H_{nm} waves in general, the particular wave being designated by replacing n or m with the order of the wave under discussion. Thus, if $n = 0$ and $m = 0$, the condition existing may be designated H_{00}. If $n = 1$, $m = 0$, the wave is designated H_{10}. In this way the H_{nm} wave stands for an infinite number of specific waves. All these waves are not possible, however; for example, the H_{00} wave is not possible since, when $n = 0$ and $m = 0$, all the components in equations 6·47 vanish. If $n = 0$ and $m \neq 0$, $E_x = E_z = H_y = 0$ and three components of the H_{0m} wave remain. Hence all waves of the H_{0m} type are possible. It is seen that in this wave, since $E = E_y$, the electric intensity is normal to the direction of the propagation, and the wave may be characterized as transverse electric. Waves of the type H_{11} and higher-order waves are also theoretically possible. In the case of waves having complementary indices, H_{12} or H_{21}, H_{13} or H_{31}, etc., the field distributions are all alike except for their orientation in the pipe. Thus it is sufficient to con-

sider one orientation only. The other orientation is obtained simply by interchanging the y_0 and z_0 dimensions of the guide. Because of the extremely high frequencies required by the higher-order waves in guides of moderate size, their practical importance at this time is questionable.

6·7 Transmission Characteristics of H_{nm} Waves

In equation 6·46 it is seen that the propagation constant is given by the relation

$$\gamma_{nm} = \sqrt{\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2 - \omega^2\mu_1\varepsilon_1} \qquad [6\cdot46]$$

A critical analysis of this constant leads to three distinct cases, which we shall now consider.

Case I. Propagation

If

$$\omega^2\mu_1\varepsilon_1 > \left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2$$

γ_{nm} is imaginary. Since $\gamma_{nm} = \alpha_{nm} + j\beta_{nm}$, it is seen that under these circumstances $\alpha_{nm} = 0$ and $\gamma_{nm} = j\beta_{nm}$. Thus,

$$\beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2 - \left(\frac{m\pi}{z_0}\right)^2} \qquad [6\cdot48]$$

and we may write from equations 6·1 and 6·2

$$E = E'e^{j\omega t - j\beta_{nm}x} = E'e^{j(\omega t - \beta_{nm}x)}$$
$$= E'\{\cos(\omega t - \beta_{nm}x) + j\sin(\omega t - \beta_{nm}x)\} \qquad [6\cdot49]$$

$$H = H'e^{j\omega t - j\beta_{nm}x} = H'e^{j(\omega t - \beta_{nm}x)}$$
$$= H'\{\cos(\omega t - \beta_{nm}x) + j\sin(\omega t - \beta_{nm}x)\} \qquad [6\cdot50]$$

Since in the final result we take only the real part of E and H as the solution, we have

$$E = E'\cos(\omega t - \beta_{nm}x) \qquad [6\cdot51]$$

$$H = H'\cos(\omega t - \beta_{nm}x) \qquad [6\cdot52]$$

if the amplitudes of E and H are real. If the amplitudes are imaginary and therefore equal to jE' and jH', then the real part of 6·49 and 6·50 is

$$E = -E'\sin(\omega t - \beta_{nm}x) \qquad [6\cdot53]$$

$$H = -H'\sin(\omega t - \beta_{nm}x) \qquad [6\cdot54]$$

giving the solution. The amplitudes of E_y and E_z as given in equation 6·47 are imaginary, and this condition will apply to them. If the amplitudes are complex, a phase angle may be introduced so that

$$E = E' \cos (\omega t - \beta_{nm} x + \theta_1) \qquad [6\cdot55]$$

$$H = H' \cos (\omega t - \beta_{nm} x + \theta_2) \qquad [6\cdot56]$$

where θ_1 and θ_2 are angles whose tangents are the imaginary part of the amplitude, divided by the real part of the amplitude of the respective vectors. When θ_1 and θ_2 are 90 degrees, i.e., when the amplitudes of E and H are purely imaginary, equations 6·55 and 6·56 reduce to equations 6·53 and 6·54.

From the above we see that, as long as the propagation constant is imaginary, E and H are periodic functions of distance and hence wave propagation in the X direction takes place with no attenuation. This is, of course, strictly true only in the ideal case where the conductivity is infinite.

Case II. No propagation

If $(n\pi/y_0)^2 + (m\pi/z_0)^2 > \omega^2\mu_1\varepsilon_1$, γ_{nm} is real, $\beta_{nm} = 0$, and $\gamma_{nm} = \alpha_{nm}$. Thus

$$\alpha_{nm} \quad \sqrt{\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2 - \omega^2\mu_1\varepsilon_1} \qquad [6\cdot57]$$

and we may write from equations 6·1 and 6·2

$$E = E'e^{-\alpha x + j\omega t} = E'e^{-\alpha x} \{\cos \omega t + j \sin \omega t\}$$

$$H = H'e^{-\alpha x + j\omega t} = H'e^{-\alpha x} \{\cos \omega t + j \sin \omega t\}$$

Since in the final result we take only the real part of E and H as the solution, we have

$$E = E'e^{-\alpha x} \cos \omega t \qquad [6\cdot58]$$

$$H = H'e^{-\alpha x} \cos \omega t \qquad [6\cdot59]$$

if the amplitude of E and H are real. If these amplitudes are complex, then a phase angle is introduced as in Case I.

The expressions 6·58 and 6·59 indicate that the wave is no longer a periodic function of distance. Propagation does not take place, but instead there is only attenuation. The attenuation factor $e^{-\alpha x}$ determines the rate at which the maximum amplitudes E' and H' are reduced as the distance x increases away from the excitation source creating the wave. Furthermore, the magnitude of α is such that practically complete extinction of the wave occurs in a distance corresponding to a few wavelengths in free space.

Case III. Critical case

If $$\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2 = \omega^2\mu_1\varepsilon_1$$

a critical case is defined, which represents the intermediate condition between Cases I and II. The critical frequency $(f_0)_{nm} = \omega_0/2\pi$ defined by this condition is thus

$$(f_0)_{nm} = \frac{1}{2\pi\sqrt{\mu_1\varepsilon_1}}\sqrt{\left(\frac{n\pi}{y_0}\right)^2 + \left(\frac{m\pi}{z_0}\right)^2} = \frac{1}{2\sqrt{\mu_1\varepsilon_1}}\sqrt{\left(\frac{n}{y_0}\right)^2 + \left(\frac{m}{z_0}\right)^2} \quad [6\cdot60]$$

For values of frequency greater than $(f_0)_{nm}$, transmission of the wave takes place. This corresponds to Case I. For values of frequency less than $(f_0)_{nm}$, the wave is attenuated and no propagation takes place. This condition corresponds to Case II.

The quantity $1/\sqrt{\mu_1\varepsilon_1}$ has been shown in Chapter 3 to be the velocity v_1 of a free wave* in a medium of dielectric constant ε_1 and permeability μ_1. In terms of this velocity, we may define a critical wavelength $(\lambda_0)_{nm}$, i.e., $(\lambda_{nm})(f_{nm}) = v_1$, so that from equation 6.60 we have

$$\text{(Critical wavelength) } (\lambda_0)_{nm} = \frac{v_1}{(f_0)_{nm}} = \frac{2}{\sqrt{\left(\frac{n}{y_0}\right)^2 + \left(\frac{m}{z_0}\right)^2}} \quad [6\cdot61]$$

It is seen under Case I that so long as γ_{nm} is imaginary (β_{nm} real) the E and H vectors vary periodically with both distance and time, and transmission through the dielectric takes place without attenuation. The velocity at which the wave is propagated may be determined as before, by finding the conditions under which the argument $\omega t - \beta_{nm}x$ remains unchanged when the time t is increased a finite amount. Thus when t is increased by an amount t', the equiphase surface of the wave will travel a distance x'. The argument then becomes

$$\omega(t + t') - \beta_{nm}(x + x')$$

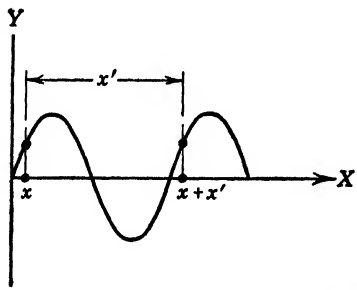


FIG. 6·2 Development of phase velocity.

If these two arguments are the same, the phase of the wave will be the same at $x + x'$ as it was at x. See Fig. 6·2. Hence we wish to determine the conditions under which

$$\omega t - \beta_{nm}x = \omega(t + t') - \beta_{nm}(x + x') \quad [6\cdot62]$$

* By free wave is meant a wave traveling in free space, in which there are no boundary conditions to be met.

These conditions, as obtained from equation 6·62, are that

$$\omega t' - \beta_{nm} x' = 0 \qquad [6\cdot63]$$

or that

$$\frac{x'}{t'} = \frac{\omega}{\beta_{nm}} \qquad [6\cdot64]$$

Now x'/t' is the average velocity with which the equiphase surface of the wave moves a distance x' in a time t'. Let us call this velocity the phase velocity v_p. Then

$$(v_p)_{nm} = \frac{\omega}{\beta_{nm}} \qquad [6\cdot65]$$

The group velocity v_g has been defined as

$$(v_g)_{nm} = \cdot \frac{1}{\dfrac{\partial \beta_{nm}}{\partial \omega}} \cdot \qquad [6\cdot66]$$

From equation 6·48

$$\frac{\partial \beta_{nm}}{\partial \omega} = \frac{1}{2}\left[\omega^2 \mu_1 \varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2 - \left(\frac{m\pi}{z_0}\right)^2\right]^{-\frac{1}{2}} 2\omega\mu_1\varepsilon_1 = \frac{\omega\mu_1\varepsilon_1}{\beta_{nm}}$$

Hence,

$$(v_g)_{nm} = \frac{\beta_{nm}}{\omega\mu_1\varepsilon_1} = \frac{v_1^2}{(v_p)_{nm}} \qquad [6\cdot67]$$

where

$$v_1 = \frac{1}{\sqrt{\mu_1\varepsilon_1}} \qquad [6\cdot68]$$

Hence

$$v_1^2 = (v_g)_{nm}(v_p)_{nm} \qquad [6\cdot69]$$

In the pipe the wavelength λ_p is given by the relation $\lambda_p f = (v_p)_{nm}$, where f is the frequency of the excitation and is equal to $\omega/2\pi$. From equation 6·65, $(v_p)_{nm} = \omega/\beta_{nm}$. Hence

$$(\lambda_p)_{nm} = \frac{(v_p)_{nm}}{f} = \frac{\omega}{\beta_{nm} f} = \frac{2\pi}{\beta_{nm}} \qquad [6\cdot70]$$

The wavelength λ_1 may be defined from the relation $\lambda_1 f = v_1$, where f, the frequency of excitation, is $\omega/2\pi$, and λ_1 and v_1 are the free wavelength and velocity in a medium having the constants μ_1 and ε_1. Thus

$$\lambda_1 = \frac{v_1}{f} = 2\pi\frac{v_1}{\omega} = \frac{2\pi}{\omega\sqrt{\mu_1\varepsilon_1}} \qquad [6\cdot70a]$$

The quantities defining the transmission properties of the H_{nm} waves are collected below.

Phase constant
$$\beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{n\pi}{y_0}\right)^2 - \left(\frac{m\pi}{z_0}\right)^2} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{n\pi}{y_0}\right)^2 - \left(\frac{m\pi}{z_0}\right)^2}$$

Critical frequency
$$(f_0)_{nm} = \frac{v_1}{2}\sqrt{\left(\frac{n}{y_0}\right)^2 + \left(\frac{m}{z_0}\right)^2}$$

Critical wavelength
$$(\lambda_0)_{nm} = \frac{2}{\sqrt{\left(\frac{n}{y_0}\right)^2 + \left(\frac{m}{z_0}\right)^2}}$$

Wavelength in pipe
$$(\lambda_p)_{nm} = \frac{2\pi}{\beta_{nm}}$$

Phase velocity
$$(v_p)_{nm} = \frac{\omega}{\beta_{nm}}$$

Group velocity
$$(v_g)_{nm} = \frac{v_1^2}{(v_p)_{nm}}$$

Free wave velocity
$$v_1 = \frac{1}{\sqrt{\mu_1\varepsilon_1}}$$

Wavelength in free space
$$\lambda_1 = 2\pi\frac{v_1}{\omega} = \frac{v_1}{f}$$

[6·71]*

6·8 The H_{01} or TE_{01} Wave

It was pointed out in section 6·6 that all waves of the type H_{0m} are possible modes of transmission, except the H_{00} wave. All waves of the H_{0m} type are transverse electric waves and may also be characterized by the designation TE_{0m} waves, $m \neq 0$. Let us investigate the special case where $n = 0$, $m = 1$, i.e., the H_{01} wave. Equations 6·47 reduce to

$$H_x = A\cos\left(\frac{\pi}{z_0}z\right)e^{j\omega t-\gamma_{01}x}$$

$$H_z = A\frac{\gamma_{01}}{\gamma_{01}^2 + \omega^2\mu_1\varepsilon_1}\left(\frac{\pi}{z_0}\right)\sin\left(\frac{\pi}{z_0}z\right)e^{j\omega t-\gamma_{01}x}$$

$$E_y = A\frac{j\omega\mu_1}{\gamma_{01}^2 + \omega^2\mu_1\varepsilon_1}\left(\frac{\pi}{z_0}\right)\sin\left(\frac{\pi}{z_0}z\right)e^{j\omega t-\gamma_{01}x}$$

$$H_y = E_x = E_z = 0$$

H_{01} or TE_{01} waves in the dielectric [6·72]

It is later shown that these equations also apply to the E_{nm} waves.

Since the propagation constant γ_{01} is given by

$$\gamma_{01}^2 = \left(\frac{\pi}{z_0}\right)^2 - \omega^2\mu_1\varepsilon_1 \qquad [6\cdot73]$$

and since

$$\gamma_{01} = j\beta_{01} \qquad [6\cdot74]$$

the final solution may be written

$$\left.\begin{aligned} H_x &= A\cos\left(\frac{\pi}{z_0}z\right)\cos(\omega t - \beta_{01}x) \\ H_z &= -A\frac{\beta_{01}z_0}{\pi}\sin\left(\frac{\pi}{z_0}z\right)\sin(\omega t - \beta_{01}x) \\ E_y &= -A\frac{\omega\mu_1 z_0}{\pi}\sin\left(\frac{\pi}{z_0}z\right)\sin(\omega t - \beta_{01}x) \\ H_y &= E_x = E_z = 0 \end{aligned}\right\} \quad H_{01} \text{ or } TE_{01} \text{ waves in the dielectric } [6\cdot75]$$

The quantities defining the transmission properties of the H_{01} wave are:

$$\left.\begin{aligned} &\text{Phase constant} && \beta_{01} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{z_0}\right)^2} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{z_0}\right)^2} \\ &\text{Critical frequency} && (f_0)_{01} = \frac{1}{2\sqrt{\mu_1\varepsilon_1}\,z_0} = \frac{v_1}{2z_0} \\ &\text{Critical wavelength} && (\lambda_0)_{01} = 2z_0 \\ &\text{Wavelength in pipe} && (\lambda_p)_{01} = \frac{2\pi}{\beta_{01}} = \frac{\lambda_1}{\sqrt{1 - \left(\frac{\lambda_1}{(\lambda_0)_{01}}\right)^2}} \\ &\text{Phase velocity} && (v_p)_{01} = \frac{\omega}{\beta_{01}} = \frac{v_1}{\sqrt{1 - \left(\frac{\lambda_1}{(\lambda_0)_{01}}\right)^2}} \\ &\text{Group velocity} && (v_g)_{01} = \frac{v_1^2}{(v_p)_{01}} = v_1\sqrt{1 - \left(\frac{\lambda_1}{(\lambda_0)_{01}}\right)^2} \end{aligned}\right\} \quad [6\cdot76]^*$$

$$^*\beta_{01} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{z_0}\right)^2} = \frac{\omega}{v_1}\sqrt{1 - \left(\frac{\pi v_1}{\omega z_0}\right)^2} = \frac{\omega}{v_1}\sqrt{1 - \left(\frac{\lambda_1}{2z_0}\right)^2}$$

$$= \frac{\omega}{v_1}\sqrt{1 - \left(\frac{\lambda_1}{(\lambda_0)_{01}}\right)^2} = \frac{\omega}{v_1}\sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2}$$

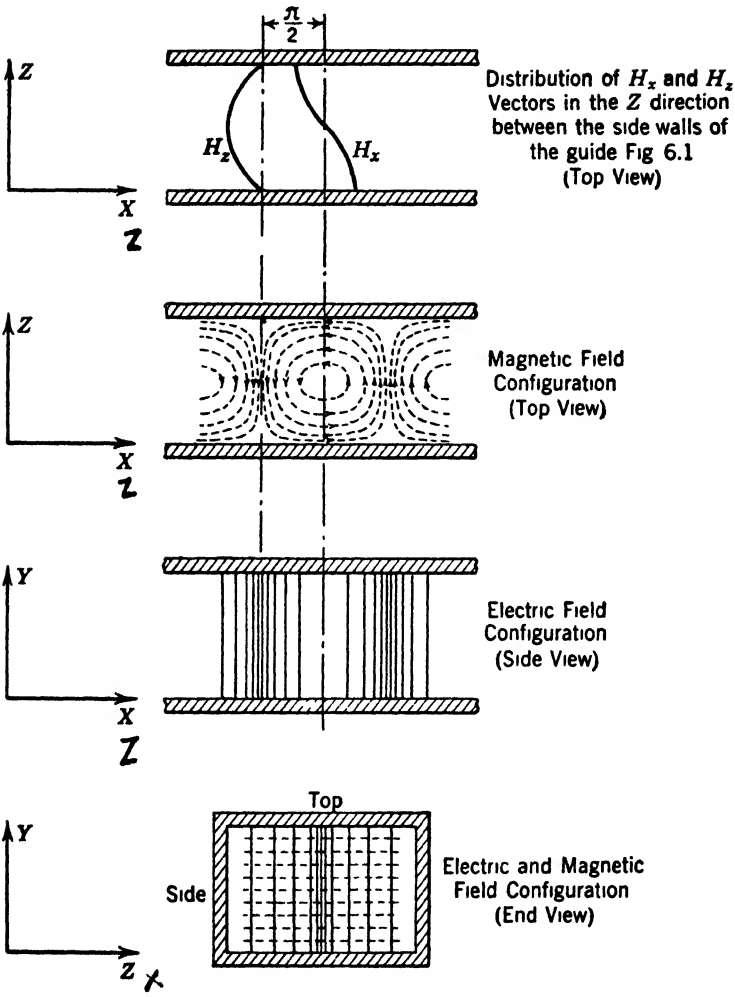


FIG. 6·3 The instantaneous field configuration for an H_{01} or TE_{01} wave in the dielectric medium which fills a rectangular wave guide. This field is propagated in the X direction through the guide.

The field configuration in the H_{01} wave may be found as in section 5·11. The lines of electric and magnetic intensity are shown in Fig. 6·3. Since the critical wavelength is independent of the y_0 dimension, either of the pipes whose cross sections are shown in Fig. 6·4 will transmit the H_{01} wave. Further discussion of this wave will be taken up in a later section.

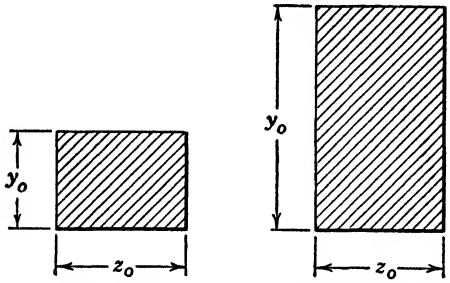


FIG. 6·4 Both pipes have the same critical frequency $(\lambda_0)_{01} = 2z_0$.

6·9 The H_{02} or TE_{02} Wave

By setting $n = 0$ and $m = 2$ we may obtain, from the general equations 6·47, the components of the H_{02} wave. They are:

$$\left.\begin{aligned} H_x &= A \cos\left(\frac{2\pi}{z_0} z\right) e^{j\omega t - \gamma_{02} x} \\ H_z &= A \frac{\gamma_{02}}{\gamma_{02}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{2\pi}{z_0}\right) \sin\left(\frac{2\pi}{z_0} z\right) e^{j\omega t - \gamma_{02} x} \\ E_y &= A \frac{j\omega\mu_1}{\gamma_{02}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{2\pi}{z_0}\right) \sin\left(\frac{2\pi}{z_0} z\right) e^{j\omega t - \gamma_{02} x} \end{aligned}\right\} \text{H_{02} or TE_{02} waves in the dielectric}$$

The propagation constant γ_{02} is given by

$$\gamma_{02}^2 = \left(\frac{2\pi}{z_0}\right)^2 - \omega^2 \mu_1 \varepsilon_1$$

and since $\gamma_{02} = j\beta_{02}$ the final solution for the H_{02} wave may be written as

$$\left.\begin{aligned} H_x &= A \cos\left(\frac{2\pi}{z_0} z\right) \cos(\omega t - \beta_{02} x) \\ H_z &= -A \frac{\beta_{02} z_0}{2\pi} \sin\left(\frac{2\pi}{z_0} z\right) \sin(\omega t - \beta_{02} x) \\ E_y &= -A \frac{\omega \mu_1 z_0}{2\pi} \sin\left(\frac{2\pi}{z_0} z\right) \sin(\omega t - \beta_{02} x) \end{aligned}\right\} \text{H_{02} or TE_{02} waves in the dielectric [6·77]}$$

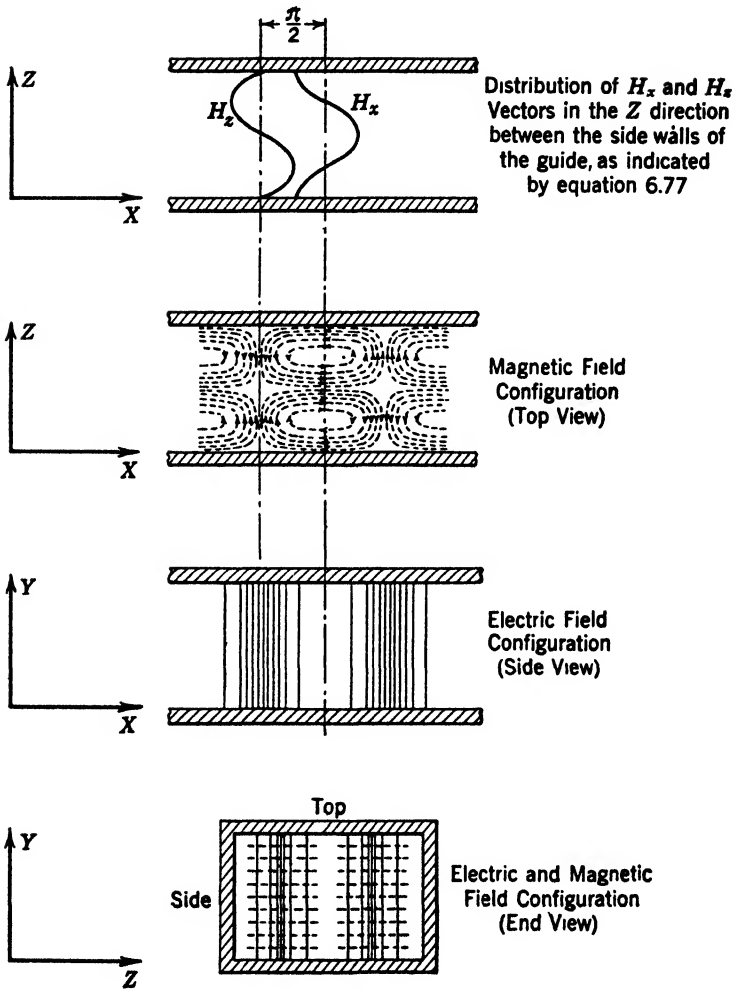


FIG. 6·5 The instantaneous field configuration for an H_{02} or TE_{02} wave in a rectangular wave guide.

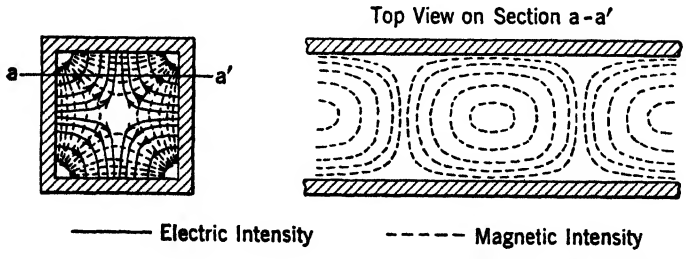
The quantities defining the transmission properties of the H_{02} wave are:

$$\left.\begin{aligned}
&\text{Phase constant} && \beta_{02} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{2\pi}{z_0}\right)^2} \\
&\text{Critical frequency} && (f_0)_{02} = \frac{1}{\sqrt{\mu_1\varepsilon_1}\, z_0} = \frac{v_1}{z_0} \\
&\text{Critical wavelength} && (\lambda_0)_{02} = z_0 \\
&\text{Wavelength in pipe} && (\lambda_p)_{02} = \frac{2\pi}{\beta_{02}} \\
&\text{Phase velocity} && (v_p)_{02} = \frac{\omega}{\beta_{02}} \\
&\text{Group velocity} && (v_g)_{02} = \frac{v_1^2}{(v_p)_{02}}
\end{aligned}\right\} \quad [6\cdot78]$$

The field configuration for the H_{02} wave is shown in Fig. 6·5. Here again the critical wavelength is independent of the y_0 dimension.

6·10 The H_{11} and H_{12} Waves

By setting $n = 1$, $m = 1$ in equations 6·47 we may obtain the expressions for the components of the H_{11} wave. The process is similar to that employed in the previous sections and will not be repeated here. The H_{11} wave configuration is more complicated than the H_{01} and H_{02}



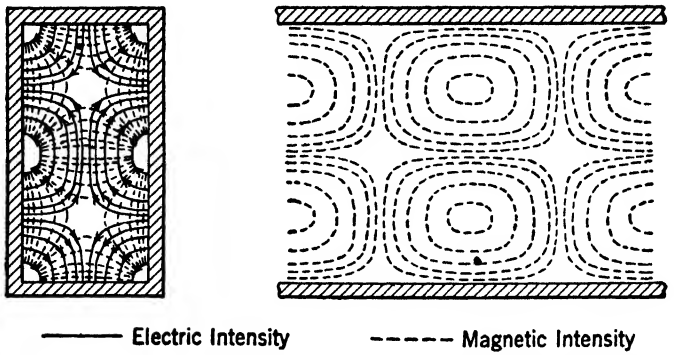
(Chu and Barrow, courtesy of IRE)

FIG. 6·6 Instantaneous field distribution for an H_{11} or TE_{11} wave in a hollow rectangular pipe.

waves. It is shown in Fig. 6·6. The quantities defining the transmission properties of this wave may be obtained from the general equations 6·71 when $n = m = 1$.

When $n = 1$, $m = 2$, the wave configuration is as shown in Fig. 6·7.

It may be noted that the wave pattern for the H_{01} wave serves as a fundamental construction unit for the higher-order H_{0m} waves. Similarly, the H_{11} wave serves as a basic unit for the construction of the higher-order H_{1m} wave patterns. The transmission properties are always obtainable from the general expressions 6·71. The critical frequency



(Chu and Barrow, courtesy of IRE)

FIG. 6·7 Instantaneous field distribution for an H_{12} or TE_{12} wave in a hollow rectangular pipe.

$(f_0)_{nm}$ and the critical wavelength $(\lambda_0)_{nm}$ for the lower-order H_{nm} waves in a square pipe are given in Table 6·1. These data were prepared by Chu and Barrow during their original experimental work on the propagation of waves in tubes.

TABLE 6·1

CRITICAL VALUES FOR AN AIR-FILLED SQUARE PIPE ($y_0 = z_0$)
(z_0 measured in centimeters)

Wave Type	Critical Frequency f_0 in Cycles	Critical Wavelength in Centimeters
H_{01}	$1.50 \times 10^{10}/z_0$	$2.000 \times z_0$
H_{11} and E_{11}	$2.12 \times 10^{10}/z_0$	$1.414 \times z_0$
H_{02}	$3\ 00 \times 10^{10}/z_0$	$1.000 \times z_0$
H_{12} and E_{12}	$3.35 \times 10^{10}/z_0$	$0.894 \times z_0$
H_{22} and E_{22}	$4.24 \times 10^{10}/z_0$	$0.707 \times z_0$
H_{03}	$4\ 50 \times 10^{10}/z_0$	$0.666 \times z_0$
H_{13} and E_{13}	$4.74 \times 10^{10}/z_0$	$0.632 \times z_0$
H_{23} and E_{23}	$5.41 \times 10^{10}/z_0$	$0.554 \times z_0$
H_{04}	$6.00 \times 10^{10}/z_0$	$0\ 500 \times z_0$

(Chu and Barrow, courtesy of IRE)

6·11 Transverse Magnetic (*TM* Mode) *E* Waves

With the restriction that $H_x = 0$, the field equations 6·5 and 6.6 reduce to

$$\left.\begin{aligned} \frac{\partial H_z'}{\partial y} - \frac{\partial H_y'}{\partial z} &= (\sigma + j\omega\varepsilon)\, E_x' \\ \gamma H_z' &= (\sigma + j\omega\varepsilon)\, E_y' \\ -\gamma H_y' &= (\sigma + j\omega\varepsilon) E_z' \end{aligned}\right\} \quad [6\cdot79] \qquad \left.\begin{aligned} \frac{\partial E_z'}{dy} - \frac{\partial E_y'}{\partial z} &= 0 \\ \frac{\partial E_x'}{\partial z} + \gamma E_z' &= -j\omega\mu H_y' \\ \gamma E_y' + \frac{\partial E_x'}{\partial y} &= j\omega\mu H_z' \end{aligned}\right\} \quad [6\cdot80]$$

By an elimination process similar to that employed in section 6·3 we may obtain an equation in E_x' alone from equations 6·79 and 6·80. This equation, in the dielectric, may be written

$$\frac{\partial^2 E_x'}{\partial y^2} + \frac{\partial^2 E_x'}{\partial z^2} = -(\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1)E_x' \qquad [6\cdot81]$$

The solution of this equation may be substituted into relations 6·79 and 6·80 to obtain the components of the E_{nm} waves in a manner similar to that employed in sections 6·4 and 6·5. These components may be written as

$$\left.\begin{aligned} E_x &= B \sin\left(\frac{n\pi}{y_0} y\right) \sin\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\ E_y &= -B \frac{\gamma_{nm}}{\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{n\pi}{y_0}\right) \cos\left(\frac{n\pi}{y_0} y\right) \sin\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\ E_z &= -B \frac{\gamma_{nm}}{\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{m\pi}{z_0}\right) \sin\left(\frac{n\pi}{y_0} y\right) \cos\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\ H_x &= 0 \\ H_y &= B \frac{j\omega\varepsilon_1}{\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{m\pi}{z_0}\right) \sin\left(\frac{n\pi}{y_0} y\right) \cos\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \\ H_z &= -B \frac{j\omega\varepsilon_1}{\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1} \left(\frac{n\pi}{y_0}\right) \cos\left(\frac{n\pi}{y_0} y\right) \sin\left(\frac{m\pi}{z_0} z\right) e^{j\omega t - \gamma_{nm} x} \end{aligned}\right\} \quad \begin{matrix} E_{nm} \text{ or} \\ TM_{nm} \\ \text{waves in} \\ \text{the} \\ \text{dielectric} \\ [6\cdot82] \end{matrix}$$

The constant B depends only upon the original excitation creating the wave and denotes an absolute magnitude only. As in the case of

the H_{nm} waves, the subscript n refers to the number of half sinusoids or maxima of field intensity in the Y direction from $y = 0$ to $y = y_0$. Also the subscript m indicates the number of half sinusoids occurring in the Z direction from $z = 0$ to $z = z_0$.

The expressions for β_{nm}, $(f_0)_{nm}$, $(\lambda_0)_{nm}$, $(\lambda_p)_{nm}$, $(v_p)_{nm}$, and $(v_g)_{nm}$ for the E waves are the same as those already given for the H waves in equations 6·71. Hence the transmission expressions are identical for E and H waves of the same order.

From equations 6·82 it may be seen that, if $n = 0$ or $m = 0$, all components vanish. Therefore, no wave of the type E_{00} is possible. By setting $n = 0$, $m = 1$, or $n = 1$, $m = 0$ it may be seen that waves of the type E_{10} or E_{01} are also not possible. The lowest-order wave which is possible is the E_{11} wave. Waves of higher order, such as E_{12}, E_{13}, etc., are all theoretically possible. Like the H waves, E waves which have complementary indices, for example, E_{12} and E_{21}, are identical except for their orientation in the tube. It is therefore sufficient to consider only one ordering of indices; for example, E_{12}, E_{21}, etc., will give the same field distribution except that the y_0 and z_0 dimensions are interchanged.

6·12 The E_{11} or TM_{11} Wave

The equations for the components of the E_{11} wave may be obtained from equations 6·82 by substituting $n = 1$, $m = 1$ therein. We obtain

$$
\begin{aligned}
E_x &= B \sin\left(\frac{\pi}{y_0} y\right) \sin\left(\frac{\pi}{z_0} z\right) e^{j\omega t - \gamma_{11} x} \\
E_y &= -B \frac{\gamma_{11}}{\gamma_{11}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{\pi}{y_0}\right) \cos\left(\frac{\pi}{y_0} y\right) \sin\left(\frac{\pi}{z_0} z\right) e^{j\omega t - \gamma_{11} x} \\
E_z &= -B \frac{\gamma_{11}}{\gamma_{11}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{\pi}{z_0}\right) \sin\left(\frac{\pi}{y_0} y\right) \cos\left(\frac{\pi}{z_0} z\right) e^{j\omega t - \gamma_{11} x} \\
H_x &= 0 \\
H_y &= B \frac{j\omega \varepsilon_1}{\gamma_{11}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{\pi}{z_0}\right) \sin\left(\frac{\pi}{y_0} y\right) \cos\left(\frac{\pi}{z_0} z\right) e^{j\omega t - \gamma_{11} x} \\
H_z &= -B \frac{j\omega \varepsilon_1}{\gamma_{11}^2 + \omega^2 \mu_1 \varepsilon_1} \left(\frac{\pi}{y_0}\right) \cos\left(\frac{\pi}{y_0} y\right) \sin\left(\frac{\pi}{z_0} z\right) e^{j\omega t - \gamma_{11} x}
\end{aligned}
\qquad [6·83]
$$

The propagation constant γ_{11}, as obtained from equation 6·46 when $n = m = 1$, is

$$\gamma_{11}^2 = \left(\frac{\pi}{y_0}\right)^2 + \left(\frac{\pi}{z_0}\right)^2 - \omega^2 \mu_1 \varepsilon_1 \qquad [6·84]$$

Since $\gamma_{11} = j\beta_{11}$, the final solution for this wave may be written

$$E_x = B\sin\left(\frac{\pi}{y_0}y\right)\sin\left(\frac{\pi}{z_0}z\right)\cos(\omega t-\beta_{11}x)$$

$$E_y = B\frac{\beta_{11}}{\left(\frac{\pi}{y_0}\right)^2+\left(\frac{\pi}{z_0}\right)^2}\left(\frac{\pi}{y_0}\right)\cos\left(\frac{\pi}{y_0}y\right)\sin\left(\frac{\pi}{z_0}z\right)\sin(\omega t-\beta_{11}x)$$

$$E_z = B\frac{\beta_{11}}{\left(\frac{\pi}{y_0}\right)^2+\left(\frac{\pi}{z_0}\right)^2}\left(\frac{\pi}{z_0}\right)\sin\left(\frac{\pi}{y_0}y\right)\cos\left(\frac{\pi}{z_0}z\right)\sin(\omega t-\beta_{11}x)$$

$$H_x = 0$$

$$H_y = -B\frac{\omega\varepsilon_1}{\left(\frac{\pi}{y_0}\right)^2+\left(\frac{\pi}{z_0}\right)^2}\left(\frac{\pi}{z_0}\right)\sin\left(\frac{\pi}{y_0}y\right)\cos\left(\frac{\pi}{z_0}z\right)\sin(\omega t-\beta_{11}x)$$

$$H_z = B\frac{\omega\varepsilon_1}{\left(\frac{\pi}{y_0}\right)^2+\left(\frac{\pi}{z_0}\right)^2}\left(\frac{\pi}{y_0}\right)\cos\left(\frac{\pi}{y_0}y\right)\sin\left(\frac{\pi}{z_0}z\right)\sin(\omega t-\beta_{11}x)$$

E_{11} or TM_{11} waves in the dielectric [6·85]

The quantities defining the transmission properties of the E_{11} wave are, from equation 6·71,

Phase constant $$\beta_{11} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{\pi}{y_0}\right)^2 - \left(\frac{\pi}{z_0}\right)^2}$$

Critical frequency $$(f_0)_{11} = \frac{v_1}{2}\sqrt{\left(\frac{1}{y_0}\right)^2+\left(\frac{1}{z_0}\right)^2}$$

Critical wavelength $$(\lambda_0)_{11} = \frac{2}{\sqrt{\left(\frac{1}{y_0}\right)^2+\left(\frac{1}{z_0}\right)^2}}$$ [6·86]

Wavelength in tube $$(\lambda_p)_{11} = \frac{2\pi}{\beta_{11}}$$

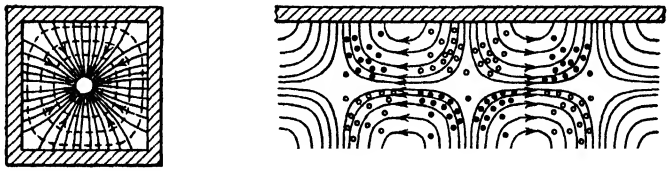
Phase velocity $$(v_p)_{11} = \frac{\omega}{\beta_{11}}$$

Group velocity $$(v_g)_{11} = \frac{v_1^2}{(v_p)_{11}}$$

where $v_1 = 1/\sqrt{\mu_1\varepsilon_1}$.

The field configuration for the E_{11} wave is shown in Fig. 6·8. It was shown in section 6·8 that the critical frequency of the lowest-order

H wave, the H_{01} wave, is $(f_0)_{01} = v_1/2z_0$. This critical frequency is lower than that for the lowest-order E wave as given in equation 6·86. For higher-order waves, the critical frequency is higher. Since the smallest possible guide for a given frequency is usually desirable, the H_{01} has the greater practical significance.

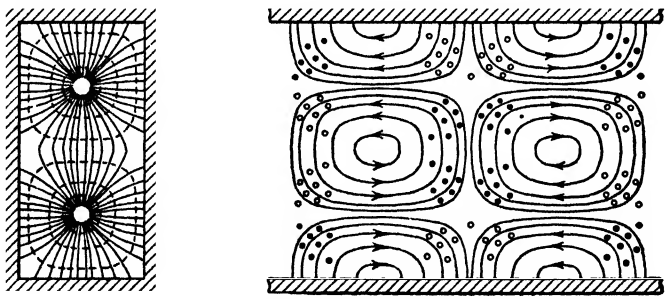


(Chu and Barrow, courtesy of IRE)

FIG. 6·8 Instantaneous field configuration for the E_{11} or TM_{11} wave in a hollow rectangular pipe.

6·13 The E_{12} or TM_{12} Wave

The components of the E_{12} wave may be obtained in a similar manner from equations 6·82 by substituting $n = 1$, $m = 2$. The wave configuration is shown in Fig. 6·9. The propagation properties are given by the general equations 6·71. The diagram, Fig. 6·8, showing the con-



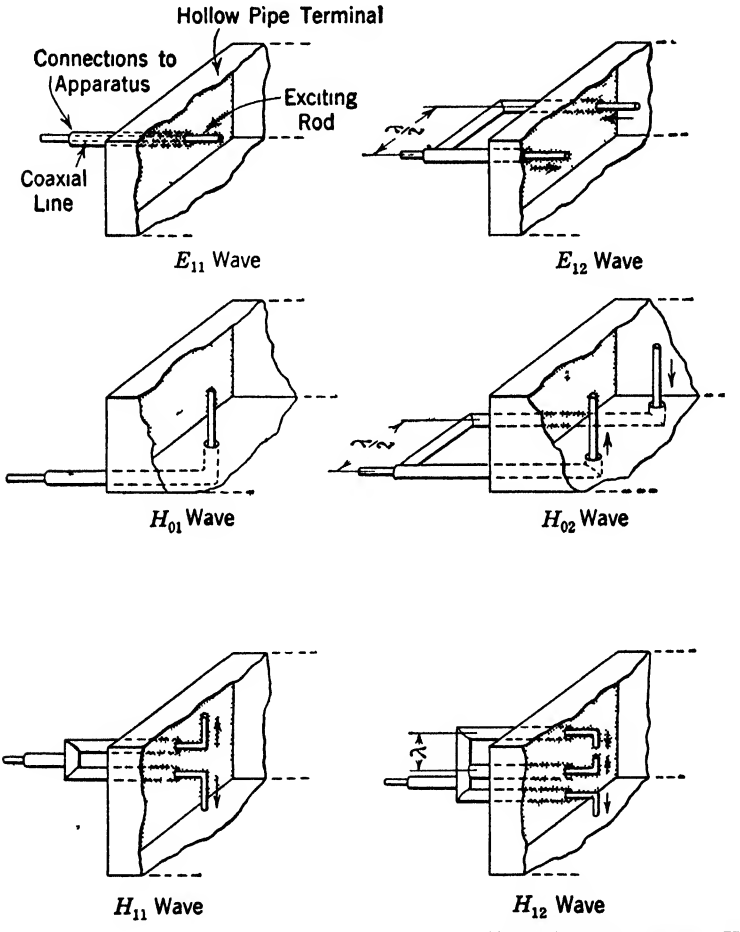
(Chu and Barrow, courtesy of IRE)

FIG. 6·9 Instantaneous field configuration for the E_{12} or TM_{12} wave.

figuration of the E_{11} wave may be regarded as a fundamental unit, from which wave configurations for the E_{1m} waves are made up. The E_{12} wave includes two such units; the E_{13} wave includes three, etc. Like the higher-order H waves, the higher-order E waves involve such extremely high frequencies in guides of reasonable size that their practical importance is doubtful. The critical frequencies and wavelengths of the lower-order E waves in square pipes are included in Table 6.1.

6·14 Terminal Devices

In exciting and receiving H_{nm} and E_{nm} waves in rectangular tubes. terminal devices may be used such as illustrated in Fig. 6·10. These devices originally suggested by Barrow and Chu serve to establish the



(Chu and Barrow, courtesy of IRE)

FIG. 6·10 Terminal devices designed by Chu and Barrow for exciting the various wave types in rectangular wave guides.

various wave configurations which have been illustrated in the figures of the chapter; they will also act as receivers of energy for the various transmission modes indicated.

It will be noted, by comparison of the particular terminating device with the wave configuration which it is designed to excite, that the

conductors injected into the pipes coincide with the lines of electric intensity E of the field pattern. This is the basic principle involved in the construction. The injected conductors are simply extensions of the central conductor of the coaxial line which connects the pipe to the exciting source or to the receiver. It will also be observed from these figures that the phase of the currents delivered to the rods, or antennas as they might be called, must be taken into consideration. By referring again to the field configuration desired we may determine the relative direction of the currents required at any given instant. This direction is indicated in the figure with small arrows shown near the antennas. In order that the antennas be excited in the proper phase, sections of coaxial line of suitable length are inserted. The length of the sections required are indicated in the figure. In a later chapter the reasons for the choice of these lengths will be explained.

In the case of the H_{nm} waves it will be seen that the antennas are at right angles to the direction in which the waves will be propagated. The antennas for exciting the E_{nm} waves are in the direction of wave propagation. The length of the antennas inside the pipe is an important consideration, and it is desirable to make provision for their adjustment. The distance from the end of the pipe to the antenna should also be adjustable, so that the exact position for best operation may be obtained easily. Other types of feeding than that employing concentric lines may be satisfactory and will be discussed later. Also, in connection with reception, it is sometimes desirable to employ a crystal detector or vacuum tube, mounted in the antenna itself inside the guide. Devices of this type will be discussed in Chapter 8.

6·15 Further Discussion of the H_{01} Wave

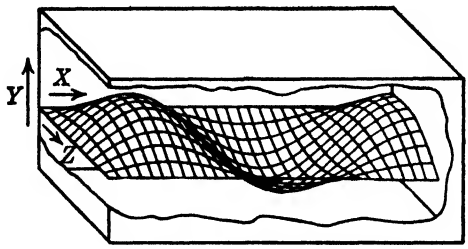
The H_{01} wave has received considerable attention by Chu and Barrow, who first called attention to its outstanding characteristics. The following discussion is based upon their research.

It may be observed by comparison of the H_{01} wave pattern shown in Fig. 6·3 with other H_{nm} and E_{nm} wave patterns that the simplest wave configuration is obtained with the H_{01} wave. This simplification is due to the absence of all the components of electric intensity except E_y. Also, this wave has the lowest critical frequency of all the possible transmission modes, and in addition has the lowest attenuation, as will be shown later.

A plot of the electric intensity vector E_y of equation 6·75 of the H_{01} wave is shown isometrically in Fig. 6·11, where the magnitude of the vector is represented by the vertical distance from a plane through the pipe, to the curved surface. This vector is independent of y but is shown

to vary sinusoidally in the X and Z directions. Time is held constant in the diagram.

We obtained in equation 6·76 a relation between the wavelength of the H_{01} wave in the pipe, $(\lambda_p)_{01}$, and the wavelength λ_1 of a free wave in a



(Chu and Barrow, courtesy of IRE)

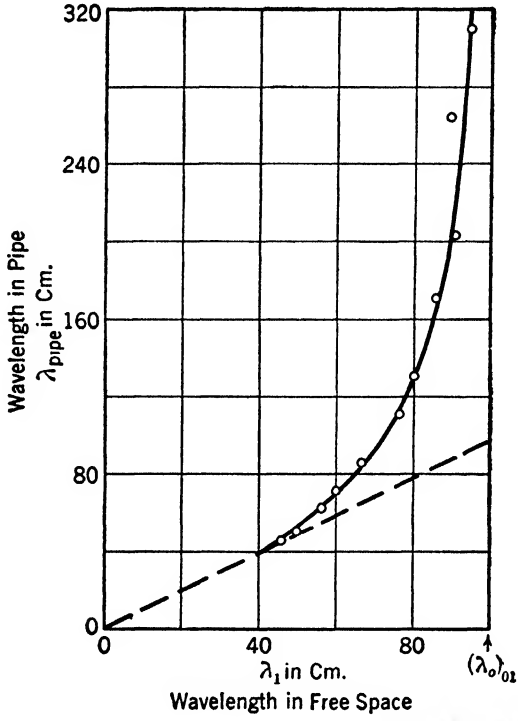
FIG. 6·11 Instantaneous distribution of the E_y vector of the H_{01} wave in a rectangular wave guide. The magnitude of the vector is represented by the vertical distance from a plane through the pipe to the curved surface. This vector is independent of y but varies sinusoidally in both the X and Z directions.

medium of dielectric constant ϵ_1 and permeability μ_1. This relation is given by

$$(\lambda_p)_{01} = \frac{\lambda_1}{\sqrt{1 - \left(\frac{\lambda_1}{(\lambda_0)_{01}}\right)^2}} \qquad [6\cdot87]$$

and is shown graphically in Fig. 6·12. It is seen that, as the wavelength λ_1 approaches the critical wavelength, the wavelength in the pipe approaches infinity. Further, as the wavelength λ_1 is decreased, the wavelength in the pipe approaches the free-space wavelength. From $(v_p)_{01}$ in equation 6·76 it is seen that the phase velocity of the wave is always greater in the pipe than the velocity of light. It is commonly believed that no velocities in nature can exceed the velocity of light. This paradox is referred to in optics as anomalous dispersion. A correct statement of the above idea is that *energy* is never propagated with a velocity greater than that of light. In free waves, the equiphase surfaces are normal to the direction of propagation, and hence the phase velocity is the velocity with which the energy of the wave is propagated. Also, with free waves, the phase velocity is the velocity of light. With bounded waves, however, the situation is entirely different, for the phase velocity no longer indicates the velocity at which the energy of the wave is propagated. If we assume the existence of a wave in a bounded medium, the equiphase surfaces of the wave will no longer be perpendicular to the direction of propagation. Thus if we are dis-

cussing the phase velocity of bounded waves, we cannot think of the energy being propagated to the point where we have already assumed it to be. This matter may be further clarified by reference to section 5·8.



(*Chu and Barrow, courtesy of IRE*)

FIG. 6·12 Relation between the wavelength in air and the wavelength in a rectangular pipe for the H_{01} wave. (The solid curve is the theoretically derived relation equation 6·87. The dots correspond to experimental points obtained by Chu and Barrow.)

6·16 Voltage, Current, and Power in Perfectly Conducting Wave Guides (H_{01} Wave)

We may, with the aid of our definition of voltage and current in terms of line integrals, obtain expressions for voltage and current in the waveguide system. The potential V between two points P and Q is defined as

$$V = \int_P^Q E \cos\theta \, dl \qquad [6\cdot88]$$

If we choose the path of integration to coincide with the direction of the electric intensity E_y, then $\theta = 0$ and $\cos\theta = 1$. This is true when P

and Q lie on a vertical line in the Y direction of Fig. 6·13. Let us choose P and Q to lie on a line cutting the Z axis at $z_0/2$, and parallel to the Y axis. Then the potential difference between the two points which are on the inside surface of the pipe is

$$V = \int_{y=0}^{y=y_0} -E_y \, dy \qquad [6\cdot89]$$

Substituting for E_y from equations 6·75

$$V = \int_{y=0}^{y=y_0} A \frac{\omega\mu_1 z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) \sin(\omega t - \beta_{01} x) \, dy \qquad [6\cdot90]$$

Since in this case $z = z_0/2$

$$\sin\left(\frac{\pi}{z_0} z\right) = \sin\left(\frac{\pi}{2}\right) = 1$$

we obtain

$$V = A \frac{\omega\mu_1 y_0 z_0}{\pi} \sin(\omega t - \beta_{01} x) \qquad [6\cdot91]$$

Thus the potential difference between the inside surface of the top and bottom walls of the pipe varies sinusoidally in the X direction down the center of the pipe. Its maximum value is given by

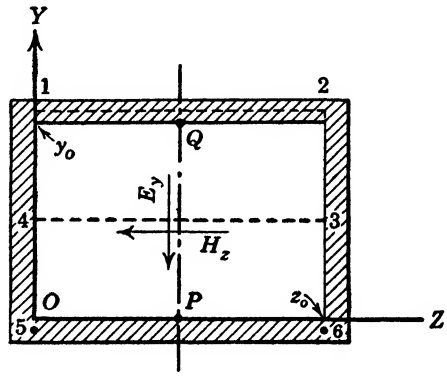


FIG. 6·13 Paths of integration for obtaining voltage and current relations in wave guide.

$$V' = A \frac{\omega\mu_1 y_0 z_0}{\pi} \qquad [6\cdot92]^*$$

and this value in turn varies sinusoidally in the Z direction.

* We may obtain any value between zero and the value defined by equation 6·91 by other choices of the path of integration. Integrating along the surface of the metal (path POy_0Q) we find the electric intensity in the direction of the path everywhere zero, since the tangential component of E is zero at the boundary. As we shorten the path of integration between P and Q we steadily increase the value of the integral.

This situation may at first be rather startling, but it is perfectly correct. The system is a dynamic rather than a static one, and the line integral of the electric intensity is never independent of the path in such a system. In particular, the voltage found by integrating around any closed loop is equal to the rate of change of magnetic flux through the loop. This is merely Faraday's law, expressed by Maxwell as

$$\text{Curl } E = -\frac{\partial B}{\partial t}$$

We may calculate the longitudinal current flowing along the surface of the top or bottom walls of the pipe by taking the line integral of the magnetic intensity along a closed path like the dotted path 1234 in Fig. 6·13. Since H is the intensity in the Z direction, the product $H \cos\theta$ is zero along the 14 and 23 sections of the path. This follows from the fact that the angle θ between H and the path is 90°. Along the path 12, the product $H \cos\theta$ is zero since H is zero in the interior of a perfect conductor. The discontinuity in H at the surface is explained by the existence of the current on the surface. The only contribution to the line integral around the closed path 1234 is made along the path 43. Hence

$$I = \int_{z=0}^{z=z_0} -H_z \, dz$$

Substituting for H_z from equations 6·75 we obtain:

$$I = \int_{z=z_0}^{z=0} A \frac{\beta_{01} z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) \sin(\omega t - \beta_{01} x) \, dz \qquad [6\cdot93]$$

$$= A \frac{\beta_{01} z_0}{\pi} \sin(\omega t - \beta_{01} x) \frac{z_0}{\pi} \cos\left(\frac{\pi}{z_0} z\right) \Bigg|_{z=z_0}^{z=0}$$

$$= 2A\beta_{01} \frac{z_0^2}{\pi^2} \sin(\omega t - \beta_{01} x) \qquad [6\cdot94]$$

The maximum value of the current is therefore

$$I' = 2A\beta_{01} \frac{z_0^2}{\pi^2} \qquad [6\cdot95]$$

and this amplitude varies sinusoidally in the X direction. The current along the inside surface of the bottom of the pipe is at any instant moving in the opposite direction to that on the inside surface of the top of the pipe.* That is, the two currents are in opposite directions, and they are also of equal magnitude for any given value of x.

We may calculate the average power flowing through the pipe by means of the Poynting vector. The power p_x streaming through a square meter of surface parallel to the YOZ plane as given by equation 3·87 is

$$p_x = \tfrac{1}{2}(E_y' H_z' - E_z' H_y') \qquad [3\cdot87]$$

Since, in the case of the H_{01} wave, $E_z' = H_y' = 0$, this reduces to

$$p_x = \tfrac{1}{2}(E_y' H_z') \qquad [6\cdot96]$$

* This may readily be verified by inspection of Fig. 6·13. For example, if we integrate around a path 4365, in the same sense as before, work is done on the field, whereas in the former case the field does work on the imaginary exploring charge.

Substituting for E_y' and H_z' from equation 6·75, we have

$$E_y'H_z' = A^2 \frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2} \sin^2 \left(\frac{\pi}{z_0} z \right)$$

so that

$$p_x = \tfrac{1}{2} A^2 \frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2} \sin^2 \left(\frac{\pi}{z_0} z \right) \qquad [6\cdot97]$$

The average power P_x flowing through the pipe is

$$P_x = \tfrac{1}{2} A^2 \frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2} \int_{y=0}^{y=y_0} \int_{z=0}^{z=z_0} \sin^2 \left(\frac{\pi}{z_0} z \right) dz\, dy \qquad [6\cdot98]$$

$$P_x = \tfrac{1}{4} A^2 \frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2} y_0 z_0 \qquad [6\cdot99]$$

It is seen that the transmitted power is directly proportional to the cross-sectional area $y_0 z_0$ of the pipe.

A characteristic impedance, Z_0, may be defined in several different ways by the use of relations for I, V, and P_x which we have obtained. Thus, on a voltage-current basis it becomes

$$Z_0 = \frac{V'}{I'} = \frac{\omega \mu_1 \pi y_0}{2 \beta_{01} z_0} \qquad [6\cdot100]$$

and on a power basis it is

$$Z_0 = \frac{P_x}{(I_{\text{rms}})^2} = \frac{2P_x}{(I')^2} = \frac{\omega \mu_1 \pi^2 y_0}{8 \beta_{01} z_0} = 0.785 \frac{\omega \mu_1 \pi y_0}{2 \beta_{01} z_0} \qquad [6\cdot101]$$

We may also define an intrinsic impedance η by the relation

$$\eta = \frac{E_y}{H_z} = \frac{\omega \mu_1}{\beta_{01}} \qquad [6\cdot102]$$

When the dielectric in the pipe is air it may be shown that η approaches the value η_0 for the intrinsic impedance of free space, i.e., $\eta_0 = 377\Omega$, as y_0 and z_0 are made very large in comparison to the wavelength λ_1 of the wave in free space.

6·17 Resolution of the H_{01} Wave into Elementary Waves

In order to obtain a better physical picture of the H_{01} wave, it is convenient to change the expressions for this wave as given in equation 6·75 to an equivalent form in which the wave appears as a superposition of two ordinary plane waves. The procedure is similar to that employed in section 5·15, for the E_1 wave between parallel planes, and will not be repeated here in detail. It will be shown that the two elementary waves are repeatedly reflected back and forth between the side walls of the pipe.

It is necessary first to express the H_{01} wave in terms of an X axis down the center of the pipe as shown in Fig. 6·14. To do this we must go back to equation 6·32 and choose as our solution

$$H_x' = A \cos \sqrt{A_1}\, y \sin \sqrt{A_2}\, z$$

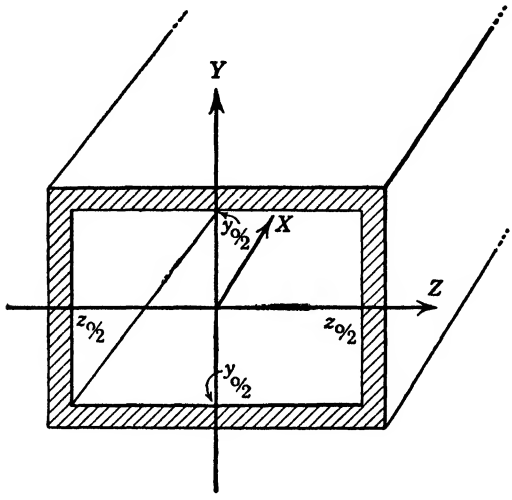


FIG. 6·14 Choice of axes in rectangular wave guide for analysis of elementary waves.

in place of the cosine solution. This is necessary since the previous solution was selected because we knew from experience that it would satisfy the required boundary conditions. We might have written down the general solution in the first place, but the extra complication of carrying the useless terms would have added unnecessary difficulty. With this choice, the general equation for the H_{nm} waves will be the same as given in equations 6·47 if, in the periodic term involving z, the sine replaces the cosine, and vice versa, and the polarity of E_y and H_z is reversed. The wave that results when $n = 0$, $m = 1$ is then

$$\left.\begin{aligned} H_x &= A \sin\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \\ H_z &= -jA \frac{\beta_{01} z_0}{\pi} \cos\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \\ E_y &= -jA \frac{\omega \mu_1 z_0}{\pi} \cos\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \end{aligned}\right\} \begin{array}{l} H_{01} \text{ wave in the} \\ \text{dielectric when} \\ X \text{ axis is chosen} \\ \text{in center of tube} \\ [6\text{·}103] \end{array}$$

Let us consider first the E_y component. We may replace the $\cos\left(\frac{\pi}{z_0}z\right)$ with its equivalent exponential form

$$\tfrac{1}{2}\left\{e^{j\left(\frac{\pi}{z_0}z\right)} + e^{-j\left(\frac{\pi}{z_0}z\right)}\right\}$$

and obtain

$$E_y = -jA\frac{\omega\mu_1 z_0}{2\pi}\left\{e^{-j\left(\beta_{01}x-\frac{\pi}{z_0}z\right)+j\omega t} + e^{-j\left(\beta_{01}x+\frac{\pi}{z_0}z\right)+j\omega t}\right\} \qquad [6\cdot104]$$

Thus E_y is separated into two elementary components $E_y^{(1)}$ and $E_y^{(2)}$ defined by

$$E_y^{(1)} = -jA\frac{\omega\mu_1 z_0}{2\pi}e^{-j\left(\beta_{01}x-\frac{\pi}{z_0}z\right)+j\omega t} = -jA\frac{\omega\mu_1 z_0}{2\pi}e^{-j\beta_{01}\left\{x-\frac{\pi}{\beta_{01}z_0}z-(v_p)_{01}t\right\}} \qquad [6\cdot105]$$

$$E_y^{(2)} = -jA\frac{\omega\mu_1 z_0}{2\pi}e^{-j\left(\beta_{01}x+\frac{\pi}{z_0}z\right)+j\omega t} = -jA\frac{\omega\mu_1 z_0}{2\pi}e^{-j\beta_{01}\left\{x+\frac{\pi}{\beta_{01}z_0}z-(v_p)_{01}t\right\}} \qquad [6\cdot106]$$

where

$$E_y = E_y^{(1)} + E_y^{(2)} \qquad [6\cdot107]$$

As in the parallel-plane case discussed in the previous chapter, it can be shown that these vector components may be interpreted as two elementary component waves traveling in different directions. This may be conveniently done by comparing the equations for the elementary components with the general equation of a plane wave, i.e.,

$$C(s - vt)$$

where C is the amplitude of the wave and $s = lx + my + nz$, l, m, and n being the direction cosines of s. Thus by comparison with equation 6·105 we see that

$$s = x - \frac{\pi}{\beta_{01}z_0}z$$

and hence $l = 1$, $m = 0$, and $n = -\frac{\pi}{\beta_{01}z_0}$. The geometrical interpretation of the exponent $x - \frac{\pi}{\beta_{01}z_0}z$ is shown in Fig. 6·15. Therefore equation 6·105 represents a component wave traveling in the $X^{(1)}$ direction at an angle

$$\theta^{(1)} = \tan^{-1}\left(-\frac{\pi}{\beta_{01}z_0}\right)$$

with the X axis. Similarly, it may be shown that equation 6·106 represents a component wave traveling in the $X^{(2)}$ direction at an angle $\theta^{(2)} = \tan^{-1}\left(\frac{\pi}{\beta_{01} z_0}\right)$ with the X axis. Hence the E_y component is expressed in terms of two component plane waves $E_y^{(1)}$ and $E_y^{(2)}$ which are traveling in the $X^{(1)}$ and $X^{(2)}$ directions, respectively. For pur-

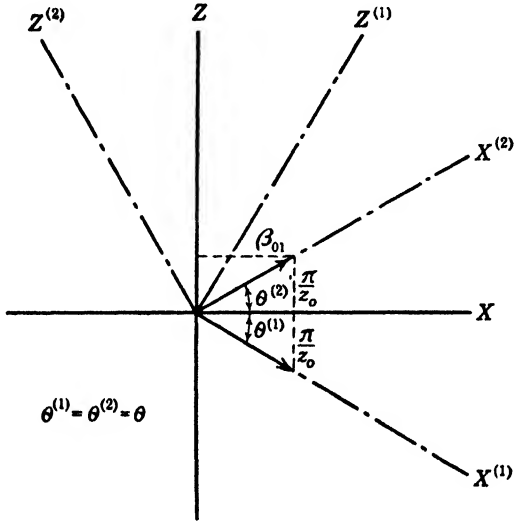


FIG. 6·15 The elementary components of the H_{01} wave are resolved along the new axes $X^{(1)}Z^{(1)}$ and $X^{(2)}Z^{(2)}$, which are rotated an angle θ from the old axes. The Y axis remains the same.

poses of analysis, we may express $E_y^{(1)}$ in terms of a new system of axes $X^{(1)}Z^{(1)}$, and $E_y^{(2)}$ in terms of the new axes $X^{(2)}Z^{(2)}$. The analytic procedure required has been explained in detail for a similar case in section 5·15. Briefly, the procedure in this case is as follows:

Let us first consider the $E_y^{(1)}$ exponent. To make the required transformation, replace x and z in equation 6·105 with the values

$$\begin{aligned} x &= x^{(1)} \cos\theta + z^{(1)} \sin\theta \\ z &= z^{(1)} \cos\theta - x^{(1)} \sin\theta \end{aligned} \qquad [6\cdot108]$$

Then

$$\beta_{01} x - \frac{\pi}{z_0} z = \left(\beta_{01} \cos\theta + \frac{\pi}{z_0} \sin\theta\right) x^{(1)} + \left(\beta_{01} \sin\theta - \frac{\pi}{z_0} \cos\theta\right) z^{(1)} \qquad [6\cdot109]$$

Substitution for the $\sin\theta$ and $\cos\theta$ from Fig. 6·15 reduces this to $(\omega/v_1)x^{(1)}$. Carrying through a similar operation on equation 6·106 and

replacing x and z by the values

$$\begin{aligned} x &= x^{(2)} \cos\theta + z^{(2)} \sin\theta \\ z &= z^{(2)} \cos\theta - x^{(2)} \sin\theta \end{aligned} \qquad [6\cdot110]$$

we find that the exponent $\beta_{01}x + \frac{\pi}{z_0} z$ reduces to $\frac{\omega}{v_1} x^{(2)}$. Thus the elementary components of $E_y^{(1)}$ and $E_y^{(2)}$ take the form

$$E_{y(1)}^{(1)} = -jA \frac{\omega\mu_1 z_0}{2\pi} e^{-j\left(\frac{\omega}{v_1}x^{(1)} - \omega t\right)} \qquad [6\cdot111]$$

$$E_{y(2)}^{(2)} = -jA \frac{\omega\mu_1 z_0}{2\pi} e^{-j\left(\frac{\omega}{v_1}x^{(2)} - \omega t\right)} \qquad [6\cdot112]$$

A similar rearrangement and transformation may be carried out for the H_x and H_z components. In this manner we may obtain for the H_{01} wave the expression

$$\begin{aligned} H_x &= [-jA\tfrac{1}{2}]\, e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} && - [-jA\tfrac{1}{2}] e^{-j\omega\left(\frac{x^{(2)}}{v_1} - t\right)} \\ H_z &= \left[-jA \frac{\beta_{01} z_0}{2\pi}\right] e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} && + \left[-jA \frac{\beta_{01} z_0}{2\pi}\right] e^{-j\omega\left(\frac{x^{(2)}}{v_1} - t\right)} \\ E_y &= \left[-jA \frac{\omega\mu_1 z_0}{2\pi}\right] e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} && - \left[-jA \frac{\omega\mu_1 z_0}{2\pi}\right] e^{-j\omega\left(\frac{x^{(2)}}{v_1} - t\right)} \end{aligned} \qquad [6\cdot113]$$

H_{01} Wave Group I — H_{01} Wave Group II

The H_{01} wave may therefore be divided into two groups, one of which represents a plane wave traveling in the $X^{(1)}$ direction and the other a plane wave traveling in the $X^{(2)}$ direction. In both these groups, E_y is transverse to the direction of propagation, being in the Y direction. It can be shown that the H vector in each group of waves is normal to its direction of propagation. As this was proved in the analogous study of the E_1 wave in section 5·15 in some detail, it will receive but brief treatment here. Consider the first group of elementary waves. Resolution of the H_x and H_z components of this group along and at right angles to the direction of propagation $X^{(1)}$ shows that the components of H_x and H_y along $X^{(1)}$ cancel and components of H_x and H_y at right angles to $X^{(1)}$ add. Thus, from Fig. 6·16

$$H_{x(1)}^{(1)} = H_x^{(1)} \cos\theta - H_z^{(1)} \sin\theta \qquad [6\cdot114]$$

and

$$H_{z(1)}^{(1)} = H_x^{(1)} \sin\theta + H_z^{(1)} \cos\theta \qquad [6\cdot115]$$

From Fig. 6·15 we have

$$\sin\theta = \frac{\pi}{z_0\sqrt{\beta_{01}^2 + \left(\frac{\pi}{z_0}\right)^2}} \quad \text{and} \quad \cos\theta = \frac{\beta_{01}}{\sqrt{\beta_{01}^2 + \left(\frac{\pi}{z_0}\right)^2}}$$

But

$$\beta_{01} = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{\pi}{z_0}\right)^2}$$

from equation 6·76. Hence

$$\sin\theta = \frac{\pi v_1}{z_0\omega} \quad \text{and} \quad \cos\theta = \frac{\beta_{01}v_1}{\omega} \qquad [6\cdot116]$$

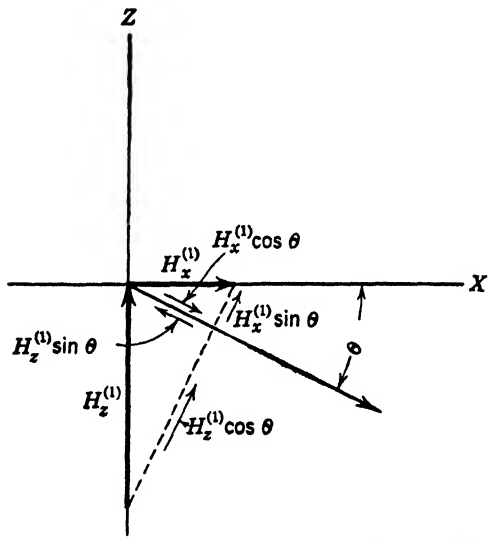


FIG. 6·16 Resolution of the first group of elementary components along and at right angles to the direction of propagation of this group.

Thus, from equations 6·115 and 6·113,

$$\begin{aligned} H_{z(1)}^{(1)} &= -jA\,\frac{1}{2}\left\{\frac{\pi v_1}{z_0\omega} + \frac{\beta_{01}^2 z_0 v_1}{\pi\omega}\right\} e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} \\ &= -j\,\frac{Az_0v_1}{2\pi\omega}\left\{\left(\frac{\pi}{z_0}\right)^2 + \beta_{01}^2\right\} e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} \qquad [6\cdot117] \\ &= -j\,\frac{Az_0v_1}{2\pi\omega}\left(\frac{\omega}{v_1}\right)^2 e^{-j\omega\left(\frac{x^{(1)}}{v_1} - t\right)} \end{aligned}$$

The final expressions for the elementary wave represented by the first group may be written

$$\left.\begin{aligned} E_{y(1)}^{(1)} &= -jA\,\frac{\omega\mu_1 z_0}{2\pi}\,e^{-j\omega\left(\frac{x^{(1)}}{v_1}-t\right)} \\ H_{z(1)}^{(1)} &= -jA\,\frac{\omega z_0}{2\pi v_1}\,e^{-j\omega\left(\frac{x^{(1)}}{v_1}-t\right)} \end{aligned}\right\} \quad \begin{array}{l} H_{01} \text{ elementary wave} \\ \text{traveling in the} \\ X^{(1)} \text{ direction} \\ [6\cdot118] \end{array}$$

In a similar manner we may obtain from the second group

$$\left.\begin{aligned} E_{y(2)}^{(2)} &= -jA\,\frac{\omega\mu_1 z_0}{2\pi}\,e^{-j\omega\left(\frac{x^{(2)}}{v_1}-t\right)} \\ H_{z(2)}^{(2)} &= -jA\,\frac{\omega z_0}{2\pi v_1}\,e^{-j\omega\left(\frac{x^{(2)}}{v_1}-t\right)} \end{aligned}\right\} \quad \begin{array}{l} H_{01} \\ \text{elementary wave} \\ \text{traveling in the} \\ X^{(2)} \text{ direction} \\ [6\cdot119] \end{array}$$

Thus these elementary waves are plane waves having the electric and magnetic vectors at right angles to one another and to their direction of propagation, and having a velocity v_1 of a free wave in a medium of dielectric constant ε_1 and permeability μ_1. The direction which these waves take as they progress in the dielectric medium within the tube is shown in Fig. 6·17.

Equations 6·118 and 6·119 for the elementary waves are entirely equivalent to equations 6·103 for the H_{01} wave from which they were obtained. We see, therefore, that the H_{01} wave consists of the superposition of two elementary plane waves, which are reflected back and forth by the metal side walls of the tube. Equation 6·118 may be considered as a wave striking the left side wall at an angle θ, and equation 6·119 may be considered as a wave being reflected from this side wall at an angle θ and traveling from left to right along the direction $X^{(2)}$. At the right side wall the roles of the waves are interchanged. We may express the angle θ, which determines the $X^{(1)}$ and $X^{(2)}$ directions as a function of the frequency of the waves. From Fig. 6·15

$$\theta = \tan^{-1}\frac{\pi}{\beta_{01}z_0} = \tan^{-1}\left(\frac{1}{\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2 - 1}}\right) \qquad [6\cdot120]^*$$

Thus, as the frequency f approaches the critical frequency $(f_0)_{01}$, $\tan\theta$ approaches ∞ and θ approaches 90°. At $\theta = 90°$ there is no transmission in the X direction since the waves bounce back and forth between the side walls. Thus the group velocity $(v_g)_{01} = 0$ at $\theta = 90°$, while the phase velocity is infinite since the plane of the wave is parallel

* See footnote on page 154 for development of an analogous expression.

to the side walls of the tube, and two points an infinite distance from one another along the X axis are in the equiphase surface of the waves. When f is slightly greater than the critical value, θ is less than 90° and there is a slow progression of the wave in the X direction as indicated in Fig. 5·18. The remarks made on page 155 in regard to the physical representation of the process taking place are equally applicable here.

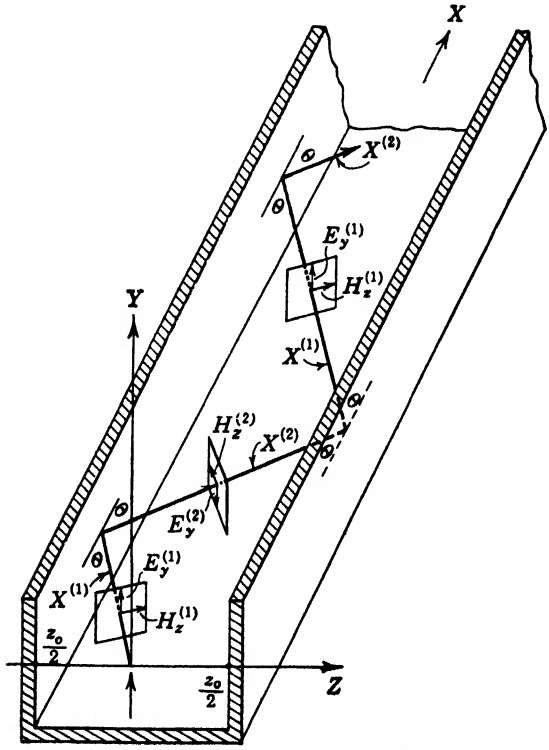


FIG. 6·17 ± Direction taken by elementary waves as they progress through the rectangular pipe.

6·18 Imperfectly Conducting Tubes

It was pointed out in section 5·16 that the problem of obtaining an exact solution of the wave equations in a region bounded by a metal of finite conductivity is extremely difficult, and as yet no complete solution has been found. The problem in the case of rectangular tubes can, however, be treated in a manner similar to that employed in Chapter 5 for the parallel-plate problem.

Again we will assume that the distribution of the field vectors in tubes of imperfect conductivity is not materially changed from that

which is obtained when the boundary walls are perfectly conducting. The imperfect conductivity permits a finite dissipation of energy in the metal. In order that energy flow from the dielectric into the metal we know from the Poynting theorem that there must be a component of electric intensity at the surface of the metal. We do not know the magnitude of this component, but we may estimate it in the following way. We know the tangential component of H at all four bounding surfaces, and we also know that these components are continuous at these boundaries. From the solution of the wave equations in an infinite metallic medium which was obtained in section 5·17 we obtained the relation $E = \eta H$, which holds wherever $\sigma \gg \omega\varepsilon$.

The fact that the relation $E = \eta H$ was derived for a medium of infinite extent is no serious objection to our adopting it in this analysis, provided that the tube wall is sufficiently thick that currents in the metal do not reach the outside surface. Also most practical metals meet the requirement that $\sigma \gg \omega\varepsilon$ at physically realizable frequencies. We may, therefore, use the relation $E = \eta H$, and in this way obtain each of the tangential components of E at the surface of the metal, from our knowledge of the tangential component of H there. Then by the use of the Poynting vector we may calculate the flow of energy or power across the surface formed by the dielectric and the metal. In this manner we may deduce the power lost from the wave as it is transmitted through the dielectric which fills the tube. Since η is quite small, the E vector introduced will be small in comparison to the general field vectors, and the vector distribution in the tube is not materially disturbed.

6·19 The Propagation of the H_{01} Wave in an Imperfectly Conducting Tube

Equations 6·75 for the H_{01} wave in the dielectric, which were derived on the assumption of perfect conductivity, may be written in exponential form as

$$\left.\begin{aligned} H_x &= A \cos\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \\ H_z &= jA \frac{\beta_{01} z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \\ E_y &= jA \frac{\omega \mu_1 z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) e^{j(\omega t - \beta_{01} x)} \end{aligned}\right\} \quad \begin{matrix} H_{01} \text{ wave in the dielectric} \\ [6\cdot121] \end{matrix}$$

If we assume that these expressions are not materially changed when the tube is made of an imperfectly conducting material the only effect

which the metal boundary of finite conductivity will have will be a reduction in the amplitude of the component of the wave as it progresses in the X direction through the tube. The introduction of the attenuation factor $e^{-\alpha x}$ into the amplitude of the various components will not affect the distribution of the wave, but will account for the loss of power during transmission. Thus for the components of the H_{01} wave in the dielectric with imperfect conductivity we may write

$$\left.\begin{aligned} H_{x1} &= A \cos\left(\frac{\pi}{z_0} z\right) e^{-\alpha x} e^{j(\omega t - \beta_{01} x)} \\ H_{z1} &= jA \frac{\beta_{01} z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) e^{-\alpha x} e^{j(\omega t - \beta_{01} x)} \\ E_{y1} &= jA \frac{\omega \mu_1 z_0}{\pi} \sin\left(\frac{\pi}{z_0} z\right) e^{-\alpha x} e^{j(\omega t - \beta_{01} x)} \end{aligned}\right\} \quad \begin{array}{l} H_{01} \text{ waves in the} \\ \text{dielectric when} \\ \text{conductivity is} \\ \text{finite} \\ [6{\cdot}122]^* \end{array}$$

These equations were discussed in section 6·8 and subsequent sections, for the case of perfect conductivity ($\sigma_2 = \infty$), where $\alpha = 0$. We will now determine the attenuation factor in the case of imperfect conductivity.

Consider the instantaneous conditions represented diagrammatically in Fig. 6·18. The tangential component of the magnetic intensity H at the upper and lower surfaces has both X and Z components. From equations 6·122 it is seen that the component H_x varies cosinusoidally in the Z direction, and the component H_z varies sinusoidally in the Z direction. Both are constant in the Y direction. Hence the H vector at either upper or lower surface is given by $H_2 = \sqrt{|H_{x1}|^2 + |H_{z1}|^2}$, where H_{x1} and H_{z1} have the values given in equations 6·122. If a component of electric intensity E_2 exists at the upper or lower surface of the metal, so that energy will flow into the metal, the value of E_2 will be given by $E_2 = \eta H_2$. At the surface of the upper plate the average power flowing through a square meter of surface, as given by equation 3·83, is

$$p = \tfrac{1}{2}\Re(\mathbf{E} \times \mathbf{H}^*) \qquad [6{\cdot}123]$$

Hence

$$p_y = \tfrac{1}{2}\Re(E_2 H_2^*) = \tfrac{1}{2}\Re[\eta (H_2')^2] \qquad [6{\cdot}124]\dagger$$

Since we will now discuss waves in the metal as well as in the dielectric, we may write H_{x1}, H_{z1}, etc., for the waves in the dielectric, and H_{x2}, H_{z2}, etc., for the waves in the metal.

† The vector product $\mathbf{A} \times \mathbf{B} = AB \sin \phi$, where ϕ is the angle between A and B. If, as here, the angle $\phi = 90°$, $\mathbf{A} \times \mathbf{B} = AB$.

From equations 6·122

$$(H_2')^2 = A^2 \cos^2\left(\frac{\pi}{z_0} z\right) e^{-2\alpha x} + A^2 \frac{\beta_{01}^2 z_0^2}{\pi^2} \sin^2\left(\frac{\pi}{z_0} z\right) e^{-2\alpha x} \quad [6\cdot125]^*$$

Hence

$$p_y = \tfrac{1}{2}\mathcal{R}\left\{\eta A^2 \left[\cos^2\left(\frac{\pi}{z_0} z\right) + \frac{\beta_{01}^2 z_0^2}{\pi^2} \sin^2\left(\frac{\pi}{z_0} z\right)\right] e^{-2\alpha x}\right\} \quad [6\cdot126]$$

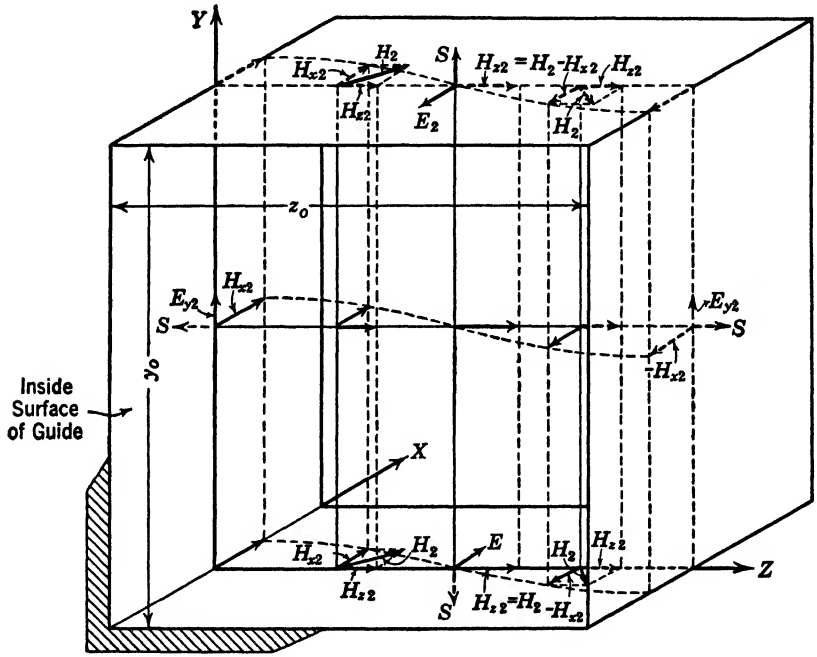


FIG. 6·18 Instantaneous distribution of magnetic intensity vectors of an H_{01} wave within a section of a rectangular wave guide. The assumed tangential component of E at the imperfectly conducting surface is indicated in several places.

From equation 5·144

$$\eta = \sqrt{\frac{j\omega\mu_2}{\sigma_2}} = \sqrt{\frac{\pi f \mu_2}{\sigma_2}} + j\sqrt{\frac{\pi f \mu_2}{\sigma_2}} \quad (\text{since } \sqrt{2j} = 1 + j) \quad [6\cdot127]$$

Taking only the real part of equation 6·126, we have

$$p_y = \frac{1}{2}\sqrt{\frac{\pi f \mu_2}{\sigma_2}} A^2 \left\{\cos^2\left(\frac{\pi}{z_0} z\right) + \frac{\beta_{01}^2 z_0^2}{\pi^2} \sin^2\left(\frac{\pi}{z_0} z\right)\right\} e^{-2\alpha x} \quad [6\cdot128]$$

* H_2', the maximum value of H_2 in distance and time, is given by $H_2' = \sqrt{(H_{s1}')^2 + (H_{s1}'')^2}$.

Hence the average power flowing into an area of surface of the upper side of the tube z_0 meters wide in the Z direction and 1 meter long in the X direction is

$$\frac{1}{2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,A^2e^{-2\alpha x}\int_{z=0}^{z=z_0}\left\{\cos^2\left(\frac{\pi}{z_0}z\right)+\frac{\beta_{01}^2z_0^2}{\pi^2}\sin^2\left(\frac{\pi}{z_0}z\right)\right\}dz \qquad [6\cdot129]$$

or

$$\frac{1}{4}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,A^2e^{-2\alpha x}z_0\left\{1+\frac{\beta_{01}^2z_0^2}{\pi^2}\right\} \qquad [6\cdot130]$$

Since H is constant in the Y direction, the power flowing into an equal area of the lower plate is also equal to 6·130. Hence P_y, the average power lost in the upper and lower walls of the tube per unit length of the tube, is

$$P_y=\frac{1}{2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,A^2e^{-2\alpha x}z_0\left\{1+\frac{\beta_{01}^2z_0^2}{\pi^2}\right\} \qquad [6\cdot131]$$

Let us now consider the power lost in the side walls of the tube. The amplitude of the tangential component of H at the surface of the right wall is, from equation 6·122, when $z = z_0$,

$$H'_{x2}=H'_{x1}=-Ae^{-\alpha x} \qquad [6\cdot132]$$

The amplitude of the required tangential E vector (Fig. 6·18) is

$$E'_{y2}=-\eta H'_{x2} \qquad [6\cdot133]$$

The power flowing into a unit area of the right side wall of the tube is, therefore,

$$p_z=\tfrac{1}{2}\mathcal{R}(-E_{y2}H^{\star}_{x2})=\tfrac{1}{2}\mathcal{R}\{\eta(H'_{x2})^2\} \qquad [6\cdot134]$$

$$=\frac{1}{2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,A^2e^{-2\alpha x} \qquad [6\cdot135]$$

The power flowing into an area y_0 meters in the Y direction and 1 meter in the X direction is y_0p_z. The power flowing into an equal area in the left wall of the tube is also y_0p_z. Hence the average power P_z lost in the two side walls of the tube per unit length is

$$P_z=2p_zy_0=\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,A^2y_0e^{-2\alpha x} \qquad [6\cdot136]$$

The average power flowing through unit area in the X direction (down the pipe) is

$$p_x=\tfrac{1}{2}(E'_{y1}H'_{z1}) \qquad [6\cdot137]$$

Substituting for $E'_{y1}H'_{z1}$ from equation 6·122, we have

$$p_x = \tfrac{1}{2}A^2\frac{\beta_{01}\omega\mu_1 z_0^2}{\pi^2}\sin^2\left(\frac{\pi}{z_0}z\right)e^{-2\alpha x} \qquad [6\cdot138]$$

Hence the average power P_x flowing through the tube in the X direction is

$$P_x = \tfrac{1}{2}A^2\frac{\beta_{01}\omega\mu_1 z_0^2}{\pi^2}e^{-2\alpha x}\int_{y=0}^{y_0}\int_{z=0}^{z_0}\sin^2\left(\frac{\pi}{z_0}z\right)dy\,dz \qquad [6\cdot139]$$

Giving

$$P_x = \left(A^2\frac{\beta_{01}z_0^2\omega\mu_1}{4\pi^2}y_0z_0\,e^{-2\alpha x}\right) \qquad [6\cdot140]$$

P_x may also be expressed as a function of (f/f_0) by appropriate algebraic manipulation.* Thus

$$P_x = \frac{y_0z_0A^2}{4}\left(\frac{f}{(f_0)_{01}}\right)\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2-1}\sqrt{\frac{\mu_1}{\varepsilon_1}}\,e^{-2\alpha x} \qquad [6\cdot140a]$$

Now the power loss in the metal boundary is equal to the rate of diminution of the power flowing through the dielectric. Since P_x is the power flowing through an area y_0z_0 normal to the Poynting vector, $-\partial P_x/\partial x$ is the rate of decrease of this power. See Fig. 6·19. This must equal the power lost by the wave in the four bounding surfaces. Thus

$$-\frac{\partial P_x}{\partial x} = P_y + P_z \qquad [6\cdot141]$$

But as shown in equation 5·158

$$\frac{\partial P_x}{\partial x} = -2\alpha P_x \qquad [6\cdot142]$$

* $\dfrac{z_0^2\omega\mu_1}{4\pi^2} = \dfrac{z_0^2 2\pi f\mu_1\varepsilon_1}{4\pi^2\varepsilon_1} = \dfrac{2z_0}{v_1}\left(\dfrac{z_0}{4\pi\varepsilon_1v_1}\right)f = \left(\dfrac{f}{(f_0)_{01}}\right)\left(\dfrac{z_0}{4\pi\varepsilon_1v_1}\right)$

Now

$$\beta_{01} = \omega\sqrt{\mu_1\varepsilon_1}\sqrt{1-\left(\frac{(f_0)_{01}}{f}\right)^2} = \omega\sqrt{\mu_1\varepsilon_1}\left(\frac{(f_0)_{01}}{f}\right)\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2-1}$$

Hence

$$\frac{z_0^2\omega\mu_1}{4\pi^2}\beta_{01} = \frac{z_0\omega\sqrt{\mu_1\varepsilon_1}}{4\pi\varepsilon_1v_1}\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2-1}$$

But

$$\frac{z_0\omega\sqrt{\mu_1\varepsilon_1}}{4\pi\varepsilon_1v_1} = \frac{z_0 2\pi f}{v_1 4\pi\varepsilon_1v_1} = \left(\frac{f}{(f_0)_{01}}\right)\frac{1}{4\varepsilon_1v_1} = \frac{f}{4(f_0)_{01}}\sqrt{\frac{\mu_1}{\varepsilon_1}}$$

giving equation 6·140*a*.

Hence

$$2\alpha P_x = P_y + P_z \qquad \text{or} \qquad \alpha = \frac{P_y + P_z}{2P_x} \qquad [6\cdot143]$$

Substituting for P_y, P_z, and P_x from equations 6·131, 6·136, and 6·140, we have

$$\alpha = \frac{\frac{1}{2}A^2 z_0 \sqrt{\frac{\pi f \mu_2}{\sigma_2}} \left\{1 + \frac{\beta_{01}^2 z_0^2}{\pi^2}\right\} e^{-2\alpha x} + A^2 y_0 \sqrt{\frac{\pi f \mu_2}{\sigma_2}}\, e^{-2\alpha x}}{\frac{1}{2}A^2 y_0 z_0 \frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2} e^{-2\alpha x}} \qquad [6\cdot144]$$

$$\alpha = \frac{1}{z_0}\sqrt{\frac{\pi f \mu_2}{\sigma_2}} \left\{ \frac{z_0}{y_0} \left(\frac{1 + \frac{\beta_{01}^2 z_0^2}{\pi^2}}{\frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2}} \right) + \left(\frac{2}{\frac{\beta_{01} z_0^2 \omega \mu_1}{\pi^2}} \right) \right\} \text{ nepers per meter} \qquad [6\cdot145]$$

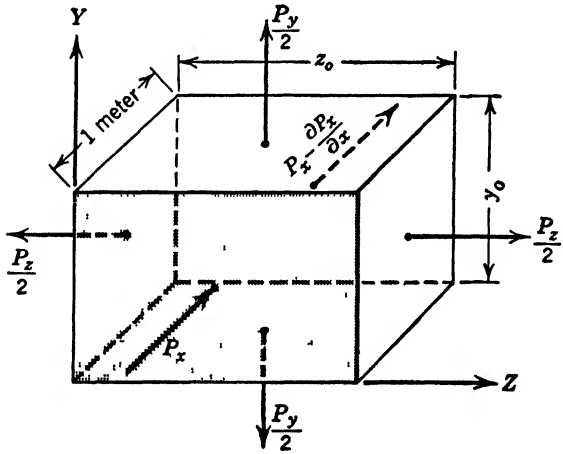


FIG. 6·19 Distribution of P_x, P_y, and P_z in the imperfectly conducting wave guide.

The attenuation constant may also be expressed as a function of $(f/(f_0)_{01})$ after certain algebraic manipulation. In this form it is

$$\alpha = \sqrt{\frac{2\pi\mu_2}{v_1 \sigma_2 \mu_1^2}}\, z_0^{-3/2} \left\{ \frac{\frac{z_0}{2y_0}\left(\frac{f}{(f_0)_{01}}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2 - 1}} + \frac{\left(\frac{f}{(f_0)_{01}}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2 - 1}} \right\} \text{ nepers per meter} \qquad [6\cdot146]$$

The attenuation constant as given in equation 6·146 is the sum of two parts, one arising from losses on the top and bottom walls of the tube, and the other from losses on the side walls of the tube. Let us denote the product of the first term within the brackets and the common factor by α', and the product of the second term and the common factor by α'', so that $\alpha = \alpha' + \alpha''$. Both α' and α'' are infinite at the critical frequency. For frequencies just above the critical frequency, both factors decrease rapidly. To determine whether or not these factors have a minimum, we may differentiate them with respect to $(f/(f_0)_{01})$ and set the derivative equal to zero. The solution of the resulting equation will give the value of $(f/(f_0)_{01})$ at the minimum, if a minimum exists. Thus

$$\frac{\partial \alpha'}{\partial\left(\frac{f}{(f_0)_{01}}\right)} = \left(\frac{f}{(f_0)_{01}}\right)^2 - 3 = 0 \qquad [6\cdot147]^*$$

$$\frac{\partial \alpha''}{\partial\left(\frac{f}{(f_0)_{01}}\right)} = 0 \qquad [6\cdot148]\dagger$$

For α' we see from equation 6·147 that the minimum exists at a value of $f = \sqrt{3}\ (f_0)_{01}$. For higher values of frequency, α' increases. At large values of frequency where $(f/(f_0)_{01})^2 \gg 1$, $\alpha' \simeq$ constant $(f/(f_0)_{01})^{\frac{1}{2}}$, and hence α' increases as the square root of the frequency. Since α'

* If we let $x = f/(f_0)_{01}$ in equation 6·146 we have

$$\alpha' = K_1\left(\frac{x^{\frac{3}{2}}}{\sqrt{x^2-1}}\right) = K_1[x^{-1} - x^{-3}]^{-\frac{1}{2}}$$

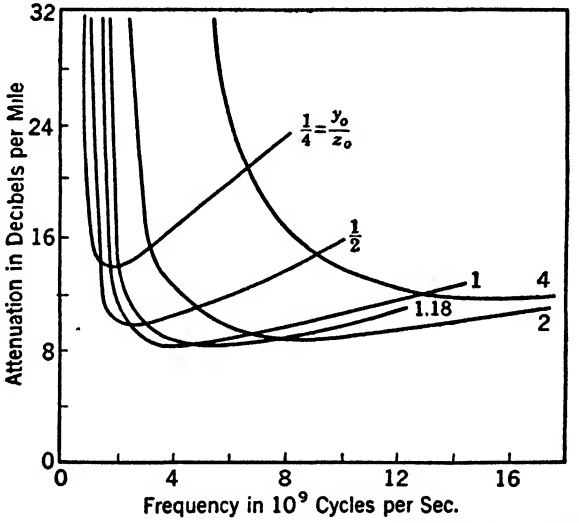
where K_1 is a constant dependent on the physical properties of the system. For a minimum $\frac{\partial \alpha'}{\partial x} = 0 = K_1\{-\frac{1}{2}(x^{-1} - x^{-3})^{-\frac{3}{2}}(-x^{-2} + 3x^{-4})\}$. Hence $x = \sqrt{1}$ or $x = \sqrt{3}$, giving $f/(f_0)_{01} = \pm 1$ or $f/(f_0)_{01} = \pm\sqrt{3}$. For all values of f less than $(f_0)_{01}$, the tube cannot support a wave. Also, at $f = (f_0)_{01}$, $\alpha = \infty$. Hence we are only concerned with values of $f/(f_0)_{01} > 1$. In this case a minimum occurs at $f = +\sqrt{3}\ (f_0)_{01}$.

† When $x = \frac{f}{(f_0)_{01}}$, $\alpha'' = K_2\left(\frac{x^{-\frac{1}{2}}}{\sqrt{x^2-1}}\right) = K_2[x^3 - x]^{-\frac{1}{2}}$ where K_2 is a constant dependent on the physical properties of the system. For a minimum

$$\frac{\partial \alpha''}{\partial x} = 0 = K_2[-\tfrac{1}{2}(x^3 - x)^{-\frac{3}{2}}(3x^2 - 1)]$$

Hence $x = \sqrt{1}$ or $x = \sqrt{\frac{1}{3}}$, giving $f/(f_0)_{01} = \pm 1$ or $f/(f_0)_{01} = \pm\sqrt{\frac{1}{3}}$. Since we are not concerned for values of $f/(f_0)_{01} \leq 1$, there is no minimum.

is that part of the attenuation constant due to the upper and lower walls of the tube, these surfaces are responsible for the increased attenuation as the frequency increases. The term α'' has no minimum for $f/(f_0)_{01} > 1$ but continues to decrease as the frequency is increased. For large values of f, α'' is proportional to $(f/(f_0)_{01})^{-3/2}$. For frequencies exceeding $\sqrt{3}\,(f_0)_{01}$, the most important contribution to the attenuation arises from the α' term. Hence we see that for frequencies greater than $\sqrt{3}\,(f_0)_{01}$ most of the loss takes place at the top and bottom walls of the tube. This was first pointed out by Chu and Barrow. When operating a rectangular pipe under these circumstances they suggested the use of a material of highest conductivity for the top and bottom walls.



(Chu and Barrow, courtesy of IRE)

FIG. 6·20 Attenuation-frequency characteristics for an H_{01} wave in a rectangular copper wave guide filled with air. The guide periphery is constant and equal to 0.4 meter. Different ratios of y_0/z_0 are shown.

A comparison of rectangular pipes for various ratios of the y_0/z_0 dimensions is shown in Fig. 6·20. The data for these curves were obtained by Chu and Barrow for air-filled copper pipes having the same periphery. In this diagram $y_0 + z_0 = 40$ cm. In this manner they found that there is an optimum ratio of $y_0/z_0 = 1.18$; that is, for a fixed perimeter and lowest attenuation the tube should be constructed so that $y_0 = 1.18\,z_0$. However, it may be noted from Fig. 6·20 that, for any ratio of y_0/z_0 between $\frac{1}{2}$ and 2, the attenuation is very close to the

minimum value. There is another factor which may be important in the choice of the y_0/z_0 ratio. It may be seen from Fig. 6·20 that, as y_0/z_0 is reduced, the required frequency is also reduced. It will be shown in a later chapter that, in the higher-frequency ranges, the efficiency of an oscillator is usually decreased as the frequency is increased. Hence it may be desirable to use values of $y_0/z_0 = \frac{1}{2}$ or less owing to

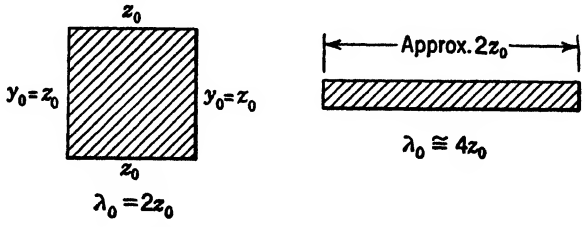
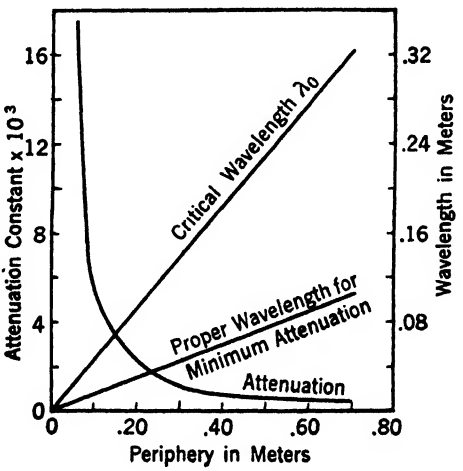


FIG. 6·21 Cross sections of two wave guides having the same periphery.

the increased efficiency of the oscillator at the lower frequencies. For a square pipe having a periphery of 46 cm, the critical wavelength is 23 cm. If this pipe is flattened as indicated in Fig. 6·21 its critical wavelength will approach 46 cm.



(*Chu and Barrow, courtesy of IRE*)

FIG. 6·22 Minimum attenuation and wavelength as a function of periphery for an H_{01} wave in a rectangular copper wave guide of optimum y_0/z_0 ratio = 1.18.

A curve showing the variation of attenuation as a function of periphery for optimum y_0/z_0 ratio is shown in Fig. 6·22. We may obtain a relation between the critical wavelength $(\lambda_0)_{01}$ and the periphery p, for optimum y_0/z_0 ratio. From equations 6·76, $(\lambda_0)_{01} = 2z_0$. Since

$y_0/z_0 = 1.18$, and $2(y_0 + z_0) = p$, we may write $(\lambda_0)_{01} = 4.36p$. This relation is also plotted in Fig. 6·22. It may be seen from Fig. 6·20 that the attenuation offered by the rectangular tubes is very nearly constant over an appreciable frequency range in the neighborhood of the frequency corresponding to minimum attenuation.

Higher-order attenuation

The attenuation constant has been calculated for E_{nm} and H_{nm} waves by Chu and Barrow. These calculations will not be given here, but the results may be expressed as:

$H_{0,m}$ wave:

$$\alpha = Kz_0^{-3/2}\sqrt{m}\left[\frac{\frac{z_0}{2y_0}\left(\frac{f}{(f_0)_{0m}}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_{0m}}\right)^2 - 1}} + \frac{\left(\frac{f}{(f_0)_{0m}}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_{0m}}\right)^2 - 1}}\right] \quad \text{nepers per meter}$$

$H_{n,m}$ wave: $n \neq 0$

$$\alpha = Kz_0^{-3/2}\sqrt{m} \times \frac{\frac{z_0}{y_0}\left\{1 + \left(\frac{n}{m}\right)^2\frac{z_0}{y_0}\right\}\left(\frac{f}{(f_0)_{nm}}\right)^{3/2} + \left\{1 + \left(\frac{n}{m}\right)^2\left(\frac{z_0}{y_0}\right)^3\right\}\left(\frac{f}{(f_0)_{nm}}\right)^{-1/2}}{\left[1 + \left(\frac{z_0}{y_0}\right)^2\left(\frac{n}{m}\right)^2\right]^{3/4}\sqrt{\left(\frac{f}{(f_0)_{nm}}\right)^2 - 1}} \quad \text{nepers per meter}$$

$E_{n,m}$ wave:

$$\alpha = Kz_0^{-3/2}\sqrt{m}\frac{\left[1 + \left(\frac{n}{m}\right)^2\left(\frac{z_0}{y_0}\right)^3\right]\left(\frac{f}{(f_0)_{nm}}\right)^{3/2}}{\left[1 + \left(\frac{z_0}{y_0}\right)^2\left(\frac{n}{m}\right)^2\right]^{3/4}\sqrt{\left(\frac{f}{(f_0)_{nm}}\right)^2 - 1}} \quad \text{nepers per meter}$$

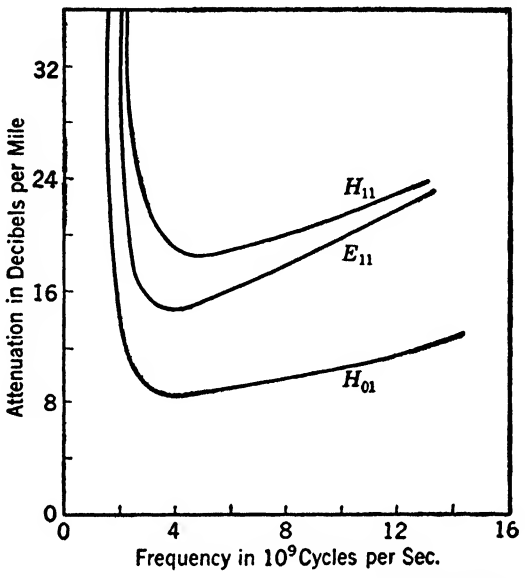
where

$$K = \sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}$$

Since the attenuation constant α for the H_{nm} wave does not reduce to the correct value when $n = 0$, it is necessary to give a separate expression for waves of the H_{0m} type.

Figure 6·23 shows the attenuation-frequency curves for H_{11}, E_{11}, and H_{01} waves in a square air-filled copper pipe. It is found that, for equal orders of harmonic variations, $n = m$, a square pipe gives the minimum attenuation. The H_{01} wave has the lowest attenuation of all

E or H waves in a square pipe. Table 6·2 gives the critical and optimum conditions for the waves shown in Fig. 6·23.



(Chu and Barrow, courtesy of IRE)

FIG. 6·23 Comparison of the E_{11}, H_{01}, and H_{11} wave attenuation characteristics for a square air-filled copper pipe. $y_0 = z_0 = 0.10$ meter.

TABLE 6·2

CRITICAL AND OPTIMUM CONDITIONS FOR H_{01}, H_{11}, E_{11} WAVES IN A SQUARE AIR FILLED COPPER PIPE 10 CM ON A SIDE

Wave type	H_{01}	H_{11}	E_{11}
Critical frequency in cycles	1.50×10^9	2.12×10^9	2.12×10^9
Frequency for minimum attenuation in cycles	4.44×10^9	5.18×10^9	3.67×10^9
Minimum attenuation in decibels per mile	8.55	18.1	14.6

Chu and Barrow, courtesy of IRE

PROBLEMS

6·1 Obtain equation 6·81 from equations 6·79 and 6·80 by an elimination process similar to that employed in section 6·3.

6·2 Obtain several solutions to equation 6·81, and show that the one involving the product of the sines will meet the boundary conditions $E_y = 0$ at $z = 0$, z_0, and

$E_z = 0$ at $y = 0, y_0$ for the E waves in the rectangular guide. Will other solutions meet these boundary conditions?

6·3 Show that the components of the E_{nm} waves are as given in equations 6·82.

6·4 Show that the propagation properties of the H_{nm} waves as given in equations 6·71 are the same as those for the E_{nm} waves.

6·5 Show by substitution of $\sin\theta$ and $\cos\theta$ from Fig. 6·15 that equation 6·109 reduces to $(\omega/v_1)x^{(1)}$.

6·6 Show that equation 6·106 for the elementary component $E_y^{(2)}$ may be expressed in terms of distance in the $X^{(2)}$ direction as given in equation 6·112.

6·7 Show that the H_x and H_z components of the H_{01} wave in the dielectric, as given in equation 6·103, may be expressed in terms of distance in the $X^{(1)}$ direction as given in equation 6·113.

6·8 Show that resolution of the H_x and H_z components of the first group of elementary waves given in equations 6·113, along the $X^{(1)}$ direction results in a zero longitudinal component of magnetic intensity. Repeat for the second group along the $X^{(2)}$ direction.

6·9 Show that the intrinsic impedance η for the H_{01} wave in a rectangular pipe, as given in equation 6·102, will approach 377Ω if the dimensions of the pipe are made very large in comparison to the wavelength of the wave and if the dielectric in the pipe is air.

6·10 A general solution of equation 6·16 is $H_x = YZ$, where Y and Z are given in equations 6·28 and 6·31. Thus $H_x' = C_1C_3 \sin\sqrt{A_1}y \sin\sqrt{A_2}z + C_2C_4 \cos\sqrt{A_1}y \cos\sqrt{A_2}z + C_1C_4 \sin\sqrt{A_1}y \cos\sqrt{A_2}z + C_2C_3 \cos\sqrt{A_1}y \sin\sqrt{A_2}z$. Show that the solution $H_x' = A \cos\sqrt{A_1}y \sin\sqrt{A_2}z$ (where, in this case, $A = C_2C_3$) will satisfy the boundary conditions for the H_{nm} waves when the X axis is chosen down the center of the pipe. See Fig. 6·14. Will any other particular solution chosen from among these satisfy the boundary conditions for this choice of axes?

6·11 Obtain the components for the H_{nm} waves based upon the choice of axes shown in Fig. 6·14.

6·12 Show how equation 6·145 reduces to 6·146 by substitution of the relations $\beta_{01} = \dfrac{\omega}{v_1}\sqrt{1 - \left(\dfrac{(f_0)_{01}}{f}\right)^2}$; $v_1 = \dfrac{1}{\sqrt{\mu_1\varepsilon_1}}$; and any other suitable relations from equations 6·76.

6·13 A wave guide of metal is 10 cm square inside. Tabulate all the modes which may be propagated at 2000; 3000; and 5000 megacycles per second.

6·14 Calculate the phase velocities of the several modes tabulated in problem 6·13.

6·15 If the guide of problem 6·13 is of copper and at room temperature, calculate the attenuation at 4000 mc/sec for the H_{01} and E_{11} modes of transmission.

6·16 Repeat problem 6·15 for a guide of aluminum.

6·17 Repeat problem 6·15 for a guide of soft iron for which $\mu_m = 1000$.

6·18 Plot the attenuation vs. frequency characteristic from $(f_0)_{01}$ up to a frequency of 50,000 mc for an H_{01} wave propagated in a copper wave guide 20 by 24 cm. The wave is oriented so as to produce the lowest possible loss.

6·19 A rectangular wave guide must be divided into two separate guides by means of a sort of Y fitting. Considering only *TE* modes of low order, discuss the mechanical construction which should be employed and the nature of the wave behavior at the junction. Need a serious reflection result? How should the wave be polarized?

6·20 A wave guide 3000 meters long must have an attenuation of only 0.6 neper. Assuming that the interior is silver-plated for highest possible conductivity, choose the mode of propagation and the dimensions of the guide to meet this requirement most compactly.

CHAPTER 7

CYLINDRICAL WAVE GUIDES

In treating the problems of the previous chapters, it was found expedient to use the cartesian coordinate system. This was convenient since the equations of the boundaries of the parallel planes and rectangular tubes are most easily expressed in this system.

We shall now take up problems in which there is axial symmetry, as in tubes of circular cross-section and concentric circular tubes. Under these circumstances it is most convenient to use the cylindrical coordinate system since the cylindrical boundaries are most easily expressed in this system by the simple equation r = constant.

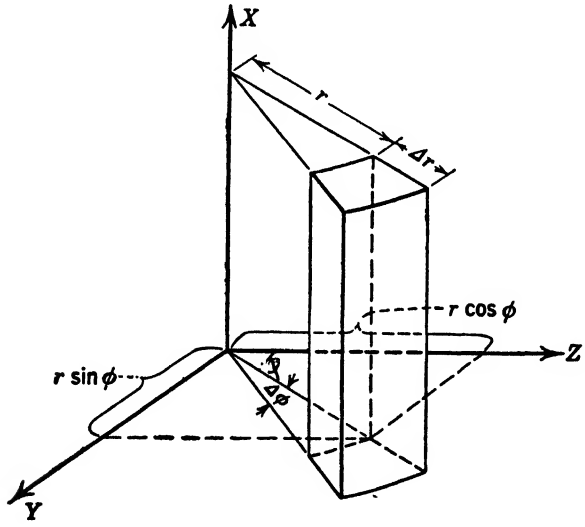


FIG. 7·1*a*

7·1 Maxwell's Equations in Cylindrical Coordinates

The choice of axes in this system as related to the axes of the cartesian coordinate system which we have used is shown in Fig. 7·1*a*. The circular cylindrical coordinates are r, ϕ, and x. These are related to the rectangular coordinates by the equations

$$x = x, \qquad y = r \sin \phi, \qquad z = r \cos \phi \qquad [7\cdot1]$$

The relation between the vectors in the two systems may be obtained in the following way. Consider a vector E, Fig. 7·1*b*, lying in the YZ plane, making an angle θ with the radius vector r through 0, and an angle $\theta + \phi$ with the Z axis. From this diagram we may write

$$E_y = E \sin(\theta + \phi) = E(\cos\theta \sin\phi + \sin\theta \cos\phi)$$

$$E_z = E \cos(\theta + \phi) = E(\cos\theta \cos\phi - \sin\theta \sin\phi)$$

Also

$$E_r = E \cos\theta \qquad \text{and} \qquad E_\phi = E \sin\theta$$

Hence

$$\begin{aligned} E_y &= E_r \sin\phi + E_\phi \cos\phi \\ E_z &= E_r \cos\phi - E_\phi \sin\phi \\ E_x &= E_x \end{aligned} \qquad [7\cdot2]$$

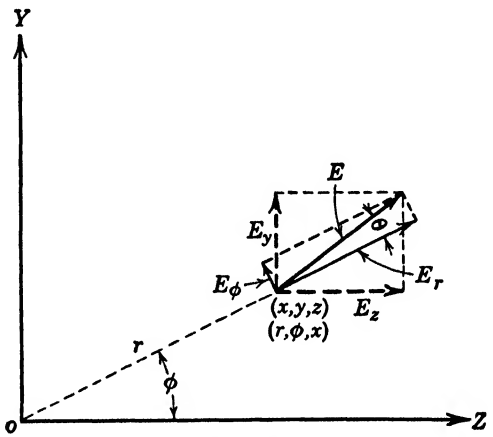


FIG. 7·1*b*

Equations 7·2 may be solved to give E_r, E_ϕ, and E_x in terms of E_x, E_y, and E_z. They yield

$$\begin{aligned} E_r &= E_y \sin\phi + E_z \cos\phi \\ E_\phi &= E_y \cos\phi - E_z \sin\phi \\ E_x &= E_x \end{aligned} \qquad [7\cdot3]$$

With the aid of relations 7·2 and 7·3, the components of a vector in cylindrical coordinates may be expressed in terms of the rectangular components, and vice versa.

Maxwell's equations, in cylindrical coordinates, may be obtained with the aid of the integral relations 3·1, 3·2, 3·3, and 3·4 by taking the appropriate line and surface integrals over a small element in cylindrical

coordinates. This may be done in a manner similar to that employed in the rectangular coordinate case of Chapter 2. Consider, for example, the volume element shown in Fig. 7·2, which is greatly enlarged for convenience. Let us imagine this volume element to be located in a homogeneous isotropic medium through which a current is flowing. We may resolve the current density ι into three components. Let these components at the point (r, ϕ, x) be ι_r, ι_ϕ, and ι_x, in the directions of increasing r, ϕ, and x. The magnetic field H in the medium at this

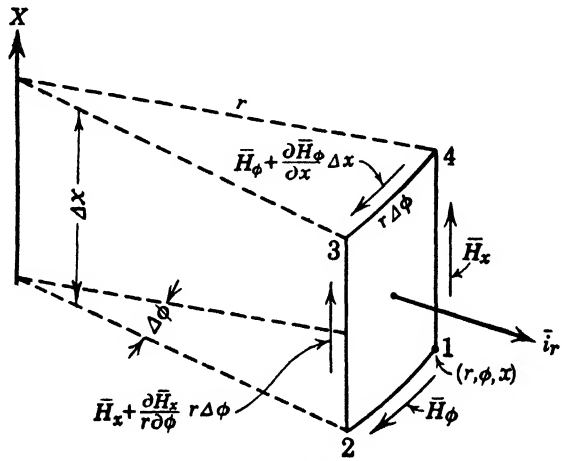


FIG. 7·2

point may similarly be resolved into the components H_r, H_ϕ, and H_x. Let $\overline{H}_\phi$ be the average value of the magnetic intensity along the side 1, 2, of the volume element. Then, assuming proper continuity, and neglecting second-order effects, the average value of the magnetic intensity along the side 4, 3 is $\overline{H}_\phi + \dfrac{\partial \overline{H}_\phi}{\partial x}\Delta x$. Similarly, if $\overline{H}_x$ is the average magnetic intensity along the side 1, 4, that along the side 2, 3 is $\overline{H}_x + \dfrac{\partial \overline{H}_x}{r\partial\phi} r\Delta\phi$. Hence, the line integral of H around the path 1, 2, 3, 4, is

$$\oint_l H \cos\theta \, dl = \overline{H}_\phi r\,\Delta\phi + \left\{\overline{H}_x + \frac{\partial \overline{H}_x}{r\partial\phi} r\,\Delta\phi\right\}\Delta x - \left\{\overline{H}_\phi + \frac{\partial \overline{H}_\phi}{\partial x}\Delta x\right\} r\,\Delta\phi$$

$$- \overline{H}_x\,\Delta x = \left\{\frac{\partial \overline{H}_x}{r\partial\phi} - \frac{\partial \overline{H}_\phi}{\partial x}\right\} r\,\Delta\phi\,\Delta x \qquad [7\cdot4]$$

But, by Ampère's law, this is equal to the surface integral of ι over the surface enclosed by the path l. If $\bar{\iota}_r$ is the average current density

over the surface $r\,\Delta\phi\,\Delta x$, normal to the surface, then

$$\int_S\int \iota\,ds = \bar{\iota}_r r\,\Delta\phi\,\Delta x \qquad [7\cdot5]$$

and the other components ι_x and ι_ϕ contribute nothing to the surface integral. Hence

$$\oint_l H\cos\theta\,dl = \bar{\iota}_r r\,\Delta\phi\,\Delta x = \left\{\frac{1}{r}\frac{\partial \overline{H}_x}{\partial\phi} - \frac{\partial \overline{H}_\phi}{\partial x}\right\} r\,\Delta\phi\,\Delta x \qquad [7\cdot6]$$

Dividing through by $r\,\Delta\phi\,\Delta x$ and taking the limit as $\Delta\phi$ and Δx approach zero, we obtain

$$\underset{\Delta\phi\,\Delta x\to 0}{\text{Limit}}\left\{\frac{\oint_l H\cos\theta\,dl}{r\,\Delta\phi\,\Delta x}\right\} = \text{curl}_r H = \frac{1}{r}\frac{\partial H_x}{\partial\phi} - \frac{\partial H_\phi}{\partial x} = \iota_r \qquad [7\cdot7]$$

where the average values of the magnetic intensity and current approach their values at the point x, r, ϕ in the limit.

By constructing similar mathematical areas $\Delta r\,\Delta x$ and $r\,\Delta\phi\,\Delta r$, and equating the surface integral of ι through these areas to the line integral of H around the areas, we obtain in the limit

$$\text{curl}_\phi H = \frac{\partial H_r}{\partial x} - \frac{\partial H_x}{\partial r} = \iota_\phi \qquad [7\cdot8]$$

$$\text{curl}_x H = \frac{1}{r}\left\{\frac{\partial (rH_\phi)}{\partial r} - \frac{\partial H_r}{\partial\phi}\right\} = \iota_x \qquad [7\cdot9]$$

Since

$$\iota = \mathbf{r}\iota_r + \boldsymbol{\phi}\iota_\phi + \mathbf{x}\iota_x \qquad [7\cdot10]$$

where $\mathbf{r}$, $\boldsymbol{\phi}$, $\mathbf{x}$, are unit vectors in the directions of increasing r, ϕ, and x, we have

$$\iota = \mathbf{r}\left\{\frac{1}{r}\frac{\partial H_x}{\partial\phi} - \frac{\partial H_\phi}{\partial x}\right\} + \boldsymbol{\phi}\left\{\frac{\partial H_r}{\partial x} - \frac{\partial H_x}{\partial r}\right\} + \mathbf{x}\frac{1}{r}\left\{\frac{\partial (rH_\phi)}{\partial r} - \frac{\partial H_r}{\partial\phi}\right\} \qquad [7\cdot11]$$

If, in addition to the conduction current i, there is also a displacement current $\partial D/\partial t$, we may write

$$\left.\begin{aligned} \frac{1}{r}\frac{\partial H_x}{\partial\phi} - \frac{\partial H_\phi}{\partial x} &= \iota_r + \frac{\partial D_r}{\partial t} = \sigma E_r + \varepsilon\frac{\partial E_r}{\partial t} \\ \frac{\partial H_r}{\partial x} - \frac{\partial H_x}{\partial r} &= \iota_\phi + \frac{\partial D_\phi}{\partial t} = \sigma E_\phi + \varepsilon\frac{\partial E_\phi}{\partial t} \\ \frac{1}{r}\frac{\partial (rH_\phi)}{\partial r} - \frac{1}{r}\frac{\partial H_r}{\partial\phi} &= \iota_x + \frac{\partial D_x}{\partial t} = \sigma E_x + \varepsilon\frac{\partial E_x}{\partial t} \end{aligned}\right\} \qquad [7\cdot12]$$

which expresses one of Maxwell's equations in cylindrical coordinates.

By taking the line integral of the electric intensity around a similar surface in cylindrical coordinates, and equating it to the surface integral of $\partial B_n/\partial t$ through the surface, we obtain in the limit, as the surface approaches zero, the relations

$$\left.\begin{aligned} \frac{1}{r}\frac{\partial E_x}{\partial \phi} - \frac{\partial E_\phi}{\partial x} &= -\frac{\partial B_r}{\partial t} = -\mu\frac{\partial H_r}{\partial t} \\ \frac{\partial E_r}{\partial x} - \frac{\partial E_x}{\partial r} &= -\frac{\partial B_\phi}{\partial t} = -\mu\frac{\partial H_\phi}{\partial t} \\ \frac{1}{r}\frac{\partial (rE_\phi)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \phi} &= -\frac{\partial B_x}{\partial t} = -\mu\frac{\partial H_x}{\partial t} \end{aligned}\right\} \qquad [7\cdot13]$$

giving another of Maxwell's equations in cylindrical coordinates. The divergence equations in these coordinates are

$$\frac{1}{r}\frac{\partial}{\partial r}(rD_r) + \frac{1}{r}\frac{\partial D_\phi}{\partial \phi} + \frac{\partial D_x}{\partial x} = \rho \text{ or } 0 \qquad [7\cdot14]$$

and

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_r) + \frac{1}{r}\frac{\partial B_\phi}{\partial \phi} + \frac{\partial B_x}{\partial x} = 0 \qquad [7\cdot15]$$

They also may be obtained by taking the required volume integrals over an infinitesimal element in the cylindrical coordinate system.

7·2 Waves Guided by Cylindrical Metallic Tubes

As in the previous cases we are interested in the propagation of waves which vary periodically with distance and time in the positive X direction. We therefore introduce the requirement that electric and magnetic intensities involve time only in the form $e^{j\omega t}$ and distance in the form $e^{-\gamma x}$ so that

$$\mathbf{E} = \mathbf{E}'e^{j\omega t - \gamma x} \qquad [7\cdot16]$$

and

$$\mathbf{H} = \mathbf{H}'e^{j\omega t - \gamma x} \qquad [7\cdot17]$$

In the final solution we shall take only the real part of 7·16 and 7·17. Thus, in the final result

$$\mathbf{E} = \mathcal{R}[\mathbf{E}'e^{j\omega t - \gamma x}] \quad \text{and} \quad \mathbf{H} = \mathcal{R}[\mathbf{H}'e^{j\omega t - \gamma x}]$$

The factor $e^{j\omega t - \gamma x}$ represents a wave traveling in the positive X direction with a frequency $f = \omega/2\pi$ and a propagation constant $\gamma = \alpha + j\beta$, α being the attenuation constant and β the phase constant. Substituting 7·16 and 7·17 into 7·12 and 7·13, and dividing through by

$e^{j\omega t-\gamma x}$ we obtain

$$\left.\begin{aligned}
&\frac{1}{r}\frac{\partial H'_x}{\partial \phi}+\gamma H'_\phi=(\sigma+j\omega\varepsilon)E'_r && (a)\\
&-\gamma H'_r-\frac{\partial H'_x}{\partial r}=(\sigma+j\omega\varepsilon)E'_\phi && (b)\\
&\frac{1}{r}\frac{\partial}{\partial r}(rH'_\phi)-\frac{1}{r}\frac{\partial H'_r}{\partial \phi}=(\sigma+j\omega\varepsilon)E'_x && (c)
\end{aligned}\right\}\ [7\cdot18]\qquad
\left.\begin{aligned}
&\frac{1}{r}\frac{\partial E'_x}{\partial \phi}+\gamma E'_\phi=-j\omega\mu H'_r && (a)\\
&-\gamma E'_r-\frac{\partial E'_x}{\partial r}=-j\omega\mu H'_\phi && (b)\\
&\frac{1}{r}\frac{\partial}{\partial r}(rE'_\phi)-\frac{1}{r}\frac{\partial E'_r}{\partial \phi}=-j\omega\mu H'_x && (c)
\end{aligned}\right\}\ [7\cdot19]$$

When applying these equations to problems involving the propagation of electromagnetic waves inside long conducting cylinders in the axial X direction it is convenient to write them in another form, in which the field is expressed in terms of E and H components along the direction of propagation, i.e., the axial components E_x and H_x. To obtain this form from equations 7·18 and 7·19 we may proceed as follows:

From equation 7·19b

$$H'_\phi=\frac{\gamma E'_r}{j\omega\mu}+\frac{1}{j\omega\mu}\frac{\partial E'_x}{\partial r}$$

Substituting this value of H'_ϕ into 7·18a we obtain

$$\frac{1}{r}\frac{\partial H'_x}{\partial \phi}+\frac{\gamma^2}{j\omega\mu}E'_r+\frac{\gamma}{j\omega\mu}\frac{\partial E'_x}{\partial r}=(\sigma+j\omega\varepsilon)E'_r$$

giving

$$E'_r=\frac{1}{j\omega\mu(\sigma+j\omega\varepsilon)-\gamma^2}\left\{\gamma\frac{\partial E'_x}{\partial r}+\frac{j\omega\mu}{r}\frac{\partial H'_x}{\partial \phi}\right\} \qquad [7\cdot20]$$

Similarly, substituting for H'_r from equation 7·19a into 7·18b we obtain

$$E'_\phi=\frac{1}{j\omega\mu(\sigma+j\omega\varepsilon)-\gamma^2}\left\{\frac{\gamma}{r}\frac{\partial E'_x}{\partial \phi}-j\omega\mu\frac{\partial H'_x}{\partial r}\right\} \qquad [7\cdot21]$$

Also substituting E'_ϕ from equation 7·18b into 7·19a gives

$$H'_r=\frac{1}{j\omega\mu(\sigma+j\omega\varepsilon)-\gamma^2}\left\{-\frac{\sigma+j\omega\varepsilon}{r}\frac{\partial E'_x}{\partial \phi}+\gamma\frac{\partial H'_x}{\partial r}\right\} \qquad [7\cdot22]$$

and E'_r from 7·18a, substituted into 7·19b, gives

$$H'_\phi=\frac{1}{j\omega\mu(\sigma+j\omega\varepsilon)-\gamma^2}\left\{(\sigma+j\omega\varepsilon)\frac{\partial E'_x}{\partial r}+\frac{\gamma}{r}\frac{\partial H'_x}{\partial \phi}\right\} \qquad [7\cdot23]$$

The equations 7·20, 7·21, 7·22, and 7·23 express E'_r, E'_ϕ, H'_r, and H'_ϕ as functions of the longitudinal or axial components E'_x and H'_x. By elimination among the equations 7·18 and 7·19 we may obtain an equation in E'_x alone or H'_x alone. These equations, together with equations

7·20–7·23, completely determine the field distribution. The elimination process to obtain the equations in E_x' or H_x' will not be carried through here, but it is similar to that employed in Chapters 5 and 6 to obtain similar relations in rectangular coordinates. The equation we seek is the wave equation; expressed in cylindrical coordinates it is

$$\frac{\partial^2 E_x'}{\partial r^2} + \frac{1}{r}\frac{\partial E_x'}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_x'}{\partial \phi^2} = [j\omega\mu(\sigma + j\omega\varepsilon) - \gamma^2]E_x' \qquad [7\cdot24]$$

Similarly,

$$\frac{\partial^2 H_x'}{\partial r^2} + \frac{1}{r}\frac{\partial H_x'}{\partial r} + \frac{1}{r^2}\frac{\partial^2 H_x'}{\partial \phi^2} = [j\omega\mu(\sigma + j\omega\varepsilon) - \gamma^2]H_x' \qquad [7\cdot25]$$

The solution of these equations may be obtained by the method of separation of variables discussed in section 6·4. The first step toward the solution of 7·25 is to set

$$H_x' = R\Phi \qquad [7\cdot26]$$

where R is a function of r alone, $R(r)$, and Φ is a function of ϕ alone, $\Phi(\phi)$. The required partial derivatives are

$$\frac{\partial H_x'}{\partial r} = \frac{\partial}{\partial r}(R\Phi) = \Phi\frac{\partial R}{\partial r}$$

$$\frac{\partial^2 H_x'}{\partial r^2} = \Phi\frac{\partial^2 R}{\partial r^2}$$

and

$$\frac{\partial^2 H_x'}{\partial \phi^2} = R\frac{\partial^2 \Phi}{\partial \phi^2}$$

Hence equation 7·25 becomes

$$\Phi\frac{\partial^2 R}{\partial r^2} + \frac{1}{r}\Phi\frac{\partial R}{\partial r} + \frac{1}{r^2}R\frac{\partial^2 \Phi}{\partial \phi^2} = [j\omega\mu(\sigma + j\omega\varepsilon) - \gamma^2]R\Phi \qquad [7\cdot27]$$

Dividing through by $R\Phi/r^2$ and rearranging

$$\frac{r^2}{R}\frac{\partial^2 R}{\partial r^2} + \frac{r}{R}\frac{\partial R}{\partial r} + r^2[\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)] + \frac{1}{\Phi}\frac{\partial^2 \Phi}{\partial \phi^2} = 0 \qquad [7\cdot28]$$

The first three terms of equation 7·28 are a function of r only, say $F(r)$; the fourth term is a function of ϕ only, say $F(\phi)$, so that

$$F(r) + F(\phi) = 0 \qquad [7\cdot29]$$

The equation therefore states that a function of r plus a function of ϕ is equal to zero, no matter what values r and ϕ may have. This is

surely not possible in general, for, if we keep r fixed and vary ϕ, $F(\phi)$ would change while $F(r)$ remained the same. Thus the equation would not be satisfied. The only exception would, of course, occur where $F(r)$ and $F(\phi)$ are each constant and independent of r and ϕ. Let us impose this condition and set

$$F(r) = p^2 \qquad \text{and} \qquad F(\phi) = -p_1^2$$

where p and p_1 are constants. Substituting for $F(r)$ and $F(\phi)$ in 7·29 we obtain $p^2 - p_1^2 = 0$, giving $p = p_1$. Thus, from the last term of equation 7·28,

$$\frac{\partial^2 \Phi}{\partial \phi^2} + p_1^2 \Phi = 0 \qquad [7\cdot30]$$

an ordinary differential equation in Φ. Its solution may be written as

$$\Phi = C_1 \sin p_1\phi + C_2 \cos p_1\phi$$

where C_1 and C_2 are arbitrary constants. Also, since $p = p_1$

$$\Phi = C_1 \sin p\phi + C_2 \cos p\phi \qquad [7\cdot31]$$

We obtain from the first three terms of equation 7·28

$$\frac{r^2}{R}\frac{\partial^2 R}{\partial r^2} + \frac{r}{R}\frac{\partial R}{\partial r} + r^2[\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)] - p^2 = 0 \qquad [7\cdot32]$$

Multiplying through by R/r^2 we obtain

$$\frac{\partial^2 R}{\partial r^2} + \frac{1}{r}\frac{\partial R}{\partial r} + \left[\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon) - \frac{p^2}{r^2}\right] R = 0 \qquad [7\cdot33]$$

This is a form of Bessel's differential equation; its general solution may be written

$$R = C_3 J_p(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)}) + C_4 Y_p(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)}) \qquad [7\cdot34]$$

where J_p and Y_p are Bessel functions of order p, of the first and second kind, and C_3 and C_4 are arbitrary constants.

The general solution of 7·25 may, from equation 7·26, be written as the product of 7·31 and 7·34.

Thus

$$H_z' = R\Phi = (C_1 \sin p\phi + C_2 \cos p\phi)[C_3 J_p(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)}) + C_4 Y_p(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)})] \qquad [7\cdot34a]$$

Before considering this solution further, we shall digress to discuss equation 7·33 briefly.

7·3 Bessel's Differential Equation and Bessel Functions*

In practice, Bessel's equation generally appears in the form 7·33. If we set

$$k^2 = \gamma^2 - j\omega\mu(\sigma + j\omega\epsilon) \qquad [7\cdot35]$$

it may be written

$$\frac{\partial^2 R}{\partial r^2} + \frac{1}{r}\frac{\partial R}{\partial r} + \left(k^2 - \frac{p^2}{r^2}\right)R = 0 \qquad [7\cdot36]$$

This equation is readily reducible to the standard form of Bessel's equation by a change of variable. For example, let $\rho = kr$. Then

$$r = \frac{\rho}{k} \quad \text{and} \quad \frac{\partial R}{\partial r} = k\frac{\partial R}{\partial \rho}, \quad \frac{\partial^2 R}{\partial r^2} = k^2\frac{\partial^2 R}{\partial \rho^2} \qquad [7\cdot37]$$

so that equation 7·36 becomes

$$k^2\frac{\partial^2 R}{\partial \rho^2} + \frac{k^2}{\rho}\frac{\partial R}{\partial \rho} + \left(k^2 - \frac{k^2 p^2}{\rho^2}\right)R = 0$$

or

$$\frac{\partial^2 R}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial R}{\partial \rho} + \left(1 - \frac{p^2}{\rho^2}\right)R = 0 \qquad [7\cdot38]$$

Equation 7·38 is the standard form of Bessel's differential equation. Each value of the parameter p is associated with a pair of fundamental solutions called Bessel functions of order p. One of these solutions is called a Bessel function of the first kind. It is finite at $\rho = 0$. The other fundamental solution is called a Bessel function of the second kind. Linear combinations of these solutions are sometimes called Bessel functions of the third kind or Hankel functions.

Let us consider first a special case for a Bessel function of the first kind where $p = 0$. Equation 7·38 reduces to

$$\frac{\partial^2 R}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial R}{\partial \rho} + R = 0 \qquad [7\cdot39]$$

This equation is known as Bessel's equation of zero order. One of the solutions of this equation, $J_0(\rho)$, is given by the infinite series

$$J_0(\rho) = 1 - \frac{\rho^2}{2^2} + \frac{\rho^4}{2^2\cdot4^2} - \frac{\rho^6}{2^2\cdot4^2\cdot6^2} + \cdots \qquad [7\cdot40]$$

* For a more complete discussion of Bessel's equation see pages 47–68, Kármán and Biot, *Mathematical Methods in Engineering*, McGraw-Hill Book Company, New York, N. Y.

and is known as the Bessel function of the first kind of zero order. Another solution of equation 7·39, $Y_0(\rho)$, is given by the relation

$$Y_0(\rho)^* = \frac{2}{\pi}\left\{J_0(\rho)\left(\log_e \frac{\rho}{2} + c'\right) + \left(\frac{\rho}{2}\right)^2 - \frac{\left(\frac{\rho}{2}\right)^4}{(2!)^2}\left(1 + \tfrac{1}{2}\right) + \frac{\left(\frac{\rho}{2}\right)^6}{(3!)^2}\left(1 + \tfrac{1}{2} + \tfrac{1}{3}\right) - \cdots\right\} \quad [7\cdot41]$$

where the constant c', called Euler's constant, has the value $c' = 0.57721\cdots$. The solution $Y_0(\rho)$ is known as the Bessel function of the second kind of zero order. We may therefore write the general solution of 7·39 in the form

$$R = C_3 J_0(\rho) + C_4 Y_0(\rho) \quad [7\cdot42]$$

where C_3 and C_4 are arbitrary constants. The zero-order Bessel functions $J_0(\rho)$ and $Y_0(\rho)$ are plotted in Fig. 7·3*a*, and their derivatives with respect to ρ in Fig. 7·3*b*. It is seen that they are oscillatory func-

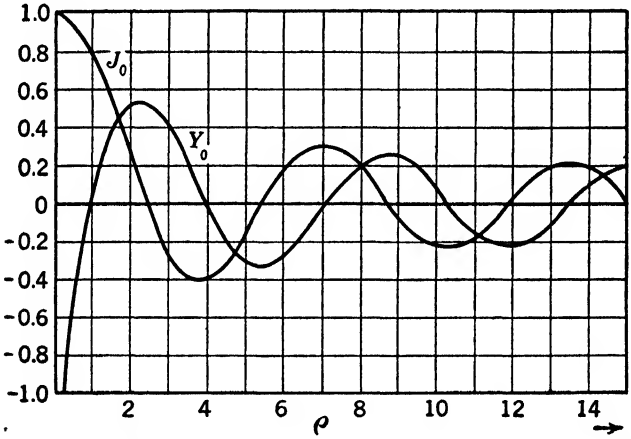


FIG. 7·3*a* Bessel function of zero order of the first and second kind.

tions, resembling damped sinusoidal and cosinusoidal functions, except that $Y_0(\rho)$ approaches $-\infty$ for small values of ρ.

The apprehension with which some engineers approach these functions is scarcely justified in view of the close relation the functions bear to trigonometric expressions. Roots of the trigonometric functions

* The solution $Y_0(\rho)$ as given here may be found in any textbook on Bessel functions such as *Bessel Functions for Engineers*, McLachlan, Oxford Press, 1934, page 69. In German textbooks and tables the notation $N_0(\rho)$ is generally used instead of $Y_0(\rho)$.

are evenly spaced at intervals of π. This is also true for the roots of higher rank in Bessel functions.

It can be shown either by an asymptotic expansion or by transformation to circular functions that, for large values of ρ, $J_0(\rho)$ and $Y_0(\rho)$ are given by the expressions

$$\left.\begin{aligned} J_0(\rho) &= \sqrt{\frac{2}{\pi\rho}}\cos\left(\rho - \frac{\pi}{4}\right) \\ Y_0(\rho) &= \sqrt{\frac{2}{\pi\rho}}\sin\left(\rho - \frac{\pi}{4}\right) \end{aligned}\right\} \text{ For large } \rho \qquad [7\cdot42a]$$

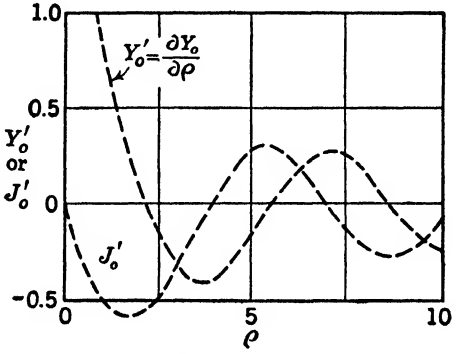


FIG. 7·3*b* The derivative of Bessel functions of zero order, of the first and second kind, with respect to their arguments.

The approximations are excellent, for at $\rho = 10$ the error is of the order of 1 per cent and steadily decreases as ρ increases. By the same expansion it could be established that lower-rank roots depart from this periodicity. Reference to Fig. 7·4 will make this immediately evident.

Bessel functions of the third kind, of zero order, are linear combinations of $J_0(\rho)$ and $Y_0(\rho)$. Thus:

$$H_0^{(1)}(\rho) = J_0(\rho) + jY_0(\rho) \qquad [7\cdot43]$$

$$H_0^{(2)}(\rho) = J_0(\rho) - jY_0(\rho) \qquad [7\cdot44]$$

where $H_0^{(1)}(\rho)$ and $H_0^{(2)}(\rho)$ are called Hankel functions of zero order. Such combinations are necessary in some problems in order that required boundary conditions be satisfied. Their presence in this section is in the interest of generality, for the special cases requiring their use will not be treated.

For large values of ρ, the asymptotic approximations of the zero-order Hankel functions are given by substitution of equations 7·42*a*

into 7·43 and 7·44.

$$\left.\begin{aligned} H_0^{(1)}(\rho) &\simeq \sqrt{\frac{2}{\pi\rho}}\, e^{j\left(\rho-\frac{\pi}{4}\right)} = \sqrt{\frac{2}{\pi\rho}}\left\{\cos\left(\rho-\frac{\pi}{4}\right)+j\sin\left(\rho-\frac{\pi}{4}\right)\right\} \\ H_0^{(2)}(\rho) &\simeq \sqrt{\frac{2}{\pi\rho}}\, e^{-j\left(\rho-\frac{\pi}{4}\right)} = \sqrt{\frac{2}{\pi\rho}}\left\{\cos\left(\rho-\frac{\pi}{4}\right)-j\sin\left(\rho-\frac{\pi}{4}\right)\right\} \end{aligned}\right\} \begin{matrix}\text{For large } \rho \\ [7.45]\end{matrix}$$

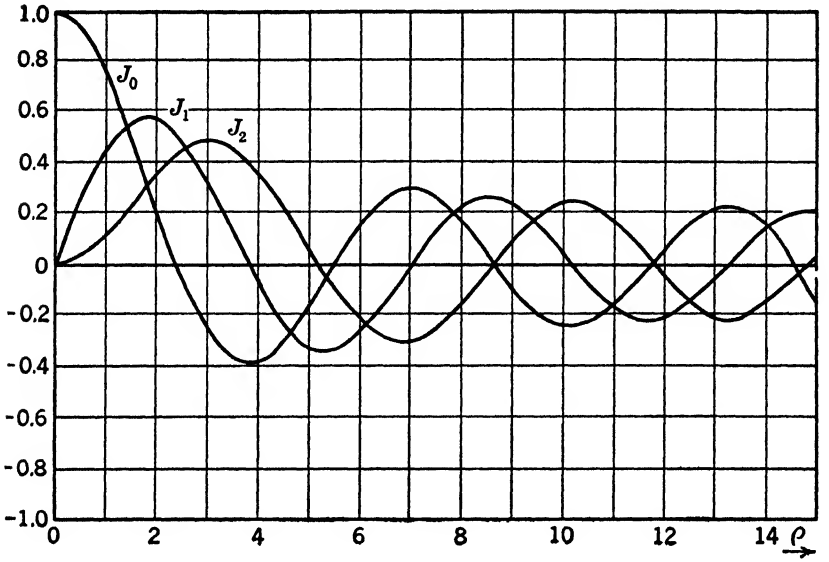


FIG. 7·4 Bessel functions of the first kind, of order zero, one, and two.

Bessel's equation of order p, as given in equation 7·38, is

$$\frac{\partial^2 R}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial R}{\partial\rho}+\left(1-\frac{p^2}{\rho^2}\right)R=0 \qquad [7.38]$$

A solution of this equation, $J_p(\rho)$, is called the Bessel function of order p defined by the infinite series

$$J_p(\rho)=\frac{1}{2^p\Gamma(p+1)}\rho^p\left\{1-\frac{\rho^2}{1!(p+1)2^2}+\frac{\rho^4}{2!(p+1)(p+2)2^4}-\cdots\right\} \qquad [7.46]$$

where $\Gamma(p+1)$ is the gamma* or factorial function defining the factorial p, i.e., $\Gamma(p+1)=p(p-1)(p-2)\cdots=p!$ $J_p(\rho)$ is sometimes ex-

* Numerical values of gamma functions have been computed and tabulated in several reference works. See B. O. Pierce, *A Short Table of Integrals*, page 140, Ginn and Co.

pressed more compactly by the summation

$$J_p(\rho) = \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\,\Gamma(p+m+1)} \left(\frac{\rho}{2}\right)^{p+2m} \qquad [7\cdot47]$$

Relation 7·46 or its equivalent 7·47 holds for integral or fractional values of p. If p is not an integer,* we may replace p by $-p$ and obtain another fundamental solution, $J_{-p}(\rho)$, which is called a Bessel function of negative order.

If p is an integer, say n, then the gamma function $\Gamma(p+1)$ may be replaced by $n!$ in equation 7·46, or $\Gamma(p+m+1)$ may be replaced by $(n+m)!$ in 7·47. Thus

$$J_n(\rho) = \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\,(m+n)!} \left(\frac{\rho}{2}\right)^{n+2m} \qquad (n = 0, 1, 2, 3, \cdots) \qquad [7\cdot48]$$

This expression is plotted in Fig. 7·4.

For sufficiently large values of ρ, $J_p(\rho)$ is given by the asymptotic expansion

$$J_p(\rho) \simeq \sqrt{\frac{2}{\pi\rho}} \cos\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right) \quad \text{for large } \rho \qquad [7\cdot49]$$

Another solution of equation 7·38 which is independent of $J_p(\rho)$ is given by

$$Y_p(\rho) = \frac{1}{\sin p\pi} \{J_p(\rho) \cos p\pi - J_{-p}(\rho)\} \qquad [7\cdot50]$$

where $Y_p(\rho)$ is a Bessel function of the second kind of order p. When p is an integer, this relation becomes indeterminate, $Y_p(\rho) = 0/0$, but can be evaluated by the usual methods. Since $Y_p(\rho)$ approaches $-\infty$ as ρ approaches zero, this function cannot be used to express physically finite fields in the neighborhood of the origin. See Fig. 7·5. We are therefore interested primarily in the asymptotic value of $Y_p(\rho)$, which is given by

$$Y_p(\rho) \simeq \sqrt{\frac{2}{\pi\rho}} \sin\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right) \quad \text{for large } \rho \qquad [7\cdot51]$$

Bessel functions of the third kind of order p, or Hankel functions of

* The gamma function is not finite for negative integral values of p.

order p as they are called, are defined by the relations

$$H_p^{(1)}(\rho) = J_p(\rho) + jY_p(\rho) \qquad [7\cdot52]$$

$$H_p^{(2)}(\rho) = J_p(\rho) - jY_p(\rho) \qquad [7\cdot53]$$

Thus the Hankel functions represent linear combinations of the first- and second-order Bessel functions just as $e^{\pm j\rho}$ represents a linear combination of cosine and sine functions; i.e.,

$$e^{j\rho} = \cos\rho + j\sin\rho$$

and

$$e^{-j\rho} = \cos\rho - j\sin\rho$$

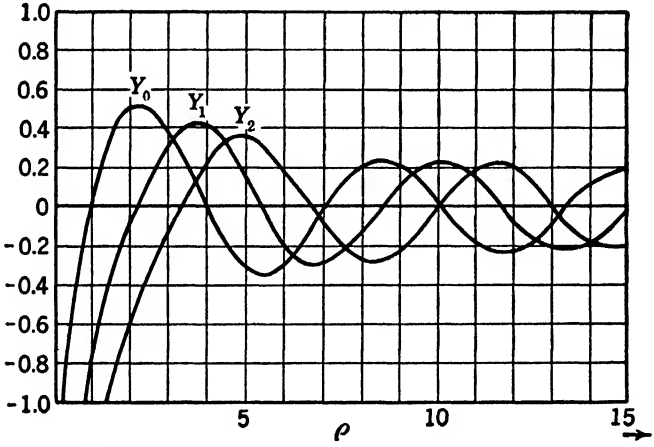


FIG. 7·5 Bessel functions of the second kind of order zero, one, and two.

For large values of the argument ρ,

$$H_p^{(1)}(\rho) \simeq \sqrt{\frac{2}{\pi\rho}}\, e^{j\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)} = \sqrt{\frac{2}{\pi\rho}}\left\{\cos\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right) + j\sin\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)\right\} \qquad [7\cdot54]$$

$$H_p^{(2)}(\rho) \simeq \sqrt{\frac{2}{\pi\rho}}\, e^{-j\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)} = \sqrt{\frac{2}{\pi\rho}}\left\{\cos\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right) - j\sin\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)\right\} \qquad [7\cdot55]$$

For convenience we may arrange the values of the various Bessel functions as ρ approaches zero or infinity as in Table 7·1.

TABLE 7·1

For Large Values of ρ	For Small Values of ρ
$J_p(\rho) = \sqrt{\frac{2}{\pi\rho}} \cos\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)$	$J_p(\rho) = \frac{\rho^p}{p!2^p}, \; p \neq 0$
$Y_p(\rho) = \sqrt{\frac{2}{\pi\rho}} \sin\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)$	$Y_p(\rho) = -\left(\frac{p!}{\pi}\right)\left(\frac{2}{\rho}\right)^p$
$H_p^{(1)}(\rho) = \sqrt{\frac{2}{\pi\rho}}\, e^{j\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)}$	$H_0^{(1)}(\rho) = \frac{2}{j\pi} J_0(\rho) \log_e \frac{1.7811}{j2}\rho$
$H_p^{(2)}(\rho) = \sqrt{\frac{2}{\pi\rho}}\, e^{-j\left(\rho - \frac{\pi}{4} - \frac{p\pi}{2}\right)}$	$H_0^{(2)}(\rho) = \frac{2}{j\pi} J_0(\rho) \log_e \frac{1.7811}{2}\rho$

It is seen from Table 7·1 that, as $\rho \to 0$, only the Bessel function of the first kind is finite, the others approaching infinity. When we attempt to satisfy the boundary conditions inside circular tubes or concentric cylinders it is generally necessary to use $J_p(\rho)$. For this reason $J_p(\rho)$ is sometimes called the internal Bessel function.

7·4 Electromagnetic Waves in the Dielectric within a Conducting Cylinder

Let us consider the propagation of electromagnetic waves through the homogeneous isotropic dielectric which fills the interior of a hollow metal cylinder. Let us take the X axis as the axis of the cylinder and consider propagation in the X direction. Let the inside radius of the cylinder be a, as indicated in Fig. 7·6.

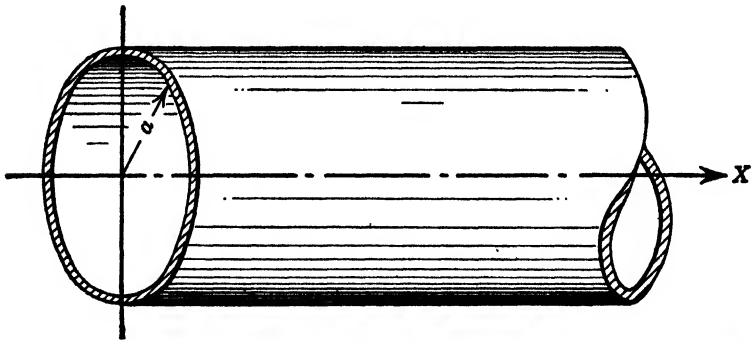


FIG. 7·6 A section of a cylindrical wave guide showing choice of axes.

We have seen in the previous section that the solution $Y_p(\rho)$ of Bessel's equation becomes infinite as ρ approaches zero. It is therefore not suitable for defining the field in the dielectric when the radius r is restricted by the relations $0 \leqslant r \leqslant a$ and $\sigma \ll j\omega\epsilon$. Recalling the

development earlier in the chapter and in particular equations 7·26, 7·31, and 7·34, it will be clear that equation 7·34*a* for the axial component of the magnetic intensity H_x may be written

$$H_x = \left\{ (A' \cos p\phi + B' \sin p\phi)\, J_p(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) \right\} e^{j\omega t - \gamma x} \quad [7\cdot56]^*$$

where $e^{j\omega t - \gamma x}$ enters as the function of distance and time, and where $C_4 = 0$.

A similar operation may be effected giving a solution for E_x in the dielectric. Thus

$$E_x = \left\{ (A \cos p\phi + B \sin p\phi)\, J_p(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) \right\} e^{j\omega t - \gamma x} \quad [7\cdot57]$$

In these equations, A, B, A', and B' are constants to be determined.

Equations 7·20–7·23, expressing the remaining components of the field in terms of E_x and H_x, become in the dielectric

$$E_r' = \frac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left\{ -\gamma \frac{\partial E_x'}{\partial r} - \frac{j\omega\mu}{r}\frac{\partial H_x'}{\partial \phi} \right\} \quad [7\cdot58]$$

$$E_\phi' = \frac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left\{ -\frac{\gamma}{r}\frac{\partial E_x'}{\partial \phi} + j\omega\mu \frac{\partial H_x'}{\partial r} \right\} \quad [7\cdot59]$$

$$H_r' = \frac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left\{ \frac{j\omega\varepsilon}{r}\frac{\partial E_x'}{\partial \phi} - \gamma \frac{\partial H_x'}{\partial r} \right\} \quad [7\cdot60]$$

$$H_\phi' = \frac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left\{ -j\omega\varepsilon \frac{\partial E_x'}{\partial r} - \frac{\gamma}{r}\frac{\partial H_x'}{\partial \phi} \right\} \quad [7\cdot61]$$

Waves in the dielectric $\sigma \ll j\omega\varepsilon$

Equations 7·58 through 7·61 define the electromagnetic field in the dielectric within the pipe. When the constants of equations 7·56 and 7·57 are chosen to fit the boundary conditions of the problem, the remaining components are determined by substituting E_x' and H_x' in equations 7·58 through 7·61.

7·5 Transmission Modes in a Perfectly Conducting Pipe

When the metal wall of the tube is a perfect conductor, the tangential components of the electric intensity, E_x and E_ϕ, must vanish there. Two types of waves can fulfill this condition, the H and the E type waves. As in the previous cases considered, the H type wave is one in which the longitudinal component of E vanishes everywhere, and the E type is one in which the longitudinal component of H vanishes every-

* Equation 7·34*a* is the general solution of the differential equation 7·25. Since the constants C_4, C_3 of 7·34*a* are arbitrary, C_4 is set = 0, yielding a particular solution which is more useful since $Y_p(\rho)$ is inoperative. Obviously equation 7·56 still satisfies equation 7·25.

where. Where the conducting wall is not a perfect conductor, it can be shown that E and H waves cannot exist alone unless the wave is circularly symmetrical. This will be discussed in a later section.

The equations for the two modes of transmission may be collected as follows:

***E* Waves in the Dielectric**

$$H_x = 0$$

$$E_x = \{(A\cos p\phi + B\sin p\phi)J_p(r\sqrt{\gamma^2+\omega^2\mu\varepsilon})\}\, e^{j\omega t-\gamma x} \qquad [7\cdot57]$$

$$\left.\begin{aligned} E_r &= -\left\{\frac{\gamma}{\gamma^2+\omega^2\mu\varepsilon}\frac{\partial E_x'}{\partial r}\right\} e^{j\omega t-\gamma x} \\ E_\phi &= -\left\{\frac{\gamma}{r(\gamma^2+\omega^2\mu\varepsilon)}\frac{\partial E_x'}{\partial \phi}\right\} e^{j\omega t-\gamma x} \\ H_r &= \left\{\frac{j\omega\varepsilon}{r(\gamma^2+\omega^2\mu\varepsilon)}\frac{\partial E_x'}{\partial \phi}\right\} e^{j\omega t-\gamma x} \\ H_\phi &= -\left\{\frac{j\omega\varepsilon}{\gamma^2+\omega^2\mu\varepsilon}\frac{\partial E_x'}{\partial r}\right\} e^{j\omega t-\gamma x} \end{aligned}\right\} \qquad [7\cdot62]$$

***H* Waves in the Dielectric**

$$E_x = 0$$

$$H_x = \{(A'\cos p\phi + B'\sin p\phi)J_p(r\sqrt{\gamma^2+\omega^2\mu\varepsilon})\}\, e^{j\omega t-\gamma x} \qquad [7\cdot56]$$

$$\left.\begin{aligned} E_r &= -\left\{\frac{j\omega\mu}{r(\gamma^2+\omega^2\mu\varepsilon)}\frac{\partial H_x'}{\partial \phi}\right\} e^{j\omega t-\gamma x} \\ E_\phi &= \left\{\frac{j\omega\mu}{(\gamma^2+\omega^2\mu\varepsilon)}\frac{\partial H_x'}{\partial r}\right\} e^{j\omega t-\gamma x} \\ H_r &= -\left\{\frac{\gamma}{\gamma^2+\omega^2\mu\varepsilon}\frac{\partial H_x'}{\partial r}\right\} e^{j\omega t-\gamma x} \\ H_\phi &= -\left\{\frac{\gamma}{r(\gamma^2+\omega^2\mu\varepsilon)}\frac{\partial H_x'}{\partial \phi}\right\} e^{j\omega t-\gamma x} \end{aligned}\right\} \qquad [7\cdot63]$$

7·6 Boundary Conditions for *E* Waves

Let us examine the solution for E_x as given in equation 7·57. It is clear that in any circular structure a reference of angular position is necessarily arbitrary. Accordingly, we are free to choose a reference in such a way as to make the mathematical analysis as simple as possible. In the present case the mathematical analysis is simplified if we choose

the reference so that $B = 0$. Our solution then becomes

$$E'_x = (A \cos p\phi) J_p(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) \qquad [7\cdot64]$$

Let us consider the term involving ϕ. It is evident from physical considerations that the electric intensity E has one and only one value at each point within the tube. Thus, when r and x are held constant, we may increase ϕ by 360° without changing the value of E.

The function $\cos p\phi$ satisfies this condition if and only if p is an integer. The situation is similar to that existing in Fourier analysis. Only harmonics of the fundamental frequency return to their proper relative position when 360° is added to the fundamental angle. Thus, in order that our solution be single valued, p must be an integer. Let us set $p = n = 0, 1, 2, 3 \cdots$, so that 7.64 becomes

$$E'_x = A \cos n\phi \; J_n(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) \quad (n = 0, 1, 2, 3 \cdots) \qquad [7\cdot65]$$

Let us consider, now, the Bessel function $J_n(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon})$ under the restriction that n is an integer. At the inside surface of the pipe $r = a$, and, since E_x is zero there, we must have

$$J_n(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) = 0 \qquad [7\cdot66]$$

From Fig. 7·4 it is evident that J_n is zero only for certain definite values of the argument, called the roots of J_n. There are an infinite number of such roots, a few of which are given in Table 7·2.

TABLE 7·2

ROOTS OF $J_n(\rho) = 0$

Rank (m)	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1	2.405	3.832	5.135	6.379	7.586	8.780
2	5.520	7.016	8.417	9.760	11.064	12.339
3	8.654	10.173	11.620	13.017	14.373	15.700
4	11.792	13.323	14.796	16.224	17.616	18.982
5	14.931	16.470	17.960	19.410	20.827	22.220
6	18.071	19.616	21.117	22.583	24.018	25.431

Let r_{nm} represent the roots of J_n, where n refers to the order of the Bessel function and m refers to the rank of the root. Thus $r_{01} = 2.405$, $r_{23} = 11.620$, etc. Hence in order that 7·66 be satisfied

$$a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = r_{nm} \qquad [7\cdot67]$$

giving

$$\gamma_{nm}^2 = \left(\frac{r_{nm}}{a}\right)^2 - \omega^2\mu_1\varepsilon_1 \qquad [7\cdot68]$$

or
$$\gamma_{nm} = \sqrt{\left(\frac{r_{nm}}{a}\right)^2 - \omega^2\mu_1\varepsilon_1} \qquad (7\cdot69]$$

Since γ_{nm} must be imaginary in order that propagation take place, it is necessary that $(r_{nm}/a)^2 < \omega^2\mu_1\varepsilon_1$. It is therefore more convenient to write

$$\gamma_{nm} = j\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r_{nm}}{a}\right)^2} \qquad [7\cdot70]$$

The propagation constant $\gamma_{nm} = \alpha_{nm} + j\beta_{nm}$, where α_{nm} is the attenuation constant and β_{nm} the phase constant. Hence

$$\alpha_{nm} + j\beta_{nm} = j\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r_{nm}}{a}\right)^2} \qquad [7\cdot71]$$

giving
$$\alpha_{nm} = 0$$

$$\beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r_{nm}}{a}\right)^2} \qquad [7\cdot72]$$

Thus we find that for perfectly conducting tubes $\alpha_{nm} = 0$ and there is no attenuation. If β_{nm} is real, i.e., if $\omega^2\mu_1\varepsilon_1 > (r_{nm}/a)^2$, the wave will be transmitted down the tube. If $\omega^2\mu_1\varepsilon_1 < (r_{nm}/a)^2$, β_{nm} as defined by equation 7·72 is imaginary and there is no transmission at all. This sets a lower limit on the frequency at which transmission will take place. As in the previous cases we can define a critical frequency $f_0 = \omega_0/2\pi$, where $\omega_0^2\mu_1\varepsilon_1 = (r_{nm}/a)^2$. Hence

$$(f_0)_{nm} = \frac{1}{2\pi\sqrt{\mu_1\varepsilon_1}}\left(\frac{r_{nm}}{a}\right) \qquad [7\cdot73]$$

For frequencies greater than $(f_0)_{nm}$, β_{nm} is real and the wave will be propagated down the tube. For a wave whose frequency is less than $(f_0)_{nm}$, β_{nm} is imaginary and there is no transmission. We may define a critical wavelength $(\lambda_0)_{nm}$ by the relation $(\lambda_0)_{nm}\,(f_0)_{nm} = v_1$, where $v_1 = 1/\sqrt{\mu_1\varepsilon_1}$ is the velocity of a free wave in a medium of dielectric constant ε_1 and permeability μ_1. Thus

$$(\lambda_0)_{nm} = \frac{1}{(f_0)_{nm}\sqrt{\mu_1\varepsilon_1}} = \frac{2\pi a}{r_{nm}} \qquad [7\cdot74]$$

The phase velocity which has been defined previously is

$$(v_p)_{nm} = \frac{\omega}{(\beta)_{nm}} = \frac{\omega}{\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r_{nm}}{a}\right)^2}} \qquad [7\cdot75]$$

and the group velocity $(v_g)_{nm}$ is

$$(v_g)_{nm} = \frac{1}{\dfrac{\partial \beta_{nm}}{\partial \omega}} = \frac{\sqrt{\omega^2 \mu_1 \varepsilon_1 - \left(\dfrac{r_{nm}}{a}\right)^2}}{\omega \mu_1 \varepsilon_1} \qquad [7\cdot76]$$

7·7 E_{nm} Waves in the Dielectric

From equation 7·67 we may write $\sqrt{\gamma^2 + \omega^2 \mu_1 \varepsilon_1} = r_{nm}/a$. Replacing the radical in equation 7·65 we obtain for the amplitude of the axial component of the electric intensity

$$E'_x = A \cos n\phi J_n\left(r \frac{r_{nm}}{a}\right) \quad (n,m = 0, 1, 2, 3, \cdots) \qquad [7\cdot77]$$

To obtain the remaining components we need the following derivatives

$$\frac{\partial E'_x}{\partial \phi} = -An \sin n\phi \, J_n\left(r \frac{r_{nm}}{a}\right) \qquad [7\cdot78]$$

and

$$\frac{\partial E'_x}{\partial r} = A \cos n\phi \left\{\frac{n}{r} J_n\left(r \frac{r_{nm}}{a}\right) - \frac{r_{nm}}{a} J_{n+1}\left(r \frac{r_{nm}}{a}\right)\right\} \qquad [7\cdot79]^*$$

* The derivative of the Bessel function $J_n(kx)$ with respect to x is

$$\frac{\partial J_n(kx)}{\partial x} = \frac{n}{x} J_n(kx) - kJ_{n+1}(kx) = kJ'_n(kx),$$

where $J'_n(kx)$ is the derivative with respect to x. This follows directly from differentiation of equation 7·47 with $\rho = kx$ and $p = n$, an integer. From equation 7·47

$$J_n(kx) = \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\Gamma(n + m + 1)} \left(\frac{kx}{2}\right)^{n+2m}$$

and

$$J_{n+1}(kx) = \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\Gamma(n + m + 2)} \left(\frac{kx}{2}\right)^{n+2m+1}$$

Differentiating $J_n(kx)$ with respect to x, we obtain

$$\begin{aligned}\frac{\partial J_n(kx)}{\partial x} &= \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\Gamma(n + m + 1)} (n + 2m) \left(\frac{k}{2}\right)\left(\frac{kx}{2}\right)^{n+2m-1} \\ &= \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\Gamma(n + m + 1)} n \left(\frac{k}{2}\right)\left(\frac{kx}{2}\right)^{n+2m-1} \\ &\quad + \sum_{m=0}^{m=\infty} \frac{(-1)^m}{m(m-1)!\Gamma(n + m + 1)} 2m \left(\frac{k}{2}\right)\left(\frac{kx}{2}\right)^{n+2m-1}\end{aligned}$$

We may now write the complete equations for the E_{nm} waves in the dielectric by substituting 7·78 and 7·79 in equations 7·62. It is convenient to replace $\gamma_{nm}^2 + \omega^2\mu_1\varepsilon_1$ with $(r_{nm}/a)^2$ and γ_{nm} with $j\beta_{nm}$. Thus the following equations result:

$$\left.\begin{aligned}
E_x &= \left\{A \cos n\phi \, J_n\left(r\frac{r_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)} \\
E_r &= -j\left\{A\beta_{nm}\left(\frac{a}{r_{nm}}\right)^2 \cos n\phi \times \right. \\
&\qquad \left.\left[\frac{n}{r} J_n\left(r\frac{r_{nm}}{a}\right) - \left(\frac{r_{nm}}{a}\right) J_{n+1}\left(r\frac{r_{nm}}{a}\right)\right]\right\} e^{j(\omega t-\beta_{nm}x)} \\
E_\phi &= j\left\{A\beta_{nm}\left(\frac{a}{r_{nm}}\right)^2 \frac{n}{r} \sin n\phi \, J_n\left(r\frac{r_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)} \\
H_x &= 0 \\
H_r &= -j\left\{A\omega\varepsilon_1\left(\frac{a}{r_{nm}}\right)^2 \frac{n}{r} \sin n\phi \, J_n\left(r\frac{r_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)} \\
H_\phi &= -j\left\{A\omega\varepsilon_1\left(\frac{a}{r_{nm}}\right)^2 \cos n\phi \times \right. \\
&\qquad \left.\left[\frac{n}{r} J_n\left(r\frac{r_{nm}}{a}\right) - \frac{r_{nm}}{a} J_{n+1}\left(r\frac{r_{nm}}{a}\right)\right]\right\} e^{j(\omega t-\beta_{nm}x)}
\end{aligned}\right\} \quad [7\cdot80]$$

Multiplying both sides by x and rearranging terms, we obtain

$$\begin{aligned}
x\frac{\partial J_n(kx)}{\partial x} &= n\sum_{m=0}^{m=\infty} \frac{(-1)^m}{m!\Gamma(n+m+1)}\left(\frac{kx}{2}\right)^{n+2m} \\
&\qquad + kx\sum_{m=0}^{m=\infty} \frac{(-1)^m}{(m-1)!\Gamma(n+m+1)}\left(\frac{kx}{2}\right)^{n+2m-1} \\
&= nJ_n(kx) - kx\sum_{q=0}^{q=\infty} \frac{(-1)^q}{q!\Gamma(n+q+2)}\left(\frac{kx}{2}\right)^{n+1+2q}
\end{aligned}$$

where $m - 1 = q$. Hence

$$x\frac{\partial J_n(kx)}{\partial x} = nJ_n(kx) - kxJ_{n+1}(kx)$$

Dividing through by x

$$\frac{\partial J_n(kx)}{\partial x} = \frac{n}{x}J_n(kx) - kJ_{n+1}(kx) = kJ_n'(kx)$$

The propagation properties of E_{nm} waves are given by the following relationships:

$$
\left.
\begin{aligned}
&\text{Propagation constant} && \beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r_{nm}}{a}\right)^2} \\
&&& = \sqrt{\left(\frac{\omega}{v_1}\right)^2 - \left(\frac{r_{nm}}{a}\right)^2} \\
&&& = \sqrt{\left(\frac{2\pi}{\lambda_1}\right)^2 - \left(\frac{2\pi}{(\lambda_0)_{nm}}\right)^2} \\
&&& = \frac{\omega}{v_1}\sqrt{1 - \left(\frac{(f_0)_{nm}}{f}\right)^2} \\
&\text{Critical frequency} && (f_0)_{nm} = \frac{v_1}{2\pi}\left(\frac{r_{nm}}{a}\right) \\
&\text{Critical wavelength} && (\lambda_0)_{nm} = 2\pi\left(\frac{a}{r_{nm}}\right) \\
&\text{Wavelength in pipe} && (\lambda_p)_{nm} = \frac{2\pi}{\beta_{nm}} \\
&\text{Free wavelength} && \lambda_1 = 2\pi\left(\frac{v_1}{\omega}\right) \\
&\text{Phase velocity} && (v_p)_{nm} = \frac{\omega}{\beta_{nm}} \\
&\text{Group velocity} && (v_g)_{nm} = \frac{v_1^2}{(v_p)_{nm}} \\
&\text{Free wave velocity} && v_1 = \frac{1}{\sqrt{\mu_1\varepsilon_1}}
\end{aligned}
\right\} \quad [7\cdot81]
$$

There is no root r_{00} of the Bessel function J_0; hence no wave of the type E_{00} exists. Waves of the type E_{0m} exist, however, and W. L. Barrow has made an exhaustive study of the E_{01} mode of transmission. This wave is usually referred to as the E_0 wave. Waves of the type E_{11} are referred to as E_1 waves, E_{21} waves as E_2 waves, etc. Here as in the plane and rectangular wave guides the designation TE is equivalent to H. Similarly the designations TM and E are equivalent. To be exact it is necessary to use double subscripts to identify any of these wave types.

7·8 The E_0 Wave. Perfect Conductivity

It may be seen from equations 7·80 that, for $n = 0$, $m = 1$, the components E_ϕ and H_r reduce to zero. The remaining components become

$$\left.\begin{aligned} E_x &= AJ_0\left(r\frac{r_{01}}{a}\right)e^{j(\omega t-\beta_{01}x)} \\ E_r &= jA\frac{\beta_{01}}{r_{01}}aJ_1\left(r\frac{r_{01}}{a}\right)e^{j(\omega t-\beta_{01}x)} \\ H_\phi &= jA\frac{\omega\varepsilon_1}{r_{01}}aJ_1\left(r\frac{r_{01}}{a}\right)e^{j(\omega t-\beta_{01}x)} \end{aligned}\right\} E_0 \text{ wave in the dielectric [7·82]}$$

where $r_{01} = 2.405$ and $\beta_{01} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{2.405}{a}\right)^2}$.

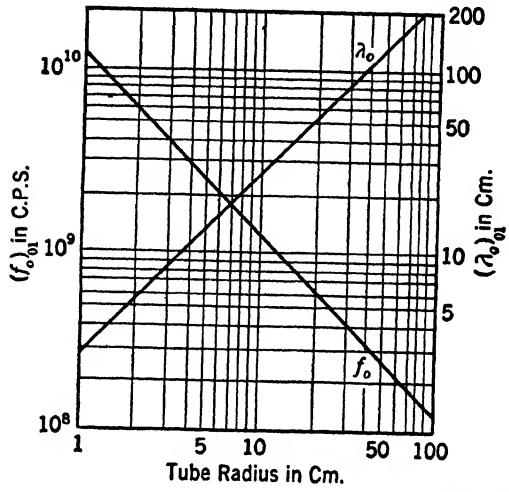
If the dielectric is air, then $\mu_1 = \mu_0$, $\varepsilon_1 = \varepsilon_0$ very nearly, and $v_1 = 1/\sqrt{\mu_0\varepsilon_0} \simeq c = 3 \times 10^8$ meters per second. Then

$$\left.\begin{aligned} \beta_{01} &= \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{2.405}{a}\right)^2} \\ (f_0)_{01} &= \frac{c}{2\pi}\left(\frac{2.405}{a}\right) \cong 1.15 \times 10^8\left(\frac{1}{a}\right) \\ (\lambda_0)_{01} &= 2\pi\left(\frac{a}{2.405}\right) \simeq 2.615a \\ (\lambda_p)_{01} &= \frac{2\pi}{\sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{2.405}{a}\right)^2}} \\ (v_p)_{01} &= \frac{\omega}{\sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{2.405}{a}\right)^2}} \\ (v_g)_{01} &= \frac{c^2}{(v_p)_{01}} \end{aligned}\right\} E_0 \text{ wave when dielectric is air } v_1 \cong c \text{ [7·83]}$$

Since the critical frequency is inversely proportional to the tube radius a, the smaller the tube, the higher the frequency required for transmission. A plot of the critical frequency and the critical wavelength as a function of tube radius is shown in Fig. 7·7 for an air dielectric. The corresponding wavelength in free space is also shown as a function of tube radius. It can be seen that tube radii less than 10 cm require frequencies in excess of 10^9 cycles per second.

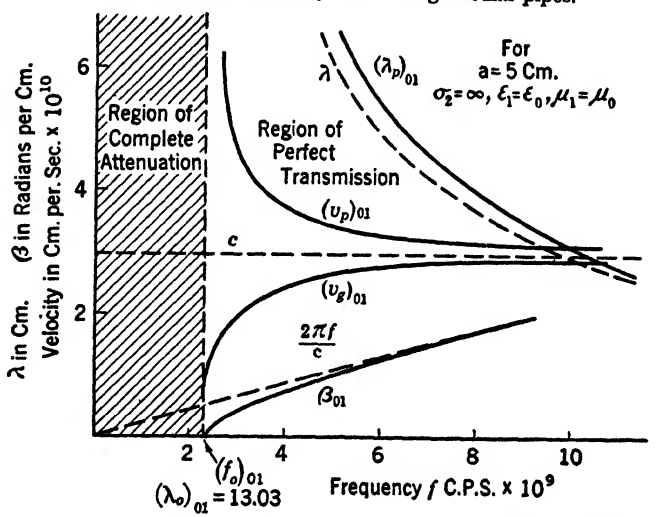
The quantities given in equation 7·83 are plotted as a function of

frequency in Fig 7·8. Although in this diagram the tube radius is held constant at 5 cm, the curves shown are typical for tubes of any



(*Barrow, courtesy of IRE*)

FIG. 7·7 Critical frequency and critical wavelength as a function of tube radius for E_0 waves in perfectly conducting circular pipes.



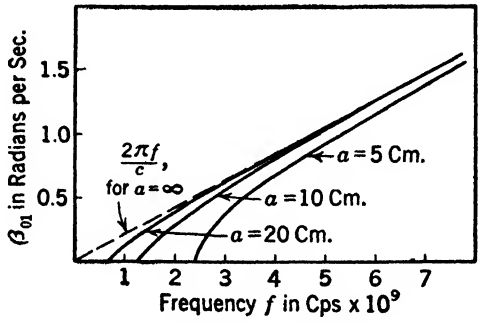
(*Barrow, courtesy of IRE*)

FIG. 7·8 Transmission characteristics of E_0 waves in a perfectly conducting circular pipe.

radius. As for the rectangular pipe, the phase velocity is infinite at the critical frequency and approaches the velocity of light as the frequency

is increased. The group velocity, on the other hand, is zero at the critical frequency, but also approaches the velocity of light with increased frequency. It is seen, also, that the wavelength in the pipe approaches the wavelength in free space as the frequency is increased.

A plot of β_{01} as a function of frequency for various tube radii is shown in Fig. 7·9. As the radius of the pipe is increased so that $a \gg \lambda_a$, β_{01} approaches the value $2\pi f/c$.



(Barrow, courtesy of IRE)

FIG. 7·9 Variation of the phase constant β_{01} with the radius of a circular guide.

As for the pipe of rectangular cross section, we may obtain a plot of the electric intensity inside the pipe by setting t equal to zero in equations for E_x and E_r in 7·82. Thus,

$$|E_x| = AJ_0\left(r\frac{r_{01}}{a}\right)\cos\beta_{01}x \qquad [7\cdot84]$$

$$|E_r| = A\frac{\beta_{01}a}{r_{01}}J_1\left(r\frac{r_{01}}{a}\right)\sin\beta_{01}x \qquad [7\cdot85]$$

The differential equation of the lines of electric intensity is thus

$$\frac{\partial(\beta_{01}x)}{\partial\left(\frac{r}{a}\right)} = \frac{|E_x|}{|E_r|} = \frac{r_{01}}{\beta_{01}a}\frac{J_0\left(r\frac{r_{01}}{a}\right)}{J_1\left(r\frac{r_{01}}{a}\right)}\cot\beta_{01}x \qquad [7\cdot86]$$

Barrow solved this equation by graphical means and obtained the distribution shown in Fig. 7·10. The picture shows the field structure at any instant of time. The propagation of the wave down the tube may be thought of as a movement of this distribution through the tube at a velocity $(v_p)_{01}$ given in equations 7·83.

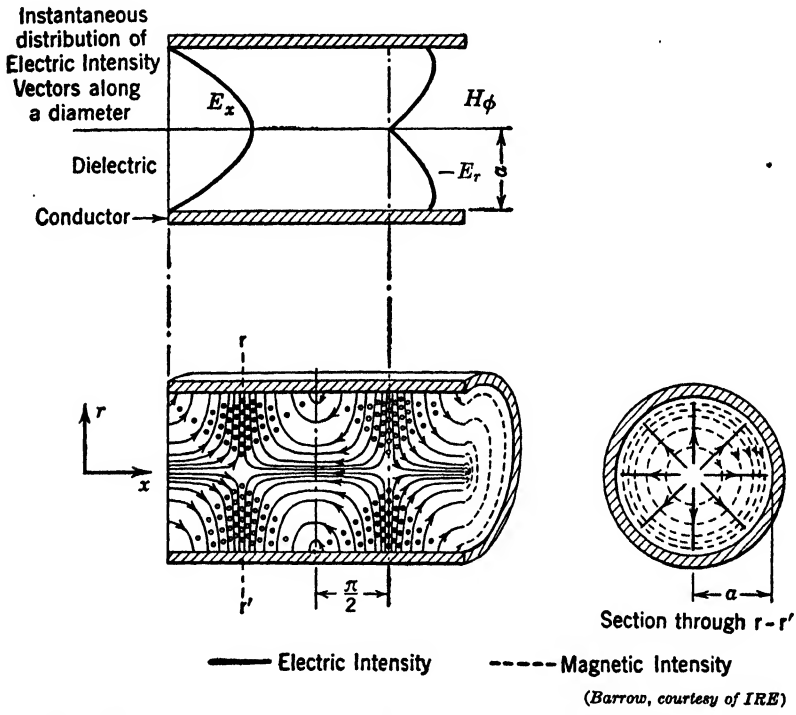
We may calculate the transverse electromotive force in volts between the tube axis and the inside surface of the tube by taking the line integral

of the electric intensity between these points. Thus,

$$V' = \int_0^a E_r' \, dr = jA \frac{\beta_{01} a}{r_{01}} \int_0^a J_1\left(r \frac{r_{01}}{a}\right) dr = jA \frac{\beta_{01} a^2}{(r_{01})^2} J_0\left(r \frac{r_{01}}{a}\right)\Bigg|_0^a \quad [7\cdot87]^*$$

Hence

$$V' = jA\beta_{01}\left(\frac{a}{r_{01}}\right)^2 \text{ volts} \quad [7\cdot88]$$



(*Barrow, courtesy of IRE*)

FIG. 7·10 Instantaneous field configuration of an E_{01} wave (TM_{01}) in a circular wave guide.

The longitudinal current in amperes along the inside surface of the tube may be calculated by taking the line integral of H_ϕ around this surface. Since there is a surface layer of current on the conductor, the tangential component of magnetic intensity is discontinuous at the

$$^* \int_0^a J_1\left(r \frac{r_{01}}{a}\right) dr = -\frac{a}{r_{01}}\left[J_0\left(r \frac{r_{01}}{a}\right)\right]\Bigg|_0^a = -\frac{a}{r_{01}}\left[J_0(r_{01}) - J_0(0)\right] = \frac{a}{r_{01}}$$

since $J_0(0) = +1$ and $J_0(r_{01}) = 0$, for r_{01} is by definition the first root of the zero-order Bessel function.

surface and is zero inside the surface. Hence the only contribution to the line integral is given by the integral of H_ϕ taken at $r = a$. Thus,

$$I' = \oint_{r=a} H'_\phi \, d\phi = jA \frac{\omega\varepsilon_1 a}{r_{01}} J_1(r_{01}) \oint_{r=a} d\phi \qquad [7\cdot89]$$

$$= jA \frac{\omega\varepsilon_1 a}{r_{01}} J_1(r_{01}) \, 2\pi a \quad \text{amperes} \qquad [7\cdot90]$$

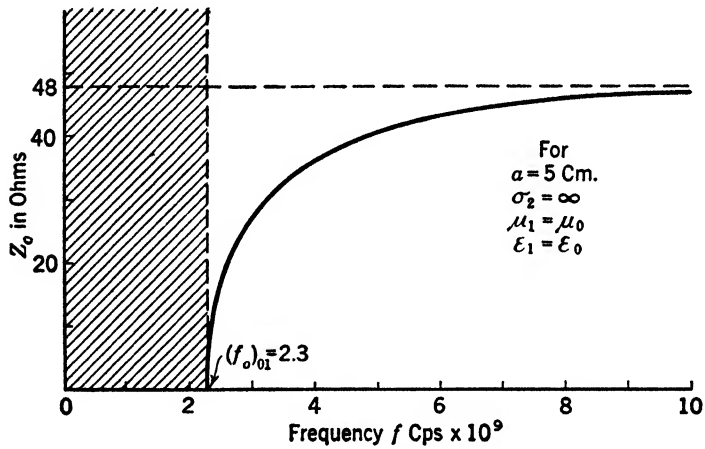
We may define a characteristic impedance in ohms as the ratio of this transverse voltage V' and the longitudinal current I'. Thus,

$$Z_0 = \frac{V'}{I'} = \frac{\beta_{01}}{2\pi\omega\varepsilon_1 r_{01} J_1(r_{01})}$$

$$= \frac{1}{2\pi r_{01} J_1(r_{01})} \sqrt{\frac{\mu_1}{\varepsilon_1}} \sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2} \qquad [7\cdot91]^*$$

For air dielectric, this becomes

$$Z_0 = 48 \sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2}$$



(Barrow, courtesy of IRE)

FIG. 7·11 Characteristic impedance vs. frequency for a 5-cm radius, perfectly conducting air-filled tube.

The characteristic impedance Z_0 is plotted in Fig. 7·11 for a pipe of 5-cm radius. Since β_{01} is zero at the critical frequency, Z_0 is zero at this frequency. As f increases above $(f_0)_{01}$, Z_0 increases and approaches

* $\frac{\beta_{01}}{\omega} = \frac{1}{v_1}\sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2} = \sqrt{\mu_1\varepsilon_1}\sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2}$

48 ohms as f approaches infinity. Similar results are obtained for other radii. On the basis of this definition it is seen that the hollow-pipe transmission system is essentially a low-impedance device.

The characteristic impedance, defined on a power basis,* is found to be

$$Z_0 = 30\sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2}$$

when the dielectric is air.

7·9 The E_1 Wave. Perfect Conductivity

The equations for the E_1 wave in the dielectric may be written from equations 7·80 by setting $n = 1$, $m = 1$. The wave structure for this mode of transmission, which is much more complicated than that for the E_0 wave, is shown in Fig. 7·12.

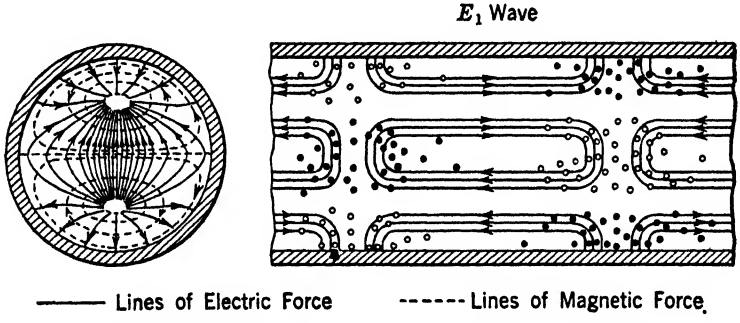


FIG. 7·12 Instantaneous field configuration of E_1 wave in cylindrical wave guide.

The critical frequency for the E_{11} wave is given by the relation

$$(f_0)_{11} = \frac{v_1}{2\pi}\left(\frac{r_{11}}{a}\right) \qquad [7·92]$$

where $r_{11} \doteq 3.83$. (See Table 7·2.) Thus this transmission mode is possible for all frequencies greater than $(f_0)_{11}$ as given in 7·92. The higher-order E waves are also possible theoretically but will not be considered here. The frequencies required for this transmission become so high as to be of little practical importance at this time.

7·10 Boundary Conditions for H Waves

By our choice of the reference of angular position we may take B' in equation 7·56 equal to zero so that our solution for H becomes

$$H_x' = A' \cos p\phi J_p(r\sqrt{\gamma^2 + \omega^2\mu_1\epsilon_1}) \qquad [7·93]$$

* Carson, Mead, and Schelkunoff, *Bell System Tech. J.*, 15, 2, p. 310, 1936.

In order that this solution be single valued, p must be an integer, n. Thus,

$$H'_x = A' \cos n\phi J_n(r\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) \quad (n = 0, 1, 2, 3 \cdots) \quad [7\cdot94]$$

The amplitude of the tangential component of the electric intensity from equation 7·63 is

$$E'_\phi = \frac{j\omega\mu}{\gamma^2 + \omega^2\mu_1\varepsilon_1}\frac{\partial H'_x}{\partial r} = A'\frac{j\omega\mu}{\gamma^2 + \omega^2\mu_1\varepsilon_1}\cos n\phi \frac{\partial}{\partial r}[J_n(r\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})] \quad [7\cdot95]$$

But E'_ϕ must be zero at the surface of the conductor at $r = a$. This requires that

$$\frac{\partial}{\partial r} J_n(r\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) = 0 \quad \text{at } r = a \quad [7\cdot96]$$

or

$$a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = r'_{nm} \quad [7\cdot97]^*$$

where r'_{mn} is the mth root of $\frac{\partial}{\partial r} J_n(r\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})$ at $r = a$. Several

* As shown for equation 7·79,

$$\frac{\partial J_n(kr)}{\partial r} = \frac{nJ_n(kr) - krJ_{n+1}(kr)}{r}$$

or

$$kJ'_n(kr) = \frac{n}{r}J_n(kr) - kJ_{n+1}(kr)$$

Then

$$\frac{\partial}{\partial r}J_n(r\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1}) = \frac{n}{r}J_n(r\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1}) - \sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1}\,J_{n+1}(r\sqrt{\gamma^2+\omega^2\mu_1\varepsilon_1})$$

This must be zero at $r = a$. Hence

$$\frac{n}{a}J_n(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) - \sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}\,J_{n+1}(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) = 0$$

This requires that

$$nJ_n(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) = a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}\,J_{n+1}(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})$$

When $n = 0$

$$J_1(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}) = 0$$

and

$$a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = r_{1m}$$

where r_{1m} is the mth root of the first-order Bessel function.

Thus, if $m = 1$, $r'_{01} = r_{11} = 3.83$; if $m = 2$, $r'_{02} = r_{12} = 7.02$, etc.

of the lower order roots are:

$$r'_{01},\ r'_{02}\cdots = 3.83,\ 7.02\cdots$$

$$r'_{11},\ r'_{12}\cdots = 1.84,\ 5.33\cdots$$

Thus

$$\gamma_{nm}^2 = \left(\frac{r'_{nm}}{a}\right)^2 - \omega^2\mu_1\varepsilon_1 \qquad [7\cdot98]$$

$$\gamma_{nm} = \sqrt{\left(\frac{r'_{nm}}{a}\right)^2 - \omega^2\mu_1\varepsilon_1} \qquad [7\cdot99]$$

Equation 7·98 can be satisfied only by values of $\gamma_{nm}^2 < \omega^2\mu_1\varepsilon_1$. Hence we may write

$$\gamma_{nm} = j\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r'_{nm}}{a}\right)^2} \qquad [7\cdot100]$$

As in the case of the E waves we may write

$$\gamma_{nm} = \alpha_{nm} + j\beta_{nm} = j\sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r'_{nm}}{a}\right)^2} \qquad [7\cdot101]$$

which leads to the conclusion that $\alpha_{nm} = 0$ where the metal of the tube is a perfect conductor, and

$$\beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r'_{nm}}{a}\right)^2} \qquad [7\cdot102]$$

Since β_{nm} must be real if propagation is to take place, $\omega^2\mu_1\varepsilon_1 > (r'_{nm}/a)^2$, and we can define a critical frequency, $\omega_0/2\pi$, below which there can be no transmission.* Thus

$$(f_0)_{nm} = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{\mu_1\varepsilon_1}}\left(\frac{r'_{nm}}{a}\right) \qquad [7\cdot103]$$

When $n = 1$,

$$J_1(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}\,J_2(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})$$

or

$$\frac{J_1(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})}{J_2(a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1})} = a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} \qquad [7\cdot97a]$$

If the lowest value of $a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1}$ that will satisfy this equation is $a\sqrt{\gamma^2 + \omega^2\mu_1\varepsilon_1} = 1.84$, then we call this the first root or $r'_{11} = 1.84$. The next highest root to satisfy 7·97a may be checked by reference to Fig. 7·4.

* The qualifying remarks given in Chapter 6 also apply here.

The remaining propagation quantities $(\lambda_0)_{nm}$, $(\lambda_p)_{nm}$, λ_1, $(v_p)_{nm}$, $(v_g)_{nm}$ may now be written for the H wave in a similar way to that employed for the E waves discussed in section 7·7.

7·11 H_{nm} Waves in the Dielectric

The equations for the H_{nm} waves may be written from equations 7·63 by substituting the required partial derivatives, $\partial H'_x/\partial\phi$ and $\partial H'_x/\partial r$. From equation 7·97 we may write $\sqrt{\lambda^2 + \omega^2\mu_1\varepsilon_1} = (r'_{nm}/a)$. Replacing the radical in equation 7·94 we obtain, for the axial component of the magnetic intensity

$$H'_x = A' \cos n\phi J_n\left(r\frac{r'_{nm}}{a}\right) \quad (n, m = 0, 1, 2, 3, \cdots)$$

Hence

$$\frac{\partial H'_x}{\partial\phi} = -A'n \sin n\phi J_n\left(r\frac{r'_{nm}}{a}\right)$$

and

$$\frac{\partial H'_x}{\partial r} = A' \cos n\phi \left\{\frac{n}{r} J_n\left(r\frac{r'_{nm}}{a}\right) - \left(\frac{r'_{nm}}{a}\right) J_{n+1}\left(r\frac{r'_{nm}}{a}\right)\right\}$$

Substituting these derivatives into equations 7·63 and replacing $\gamma^2 + \omega^2\mu_1\varepsilon_1$ with $(r'_{nm}/a)^2$ and γ with $j\beta$, we obtain

$$\left.\begin{aligned}
E_x &= 0\\
E_r &= j\left\{A'\omega\mu_1\left(\frac{a}{r'_{nm}}\right)^2 \frac{n}{r} \sin n\phi\, J_n\left(r\frac{r'_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)}\\
E_\phi &= j\left\{A'\omega\mu_1\left(\frac{a}{r'_{nm}}\right)^2 \cos n\phi\left[\frac{n}{r} J_n\left(r\frac{r'_{nm}}{a}\right)\right.\right.\\
&\qquad \left.\left. - \left(\frac{r'_{nm}}{a}\right) J_{n+1}\left(r\frac{r'_{nm}}{a}\right)\right]\right\} e^{j(\omega t-\beta_{nm}x)}\\
H_x &= \left\{A' \cos n\phi J_n\left(r\frac{r'_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)}\\
H_r &= -j\left\{A'\beta_{nm}\left(\frac{a}{r'_{nm}}\right)^2 \cos n\phi\left[\frac{n}{r} J_n\left(r\frac{r'_{nm}}{a}\right)\right.\right.\\
&\qquad \left.\left. - \left(\frac{r'_{nm}}{a}\right) J_{n+1}\left(r\frac{r'_{nm}}{a}\right)\right]\right\} e^{j(\omega t-\beta_{nm}x)}\\
H_\phi &= j\left\{A'\beta_{nm}\left(\frac{a}{r'_{nm}}\right)^2 \frac{n}{r} \sin n\phi\, J_n\left(r\frac{r'_{nm}}{a}\right)\right\} e^{j(\omega t-\beta_{nm}x)}
\end{aligned}\right\} \begin{array}{l}H_{nm}\text{ waves}\\ \text{in the}\\ \text{dielectric}\\ [7\cdot104]\end{array}$$

The propagation properties are given by:

$$\left.\begin{aligned}
&\text{Propagation constant} && \beta_{nm} = \sqrt{\omega^2\mu_1\varepsilon_1 - \left(\frac{r'_{nm}}{a}\right)^2} \\
&\text{Critical wavelength} && (\lambda_0)_{nm} = 2\pi\left(\frac{a}{r'_{nm}}\right) \\
&\text{Critical frequency} && (f_0)_{nm} = \frac{v_1}{2\pi}\left(\frac{r'_{nm}}{a}\right) \\
&\text{Wavelength in pipe} && (\lambda_p)_{nm} = \frac{2\pi}{\beta_{nm}} \\
&\text{Free wavelength} && \lambda_1 = 2\pi\frac{v_1}{\omega} \\
&\text{Phase velocity} && (v_p)_{nm} = \frac{\omega}{\beta_{nm}} \\
&\text{Group velocity} && (v_g)_{nm} = \frac{v_1^2}{(v_p)_{nm}} \\
&\text{Free wave velocity} && v_1 = \frac{1}{\sqrt{\mu_1\varepsilon_1}}
\end{aligned}\right\} \quad [7\cdot105]$$

Since there is no root of the type r'_{00}, no wave of the type H_{00} exists. Waves of higher order all exist theoretically, but owing to the extremely high frequencies involved we will consider only the lower-order H_{01} and H_{11} waves. These waves are usually referred to as the H_0 and H_1 waves, respectively.

7·12 The H_0 Wave. Perfect Conductivity

Substituting $n = 0$, $m = 1$ into the general equations 7·104 for the H_{nm} waves, the components E_r and H_ϕ vanish and we obtain

$$\left.\begin{aligned}
E_\phi &= -j\left\{A'\omega\mu_1\frac{a}{r'_{01}}J_1\left(r\frac{r'_{01}}{a}\right)\right\}e^{j(\omega t-\beta_{01}x)} \\
H_x &= \left\{A'J_0\left(r\frac{r'_{01}}{a}\right)\right\}e^{j(\omega t-\beta_{01}x)} \\
H_r &= j\left\{A'\beta_{01}\frac{a}{r'_{01}}J_1\left(r\frac{r'_{01}}{a}\right)\right\}e^{j(\omega t-\beta_{01}x)}
\end{aligned}\right\} \quad [7\cdot106]$$

where $r'_{01} = 3.83$ and $\beta_{01} = \sqrt{\omega^2\mu_1\varepsilon_1 - (3.83/a)^2}$. If the dielectric is air, $\mu_1 = \mu_0$ and $\varepsilon_1 = \varepsilon_0$, very nearly, and $v = 1/\sqrt{\mu_0\varepsilon_0} = c = 3 \times 10^8$

meters per second. Then the propagation properties become

$$\left.\begin{aligned}
\beta_{01} &= \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{3.83}{a}\right)^2} \\
(f_0)_{01} &= \frac{c}{2\pi}\frac{3.83}{a} \simeq 1.829 \times 10^8 \times \frac{1}{a} \\
(\lambda_0)_{01} &= 2\pi\left(\frac{a}{3.83}\right) \simeq 1.640a \\
(\lambda_p)_{01} &= \frac{2\pi}{\sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{3.83}{a}\right)^2}} \\
(v_p)_{01} &= \frac{\omega}{\sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{3.83}{a}\right)^2}} \\
(v_g)_{01} &= \frac{c^2}{(v_p)_{01}}
\end{aligned}\right\} \qquad [7\cdot107]$$

The distribution of the field intensity vectors for this type of wave is shown in Fig. 7·13. Equations defining this distribution may be obtained in a manner similar to that employed in section 7·8, from the relations

$$\left.\begin{aligned}
|H_x| &= A' J_0\left(r\frac{r'_{01}}{a}\right)\cos\beta_{01}x \\
|H_r| &= A'\beta_{01}\left(\frac{a}{r'_{01}}\right) J_1\left(r\frac{r'_{01}}{a}\right)\sin\beta_{01}x \\
|E_\phi| &= A'\omega\mu_1\left(\frac{a}{r'_{01}}\right) J_1\left(r\frac{r'_{01}}{a}\right)\sin\beta_{01}x
\end{aligned}\right\} \qquad [7\cdot108]$$

where $t = 0$.

The characteristic impedance for the H_0 wave, defined on a power basis, has been found to be

$$Z_0 = \frac{2\pi a^2}{r'_{01}} f^2 \sqrt{1 - \left(\frac{(f_0)_{01}}{f}\right)^2} \qquad [7\cdot109]^*$$

when the dielectric is air. On comparison of this equation with that of Z_0 for the H_0 wave (equation 7·91), it is seen that at frequencies large in comparison to the critical frequency the characteristic im-

* J. R. Carson, S. P. Mead, and S. A. Schelkunoff, "Hyper-Frequency Wave Guides — Mathematical Theory," *Bell System Tech. J.*, **15**, 2, p. 310, 1936.

pedance of the H_0 wave is proportional to the square of the frequency, whereas with the E_0 wave it approaches a constant value.

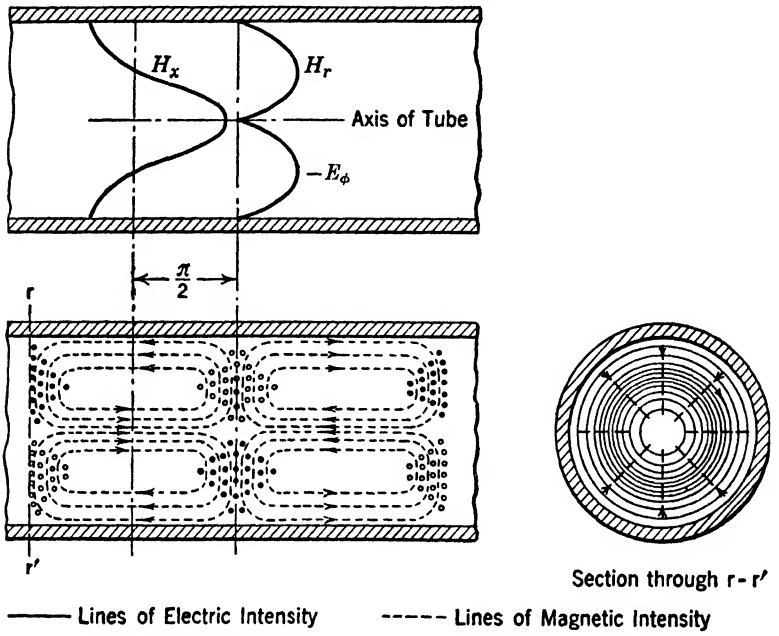


FIG. 7·13 Instantaneous field configuration for H_0 wave (TE_{01}) in a cylindrical wave guide.

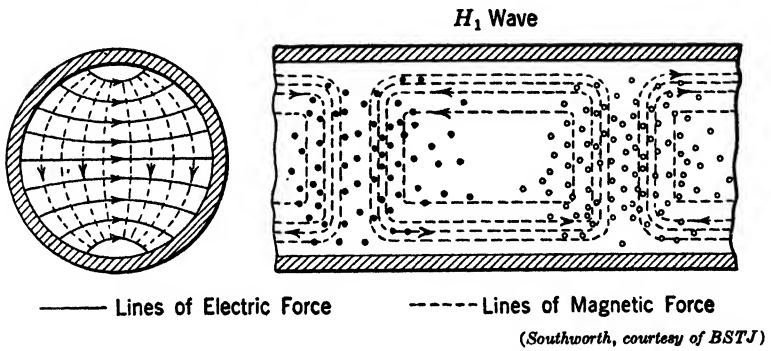


FIG. 7·14 Instantaneous field configuration for H_1 wave (TE_{11}) in a cylindrical wave guide.

7·13 The H_1 Wave. Perfect Conductivity

The equations for the H_1 wave in the dielectric may be written from equations 7·104 by setting $n = 1$, $m = 1$. The wave structure for

this mode of transmission, at any instant of time, is shown in Fig. 7·14. For the higher-order E and H waves, the currents vary as $\cos n\phi$ around the periphery of the inside surface of the conducting tube.

TABLE 7·3

Wave Type	Bessel Function	Root	Cut-off Wavelength meters $= \frac{2\pi a\sqrt{\kappa_e}}{x}$	Cut-off Frequency megacycles $= \frac{xc}{2\pi a\sqrt{\kappa_e}}$
E_0	$J_0(x)$	$x = 2.41$	$2.62\, a\sqrt{\kappa_e}$	$115 \times \frac{1}{a\sqrt{\kappa_e}}$
E_1	$J_1(x)$	$x = 3.83$	$1.64\, a\sqrt{\kappa_e}$	$183 \times \frac{1}{a\sqrt{\kappa_e}}$
H_0	$J_0'(x)$	$x = 3.83$	$1.64\, a\sqrt{\kappa_e}$	$183 \times \frac{1}{a\sqrt{\kappa_e}}$
H_1	$J_1'(x)$	$x = 1.84$	$3.42\, a\sqrt{\kappa_e}$	$88 \times \frac{1}{a\sqrt{\kappa_e}}$

where κ_e is the dielectric constant, and a is the radius of the guide in meters.

The critical frequency for the H_1 wave is given by the relation

$$(f_0)_{11} = \frac{v_1}{2\pi}\left(\frac{r'_{11}}{a}\right) \qquad [7\cdot110]$$

where $r'_{11} = 1.84$. It may be seen by comparing this critical frequency with that of the E_0, E_1, and H_0 waves that this one represents the lowest frequency for a given dielectric material and given tube radius. See Table 7·3. From this point of view it might seem to be the most desirable mode of transmission. It will be seen, however, that when the attenuation is considered this advantage of the H_{11} wave may be of secondary importance.

7·14 Imperfectly Conducting Pipes

The solution to the problem when the walls of the pipe are imperfect conductors may be obtained in a manner similar to that employed in sections 5·18 and 6·18. In these sections it was assumed that the general field structure was not appreciably changed from its form when the boundary surfaces were perfectly conducting. As was pointed out before, this assumption is valid for metals of high conductivity such as copper and aluminum. It is not valid for poor conductors, or for extremely high frequencies of the order of visible light. The degree of validity is indeed dependent upon how well the inequality $\sigma \gg \omega\epsilon$ is

met. Since in practice metals of high conductivity are available, and it is in general desirable to keep the losses as low as possible, the attenuation calculated under the above assumption is entirely satisfactory.

In order that energy flow into the metal, we know from Poynting's theorem that there must be a tangential component of both electric and magnetic intensity at the surface of the metal. In the perfectly conducting case there is a component of magnetic intensity at the surface, but the tangential component of electric intensity is zero there. Since we know physically that the metal can support a small component of E along the surface, we may use the concept of the intrinsic impedance of a metal, η, to determine the value of $E_{\tan}$ at the surface. Thus

$$E_{\tan} = \eta H_{\tan} \qquad [7\cdot111]$$

and the direction of $E_{\tan}$ is chosen in such a way that the computed Poynting vector is directed into the metal. The component $E_{\tan}$ is assumed to be at right angles to the component $H_{\tan}$. The average power absorbed by the tube per square meter of surface is given by the Poynting theorem as

$$\begin{aligned} p_r &= \tfrac{1}{2}\mathcal{R}(\mathbf{E}_{\tan} \times \mathbf{H}^*_{\tan}) = \tfrac{1}{2}\mathcal{R}(\eta\mathbf{H}_{\tan} \times \mathbf{H}^*_{\tan}) \\ &= \tfrac{1}{2}\mathcal{R}\{\eta(H'_{\tan})^2\} \end{aligned} \qquad [7\cdot112]$$

where $H_{\tan} = H'_{\tan}\, e^{j(\omega t - \beta x)}$. In this case of practical conductors, all of the amplitude terms H' and E' contain the attenuation factor $e^{-\alpha x}$.

The average power P_r absorbed by the pipe per unit length is

$$\begin{aligned} P_r &= \tfrac{1}{2}\int\int_{\text{surface of pipe}} p_r\, ds \\ &= \tfrac{1}{2}\int_0^{2\pi} \mathcal{R}\{\eta(H'_{\tan})^2\}\, d\phi \end{aligned} \qquad [7\cdot113]$$

Also the average power transmitted in the X direction through the tube per square meter of surface normal to the axis of the tube is

$$p_x = \tfrac{1}{2}(E'_r H'_\phi - E'_\phi H'_r)$$

The average power transmitted through the tube, P_x, is then

$$P_x = \tfrac{1}{2}\int\int_{\substack{\text{surface normal} \\ \text{to axis of tube}}} p_x\, ds$$

$$P_x = \tfrac{1}{2}\int_{\phi=0}^{\phi=2\pi}\int_{r=0}^{r=a} (E'_r H'_\phi - E'_\phi H'_r) r\, dr\, d\phi \qquad [7\cdot114]$$

Now the power loss in the metal boundary is equal to the rate of diminution of power flowing through the dielectric. Since P_x is the

power flowing through the dielectric $-\partial P_x/\partial x$ is the rate of decrease of this power. See Fig. 7·15. This must equal the power loss P_r in the radial direction into the metal. Thus,

$$P_r = -\frac{\partial P_x}{\partial x}$$

In Chapter 5 it is shown that

$$\frac{\partial P_x}{\partial x} = -2\alpha P_x \qquad [7\cdot115]$$

Hence

$$\alpha = \frac{P_r}{2P_x}$$

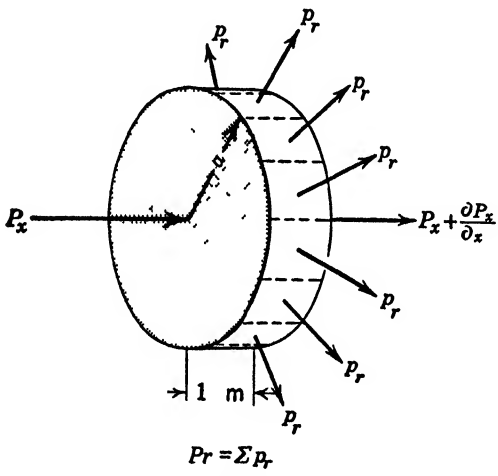


FIG. 7·15

Substituting for P_r and P_x from equations 7·113 and 7·114 we obtain

$$\alpha = \frac{\int_0^{2\pi} \mathcal{R}(\eta H'^2_{\tan})\, d\phi}{2\int_0^{2\pi}\int_0^{a} (E'_r H'_\phi - E'_\phi H'_r) r\, dr\, d\phi} \qquad [7\cdot116]$$

Since the factor $e^{-\alpha x}$ which is contained in the amplitudes E' and H' in the imperfectly conducting case cancels out upon substitution in 7·116, we may substitute directly into this equation from the general equations 7·80 and 7·104.

In the transverse magnetic case (E waves), we obtain upon substitution in 7·116 from 7·80

$$(E \text{ waves}) \quad \alpha_{nm} = \frac{1}{a}\sqrt{\frac{\pi\mu_2 f\varepsilon_1}{\sigma_2\mu_1}}\left[1 - \left(\frac{(f_0)_{nm}}{f}\right)^2\right]^{-\frac{1}{2}} \qquad [7\cdot117]$$

For the transverse electric case (H waves) we obtain, upon substitution in 7·116 from 7·104

$$(H \text{ waves})\alpha_{nm} = \frac{1}{a}\sqrt{\frac{\pi\mu_2 f\varepsilon_1}{\sigma_2\mu_1}}\left[\frac{n}{r_{nm}^2 - n^2} + \left(\frac{(f_0)_{nm}}{f}\right)^2\right]\left[1 - \left(\frac{(f_0)_{nm}}{f}\right)^2\right]^{-\frac{1}{2}} \qquad [7\cdot118]$$

For the E_0, E_1, H_0, and H_1 waves these expressions take the following form

$$E_0 \text{ wave} \quad \alpha_{01} = 0.438a^{-3/2}K\frac{\left(\frac{f}{(f_0)_{01}}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2 - 1}} \Bigg\} \text{ nepers per meter} \qquad [7\cdot119]$$

$$E_1 \text{ wave} \quad \alpha_{11} = 0.552a^{-3/2}K\frac{\left(\frac{f}{(f_0)_{11}}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_{11}}\right)^2 - 1}} \Bigg\} \text{ nepers per meter} \qquad [7\cdot120]$$

$$H_0 \text{ wave} \quad \alpha_{01} = 0.552a^{-3/2}K\frac{\left(\frac{f}{(f_0)_{01}}\right)^{-1/2}}{\sqrt{\left(\frac{f}{(f_0)_{01}}\right)^2 - 1}} \Bigg\} \text{ nepers per meter} \qquad [7\cdot121]$$

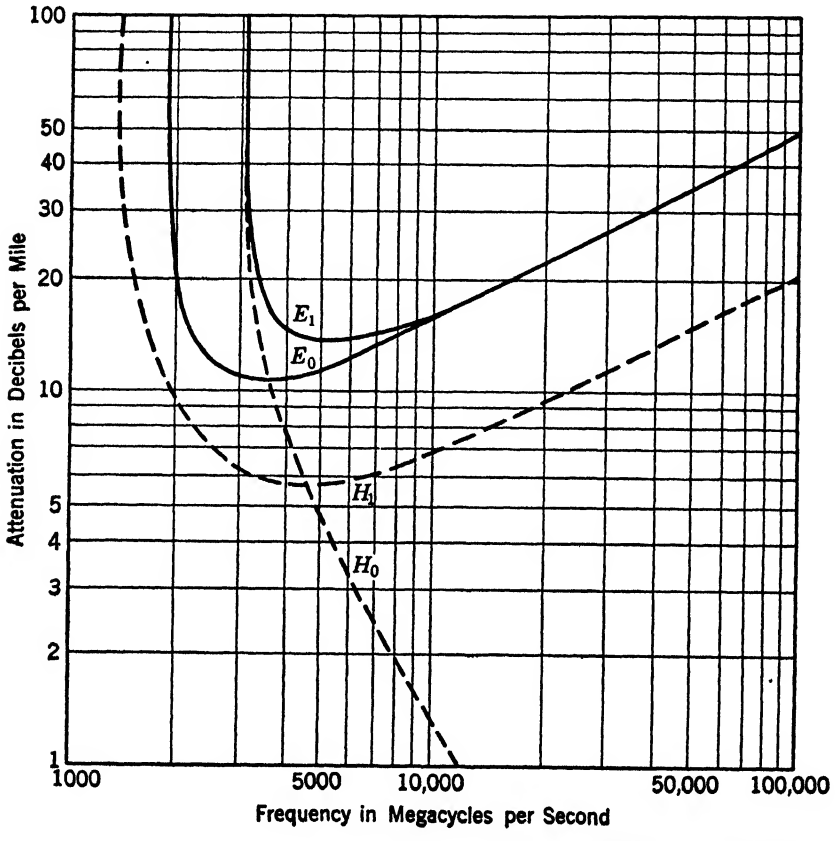
H_1 wave

$$\alpha_{11} = 0.382a^{-3/2}K\frac{\left(\frac{f}{(f_0)_{11}}\right)^{-1/2} + \frac{1}{2.38}\left(\frac{f}{(f_0)_{11}}\right)^{3/2}}{\sqrt{\left(\frac{f}{(f_0)_{11}}\right)^2 - 1}} \Bigg\} \text{ nepers per meter} \qquad [7\cdot122]$$

where $K = \sqrt{\frac{2\pi\mu_2}{v_1\sigma_2\mu_1^2}}$ nepers per meter. Also the critical frequency $(f_0)_{nm}$ in the various expressions refers to the particular wave under consideration.

The attenuation factor $e^{-\alpha x}$ is plotted as a function of frequency in Fig. 7·16 for each of the above waves existing in a 5-inch copper tube

filled with air. As the cut-off frequency is approached, the attenuation, as given by equations 7·119–7·122, increases without limit. Actually, this is not the case, since these equations are not valid for frequencies less than or equal to the critical frequency. The attenuation does, however, increase to a very high value as the frequency approaches



(*Southworth, courtesy of BSTJ*)

FIG. 7·16 Attenuation characteristic of the E_0, H_0, E_1, and H_1 waves in an air-filled circular wave guide 5 inches in diameter.

the critical value. Linder* has shown that the attenuation is over ten thousand times greater than the minimum attenuation for frequencies 1 per cent less than the critical frequency, for waves of the E_{01} type. He has obtained expressions for the propagation constant which

* E. G. Linder, "Attenuation of Electromagnetic Fields in Pipes Smaller Than the Critical Size," *Proc. IRE*, Vol. 30, pp. 554–556, December, 1942.

apply for frequencies equal to, and less than, the critical frequency of the guide.

For each of the E waves, the attenuation reaches a minimum at a frequency equal to $\sqrt{3}$ times the cut-off frequency.* For frequencies greater than the minimum the attenuation rises again and approaches a linear relation with frequency as the frequency continues to increase. In this respect the hollow conductor behaves like ordinary conductors. The minimum for the H_1 wave occurs at a frequency equal to $2.6\sqrt{3}$ times the cut-off frequency. This wave has the lowest cut-off frequency, and over a band width of 4000 megacycles the attenuation is constant within $\frac{1}{2}$ db per mile. Above the critical frequency the attenuation again increases normally, linear with frequency. The properties of the H_0 wave are indeed unique in that the attenuation decreases with increased frequency for all values of frequency. This is the anomaly referred to in connection with hollow-pipe transmission. It means that we can achieve extremely low attenuation by the simple expedient of increasing the frequency. Unfortunately, however, the frequencies required to reduce the attenuation appreciably are extremely high with pipes having small physical dimensions. The stability of the various wave types is discussed later in this chapter.

7·15 Characteristic Impedance of Hollow Pipes

By integrating the Poynting vector over the cross-sectional area of the pipes and dividing by the square of the effective current, Carson, Mead, and Schelkunoff† have obtained the following expression for the characteristic impedance of hollow pipes.

E waves
$$Z_0 = (2\pi a)^2 \sqrt{\frac{\mu_1}{\epsilon_1}} \sqrt{1 - \left(\frac{(f_0)_{nm}}{f}\right)^2} \quad \text{ohms} \qquad [7\cdot123]$$

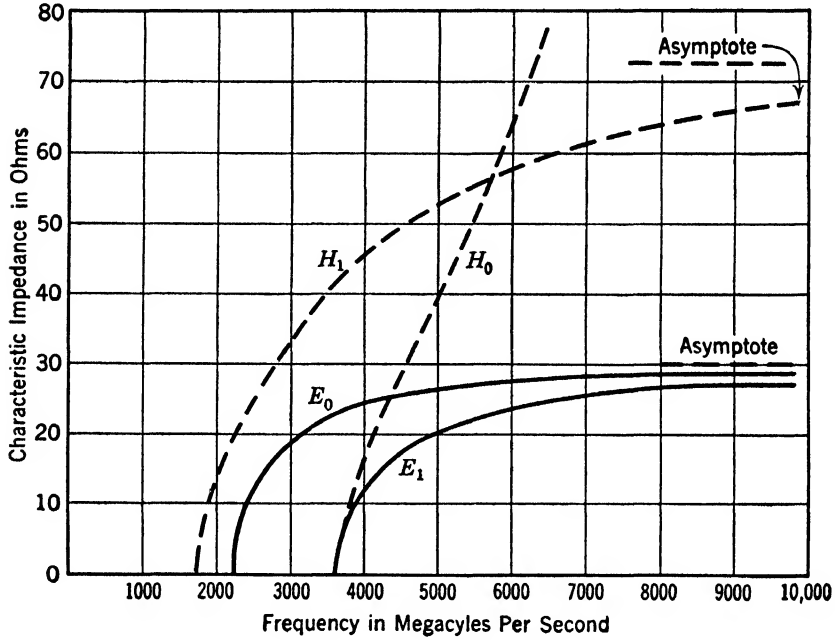
H waves
$$Z_0 = (2\pi a)^4 \left[1 - \frac{n^2}{(r'_{nm})^2}\right] \sqrt{\frac{\mu_1}{\epsilon_1}} \left(\frac{f}{r'_{nm}}\right)^2 \left[1 - \left(\frac{(f_0)_{nm}}{f}\right)^2\right]^{\frac{1}{2}} \quad \text{ohms} \qquad [7\cdot124]$$

These values of characteristic impedance were obtained neglecting attenuation. The critical frequency in each expression refers to the particular wave under consideration.

* This value may be obtained from equations 7·119–7·120 by differentiating and noting the condition for minimum to be $f = \sqrt{3} f_0$.

† J. R. Carson, S. P. Mead, and S. A. Schelkunoff, "Hyper-Frequency Wave Guides — Mathematical Theory," *Bell System Tech. Jour.*, Vol. 15, pp. 310–333, April, 1936

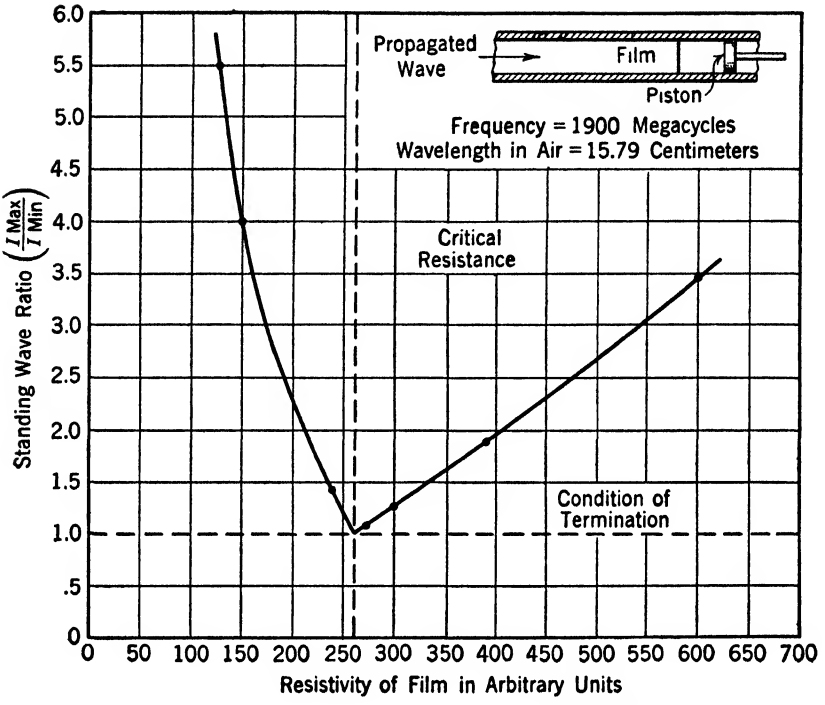
The characteristic impedance for the E_0, E_1, H_0, and H_1 waves is plotted as a function of frequency in Fig. 7·17. These curves were calculated for air-filled copper pipe having a diameter of 4 inches. It is seen that the characteristic impedance of all except the H_0 wave approaches a constant value asymptotically as the frequency is increased. By terminating these pipes with the appropriate characteristic impedance no standing waves result and a maximum of power



(Southworth, courtesy of BSTJ)

FIG. 7·17 Calculated values of the characteristic impedance of a 4-inch air-filled pipe for the E_0, H_0, E_1, and H_1 waves.

transfer takes place. Methods for obtaining the required impedance follow conventional procedures employed with parallel wires. A method suggested by Southworth to avoid standing waves in a line is to place a thin film of resistance material across the tube at right angles to the tube axis, and behind the film a suitable distance, a metal reflector. This reflector may take the form of a movable piston. Other methods employ some form of resistance material in a resonant chamber. Experimental methods for obtaining the required impedance may be easily devised on a trial-and-error basis. A typical set of data obtained by Southworth is shown in Fig. 7·18.



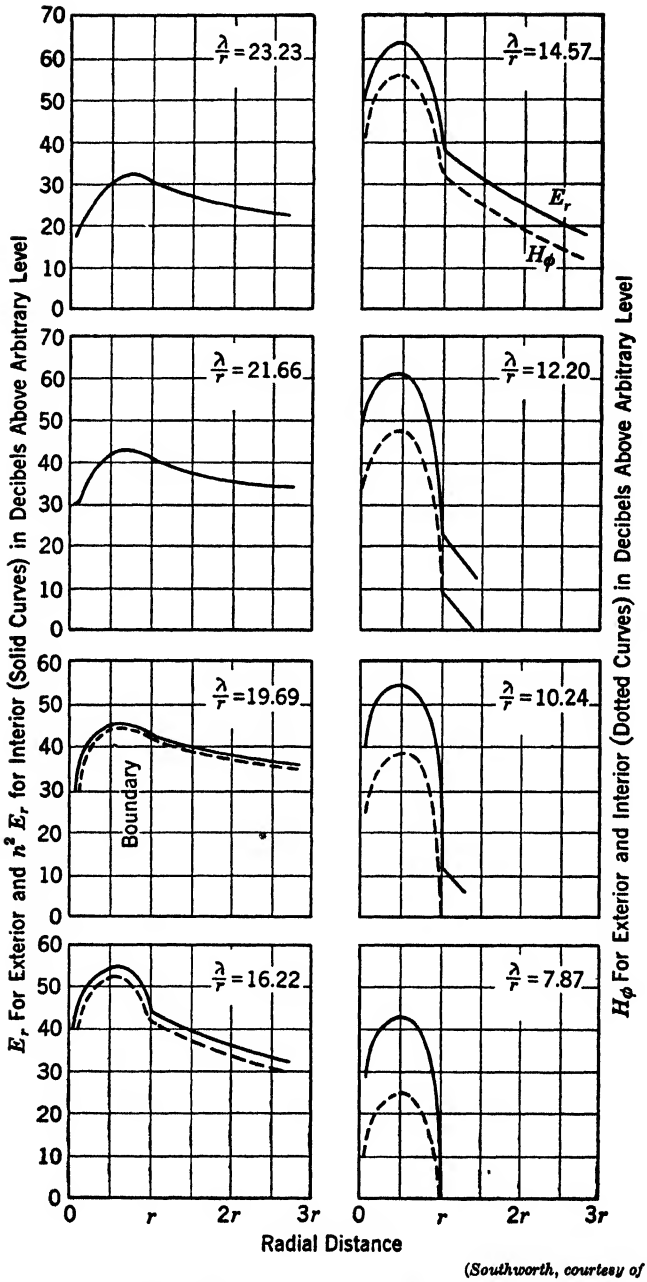
(Southworth, courtesy of BSTJ)

FIG. 7·18 Typical set of experimental data as various degrees of impedance match are obtained.

7·16 Dielectric Wave Guides

If the conducting wall which constitutes the metallic guide is removed from the dielectric material which it contains, the lines of electric and magnetic intensity which were previously confined inside the metal surface will in general extend into the surrounding space. Thus the waves are propagated partly through the dielectric and partly through the surrounding space. Theoretical calculations pertaining to this phenomenon have been made recently by Carson, Mead, and Schelkunoff and by others. Experimental data on this problem have been obtained by Southworth* who has found excellent agreement with theoretical results. Using a cylindrical dielectric wire, having a dielectric constant of 81, he obtained the results shown graphically in Fig. 7·19. In these curves λ is the free-space wavelength.

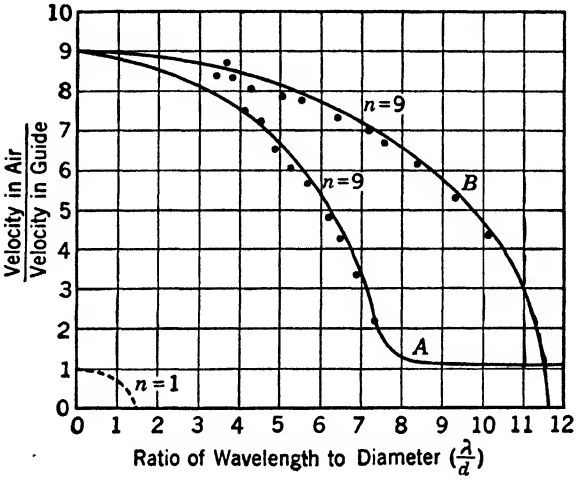
* G. C. Southworth, "Hyper-Frequency Wave Guides — General Considerations and Experimental Results," *Bell System Tech. Jour.*, Vol. 15, pp. 284–309, April, 1936.



(Southworth, courtesy of BSTJ)

FIG. 7·19 Relative intensities of electric and magnetic fields inside and outside a dielectric wire while propagating the E_0 wave. A dielectric constant of 81 is assumed. The index of refraction $n=\sqrt{\epsilon}$

With materials having a large dielectric constant and using frequencies well above cut-off frequency, it is seen that the power is propagated largely within the guide. Conversely, with low dielectric materials, and with frequencies just above critical, it is found that a large portion of the power travels outside the guide. The relative intensities of the electric and magnetic vectors inside and outside the guides are indicated in the figure. It is seen in the case of the highest frequencies that the guide substantially shields the energy which is propagated. The critical frequency as applied to dielectric guides indicates the



(Southworth, courtesy of BSTJ)

FIG. 7·20. Curve *A*. Velocity ratio for the E_0 wave in a dielectric wire having a dielectric constant $\kappa_e = 81$. Curve *B*. Velocity ratio for the E_0 wave in a metal pipe, the interior of which is filled with the same dielectric material ($\kappa_e = 81$).

It is seen that as λ approaches zero the velocity ratio approaches $\sqrt{81} = 9$. If the dielectric constant of the material were progressively reduced curve *B* would shrink to the dotted curve shown in the lower left corner.

frequency at which the most rapid transition takes place between propagation inside and outside the guide.

A comparison of the velocity of the E_0 wave in the dielectric guide with its velocity when a metal sheath surrounds the dielectric is shown in Fig. 7·20. As the wavelength approaches zero, i.e., as the frequency becomes infinite, it is seen that the velocity of the wave in either case approaches the free wave velocity in a medium having the same dielectric constant as that of the material used. When the metal sheath surrounds the dielectric, the velocity of the wave approaches infinity at $\lambda/d = 11.6$. When the sheath is removed, however, the

velocity in the guide approaches the velocity in free space as the wavelength is increased.

7·17 Coaxial Lines

From the preceding sections it has been found that, for hollow-tube transmission systems, wavelengths of the order of the tube cross-sectional dimensions are required in order that propagation take place. In Table 7·3, page 262, are shown the minimum or cut-off wavelengths associated with several types of waves in circular pipes.

Since the hollow-tube transmission systems require such extremely high frequencies their practical applications are quite limited. The frequency limitations of the hollow-tube transmission systems are attributed to the lack of a principal mode of transmission. This mode, which is characterized by the *TEM* transmission mode, is not possible with the hollow tube, since, as has been shown, no wave can exist with both longitudinal components zero. The plane parallel plates permit the existence of a principal mode, and hence all frequencies are transmitted through the dielectric between them. The principal mode may also exist in coaxial cylindrical conductors. In fact, by extending the respective radii of the coaxial system to infinity it may be shown that the plane parallel plates are but a special case of coaxial conductors.

Wave guides of the hollow-tube and coaxial types have a common advantage in that the outside conductor serves both as a conductor and a shield. At the lower frequencies used in telephone work, the field may penetrate the outside conductor and cause interference in other cables and wires in the neighborhood of the cable. Such interference, known as cross-talk, represents a very important problem in low-frequency communications. At higher frequencies the field penetration into the conductor is reduced and no appreciable field reaches the outside. The cable wall is then said to be electrically thick.

Losses in coaxial conductors tend to be larger than the losses in hollow wave guides. This may be attributed to skin effect in the central conductor and to losses in the insulating supports provided for it.

We shall be concerned here with frequencies greater than a megacycle, in which case the penetration of the field into the metal is quite small. See Table 5·1. Under these circumstances we may consider the outside conductor of the guide to have infinite thickness. By the same reasoning, the radius of the inner conductor is also electrically thick. In many types of coaxial lines now in use the inner conductor is supported at intervals along the line, whereas in others a dielectric material completely fills the space between the two conductors.

We shall consider the latter case, with the radius of the inner conductor a and the inner radius of the outer conductor b. Where necessary we shall distinguish between the properties of the two media with the subscripts 1 and 2. See Fig. 7·21.

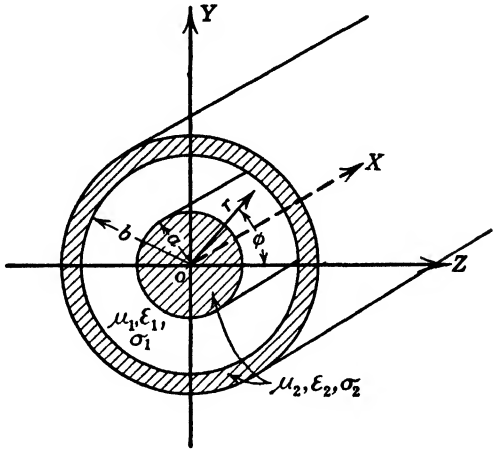


FIG. 7·21 Choice of axes in a coaxial line.

Where high-frequency waves are to be transmitted it may be desirable to exclude low-frequency components. The hollow wave guide is a logical choice since its filtering action removes any undesirable components. In other applications the coaxial transmission line excited in its principal mode is usually a more compact and satisfactory structure. Moreover, a small coaxial structure provides low losses at frequencies for which hollow guides are physically prohibitively large. Inspection of the field equations as modified by the appropriate boundary conditions shows that a variety of transmission modes is possible in the coaxial structure.

In addition to the principal mode of transmission many other modes are theoretically possible in coaxial lines. The critical wavelength required with the higher-order modes is of the order of the difference in radii, $b - a$. Thus the frequency requirements for transmission are at least higher than those of a hollow guide having the same outside radius. Since the frequency requirements for exciting the higher-order modes are in general less favorable and the losses higher, we shall limit our discussion entirely to circularly symmetrical transverse magnetic fields. This will result in considerable simplification of the analysis. We shall assume here, as before, that the media with which we deal are homogeneous and isotropic.

When the electromagnetic field is circularly symmetrical, that is, when the components of electric and magnetic intensity are independent of ϕ, all partial derivatives with respect to ϕ vanish. Under these conditions the field equations 7·18 and 7·19 reduce to

$$\left.\begin{aligned} \gamma H'_\phi &= (\sigma + j\omega\varepsilon)E'_r \quad (a) \\ -\gamma H'_r - \frac{\partial H'_x}{\partial r} &= (\sigma + j\omega\varepsilon)E'_\phi \quad (b) \\ \frac{1}{r}\frac{\partial}{\partial r}(rH'_\phi) &= (\sigma + j\omega\varepsilon)E'_x \quad (c) \end{aligned}\right\} \text{[7·125]} \qquad \left.\begin{aligned} \gamma E'_\phi &= -j\omega\mu H'_r \quad (a) \\ \gamma E'_r + \frac{\partial E'_x}{\partial r} &= j\omega\mu H'_\phi \quad (b) \\ \frac{1}{r}\frac{\partial}{\partial r}(rE'_\phi) &= -j\omega\mu H'_x \quad (c) \end{aligned}\right\} \text{[7·126]}$$

These equations define two separate transmission modes. If we choose H entirely in the ϕ direction we obtain the circular magnetic or transverse magnetic transmission mode. Waves satisfying this mode are also called E waves. Setting $H_r = H_x = 0$ we obtain

$$\left.\begin{aligned} \gamma H'_\phi &= (\sigma + j\omega\varepsilon)\, E'_r \quad (a) \\ 0 &= (\sigma + j\omega\varepsilon)\, E'_\phi \quad (b) \\ \frac{1}{r}\frac{\partial}{\partial r}(rH'_\phi) &= (\sigma + j\omega\varepsilon)\, E'_x \quad (c) \end{aligned}\right\} \text{[7·127]} \qquad \left.\begin{aligned} \gamma E'_\phi &= 0 \quad (a) \\ \gamma E'_r + \frac{\partial E'_x}{\partial r} &= j\omega\mu\, H'_\phi \quad (b) \\ \frac{1}{r}\frac{\partial}{\partial r}(rE'_\phi) &= 0 \quad (c) \end{aligned}\right\} \begin{matrix} E \text{ waves} \\ \text{in the} \\ \text{dielectric} \\ \text{[7·128]} \end{matrix}$$

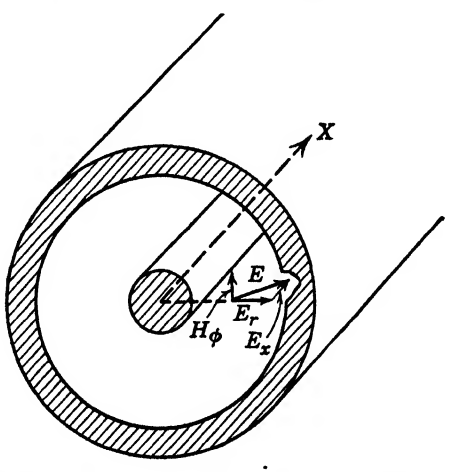


FIG. 7·22 Circular magnetic or transverse magnetic transmission mode in a coaxial guide. E waves.

It is evident from these relations that $E'_\phi = 0$ in this type of transmission. Figure 7·22 shows the orientation of the TM or E wave field vectors in the dielectric.

If we choose E entirely in the ϕ direction, the circular electric or

transverse electric transmission mode results. Waves satisfying this mode are also called H waves. Substituting $E_r = E_x = 0$ into equations 7·125 and 7·126 we obtain

$$\left.\begin{aligned} \gamma H'_\phi &= 0 && (a)\\ -\gamma H'_r - \frac{\partial H'_x}{\partial r} &= (\sigma + j\omega\varepsilon)\, E'_\phi && (b)\\ \frac{1}{r}\frac{\partial}{\partial r}(rH'_\phi) &= 0 && (c) \end{aligned}\right\} \quad [7\cdot129]$$

$$\left.\begin{aligned} \gamma E'_\phi &= -j\omega\mu H'_r && (a)\\ 0 &= -j\omega\mu H'_\phi && (b)\\ \frac{1}{r}\frac{\partial}{\partial r}(rE'_\phi) &= -j\omega\mu H'_x && (c) \end{aligned}\right\} \quad \begin{matrix} H \text{ waves} \\ \text{in the} \\ \text{dielectric} \\ [7\cdot130] \end{matrix}$$

Hence $H'_\phi = 0$. The orientation of the field vectors in the dielectric in this type of transmission are shown in Fig. 7·23.

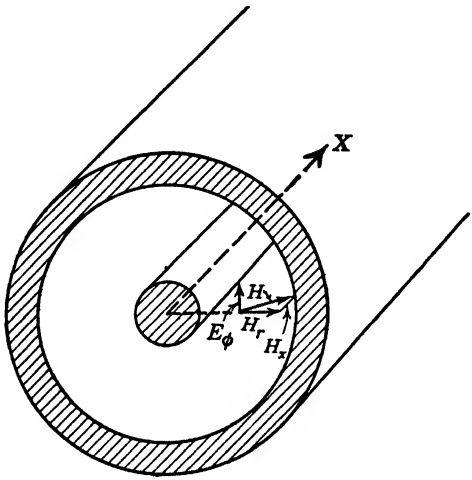


FIG. 7·23 Circular electric or transverse electric transmission mode in a coaxial wave guide. H waves.

Equations 7·127, 7·128, 7·129, and 7·130 for the two transmission modes are expressed in terms of the maximum values of the E and H vectors in distance and time. With the H waves the lines of electric intensity form a system of concentric circles about the X axis in the dielectric. This is associated with uniform current flow in the ϕ direction on the surface of the perfect conductors. With the E waves the lines of magnetic intensity form concentric circles about the X axis in the dielectric. The associated current flow, therefore, is in the longitudinal direction on the conductor surface. It is thus seen that the transverse magnetic mode of transmission corresponds to the conditions under which coaxial transmission lines are usually excited in practice. To obtain a solution to the differential equations describing this mode,

it is necessary to obtain an equation in one of the vector components alone. To do this, let us first obtain an equation relating H'_ϕ and E'_r by substituting E'_r from 7·127*a* into 7·128*b*. This gives

$$\frac{\gamma^2}{\sigma + j\omega\varepsilon} H'_\phi + \frac{\partial E'_x}{\partial r} = j\omega\mu H'_\phi \qquad [7\cdot131]$$

or

$$H'_\phi = \frac{\sigma + j\omega\varepsilon}{j\omega\mu(\sigma + j\omega\varepsilon) - \gamma^2} \frac{\partial E'_x}{\partial r} \qquad [7\cdot132]$$

Differentiating partially with respect to r we have

$$\frac{\partial H'_\phi}{\partial r} = \frac{\sigma + j\omega\varepsilon}{j\omega\mu(\sigma + j\omega\varepsilon) - \gamma^2} \frac{\partial^2 E'_x}{\partial r^2} \qquad [7\cdot133]$$

But from 7·127*c*

$$\frac{1}{r} \frac{\partial(rH'_\phi)}{\partial r} = \frac{1}{r}\left(r \frac{\partial H'_\phi}{\partial r} + H'_\phi\right) = (\sigma + j\omega\varepsilon) E'_x \qquad [7\cdot134]$$

Substituting for H'_ϕ from 7·132 and for $\partial H_\phi/\partial r$ from 7·133 into 7·134, we have

$$\frac{\partial^2 E'_x}{\partial r^2} + \frac{1}{r} \frac{\partial E'_x}{\partial r} + (\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)) E'_x = 0 \qquad [7\cdot135]$$

We may have obtained this equation directly from equation 7·24, setting $\partial/\partial\phi = 0$. This is a form of Bessel's equation of zero order. Its solution is

$$E'_x = AJ_0(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)}) + BY_0(r\sqrt{\gamma^2 - j\omega\mu(\sigma + j\omega\varepsilon)})$$

where J_0 and Y_0 are Bessel functions of zero order of the first and second kinds, respectively, and A and B are constants to be determined. If in the dielectric $\sigma \ll j\omega\varepsilon$ we may write

$$E'_x = AJ_0(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) + BY_0(r\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) \qquad [7\cdot136]$$

Since $E_x = 0$ at the surface of the perfectly conducting cylinders it is necessary that

$$AJ_0(a\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) + BY_0(a\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) = 0 \qquad [7\cdot137]$$

and

$$AJ_0(b\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) + BY_0(b\sqrt{\gamma^2 + \omega^2\mu\varepsilon}) = 0 \qquad [7\cdot138]$$

Solving these equations for A/B we obtain

$$-\frac{A}{B} = \frac{Y_0(a\sqrt{\gamma^2 + \omega^2\mu\varepsilon})}{J_0(a\sqrt{\gamma^2 + \omega^2\mu\varepsilon})} = \frac{Y_0(b\sqrt{\gamma^2 + \omega^2\mu\varepsilon})}{J_0(b\sqrt{\gamma^2 + \omega^2\mu\varepsilon})} \qquad [7\cdot139]$$

This equation has an infinite number of roots, the approximate values of which may be obtained if we replace the Bessel functions with their approximate values for large values of the argument. Making this substitution, we obtain

$$\sin\{(b-a)\sqrt{\gamma^2+\omega^2\mu\varepsilon}\} = 0 \qquad [7\cdot140]^*$$

This is true if

$$(b-a)\sqrt{\gamma^2+\omega^2\mu\varepsilon} = n\pi \qquad n = 1, 2, 3, \cdots \qquad [7\cdot141]$$

giving

$$\sqrt{\gamma^2+\omega^2\mu\varepsilon} = \frac{n\pi}{b-a} \qquad [7\cdot142]$$

This is an excellent approximation for all roots of equation 7·139 if $b < 3a$. Also, the larger the value of n, the better the approximation.

Squaring both sides and solving for γ we have

$$\gamma_n = \sqrt{\left(\frac{n\pi}{b-a}\right)^2 - \omega^2\mu\varepsilon} \qquad [7\cdot143]$$

If $\omega^2\mu\varepsilon > \left(\frac{n\pi}{b-a}\right)^2$, γ_n is imaginary, and the wave is propagated without

* This relation is obtained as follows: The equation for J_p and Y_p for large values of the argument are given in Table 7·1, page 242. Setting $p = 0$ we have

$$J_0(\rho) = \sqrt{\frac{2}{\pi\rho}}\cos\left(\rho - \frac{\pi}{4}\right) \qquad \text{and} \qquad Y_0(\rho) = \sqrt{\frac{2}{\pi\rho}}\sin\left(\rho - \frac{\pi}{4}\right)$$

Hence

$$\frac{\sin\left(a\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)}{\cos\left(a\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)} = \frac{\sin\left(b\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)}{\cos\left(b\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)}$$

Then

$$\sin\left(b\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)\cos\left(a\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right) -$$
$$\sin\left(a\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right)\cos\left(b\sqrt{\gamma^2+\omega^2\mu\varepsilon} - \frac{\pi}{4}\right) = 0$$

But

$$\sin A\cos B - \sin B\cos A = \sin(A-B)$$

Hence

$$\sin\{(b-a)\sqrt{\gamma^2+\omega^2\mu\varepsilon}\} = 0$$

attenuation. Thus

$$\gamma_n = \alpha_n + j\beta_n = j\sqrt{\omega^2\mu\varepsilon - \left(\frac{n\pi}{b-a}\right)^2} \qquad [7\cdot144]$$

giving $\alpha_n = 0$ and

$$\beta_n = \sqrt{\omega^2\mu\varepsilon - \left(\frac{n\pi}{b-a}\right)^2} \qquad [7\cdot145]$$

But if $\left(\frac{n\pi}{b-a}\right)^2 > \omega^2\mu\varepsilon$, γ_n is real and there is no propagation at all. We may define a critical frequency $(f_0)_n$ as in the hollow-tube system. Setting $\beta_n = 0$

$$(f_0)_n = \frac{n}{2(b-a)\sqrt{\mu\varepsilon}} \qquad [7\cdot146]$$

and the critical wavelength is

$$(\lambda_0)_n = \frac{2}{n}(b-a) \qquad [7\cdot147]$$

Since in most coaxial conductors $b - a$ is less than a centimeter, the higher-order E_n waves require such enormous frequencies that they represent impractical modes of transmission. These remarks are equally applicable to H_n waves in coaxial lines.

There is another solution of equations 7·137 and 7·138 which is very important, but which appears at first glance to be trivial. This solution, which corresponds to the principal mode of transmission, is obtained by setting A and B equal to zero. From equation 7·136 we see that this requires that E_x vanish identically. Now, when $\sigma \ll \omega\varepsilon$, equation 7·132 becomes

$$H'_\phi = -\frac{j\omega\varepsilon}{\omega^2\mu\varepsilon + \gamma^2}\frac{\partial E'_x}{\partial r} \qquad [7\cdot148]$$

and replacing H'_ϕ in 7·127a we have

$$E'_r = -\frac{\gamma}{\omega^2\mu\varepsilon + \gamma^2}\frac{\partial E'_x}{\partial r} \qquad [7\cdot149]$$

Hence H'_ϕ and E'_r also vanish if E_x is zero unless

$$\gamma^2 + \omega^2\mu\varepsilon = 0 \qquad [7\cdot150]$$

giving

$$\gamma = \alpha + j\beta = j\omega\sqrt{\mu\varepsilon} \qquad [7\cdot151]$$

Hence

$$\alpha = 0, \qquad \beta = \omega\sqrt{\mu\varepsilon} \tag{7·152}$$

This value of β indicates that propagation in the principal mode has no restrictions in regard to frequency. Since $E'_x = 0$, equation 7·126*b* becomes

$$\gamma E'_r = j\omega\mu H'_\phi \tag{7·153}$$

Replacing γ from 7·151,

$$E'_r = \sqrt{\frac{\mu}{\varepsilon}} H'_\phi \tag{7·154}$$

Also equation 7·127*c* becomes

$$\frac{\partial(rH'_\phi)}{\partial r} = 0 \tag{7·155}$$

giving $rH'_\phi = 0$ or a constant, C, independent of r. Taking it as C,

$$H'_\phi = \frac{C}{r} \tag{7·156}$$

We may evaluate the constant C by Ampère's law. Let us take the line integral of H'_ϕ around the surface of the inner conductor. This is $2\pi a H'_\phi$ and is equal to the longitudinal current I carried by the inner conductor. Hence

$$I' = 2\pi a H'_\phi \tag{7·157}$$

But, at $r = a$, equation 7·156 gives

$$C = aH'_\phi \tag{7·158}$$

Hence

$$I' = 2\pi C \qquad \text{or} \qquad C = \frac{I'}{2\pi} \tag{7·159}$$

and from 7·156 the equation for the magnetic intensity may be written

$$H_\phi = \frac{I'}{2\pi r} e^{j(\omega t - \beta x)} = \frac{I}{2\pi r} \tag{7·160}$$

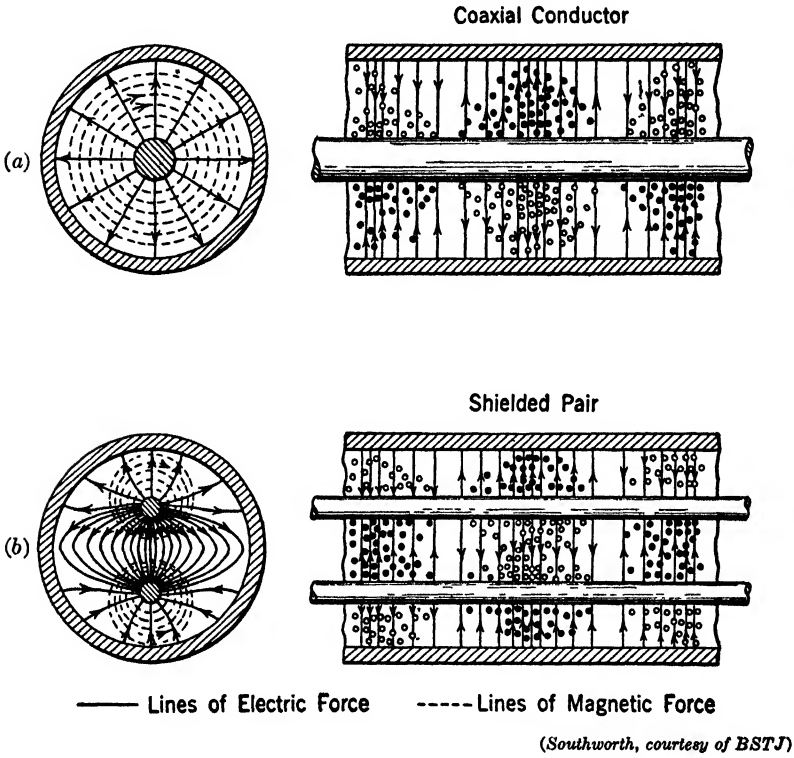
and from 7·154 the electric intensity becomes

$$E_r = \frac{I'}{2\pi r}\sqrt{\frac{\mu}{\varepsilon}}\, e^{j(\omega t - \beta x)} = \frac{I}{2\pi r}\sqrt{\frac{\mu}{\varepsilon}} \tag{7·161}$$

The field distribution for the principal mode defined by equations 7·160 and 7·161 is shown in Fig. 7·24*a*. Since E_x and H_x are zero, this mode is sometimes referred to as the transverse electromagnetic (*TEM*) mode.

It will be noted that the wave configuration of Fig. 7·24*a* is similar to that shown in Fig. 7·10 for the E_0 wave in a hollow circular tube.

For purposes of comparison the field distribution for a pair of conductors enclosed in a circular shield is shown in Fig. 7·24*b*. The mathematical theory for this configuration of conductors will not be considered



(*Southworth, courtesy of BSTJ*)

FIG. 7·24 Instantaneous field configuration of the lines of electric and magnetic field intensity in a coaxial conductor and also in a shielded pair of conductors. Note similarity to E_0 and E_1 waves of Fig. 7·10 and 7·12.

here. It may be seen, however, that this field structure is very similar to that of the E_1 wave shown in Fig. 7·12. The propagation properties of these waves and of the hollow-tube waves are quite different. In particular, there is no low-frequency cut-off for the modes shown in Fig. 7·24.

We may calculate the potential difference between the inner and outer cylinders of the coaxial transmission line by taking the line integral of E_r along a radius vector.

$$V = \int_a^b E_r \, dr = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\int_a^b I\frac{dr}{r} \qquad [7\cdot162]$$

Hence

$$V = \frac{I}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\log_e\left(\frac{b}{a}\right) \qquad [7\cdot163]$$

The characteristic impedance of the coaxial cylinders may then be expressed as

$$Z_0 = \frac{V}{I} = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\log_e\frac{b}{a} \text{ ohms} \qquad [7\cdot164]$$

which may be written

$$Z_0 = \frac{1}{2\pi}K\log_e\frac{b}{a} \text{ ohms} \qquad [7\cdot165]$$

where $K = \sqrt{\mu/\varepsilon}$ is the characteristic impedance of an infinite medium of dielectric constant ε and permeability μ.

Let us differentiate equation 7·163 with respect to x. Then

$$\frac{\partial V}{\partial x} = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\log_e\left(\frac{b}{a}\right)\frac{\partial I}{\partial x} \qquad [7\cdot166]$$

Now $I = I'e^{j(\omega t-\beta x)}$; hence

$$\frac{\partial I}{\partial x} = -j\beta I'e^{j(\omega t-\beta x)} = -j\beta I$$

giving, when β is replaced by $\omega\sqrt{\mu\varepsilon}$ from equation 7·152

$$\frac{\partial V}{\partial x} = -j\frac{\omega\mu}{2\pi}\log_e\left(\frac{b}{a}\right)I \qquad [7\cdot167]$$

It is shown in Chapter 9 that, when we have a line with distributed parameters, the equations governing the transverse voltage V and current I along the line are

$$\frac{\partial V}{\partial x} = -ZI \qquad [7\cdot168]$$

and

$$\frac{\partial I}{\partial x} = -YV \qquad [7\cdot169]$$

where $Z = R + j\omega L$ and $Y = G + j\omega C$. L and R are the series inductance and resistance, respectively, and G and C are the shunt conductance and capacitance, respectively. The values of R, L, G, and C in these relations are expressed per unit length of line.

The telegrapher's differential equation is obtained by eliminating I

from equations 7·168 and 7·169; it is

$$\frac{\partial^2 V}{\partial x^2} = ZYV \qquad [7\cdot170]$$

The solution of this equation is of the form $e^{\sqrt{ZY}}$, as substitution will show. Thus the propagation constant is

$$\gamma = \sqrt{ZY} \qquad [7\cdot171]$$

On comparing equations 7·167 and 7·168 we see that

$$Z = R + j\omega L = j\omega \frac{\mu}{2\pi} \log_e \frac{b}{a} \qquad [7\cdot172]$$

Since with perfect conductivity $R = 0$, we have

$$L = \frac{\mu}{2\pi} \log_e \frac{b}{a} \quad \text{henrys per meter} \qquad [7\cdot173]$$

giving the series inductance per unit length of coaxial line.

We may calculate the admittance Y with the aid of 7·171. The propagation constant γ, when σ is not neglected, is

$$\gamma = \sqrt{j\omega\mu\,(\sigma + j\omega\varepsilon)}$$

Setting this equal to γ in 7·171 we obtain

$$Y = \frac{1}{Z}[j\omega\mu(\sigma + j\omega\varepsilon)]$$

$$Y = \frac{2\pi}{\log_e \frac{b}{a}}[\sigma + j\omega\varepsilon] = G + j\omega C \qquad [7\cdot174]$$

Thus the shunt capacity and conductance are

$$G = \frac{2\pi\sigma}{\log_e \frac{b}{a}} \text{ mhos/meter} \quad \text{and} \quad C = \frac{2\pi\varepsilon}{\log_e \frac{b}{a}} \text{ farads/meter} \qquad [7\cdot174a]$$

7·18 Coaxial Lines — Finite Conductivity

The average power P_x flowing through the dielectric between the cylinders may be calculated from the Poynting theorem as

$$P_x = \int_0^{2\pi} \int_a^b \tfrac{1}{2}\mathcal{R}\,[E_r' H_\phi']r\,d\phi\,dr \qquad [7\cdot175]$$

Substituting for E_r' and H_ϕ' from equations 7·160 and 7·161

$$= \frac{1}{2}\left(\frac{I'}{2\pi}\right)^2 \sqrt{\frac{\mu_1}{\mathfrak{e}_1}}\, 2\pi \int_a^b \frac{dr}{r} \qquad [7\cdot176]$$

Thus

$$P_x = \tfrac{1}{2}V'I' = \frac{I'^2}{4\pi}\sqrt{\frac{\mu_1}{\varepsilon_1}}\log_e\frac{b}{a} \qquad [7\cdot177]$$

Where the conductivity is not infinite we may calculate the average power loss into the outer and inner cylinders by the method employed previously. Thus the average power absorbed by the inner conductor per unit length is

$$P_a = \int_0^{2\pi} \tfrac{1}{2}\mathcal{R}\,[\eta H'^2_{\tan}]a\,d\phi \qquad [7\cdot178]$$

Substituting for $H'_{\tan} = H'_\phi$ from equation 7·160 at $r = a$, we have

$$P_a = \frac{1}{2}\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\int_0^{2\pi}\frac{(I')^2}{4\pi^2}\frac{d\phi}{a} = \left(\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\right)\frac{(I')^2}{4\pi a} \qquad [7\cdot179]$$

But

$$P_a = \tfrac{1}{2}(I')^2R_a \qquad [7\cdot180]$$

when R_a is the resistance per unit length of the inner conductor. Hence

$$R_a = \frac{2P_a}{(I')^2} = \frac{1}{2\pi a}\sqrt{\frac{\pi f\mu_2}{\sigma_2}} \qquad [7\cdot181]$$

In a similar manner we may show that the resistance of the outer conductor R_b per unit length is

$$R_b = \frac{1}{2\pi b}\sqrt{\frac{\pi f\mu_2}{\sigma_2}} \qquad [7\cdot182]$$

The total resistance per unit length, i.e., the loop unit resistance, $R = R_a + R_b$. Thus

$$R = \frac{1}{2\pi}\left(\frac{1}{a}+\frac{1}{b}\right)\sqrt{\frac{\pi f\mu_2}{\sigma_2}} \qquad [7\cdot183]$$

The total average power lost in the inner and outer cylinders per unit length is thus

$$P_r = \tfrac{1}{2}R(I')^2 \qquad [7\cdot184]$$

$$= \frac{(I')^2}{4\pi}\left(\frac{1}{a}+\frac{1}{b}\right)\sqrt{\frac{\pi f\mu_2}{\sigma_2}} \qquad [7\cdot185]$$

As in equation 7·115 we may write

$$P_r = -\frac{\partial P_x}{\partial x} = 2\alpha P_x \qquad [7\cdot186]$$

Hence

$$\alpha = \frac{P_r}{2P_x} \qquad [7\cdot187]$$

Substituting for P_x and P_r we obtain for the attenuation constant

$$\alpha = \frac{\sqrt{\frac{\pi f \mu_2}{\sigma_2}}\left(\frac{1}{a}+\frac{1}{b}\right)}{2\sqrt{\frac{\mu_1}{\varepsilon_1}}\log_e\frac{b}{a}} = \frac{1}{2}\sqrt{\frac{\pi f \mu_2 \varepsilon_1}{\sigma_2 \mu_1}}\left(\frac{1}{a}+\frac{1}{b}\right)\frac{1}{\log_e\frac{b}{a}} \text{ nepers per meter} \qquad [7\cdot188]$$

It is possible to determine, for a given frequency, the optimum ratio for minimum attenuation by differentiating equation 7·188 with respect to b/a and setting this differential equal to zero. This gives for the optimum ratio

$$\frac{b}{a} = 3.6 \qquad [7\cdot189]$$

This is considered in more detail in Chapter 9.

7·19 Elliptic Wave Guides and Stability

The problem of the propagation of electromagnetic waves in hollow metal pipes of elliptic cross section has been investigated theoretically by Chu.* Since some deformation of the pipes of circular cross section into a slightly elliptic form is almost unavoidable, this problem has important practical applications. In his investigation Chu obtained significant information relative to the stability of waves in pipes of circular cross section. The mathematical argument associated with this problem will not be given here, although the method will be discussed briefly.

The procedure followed by Chu is similar to that employed in the preceding sections for the circular pipe except that it is far more complicated. The use of the elliptic coordinate system is necessary, and the first step toward the solution is to set up Maxwell's equations in this system. The general equations for the E and H waves are obtained as before and are expressed in terms of Mathieu† functions. Since the

* Lan Jen Chu, "Electromagnetic Waves in Hollow Elliptic Pipes of Metal," *J. Applied Phys.*, Vol. 9, No. 9, September, 1938.

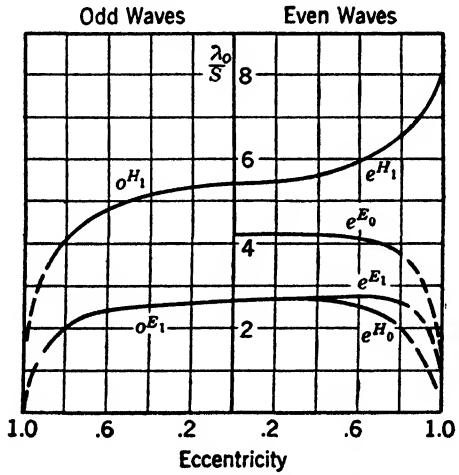
† Mathieu functions are the solutions to Mathieu equations, just as Bessel functions are solutions to Bessel's equations. Mathieu equations arise in this case from the expression of the wave equation in elliptic coordinates. For discussion of these functions see E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 1927, S. Goldstein, *Trans. Camb. Phil. Soc.*, Vol. 23, No. 11, 1927.

E and H waves in the elliptic pipe are described in both even and odd Mathieu functions, the prefix e or o is used. Thus eH or eE indicates that the wave is described by even Mathieu functions; oH or oE indicates that the wave is described by odd functions. Also in elliptic pipes it is convenient to use the double subscript n, m, where n refers to the order of the Mathieu function and m denotes the rank of the root. Since the present discussion is entirely limited to the first root, $m = 1$, this distinguishing subscript may be omitted. Only the six lowest-order waves were studied by Chu. These are the eH_0, eH_1, oH_1, eE_0, eE_1, and oE_1 waves.

The waves which we have studied in circular pipes are closely related to those in the elliptic pipes. When the distance between the foci approaches zero, the elliptic cylinder degenerates to a circular cylinder. Thus the Mathieu functions degenerate into circular functions. Elliptic and circular waves of similar types are compared in Table 7·4.

TABLE 7·4

Waves in elliptic pipe	eE_0	eE_n oE_n	eH_0	eH_n oH_n
Corresponding waves in circular pipe	E_0	E_n	H_0	H_n

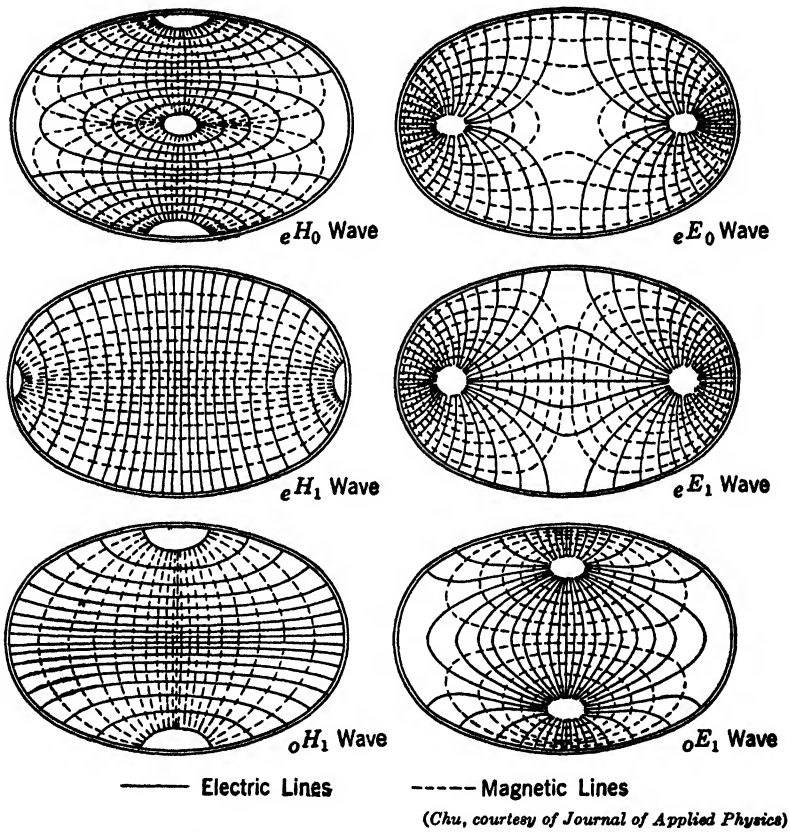


(Chu, courtesy of Journal of Applied Physics)

FIG. 7·25 Critical wavelengths of waves in elliptic pipes.

In Fig. 7·25 the ratio of the critical wavelength λ_0 to the periphery S of the elliptic pipe is plotted as a function of the eccentricity of the pipe. Curves on the left side of this figure are for odd waves; curves on the right side, for even waves. When the eccentricity is zero, corresponding to a circle, the critical wavelengths indicated in the figure

correspond to those for the circular wave guides. As a tube of elliptic cross section is deformed along its major axis into a circle, and again elongated into an ellipse along an axis making an angle $\pi/2n$ radians with the original axis, the critical wavelengths vary in the manner shown. Since only even waves of the zero order exist, the curves for the eH_0 and eE_0 waves have zero slope at zero eccentricity and turn



(*Chu, courtesy of Journal of Applied Physics*)

FIG. 7·26 Field configurations of waves in elliptic pipe having eccentricity of 0.75.

back on themselves. For the higher-order waves ($n \neq 0$) the curves are not symmetrical about the line corresponding to an eccentricity of zero.

The field distribution for the six waves considered by Chu are shown in Fig. 7·26. In these figures, as in the previous cross-sectional configurations shown for the various types of guides, the lines which seem to converge toward certain points actually bend in the longitudinal direc-

tion along the axis of the elliptical cylinder. The diagrams illustrating the configuration in the longitudinal direction are not shown.

For the circular pipe the solution for the E_x or H_x components may be written

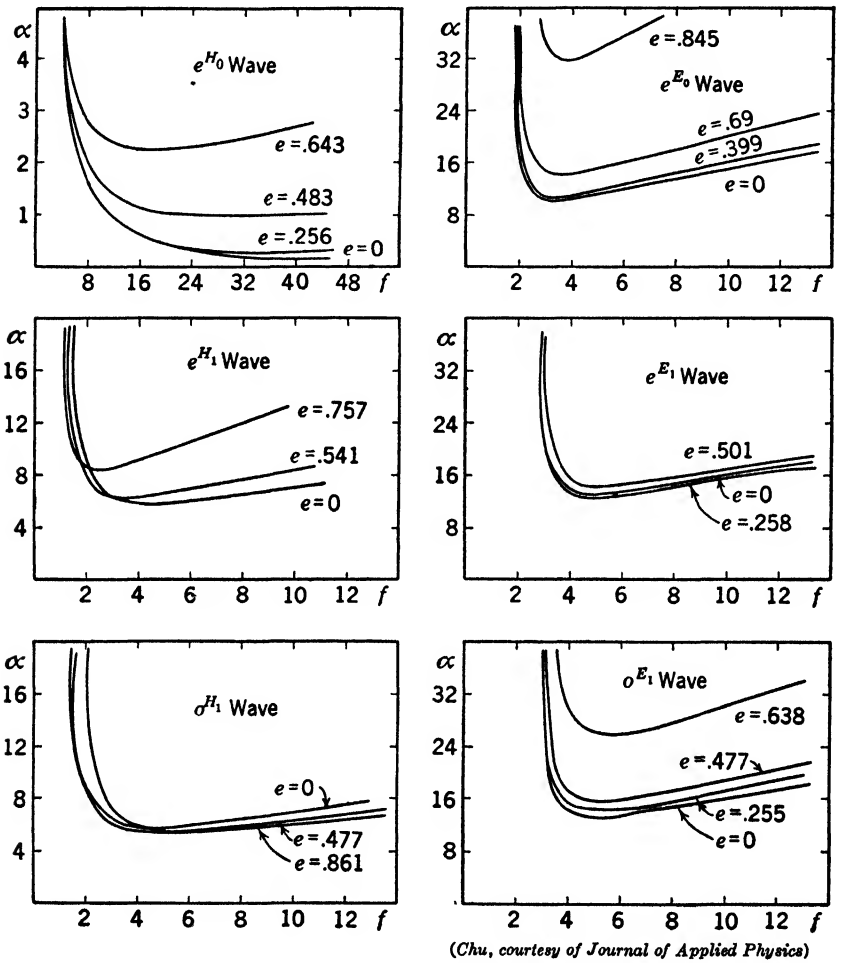
$$\left.\begin{matrix} E_x \\ H_x \end{matrix}\right\} = [A \cos(n\phi) + B \sin(n\phi)] J_n(kr)\, e^{j(\omega t - \beta x)} \qquad [7\cdot190]$$

where $k = \sqrt{\omega^2 \mu\varepsilon - \beta^2}$ and A and B are constants.

Let us assume that the pipe of circular cross section is deformed into an ellipse whose major axis is the $\phi = 0$ axis. A wave in which E_x or H_x is proportional to $\cos n\phi$, i.e., $B = 0$, becomes an even wave in the elliptic pipe. On the other hand, a wave with E_x or H_x proportional to $\sin n\phi$, i.e., $A = 0$, becomes an odd wave in the elliptic pipe. When $\beta = 0$, waves cannot be propagated through the pipe. This condition, as before, defines the critical frequency. We may consider that the two terms, $\cos n\phi$ and $\sin n\phi$, in the general solution 7·190 represent two distinct components of the wave having the same phase constants and hence traveling with the same phase velocity in the circular pipe. When the circular pipe is deformed by shortening the diameter along the $\phi = \pi/2$ axis, the circular cross section becomes elliptical. Under this condition the component wave containing the $\cos n\phi$ term becomes an even wave and that containing the $\sin n\phi$ term becomes an odd wave. But the odd and even waves do not have the same critical wavelengths, since the phase constants differ for the odd and even waves. Hence one wave travels with a higher phase velocity than the other, and the two cannot remain together. Under these circumstances, as was pointed out by Chu, all waves, in circular pipes (except for two important cases) tend to be unstable under slight deformation of cross section. The two exceptions are stable for small deformations. The first case is obtained when either A or B in equation 7·190 is zero. Then, if the deformation takes place along one of the axes of symmetry, the deformed wave will be entirely even or odd, and there will be nothing to divide. The second case is obtained if $n = 0$, $A \neq 0$. Under this condition the field distribution in any circular cross section is independent of ϕ, that is, the field is circularly symmetrical. Since there is only one type of elliptic pipe wave into which such a field distribution can be deformed, there can be no splitting. Thus we may conclude that for a slight deformation of the physical shape of a circular pipe the H_0 and E_0 waves are stable, and all other waves are unstable except when the deformation occurs along an axis of symmetry.

The treatment of imperfectly conducting pipes of elliptical cross section is analogous to the treatment of imperfectly conducting circular

pipes. It requires that the physical conductors have a high conductivity $\sigma \gg \omega\varepsilon$ so that the wave shape in the practical metal pipes is not materially changed from the wave configuration in the perfectly conducting pipes. The attenuation constants for the various waves ob-



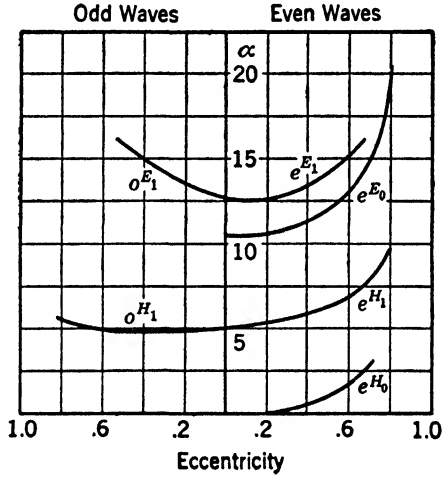
(*Chu, courtesy of Journal of Applied Physics*)

FIG. 7·27 Attenuation of waves in air-filled copper pipes with periphery of 40 cm. Abscissas — frequency in 10^9 cycles per second. Ordinates — attenuation in decibels per mile.

tained by Chu are given as a function of frequency in Fig. 7·27. These curves are for air-filled pipes having several values of eccentricity e and equal peripheries (S = 40 cm). In all cases except one, representing the eH_0 wave with $e = 0$ the attenuation rises with increasing frequency.

Chu points out that the curve for eH_0 with $e = 0.256$ will rise for frequencies beyond the range of the figure. If the eccentricity is increased, the frequency being held constant, the attenuation constant for the eE_0 and eH_0 waves always increases. On the other hand the attenuation constants for the oH_1 and eE_1 waves are not appreciably affected by a change in eccentricity. In the slightly deformed circular pipe the E_0 and H_0 waves have higher attenuations than when the pipe is perfectly circular. It is seen in Fig. 7·27 that the H_0 wave loses its anomalous attenuation characteristic when the circular pipe is deformed.

The minimum attenuation constants for different waves, independent of frequency, are plotted in Fig. 7·28 as a function of eccentricity. These curves represent elliptic pipes of equal peripheries ($S = 40$ cm). It is



(*Chu, courtesy of Journal of Applied Physics*)

FIG. 7·28 Minimum attenuation of waves in air-filled copper pipes with periphery of 40 cm. Ordinates — minimum attenuation in decibels per mile.

seen that the lowest attenuation for the oH_1 wave occurs for a value of eccentricity equal to 0.5, whereas the lowest for the eE_1 occurs for an eccentricity of approximately 0.2.

The following general conclusions were reached by Chu as a result of his work on elliptic pipes. In regard to stability: (1) When pipes of circular cross section are slightly deformed along one of the axes of symmetry, all waves are stable. (2) For any slight deformation of the cross section of a circular pipe, waves of the type H_0 and E_0 are stable. (3) Except for these two types, all waves in circular pipes are unstable when the cross section is deformed.

In regard to critical wavelength for waves in elliptic pipes: (1) The eH_1 wave possesses the longest critical wavelength for a given length of periphery. This wavelength increases with increased eccentricity and approaches the value $\lambda_0 = 0.83S$ for unit eccentricity. (2) For the other five types of waves investigated, the critical wavelength either remains unchanged or decreases with increased eccentricity.

In regard to attenuation: (1) The anomalous attenuation characteristic of the H_0 wave in circular pipes is lost under small deformation. Also its attenuation is always increased by such deformation. (2) The wave having the next lowest attenuation to that of the eH_0 wave is the oH_1 wave. For a given periphery the minimum attenuation for this wave occurs, not in the circular pipe, but in an elliptic pipe having an eccentricity of 0.5, as is seen in Fig. 7·27 or Fig. 7·28.

In this work the eccentricity e is related to N, the ratio of the minor to the major axis of the ellipse, by the equation

$$N = \tanh (\operatorname{sech}^{-1} e)$$

Thus if $e = 0.5$, $\operatorname{sech}^{-1} e = 1.320$, and $N = 0.867$. For $N = 0.99$, representative of relatively poor commercial "circular" tubing, $e = 0.140$.

PROBLEMS

7·1 By constructing mathematical areas $\Delta r\, \Delta x$ and $r\Delta\varphi\, \Delta r$, and equating the surface integral of the current density ι through these areas to the line integral of H around these areas, obtain equations 7·8 and 7·9.

7·2 What assumptions are necessary in order to obtain the curl components in the manner described in problem 7·1? Are these assumptions consistent with those made when Taylor's theorem for the expansion of a function about a point is applied?

7·3 Obtain equation 7·24 by an elimination process among equations 7·18 and 7·19 similar to that employed in Chapter 6 to obtain the wave equation in rectangular coordinates.

7·4 Obtain the attenuation constant for E waves as given in equation 7·117 by substituting the appropriate field vectors.

7·5 Obtain equation 7·118 for the attenuation constant for H waves.

7·6 Evaluate $Y_0(\rho)$ for $\rho = 0.6$ accurate to three decimal places by use of equation 7·41.

7·7 Evaluate $J_0(\rho)$ for $\rho = 2.0$ accurate to three decimal places by use of equation 7·40.

7·8 Evaluate to slide-rule accuracy all roots (zeros) lying between 30 and 100 of the function $J_p(\rho)$ for $p = 5$ using equation 7·49.

7·9 Evaluate to slide-rule accuracy all roots (zeros) lying between 30 and 100 of the function $Y_p(\rho)$ for $p = 7$ using equation 7·51.

7·10 We have shown in Chapters 5 and 6 how guided waves may be analyzed into

components which are interpreted as transverse electromagnetic waves reflected repeatedly by the conducting guides. Examine the possibility of applying a similar analysis to cylindrical guides.

7·11 A copper tube has an inside diameter of 3 inches. Evaluate the cut-off frequencies of E_0, E_1, H_0, and H_1 waves in such a guide if the dielectric is air.

7·12 A wave of 5×10^9 cycles per second is propagated through the guide of problem 7·11. Evaluate the attentuation in decibels per mile for each of the four modes listed (1 neper = 8.7 db).

7·13 A copper pipe has a perimeter of 30 cm and may be deformed into an ellipse if desired. At what frequency and with what deformation does the H_1 wave show a minimum of attenuation? What is this attenuation?

7·14 What assumptions are made in deriving the equations for L, G, and C as given in equations 7·173 and 7·174*a*?

7·15 Prove that equation 7·189 giving the optimum ratio of b/a for a coaxial line is correct. Does this optimum hold for all frequencies, or is it correct only for hyperfrequencies?

7·16 A high-quality dielectric has negligible conductivity and a dielectric constant of $\kappa_e = 5$. A rod of this material is 15 cm in diameter and serves as a wave guide at a frequency of 6×10^9 cycles per second. (No metal tube is used.) Evaluate the ratio of the velocity of the wave in the guide to the free-space velocity.

7·17 Evaluate the dielectric constant κ_e necessary if a wave of 500-mc frequency is to be propagated within a metal guide of radius 2 cm in the TE_{01} mode. Repeat for the TM_{01} mode. Are such dielectrics now common?

7·18 An air-filled pipe is formed from relatively thin sheet copper. The perimeter is 50 cm and the frequency 3×10^9 cycles per second. Compare the attenuation constants of the most favorable modes if the pipe is made round or square.

7·19 Repeat problem 7·18, but generalize by considering the most favorable rectangular and elliptic shapes.

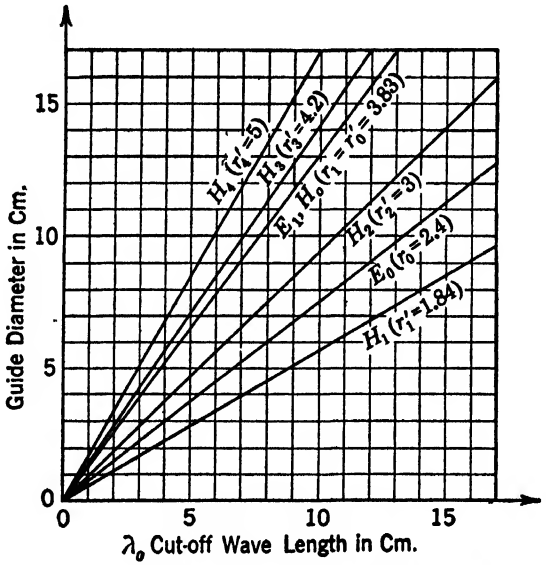
7·20 An air-filled coaxial line is made of copper. The radii a and b are 1 and 3.6 cm respectively. Evaluate the attenuation constant for this line and the components of R contributed by the inner and outer conductors separately.

CHAPTER 8

WAVE GUIDE EXPERIMENTAL APPARATUS

8·1 Introduction

It was seen in the preceding sections that there are a great many types of waves that may be transmitted through wave guides. Each of these waves is characterized by the orientation of the lines of electric and magnetic intensity within the guide. Each is also characterized by a critical



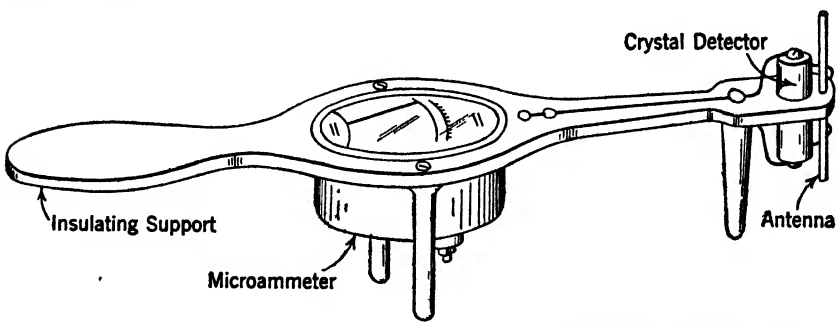
(Clavier, courtesy of Electrical Communication)

FIG. 8·1 Guide diameter as a function of the critical wavelength for circular pipes having air dielectric.

wavelength above which a particular type may not be transmitted. A comparison of the cut-off wavelengths for several wave types as a function of the guide diameter is shown in Fig. 8·1 for circular pipes. Since the guide diameter is of the order of the magnitude of the critical wavelength required for transmission, the practical application of wave guides is restricted to the use of hyper frequencies. Consider, for example, the transmission of the H_1 wave in a hollow circular pipe filled with air. From Fig. 8·1 we see that for a frequency of 3000 megacycles, corre-

sponding to a wavelength of 10 cm, the critical pipe diameter required is 5.85 cm. Since the critical diameter is inversely proportional to the square root of the dielectric constant of the medium within the pipe, the choice of a medium of high dielectric constant reduces the guide diameter appreciably. However, we must remember that the choice of the dielectric is important from another point of view, namely, attenuation. Materials are available which have a dielectric constant of 4 and over, and at the same time introduce no serious loss. In this way the critical diameter may be reduced, for this 3000-megacycle wave, to approximately 3.0 cm or 1.2 inches. For higher frequencies, or for materials having still higher dielectric constants, the required diameter may be further reduced. In this way the hollow wave guide may require dimensions of the same order of magnitude as are required for conventional coaxial conductors. For waves of the E_0 and H_0 types, and for higher-order waves, the critical dimensions are larger.

At relatively low frequencies, filters, equalizers, resonators, and impedance transformers are readily constructed from lumped circuit elements according to well-known design methods. At higher frequencies these techniques fail in various ways. The hollow conducting wave guide is of great importance in that it performs the functions of these various devices at the highest frequencies now of interest for communication work.



(*Southworth, courtesy of BSTJ*)

FIG. 8·2 A convenient probe for exploring the field around a source of electromagnetic radiation. The insulating shaft between the crystal and meter should be a foot or more in length.

8·2 Probes and Detectors

The field configuration of the lower-order E and H waves may readily be explored by means of a probe detector. One of the simplest of these is the hand probe shown in Fig. 8·2. This consists of a small crystal detector to which are attached two short antennas. When dealing

with waves of the order of 3000 megacycles, the length of each antenna should be less than an inch, in order that the field configuration be disturbed as little as possible. The detector and antennas are mounted, in the figure, on a strip of Bakelite a foot or so in length. The wires from the crystal detector run down the strip to a meter which is mounted just above the handle on the strip. If desired, the meter may be placed on a table, connected to the antennas through a twisted pair of flexible leads. Figure 8·3 shows a detail of the crystal mounting. A

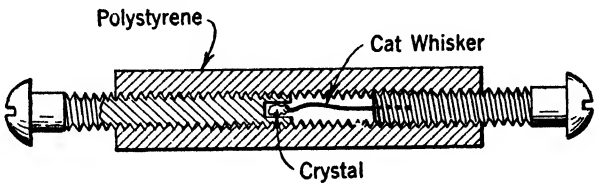


FIG. 8·3 A crystal detector mounting suggested by Barrow. The polystyrene rod may be approximately a half inch long.

fragment of silicon crystal is mounted in a small hole drilled in the end of an 8–32 screw. A tungsten cat-whisker is mounted in the upper screw. Contact between cat-whisker and crystal is made by adjusting one or the other of the screws until a sensitive spot is found. The holder for the two screws is a small polystyrene rod drilled and tapped for the correct screw size. The crystal detector makes a very satisfactory and convenient unit for laboratory work. Thermocouples, diodes, and triodes may also be used as detectors for waves of these frequencies.

In determining the configuration of the wave pattern at any point in the field, the antennas of the probe are rotated until a maximum reading is obtained on the meter. The axis of the two antennas then defines the direction of the lines of electric intensity in the neighborhood of the point. The relative intensity of the field at different points is indicated by the amplitude of the meter indications. Hand probes are useful generally for detecting the presence of a field and for determining its relative amplitude. When carefully constructed they will respond to a few milliwatts of power. It is usually desirable to use a microammeter as the indicating device. A shunt may be provided to keep the meter " on scale " in the presence of strong fields. The hand probe may be mounted in a fixed position at an appropriate spot when various tuning adjustments are being made on concentric line stubs or oscillators, etc. Under these circumstances it serves as a tuning indicator.

A very serviceable piece of laboratory apparatus is known as a traveling detector. With its aid, wavelengths and field configuration of waves inside a pipe may be determined experimentally. Such a unit is shown

in Fig. 8·4. It is similar to designs suggested by Barrow and Southworth, in which a probe similar to that described in the preceding paragraph is mounted on a slider which may be moved along a longitudinal slot cut in a section of a wave guide. The slotted section of wave guide telescopes the adjoining sections and is free to turn 360° about a horizontal axis. In this way the probe may be carried longitudinally along the wave guide or transversely around it. Adjustment for radial motion of the probe may be made by extending the probe into the pipe any desired distance. A very enlightening diagram illustrating the orientation of the field vectors inside a circular pipe has been prepared by Clavier and Altovsky; it is shown in Fig. 8·5*a* and *b*. Figure 8·5*a* illustrates the orientation for the E_0 wave, and Fig. 8·5*b* that for the H_0. These field distributions represent an instantaneous orientation of the field structure, and the patterns shown are, of course, propagated through the tube with the velocity v_p. If, however, a conducting plane perpendicular to the axis is placed across the guide the electrical intensity tangential to this plane is reduced to zero at the surface. A reflected wave similar to the initial or incident wave results, and standing waves are produced. At surfaces parallel to the reflecting plane the electric intensity in the plane is again zero if these planes are separated by one-half of the wavelength in the guide. The field distribution observed with a probe in terms of steady meter readings is similar to that shown as an instantaneous distribution in Fig. 8·5.

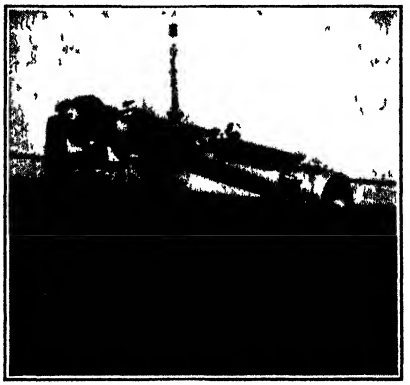


FIG. 8·4 A traveling detector arrangement used at the Illinois Institute of Technology for exploring the wave patterns of waves in circular guides. Similar detectors for use with rectangular or other forms of wave guide are readily constructed.

By the appropriate design of the antennas of the probe, E or H vectors may be identified. For example, if the antenna is arranged to coincide with a line of electric intensity, a maximum deflection due to the E vector may be indicated. On the other hand if the antenna is arranged in the form of a small loop (see Fig. 8·6) and is adjusted so that the plane of the loop is normal to the H vector, then a maximum current will be induced in the loop circuit indicating the presence of a maximum H vector.

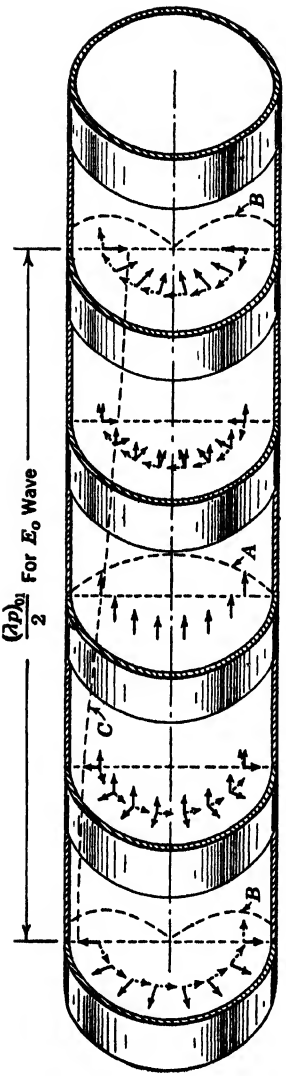
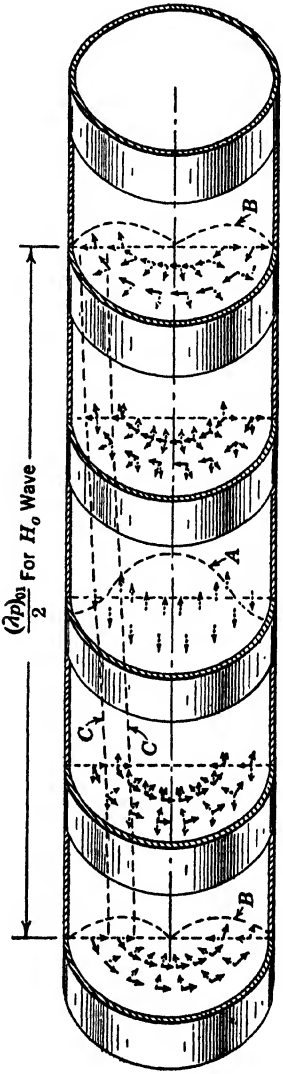


FIG. 8·5a E_0 wave: electric (full lines) and magnetic (dotted lines) vectors.



(*Clavier, courtesy of Electrical Communication*)

FIG. 8·5b H_0 wave: electric (full lines) and magnetic (dotted line) vectors.

Curve A: distribution of longitudinal vectors along the radius (Bessel function J_0). Curve B: distribution of transverse vectors along the radius (Bessel function J_1). Curve C: distribution of all vectors along the axis (sinusoidal function).

By sliding the probe along the slot in the longitudinal direction, curves such as those shown in Fig. 8·7 may be obtained. The field intensity as indicated by the rectified current is plotted as a function of longitudinal displacement of the probe in Fig. 8·7*a*. Figure 8·7*b* is obtained by leaving the detector probe fixed and moving the reflector or piston, as it is usually called, along the tube. The sharpness of resonance or high

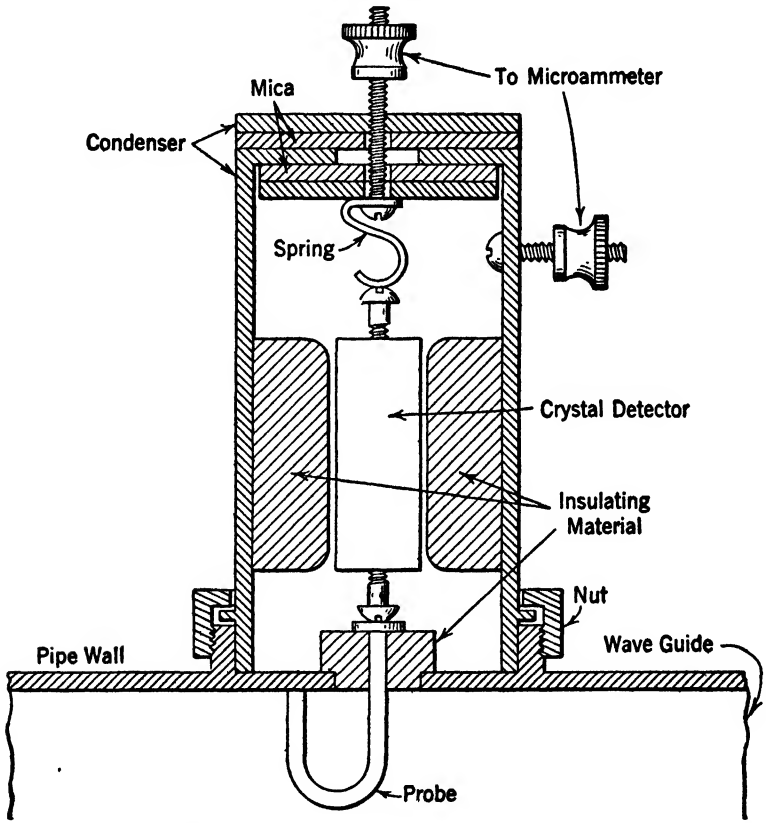
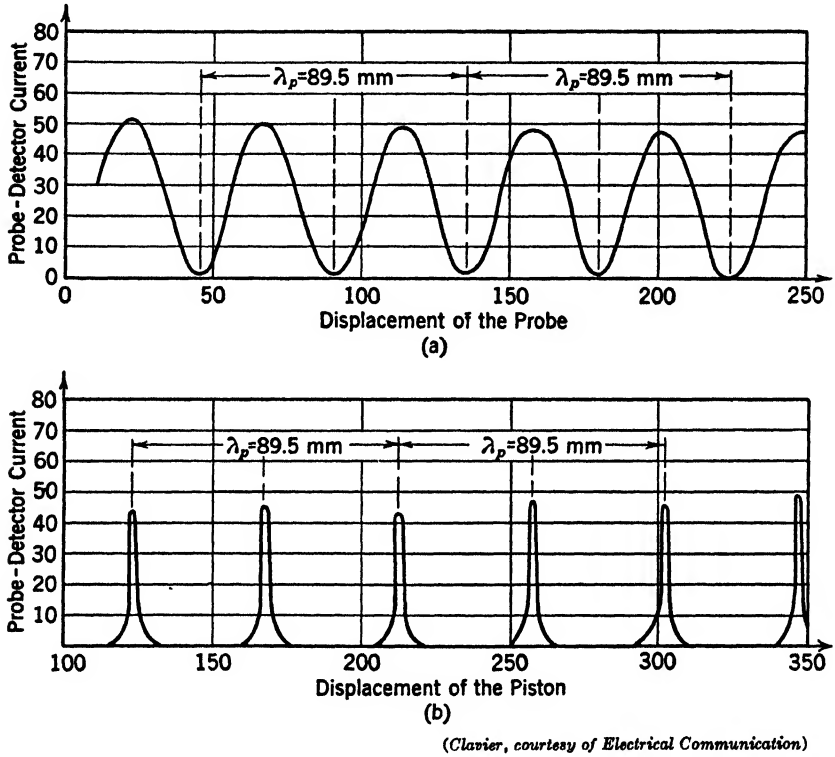


FIG. 8·6 Probe and detector arrangement for determination of field distribution in wave guides.

selectivity, is evident from the sharp peaks that are obtained. The resonant points or peaks determine guide lengths that are an integral number of half wavelengths apart. Since the maximum deflections are more sharply indicated in Fig. 8·7*b*, it is evident that by leaving the probe fixed and moving the piston, the wavelength in the pipe may be determined with greater accuracy than in the reverse process indicated in Fig. 8·7*a*. The selectivity of the system is reduced if a receiver or

other absorber of energy is placed in the resonant chamber. Evidence of the simultaneous existence of two different wave configurations within the guide is presented in Fig. 8·8, due to Clavier.* An axial probe in a traveling piston served as a receiver for the E_0 wave. The



(*Clavier, courtesy of Electrical Communication*)

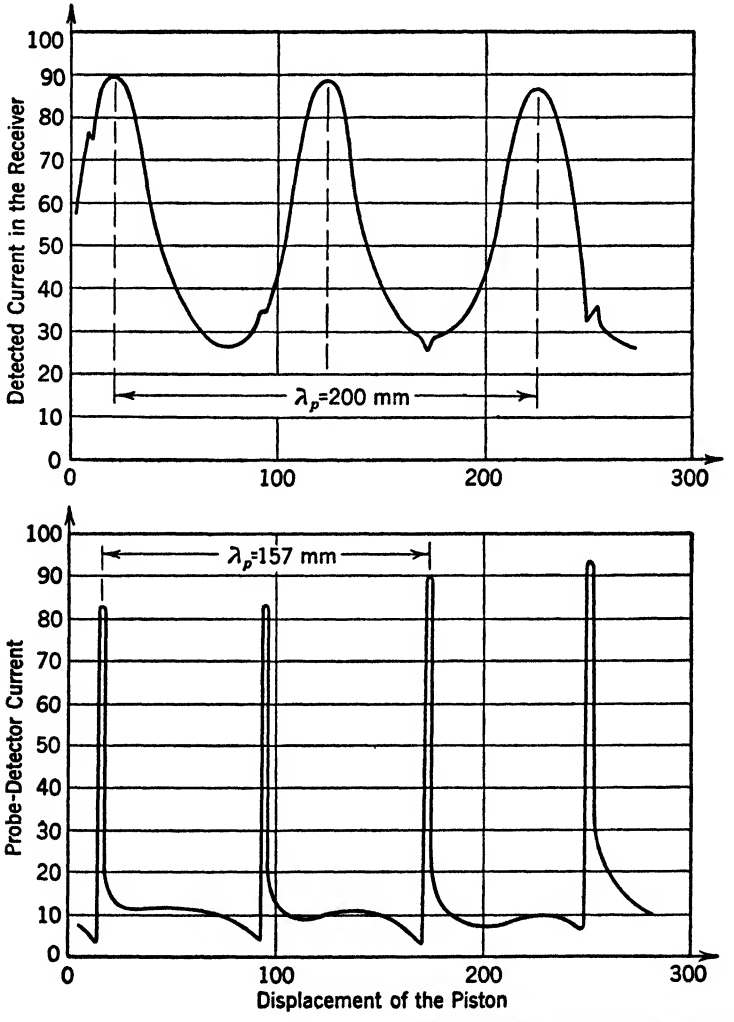
FIG. 8·7 Probe detector current in arbitrary units as a function of: (*a*) displacement of the probe along the tube in millimeters; (*b*) displacement of the piston along the tube in millimeters. The wavelength in the pipe is represented by λ_p.

rectified current in this axial probe is plotted as receiver current in the upper half of Fig. 8·8. The broad character of the maxima indicates that this probe was relatively tightly coupled to the E_0 field so that the resulting resonance was of low selectivity.

The rectified current observed by a radial probe inserted a small distance through the side of the tube is plotted in the lower half of

* A. G. Clavier and V. Altovsky, "Experimental Researches on the Propagation of Electromagnetic Waves in Dielectric Guides," *Electrical Communication*, **18**, 81, 1939.

Fig. 8·8. The sharpness of these curves is evidence of the fact that the axial probe or receiver does not affect the selectivity for this mode of transmission. The mode is identified as the H_1 by the frequency of 2370 mc and the wavelength $\lambda = 157$ mm in the guide.



(Clavier, courtesy of Electrical Communication)

FIG. 8·8

Indication of the simultaneous presence of two different wave types in a circular resonant chamber. The upper curve shows the current rectified from an axial probe mounted in the piston. The lower curve represents the current rectified from a small radial probe fixed in the side of the guide. Displacements measured in millimeters. The wavelength in the pipe is represented by λ_p.

The guide was energized by an axial probe similar to that mounted in the piston, and under ideal conditions only the E_0 wave would be excited. Slight dissymmetry in the system suffices to explain the establishment of a small component of the H_1 wave. Because the axial probes do not serve as loading to such a mode the resonant selectivity is quite high and a considerable field is established with only small excitation. Accordingly energy is propagated along the guide from the source to the receiving piston in the E_0 mode with a relatively small standing wave ratio. Simultaneously a standing wave pattern of H_1 mode of considerable amplitude is established for each favorable position of the piston.

8·3 Wave Filters in Guides

In 1888 Hertz described an experiment with plane-polarized electromagnetic waves.* He used a frame of insulating material which supported a large number of separate parallel wires. It was found that this

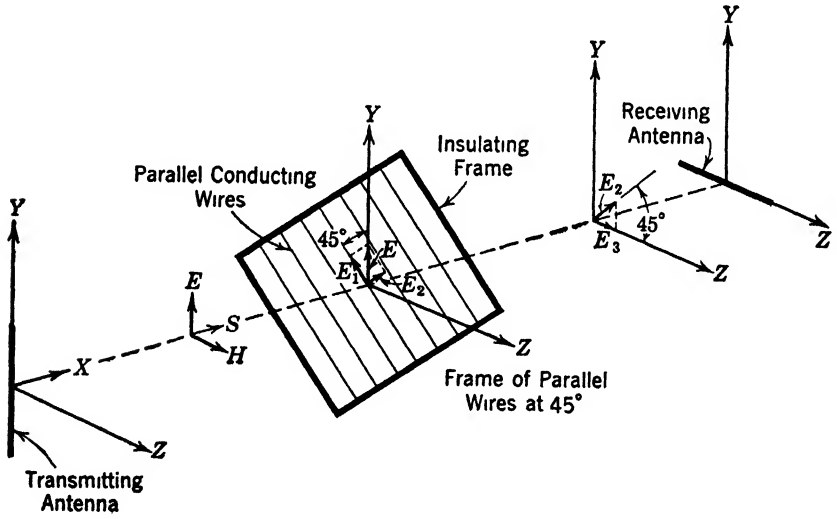


FIG. 8·9 A vertically polarized wave is changed to one having a horizontal component by the action of the parallel conducting wires in the frame.

frame produces no effect upon a wave system when the wires are at right angles to the electric intensity but that it acts as a solid conducting sheet when the wires are parallel to the electric intensity. Such a screen may be used as a filter to produce a plane-polarized wave, the component of electric intensity along the wires being reflected, the com-

* H. Hertz, *Electric Waves*, p. 179, Macmillan Company, 1893.

ponent of intensity perpendicular to the wires being transmitted freely.

Hertz also used the screen as a wave converter. He observed that a vertically polarized wave, established by a vertical radiating conductor, produced no effect upon a particular horizontal antenna. The frame was then inserted between radiator and detector, the wires being inclined at 45°. See Fig. 8·9. A strong signal in the horizontal receiving antenna was observed. This effect is explained by the method of resolution of vectors. The electric intensity at the wire frame is resolved into two

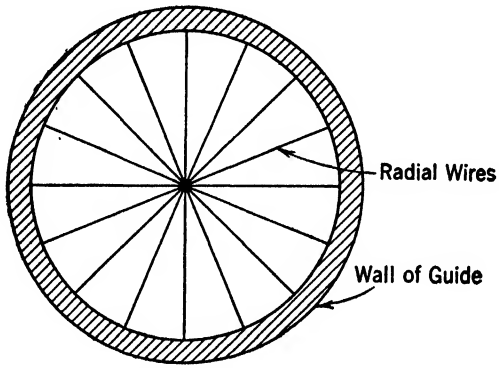


FIG. 8·10 Filter for the elimination of an E_0 wave in a circular wave guide.

components at 45° from the vertical, one parallel and one perpendicular to the wires. The perpendicular component is freely transmitted and arrives at the receiving antenna. The receiving antenna makes an angle of 45° with this wave and thus receives a considerable signal. In this case each resolution is at 45° and the factor which applies is $\cos 45° = .707$. Since two successive resolutions are necessary the voltage delivered to the receiving antenna is $.707^2 = \frac{1}{2}$ as great as it would be if the two antennas were parallel and no screen were present.

The effect may also be explained on a somewhat more physical basis. The wires have a component of length along the incident electric field, and a current is established in each wire by its action. In turn the wires have a component of their length along the receiving antenna, and the currents produced in the wires are thus able to produce a voltage in that antenna.

More recently Clavier* and Brillouin† have described several devices

* A. G. Clavier and V. Altovsky. "Experimental Researches on the Propagation of Electromagnetic Waves in Dielectric (Cylindrical) Guides," *Rev. gén. élec.*, May 27 and June 3, 1939.

† Leon Brillouin, "Hyperfrequency Waves and Their Practical Use," *Electrical Communication*, 19, 18, 1941.

for use in wave guides which utilize the principle described by Hertz. Perhaps the simplest of these is the radial wire structure of Fig. 8·10. Such a structure, serving as a virtually perfect reflector to the E_0 wave in a circular wave guide because the conductors are everywhere parallel to the electric intensity, prevents the propagation of such a wave. If one of these filters is not adequate to produce the required attenuation two or more may be used. A spacing of approximately a quarter wavelength is appropriate.

The filter just described produces negligible attenuation to the H_0 wave since the conductors are everywhere normal to the electric inten-

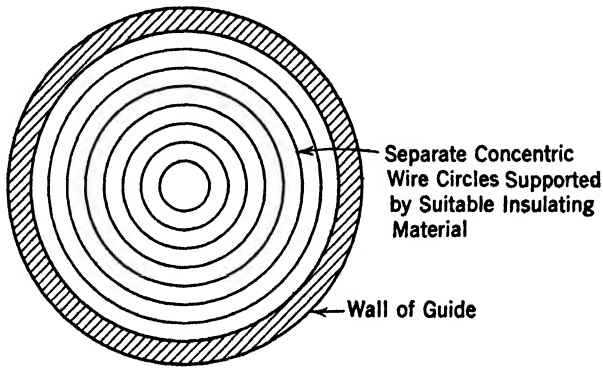


FIG. 8·11 Filter for the elimination of an H_0 wave in a circular wave guide.

sity of such a wave. Appreciable but not extremely large attenuations to waves of the E_1, H_1, and other modes are produced by this filter. Since the action is reflection rather than absorption it is feasible to transmit power in these modes through the filter by a suitable readjustment of the impedance of the transmitting antenna.

A filter similar to the above is shown in Fig. 8·11. Here, however, the conductors take the form of concentric circles each insulated from the others. Such a filter passes the E_0 wave with negligible effect but strongly attenuates the H_0 wave because the conductors are everywhere parallel to the electric intensity of such a wave. Again, waves of higher order like the E_1 and H_1 are partially reflected and are distorted in passing through such a structure.

A third device, shown in Fig. 8·12, may well be called a wave transducer. In the presence of an E_0 wave the electric intensity along the radial portions of the conductors produces a current in each conductor. This current, flowing in the angular portion of the conductors, simulates a continuous ring of current flow. Such a current distribution is favorable to the establishment of an H_0 wave. Accordingly this device serves

as an excellent transducer of E_0 into H_0 waves. The converse action is equally well performed. The circular electric field of the H_0 wave induces current in the wires by its action on the angular portion *ab*, and this current, which also flows in the radial portion of the conductors, establishes an E_0 wave in the pipe.

A complete wave filter and transducer system is shown in Fig. 8·13. It is seen that two radial filters, a wave transducer, and two circular

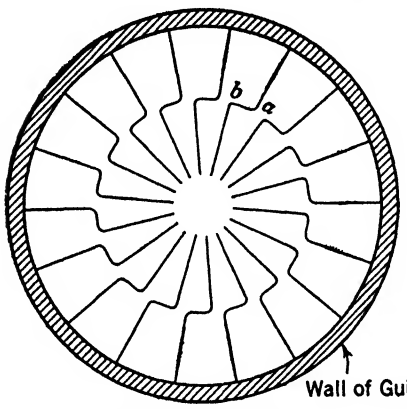


FIG. 8·12 Wave transducer.

(*Clavier, courtesy of Electrical Communication*)

FIG. 8·13 A complete wave filter and transducer system installed in a circular wave guide.

filters are employed. Small slats of polystyrene or other low-loss dielectric are required to support the circular rings of the second filter. An E_0 wave generated at the transmitter end of the system is freely propagated through the circular filters and strikes the transducer where it is partially reflected, partially transformed, and partially transmitted. The reflected portion is returned to the sending device and so is not lost. The transformed portion establishes an H_0 wave, the principal function of the device. The transmitted wave is reflected by the radial filter and, if the adjustment is appropriate, returns to the transducer in phase with the incident wave. The H_0 wave established at the transducer is radiated in both directions. The forward-propagated component passes freely through the radial filter and travels down the guide; that which is propagated backward is reflected by the circular filter and reinforces the forward, useful, wave. The operation is clarified by reference to Fig. 8·14.

The operation of such a wave transducer is quite efficient, only a small portion of the supplied power being lost. Also, the waves produced are quite pure. The ability to convert one form of wave into the other is of considerable importance, particularly because it is difficult to excite

the H_0 mode directly from an oscillator. With this device an E_0 wave which is readily produced is converted into a pure H_0 wave.

Any wave configuration in a conducting guide system may be reflected by an arrangement of wire conductors which coincide with the lines of electric intensity. Such an arrangement constitutes a filter for the elimination of that particular wave type. Similarly various forms of

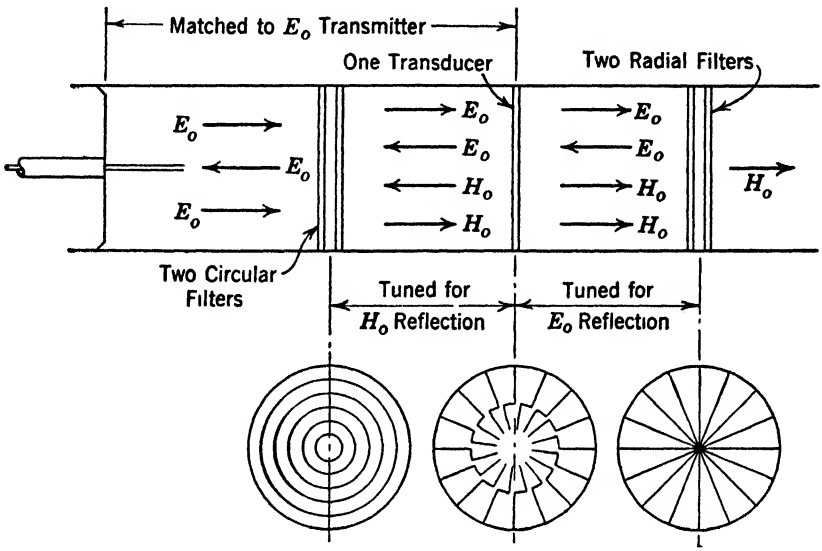


FIG. 8·14 Detail of the location and action of the filters and transducer in the system shown in Fig. 8·13.

wire transducers may be devised by applying the principle that each conductor must have part of its length along the electric field of the incident wave and part along the field to be produced. By suitable combinations of these filters other transformations similar to the $E_0 - H_0$ just described may be worked out.

8·4 Use of Wave Guides

Wave guides have at least three properties which are of practical importance. Most basic of these is the property of transmitting power from one point to another with low loss. This property is fundamental in that the other properties depend upon it. A suitable wave guide transmits power with considerably less loss than is associated with conventional parallel wires or coaxial lines. This property is not of great practical importance at the moment because of the expense of suitable guides and because the inflexible nature of guides introduces much

mechanical difficulty. W. L. Barrow has, however, demonstrated the transmission properties of wave guides made of flexible tubing similar in construction to standard BX cable used for house wiring.

The second property and perhaps the most significant one is resonance. This results from the fact that waves are propagated with small values of attenuation in suitable guides and that very complete reflections occur if a guide is bounded by a metal barrier. For convenience in tuning this reflecting barrier is often made in the form of a moving metal piston which makes a dependable contact to the guide at many distinct points. A useful resonant chamber results if a short section of a wave guide is bounded at one end by a fixed metallic plate with a circular hole or iris and at the other end is fitted with a movable piston. The iris allows the entrance of a certain amount of power and at the same time offers a relatively complete reflection to waves within the chamber. Such a chamber resonates strongly whenever the length is effectively an integral number of half wavelengths. The iris affects the relation somewhat, because the reflection produced is not exactly at the plane of the iris surface.

Such a resonant chamber may be made to serve any of the functions which at lower frequencies are performed by a simple tuned circuit consisting of a coil and a condenser. The motion of the piston may be calibrated in frequency or wavelength so that the device serves as a wavemeter. Even more important, a resonant chamber may be used as an impedance transformer in a variety of applications. By its use two wave guides of different characteristic impedance may be matched to each other, or a load or generator of arbitrary impedance may be matched to a guide.

The third property is radiation. A portion of the power in a wave which reaches the open end of a metallic guide is radiated into the air beyond and is lost. If the end of the guide, either round or rectangular, is flared in a suitable fashion this radiation is rendered highly directional and the reflection back into the guide becomes small. Such a flared conductor is referred to as an electromagnetic horn; it is discussed further in Chapter 11. It may be thought of as a device which matches the characteristic impedance of the guide to that of free space.

8·5 The Wave Guide as a Filter

One of the most fundamental properties of the hollow conducting wave guide is that frequencies below some critical value are not transmitted. Accordingly every wave-guide section is intrinsically a high-pass filter. In the case of perfectly conducting wave guides, the equations indicate that all frequencies above the critical value are trans-

mitted without loss and all frequencies below that value are completely attenuated.

Actually the transmission between two suitable antennas in a wave guide of reasonable length does not show this idealized form. Some coupling due to the familiar induction field exists at frequencies below the cut-off, and the measured characteristic is somewhat rounded. Nevertheless, the cut-off produced by a section only a few wavelengths long is satisfactory for most purposes.

Filters having band-pass and band-elimination characteristics may be constructed by combining wave-guide sections in suitable combinations. The method is quite similar to that used in the design of acoustical filters. Resonant cavities as series and shunt elements are employed in such a

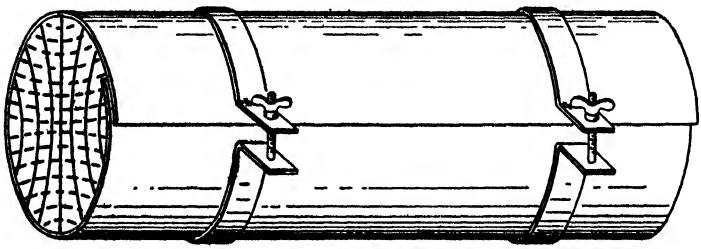


FIG. 8·15 " Pinch pipe " for demonstrating cut off frequencies.

combination to achieve the desired result. No considerable amount of work in this particular field has yet been published, but great possibilities exist.

The high-pass characteristic of a circular guide is readily demonstrated by varying the frequency of the transmitted wave and observing the deflection of a detector several wavelengths distant. A very sharp decrease in the deflection occurs when the frequency passes the cut-off value.

It is not always practical to vary the transmitted frequency in this way. An alternative procedure is to leave the frequency fixed and to vary the diameter of the tube. If a small transmitting antenna and the antenna of a hand probe, as previously described, are separated by a distance of a few feet the signal received at the probe may be quite small. When a wave-guide section is inserted between the two a strong signal is received, indicating that the transmission is largely through the guide. If the diameter of the guide is now gradually reduced a point is reached at which the received signal drops sharply. This is the cut-off diameter for the frequency in question as given by Fig. 8·1. Such a tube is readily constructed by rolling up a sheet of metal some 2 feet square into the

form of a cylinder. The diameter is then adjusted by varying the degree of overlap at the edges. Experience shows that such a guide is electrically tight even though no special effort is made to secure contact where the edges overlap. If the H_1 wave is used with the electric lines along a diameter which includes the inner edge of the rolled sheet no current need flow at the overlap and the behavior is accordingly regular. Although the experiment as here described is scarcely quantitative it does illustrate very clearly the principle involved.

8·6 Experiments with a Resonant Chamber

Figure 8·16 shows a resonant chamber as constructed at the Illinois Institute of Technology. A brass tube approximately 3 inches in diameter and 10 inches long is closed at one end by a brass plate having a central opening or iris approximately an inch in diameter. A movable brass piston provided with a large number of spring brass fingers or con-

FIG. 8·16*a* Cylindrical resonant chamber following the design of Barrow and Southworth. The piston, shown in Fig. 8·16*b*, is displaced by means of the rack-and-pinion arrangement.

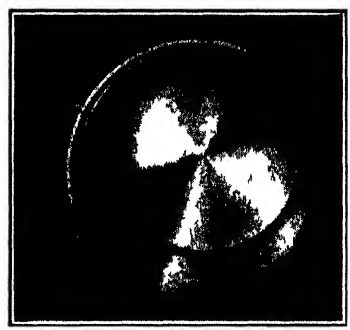


FIG. 8·16*b* End view of the resonant chamber of Fig. 8·16*a*, with metal plug removed. The spring contacts between the piston and inside wall of the guide are shown.

tactors closes the other end of this cavity. The position of the piston is adjustable by means of a rack-and-pinion arrangement as shown. The large number of contact springs is necessary in order to obtain good contact with the walls of the tube. If a secure contact is not provided the piston does not act as a true closure of the cavity and various undesirable effects occur. In particular a loss of power results, and the observations become variable, depending upon the exact degree of contact existing at the moment.

The operation of such a variable cavity resonator is most simply demonstrated by bringing it close to a suitable probe-detector. Now if a high-frequency field is produced in the region a considerable deflection of the detector is observed. This deflection is markedly decreased by tuning the cavity to resonance with the field, indicating a strong absorption of energy. (This is analogous to shunting a series resonant circuit across some form of load fed from a high-impedance lumped circuit.) With the unit described it is possible to obtain two or more positions of the piston which give minimum deflection of the meter associated with the detector. These positions differ by a half wavelength of the wave within the guide, as specified by the phase velocity for the particular wave and the dimensions of the guide.

A more precise study of the resonance is made by replacing the external detector with a suitable internal probe. A small wire extending into the guide a centimeter or so serves as an antenna to operate a crystal detector mounted just outside the conductor. In this case the deflection of the meter is a maximum when the piston is in the resonant position. This detecting system serves as a high-impedance shunt across the circuit and should therefore be located approximately a quarter wavelength from the iris. With this arrangement two successive positions of the piston may be located with quite high accuracy and the wavelength determined to a few tenths of a per cent. Knowing the diameter of the guide and the wave pattern involved (usually H_1) we may calculate the frequency to the same accuracy.

Such a resonator has a relatively high Q. The Q or sharpness of resonance is decreased as the iris is made larger, as the power required to operate the detector is increased, and as the resistivity of the material is increased. Values of Q greater than 1000 are readily obtained. Values in the order of 10,000 are obtained if the conductor is copper and if other losses are held to a minimum. Usually it is simpler to build the resonant chamber from circular material, but comparable results are obtained with rectangular units.

A number of very significant experiments may be performed with such a resonant chamber. Resonance curves may be plotted for various sizes of iris openings, illustrating the increase of selectivity which results when the opening is small. Holding the iris opening and other conditions constant we may compare the properties of various dielectric materials by placing them at a point of voltage maximum within the chamber and comparing the sharpness of the selectivity curves produced. The superiority of polystyrene and a few other materials over other familiar dielectrics is evident from such a test.

8·7 Impedance Matching

The traveling detector described in section 8·2 is an extremely valuable tool for the adjustment of wave-guide systems. Where tubular guides are to be used for the transmission of power from one point to another it is usually desirable to avoid the presence of standing waves. Such standing wave patterns are readily investigated by means of the traveling detector. When no variation in deflection is observed as the probe is moved lengthwise along the guide it is safe to assume that no appreciable reflection exists.

The simplest possible means of absorbing energy in a wave guide so that no reflection occurs is to close the cross section of the guide with a sheet of material having suitable resistivity. (This is equivalent to terminating a low-frequency transmission line with a resistor equal to the characteristic impedance of the line.) A sheet of dielectric coated with a thin film of graphite is a practical device for this purpose. A sheet of carbon paper may serve as an approximately correct termination.

The cavity resonator using an iris opening, as described in the previous section, may also be used as a non-reflecting termination. Some dissipative material such as a block of Bakelite or a sponge of slightly moistened cotton in the chamber absorbs the energy as the waves are successively reflected past it. By appropriate tuning of the piston the effective characteristic impedance of this unit is made equal to that of the guide proper. A flared section of guide, referred to as an electromagnetic horn, also serves as a termination which propagates rather than absorbs the power supplied. No appreciable reflection within the guide results if the horn design is correct.

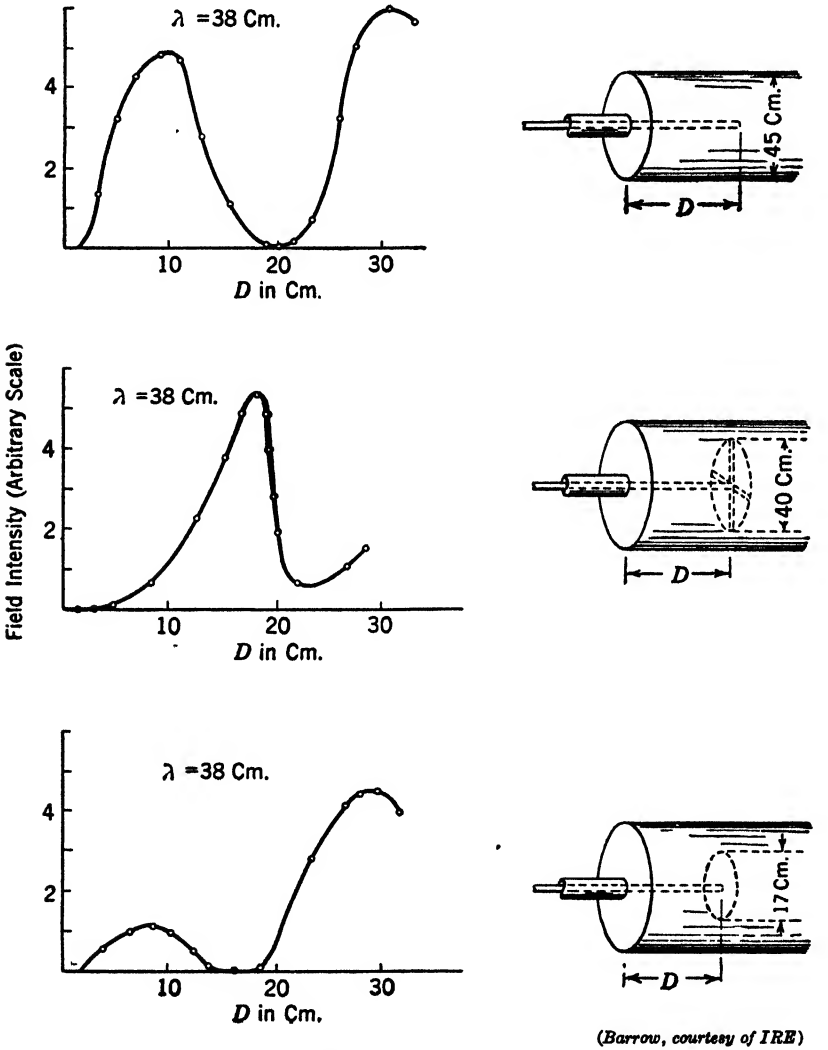
8·8 Terminal Devices

In their extensive research on hollow-tube elements Barrow and Southworth have devised suitable terminal devices for exciting and receiving the various types of wave configurations. These devices make it possible to connect conventional lines and networks to the hollow tubes and to utilize the attenuation and phase-shift properties of these devices for band-pass, corrective, or other functions.

The mathematical analysis of these terminal devices is extremely complex, and as yet no analytical solution is available. Hence, they cannot be discussed in the same exact and quantitative manner as the wave guides themselves. In order to verify the anticipated behavior of a given terminal device it is necessary to check it experimentally.

Terminal devices that have been devised by Barrow for exciting waves of the E_0 type or transverse magnetic (TM_{01}) waves are shown in

Fig. 8·17. These devices show means for connecting a conventional coaxial line (biconductor) to the hollow wave guide (uniconductor). They are equally well suited for transmitting or receiving since any



(*Barrow, courtesy of IRE*)

FIG. 8·17 Terminal devices designed by Barrow for exciting E waves in hollow pipes, together with their respective characteristics. These show the variation of the field intensity at the far end of the tube as a function of the length D of the coaxial rod projected into the tube. The curves illustrate the adjustment of the terminal for maximum energy transfer, and also show the resonant properties of the different terminals.

device which will serve as an effective radiator of waves in the hollow tube also serves as an effective receiver of them. In the conventional

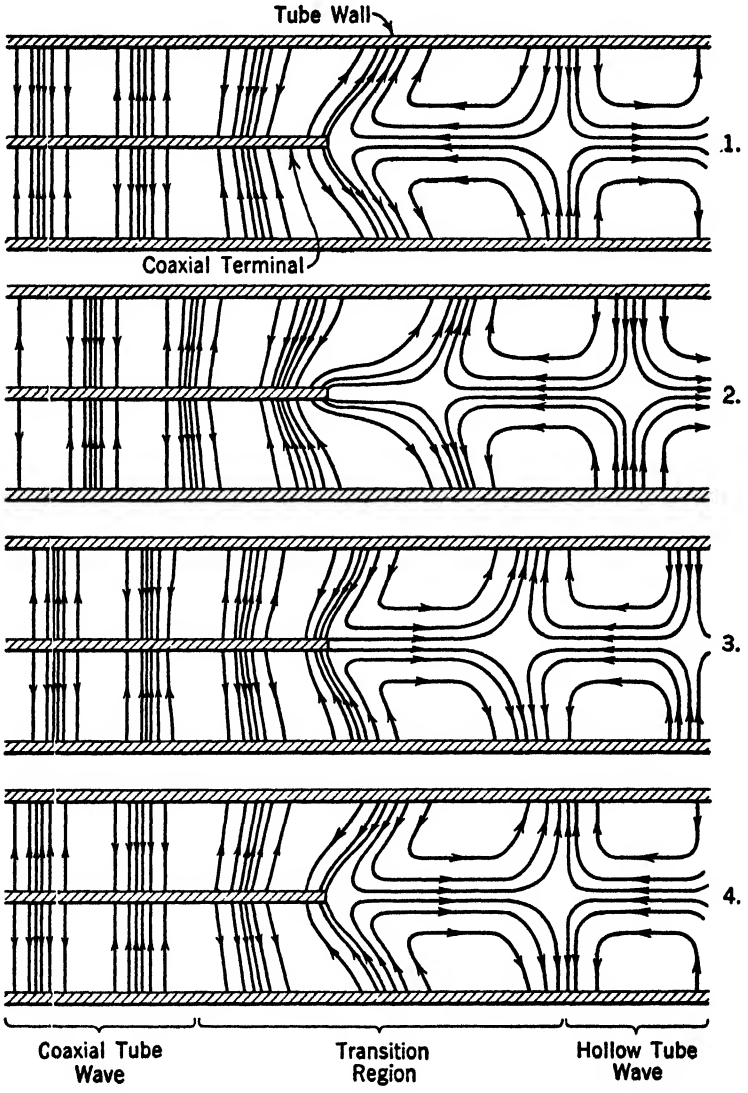
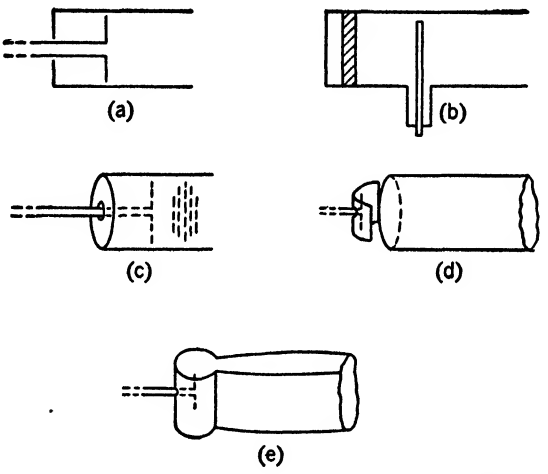


FIG. 8·18 Probable shape of the lines of electric intensity at a coaxial line terminal.

transmission line it is possible to propagate only one kind of wave, whereas in the hollow-tube system several distinct types of waves are possible. Terminal devices may be designed in such a way as to excite

one type of wave to the exclusion of other types. In general it is found that for best operation both transmitter and receiver terminals should transmit and receive waves of a single type only.

All the terminal devices shown in Fig. 8·17 employ a coaxial line. The curves associated with the various devices show the variation of the field intensity at the far end of the tube as a function of the length D of the coaxial rod. It is seen that these various terminals have sharp resonant properties, and that for a maximum energy transfer it is desirable to adjust the length D carefully. The probable shape of the lines of electric intensity at a coaxial line terminal as suggested by Barrow is shown in Fig. 8·18. The velocity of propagation of equiphase surfaces is



(Barrow, courtesy of IRE)

FIG. 8·19 Several types of terminal devices for exciting waves of the H_1 type in hollow pipes.

greater in the hollow tube than in the coaxial line. Hence, as the waves break away from the coaxial terminal we may expect that the field will be distorted to allow the formation of closed loops. The wavelength of a wave in the coaxial section is very nearly that of a wave in free space, whereas in the hollow tube the wavelength is always greater than in free space.

It is seen in Fig. 8·17 that the second terminal arrangement employing radial wires at the end of the coaxial line gives a very sharp resonance phenomenon and a higher intensity amplitude at the end of the line. This may be considered as a better impedance match between the coaxial section and the hollow-tube section. This arrangement allows reflection from the closed end of the tube to reinforce the forward radia-

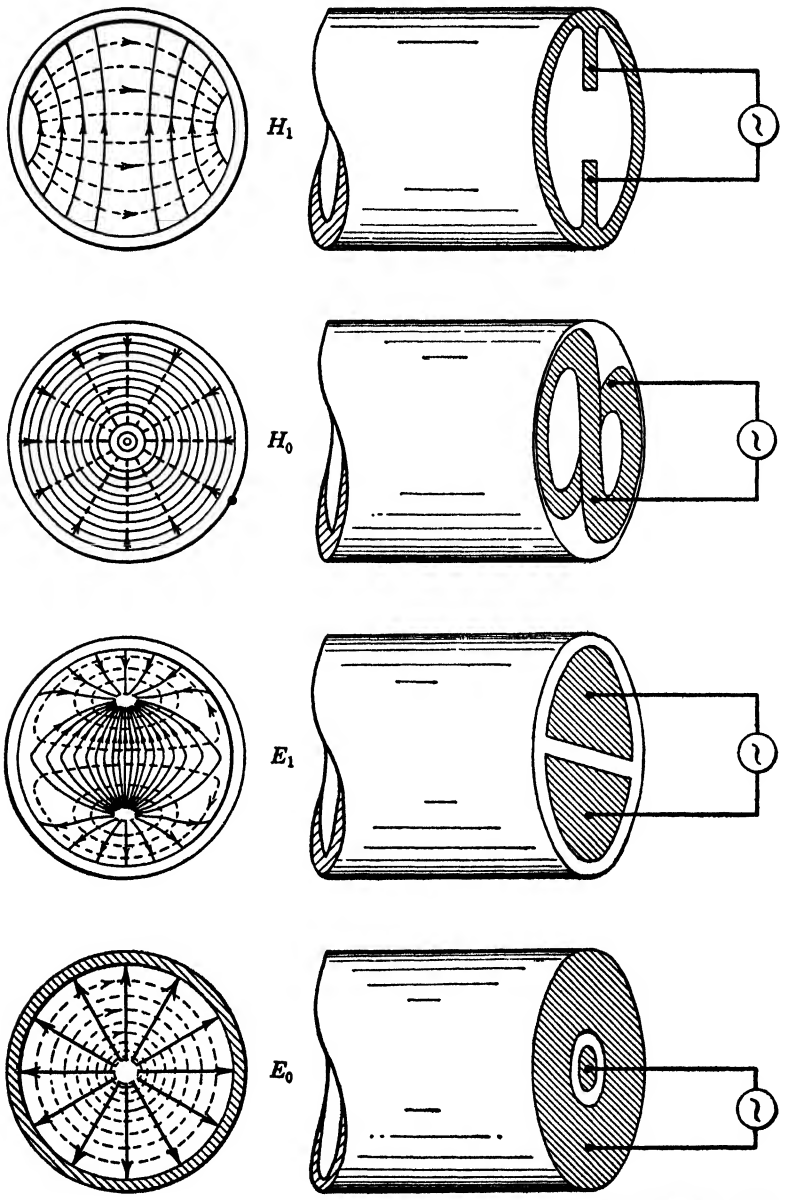
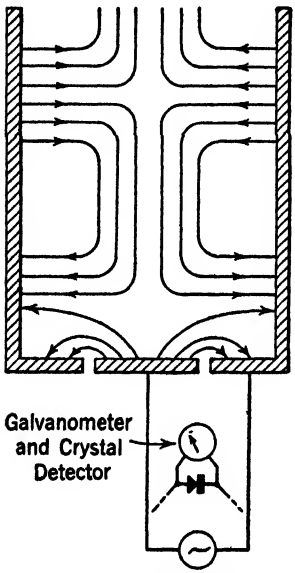


FIG. 8·20 Terminal devices designed by Southworth for exciting the H_1, H_0, E_1, and E_0 modes in circular wave guides.

tion in the same way that a reflector acts in a directive radiating system. The lower terminal in Fig. 8·17 is simply a metal disk attached to the end of the coaxial rod.

In Fig. 8·19 are shown terminal devices suggested by Barrow for exciting waves of the H_1 type in hollow tubes. It is seen that these are fundamentally different in construction from those for the E_0 waves shown in Fig. 8·17. As in the previous case, the field configuration of the wave determines the terminal design. Since the electric intensity vector of the H_1 wave is directed across the center of the pipe, the terminal "antenna" is oriented so as to coincide with these lines: in Fig. 8·19*a* a coaxial pair or open wire pair may be used to excite the rod. In Fig. 8·19*b* a coaxial line is used. Figure 8·19*c* shows the correct orientation of a configuration of parasitic antennas which may be used to reinforce the radiation. A parabolic reflector as in Fig. 8·19*d* may be used successfully externally or internally to the hollow-pipe system. In each of these cases it is necessary to provide tuning arrangements, that is, to provide facilities for the adjustment of the distance between the closed end of the pipe and the antenna. In Fig. 8·19*e* a cylindrical resonator is employed to excite the H_1 wave. The subject of resonators will be taken up in a later chapter.



(Southworth, courtesy of IRE)

FIG. 8·21 Suggested field configuration in the neighborhood of the exciter of E_0 waves.

Figure 8·20 shows Southworth's arrangement of conductors for exciting the various wave types indicated. The wave configurations excited are shown in cross-sectional view to the left of the terminal devices. A suggested field configuration in the neighborhood of the exciter of the E_0 wave is shown in Fig. 8·21. It is seen from these figures that the arrangement of the exciting system is adaptable to parallel lines, coaxial lines, or shielded pairs.

If possible, it is desirable to mount the exciting vacuum-tube oscillator directly in the hollow pipe. Such an arrangement as suggested by Southworth is shown in Fig. 8·22. The tube shown here is a spiral grid retarding field oscillator, not commercially available. Its operating principles are discussed in a later chapter, where information regarding design requirements is presented.

The primary consideration in the design of these terminal devices is to achieve a suitable impedance match between the coaxial or parallel wire line and the wave guide excited in the appropriate mode. Often it is

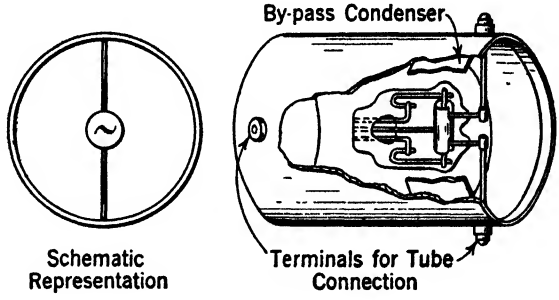


FIG. 8·22 Vacuum tube mounted directly in wave guide following Southworth's technique.

expedient to use a resonant section of coaxial line between the guide and the feeding line or to use a tuned coaxial structure as a variable shunt impedance. The exactness of this impedance match is readily tested by means of a traveling detector in the wave-guide system.

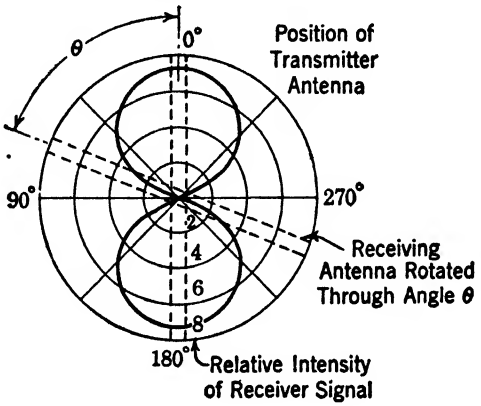
8·9 Multiplex Transmission

It was pointed out by Barrow* that multiplex transmission of waves in hollow pipes is possible. With the H_1 wave it may be demonstrated experimentally that when the orientation of the receiving terminal is at right angles to the sending terminal, and ideal conditions prevail, no voltage will be induced in the receiving antenna. When the receiving and transmitting antennas are parallel, a maximum voltage will be induced in the receiving rod. This is illustrated in Fig. 8·23, where the relative intensity of the induced voltage is plotted as a function of angular position of the receiving antenna relative to the transmitting antenna. The plane of rotation of the antenna is perpendicular to the axis of the tube. It is seen that the induced voltage is a maximum in the $\theta = 0$ or 180° positions and zero in the 90° or 270° position.

It will also be observed, when the constructional symmetry is sufficiently good, that a receiving terminal for H_0 waves will not respond to E_0 waves, and vice versa. Since both types of waves may be excited and transmitted simultaneously, and since they are mutually independent and may be received independently, Barrow has pointed out

* W. L. Barrow, "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," *I.R.E.*, **24**, 1298, 1936.

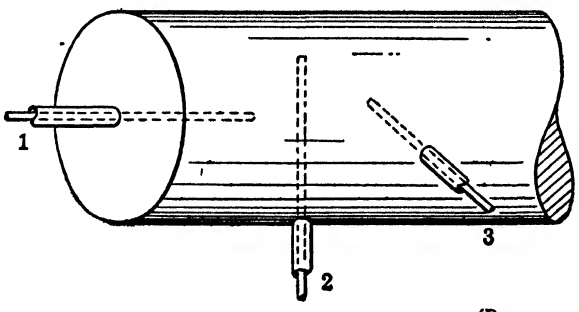
that the hollow-tube transmission system may be operated as a unique type of multiplex system. A possible arrangement for the antennas in a hollow-tube multiplex system, suggested by Barrow, is shown in



(Barrow, courtesy of IRE)

FIG. 8·23 Received signal strength vs. probe orientation for transverse magnetic waves polarized in 0° to 180° plane.

Fig. 8·24. This multiplex terminal device employs three distinct channels. Antenna 1 provides one channel which excites or receives E_0 waves. Antennas 2 and 3 provide two other channels for H_1 waves.



(Barrow, courtesy of IRE)

FIG. 8·24 Terminal device for multiplex operation of a hollow-tube transmission system. Terminal 1 excites or receives E waves, and terminals 2 and 3 excite and receive H waves. Terminals 1 and 2 are at right angles to each other. This system provides three communication channels in the same pipe.

Antenna 2 uses vertical polarization, and the antenna 3 uses horizontal polarization. Thus antennas 2 and 3 utilize the features brought out in Fig. 8·23. Such a multiplex system adds another unique feature to the practical applications of hollow-tube systems.

PROBLEMS

8·1 A section of round copper pipe of 10-cm inside diameter is to be used as a resonant chamber by closing both ends with plane copper sheets. From the phase velocity relations of Chapter 7 calculate several suitable lengths. The frequency is 2500 mc, and the interior dielectric is air. The oscillatory mode is H_1.

8·2 Repeat problem 8·1 for the H_0 mode.

8·3 The meter used on a certain hand probe has 1000 ohms resistance and a full-scale deflection of 100 microamperes. Assuming that the antenna and detector system is 20 per cent efficient, calculate the high-frequency power required to produce full-scale deflection.

8·4 A loop probe similar to that of Fig. 8·6 is used with a traveling detector. The wire of the loop is very slender and forms an area equivalent to a circle of 2-mm radius. Assuming the field to be practically uniform over the area of the loop, evaluate H', the maximum instantaneous magnetic intensity in that region, if the emf induced in the loop is 1 volt rms.

8·5 A circular wave guide is 20 cm in diameter and propagates an H_1 wave of 4×10^9 cycles per second. A straight probe 0.5 cm long is inserted radially into this guide from the outer surface in such a way as to correspond to the internal electric field. If the emf induced in the probe is 1 volt rms, calculate the power being propagated along the guide. Assume that no reflections exist.

8·6 A closed chamber is formed from a circular wave guide, and a probe is inserted in a fixed location. The piston which forms one end of the chamber is then moved so as to produce several successive maxima of deflection of the associated meter. What is the distance between the successive positions of the piston in terms of the wave propagation factors? Will the precision be relatively high, or will important second-order effects be present?

8·7 An E_0 to H_0 transducing system similar to that of Fig. 8·14 is to be designed. The frequency is 3500 mc; the guide diameter, 12 cm. Only one circular and one radial filter are to be used. Determine the optimum spacing each should have from the transducer.

8·8 Will the wave transducer system of Fig. 8·14 produce an appreciable output of either E_0 or H_0 from an incident H_1 wave? Why?

8·9 Sketch a filter and transducer system for producing H_0 from E_1 waves, and vice versa. Refer to Fig. 8·20. How does this unit illustrate the principles stated in section 8·3?

8·10 Sketch a filter and transducer system for producing H_1 from E_1 waves, and vice versa.

8·11 Sketch the probable field configuration of an H_0 wave in the vicinity of an iris in a circular wave guide. The wave guide has an inner diameter of 3 inches, and the iris is thin and has an axial hole $1\frac{1}{2}$ inches in diameter.

8·12 Design a "pinch pipe" similar to that of Fig. 8·15 to illustrate the high-pass filter characteristics of the wave guide. Show sufficient detail to make the unit a practical laboratory device.

8·13 Design a wave filter and transducer system for shifting the orientation of an H_{01} wave by 90° in a square metal wave guide. Express spacing in terms of the pipe wavelength.

8·14 Compare the theory of impedance matching in hollow wave guides with that which applies to lumped circuits at lower frequencies.

8·15 Consider the possibility of employing other systems of multiplex transmission similar to that of section 8·9.

8·16 Sketch terminal devices suitable for exciting the H_{01}, H_{02}, E_{11}, and E_{12} modes in rectangular wave guides.

8·17 Sketch terminal devices suitable for exciting TE and TM modes in the plane-parallel wave guides of Chapter 5.

8·18 Sketch terminal devices suitable for exciting E waves corresponding to equations 7·127 and 7·128 in a coaxial wave guide.

8·19 Sketch terminal devices suitable for exciting H waves corresponding to equations 7·129 and 7·130 in a coaxial wave guide.

CHAPTER 9

TRANSMISSION LINE THEORY

9·1 Introduction

Long uniform transmission lines, especially the transatlantic telegraph cables, presented a baffling problem to early communication engineers. The difficulty lay in the fact that the electrical constants of such transmission lines are distributed in space rather than lumped at one point. Accordingly the situation could not be dealt with by familiar circuit-analysis methods. Only when the brilliant minds of Heaviside, Pupin, and Kelvin were brought to bear upon the problem was a satisfactory solution reached. Fortunately the result is complete and highly accurate. We shall present such a solution in the following sections. The development is based upon the assumption of low-frequency circuit theory, but the same results may be developed directly from Maxwell's equations.

The operation of a transmission line, such as that formed by a pair of parallel round wires, may be explained in several ways. At low frequencies it is simplest to think of a current flowing out in one wire and back in the other. The product of the current and the potential difference at the sending end then defines the power delivered to the line.

It is more correct, if less convenient, to think of the parallel wire line as a special form of wave guide in which the guiding conductors are of relatively small size and the extent of the wave in the dielectric surrounding the conductors is not limited. The propagation of a wave down such a guide is illustrated in Fig. 9·1. It is assumed that the line extends indefinitely to the right or that such a termination is applied that no wave is reflected from the end of the line. The effects of reflection will be studied later.

The sectional view of Fig. 9·1 shows the distribution of the electric and magnetic fields. It may be shown that both magnetic and electric lines form systems of circles, called dipolar circles. The lines of electric and magnetic flux lie in planes perpendicular to the direction of the transmission line and are everywhere perpendicular to each other. Such a distribution represents a plane transverse electromagnetic wave propagating power in the direction of the line. Integration of the Poynting vector over any plane perpendicular to the direction of the line yields a

power transmitted along the line equal to VI, where V is the potential between the wires in that plane and I is the current in each wire at the plane. The conclusion necessarily drawn from this equality is that all the power is transmitted in the diclectric surrounding the conductors and that the conductors themselves serve only to guide the wave in the desired direction.

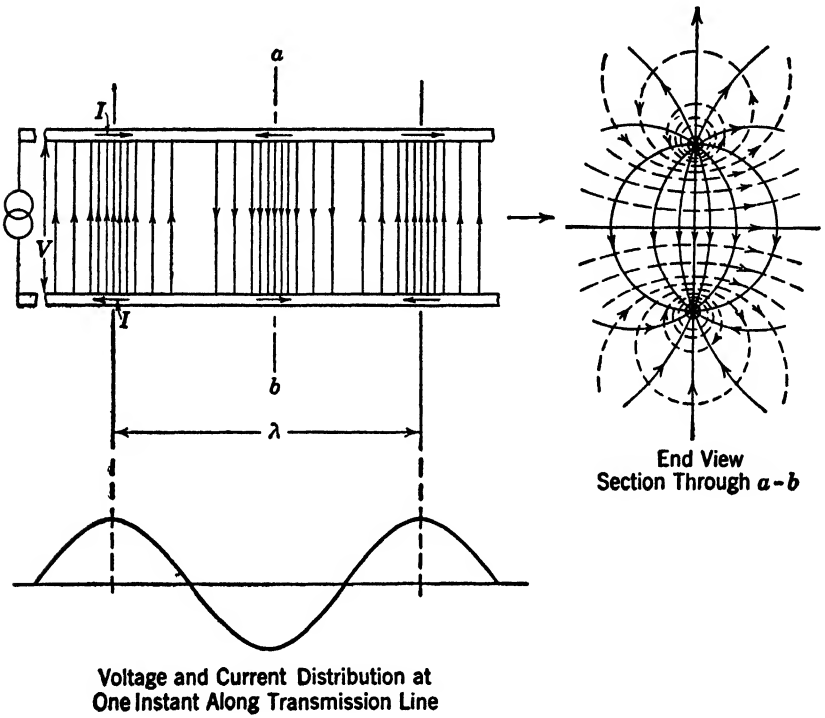


FIG. 9·1 Propagation of a wave down a pair of wires.

9·2 The Line Constants

Any form of two-conductor transmission line has four fundamental electrical constants or parameters. They are R, the series resistance; L, the series inductance; G, the shunt conductance; and C, the shunt capacitance. In general these are expressed in terms of a loop or pair of conductors of fixed length. Other constants, such as shunt inductance or series capacitance, may be shown to be unimportant. For example, any series capacitance is shunted by the conductor itself and is rendered essentially ineffective.

The significance of these parameters may be clarified by an explanation of how they may be determined. Let us consider a pair of parallel

wires a mile long supported in air on insulators. A Wheatstone bridge connected to the two wires at one end will indicate a high resistance if the other ends of the wires are open. This observation gives the shunt conductance per mile. The shunt capacitance is measured at some low frequency with the far end of the line open-circuited. In this measurement the effect of the shunt conductance must be eliminated by means of a suitable technique. The series inductance and series resistance are measured by short-circuiting the far end of the line.

It will immediately be recognized that the procedure just described is not exact. The shunt conductance of the line causes the observation of series resistance to be low and the series resistance causes the observation of shunt conductance to be too low. Similarly, the inductance of the line affects the capacitance observation, and vice versa. In normal lines, however, the conductance is so small compared to the series resistance that the error is negligible. And the observations on inductance and capacitance are less and less interdependent as the frequency is lowered. In any event the errors approach zero as the length of the line under measurement approaches zero. It is then possible to express these parameters in terms of a mile length even though they are measured on a specimen much shorter than one mile. It may be shown that these measurements of impedance with the far end alternately open- and short-circuited suffice to determine the transmission properties of any network.

9·3 Derivation of the Telegrapher's Equations

Let us examine the behavior of an infinitesimal length, dl, of an infinitely long transmission line. It will be assumed that the basic line parameters, R, L, G, and C, are known and specified in terms of some unit length such as the meter. The length dl is then referred to this same unit. For convenience we may break the line down into a succession of these equal infinitesimal lengths and in turn separate the resistive and reactive components of the impedance in each unit length. The resulting structure is shown in Fig. 9·2.

The sections, which are all alike, are characterized by a shunt conductance $dG = G\,dl$ and a shunt capacitance $dC = C\,dl$. The series elements apply to the pair of wires. That is, $dL = L\,dl$ and $dR = R\,dl$.

The behavior of the line may be deduced from the behavior of any one elementary section. Some definite input current and voltage are assumed, and the output current and voltage are deduced from the constants of the section. Since the line sections are already reduced to infinitesimal sections we may neglect the shunt elements in the computation of the effect of the series elements, and vice versa. Since we shall confine our attention to sinusoidal waves we may replace the series

elements by the impedance $Z\,dl$ and the shunt elements by the admittance $Y\,dl$. The behavior of a line when the impressed voltage is non-sinusoidal may be calculated by separating the voltage wave into a series of sinusoidal waves by Fourier analysis. The total effect produced is then the sum of the effects produced by these sinusoidal components.

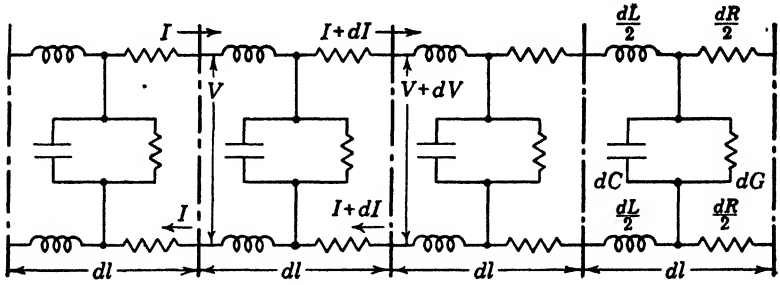


FIG. 9·2 Infinitesimal section of a transmission line expressed in terms of lumped parameters.

In each section of the line the voltage is decreased because of the series impedance, and the current is decreased by the action of the shunt admittance. Accordingly we may write

$$dI = -VY\,dl \qquad [9\cdot1]$$

and

$$dV = -IZ\,dl \qquad [9\cdot2]$$

where $Z = R + jX$ and $Y = G + jB$

Expressed as derivatives these become

$$\frac{dI}{dl} = -VY \qquad [9\cdot3]$$

$$\frac{dV}{dl} = -IZ \qquad [9\cdot4]$$

Differentiating 9·3 and 9·4 with respect to l

$$\frac{d^2I}{dl^2} = -Y\frac{dV}{dl} \qquad [9\cdot5]$$

$$\frac{d^2V}{dl^2} = -Z\frac{dI}{dl} \qquad [9\cdot6]$$

Substitution of 9·3 and 9·4 in 9·5 and 9·6 gives the symmetrical equations

$$\frac{d^2I}{dl^2} = YZI \qquad [9\cdot7]$$

$$\frac{d^2V}{dl^2} = YZV \qquad [9\cdot8]$$

9·4 Solution of the Telegrapher's Equations

Equations 9·7 and 9·8 are commonly referred to as the telegrapher's equations since they were first developed in connection with long-distance telegraphy over cables. Because of their symmetry it is evident that these equations have solutions of the same form. Therefore the solution of only one will be sought, the solution of the other being written down from similarity.

Here, as in most differential equations of electricity, we assume a solution of exponential form. That is, the form of the solution is assumed, and the appropriate constants are developed. Let

$$I = Ae^{\gamma l} \qquad [9\cdot9]$$

then

$$\frac{dI}{dl} = \gamma Ae^{\gamma l} \qquad [9\cdot10]$$

and

$$\frac{d^2I}{dl^2} = \gamma^2 Ae^{\gamma l} \qquad [9\cdot11]$$

To satisfy 9·7 it is necessary that

$$\gamma^2 Ae^{\gamma l} = ZYAe^{\gamma l} \qquad [9\cdot12]$$

Or merely that

$$\gamma^2 = ZY, \text{ giving } \gamma = \pm\sqrt{ZY} \qquad [9\cdot13]$$

The complete solution of equations 9·7 and 9·8 may be written

$$I = A_1e^{\gamma l} + B_1e^{-\gamma l} \qquad [9\cdot14]$$

$$V = A_2e^{\gamma l} + B_2e^{-\gamma l} \qquad [9\cdot15]$$

where γ is now defined as the positive square root of ZY.

The parameters A_1, A_2, B_1, B_2 are constants of integration. They may be evaluated by differentiating 9·14 and 9·15.

$$\frac{dI}{dl} = \gamma A_1e^{\gamma l} - \gamma B_1e^{-\gamma l} \qquad [9\cdot16]$$

$$\frac{dV}{dl} = \gamma A_2e^{\gamma l} - \gamma B_2e^{-\gamma l} \qquad [9\cdot17]$$

Application of the basic equations 9·3 and 9·4 to 9·16 and 9·17 gives

$$-VY = \gamma A_1e^{\gamma l} - \gamma B_1e^{-\gamma l} \qquad [9\cdot18]$$

$$-IZ = \gamma A_2e^{\gamma l} - \gamma B_2e^{-\gamma l} \qquad [9\cdot19]$$

Equations 9·18 and 9·19 express the behavior of any line subject to existing boundary conditions. It is convenient to consider the line connected to a physical generator at the *sending end* and to measure the distance from the generator. At the sending end of the line we have $l = 0$, $V = V_s$, and $I = I_s$. Equations 9·14, 9·15, 9·18, and 9·19, respectively, give us

$$I_s = A_1 + B_1 \qquad [9\cdot20]$$

$$V_s = A_2 + B_2 \qquad [9\cdot21]$$

$$-V_sY = \gamma A_1 - \gamma B_1 \qquad [9\cdot22]$$

$$-I_sZ = \gamma A_2 - \gamma B_2 \qquad [9\cdot23]$$

Solving for A_1, A_2, B_1, and B_2 in terms of Y, Z, I_s, V_s, and γ, we obtain

$$A_1 = \tfrac{1}{2}\left(I_s - V_s\frac{Y}{\gamma}\right) = \tfrac{1}{2}\left(I_s - V_s\sqrt{\frac{Y}{Z}}\right) \qquad [9\cdot24]$$

$$B_1 = \tfrac{1}{2}\left(I_s + V_s\frac{Y}{\gamma}\right) = \tfrac{1}{2}\left(I_s + V_s\sqrt{\frac{Y}{Z}}\right) \qquad [9\cdot25]$$

$$A_2 = \tfrac{1}{2}\left(V_s - I_s\frac{Z}{\gamma}\right) = \tfrac{1}{2}\left(V_s - I_s\sqrt{\frac{Z}{Y}}\right) \qquad [9\cdot26]$$

$$B_2 = \tfrac{1}{2}\left(V_s + I_s\frac{Z}{\gamma}\right) = \tfrac{1}{2}\left(V_s + I_s\sqrt{\frac{Z}{Y}}\right) \qquad [9\cdot27]$$

We have already defined $\gamma = \sqrt{ZY}$. Let us introduce the additional definition $Z_0 = \sqrt{Z/Y}$.

In terms of this definition

$$A_1 = \tfrac{1}{2}\left(I_s - \frac{V_s}{Z_0}\right) \qquad [9\cdot28]$$

$$B_1 = \tfrac{1}{2}\left(I_s + \frac{V_s}{Z_0}\right) \qquad [9\cdot29]$$

$$A_2 = \tfrac{1}{2}(V_s - I_sZ_0) \qquad [9\cdot30]$$

$$B_2 = \tfrac{1}{2}(V_s + I_sZ_0) \qquad [9\cdot31]$$

Substituting 9·30 and 9·31 in 9·14 and 9·15,

$$I = \tfrac{1}{2}I_se^{\gamma l} - \tfrac{1}{2}\frac{V_s}{Z_0}e^{\gamma l} + \tfrac{1}{2}I_se^{-\gamma l} + \tfrac{1}{2}\frac{V_s}{Z_0}e^{-\gamma l} \qquad [9\cdot32]$$

$$V = \tfrac{1}{2}V_se^{\gamma l} - \tfrac{1}{2}I_sZ_0e^{\gamma l} + \tfrac{1}{2}V_se^{-\gamma l} + \tfrac{1}{2}I_sZ_0e^{-\gamma l} \qquad [9\cdot33]$$

Rearranging terms and introducing the standard definitions

$$\frac{e^x + e^{-x}}{2} = \cosh x \qquad [9\cdot34]$$

and

$$\frac{e^x - e^{-x}}{2} = \sinh x \qquad [9\cdot35]$$

we obtain

$$I = I_s \cosh \gamma l - \frac{V_s}{Z_0} \sinh \gamma l \qquad [9\cdot36]$$

$$V = V_s \cosh \gamma l - I_s Z_0 \sinh \gamma l \qquad [9\cdot37]$$

where l is measured from the *sending end* of the line, $\gamma = \sqrt{ZY}$, and $Z_0 = \sqrt{Z/Y}$. These equations, which are perfectly general within the limits of the original assumptions, contain a wealth of information. They apply to any transmission-line system propagating transverse electromagnetic waves. Since at the present time no other transmission modes are of practical significance we may regard these equations as descriptive of all the phenomena of transmission lines.

A similar and equally basic pair of equations are developed from equations 9·14, 9·15, 9·18, and 9·19 by considering the *receiving end* of an infinite transmission line. The subscript r will be used to designate voltages and currents at the receiving end. The constants are evaluated by the same process, and a similar mathematical procedure leads us to the following:

$$I = I_r \cosh \gamma l + \frac{V_r}{Z_0} \sinh \gamma l \qquad [9\cdot38]$$

$$V = V_r \cosh \gamma l + I_r Z_0 \sinh \gamma l \qquad [9\cdot39]$$

where l is measured from the *receiving end* of the line, $\gamma = \sqrt{ZY}$, and $Z_0 = \sqrt{Z/Y}$.

9·5 Impedance of the Infinite Line

Equation 9·37 may be interpreted to define the input impedance of an infinitely long line.

Dividing through by $\cosh \gamma l$ we have

$$\frac{V}{\cosh \gamma l} = V_s - I_s Z_0 \frac{\sinh \gamma l}{\cosh \gamma l} \qquad [9\cdot40]$$

Referring to equation 9·34 we see that $\cosh \gamma l$ approaches ∞ as l approaches infinity. From equation 9·35 we see that $\sinh \gamma l$ also approaches ∞ as l approaches infinity. The ratio $\sinh \gamma l / \cosh \gamma l$ ap-

proaches unity, however; and therefore we have

$$0 = V_s - I_s Z_0 \qquad [9\cdot41]$$

$$\frac{V_s}{I_s} = Z_0 = \sqrt{Z/Y} \qquad [9\cdot42]$$

That is, a finite impedance known as the characteristic impedance of the line is observed at the end of an infinitely long line regardless of the conditions at the other end of the line. This same conclusion is reached by appropriate manipulation of equations 9·36, 9·38, or 9·39.

With the qualification that the line is very long and using the equality of 9·42 we may rewrite 9·36 and 9·37 in the form

$$I = I_s (\cosh \gamma l - \sinh \gamma l) = I_s e^{-\gamma l} \qquad [9\cdot43]$$

and

$$V = V_s (\cosh \gamma l - \sinh \gamma l) = V_s e^{-\gamma l} \qquad [9\cdot44]$$

again referring to equations 9·34 and 9·35 for the relation of the hyperbolic functions.

The symbol Z_0 is variously referred to as the characteristic impedance, surge impedance, image impedance, and iterative impedance of the line. The symbol γ is universally referred to as the propagation constant because it describes the propagation of a wave along the line. Both are essentially complex numbers although it happens that Z_0 is nearly a pure real and γ nearly a pure imaginary for many important practical cases.

9·6 The Propagation Constant

We have defined $\gamma = \sqrt{ZY}$. We have in addition $Z = R + jX$ and $Y = G + jB$, and we may define $\gamma = \alpha + j\beta$, where α and β are respectively the attenuation and phase constants of the line. The resulting equation is

$$\alpha + j\beta = \gamma = \sqrt{(R + jX)(G + jB)} \qquad [9\cdot45]$$

Squaring and collecting real and imaginary terms,

$$\alpha^2 - \beta^2 + 2j\alpha\beta = RG - BX + j(RB + GX) \qquad [9\cdot46]$$

Such an equation is correct only if the real and imaginary terms are separately equal. That is

$$\alpha^2 - \beta^2 = RG - BX \qquad [9\cdot47]$$

$$2\alpha\beta = RB + GX \qquad [9\cdot48]$$

Solving 9·48 for α we have

$$\alpha = \frac{RB + GX}{2\beta} \qquad [9\cdot49]$$

Substituting 9·49 in 9·47,

$$\frac{(RB + GX)^2}{4\beta^2} - \beta^2 = RG - BX \qquad [9\cdot50]$$

giving

$$\frac{(RB + GX)^2}{4} - \beta^4 = (RG - BX)\beta^2 \qquad [9\cdot51]$$

or

$$\beta^4 + \beta^2(RG - BX) - \frac{(RB + GX)^2}{4} = 0 \qquad [9\cdot52]$$

Solving for β^2 by the quadratic formula and extracting the root

$$\beta = \sqrt{\tfrac{1}{2}[\sqrt{(R^2 + X^2)(G^2 + B^2)} - (RG - BX)]}$$

By a similar development

$$\alpha = \sqrt{\tfrac{1}{2}[\sqrt{(R^2 + X^2)(G^2 + B^2)} + (RG - BX)]}$$

Substituting $X = \omega L$ and $B = \omega C$, these equations become

$$\beta = \sqrt{\tfrac{1}{2}[\sqrt{(R^2 + \omega^2L^2)(G^2 + \omega^2C^2)} - (GR - \omega^2LC)]} \qquad [9\cdot53]$$

$$\alpha = \sqrt{\tfrac{1}{2}[\sqrt{(R^2 + \omega^2L^2)(G^2 + \omega^2C^2)} + (GR - \omega^2LC)]} \qquad [9\cdot54]$$

In general, the dissipation associated with physical lines is small. In most physical lines R is small compared with ωL and G is small compared with ωC. Under these circumstances, it is convenient to write equation 9·45 as

$$\alpha + j\beta = j\omega\sqrt{LC}\left(1 + \frac{R}{j\omega L}\right)^{\frac{1}{2}}\left(1 + \frac{G}{j\omega C}\right)^{\frac{1}{2}}$$

Expanding the latter terms in powers of $\dfrac{R}{j\omega L}$ and $\dfrac{G}{j\omega C}$ according to the binomial theorem,* we obtain

$$\alpha + j\beta = j\omega\sqrt{LC}\left\{1 + \frac{R}{2j\omega L} - \frac{R^2}{8j^2\omega^2L^2} + \cdots\right\}\left\{1 + \frac{G}{2j\omega C} - \frac{G^2}{8j^2\omega^2C^2} + \cdots\right\}$$

The binomial theorem may be expressed as

$$(a + b)^n = k(1 + x)^n = k\left(1 + nx + \frac{n(n-1)}{2!}x^2 + \cdots\right)$$

where the series is rapidly convergent for $|x| \ll 1$. A complete discussion of this theorem may be found in any advanced algebra book.

or, neglecting squared terms and transferring j inside the parentheses

$$\alpha + j\beta \simeq \omega\sqrt{LC}\left(j + \frac{R}{2\omega L} + \frac{G}{2\omega C}\right)$$

Equating real and imaginary parts

$$\alpha \simeq \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \qquad [9\cdot55]$$

and

$$\beta \simeq \omega\sqrt{LC} \qquad [9\cdot56]$$

In these equations all quantities must be expressed in terms of the same unit distance. If the line constants are expressed per loop mile the constants α and β will be in terms of nepers per mile and radians per mile, respectively.

Equations 9·43 and 9·44 are now rewritten

$$I = I_s e^{-\alpha l} e^{-j\beta l} \qquad [9\cdot57]$$

$$V = V_s e^{-\alpha l} e^{-j\beta l} \qquad [9\cdot58]$$

If we substitute $V_s = V'_s e^{j\omega t}$, where V'_s is the maximum time variation of V_s, in equation 9·58, we obtain

$$V = V'_s e^{-\alpha l} e^{j\omega t} e^{-j\beta l} = V'_s e^{-\alpha l} e^{j(\omega t - \beta l)} \qquad [9\cdot59]$$

Equation 9·59 is recognized as a wave equation, very similar to those previously considered. The attenuation constant α is simply a multiplying factor which acts to diminish the magnitude of the wave as it progresses. The term $\omega t - \beta l$ establishes the wave motion and also that it is sinusoidal in both time and space variation.

If the applied voltage is non-sinusoidal it is resolved into sinusoidal components by Fourier analysis. The separate voltage waves which result establish these separate traveling waves on the line, the actual wave being the sum of these separate components by the principle of superposition.

Figure 9·3 shows the wave motion on a line for two successive instants. At time $t = 0$ the input voltage is zero and a certain definite wave system exists along the wire. At a time corresponding to a quarter of a cycle later the input voltage has reached a negative maximum and the entire wave system has moved to the right a distance $\lambda/4$.

The velocity of propagation is obtained by setting $(\omega t - \beta l)$ equal to a constant. Then

$$(\omega t - \beta l) = \omega(t + \Delta t) - \beta(l + \Delta l) \qquad [9\cdot60]$$

or

$$0 = \omega\Delta t - \beta\Delta l \qquad [9\cdot61]$$

giving

$$\frac{\omega}{\beta} = \frac{\Delta l}{\Delta t} = v \qquad [9\cdot62]$$

Evidently equation 9·53 expressing the phase constant in terms of ω can be combined with 9·62 to give an explicit solution for the velocity. It should be noted that the velocity given by 9·62 is a phase velocity.

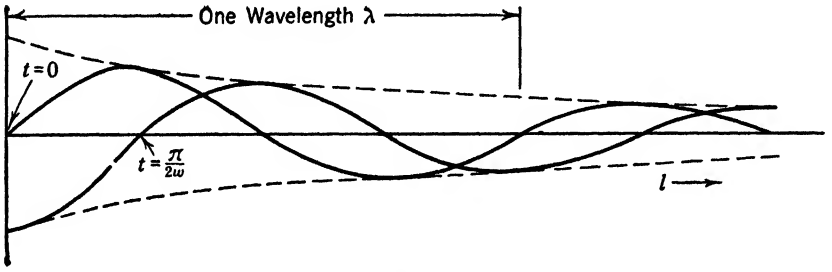


FIG. 9·3 Wave motion on a line having attenuation. (Two successive instants of time shown.)

For most radio transmission lines, however, v is independent of ω, and therefore the signal and group velocities are equal to the phase velocity.

The wavelength is given by the basic equation

$$v = f\lambda \qquad [9\cdot63]$$

Combining with 9·62 to eliminate v we obtain

$$\frac{\omega}{\beta} = \frac{2\pi f}{\beta} = f\lambda \qquad [9\cdot64]$$

or

$$\lambda = \frac{2\pi}{\beta} \qquad [9\cdot65]$$

9·7 Lines of Finite Length

We have already seen that an infinite transmission line presents a finite impedance equal to Z_0 to the generator at the sending end. This impedance will not be altered if a finite amount of line is inserted between the generator and the infinite line since a finite addition does not affect an infinite quantity. From this we must conclude that the finite section of a line if terminated in a lumped impedance equal to Z_0 has an input impedance again equal to Z_0.

In Fig. 9·4 the behavior is very simple and regular. The wave sent out from the generator arrives at the receiving termination somewhat later and is completely absorbed. If, however, the termination is unequal to the characteristic impedance the behavior is much more complex.

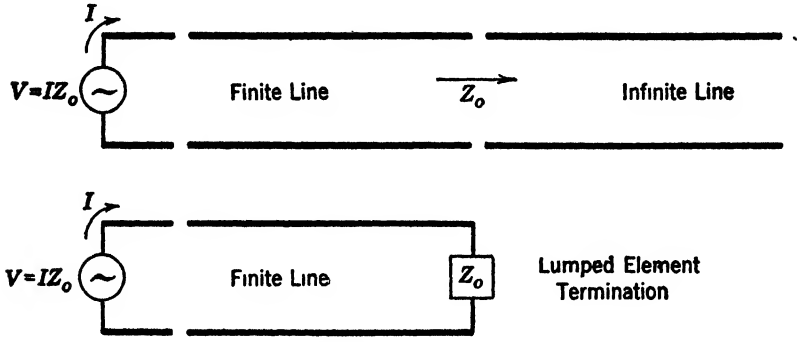


FIG. 9·4 Equivalence of finite terminated line to an infinite line.

Let us consider an example in which the receiving terminals are short-circuited; that is, the receiving impedance is equal to zero. Under these conditions we may most conveniently set $V_r = 0$ in equations 9·38 and 9·39. The current and voltage at any distance l from the *receiving end* are then

$$I = I_r \cosh \gamma l \qquad [9\cdot66]$$

$$V = I_r Z_0 \sinh \gamma l \qquad [9\cdot67]$$

The sending end impedance for any particular section of line is simply the ratio of these two quantities evaluated for the appropriate length l.

$$Z_{sc} = \frac{V}{I} = Z_0 \frac{\sinh \gamma l}{\cosh \gamma l} = Z_0 \tanh \gamma l \qquad [9\cdot68]$$

An alternative condition of equal interest is that in which the receiving terminals are open-circuited. Here $I_r = 0$, and equations 9·38 and 9·39 become

$$I = \frac{V_r}{Z_0} \sinh \gamma l \qquad [9\cdot69]$$

$$V = V_r \cosh \gamma l \qquad [9\cdot70]$$

The sending end impedance is now

$$Z_{oc} = \frac{V}{I} = Z_0 \coth \gamma l \qquad [9\cdot71]$$

9·8 Reflection

We saw in Chapter 4 that a plane electromagnetic wave is reflected, at least in part, at any discontinuity of the medium through which it is propagated. Sometimes the entire incident energy is reflected. In extreme cases virtually none is reflected, and all intermediate degrees are possible. A very similar situation exists in respect to waves on lines. If the transmission line is perfectly uniform or smooth there is no reflection and the entire energy is propagated normally. If some irregularity exists, as for example a high-resistance splice or a change of mechanical dimension, a reflection of smaller or larger magnitude occurs. Part of the incident energy is returned toward the sending end, the remainder proceeding along the line.

Complete discontinuity exists if the line is open-circuited or short-circuited. The open-circuit case is somewhat analogous to total internal reflection previously discussed; the short-circuit case is very similar to reflection of a wave in air by a metal surface. In both, the entire incident energy is reflected.

Reflection from a short circuit is most easily treated if we consider a line of infinite extent with a short circuit applied at some finite length from the sending end. By definition the voltage at the short circuit is zero. This condition is satisfied if we assume a voltage equal and opposite to the incident voltage wave. The conditions are shown in Fig. 9·5. A wave traveling to the left in the figure having such a magnitude and phase as to cancel the incident wave at the short circuit is assumed. Since the line has attenuation this wave becomes smaller as it travels to the left, just as the original wave is attenuated as it travels to the right. The auxiliary waves assumed to the right of the short circuit have no physical significance and are shown for geometrical purposes. In fact the reflection is not altered if the line to the right of the short circuit is removed.

The sketches of Fig. 9·5 are all in terms of instantaneous voltages and therefore do not represent readings of meters. Figure 9·6 is drawn in terms of effective voltage and currents. It therefore represents a phenomenon which may be observed with instruments.

The curves of Fig. 9·6 are readily developed by a rigorous mathematical argument. From equation 9·39 we have

$$V = V_r \cosh \gamma l + I_r Z_0 \sinh \gamma l \qquad [9\cdot39]$$

where l is measured from the *receiving end* of the line, in this case, the short-circuited end. Accordingly the resulting equation will be expressed in terms of distance from the short circuit.

For a short-circuited line $V_r = 0$. In section 9·7 we have shown that

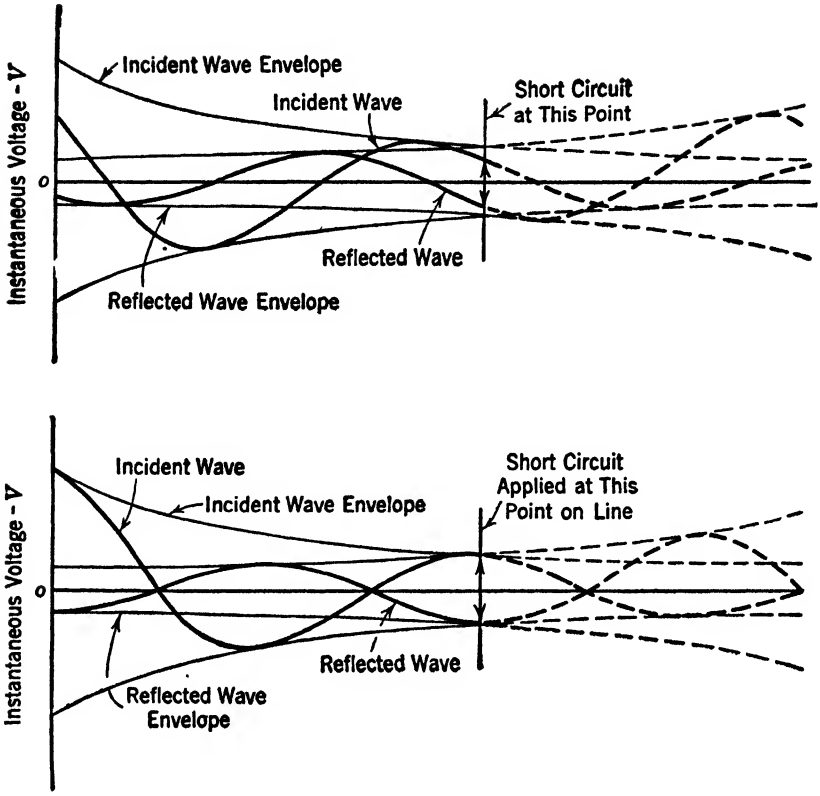


FIG. 9·5 Reflection from short circuit at two different instants of time.

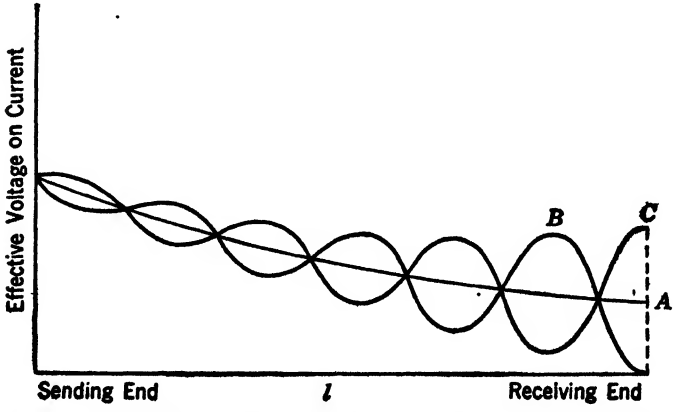


FIG. 9·6 Voltage and current distribution of line having attenuation.

the voltage at any distance l from a short circuit may be expressed in terms of the receiving current through the short circuit by the relation

$$V = I_r Z_0 \sinh \gamma l \qquad [9\cdot67]$$

Substituting $\gamma = \alpha + j\beta$

$$V = I_r Z_0 \sinh (\alpha l + j\beta l)$$

or

$$V = I_r Z_0 (\sinh \alpha l \cosh j\beta l + \cosh \alpha l \sinh j\beta l)$$

But

$$\cosh j\beta l = \cos \beta l \qquad \text{and} \qquad \sinh j\beta l = j \sin \beta l$$

Hence

$$V = I_r Z_0 (\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l)$$

The voltage V is seen to have a real part, $I_r Z_0 \sinh \alpha l \cos \beta l$, and an imaginary part, $I_r Z_0 \cosh \alpha l \sin \beta l$. The absolute magnitude of V is thus the square root of the sum of the squares of the real and imaginary parts.

$$|V| = I_r Z_0 \sqrt{\sinh^2 \alpha l \cos^2 \beta l + \cosh^2 \alpha l \sin^2 \beta l}$$

We now use the identities

$$\cosh^2 \theta = \sinh^2 \theta + 1$$

and

$$\cos^2 X + \sin^2 X = 1$$

giving

$$|V| = I_r Z_0 \sqrt{\sinh^2 \alpha l + \sin^2 \beta l}$$

This function is zero for $l = 0$ as required, and is never negative. As l is increased the first term steadily increases in size, whereas the second term oscillates with constant magnitude.

The effect of the undulations diminishes rapidly with distance because of the effect of extracting the square root. For $\sinh \alpha l = 10$ and $\sin \beta l = \pm 1$ (its maximum value) the deviation is only from $\sqrt{101}$ to $\sqrt{99}$, a matter of $\pm$ one-half of one per cent.

The curves of Fig. 9·6 are considerably more general than has so far been stated. Curve A is a simple exponential, representing equation 9·57 or 9·58, and applies to the voltage or current in a line terminated in its characteristic impedance. Curve B represents the relation just developed and applies to the voltage on a short-circuit line or the current in an open-circuit line. Curve C represents the function

$\sqrt{\sinh^2 \alpha l + \cos^2 \beta l}$ and applies to the voltage on an open-ciruit line or the current in a short-circuit line.

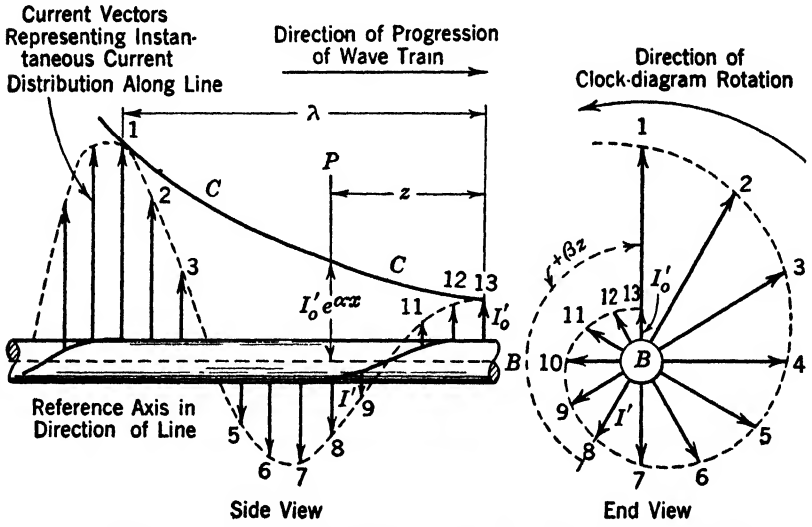


FIG. 9·7 Sketch of a physical model of wave propagation on a transmission line.

An excellent concept of the propagation of waves on a transmission line is gained by study of a circuit-vector model in three dimensions. Figure 9·7 is a sketch of such a model. Figure 9·8 is a perspective drawing of the same model.

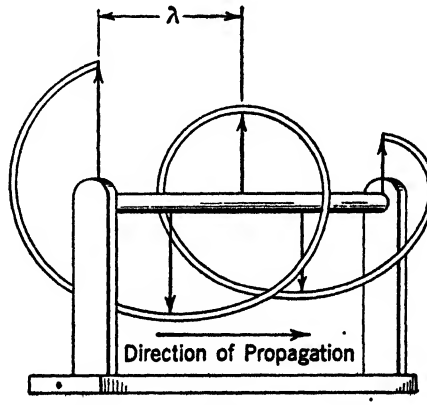


FIG. 9·8 Sketch of a physical model of wave propagation on a transmission line.

9·9 Dissipationless Lines

At radio frequencies suitable lines have the property that $R \ll L$ and $G \ll C$. Such lines have very low attenuation and may be used as

resonators, to good advantage. Let us examine the general equations as simplified by the approximation $R = G = 0$. Referring to equations 9·55 and 9·56 we find

$$\beta = \omega\sqrt{LC} \qquad [9·72]$$

and

$$\alpha = 0 \qquad [9·73]$$

Substituting in equations 9·38 and 9·39

$$I = I_r \cosh j\beta l + \frac{V_r}{Z_0} \sinh j\beta l \qquad [9·74]$$

$$V = V_r \cosh j\beta l + I_r Z_0 \sinh j\beta l \qquad [9·75]$$

But* $\cosh jX = \cos X$ and $\sinh jX = j \sin X$. Accordingly we write

$$I = I_r \cos \beta l + j \frac{V_r}{Z_0} \sin \beta l \qquad [9·76]$$

$$V = V_r \cos \beta l + jI_r Z_0 \sin \beta l \qquad [9·77]$$

The characteristic impedance $\sqrt{Z/Y}$ takes the simple form

$$Z_0 = \sqrt{\frac{L}{C}} \qquad [9·78]$$

The input impedance of a section of dissipationless line terminated in an arbitrary receiving impedance Z_r is of great interest since it is readily simplified to yield all the familiar special cases. Given the receiving impedance Z_r, we have

$$I_r Z_r = V_r \qquad [9·79]$$

where Z_r, the receiving impedance, may have any magnitude and phase angle. Accordingly 9·76 and 9·77 become

$$I = I_r \cos \beta l + j \frac{I_r Z_r}{Z_0} \sin \beta l \qquad [9·80]$$

$$V = I_r Z_r \cos \beta l + jI_r Z_0 \sin \beta l \qquad [9·81]$$

The ratio V/I defines the input impedance of a section of line of length l. This ratio is

$$\frac{V}{I} = Z_{in} = \frac{Z_r \cos \beta l + jZ_0 \sin \beta l}{\cos \beta l + j \dfrac{Z_r}{Z_0} \sin \beta l} \qquad [9·82]$$

* These identities follow from the definitions of the functions

$$\cosh x = \frac{e^x + e^{-x}}{2}; \cos x = \frac{e^{jx} + e^{-jx}}{2}; \sinh x = \frac{e^x - e^{-x}}{2}; \text{ and } \sin x = \frac{e^{jx} - e^{-jx}}{2j}.$$

which is more symmetrical in the form

$$Z_{in} = Z_0 \frac{Z_r \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_r \sin \beta l} \qquad [9\cdot83]$$

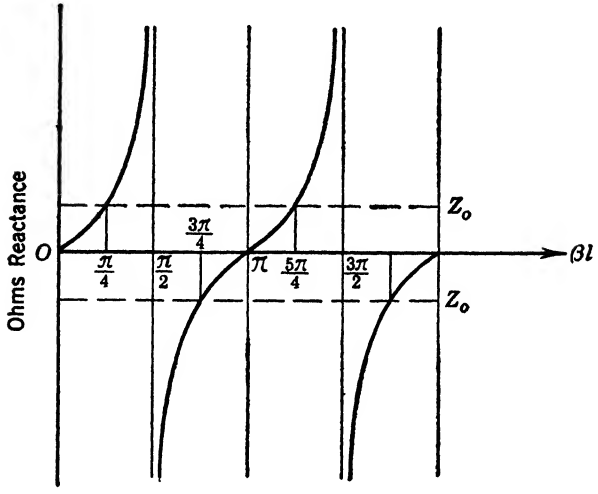


FIG. 9·9 Input impedance of a finite short-circuited dissipationless line.

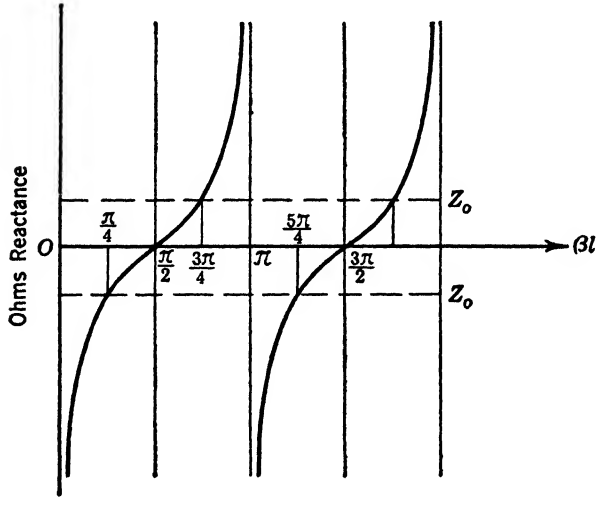


FIG. 9·10 Input impedance of a finite open-circuited dissipationless line.

With the special case for $Z_r = 0$, a short circuit

$$Z_{in} = jZ_0 \tan \beta l \qquad [9\cdot84]$$

And if $Z_r = \infty$ corresponding to an open circuit

$$Z_{in} = -jZ_0 \cot \beta l \qquad [9\cdot85]$$

Equations 9·84 and 9·85 are plotted in Figs. 9·9 and 9·10.

The abscissa of Figs. 9·9 and 9·10 may be interpreted equally well as length of line or as frequency since the two enter as a product in the expression βl. It is the similarity of these curves to curves of reactance *vs.* frequency of lumped reactive networks which permits us to use sections of transmission lines as resonators.

9·10 Equivalence of Dissipationless Lines to Resonant Circuits

It is well known that an open-circuited line a quarter wavelength long is approximately equivalent to a single coil and condenser in series. Let us examine the conditions under which this is true and the degree of the approximation which results. Referring again to equation 9·85 and using the identity of equation 9·62 the input impedance of the line is a pure reactance given by

$$X_{in} = -Z_0 \cot \omega \frac{l}{v} \qquad [9\cdot86]$$

From Fig. 9·10 it is seen that X_{in} is zero for an indefinite number of values of $\omega l/v$. The first zero, which corresponds to $\omega l/v = \pi/2$, is the one of greatest interest to us. We may define the resonant angular velocity for the line from the relation

$$\omega = \omega_0 = \frac{\pi v}{2l} \qquad [9\cdot87]$$

The impedance of a coil L in series with a condenser C is a pure reactance given by the expression

$$X = \omega L - \frac{1}{\omega C} \qquad [9\cdot88]^*$$

We wish to choose L and C in terms of the line properties l, v, and Z_0 so that the two impedances are closely equal, at least in the region of resonance. The resonant frequencies are made to correspond if

$$\omega_0 = \frac{\pi v}{2l} = \frac{1}{\sqrt{LC}} \qquad [9\cdot89]$$

The two curves will cross the axis with the same slope if $dX/d\omega$ is the same for the two. From equation 9·86 which applies to the trans-

* The symbols L and C here refer to a lumped network equivalent to the line and must not be confused with the basic line constants.

mission line we have

$$\frac{dX_{in}}{d\omega} = \frac{l}{v} Z_0 \csc^2 \frac{\omega l}{v} \qquad [9\cdot90]$$

which at $\omega = \omega_0 = \pi v/2l$ reduces to

$$\frac{dX_{in}}{d\omega} = \frac{lZ_0}{v} \qquad [9\cdot91]$$

For the lumped resonant circuit

$$\frac{dX}{d\omega} = L + \frac{1}{\omega^2 C} \qquad [9\cdot92]$$

which at $\omega = \omega_0 = 1/\sqrt{LC}$ reduces to

$$\frac{dX}{d\omega} = 2L \qquad [9\cdot93]$$

Using the equalities of the resonant frequency and the slope we may write

$$\omega_0 = \frac{\pi v}{2l} = \frac{1}{\sqrt{LC}} \qquad [9\cdot89]$$

and

$$\frac{dX}{d\omega} = 2L = \frac{lZ_0}{v} \qquad [9\cdot94]$$

Let us again introduce the relation $v = f\lambda$, which in this case is $v = f_0\lambda_0$, where f_0 and λ_0 are the resonant frequency and wavelength, respectively. From 9·87 we have

$$\omega_0 = 2\pi f_0 = \frac{\pi f_0 \lambda_0}{2l} \qquad [9\cdot95]$$

which reduces to

$$l = \frac{\lambda_0}{4} \qquad [9\cdot96]$$

Equation 9·96 establishes the fact that the transmission line is resonant when its length is one-fourth of a wavelength. Equation 9·94 may be written

$$L = \frac{lZ_0}{2v} = \frac{lZ_0}{2f_0\lambda_0} = \frac{\pi Z_0}{4\omega_0} \qquad [9\cdot97]$$

Substituting 9·97 in 9·89 we find

$$C = \frac{4}{\pi\omega_0 Z_0} = \frac{8l}{\pi^2 v Z_0} \qquad [9\cdot98]$$

Equivalence of open-circuit line and series resonant circuit

Equations 9·97 and 9·98 define the values of a coil L and a condenser C which, when connected in series, are equivalent to a line that is open-circuited and one-fourth wavelength long. The impedance curves resulting in the two cases are plotted in Fig. 9·11. It is seen that the approximation is excellent in the region of resonance but fails at higher or lower frequencies.

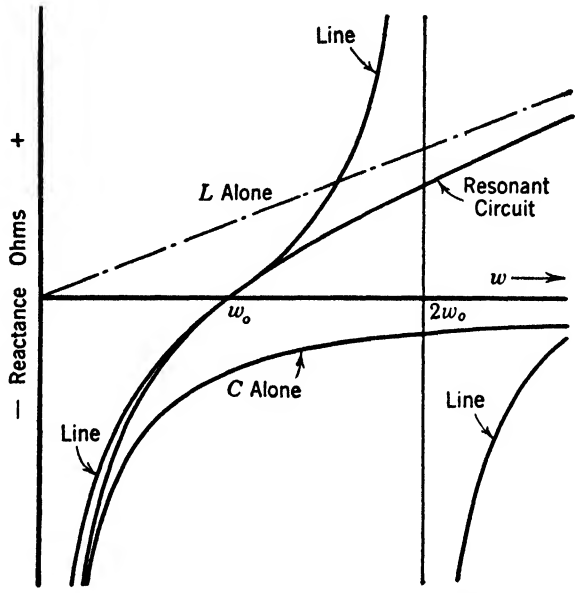


FIG. 9·11 Equivalence of resonant circuit to an open-circuited line.

In a similar way we may evaluate the constants of an anti-resonant circuit which approximates a short-circuited line one-fourth wavelength long. The mathematical development is facilitated here by the use of admittances, because the impedance becomes infinite at the critical frequency.

From equation 9·84 the input reactance of a short-circuited line is given by $X_{in} = Z_0 \tan \beta l$, and the admittance becomes

$$B_{in} = -\frac{1}{X_{in}} = -\frac{1}{Z_0} \cot \beta l = -\frac{1}{Z_0} \cot \frac{\omega l}{v} \qquad [9\cdot99]$$

This expression reduces to zero when $\omega l/v = \pi/2$, defining the resonant angular velocity as before

$$\omega = \omega_0 = \frac{\pi v}{2l} \qquad [9\cdot87]$$

Hence $2\pi f_0 = \pi f_0 \lambda_0 / 2l$, which reduces to $l = \lambda_0/4$. That is, the line is one-fourth of a wavelength long for parallel resonance. Substituting this value of l into equation 9·87, $\omega_0 = 2\pi v/\lambda_0$.

The admittance curve from 9·99 has the slope

$$\frac{dB_{\text{in}}}{d\omega} = \frac{1}{Z_0} \cdot \frac{l}{v} \csc^2 \frac{\omega l}{v} \qquad [9\cdot100]$$

which reduces to

$$\frac{dB_{\text{in}}}{d\omega} = \frac{l}{Z_0 v} \quad \text{for} \quad \omega = \omega_0 \qquad [9\cdot101]$$

The admittance of the parallel L, C circuit is given by

$$B = \left(\omega C - \frac{1}{\omega L}\right) \qquad [9\cdot102]$$

which is zero if

$$\omega = \omega_0 = \frac{1}{\sqrt{LC}} \qquad [9\cdot103]$$

and has the slope

$$\frac{dB}{d\omega} = C + \frac{1}{\omega^2 L} \qquad [9\cdot104]$$

which becomes

$$\frac{dB}{d\omega} = 2C \quad \text{for} \quad \omega = \omega_0 \qquad [9\cdot105]$$

Equating 9·105 and 9·101 we have

$$2C = \frac{l}{Z_0 v} \qquad [9\cdot106]$$

or using 9·87

$$C = \frac{l}{2Z_0 v} = \frac{\pi}{4Z_0 \omega_0} \qquad [9\cdot107]$$

giving with 9·103

$$L = \frac{4Z_0}{\pi\omega_0} = \frac{8lZ_0}{\pi^2 v} \qquad [9\cdot108]$$

Equivalence of short-circuit line and anti-resonant circuit

Equations 9·107 and 9·108 define the values of a coil L and condenser C which, when connected in parallel, are equivalent to a dissipationless line that is short-circuited and one-fourth wavelength long. These equations, together with equations 9·97 and 9·98, are extremely useful

since they permit us to design simple lumped networks which are equivalent, in the region of resonance, to transmission lines.

The equations may be arranged in the reverse order so as to give explicit equations for the line constants Z_0, l, and v in terms of the equivalent lumped elements L and C. In the series resonant case we divide 9·97 by 9·98, giving

$$\left.\begin{aligned} \frac{L}{C} &= \frac{\pi^2 Z_0^2}{16} \\ \text{or} \\ Z_0 &= \frac{4}{\pi}\sqrt{\frac{L}{C}} \\ \text{and from 9·89} \\ \frac{l}{v} &= \frac{\pi}{2}\sqrt{LC} \end{aligned}\right\} \quad \begin{matrix}\text{Equivalence of open-circuit line and series resonant circuit}\end{matrix} \qquad [9\cdot109]$$

In the anti-resonant case we divide 9·108 by 9·107, obtaining

$$\left.\begin{aligned} \frac{L}{C} &= \frac{16 Z_0^2}{\pi^2} \\ \text{or} \\ Z_0 &= \frac{\pi}{4}\sqrt{\frac{L}{C}} \\ \text{and again} \\ \frac{l}{v} &= \frac{\pi}{2}\sqrt{LC} \end{aligned}\right\} \quad \begin{matrix}\text{Equivalence of short-circuit line and anti-resonant circuit}\end{matrix} \qquad [9\cdot110]$$

Equations 9·109 and 9·110 are of great practical importance in that calculations of amplifiers, oscillators, etc., are most conveniently carried out in terms of lumped elements. These equations permit a direct transfer of such designs into systems employing transmission lines.

9·11 Practical Considerations in the Choice of Transmission Lines

If a source of power or voltage is connected to a transmission line at one end, and a load or receiver of power is connected to the other end, there is a flow of energy along the direction of the line. The line serves as a *four-terminal* network and is truly a transmission line in that its function is to convey or transmit power from one point to another.

In an alternative arrangement the receiver is either an open circuit or a short circuit. In either event no power is absorbed by the receiver. Our interest is confined to the relation of the current and the voltage at the sending end of the line. The line now serves as a reactor, or

more exactly as a *two-terminal* network having both resistance and reactance. The term two-terminal network signifies that external connections to the network are made to only two separate terminals.

The form taken by a transmission line depends upon the functions which it must perform. Factors which must be considered in choosing a transmission line include the frequency or frequencies of operation, the amount of power to be conveyed, the current which must be delivered, the highest voltage which may be encountered, the amount of room available, allowable coupling with other circuits, and cost.

Electric power in large amounts is usually transmitted at 60 cycles per second over three wires arranged in a more or less symmetrical fashion. Often the diameter of the conductors is artificially increased by making them hollow, not because of skin effect, but in order to decrease the corona loss caused by the voltage gradient at the surface.

Power at radio frequencies is seldom propagated over wires for great distances. An interesting exception is the so-called coaxial system used for the transmission of speech and television signals over distances of many hundreds of miles. Frequencies as high as 3 megacycles are now in use, and higher frequencies are under consideration. Because of the relatively high attenuation of these signals it is necessary to provide repeaters or amplifiers at intervals of only a few miles.

Parallel wires in air give results comparable to those achieved by coaxial systems. In general a smaller-weight conductor serves to produce a given transmission efficiency in the parallel line. However, the risk of undesired coupling with other systems is much greater than in coaxial lines. Also the total space occupied by the line is greater than for the coaxial conductors.

The following sections discuss some relations which are particularly helpful in the choice of transmission lines for many practical applications. It should be borne in mind that the problem is one of design and that a compromise between several conflicting requirements must usually be made.

9·12 Design of Coaxial Lines for Maximum Dielectric Strength

If the diameter of the outer conductor of a coaxial line is fixed by economic or other considerations there is a certain diameter of inner conductor which gives maximum dielectric strength. That is, one specific design will lead to the highest voltage required for breakdown. If the central conductor is made larger the length of the dielectric path is unduly shortened and breakdown is made easier. If the central conductor is made smaller the electric gradient at its surface becomes excessive owing to the high concentration of electric flux there.

Let us consider a unit length of the very long coaxial structure of Fig. 9·12. If a charge q per unit length is applied to the inner conductor the electric intensity produced between the two conductors is radial and the total flux per unit length is

$$\psi = q \qquad [9\cdot111]$$

The flux density at any distance r from the axis is equal to

$$\frac{\psi}{2\pi r} = \frac{q}{2\pi r} \qquad [9\cdot112]$$

The associated electric intensity is given by the equation

$$E = \frac{q}{2\pi\varepsilon r} \qquad [9\cdot113]$$

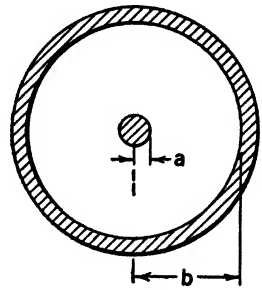


FIG. 9·12 Coaxial conductor.

where $\varepsilon = \kappa_e\varepsilon_0$ and E is the intensity in volts per meter. The potential difference between inner and outer conductors is given by the integral

$$V = \int_a^b \frac{q}{2\pi\varepsilon r}\,dr \qquad [9\cdot114]$$

which, when evaluated, yields

$$V = \frac{q}{2\pi\varepsilon}\log_e\frac{b}{a} \qquad [9\cdot115]$$

The capacitance per unit length is defined as the ratio of charge to voltage; it is given by the expression

$$C = \frac{q}{V} = \frac{2\pi\varepsilon}{\log_e\dfrac{b}{a}} \quad \text{farads per meter} \qquad [9\cdot116]^*$$

The maximum value of electric intensity, E_1, occurs at the surface of the inner conductor because the electric flux density is greatest there. Its value is

$$E_1 = \frac{q}{2\pi\varepsilon a} \qquad [9\cdot117]$$

If V and b are fixed we find that E_1 varies as a varies. Eliminating q

* This relation was developed by a somewhat different method in Chapter 7 as equation 7·174*a*.

between equations 9·115 and 9·117 we have

$$E_1 = \frac{V}{a \log_e \frac{b}{a}} \qquad [9\cdot118]$$

Differentiating E_1 with respect to a so as to obtain a minimum value for this quantity, we have

$$\frac{dE_1}{da} = V \frac{\log_e \frac{b}{a} - \left(\frac{a^2}{b} \cdot \frac{b}{a^2}\right)}{\left(a \log_e \frac{b}{a}\right)^2} \qquad [9\cdot119]$$

For a minimum the derivative must be zero, requiring

$$0 = \log_e \frac{b}{a} - 1$$

or

$$\frac{b}{a} = e = 2.718 \cdots \qquad [9\cdot120]$$

That is, the ratio of radii of outer and inner conductors must be equal to e if the greatest dielectric strength is to be achieved for a given outer radius.

9·13 Design of Coaxial Lines for Minimum Attenuation

The high-frequency resistance of coaxial lines depends not only upon the resistivity and size of the conductors but also upon the frequency and permeability. In the limiting case the current flows only in a thin surface layer, and the resistance depends not only upon the volume but upon the surface area of the conductor. The resistance, as given in equation 7·183 is

$$R = \frac{1}{2\pi} \sqrt{\frac{\mu f \pi}{\sigma}} \left(\frac{1}{a} + \frac{1}{b}\right) \text{ ohms per loop meter} \qquad [9\cdot121]$$

where σ is the conductivity in mhos per meter.
μ is the permeability.
f is the frequency in cycles per second.
a is the outer radius of the inner conductor in meters.
b is the inner radius of the outer conductor in meters.

For pure copper at ordinary temperatures $\sigma = 5.8 \times 10^7$, $\mu = 1.25 \times 10^{-6}$, and the expression reduces to

$$R = 4.16\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right) 10^{-8} \text{ ohms per loop meter} \qquad [9\cdot122]$$

The capacitance of a coaxial line as given in equation 7·174*a* is

$$C = \frac{2\pi\varepsilon}{\log_e \frac{b}{a}} = \frac{\kappa_e}{1.8 \times 10^{10} \log_e \frac{b}{a}} \text{ farads per meter} \qquad [9\cdot123]$$

The inductance as given in equation 7·173 is*

$$L = \frac{\mu}{2\pi} \log_e \frac{b}{a} = 2 \times 10^{-7} \mu_m \log_e \frac{b}{a} \text{ henries per meter} \qquad [9\cdot124]$$

The characteristic impedance thus becomes

$$\sqrt{\frac{L}{C}} = Z_0 = \sqrt{3.6 \times 10^3} \sqrt{\frac{\mu_m}{\kappa_e}} \log_e \frac{b}{a}$$

In all dielectrics $\mu_m = 1$, and in air $\kappa_e = 1$, giving

$$Z_0 = 60 \log_e \frac{b}{a} \text{ ohms} \qquad [9\cdot125]$$

The attenuation, as given in equation 7·188 is

$$\alpha = \frac{3.47\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right) 10^{-10}}{\log_e \frac{b}{a}} \text{ nepers per meter} \qquad [9\cdot126]$$

where the conductors are of copper and the dielectric is air. Again assuming b fixed, and differentiating with respect to a for a minimum attenuation,

$$0 = \frac{d\alpha}{da} = \frac{-\frac{\sqrt{f}}{a^2} \log_e \frac{b}{a} + \sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right)\frac{a}{b}\frac{b}{a^2}}{\left(\log_e \frac{b}{a}\right)^2} 3.47 \times 10^{-10} \qquad [9\cdot127]$$

which requires

$$\log_e \frac{b}{a} = 1 + \frac{a}{b} \qquad [9\cdot128]$$

This equation may be evaluated graphically, yielding $b = 3.6a$.

That is, a coaxial system whose outer conductor is fixed in size has minimum attenuation when the radius of the outer conductor is 3.6 times as large as that of the inner conductor.

* The inductance defined by this equation includes only the flux external to the conductors. It is correct at hyper frequencies but too low at lower frequencies.

9·14 Parallel Wire Lines

Equations similar to those obtained in the preceding section apply to the transmission line composed of two parallel round wires in air. In this case the resistance for copper conductors is given by the formula

$$R = 8.32 \frac{\sqrt{f}}{a} \times 10^{-8} \quad \text{ohms per meter} \qquad [9\cdot129]$$

The capacitance is

$$C = \frac{1}{3.6 \times 10^{10} \log_e \dfrac{b-a}{a}} \quad \text{farads per meter} \qquad [9\cdot130]$$

The inductance is

$$L = 4 \times 10^{-7} \log_e \frac{b-a}{a} \quad \text{henries per meter} \qquad [9\cdot131]^*$$

The characteristic impedance

$$Z_0 = \sqrt{\frac{L}{C}} = 120 \log_e \frac{b-a}{a} \quad \text{ohms} \qquad [9\cdot132]$$

The attenuation, from equation 9·55 when $G = 0$, is

$$\alpha \simeq \frac{R}{2Z_0} \qquad [9\cdot133]$$

Substituting for R and Z_0 from equations 9·129 and 9·132, we obtain

$$\alpha \simeq \frac{3.47\sqrt{f}}{a \log_e \dfrac{b-a}{a}} 10^{-10} \quad \text{nepers per meter} \qquad [9\cdot134]$$

where a is the radius of each conductor and b is the distance between the centers of the conductors, expressed in meters.

Except for the resistance these results are very nearly the same as those that would be obtained if two similar coaxial lines with very thin outer conductors were arranged parallel to and touching each other. Upon

* Equation 9·130 is valid at all frequencies provided that b is relatively large in comparison to a. Equations 9·129 and 9·131 have the additional qualification that the frequency is high so that current flows only in a thin skin upon the conductor. Equation 9·129 follows directly from the development of Chapter 7 on the assumption that current flows uniformly over the conductor surface. Equations 9·130 and 9·131 appear in most books on electrical power transmissions, for example, E. A. Loew, *Electrical Power Transmission*, McGraw-Hill Book Company, New York.

the special assumption of a fixed spacing b, therefore, it is to be expected that maximum dielectric strength will occur for

$$\frac{b}{a} \simeq 2e \simeq 5.4 \quad \text{approximately}$$

or twice that for the single line.

The attenuation is a minimum when

$$\frac{b}{a} \simeq 4.6 \qquad [9\cdot135]$$

9·15 Effective Q of Lines as Resonators

We have already shown that sections of line an integral number of quarter wavelengths long may serve as resonant or anti-resonant circuits. In the development of these basic properties it was convenient to consider both lines and lumped circuits as free from the effects of dissipation. Let us now determine the effective Q of such a system.

Consider an open-circuited line which is approximately a quarter wavelength long. The sending-end impedance is, from equation 9·71,

$$Z_{oc} = Z_0 \coth (\gamma l) = Z_0 \coth (\alpha l + j\beta l) \qquad [9\cdot136]$$

But

$$\coth (A + B) = \frac{1 + \tanh A \tanh B}{\tanh A + \tanh B} \quad \text{and} \quad \tanh jX = j \tan X$$

Accordingly

$$Z_{oc} = Z_0 \frac{1 + j \tanh \alpha l \tan \beta l}{\tanh \alpha l + j \tan \beta l} \qquad [9\cdot137]$$

or, dividing by $\tan \beta l$,

$$Z_{oc} = Z_0 \frac{\cot \beta l + j \tanh \alpha l}{\cot \beta l \tanh \alpha l + j} \qquad [9\cdot138]$$

For the resonant frequency, f_0, $\beta l = \pi/2$, $\cot \beta l = 0$, and the input impedance becomes a relatively low pure resistance

$$Z_{oc} = Z_0 \tanh \alpha l \qquad [9\cdot139]$$

At two frequencies near f_0 the impedance is larger and has a phase angle of 45°. We shall determine one of these frequencies, designating it f_1. We are dealing with a high Q system in which $\alpha \ll \beta$. Moreover, in general $\tanh X < \tan X$. Accordingly the first term in the denominator of 9·137 is negligible in comparison to the second. If the phase angle of the input impedance of equation 9·137 is to be 45° it is necessary that

$$\tanh \alpha l \tan \beta l = \pm 1 \qquad [9\cdot140]$$

Since αl is small we may write $\tanh \alpha l = \alpha l$. In general, $\beta = 2\pi f/c$, where c is the velocity of light in meters per second.* Arbitrarily choosing the positive sign of 9·140 and making the above substitutions of β and $\tanh \alpha l$ in it, we obtain, on dividing through by $\tan \beta l$,

$$\cot \frac{2\pi l}{c} f_1 = \alpha l \qquad [9\cdot141]$$

For angles near $\pi/2$ we may write $\cot X = \pi/2 - X$. Hence

$$\frac{\pi}{2} - \frac{2\pi l}{c} f_1 = \frac{2\pi l}{c}\left(\frac{c}{4l} - f_1\right) = \alpha l$$

From equation 9·87, the resonant frequency f_0 is given by

$$f_0 = \frac{v}{4l} = \frac{c}{4l} \qquad [9\cdot141a]$$

when the dielectric medium is air. Thus

$$\frac{2\pi l}{c}(f_0 - f_1) = \alpha l \qquad [9\cdot142]$$

For the standard series resonant circuit where $Q = \omega L/R$, in the notation of the present development it may be shown that

$$\frac{f_0 - f_1}{f_0} = \frac{1}{2Q} \qquad [9\cdot143]$$

We shall define the Q of a resonant line by this equation. Multiplying both terms of 9·142 by the factor $c/(2\pi l f_0)$, we have

$$\frac{f_0 - f_1}{f_0} = \frac{\alpha c}{2\pi f_0} = \frac{1}{2Q} \qquad [9\cdot144]$$

or

$$Q = \frac{\pi f_0}{\alpha c} \qquad [9\cdot145]$$

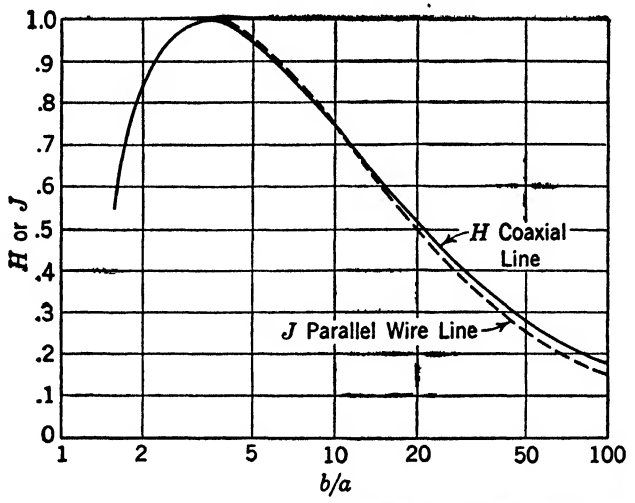
Substituting $\alpha = \dfrac{R}{2Z_0}$ from equation 9·133

$$Q = \frac{2\pi f_0 Z_0}{Rc} \qquad [9\cdot146]$$

where f_0 is the frequency at which the line becomes series resonant, Z_0 is the characteristic impedance, R is the distributed resistance of the line, and c is the velocity of light.

* This substitution requires that the dielectric constant κ_e be equal to 1. This is ordinarily true because few solid dielectrics are known which have low enough losses to make satisfactory resonators.

From equation 9·145 it is clear that the Q is a maximum for the line proportions that give minimum attenuation. That is, maximum Q calls for $b/a = 3.6$ in coaxial structures and 4.0 for parallel wire systems.



(Terman, courtesy of Electrical Engineering)

FIG. 9·13 Selectivity or attenuation of lines.

Figure 9·13 shows the degradation of Q or increase of α that results if other ratios are used.

When reduced to actual line constants the equations become

$$\text{For coaxial lines } Q = 8.39\sqrt{f}\,bH \qquad [9\cdot147]$$

$$\text{For parallel wire lines } Q = 8.87\sqrt{f}\,bJ \qquad [9\cdot148]$$

where b is in meters and H and J are shown in Fig. 9·13.

9·16 Impedance of Resonant Lines

A short-circuited line of any arbitrary length shows an impedance given by equation 9·68.

$$Z_{sc} = Z_0 \tanh \gamma l \qquad [9\cdot149]$$

By a process similar to that used in the development of equation 9·137 we may show that if $\beta l = \pi/2$ corresponding to a quarter wavelength the input impedance at anti-resonance reduces to

$$\frac{Z_0}{\tanh \alpha l}, \text{ a pure resistance} \qquad [9\cdot150]$$

Since α is small we may write $\tanh \alpha l = \alpha l$, so that

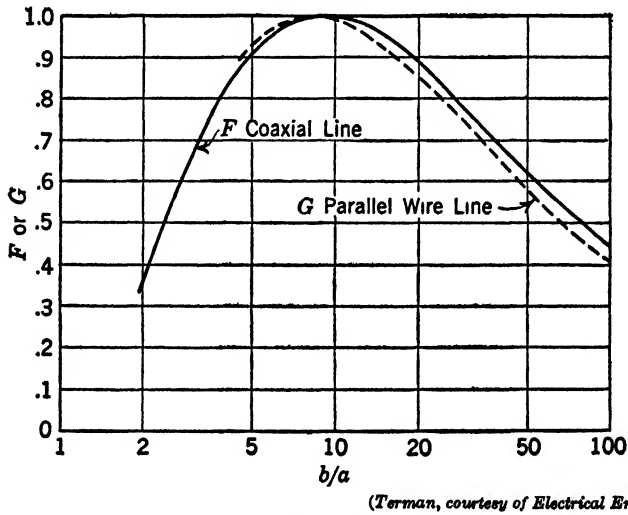
$$Z_{sc} = R_{sc} = \frac{Z_0}{\alpha l} \quad \text{if} \quad \beta l = \frac{\pi}{2} \qquad [9\cdot151]$$

Referring to equation 9·125 and 9·126, we see that the ratio b/a enters into both numerator and denominator of this expression. Terman* shows that the input impedance becomes a maximum for $b/a = 8.0$ in parallel wire lines and 9.2 for coaxial lines. The results for other ratios are shown in Fig. 9·14. For coaxial lines

$$Z_{sc} = \frac{1111\sqrt{f}\, bF}{n} \qquad [9\cdot152]$$

For parallel wire lines

$$Z_{sc} = \frac{2395\sqrt{f}\, bG}{n} \qquad [9\cdot153]$$



(*Terman, courtesy of Electrical Engineering*)

FIG. 9·14 Input impedances of resonant lines.

where b is in meters, n is the number of quarter wavelengths in the line, and F and G depend upon b/a as shown in Fig. 9·14.

9·17 Quarter Wave Line as Impedance Inverter

A very interesting and useful property of lines may be developed from equation 9·83. Substituting $\beta l = \pi/2$, the condition for a quarter wave-

* F. E. Terman, "Resonant Lines in Radio Circuits," *Electrical Engineering*, **53**, 1046, July, 1934.

length, the input impedance of a dissipationless line becomes

$$Z_{in} = \frac{Z_0^2}{Z_r} \qquad [9\cdot154]$$

That is, the input impedance of a section of line varies inversely with the load impedance Z_r. And, as always, the input impedance is equal to Z_0 if the load impedance is equal to Z_0. In dissipationless lines the characteristic impedance is always a pure resistance. In many practical cases Z_r is also a pure resistance and accordingly Z_{in} is a pure resistance as defined by 9·154.

It is an important fact that *any* dissipationless network which produces 90° phase shift shows this property of impedance inversion. Accordingly the choice between a distributed line and a network of lumped reactances will depend upon the conditions of the problem, principally the frequency involved. At high frequencies the line is generally preferable. At low frequencies the reverse is true.

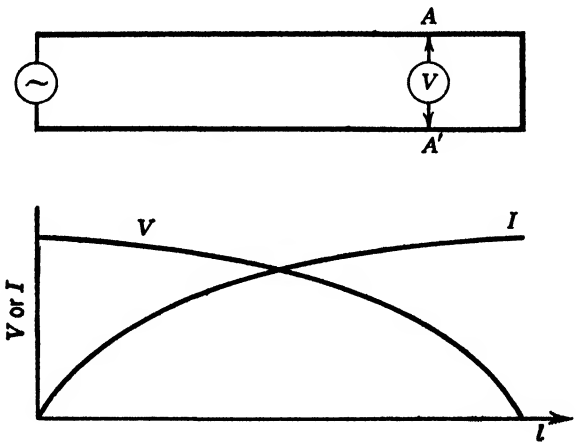


FIG. 9·15 Voltage and current distribution on short circuited line one-fourth wavelength long.

9·18 Use of Transmission Lines as Transformers

Lines which are an odd number of quarter wavelengths long may be used as transformers to increase or decrease voltage or current. The operation is very similar to that of resonant or anti-resonant circuits. Again it is simplest to explain the operation in terms of a line that has no loss, and later to extend the analysis to dissipative lines. Consider the arrangement of Fig. 9·15.

It is seen that the voltage between the conductors varies according to

a cosine relation, being zero at the location of the short circuit. Under the conditions assumed the input or exciting current of the system is zero, corresponding to those in an ideal transformer. If some load impedance Z is connected between the points A, A' it will be subjected to a voltage less than the input voltage depending upon the location of the load tap. Ordinarily the load added is a pure resistance, in which case the input current of the system is nearly in phase with the input voltage unless Z is very low in comparison with the characteristic impedance of the line. That is, the dissipationless line acts almost as an ideal transformer, delivering a load current magnified by the same ratio that the load voltage is decreased.

The operation is closely analogous to that of the lumped circuit shown in Fig. 9·16. A condenser and coil of finite values are anti-resonant at

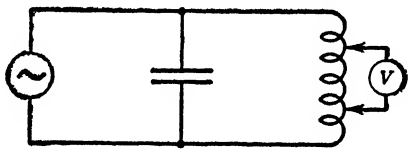


FIG. 9·16 Lumped circuit transformer equivalent to Fig. 9·15.

the operating frequency. The load is then tapped across some part of the coil. If the turns of the coil are closely coupled to each other the leakage flux is small and the system closely approximates an ideal transformer.

If dissipation is present in either the line system or the lumped circuit just described its effect can be accounted for to a good approximation by assuming a resistor shunted across the high-impedance terminals. For the lumped-circuit case this resistor must be equal to QX, where X is the reactance of either element and Q is the selectivity factor of the entire system.* For the line this value has already been shown, from equation 9·151, to be

$$R_{sc} = \frac{Z_0}{\alpha l}$$

Evidently lines may be used as step-up transformers by interchanging the position of generator and load. A limiting case of some interest arises when a generator of low voltage and low internal impedance is connected to one end of a quarter wave line and a high impedance, such as the input of a vacuum tube, is connected to the other end.

* It should be recalled that $Q = R/\omega L$ for the selectivity of a loss-free coil shunted by a resistor, whereas $Q = \omega L/R$ for the familiar case of a loss-free coil in series with a resistor.

Recalling that at resonance $\beta l = \pi/2$ and $\cot \pi/2 = 0$ it is seen from equation 9·85 that the input impedance of a quarter wave line is zero if the far end is open-circuited and dissipation is negligible. Theoretically, therefore, an infinite voltage step-up is possible, exactly as with a loss-free series-resonant circuit.

When dissipation is present in the line its effect is best accounted for by assuming a lumped resistor equal to

$$R_0 = \frac{Z_0}{\alpha l} \qquad [9\cdot155]$$

at the open-circuit end as in Fig. 9·17. Since the dissipation of the line has been accounted for we may now calculate the behavior as if the lines

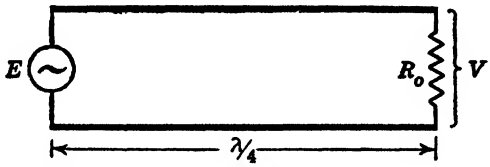


FIG. 9·17 Quarter wave line with dissipation, as step-up transformer.

were free from loss. Using equation 9·154, the impedance seen by the generator is

$$Z = \frac{Z_0^2}{R_0} \qquad [9\cdot156]$$

Combining these two equations the input impedance is

$$Z = Z_0 \alpha l, \quad \text{a pure resistance} \qquad [9\cdot157]$$

Since all the energy drawn from the generator is expended in the dissipation of the line and since the effect of this dissipation is equivalent to that of the lumped resistor R_0 we may apply the principle of conservation of energy to Fig. 9·17 yielding

$$\frac{E^2}{Z} = \frac{V^2}{R_0} \quad \text{or} \quad \frac{V^2}{E^2} = \frac{R_0}{Z} \qquad [9\cdot158]$$

But $R_0 = Z_0/\alpha l$ and $Z = Z_0 \alpha l$. Accordingly

$$\frac{V^2}{E^2} = \frac{Z_0}{\alpha l}\frac{1}{Z_0 \alpha l} = \frac{1}{(\alpha l)^2} \qquad [9\cdot159]$$

and the voltage step-up is

$$\frac{V}{E} = \frac{1}{\alpha l} \qquad [9\cdot160]$$

But $\alpha \simeq R/2Z_0$ from equation 9·133, and the resonant frequency $f_0 = c/4l$, from equation 9·141*a*.

The voltage step-up becomes

$$\frac{V}{E} = \frac{8Z_0 f_0}{cR} \qquad [9\cdot161]$$

where c is the velocity of light.

The form of 9·161 is very interesting since it shows that the step-up ratio is proportional to the frequency and to the characteristic impedance and is inversely proportional to the line resistance.

It has been shown that the conditions for maximum Q do not permit flexible adjustment of the impedance level of lines as anti-resonant circuits.* By means of the device illustrated in Fig. 9·15 it is possible to design the line for maximum selectivity and then to adjust the effective impedance by tapping on to the line. This is unavailing, of course, if the required impedance is higher than that offered by a quarter wave line of maximum Q. Fortunately such is seldom the case. Figure 9·18 shows resonant lines operating in an oscillator on this basis.

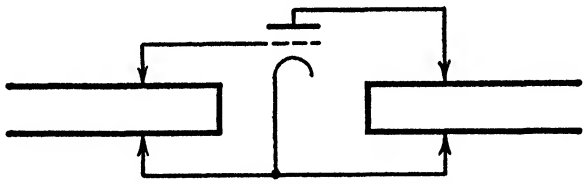


FIG. 9 18 Resonant line oscillator.

9·19 Impedance Transformation by Means of Stubs

By the use of a separate section of transmission line, referred to as a stub line or stub, it is possible to secure the most general form of impedance transformation. In practice the method is most frequently applied to antennas. By a proper choice of the line lengths involved it is possible to secure a true impedance match between the feeding line and an antenna of arbitrary impedance. The arrangement is shown in Fig. 9·18*a*.

The antenna or other load is represented by the general impedance Z_r. It may have any phase angle and any impedance whatever. Line 3 is usually relatively long and must have a low attenuation constant in order to deliver power efficiently from the generator to the antenna. Lines of low attenuation are characterized by small values of series resistance and shunt conductance, and therefore have a characteristic

* Section 9·15 equations 9·147 and 9·148.

impedance, Z_0, which is nearly a pure resistance. For the sake of simplicity it will be assumed that all three sections of line have the same characteristic impedance and that it is a pure resistance. An analysis of the more general case is not extremely difficult but does not yield results which are much more useful than these.

The length L_1 is first chosen so that the impedance presented to the junction by line 1 is equivalent to a pure resistance equal to Z_0 shunted by some value of reactance. The value of L_1 depends upon the magnitude and phase angle of Z_r. The reactance presented to the junction depends, in turn, upon the value of L_1. The important fact, however, is

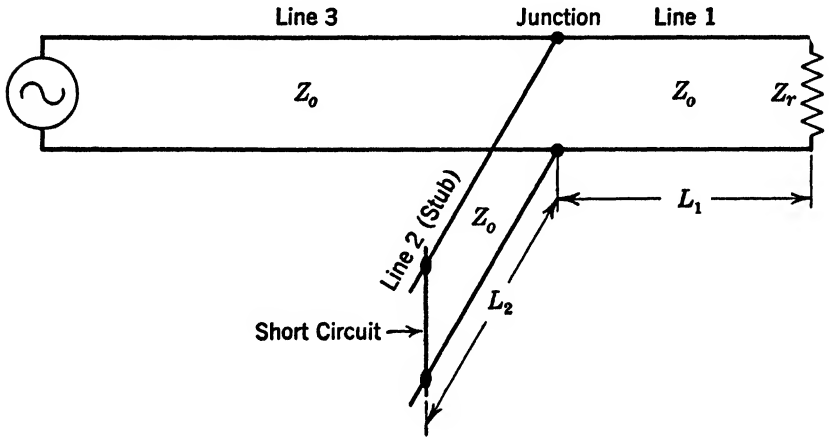


FIG. 9·18a Impedance matching by means of a stub line.

that there is always a value of L_1 for which the resistive component of the impedance does match Z_0. The length L_2 is then adjusted so that the impedance presented to the junction by line 2 is a pure reactance equal and opposite to the reactive component of the parallel impedance of line 1.

Under these conditions line 3 is perfectly matched by the combined actions of lines 1 and 2. Accordingly no standing waves exist in line 3, and energy is transmitted with maximum efficiency from the generator to the load. Line 1 produces all the resistance transformation, and line 2 produces the reactance which is necessary to cancel the undesired reactance which unavoidably accompanies the action of line 1. We may think of the operation as a sort of complex parallel resonance in which the stub line cancels the reactive charging current of the load which comprises Z_r and line 1.

In practice lines 1 and 3 are usually continuous, and line 2, the stub,

is at right angles to the main line. It is sometimes more convenient, and it is equally correct, to make lines 2 and 3 continuous and to apply the load to the right-angle branch. The length L_1 is usually approximately a quarter of a wavelength. The stub line is usually short-circuited for greatest mechanical convenience and is then approximately a quarter wavelength long. If for some reason the stub line must be open-circuited it should be approximately a half wavelength long.

If the impedance of Z_r is not in the same order of magnitude as Z_0 it may be found that appreciable losses occur owing to excessive currents in the matching lines. Under these conditions it may prove desirable to adjust the characteristic impedance of line 1 or line 2 or both so as to minimize this loss.

9·20 Use of Transmission Lines in Filters

Low-pass filters are structures which transmit freely all frequencies from direct current to some limiting frequency called the cut-off frequency. All frequencies above the cut-off value are attenuated. Such a filter necessarily has coils for the series elements and condensers for the shunt elements.

Band-pass filters, perhaps the most important of all in communication systems, attenuate both high and low frequencies and pass freely only frequencies lying between two definite values. Such a filter uses a series-resonant circuit wherever a simple coil appears in the low-pass filter and an anti-resonant circuit wherever a condenser appears in the low-pass design.*

We have already shown that transmission lines of suitable length may serve to simulate resonant or anti-resonant combinations of high Q.

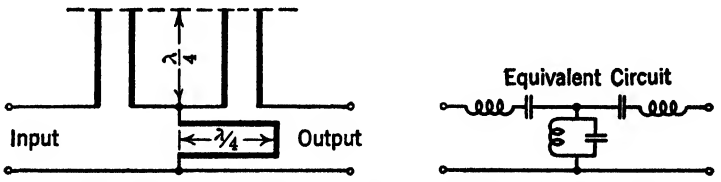


FIG. 9·19 Prototype band-pass filter using resonant lines.

This statement applies with equal force over a wide range of frequencies At the lower frequencies, transmission lines are relatively bulky, and lumped coils and condensers are used. At frequencies above a few megacycles, however, the situation is reversed and transmission lines

* These statements apply rigorously only to the so-called prototype filter. However, they are indicative of the general mode of operation of all filters.

offer the most satisfactory filter networks now known. In the hyper-frequency region, hollow tubes may be employed to advantage.

Although parallel wire lines may be used, coaxial structures are

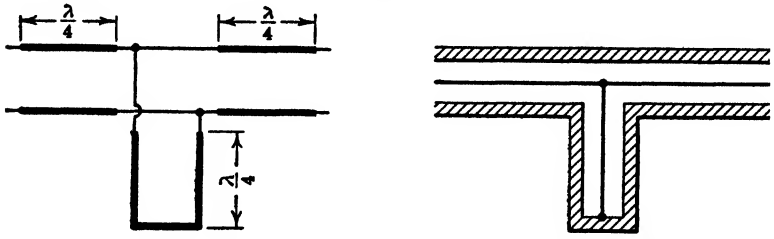
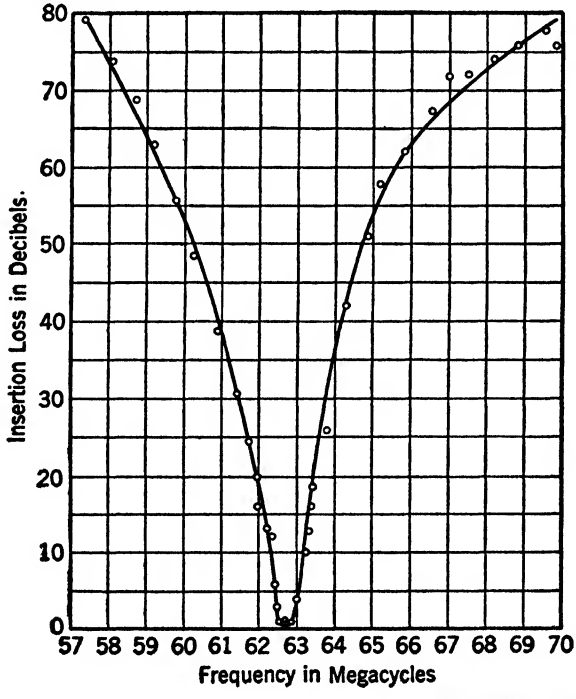


FIG. 9·20 Band-pass filter using coaxial line structure.

generally more convenient. The mechanical stability is usually higher, and the risk of undesired coupling between the several lines and neighboring circuits is reduced.



(Mason and Sykes, courtesy of BSTJ)

FIG. 9·21 Measured insertion loss of filter used on Green Harbour-Provincetown radio link.

The arrangement of Fig. 9·19 has been successful as a band-pass filter. The two open-circuited lines are identical and serve as series resonant

circuits. The short-circuited line, which serves as the shunt anti-resonant combination, may need to have a different characteristic impedance in order to satisfy the filter design equations.

A rather different arrangement is shown in Fig. 9·20. Here the transmission lines serve as four-terminal rather than as two-terminal networks. The physical arrangement, however, is simpler to construct than that of Fig. 9 19. The electrical performances are comparable.

Although suitable high-Q coils are not available at these frequencies it is possible to construct excellent condensers of appropriate capacitance. Accordingly an important class of filters exists in which resonant lines and condensers serve as the elements. Mason and Sykes have published a very complete account of such filters.* The necessary theoretical background and design formulas are included. Figure 9·21 is a measured characteristic reproduced from their paper.

9·21 Radiation from Transmission Lines

Let us first consider a coaxial line at relatively high frequencies. We have shown that with perfect conductivity both electric and magnetic fields are normal to the direction of propagation. In other words, a true transverse electromagnetic wave system exists. Energy evidently is not dispersed laterally from such a system. It is clear that even with finite conductivity of the guiding conductors no appreciable amount of power can be transmitted through the outer conductor to be radiated from it. Therefore we may conclude that no significant amount of power is radiated from a coaxial line at high frequency except from the open ends. This conclusion is supported by the fact that experimental data accurately checks the attenuation as calculated from equation 9·126.

With the parallel wire line the situation is more involved. Here the electric and magnetic fields have no finite boundary and accordingly the wave system extends to infinity. Again, however, we may simplify the problem by assuming the wires to have infinite conductivity. The electric and magnetic vectors are everywhere perpendicular to each other and to the direction of propagation. The resulting system is very similar to the simple plane-polarized wave dealt with in earlier chapters. The only significant difference is that the lines of electric and magnetic intensity are closed curves rather than infinite straight lines. Both wave systems share the important property that they are propagated with the speed of light and without attenuation. In particular the wave system is guided along the conductors with a wave front that is

* W. P. Mason and R. A. Sykes, "The Use of Coaxial and Balanced Transmission Lines in Filters and Wide Band Transformers for High Radio Frequencies," *Bell System Tech. J.*, 16, 275, June, 1937.

perpendicular to the wires. Thus the Poynting vector shows no component except in the direction of the guides, and accordingly there is no radiation loss from a very long parallel wire line.

The conditions which exist at the ends of a finite line require separate investigation. From our previous study of wave guides we may conclude that no energy can be radiated from the end of a coaxial line if the end is closed with a solid conducting plug as shown in Fig. 9·22.

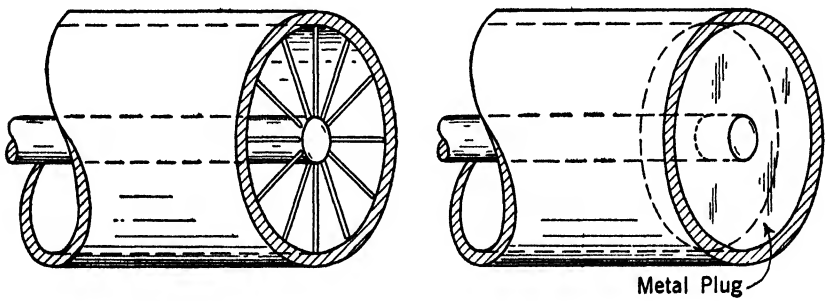


FIG. 9·22 Coaxial line short-circuited by bars and a disk.

Accordingly such a plug is a perfect short circuit and reflects all the incident energy back into the line. In the principal mode of coaxial transmission the electric field is radial, and the wire grid structure of Fig. 9·22 forms an entirely satisfactory short circuit.

The parallel wire line is commonly regarded as short-circuited if the two conductors are joined by a portion of a similar conductor. From the previous paragraphs it is clear that such an arrangement can, at best, reflect only a portion of the unbounded wave propagated down the line. For perfect reflection of the wave system an infinite conducting sheet is necessary, but very satisfactory results are obtained if a circular disk a few times as large in diameter as the line spacing is used.

A system comprising a large number of wires designed to coincide with the electric field configuration serves as an excellent reflector or short-circuit and may require less material and weight than the solid sheet.

If only the single wire is used as a short circuit, the electromagnetic field is distorted at the end of the line. A considerable portion of the power continues in the general direction of the line as radiation in free space and is lost. Thus the effective Q of the line as a resonator is diminished.

In making measurement on parallel lines, it is usually convenient to provide an adjustable shorting device which may be placed in any desired position along the line. A conducting disk illustrated in Fig.

9·23*b* may be used. To aid in making good electrical connection the edges of the disk in contact with the wires may be filed to a sharp edge. Often the section of the line beyond the short circuit may be excited by

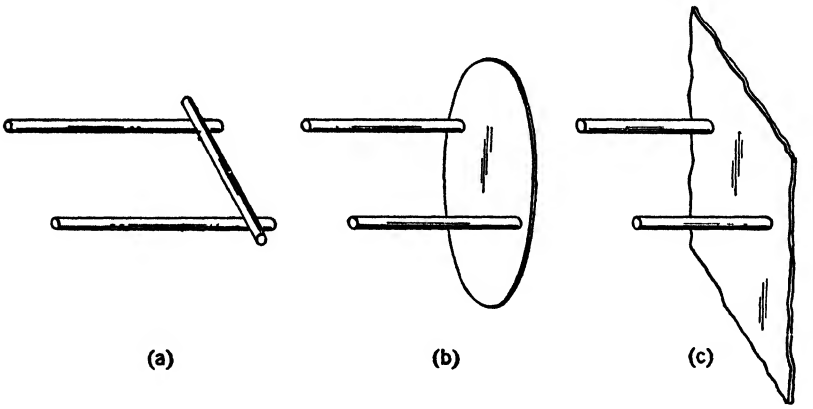


FIG. 9·23 Parallel wire line short-circuited by wire, by a circular disk and by an infinite conducting sheet.

induction from the section in use. This auxiliary or secondary circuit may react upon the primary section or circuit, providing a coupled circuit phenomenon. The amplitude characteristics of such coupled circuits have been studied by King.* In order to avoid this reaction, he

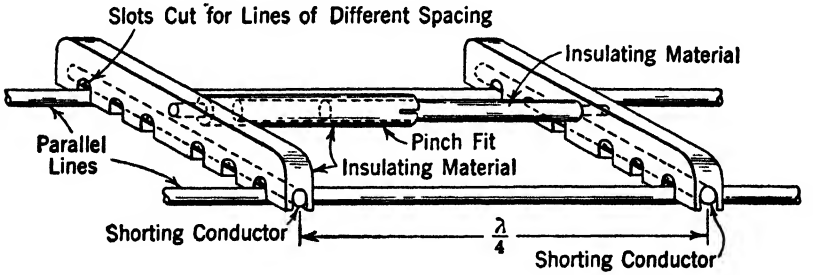


FIG. 9·24 Shorting bars in tandem arrangement devised by King. This device provides decoupling between used and unused sections of line.

has designed a tandem shorting device shown in Fig. 9·24. Two wires are mounted in insulating supports which are separated from each other by a dielectric rod of adjustable length. In operation, the length of this rod is made equal to approximately one-quarter the wavelength of the voltage being used in the primary circuit. In this way the section of

* Ronald King, "Amplitude Characteristics of Coupled Circuits Having Distributed Constants," *Proc. IRE*, Vol. 21, August, 1933.

line directly behind the short circuit is detuned, and thus any reaction on the primary circuit is reduced.

It is often desirable to terminate a finite line section in a lumped resistance equal to the characteristic impedance. When the line spacing is very small compared to the wavelength, that is, at relatively low frequencies, a normal two-terminal resistor serves admirably. At higher frequencies, however, the concept of a guided wave becomes necessary. As in the short-circuited line, satisfactory results are attained only if the termination is distributed. In the coaxial line a small disk of material of suitable resistivity serves as a perfect termination to very high frequencies. In the open-wire line it is again necessary to employ a relatively large sheet if the radiated energy is to be held to a very low value. In practice these distributed terminations are usually made by spraying some form of carbon on a dielectric support. Under certain conditions a piece of carbon-coated paper serves as a perfect resistive termination.

The open-circuited line is difficult to achieve in practice. The reason for this is made clear if we examine a coaxial structure in which the outer conductor is large enough to serve as a circular wave guide for the frequency being transmitted. Now if we cut off the central conductor in the effort to achieve an open-circuited line, we merely obtain a partial reflection, and an E_0 wave is propagated along the guide. If the outer conductor is cut off, leaving the inner conductor extending, we meet a similar difficulty. Part of the energy is reflected, but again much of it is radiated and lost. If both are cut at the same point we are, perhaps, better off, but certainly significant radiation losses exist. For frequencies below the cut-off of the outer conductor it is possible to achieve nearly perfect reflection if the outer conductor extends somewhat beyond the inner one.

In the parallel wire line it is again difficult to achieve anything closely approximating open circuit unless the wavelength is great compared to the line spacing. For these reasons it is usually desirable to use short-circuited lines for resonators whenever possible.

PROBLEMS

9·1 A common form of telephone circuit consists of two hard-drawn copper wires, 12 gage, spaced 8 in. on center. Calculate the line constants R, L, and C per loop meter. Ignore the effect of insulators and the conductance of the air.

9·2 A particular cable composed of 22-gage wires has the line parameters per loop mile $R = 171$ ohms, $L = 0.001$ henry, $C = 0.073$ microfarad, $G = 1.75$ micromhos. Calculate the characteristic impedance of this line at frequencies of 30 and 300 cycles per second.

9·3 Calculate the characteristic impedance of the line of problem 9·2 at frequencies of 3000 and 30,000 cycles per second.

9·4 Calculate the propagation constant of the line of problem 9·2 at frequencies of 30 and 300 cycles per second.

9·5 Calculate the propagation constant of the line of problem 9·2 at frequencies of 3000 and 30,000 cycles per second.

9·6 A particular cable has a characteristic impedance of $100 + j30$ ohms and a propagation constant of $0.0001 + j0.0025$ (nepers and radians per meter) at a frequency of 1 million cycles per second. A generator having a terminal potential of 10 volts at this frequency is connected to a very long line of this sort. Calculate the input current and the current at distances of 1000 and 10,000 meters from the generator. Make a vector diagram of the currents and voltages involved.

9·7 A cable 1000 meters long having the constants given in problem 9·6 is terminated in a lumped impedance of $20 - j30$ ohms at 1 million cycles per second. What is the input impedance of the cable?

9·8 The center-to-center spacing of a parallel wire transmission line is fixed at the value D. Determine the conductor diameter d for which the breakdown voltage will be greatest.

9·9 A particular line has the constants $R = 0.009$ ohm, $L = 2$ microhenries, $C = 5\mu\mu f$, $G = 0$ per meter. Calculate the characteristic impedance at a frequency of 2 megacycles per second.

9·10 Calculate the effective Q of a quarter-wave resonator for 2 megacycles per second formed from the line of problem 9·9.

9·11 A coaxial line of copper 3 meters long is short-circuited and serves as a quarter-wave resonator equivalent to a parallel resonant circuit. If the inner radius of the outer conductor is 3 cm and the radius of the inner conductor is 0.8 cm, calculate the resonant frequency and the impedance at the resonant frequency.

9·12 Calculate the Q of the above resonator.

9·13 Design a parallel wire open-circuit resonator to operate at 12 megacycles. The material is copper, and the line must not be unnecessarily long. Choose constants for which the input impedance is approximately 20,000 ohms.

9·14 A resonant line has a characteristic impedance of 75.0 ohms, a velocity of propagation of 3×10^8 meters per second, and is 4 meters long. It is short-circuited. Calculate the constants of the parallel resonant circuit equivalent to this resonator.

9·15 The line of problem 9·14 is open-circuited. Determine the constants of the equivalent series resonant circuit.

9·16 At a point near a coaxial conductor the magnetic field may be thought of as a superposition of two fields, one due to the inner conductor, one due to the outer conductor. The currents in inner and outer conductors are in phase at any particular point along the line. Because of the unequal distances of the point from the various current elements, the fields due to these elements at the point are not in time phase. It thus appears that a net field at the point results and that energy is radiated from the line. Discuss this concept.

9·17 A portion of a transmission line is $\frac{1}{8}$ wavelength long and has negligible attenuation. That is, $\alpha = 0$, $\beta l = 45°$. If such a line is terminated in a pure resistance R equal to the characteristic impedance Z_0 the input impedance is also equal

to R. Demonstrate that the input impedance is modified by an increment of reactance numerically equal to ΔR if the terminal resistor R is increased by a small increment of resistance ΔR.

9·18 Determine the effect upon the input impedance of the line of problem 9·17 if the terminating resistance is modified by the addition of a small reactance ΔX.

9·19 A section 1 meter long of the line of problem 9·9 is used as a quarter-wave line to couple a low-impedance generator to a very high-impedance load (vacuum-tube grid). Assuming the velocity of propagation to be that of light, calculate the voltage step-up.

9·20 When resonant parallel wire lines (lecher lines) are used for laboratory measurement it is common practice to short-circuit the line with a narrow bar or wire. Why is it desirable, as is usually done, to short-circuit the line with a second conductor located approximately $\frac{1}{4}$ wavelength beyond the other? Consider in terms of pure circuit theory and in terms of a propagated electromagnetic wave.

9·21 A condenser is formed from two strips of aluminum foil separated by a sheet of paper. The electrical connections are made at one end of the system. The aluminum sheets are 10 cm wide and 100 meters long, and each has a resistance of 10 ohms per meter of length. The paper, which has become wet, has a conductivity of 10^{-8} mho per meter and is 0.01 cm thick. Calculate the net resistance of this system.

9·22 Two round copper wires 0.1 cm in diameter are spaced a distance of 1 meter apart and are immersed in sea water for a length of 1000 meters. The conductivity of the copper is 5.8×10^7 mhos per meter. The conductivity of the water is 5 mhos per meter. Calculate the resistance observed between the two wires, neglecting end effects.

CHAPTER 10

CAVITY RESONATORS

10·1 Introduction

This chapter is concerned with resonant electric circuit elements of the cavity or enclosed chamber type in which a dielectric medium is enclosed by a conducting material. Physicists have often postulated enclosures of this type and have given them the name " hohlraums."*

The principle of cavity resonators has been used for some time in communication circuits in the form of the familiar concentric line resonator. More recently other hollow-cavity types have been proposed. These resonators are approximately equivalent to a conventional resonant circuit consisting of inductance and capacity. In general the mathematical problem involved in obtaining a complete solution of these elements is extremely complex, and only the simplest structures have been solved completely. In the following sections a few of the more elementary structures will be considered.

It has been known for some time that solutions satisfying Maxwell's equations may be obtained which establish the possibility of electromagnetic waves in the dielectric medium of the cavity. These solutions have a definite frequency and vanishing tangential component of electric intensity at the conducting surface. The requirement that the tangential electric intensity be zero at the boundary between the dielectric and the metal restricts the allowable frequencies. Thus only certain definite frequencies have associated with them electromagnetic fields which meet the required boundary conditions. The value of these definite or resonant frequencies is determined by the size and shape of the conducting surface. Since the hohlraum can contain an electromagnetic field, which varies periodically when and only when the frequency of the field has certain definite values, it is appropriately called a resonator. Because of the oscillations existing inside the enclosure, the name rhumbic enclosure or Rhumbatron† has been used to identify these electrical structures.

* It is assumed that these hohlraums are surrounded by a conducting medium of infinite extent. At the frequencies with which we are concerned here, the penetration of the waves into the metal conducting surface is extremely small. Hence for all practical purposes the thickness of the metal enclosure may be physically thin, yet electrically it may be infinitely thick.

† The word " Rhumbatron " is a trademark owned by the Sperry Gyroscope Company, Inc.

10·2 Resonance in a Rectangular Cavity (Perfect Conductivity)

In the discussion of the rectangular wave guide it was shown that there is no propagation in the X direction through the guide at the critical frequency. At this frequency, the angle θ defining the direction of the elementary waves becomes 90°, and the electromagnetic wave may be thought of as traveling back and forth between the conducting surfaces, in a direction normal to the axis of the tube. This is one form of the cavity-resonance principle.

The phenomenon of resonance is associated with the production of standing waves. If we insert a conducting boundary into the rectangular pipe of Fig. 6·1, the waves striking this boundary will be reflected. If the reflected magnetic and electric components are in proper phase to reinforce those of the direct wave, standing waves will be set up. If some means is provided for exciting waves in a pipe closed at both ends by conducting surfaces which are adjusted for reflection in proper

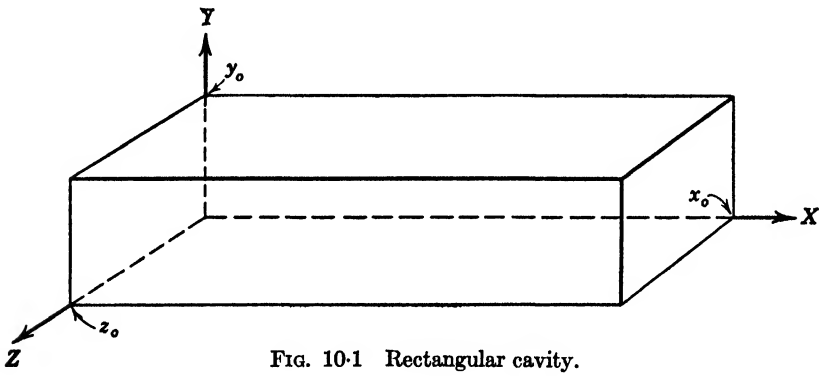


FIG. 10·1 Rectangular cavity.

phase, the waves will travel back and forth within the metal enclosure between the reflecting surfaces. Adjustment of the reflecting surfaces to give the correct phase relations is called tuning in the same sense that this word is used with reference to series or parallel resonant L-C circuits.

Let us consider the conditions for resonance in the rectangular cavity shown in Fig. 10·1, having the dimensions x_0, y_0, and z_0 as indicated in the figure. This cavity may be considered as a section of the rectangular pipe of Fig. 6·1, x_0 units in length. We have already developed, in Chapter 6, expressions for traveling waves in a pipe of infinite length. Let us determine under what conditions these waves will be reflected from a conducting boundary surface inserted in the pipe.

Let us consider, in particular, waves of the H_{0m}, or transverse electric type. The equations for these waves may be obtained from the general

equations 6·47. Thus, when $n = 0$, they may be written

$$\left.\begin{aligned} H_{x+} &= A \cos\left(\frac{m\pi}{z_0} z\right) e^{j(\omega t - \beta_{0m} x)} \\ H_{z+} &= jA\left(\frac{z_0}{m\pi}\right)\beta_{0m} \sin\left(\frac{m\pi}{z_0} z\right) e^{j(\omega t - \beta_{0m} x)} \\ E_{y+} &= jA\left(\frac{z_0}{m\pi}\right)\omega\mu_0 \sin\left(\frac{m\pi}{z_0} z\right) e^{j(\omega t - \beta_{0m} x)} \end{aligned}\right\} \quad H_{0m} \text{ waves traveling in the plus } X \text{ direction } [10\cdot1]^*$$

The fact that these waves are traveling in the plus X direction is indicated by the subscript $+$. Components striking the conducting boundary will undergo certain changes in phase and direction as indicated in Fig. 10·2. The subscript $-$ indicates the reflected component traveling in the $-X$ direction.

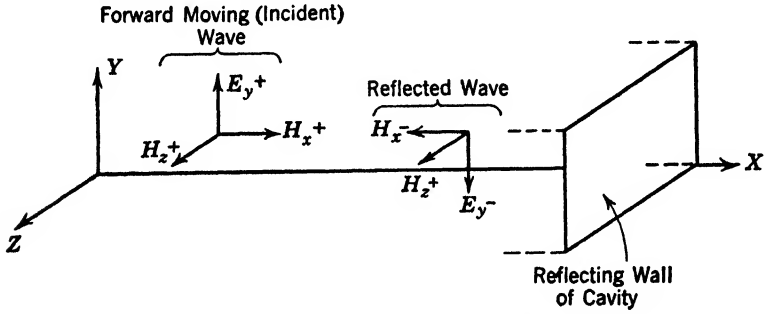


FIG. 10·2 Orientation of the sectors in an H_{0m} wave before and after striking the reflecting wall inserted in the rectangular pipe.

It is seen that E_y and H_x are changed in phase 180° after striking the conducting surface, whereas H_z maintains its same relative phase position. The relationships are

$$\begin{aligned} H_{x-} &= -H_{x+} \\ H_{z-} &= +H_{z+} \\ E_{y-} &= -E_{y+} \end{aligned} \qquad [10\cdot2]$$

The standing wave developed is the sum of the direct and reflected waves. Using the subscript s to refer to the standing wave produced,

$$\begin{aligned} H_{xs} &= H_{x+} + H_{x-} \\ H_{zs} &= H_{z+} + H_{z-} \\ E_{ys} &= E_{y+} + E_{y-} \end{aligned} \qquad [10\cdot3]$$

where addition is in a vector sense, since distance enters in the general

* In writing these equations, when the dielectric is air, $\mu_1 = \mu_0$, $\epsilon_1 = \epsilon_0$. Also γ_{0m} has been replaced by $j\beta_{0m}$ and $\sqrt{\gamma_{0m}^2 + \omega^2\epsilon\mu}$ by $m\pi/z_0$ from equation 6·45, when $n = 0$.

wave equations as an exponential function. Substituting 10·1 in 10·3 and using 10·2 we have

$$\left.\begin{aligned} H_{xs} &= A \cos\left(\frac{m\pi}{z_0} z\right) \left\{e^{-j\beta_{0m}x} - e^{j\beta_{0m}x}\right\} e^{j\omega t} \\ H_{zs} &= jA \left(\frac{z_0}{m\pi}\right) \beta_{0m} \sin\left(\frac{m\pi}{z_0} z\right) \left\{e^{-j\beta_{0m}x} + e^{j\beta_{0m}x}\right\} e^{j\omega t} \\ E_{ys} &= jA \left(\frac{z_0}{m\pi}\right) \omega\mu_0 \sin\left(\frac{m\pi}{z_0} z\right) \left\{e^{-j\beta_{0m}x} - e^{j\beta_{0m}x}\right\} e^{j\omega t} \end{aligned}\right\} \quad [10\cdot4]^*$$

or

$$\left.\begin{aligned} H_{xs} &= -j2A \cos\left(\frac{m\pi}{z_0} z\right) \sin(\beta_{0m}x) e^{j\omega t} \\ H_{zs} &= j2A \left(\frac{z_0}{m\pi}\right) \beta_{0m} \sin\left(\frac{m\pi}{z_0} z\right) \cos(\beta_{0m}x) e^{j\omega t} \\ E_{ys} &= 2A \left(\frac{z_0}{m\pi}\right) \omega\mu_0 \sin\left(\frac{m\pi}{z_0} z\right) \sin(\beta_{0m}x) e^{j\omega t} \end{aligned}\right\} \quad \begin{matrix} H_{0m} \text{ waves} \\ \text{standing in} \\ \text{pipe} \\ [10\cdot5]\dagger \end{matrix}$$

It may be noted that, in the case of traveling waves, H_z and E_y are in phase in both time and space, whereas for standing waves H_{zs} and E_{ys} are 90° out of phase in both time and space. This is indicated in Fig. 10·3. Equations 10·5 are general forms for the standing-wave phenomena in a tube and are subject to the boundary conditions of the resonator which will now be imposed.

At the boundary planes, $x = 0$ and $x = x_0$, E_{ys} must be zero. Let us set

$$\beta_{0m} = \frac{l\pi}{x_0} \qquad [10\cdot6]$$

where l is an integer; $l = 1, 2, 3$, etc. Then equations 10·5 become

$$\left.\begin{aligned} H_{xr} &= -j2A \cos\left(\frac{m\pi}{z_0} z\right) \sin\left(\frac{l\pi}{x_0} x\right) e^{j\omega t} \\ H_{zr} &= j2A \left(\frac{z_0}{m\pi}\right) \beta_{0m} \sin\left(\frac{m\pi}{z_0} z\right) \cos\left(\frac{l\pi}{x_0} x\right) e^{j\omega t} \\ E_{yr} &= 2A \left(\frac{z_0}{m\pi}\right) \omega\mu_0 \sin\left(\frac{m\pi}{z_0} z\right) \sin\left(\frac{l\pi}{x_0} x\right) e^{j\omega t} \end{aligned}\right\} \quad \begin{matrix} H_{0ml} \\ \text{resonant waves} \\ \text{in pipe} \\ [10\cdot7] \end{matrix}$$

where the subscript r indicates the resonant wave.

* It will be recalled, in writing these equations, that a wave moving in the $+X$ direction is represented by the function $C(\omega t - \beta x)$ whereas one moving in the $-X$ direction is represented by the function $C(\omega t + \beta x)$.

† Equations 10·5 are deduced immediately with the well-known equalities

$$\frac{e^{-j\beta x} - e^{j\beta x}}{-2} = -j \sin \beta x \qquad \text{and} \qquad \frac{e^{j\beta x} + e^{-j\beta x}}{2} = \cos \beta x$$

Hence, when $x = 0$ or x_0, the boundary conditions are satisfied, no matter what integral value l may have. This particular wave may be designated H_{0ml}; it represents a special case of the general H_{nml} mode. These equations completely define the standing waves of the rectangular cavity and as such establish mathematical conditions of resonance for

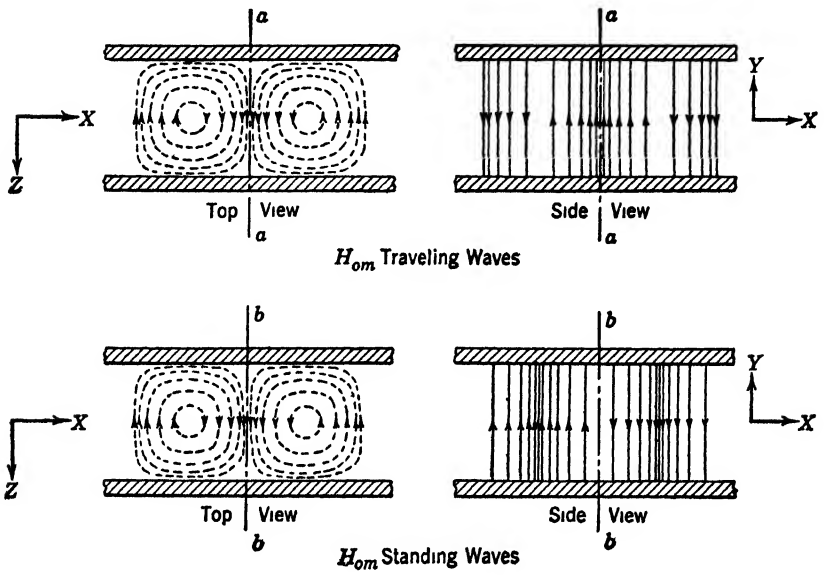


FIG. 10·3 Comparison of time and space phase relations between **E** and **H** vectors of the H_{0m} wave in a rectangular wave guide and in a rectangular cavity resonator.

the three components of the H_{0ml} mode. When the frequency and dimensions of the cavity satisfy these conditions, equations 10·7 apply, and the structure is said to be in resonance. The presence of such resonant oscillations may be indicated by methods described in Chapter 8.

The phase constant in the general case, β_{nm}, is given in equations 6·71, and when the dielectric is air, we may replace v_1 with c, giving

$$\beta_{nm} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{n\pi}{y_0}\right)^2 - \left(\frac{m\pi}{z_0}\right)^2} \qquad [10\cdot8]$$

For the H_{01} wave, $n = 0$, $m = 1$, giving

$$\beta_{01} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{\pi}{z_0}\right)^2} \qquad [10\cdot9]$$

The critical frequency, f_0, for the propagation of the H_{01} wave in a

rectangular guide is obtained by setting $\beta_{01} = 0$. Thus

$$(f_0)_{01} = \frac{c}{2\pi}\left(\frac{\pi}{z_0}\right) = \frac{c}{2z_0} \qquad [10\cdot10]$$

Since the critical wavelength $(\lambda_0)_{01}$ is defined by the relation $(\lambda_0)_{01}(f_0)_{01} = c$, we have

$$(\lambda_0)_{01} = \frac{c}{(f_0)_{01}} = 2\pi\left(\frac{z_0}{\pi}\right) = 2z_0 \qquad [10\cdot11]$$

The resonant frequency of the cavity operating in the H_{01l} mode is readily obtained by equating 10·9 and 10·6. Thus

$$\frac{l\pi}{x_0} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{\pi}{z_0}\right)^2} \qquad [10\cdot12]$$

Squaring both sides and transposing

$$\left(\frac{\omega}{c}\right)^2 = \left(\frac{l}{x_0}\right)^2\pi^2 + \left(\frac{1}{z_0}\right)^2\pi^2 \qquad [10\cdot13]$$

Multiplying by $\left(\frac{c}{2\pi}\right)^2$ and extracting the square root

$$f_{01l} = c\sqrt{\left(\frac{l}{2x_0}\right)^2 + \left(\frac{1}{2z_0}\right)^2} \qquad [10\cdot14]$$

Equation 10·14 defines the resonant frequency f_{01l} of any rectangular cavity resonator operating in the H_{01l} mode in terms of the dimensions x_0 and z_0 and the index number l. It is independent of y_0 because the field does not vary in the Y direction, just as the critical frequency of the traveling H_{01} wave in a guide is independent of y_0.

Let us determine the resonant frequency of a perfect cube. This is accomplished by setting $x_0 = y_0 = z_0$ in equation 10·14. The resulting equation is

$$f_{01l} = c\sqrt{\left(\frac{l}{2z_0}\right)^2 + \left(\frac{1}{2z_0}\right)^2} \qquad [10\cdot15]$$

$$= \frac{c}{2z_0}\sqrt{l^2+1} \qquad [10\cdot16]$$

The lowest of all resonant frequencies associated with a cube is obtained by setting $l = 1$.* Then

$$f_{011} = \frac{c}{\sqrt{2}z_0} \qquad [10\cdot17]$$

* Equations 10·7 show that H_{xr} and E_{yr} are zero if $l = 0$ and x_0 is finite. Thus the wave cannot exist and H_{sr} is necessarily zero.

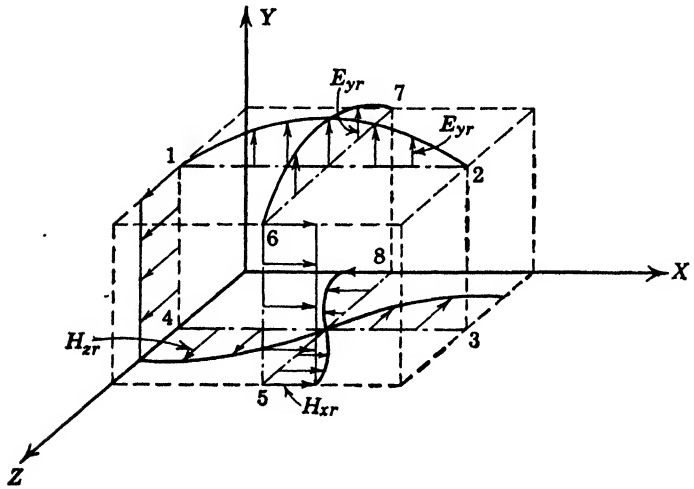
or

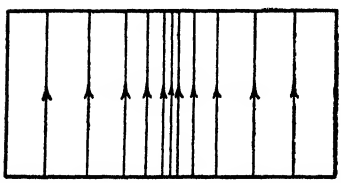
$$f_{011} = \sqrt{2}(f_0)_{01} \qquad [10\cdot18]$$

where $(f_0)_{01}$, associated with the traveling wave, is given in equation 10·10.

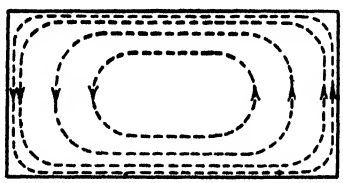
In the case of the H_{011} mode, for $m = 1$ and $l = 1$, equations 10·7 reduce to

$$\left.\begin{aligned} H_{xr} &= -j2A \cos\left(\frac{\pi}{z_0} z\right) \sin\left(\frac{\pi}{x_0} x\right) e^{j\omega t} \\ H_{zr} &= j2A \left(\frac{z_0}{\pi}\right) \beta_{01} \sin\left(\frac{\pi}{z_0} z\right) \cos\left(\frac{\pi}{x_0} x\right) e^{j\omega t} \\ E_{yr} &= 2A \left(\frac{z_0}{\pi}\right) \omega\mu_0 \sin\left(\frac{\pi}{z_0} z\right) \sin\left(\frac{\pi}{x_0} x\right) e^{j\omega t} \end{aligned}\right\} \begin{array}{l} H_{011} \\ \text{resonant waves} \\ \text{in pipe} \\ [10\cdot19] \end{array}$$





Electric Intensity Vector Distribution in Plane 1234 or 5678



Magnetic Intensity Vector Distribution in Horizontal Plane Top View

H_{011} Wave

FIG. 10·4 Instantaneous distribution of the intensity vectors in a rectangular cavity resonator when excited in the H_{011} mode.

The distribution of H_{011} wave vectors in the rectangular cavity is shown in Fig. 10·4; higher-order H_{0ml} waves may be determined directly from equations 10·7. The distribution of the vectors inside the cavity may be determined directly since the number for the indices of the H_{nml} wave indicate immediately the number of half wavelengths along the respective directions. As an example, the H_{023} wave distribution for the E_{yr} vector is given in Fig. 10·5.

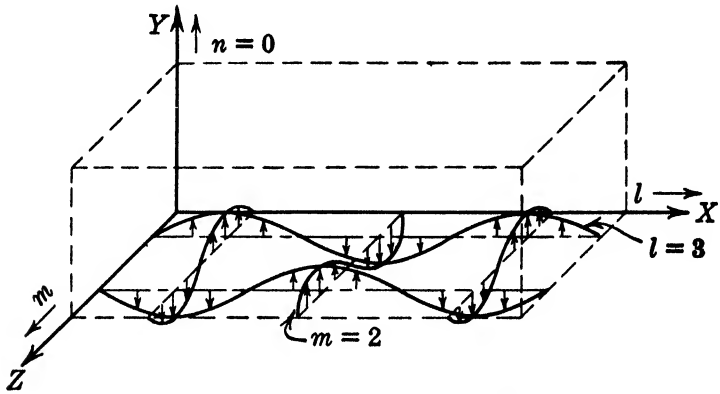


FIG. 10·5 H_{023} wave resonance in a rectangular cavity.

The general equations for the E_{nml} or H_{nml} modes may be obtained from the equations of Chapter 6, and will not be given here.

It is seen from the preceding that, if a section of rectangular pipe of length x_0 is made into a resonant chamber by closing both ends with conducting plates, the boundary conditions can, in the general case, be satisfied by a proper choice of frequency.

The resonant frequency of the waves which exist in the general H_{nml} case is obtained by equating equations 10·8 and 10·16; thus

$$f_{nml} = c\sqrt{\left(\frac{n}{2y_0}\right)^2 + \left(\frac{m}{2z_0}\right)^2 + \left(\frac{l}{2x_0}\right)^2} \qquad [10\cdot20]$$

where n, m, and l are integers. A distinct resonance will exist for each combination of n, m, and l, with the qualification that only one of these integers may be zero for H modes and none may be zero for E modes. Hence an infinite number of oscillatory modes is theoretically possible. Usually only the lowest modes are of practical interest in view of the extremely high frequencies involved. To distinguish between the various resonant modes, the notation TE_{nml} or H_{nml} for the transverse electric modes, and TM_{nml} or E_{nml} for the transverse magnetic modes,

may be used. The subscripts n, m, and l refer to the number of half-period variations of the intensity components in the y, z, and x directions, respectively.

10·3 Power Relations in a Rectangular Cavity (Imperfect Conductivity)

The energy stored in electric and magnetic fields per unit volume as determined in Chapter 1 is given by the expressions

$$\tfrac{1}{2}\varepsilon E^2 \qquad \text{and} \qquad \tfrac{1}{2}\mu H^2 \tag{10·21}$$

respectively. In air, the quantities become $\frac{1}{2}\varepsilon_0 E^2$ and $\frac{1}{2}\mu_0 H^2$. By integrating these relations over the volume of the cavity we may determine the energy stored in the electric and magnetic fields within the cavity. Thus

$$W_E = \frac{\varepsilon_0}{2}\int_{x=0}^{x_0}\int_{y=0}^{y_0}\int_{z=0}^{z_0} |E_r'|^2\, dx\, dy\, dz \tag{10·22}$$

and

$$W_M = \frac{\mu_0}{2}\int_{x=0}^{x_0}\int_{y=0}^{y_0}\int_{z=0}^{z_0} |H_r'|^2\, dx\, dy\, dz \tag{10·23}$$

where W_E and W_M represent respectively the maximum electric or magnetic energy stored in the cavity. In the case of the H_{0ml} wave considered in the previous section, the electric vector $E = E_{yr}$, and by substitution in 10·22 the maximum value of the energy stored in the electric field becomes

$$W_E = \frac{\varepsilon_0}{2}\left[2A\left(\frac{z_0}{m\pi}\right)\mu_0\omega_{0ml}\right]^2 \int_0^{x_0}\int_0^{y_0}\int_0^{z_0} \sin^2\left(\frac{m\pi}{z_0}z\right)\sin^2\left(\frac{l\pi}{x_0}x\right) dx\, dy\, dz^*$$

$$W_E = \frac{\varepsilon_0}{2}\left[4A^2\frac{z_0^2}{m^2\pi^2}\mu_0^2\omega_{0ml}^2\right]\left[(y_0)\left(\frac{z_0}{2}\right)\left(\frac{x_0}{2}\right)\right]$$

From Chapter 3

$$c^2 = \frac{1}{\varepsilon_0\mu_0}$$

Then,

$$W_E = \frac{\mu_0 y_0 z_0^3 x_0 A^2}{\pi^2 2m^2}\left(\frac{\omega_{0ml}}{c}\right)^2.$$

$$\int_0^{z_0}\sin^2\left(\frac{m\pi}{z_0}z\right)dz = \left(\frac{z_0}{m\pi}\right)\left[\left(\frac{m\pi}{z_0}\right)\frac{z}{2} - \frac{1}{4}\sin^2\left(\frac{m\pi}{z_0}z\right)\right]\Bigg|_0^{z_0} = \frac{z_0}{2}$$

From 10·20 with $n = 0$, we have:

$$\left[\frac{2\pi f_{0ml}}{c}\right]^2 = \left[\frac{\omega_{0ml}}{c}\right]^2 = 4\pi^2\left[\left(\frac{m}{2z_0}\right)^2 + \left(\frac{l}{2x_0}\right)^2\right]$$

Therefore

$$W_E = \frac{A^2 z_0^3 \mu_0 y_0 x_0 \pi^2}{2m^2\pi^2}\left[\frac{m^2}{z_0^2} + \frac{l^2}{x_0^2}\right] = \frac{A^2\mu_0 y_0 z_0 x_0}{2}\left[1 + \left(\frac{z_0 l}{x_0 m}\right)^2\right] \quad [10\cdot24]$$

giving the maximum instantaneous value of electrical energy stored in the cavity.

The magnetic intensity vector is given in magnitude by

$$|H| = \sqrt{|H_{xr}|^2 + |H_{zr}|^2}$$

Hence the maximum value of the energy stored in the magnetic field is

$$W_M = \frac{\mu_0}{2}\int_0^{x_0}\int_0^{y_0}\int_0^{z_0} 4A^2\cos^2\left(\frac{m\pi}{z_0}z\right)\sin^2\left(\frac{l\pi}{x_0}x\right)dx\,dy\,dz$$
$$+ \frac{\mu_0}{2}\int_0^{x_0}\int_0^{y_0}\int_0^{z_0} 4A^2\left(\frac{z_0}{m\pi}\right)^2\beta_{0m}^2\sin^2\left(\frac{m\pi}{z_0}z\right)\cos^2\left(\frac{l\pi}{x_0}x\right)dx\,dy\,dz \quad [10\cdot25]$$

By an evaluation similar to that used in 10·24 with m and l as integers, the first term of the right member of 10·25 reduces to

$$\frac{y_0\mu_0 4A^2}{2}\left(\frac{z_0}{2}\right)\left(\frac{x_0}{2}\right) \quad [10\cdot25a]$$

and the second term, using 10·6 becomes

$$\frac{y_0\mu_0 4A^2}{2}\left(\frac{z_0}{m\pi}\right)^2\left(\frac{l\pi}{x_0}\right)^2\left(\frac{z_0}{2}\right)\left(\frac{x_0}{2}\right) \quad [10\cdot26]$$

Collecting, the maximum instantaneous value of magnetic energy stored in the cavity is

$$W_M = \frac{A^2\mu_0 y_0 z_0 x_0}{2}\left[1 + \left(\frac{z_0 l}{x_0 m}\right)^2\right] \quad [10\cdot27]$$

which, it will be observed, is identical with 10·24, thus satisfying the basic requirement for the equality of the magnetic and electric energies. The energy within the cavity oscillates from the electric to the magnetic form. At the instant when the magnetic energy is maximum the electric energy is zero, and vice versa. Thus the energy at any instant is the sum of instantaneous electric and instantaneous magnetic energies.

We may calculate the average power lost in the side walls of the cavity by means of the Poynting vector. Let us designate the power

flowing into the respective surfaces, Fig. 10·6, as follows:

$$\text{Power loss} \begin{cases} P_x = \text{loss at both ends} \\ P_y = \text{loss at top and bottom} \\ P_z = \text{loss at both sides} \end{cases} \qquad [10\cdot28]$$

From the Poynting theorem we know that the power loss into a metal surface* may be defined as

$$\frac{1}{2}\sqrt{\frac{\pi f \mu_2}{\sigma_2}} \iint_s |H'_{\tan}|^2 \, ds \qquad [10\cdot29]$$

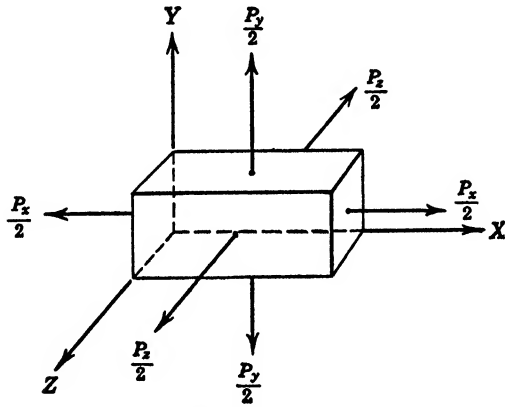


FIG. 10·6 Components of the average power flowing into the walls of a rectangular cavity resonator.

From equations 5·128, 5·129, and 5·130,

$$E = \eta H = \sqrt{\frac{j\omega\mu_2}{\sigma_2}} H$$

where η is the intrinsic impedance of the metal. Since $\sqrt{j2\pi f} = (1+j)\sqrt{\pi f}$, and only the real part is considered,

$$E = \eta H = \sqrt{\frac{\pi f \mu_2}{\sigma_2}} H$$

The power flowing (equation 5·153) into the metal per square meter of surface is

$$p = \mathcal{R}\,[\tfrac{1}{2}(\mathbf{E}_{\tan} \times \mathbf{H}^*_{\tan})] = \mathcal{R}[\tfrac{1}{2}\eta \mathbf{H}'_{\tan} \times \mathbf{H}'_{\tan}] = \mathcal{R}\tfrac{1}{2}\eta \, |H'_{\tan}|^2$$

and for the entire surface

$$P = \frac{1}{2}\iint_s \eta \, |H'_{\tan}|^2 \, ds = \frac{1}{2}\sqrt{\frac{\pi f \mu_2}{\sigma_2}} \iint_s |H'^2_{\tan}| \, ds$$

where $H'_{\tan}$ is the amplitude of the tangential component of H at the surface.

It will be observed from this definition that only tangential components in the particular face being studied are those contributing to the integral. These components in the positive directions are shown in Fig. 10·7. Their magnitude upon reflection is the same, and twice the power computed for one face will then be the power dissipation for both.

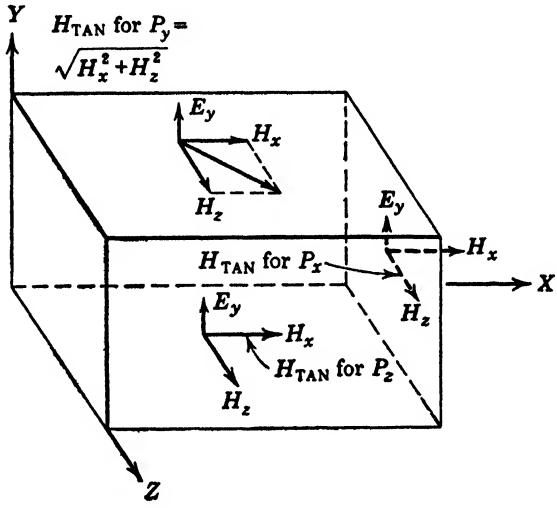


FIG. 10·7 Instantaneous distribution of the vectors of an H_{nm} wave just inside the various surfaces of a rectangular cavity.

Therefore with H_z as the tangential component to the YZ surface, we have for both faces the power loss, using equation 10·7 for the H_{0ml} modes,

$$P_x = 2(\tfrac{1}{2})\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\int_{y=0}^{y_0}\int_{z=0}^{z_0} 4A^2\left(\frac{z_0}{m\pi}\right)^2 \beta_{0m}^2 \sin^2\left(\frac{m\pi}{z_0}z\right)\cos^2\left(\frac{l\pi}{x_0}x\right)dz\,dy$$

Integrating, and substituting for β_{0m} its value as given in equation 10·6,

$$P_x = 2(\tfrac{1}{2})\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left[\frac{4A^2z_0^2}{m^2\pi^2}\right]\left(\frac{l^2\pi^2}{x_0^2}\right)(1)(y_0)\left(\frac{z_0}{2}\right) \qquad [10\cdot29a]^*$$

or, cancelling and rearranging

$$P_x = \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,(2A^2y_0z_0)\left(\frac{z_0l}{x_0m}\right)^2 \qquad [10\cdot30]$$

* $\cos^2\left(\frac{l\pi}{x_0}x\right) = 1$, since at either face when $x = 0$, or x_0, and l is an integer, the result is unity.

Since H_x is the only component tangential to the XY surface, we have, for the power loss in these surfaces for the H_{oml} modes,

$$P_z = 2(\tfrac{1}{2})\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\int_{x=0}^{x_0}\int_{y=0}^{y_0} 4A^2\cos^2\left(\frac{m\pi}{z_0}z\right)\sin^2\left(\frac{l\pi}{x_0}x\right)dx\,dy$$

$$= \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,(4A^2)\left(\frac{x_0}{2}\right)(1)(y_0) = \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\,(A^2 2y_0x_0) \qquad [10\cdot31]$$

The tangential component of H in the XZ surfaces is the resultant of H_x and H_z. Power dissipated in the total Y surface is

$$P_y = 2\left(\frac{\eta}{2}\right)\int_s\int \{|H'_{xr}|^2 + |H'_{zr}|^2\}\,ds$$

$$= \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left\{\int_{z=0}^{z_0}\int_{x=0}^{x_0} 4A^2\left(\frac{z_0}{m\pi}\right)^2\beta_{0m}^2\sin^2\left(\frac{m\pi}{z_0}z\right)\cos^2\left(\frac{l\pi}{x_0}x\right)dx\,dz\right.$$

$$\left. + \int_{z=0}^{z_0}\int_{x=0}^{x_0} 4A^2\cos^2\left(\frac{m\pi}{z_0}z\right)\sin^2\left(\frac{l\pi}{x_0}x\right)dx\,dz\right\}$$

$$= \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left\{\left[\frac{4A^2z_0^2}{m^2\pi^2}\cdot\frac{l^2\pi^2}{x_0^2}\right]\left[\frac{x_0}{2}\right]\left[\frac{z_0}{2}\right] + 4A^2\left[\frac{x_0}{2}\right]\left[\frac{z_0}{2}\right]\right\}$$

$$P_y = \sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left\{A^2z_0x_0\left[1+\left(\frac{z_0l}{x_0m}\right)^2\right]\right\} \qquad [10\cdot32]$$

The total loss P is therefore given by the sum of P_x, P_y, and P_z. Thus

$$P = \sqrt{\frac{\pi f\mu_2}{\sigma_2}}A^2\left\{2y_0z_0\left(\frac{z_0l}{x_0m}\right)^2 + z_0x_0\left[1+\left(\frac{z_0l}{x_0m}\right)^2\right] + 2y_0x_0\right\} \qquad [10\cdot33]$$

The selectivity or Q of the resonator is defined in an analogous way to that of the normal series resonant circuit. The Q of a coil may be expressed as

$$Q = \frac{\omega L}{R} = \omega\frac{[\frac{1}{2}LI^2]}{[\frac{1}{2}RI^2]} \qquad [10\cdot34]$$

Thus

$$Q = \omega\frac{W_M}{P} = \frac{2\pi}{T}\frac{W_M}{P} \qquad [10\cdot35]$$

where R and L are the resistance and inductance of the coil, I is the maximum instantaneous value of current, W_M is the maximum energy stored during the cycle, P is the power loss, and $T = 1/f$ is the period.

Now PT is the energy loss in one period, and hence in general

$$Q = 2\pi \frac{\text{Maximum energy stored in} \left\{ \begin{matrix} \text{electric} \\ \text{or} \\ \text{magnetic} \end{matrix} \right\} \text{field}}{\text{Energy loss in one period}} \qquad [10\cdot36]$$

Applying this definition of Q to the rectangular-cavity resonator we have

$$Q = \frac{2\pi}{T} \frac{W_M}{P_x + P_y + P_z}$$

Replacing W_M by its value in equation 10·27 and P_x, P_y, P_z by equation 10·33, Q becomes for any resonance of type H_{oml}

$$\frac{2\pi \dfrac{A^2}{2} \mu_0 y_0 z_0 x_0 \left[1 + \left(\dfrac{z_0 l}{x_0 m}\right)^2\right]}{\dfrac{1}{f}\sqrt{\dfrac{\pi f \mu_2}{\sigma_2}} A^2 \left\{2y_0 z_0 \left(\dfrac{z_0 l}{x_0 m}\right)^2 + 2y_0 x_0 + z_0 x_0 \left[1 + \left(\dfrac{z_0 l}{x_0 m}\right)^2\right]\right\}} \qquad [10\cdot37]$$

which upon cancellation and rearrangement becomes

$$\sqrt{\frac{\pi \mu_0^2 \sigma_2}{\mu_2}} \sqrt{f} \left\{ \frac{y_0 z_0 x_0 \left[1 + \left(\dfrac{z_0 l}{x_0 m}\right)^2\right]}{2y_0 z_0 \left(\dfrac{z_0 l}{x_0 m}\right)^2 + 2y_0 x_0 + z_0 x_0 \left[1 + \left(\dfrac{z_0 l}{x_0 m}\right)^2\right]} \right\} \qquad [10\cdot38]$$

In the case of the H_{011} mode, for $l = m = 1$, and in the particular case for $z_0 = x_0$, 10·30 reduces to

$$Q = \sqrt{\frac{\pi \mu_0^2 \sigma_2}{\mu_2}} \sqrt{f} \left\{ 2 \frac{y_0 z_0 x_0}{2y_0 z_0 + 2y_0 x_0 + 2z_0 x_0} \right\} \qquad [10\cdot39]$$

$$Q = \sqrt{\frac{\pi \mu_0^2 \sigma_2}{\mu_2}} \sqrt{f} \left\{ 2 \frac{\text{Volume}}{\text{Surface area}} \right\} \qquad [10\cdot40]$$

At the critical frequency for the H_{011} mode of oscillation

$$f_{011} = \frac{1}{z_0 \sqrt{2\mu_0 \varepsilon_0}} \qquad [10\cdot41]$$

At this critical frequency, Q may be expressed as

$$Q_0 = 28.9 \sqrt{\sigma_2} \frac{y_0 \sqrt{z_0}}{2y_0 + z_0} \qquad z_0 = x_0 \quad \text{and} \quad \mu_2 = \mu_0 \qquad [10\cdot42]$$

A plot of Q_0 as a function of y_0 and of z_0 is shown in Fig. 10·8 for a rectangular cavity of copper.

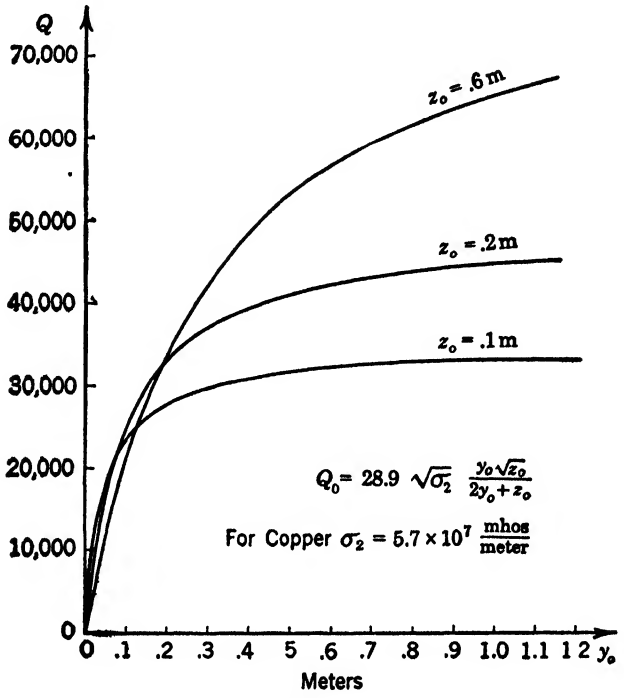


FIG 10 8 Selectivity Q of a rectangular copper cavity.

10·4 Resonance in Cylindrical Cavities

Resonance in cylindrical cavities may be treated in a manner similar to that employed for rectangular cavities in the preceding sections. Let us consider waves of the H_{0m} mode in the cylindrical cavity. Equations for these waves may be obtained from the general equations 7·104 for the H_{nm} waves in the dielectric. Setting $n = 0$, $\mu_1 = \mu_0$, and $\varepsilon_1 = \varepsilon_0$, we obtain, for the components of the traveling waves in the positive X direction, the relations

$$\left.\begin{aligned} H_{x+} &= \left\{A' J_0\left(r\frac{r'_{01}}{a}\right)\right\} e^{j(\omega t-\beta_{01}x)} \\ H_{r+} &= j\left\{A'\beta_{01}\left(\frac{a}{r'_{01}}\right) J_1\left(r\frac{r'_{01}}{a}\right)\right\} e^{j(\omega t-\beta_{01}x)} \\ E_{\phi+} &= -j\left\{A'\omega\mu_0\left(\frac{a}{r'_{01}}\right) J_1\left(r\frac{r'_{01}}{a}\right)\right\} e^{j(\omega t-\beta_{01}x)} \end{aligned}\right\} \begin{matrix} H_{0i} \\ \text{traveling} \\ \text{waves} \\ [10.43] \end{matrix}$$

Let us insert in the pipe the circular conducting plugs shown in Fig. 10·9, at $x = 0$ and $x = x_0$. Then, by virtue of reflection, we have at the conducting surface of these plugs the relations

$$\left.\begin{aligned} H_{x+} &= -H_{x-} \\ H_{r+} &= +H_{r-} \\ E_{\phi+} &= -E_{\phi-} \end{aligned}\right\}^{*} \qquad [10\cdot44]$$

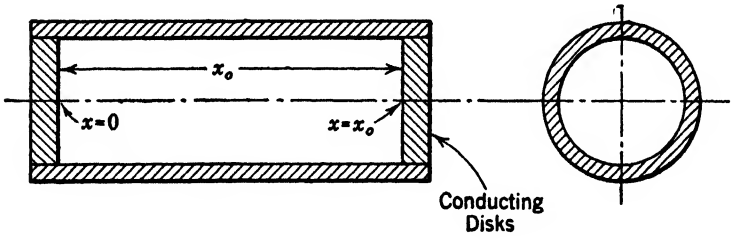


FIG. 10·9 Cylindrical cavity resonator.

The standing waves that result are, on a vector basis,

$$\begin{aligned} H_{xs} &= H_{x+} + H_{x-} \\ H_{rs} &= H_{r+} + H_{r-} \\ E_{\phi s} &= E_{\phi+} + E_{\phi-} \end{aligned} \qquad [10\cdot45]$$

Hence substituting equation 10·43 in 10·44 and using 10·45 we obtain†

$$\left.\begin{aligned} H_{xs} &= -j2A'J_0\left(r\frac{r'_{01}}{a}\right)\sin(\beta_{01}x)e^{j\omega t} \\ H_{rs} &= j2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)J_1\left(r\frac{r'_{01}}{a}\right)\cos(\beta_{01}x)e^{j\omega t} \\ E_{\phi s} &= -2A'\omega\mu_0\left(\frac{a}{r'_{01}}\right)J_1\left(r\frac{r'_{01}}{a}\right)\sin(\beta_{01}x)e^{j\omega t} \end{aligned}\right\} \begin{matrix} H_{01} \\ \text{standing} \\ \text{waves} \\ [10\cdot46] \end{matrix}$$

At the boundary plates, $x = 0$ or x_0 and $E_{\phi s} = 0$. Let us set $\beta_{01} = l\pi/x_0$, where $l = 1, 2, 3, \cdots$. Then the equation for $E_{\phi s}$ is satisfied at the boundary since sin $(\beta_{01}x)$ is zero there, and we may write the equa-

* Subscripts + and − indicate components traveling in $+x$ and $-x$ directions respectively.

† Equations 10·46 are obtained as were those of 10·5 with proper substitution and with aid of the equalities given in the footnote on page 367.

tions for the resonant waves as

$$\left.\begin{aligned} H_{xr} &= -j2A' J_0\left(r\frac{r'_{01}}{a}\right)\sin\left(\frac{l\pi}{x_0}x\right)e^{j\omega t} \\ H_{rr} &= +j2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)J_1\left(r\frac{r'_{01}}{a}\right)\cos\left(\frac{l\pi}{x_0}x\right)e^{j\omega t} \\ E_{\phi r} &= -2A'\omega\mu_0\left(\frac{a}{r'_{01}}\right)J_1\left(r\frac{r'_{01}}{a}\right)\sin\left(\frac{l\pi}{x_0}x\right)e^{j\omega t} \end{aligned}\right\} \begin{matrix} H_{01l} \\ \text{resonant} \\ \text{waves} \\ [10\cdot47] \end{matrix}$$

The phase constant in the general case for waves of the H_{nm} type, as given in equation 7·105, is

$$\beta_{nm} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{r'_{nm}}{a}\right)^2} \qquad [10\cdot48]$$

For the H_{01} wave, $n = 0$, $m = 1$, and

$$\beta_{01} = \sqrt{\left(\frac{\omega}{c}\right)^2 - \left(\frac{r'_{01}}{a}\right)^2} \qquad [10\cdot48a]$$

The critical frequency f_0 for the cylindrical H_{01} wave is obtained, as before, by setting $\beta_{01} = 0$. Then

$$(f_0)_{01} = \frac{c}{2\pi}\left(\frac{r'_{01}}{a}\right) \qquad [10\cdot49]$$

The critical wavelength λ_0 is given by the relation $\lambda_0 f_0 = c$. Hence

$$(\lambda_0)_{01} = \frac{c}{f_0} = 2\pi\left(\frac{a}{r'_{01}}\right) \qquad [10\cdot50]$$

The wavelength λ of a wave in free space of frequency f is $\lambda = c/f$ Hence $\lambda/2\pi = c/\omega$, and with the aid of 10·50, equation 10·48a may be written

$$\beta_{01} = \sqrt{\left(\frac{2\pi}{\lambda}\right)^2 - \left(\frac{2\pi}{\lambda_0}\right)^2} = 2\pi\sqrt{\left(\frac{1}{\lambda}\right)^2 - \left(\frac{1}{\lambda_0}\right)^2} = \frac{2\pi}{c}\sqrt{f^2 - (f_0)^2_{01}} \qquad [10\cdot51]$$

where λ_0 is the critical wavelength and λ is the wavelength in free space. From the boundary condition at the end disks closing the cylinder, $\beta_{01} = l\pi/x_0$. Equating this and 10·51,

$$\left(\frac{l\pi}{x_0}\right)^2 = \left(\frac{2\pi}{c}\right)^2\{f^2 - (f_0)^2_{01}\}$$

Solving for f we obtain

$$f_{01l} = \sqrt{(f_0)^2_{01} + \left(\frac{lc}{2x_0}\right)^2} = \frac{c}{2}\sqrt{\left(\frac{r'_{01}}{\pi a}\right)^2 + \left(\frac{l}{x_0}\right)^2} \qquad [10\cdot52]$$

For the more general case of the H_{nml} mode, the waves at the boundary require that

$$\beta_{nm} = \frac{l\pi}{x_0}$$

Equating this to 10·48,

$$\left(\frac{l\pi}{x_0}\right)^2 = \left(\frac{\omega}{c}\right)^2 - \left(\frac{r'_{nm}}{a}\right)^2 \qquad [10\cdot53]$$

Solving for ω we obtain

$$\omega_{nml} = c\sqrt{\left(\frac{l\pi}{x_0}\right)^2 + \left(\frac{r'_{nm}}{a}\right)^2} \qquad [10\cdot54]$$

or

$$f_{nml} = \sqrt{\left(\frac{cl}{2x_0}\right)^2 + \left[\left(\frac{c}{2\pi}\right)\left(\frac{r'_{nm}}{a}\right)\right]^2} = \sqrt{\left(\frac{cl}{2x_0}\right)^2 + (f_0)^2_{nm}} \qquad [10\cdot55]$$

where

$$(f_0)_{nm} = \frac{c}{2\pi}\left(\frac{r'_{nm}}{a}\right) \quad \text{in the general } n,m, \text{ case} \qquad [10\cdot56]$$

Equation 10·55 is readily converted to a form very convenient for calculation or graphic representation

$$(2af_{nml})^2 = \left(\frac{cr'_{nm}}{\pi}\right)^2 + \left(\frac{cl}{2}\right)^2\left(\frac{a}{x_0}\right)^2 \qquad [10\cdot57]$$

10·5 Power Relations in a Cylindrical Cavity

Let us calculate the maximum energy W_E stored in the electric field in the H_{011} case.

$$W_E = \frac{\varepsilon_0}{2}\int_{\phi=0}^{2\pi}\int_{r=0}^{a}\int_{x=0}^{x_0} |E'_{\phi r}|^2\, r\, d\phi\, dr\, dx \qquad [10\cdot58]$$

From 10·47 we substitute $E_{\phi r}$ and obtain, when $l = 1$

$$W_E = \frac{\varepsilon_0}{2}\left[4A'^2\omega^2\mu_0^2\left(\frac{a}{r'_{01}}\right)^2\right]\int_0^{2\pi}\int_0^a\int_0^{x_0} rJ_1^2\left(r\frac{r'_{01}}{a}\right)\sin^2\left(\frac{\pi}{x_0}x\right)d\phi\, dr\, dx$$

Writing $\mu_0\varepsilon_0 = 1/c^2$, evaluating $\int_{\phi=0}^{2\pi} d\phi$ as 2π, and writing the integral for $\sin^2\left(\frac{\pi}{x_0}x\right)$ as shown, we have

$$W_E = 2\pi 2A'^2\mu_0\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2\frac{x_0}{\pi}\left[\frac{\pi}{2}\cdot\frac{x}{x_0} - \frac{1}{4}\sin\left(\frac{2\pi x}{x_0}\right)\right]\Bigg|_0^{x_0}\int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right)dr$$

$$= 2A'^2\mu_0\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2\pi x_0\int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right)dr \qquad [10\cdot58a]$$

Now

$$\int_0 rJ_1^2\left(r\frac{r'_{01}}{a}\right)dr = \frac{r^2}{2}\left[\left\{J'_1\left(r\frac{r'_{01}}{a}\right)\right\}^2 + \left(1 - \frac{1}{\left(r\frac{r'_{01}}{a}\right)^2}\right)\left\{J_1\left(r\frac{r'_{01}}{a}\right)\right\}^2\right]\Bigg|_0^a *$$

which for $r = a$ becomes

$$\cdot\ \frac{a^2}{2}\left[\left\{J'_1(r'_{01})\right\}^2 + \left(1 - \frac{1}{(r'_{01})^2}\right)\left\{J_1(r'_{01})\right\}^2\right]$$

But†

$$k\,J'_n(kx) = \frac{n}{x}J_n(kx) - J_{n+1}\,k(kx)$$

$$J'_1(r'_{01}) = \frac{1}{r'_{01}}J_1(r'_{01}) - J_2(r'_{01}) \qquad [10\cdot58b]$$

Then,

$$\int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right)dr = \tfrac{1}{2}a^2\left\{\frac{1}{(r'_{01})^2}J_1^2(r'_{01}) - \frac{2}{r'_{01}}J_1(r'_{01})J_2(r'_{01}) + J_2^2(r'_{01}) + \left[1 - \left(\frac{1}{r'_{01}}\right)^2\right]\left[J_1^2(r'_{01})\right]\right\}$$

But‡ $J_1(r'_{01}) = 0$; then, since only the second-order function is not zero, the evaluation of the integral becomes

$$\int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right)dr = \frac{a^2}{2}J_2^2(r'_{01}) \qquad [10\cdot58c]$$

Substituting 10·58c for the integral in 10·58a,

$$W_E = 2A'^2\mu_0\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2\pi x_0\left[\frac{a^2}{2}J_2^2(r'_{01})\right]$$

$$= A'^2\mu_0\left(\frac{\omega}{c}\right)^2\pi x_0 a^2\left[J_2^2(r'_{01})\right]\left(\frac{a}{r'_{01}}\right)^2$$

* This evaluation is a special form of the Lommel integral and is valid for an order of the Bessel function not less than -1. For further reference see McLachlan, *Bessel Functions for Engineers*, Oxford Press, 1934, Chapter VI, pages 94–97

† Cf. page 247, footnote.

‡ From above footnote, when $n = 0$ and with $k = 1$ and $x = r'_{01}$ we have

$$J'_0(r'_{01}) = -J_1(r'_{01})$$

But from page 256, footnote, r'_{01} is a root of $J'_0(r'_{01})$. Hence $J'_0(r'_{01}) = -J_1(r'_{01}) = 0$.

Now $J_0(r'_{01}) = -J_2(r'_{01})$,* so that

$$W_E = A'^2\mu_0\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2 \pi x_0 a^2 J_0^2(r'_{01})$$

Since $\pi^2 a x_0$ is the volume V of the cylindrical cavity

$$W_E = \left(\frac{a}{r'_{01}}\right)^2 VA'^2\mu_0\left(\frac{\omega}{c}\right)^2 J_0^2(r'_{01}) \qquad [10\cdot59]$$

giving an expression for the maximum energy stored in the electric field.

Let us now calculate, with the aid of the Poynting vector, the average power entering the various walls of the tube. Let P_x be the power entering the two end conductors, and P_r the power entering the cylindrical surface. Then P, the total power lost in the cavity, is

$$P = P_x + P_r$$

Now

$$P_r = \tfrac{1}{2}\eta \int_{x=0}^{x=x_0}\int_{\phi=0}^{\phi=2\pi} |H'_{\tan}|^2\, a\, d\phi\, dx \qquad [10\cdot60]\dagger$$

where the intrinsic impedance $\eta = \sqrt{\pi f\mu_2/\sigma_2}$ for the metal, and $a\, d\phi\, dx$ is the element of surface of the cylindrical wall. The maximum value of the tangential component at the surface may be found from equations 10·47. Setting $l = 1$ and $r = a$, $|H'_{\tan}|$ becomes in this case

$$|H'_{\tan}| = |H'_{xr}| = 2A' \sin\left(\frac{\pi}{x_0}x\right) J_0(r'_{01})$$

* From footnote on page 248 with $k = 1$ and $x = r'_{01}$

$$J'_n(r'_{01}) = \frac{n}{r'_{01}} J_n(r'_{01}) - J_{n+1}(r'_{01}) \qquad (a)$$

In like manner it can be shown that

$$J'_n(r'_{01}) = -\frac{n}{r'_{01}} J_n(r'_{01}) + J_{n-1}(r'_{01}) \qquad (b)$$

Subtracting (a) and (b) and rearranging

$$\frac{2n}{r'_{01}} J_n(r'_{01}) = J_{n+1}(r'_{01}) + J_{n-1}(r'_{01})$$

When $n = 0$ this reduces to $J_2(r'_{01}) = -J_0(r'_{01})$.

† See footnote on page 374.

Hence

$$P_r = \tfrac{1}{2}\eta[2A'J_0(r'_{01})]^2 a \int_{x=0}^{x_0} \sin^2\left(\frac{\pi}{x_0}x\right) dx \int_{\phi=0}^{2\pi} d\phi$$

$$= \tfrac{1}{2}\eta[2A'J_0(r'_{01})]^2 2\pi a \left[\frac{x_0}{\pi}\left\{\left(\frac{\pi x}{2x_0}\right) - \tfrac{1}{4}\sin\frac{2\pi x}{x_0}\right\}\right]_{x=0}^{x=x_0}$$

$$= \eta A'^2(2\pi a x_0)J_0^2(r'_{01}) \qquad [10\cdot60a]$$

Since the cylindrical surface area $A_r = 2\pi a x_0$, we may write

$$P_r = \eta A'^2 A_r J_0^2(r'_{01}) \qquad [10\cdot60b]$$

Also, for P_x, we write for both surfaces

$$P_x = 2\left\{\tfrac{1}{2}\eta \int_{\phi=0}^{2\pi}\int_{r=0}^{a} |H'_{rr}|^2 r\, d\phi\, dr\right\} \qquad [10\cdot61]$$

where H'_{rr} is the amplitude of the tangential component of the magnetic intensity at the surface of the conducting end disks. From equations 10·47

$$H_{rr} = j2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)\cos\left(\frac{l\pi}{x_0}x\right) J_1\left(r\frac{r'_{01}}{a}\right) e^{j\omega t} \qquad [10\cdot62]$$

At the end disks $x = 0$ or x_0, and the amplitude H'_{rr} may be written

$$|H'_{rr}| = 2A'\beta_{01}\left(\frac{a}{r'_{01}}\right) J_1\left(r\frac{r'_{01}}{a}\right) \qquad [10\cdot63]^*$$

Hence equation 10·61 becomes

$$P_x = (2)\tfrac{1}{2}\eta\left[2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)\right]^2 \int_{r=0}^{a} rJ_1^2\left(r\frac{r'_{01}}{a}\right) dr \int_{\phi=0}^{2\pi} d\phi \qquad [10\cdot64]$$

$$= \eta\left[2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)\right]^2 2\pi \int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right) dr \qquad [10\cdot65]$$

$$= \eta\left[2A'\beta_{01}\left(\frac{a}{r'_{01}}\right)\right]^2 \pi a^2 J_2^2(r'_{01}) \qquad [10\cdot66]\dagger$$

* Since $\cos\left(\frac{l\pi}{x_0}x\right) = \pm 1$ for $x = 0$ or x_0.

† Since $\int_0^a rJ_1^2\left(r\frac{r'_{01}}{a}\right) dr = \frac{a^2}{2}J_2^2(r'_{01})$ as given in equation 10·58c and preceding development.

Writing A_x as the area πa^2 of the circular end disks, we have

$$P_x = \eta 4A'^2\beta_{01}^2 \left(\frac{a}{r'_{01}}\right)^2 A_x J_2^2(r'_{01}) \qquad [10\cdot67]$$

$$= \eta 4A'^2\beta_{01}^2 \left(\frac{a}{r'_{01}}\right)^2 A_x J_0^2(r'_{01}) \qquad [10\cdot68]^*$$

We may now calculate the Q of the cylindrical cavity. Using the definition of 10·35 we have

$$Q = \frac{\omega W_E}{P} = \frac{\omega W_E}{P_r + P_x} \qquad [10\cdot69]$$

Hence from 10·59, 10·60*b*, and 10·68* the selectivity for the H_{011} mode is

$$Q = \frac{\omega\left[\left(\frac{a}{r'_{01}}\right)^2 VA'^2\mu_0\left(\frac{\omega}{c}\right)^2 J_0^2(r'_{01})\right]}{\eta\left[\left\{A'^2A_r J_0^2(r'_{01})\right\} + \left\{4A'^2\beta_{01}^2\left(\frac{a}{r'_{01}}\right)^2 A_x J_0^2(r'_{01})\right\}\right]} \qquad [10\cdot70]$$

Writing η as in the footnote on page 374, equation 10·70, becomes, after cancellation,

$$Q = \frac{\left\{\omega\mu_0\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2\right\} V}{\sqrt{\frac{\pi f\mu_2}{\sigma_2}}\left\{A_r + 4A_x\beta_{01}^2\left(\frac{a}{r'_{01}}\right)^2\right\}} \qquad [10\cdot71]$$

Further combination may be effected. $1/\eta$ together with $\omega\mu_0$ gives

$$2\sqrt{\frac{\pi f\sigma_2\mu_0^2}{\mu_2}} \qquad [10\cdot72]$$

also, from 10·48*a*

$$\beta_{01}^2 = \left(\frac{\omega}{c}\right)^2 - \left(\frac{r'_{01}}{a}\right)^2 = \left(\frac{\pi}{x_0}\right)^2$$

giving

$$\left(\frac{\omega}{c}\right)^2 = \left(\frac{\pi}{x_0}\right)^2 + \left(\frac{r'_{01}}{a}\right)^2 \qquad [10\cdot73]$$

Now $f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{\mu_0\varepsilon_0}}\left(\frac{\omega}{c}\right)$, so that, together with the above two rela-

* $J_2(r'_{01}) = \frac{2}{r'_{01}} J_1(r'_{01}) - J_0(r'_{01}) = -J_0(r'_{01})$

tions, 10·71 becomes

$$Q = 2\sqrt{\frac{\pi f \sigma_2 \mu_0^2}{\mu_2}} \left\{ \frac{\left[\left(\frac{\pi}{x_0}\right)^2 + \left(\frac{r'_{01}}{a}\right)^2\right]\left[\left(\frac{a}{r'_{01}}\right)^2\right]}{A_r + 4A_x\left[\left(\frac{\pi}{x_0}\right)^2\left(\frac{a}{r'_{01}}\right)^2\right]} \right\} V \qquad [10\cdot74]$$

And from the preceding we write for $f = \frac{1}{2\pi\sqrt{\mu_0\varepsilon_0}}\left(\frac{\omega}{c}\right)$ the expression

$$f = \frac{1}{2\pi\sqrt{\mu_0\varepsilon_0}}\sqrt{\left(\frac{\pi}{x_0}\right)^2 + \left(\frac{r'_{01}}{a}\right)^2} \qquad [10\cdot75]$$

Finally, observing that

$$\left(\frac{\omega}{c}\right)^2\left(\frac{a}{r'_{01}}\right)^2 = \left[\left(\frac{\pi a}{x_0 r'_{01}}\right)^2 + 1\right] \qquad [10\cdot76]$$

we obtain, after these substitutions are made, a general form for the Q of a cylindrical cavity oscillating in the H_{011} mode. It is

$$Q = \sqrt{2}\sqrt{\frac{\sigma_2\mu_0^2}{\mu_2\sqrt{\mu_0\varepsilon_0}}}\left\{\left(\frac{\pi}{x_0}\right)^2 + \left(\frac{r'_{01}}{a}\right)^2\right\}^{1/4}\left\{\left(\frac{\pi a}{x_0 r'_{01}}\right)^2 + 1\right\}\frac{V}{A_r + 4A_x\left\{\frac{a\pi}{x_0 r'_{01}}\right\}^2} \qquad [10.77]$$

It is seen from the above discussion that the resonant phenomena existing in cylindrical tubes may be thought of as produced by multiple reflections of traveling waves. The general equations for the transverse electric (H waves) and transverse magnetic (E waves) may be obtained in a similar manner to that employed in section 10·4.

The critical frequencies for the transmission of the E or H waves as given in Chapter 7 are different for the different transmission modes. Thus for the two classes of waves

$$(f_0)_{nm} = \frac{cr_{nm}}{2\pi a} \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot78]$$

$$(f_0)_{nm} = \frac{cr'_{nm}}{2\pi a} \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot79]$$

The quantities r_{nm} and r'_{nm} are the roots of the equations.

$$J_n(r_{nm}) = 0 \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot80]$$

$$J'_n(r'_{nm}) = 0 \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot81]$$

J_n is a Bessel function of the first kind of order n, and J'_n is its derivative with respect to the argument.

When a section of hollow tube having a length x_0 is closed at both ends by conducting surfaces, standing waves may be excited in the tube when their wavelength in the tube is given by equation 10·57.

$$\lambda = \frac{2x_0}{l} \qquad [10\cdot57]$$

where l is an integer and denotes the number of half-period variations of the intensity vectors along the axial x direction. The critical or resonant frequencies are given by equation 10·55 as

$$f_{nml} = \sqrt{\left(\frac{cl}{2x_0}\right)^2 + (f_0)^2_{nm}} \qquad [10\cdot55]$$

where the value of f_0 is given in equation 10·78 or 10·79, depending upon the type of wave being considered. As in the rectangular resonator, there will be a definite resonant frequency for each value of n, m, or l. Because of practical considerations only the lower-order modes of oscillation are important at the present time, although all are theoretically possible. In distinguishing between the different modes the notation TE_{nml} or H_{nml} and TM_{nml} or H_{nml} may be used, where n represents the number of full-period variations of the radial electric field along the ϕ direction, m represents the numbers of half-period variations of the angular component of the electric field in the radial r direction, and l represents the number of half-period variations of the radial component of electric field in the axial x direction. Since the TE_{00} or TM_{00} modes of transmission corresponding to a principal mode are not possible in cylindrical tubes, no resonant modes of the TE_{00} or TM_{00} type exist.

10·6 Resonance in Coaxial Cylindrical Cavities

As was shown in Chapter 7, the principal mode of transmission, the *TEM* mode, is possible in coaxial transmission lines. Resonance phenomena associated with this mode were discussed in Chapter 9. Other modes, corresponding to the transverse electric and transverse magnetic, are also possible modes of transmission. Resonance phenomena in coaxial cylindrical lines are similar in many respects to those of the hollow cylindrical type discussed in the preceding section and may be described in terms of standing-wave systems. The standing waves produced within the coaxial transmission line are also described by variations of the field intensities along the axial, radial, and angular coordinates. The three indices n, m, and l are required to distinguish between the various resonant modes. The expression for the frequency f_{nml} in the

coaxial system is identical with 10·55 for the hollow-pipe system. The critical frequency f_0 depends on the radii of the inner and outer conductors a and b, respectively, as well as on the order of the particular mode. As in the hollow-tube case, it also differs for waves of the *TE* and *TM* modes and is given by

$$(f_0)_{nm} = \frac{c[r_{nm}]}{2\pi b} \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot82]$$

$$(f_0)_{nm} = \frac{c[r'_{nm}]}{2\pi b} \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot83]$$

The quantities $[r_{nm}]$ and $[r'_{nm}]$ are roots of the equations

$$\frac{J_n([r_{nm}])}{Y_n([r_{nm}])} - \frac{J_n\left(\frac{a}{b}[r_{nm}]\right)}{Y_n\left(\frac{a}{b}[r_{nm}]\right)} = 0 \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot84)^*$$

$$\frac{J'_n([r'_{nm}])}{Y'_n([r'_{nm}])} - \frac{J'_n\left(\frac{a}{b}[r'_{nm}]\right)}{Y'_n\left(\frac{a}{b}[r'_{nm}]\right)} = 0 \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot85]^*$$

J_n and Y_n are Bessel functions of order n of the first and second kinds, respectively, and J'_n and Y'_n are their derivatives with respect to the arguments. It may be shown that the expressions 10·84 and 10·85 reduce to 10·80 and 10·81, when a/b approaches zero.

The resonant frequency of the ordinary coaxial mode (principal mode) is obtained by setting $n = 0 = m$. Then f_0 reduces to 0, and we obtain from 10·55

$$f_{00l} = \frac{cl}{2x_0} \qquad TEM \text{ mode} \qquad [10\cdot86]$$

Thus the resonant frequency f_{00l} is seen to be independent of the transverse dimensions of the resonator.

A special case pointed out by Barrow arises when the integer l is set equal to zero in the transverse magnetic case. From equation 10·55 we have

$$f_{nm0} = (f_0)_{nm} = \frac{c[r_{nm}]}{2\pi b} \cdot \qquad [10\cdot87]$$

* Barrow has pointed out that, for convenience in using tables, equations 10 84 and 10 85 are best rewritten with the substitution $[r_{nm}] = bx_{nm}/a$ and $[r'_{nm}] = bx'_{nm}/a$. The new roots x_{nm} and x'_{nm} are then found from the tables.

and the wavelength as given in equation 10·57 approaches infinity. From the differential equations describing this mode it can be shown that the transverse component of the electric field vanishes, leaving only the longitudinal component. Therefore no tangential component of electric intensity exists at the metal surfaces of the line section terminations. Hence their position does not affect the resonant frequency. Therefore it is independent of the length x_0 of the section. It appears that this mode is formed by transverse reflections of the waves between the cylindrical surfaces of the resonator. This degenerate case cannot apply to the transverse electric mode, since the vanishing of the transverse electric vector annihilates the field.

The configurations of the field structure for a number of associated modes in perfectly conducting hollow cylindrical and coaxial resonators as prepared by Barrow and Mieher* are shown in Fig. 10·10. These associated modes have been grouped together in the figure for ready comparison.

10·7 Spherical Cavity Resonators

Another type of resonator which lends itself to comparatively simple mathematical analysis is the spherical cavity. The solution of Maxwell's equations in the dielectric medium inside such a cavity has been known for some time.† It can be shown that oscillations of either the transverse electric (H waves) or transverse magnetic (E waves) can exist in the cavity. A critical frequency f_0 exists which depends upon the radius of the sphere as well as on the order of the particular wave type. As in the previous cases considered, we may write

$$(f_0)_{nm} = \frac{c[u_{nm}]}{2\pi a} \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot88]$$

$$(f_0)_{nm} = \frac{c[u'_{nm}]}{2\pi a} \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot89]$$

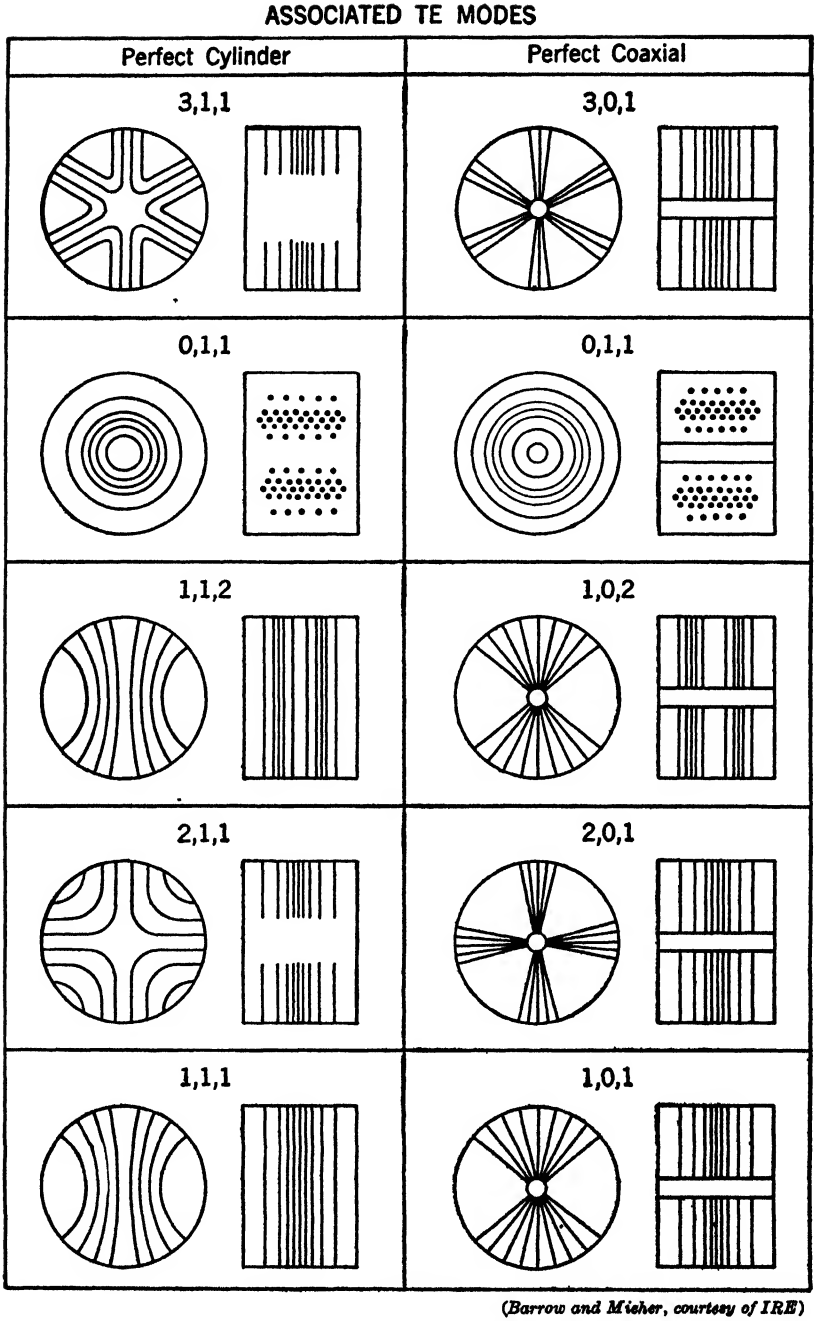
The quantities $[u_{nm}]$ and $[u'_{nm}]$ are roots of the equations

$$j_n[u_{nm}] = 0 \qquad TM_{nm} \text{ or } E_{nm} \text{ waves} \qquad [10\cdot90]$$

$$j'_n[u'_{nm}] = 0 \qquad TE_{nm} \text{ or } H_{nm} \text{ waves} \qquad [10\cdot91]$$

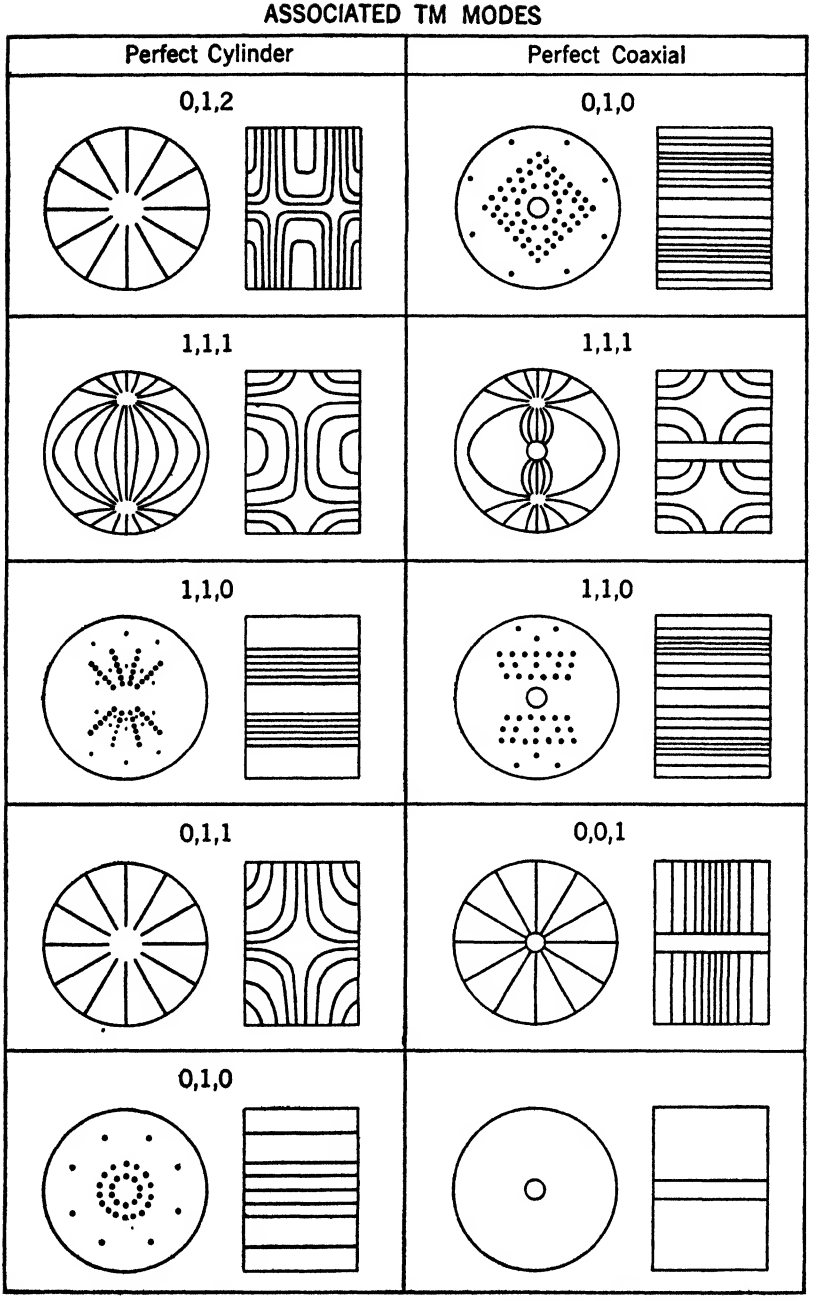
* W. L. Barrow and W. W. Mieher, "Natural Oscillations of Electrical Cavity Resonators," *Proc. IRE*, **28**, 184, 1940.

† G. Mie, *Ann. Physik*, **25**, 377, 1908; H. Bateman, *Electrical and Optical Wave Motion;* M. Born, *Optik*, p. 274 et seq.; G. Wolfsohn, *Handbuch der Physik*, **20**, 307 et seq.; W. W. Hansen, *Phys. Rev.* **47**, 129, 1935; W. W. Hansen and J. G. Beckerley, *Proc. IRE*, **24**, 1594, 1936.



(Barrow and Mieher, courtesy of IRE)

FIG. 10·10 (Part I) Sketches of the configuration of the electric field for the associated transverse electric modes in perfect cylindrical and perfect coaxial resonators.



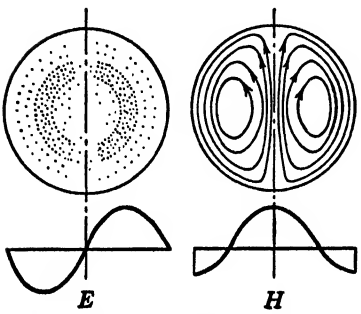
(Barrow and Mieher, courtesy of IRE)

FIG. 10·10 (Part II) Sketches of the configuration of the electric field for the associated transverse magnetic modes in perfect cylindrical and perfect coaxial resonators.

where $j_n(\rho) = \sqrt{\pi/2\rho} J_{n+1/2}(\rho)$ is a spherical Bessel function and $j_n'(\rho)$ is the derivative with respect to the argument. It may be seen on comparison of 10·88 and 10·89 with 10·78 and 10·79 that these relations are very similar to those for the hollow cylindrical cavity.

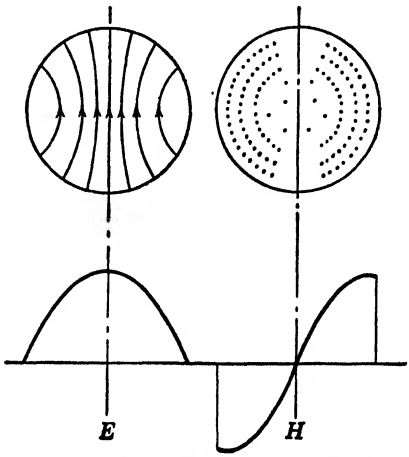
In the transverse electric mode, the wavelength in free space of the wave corresponding to the critical frequency $(f_0)_{nm}$ of equation 10·88 is

$$\lambda_{nm} = \frac{2\pi a}{[u_{nm}]} \qquad [10\cdot92]$$



(*Hansen, courtesy of J. Applied Phys.*)

FIG. 10·11 The drawing on the right represents a possible field inside a spherical conducting shell at a time when $E = 0$. The two upper drawings are cross sections of the sphere; the two lower ones show the variation with r of the fields in the equatorial plane. The drawing on the left represents the situation ¼ cycle later when the magnetic field has disappeared and all the energy exists in the electric field. Dots in the upper left figure represent ends of lines of E which is entirely in the ϕ direction.



(*Hansen, courtesy of J. Applied Phys.*)

FIG. 10·12 The drawing on the left represents a possible field inside a spherical conducting shell at a time when $H = 0$. The upper picture is a cross section of the sphere; the lower is a graph showing the variation with r of the electric field in the equatorial plane. The right-hand drawing shows the situation ¼ cycle later when the electric field has disappeared and the magnetic field has taken up the energy. The vector H is entirely in the ϕ direction, and its variation with r is shown in the graph below.

A few of the lower-order roots $[u_{nm}]$ are: $u_{11} = 4.50$, $u_{21} = 5.8$, and $u_{12} = 7.64$. Hence the longest allowable H_{nm} wave has a wavelength given by

$$\lambda_{11} = \frac{2\pi a}{4.5} = 1.4a \qquad [10\cdot93]$$

where a is the radius of the spherical cavity. The only components of the field that exist with this mode of oscillation are E_ϕ, H_r, and H_θ, having the directions of the spherical coordinates ϕ, r, and θ. Thus the charges on the conducting surface of the sphere flow along parallels of latitude. The field structure for this mode is shown in Fig. 10·11. The dots in the left side of the figure represent the electric intensity lines for the vector E_ϕ, which is entirely in the ϕ direction.

The free wavelength of the fundamental electric mode for the spherical cavity is given by the relation

$$\lambda'_{nm} = \frac{2\pi a}{[u'_{nm}]} \qquad [10\cdot94]$$

The lowest-order root $[u'_{nm}]$ is $[u'_{11}] = 2.75$, giving for the longest wavelength of the transverse magnetic mode E_{11}

$$\lambda'_{11} = \frac{2\pi a}{2.75} = 2.28a \qquad [10\cdot95]$$

This is the lowest of all possible modes of oscillation, and represents the movement of charges along meridian lines. The field distribution in a meridian plane is shown in Fig. 10·12. The components of the field that exist in this mode of oscillation are E_r, E_ϕ, and H_ϕ.

10·8 Coupling to Cavity Resonator

In making connections to external circuits for exciting electromagnetic waves in cavity resonators, and for absorbing energy from them, small coupling probes may be inserted through holes cut in the metal enclosure. Since energy will be radiated through these holes it is desirable to keep them as small as physically possible.

In general, two alternative types of coupling probes may be used. One of these consists of a loop which may be considered to provide coupling mainly with the magnetic field within the cavity. The other consists of a straight or curved rod which, acting as a small antenna, may be considered to provide coupling primarily with the electric field. The connections between the probes and external apparatus is conveniently made through coaxial lines. Illustrations of the two types of probes are shown in Fig. 10·13.

To obtain maximum coupling with the electric field, the probe of Fig. 10·13*a* should be made to coincide with the line of electric intensity. If this line is curved, the probe should be curved so as to coincide with it. The orientation of the electric field, for the particular mode used, may be obtained from a study of the field pattern. To obtain maximum coupling with the magnetic field, the plane of the loop probe of Fig.

10·13*b* must be orthogonal to the magnetic intensity vectors. To obtain a maximum linkage of magnetic flux with the loop, the correct position for the probe can be obtained as before, by a study of the field configura-

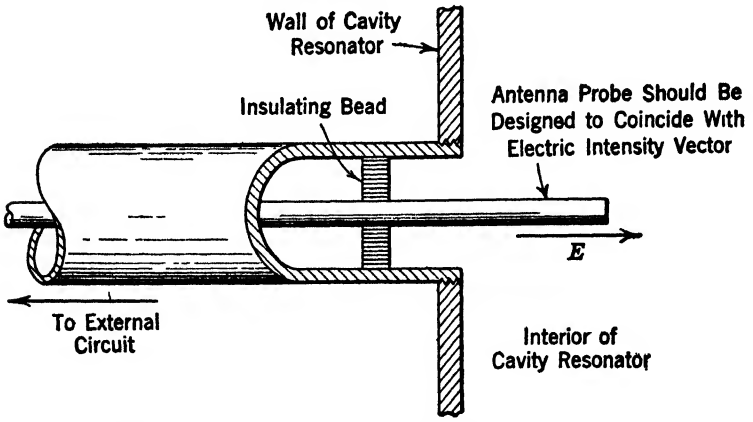


FIG. 10·13*a* Probe designed to provide coupling with an electric field.

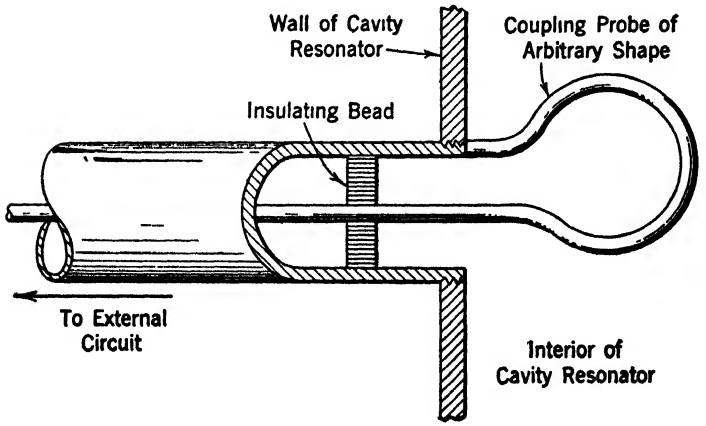


FIG. 10·13*b* Probe designed to provide coupling with a magnetic field.

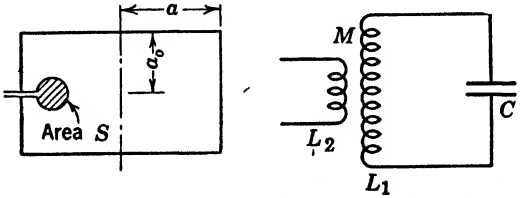
tion within the cavity resonator. Actually, the field distribution in the resonator will be influenced somewhat by the presence of the probe. This effect can be lessened by making the probe small in comparison with the wavelength of the waves in the cavity. The crystal detectors described in Chapter 8 may be employed as detecting devices.

Probes of the type shown in Fig. 10·13 may be employed in studying the field distribution in the cavity. By suitably locating and orienting these probes, the field structure may be mapped. This process is the same as that described in Chapter 8 for mapping the field configura-

tions in the various wave guides. On the other hand, when the field configuration is known, it is possible to design the probes so that they respond to one mode of oscillation and reject other modes. In Chapter 17 coupling between an electron stream and cavity resonators is described.

10·9 Equivalent Lumped Constant Circuits

The cavity resonator is equivalent in many respects to the conventional resonant circuit consisting of a coil and condenser in parallel. W. W. Hansen* has obtained expressions for the equivalent lumped circuit constants of cavity resonators when oscillating at some low-order mode. These expressions are given in Table 10·1. It is assumed in calculating this table that the loop is small compared to the wavelength and that the field and current distributions in the resonator are unaffected by the presence of the loop, and conversely. These assump-



(*Hansen, courtesy of J. Applied Phys.*)

FIG. 10·14 Resonator with coupling loop and lumped constant circuit which is equivalent.

tions are not independent, of course, but they may all be satisfied provided that the loop is very small in comparison to the size of the resonator and the wavelength. In Fig. 10·14 is shown a resonator with coupling loop and the equivalent lumped constant circuit. The mutual inductance M in Table 10·1 refers to the mutual coupling between the cavity and the loop.

10·10 Comparison of Cavity Resonators

Hansen has found that a cylinder gives about 8 per cent higher Q than a square working at the same wavelength. Also a cylinder having a height equal to its diameter gives a slightly lower Q than a sphere, although a very long cylinder is better than a sphere. It may be seen from Table 10·1 that Q is inversely proportional to δ for each of the cavities. Since δ is inversely proportional to $\sqrt{\lambda}$, Q varies as $\sqrt{\lambda}$ in each case.

It is not possible to state definitely at this time the most desirable

* W. W. Hansen, *J. Applied Physics*, 9, 654, 1938.

TABLE 10·1

COMPARISON OF VARIOUS FORMS OF CAVITY RESONATORS

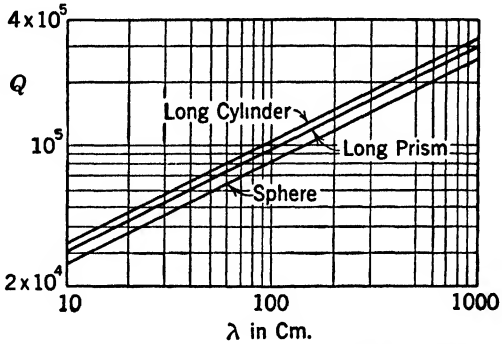
	Cylinder (E_{010})	Square (E_{111})	Sphere (E_{11})
Longest wavelength $= 2\pi c\sqrt{(LC)}$	$2\ 62\,a$	$2\ 83a$	$2\ 28a$
Equivalent $\sqrt{L/C}$	$4\ 81\dfrac{a_0}{a}c$	$4\ 29\dfrac{a_0}{a}c$	$2\ 66c$
Selectivity Q	$\sqrt{2}\dfrac{a}{\delta b}$	$\sqrt{2}\dfrac{a}{\delta b}$	$1\ 024\dfrac{a}{\delta}$
Equivalent L	$4\pi(0\ 159a_0)$	$4\pi(0\ 154a_0)$	$4\pi(0\ 0768a)$
Equivalent C	$\dfrac{1}{4\pi c^2}\left(1\ 09\dfrac{a^2}{a_0}\right)$	$\dfrac{1}{4\pi c^2}\left(1\ 32\dfrac{a^2}{a_0}\right)$	$\dfrac{1}{4\pi c^2}(1\ 72a)$
Equivalent series R	$0\ 450\dfrac{a_0\rho b}{a\delta}$	$0.435\dfrac{a_0\rho b}{a\delta}$	$0\ 300\dfrac{\rho}{\delta}$
Equivalent shunt R	$6\ 80\dfrac{a_0c}{b\delta}$	$6\ 07\dfrac{a_0c}{b\delta}$	$2.72\dfrac{ac}{\delta}$
$\dfrac{M}{L}$	$\dfrac{S}{a\,a_0}$	$1.272\dfrac{S}{aa_0}$	$2.08\dfrac{S}{a^2}$
Coupling impedance $= (M/L)^2 Q\sqrt{L/C}$	$6\ 8\left(\dfrac{S}{aa_0}\right)^2\dfrac{a_0c}{b\delta}$	$9\ 87\left(\dfrac{S}{aa_0}\right)^2\dfrac{a_0c}{b\delta}$	$11\ 8\dfrac{S^2c}{a^3\delta}$

(*Hansen, courtesy of J Applied Phys*)

a = radius of cylinder or sphere or one-half the side of the square, in meters
a_0 = one-half the height of cylinder or square, in meters.
ρ = resistivity of conducting material, ohms per cubic meter.
δ = depth of penetration of the waves into the metal enclosure, in meters.
$\delta = \sqrt{\rho/\pi\omega}$
$\omega = 2\pi f = 2\pi$ times the frequency of the waves
c = velocity of light, in meters per second
S = area of surface of probe loop in square meters See Fig 10·14.
$b = 1 + a/2a_0$
L, C, R in henrys, farads, and ohms respectively
M = equivalent mutual inductance in henrys

shape for cavity resonators, although Hansen has given some useful qualitative information in this connection. Let us deform a long circular cylinder in various manners and analyze the effects of each deformation. Any deformation toward a non-circular cylinder will make the current distribution on the inside surface non-uniform. Since the losses are proportional to the square of the current, the losses will thereby be increased. On the other hand, the field inside would not be greatly disturbed and hence the energy stored in the cavity would remain approximately the same. Thus the Q would be decreased. Now, if we deform the cylinder by converting it into a surface of revolution like an hour glass, there will again be a concentration of current in the constricted section which, it seems, would decrease the Q. Again, if one of the closed ends were indented in such a way as to form a concentric

closed tube, the Q would be decreased. This follows from the fact that it would be necessary to reduce the diameter in order to keep the wavelength constant, and also the length of the section could not exceed $\lambda/4$. Thus the volume-to-surface ratio would be reduced, reducing the selectivity. Now, as pointed out above, the sphere has a higher Q than that of a cylinder of approximately equal size. Hence, Hansen argues, it appears to be a good idea to round the ends of the cylinder. Quantitative arguments similar to those above substantiate this suggestion. Thus it appears that a long prolate spheroid is a most efficient possible shape. Hansen finds this type of cavity to have a selectivity 1.2 times better than that of a cylinder of the same length and resonant frequency. Whether the spheroid is the most desirable shape is extremely difficult to prove. If, however, the material and wavelength are fixed, it appears that no cavity resonator can operate with a Q greater than λ/δ at its lowest frequency. At higher modes of oscillation it may be shown that arbitrarily high values of Q may be obtained.



(Hansen, courtesy of J. Applied Phys.)

FIG. 10·15 Values of Q for various shapes of resonators are plotted against λ. The resonator material is copper, $\sigma = 5.7 \times 10^7$ mhos per meter.

In Fig. 10·15, values of Q are plotted as functions of λ for copper cavities. For prismatic and cylindrical cavities it has been assumed that $a_0 \gg a$. Where this condition is not met the value of Q indicated should be divided by the factor $b = \left(1 + \frac{a}{2a_0}\right)$.

PROBLEMS

10·1 A rectangular cavity resonator when excited in the H_{0m} mode is to have only one resonant frequency of 2000 megacycles. What dimensions must the cavity have?

10·2 A rectangular cavity resonator when excited in the H_{0m} mode is to have two possible resonant frequencies of 2000 and 3000 megacycles. What dimensions must the cavity have? (Two possible answers.)

10·3 A rectangular cavity resonator when excited in the H_{0m} mode is to have three possible resonant frequencies of 2000, 2500, and 3000 megacycles. What dimensions must the cavity have?

10·4 A rectangular chamber is 3 by 4 by 5 cm. Calculate at least eight resonant frequencies. Identify the mode of oscillation in each case.

10·5 A rectangular chamber is 3 by 3 by 14 cm. It is divided into two identical chambers 3 by 3 by 7 cm by a thin metallic diaphragm pierced by a small hole. What sort of performance is to be expected of this system?

10·6 A hollow ball, resembling a tennis ball, is made of copper and is placed within a one-turn coil which carries a heavy direct current. A magnetic field within as well as without the ball results from this current. If the electric current is very quickly interrupted what behavior within the ball is to be expected?

10·7 If the chamber of problem 10·5 is made of copper, how high a Q is to be expected of each of the two separate resonators so formed? (Consider the principal mode only.)

10·8 The ball of problem 10·6 has an inside diameter of 5 cm. What is its Q at the principal mode of oscillation? How long will an initial oscillation persist before H decreases to $1/e$ of its original magnitude?

10·9 Two identical resonators are made of copper and are 2 by 2 by 3 cm. They are adjacent to each other and are coupled by means of a closed conducting link. Using the ordinary definitions of low-frequency coupled-circuit theory, calculate the area of the loop required in each resonator in order to produce critical coupling.

10·10 A cylindrical cavity resonator is 10 cm in diameter and 20 cm long. Compare its principal resonant frequency with that which results if it is converted into a coaxial resonator by the addition of a central conductor.

10·11 If the material is of copper, compare the values of Q obtained in the two cases of problem 10·10.

10·12 The resonant cavity of problem 10·4 is oriented as shown in Fig. 10·1. If H_{0m} waves of the resonant frequency exist within the cavity, determine the point at which the greatest electrical intensity occurs.

10·13 If the greatest field intensity in problem 10·12 is 10 volts per centimeter, calculate the total stored electrical energy within the cavity.

10·14 A cylindrical cavity resonator in its principal mode is to have an equivalent inductance of 0.1 henry and a capacitance of 0.01 $\mu\mu$f. Determine the dimensions.

10·15 Calculate a square cavity resonator equivalent to the cylindrical resonator of problem 10·14.

10·16 Calculate the selectivity Q and equivalent shunt resistance of the resonator of problem 10·14.

10·17 Calculate the selectivity Q and equivalent series resistance of the resonator of problem 10·15.

10·18 Consider the possibility of using a cavity resonator with a thin elastic end as a microphone for the production of frequency-modulated waves

10·19 A cavity in the form of an annular torus is to serve as a resonator. Sketch the electric and magnetic lines of one possible mode of oscillation.

10·20 Obtain from the general equation 6·47, for the H_{nm} waves in a rectangular wave guide, the resonant H_{nml} waves for a rectangular cavity, in a manner similar to that employed in section 10·2.

10·21 Repeat problem 10·20 for the general E_{nml} mode. The general equation for E_{nm} waves in rectangular wave guides are given in equations 6·82.

10·22 Obtain, from the general equations for TM_{nm} waves in a cylindrical wave guide, the resonant TM_{nml} waves for a cylindrical cavity.

10·23 Repeat problem 10·22 for TE_{nml} waves in a cylindrical cavity.

10·24 Determine the selectivity Q for the general TE_{nml} waves in a rectangular cavity.

10·25 Obtain an expression for the selectivity of a rectangular cavity, oscillating in the TM_{nml} mode.

10·26 Repeat problem 10·25 for the TE_{nml} mode in a cylindrical cavity.

10·27 Repeat problem 10·25 for the TM_{nml} mode in a cylindrical cavity.

CHAPTER 11

RADIATION FROM HORNS AND REFLECTORS

11·1 Introduction

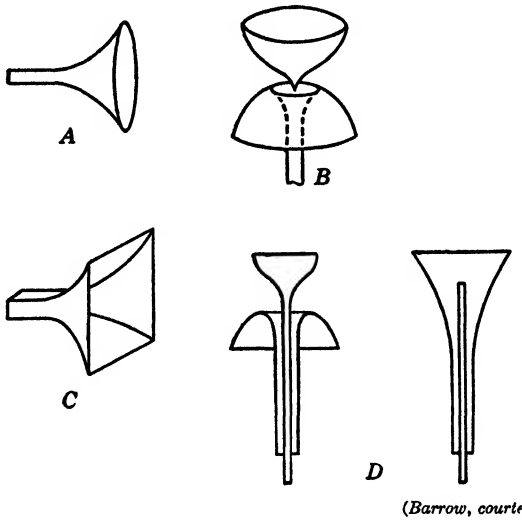
The hollow conducting wave guide has been found a convenient means of transmitting energy at frequencies above about 1000 megacycles. Closely allied to the wave guide is the electromagnetic horn, which serves as a sort of antenna to produce electromagnetic radiation in free space. At the present time, the horn finds its greatest application in the production of relatively narrow beams of high intensity; however, wide patterns for broadcast purposes may also be obtained.

Relatively narrow beams may be produced by means of metallic reflectors associated with a single antenna element of conventional form, or by the use of antenna arrays. The reflector is usually of parabolic form, either cylindrical or of the usual rotated form familiar in searchlamps. Antenna arrays are composed of a relatively large number of elementary antennas, 10 to 20 being a typical figure, arranged in a definite geometrical fashion and excited in suitable time phase.

At frequencies between about 100 and 1000 megacycles the array is of convenient mechanical size and is readily adjusted, whereas the horn or reflector becomes inconveniently large and expensive. At the higher frequencies, however, the situation may be reversed. The array is more difficult to manage and to adjust, while the horn reflector is of most convenient proportions.

In the hollow conducting wave guides, discussed in Chapters 6 and 7, electromagnetic radiation is emitted from an open end in much the same way as sound waves from a hollow tube. Barrow and Southworth suggested the possibility of flaring the ends of the tube into a horn-shaped radiator. It is found that such horns increase the directivity of the radiation and provide a good termination for the tube. In this respect the analogy with the acoustical horn is quite close. The wavelengths employed in acoustics, however, are long in comparison with the throat diameter of the horn, whereas the throat of the electromagnetic horn is of the same order of magnitude as the wavelength. The conducting horn and guide to which it is connected are, in general, many wavelengths long. The closed end is generally adjustable for tuning purposes. A piston arrangement such as that described in Chapter 8

is ordinarily used. Electromagnetic waves are excited by some form of antenna near the closed end. These waves are propagated along the tube and horn surfaces, emerging from the horn mouth to be propagated through space as free waves. Several types of electromagnetic horns first suggested by Barrow are shown in Fig. 11·1. *A* and *B* are designed



(Barrow, courtesy of IRE)

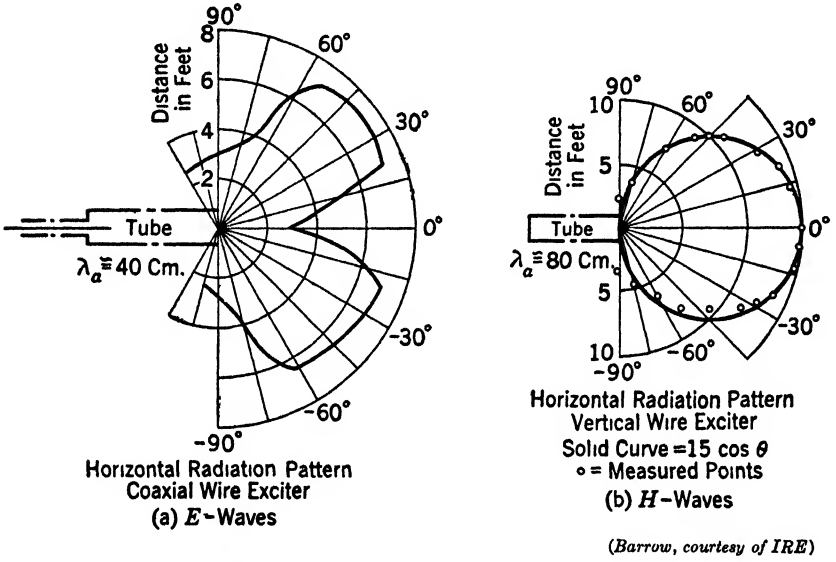
FIG. 11·1 Electromagnetic horns. *A*, *B*, and *C* — hollow tube feed; *D* — coaxial tube feed.

for connection to circular pipes, *C* for rectangular pipes, and *D* for a coaxial line. Both *A* and *C* are suitable for beam radiation of transverse electric (H_1) waves, while *B* and *D* are adaptable to broadcast radiation with longitudinal electric (E_0) waves. It may be pointed out that radiators of this type are in no sense restricted to hollow-pipe systems, for any of these horns may be fed directly by coaxial lines or by other means.

11·2 Radiation from Tube End

Radiation patterns from the open end of a circular wave guide are shown in Fig. 11·2. The electromagnetic energy propagated down the tube flows from the tube end and is distributed in the pattern shown. Because of reflections from the end of the tube, not all the energy is radiated, and standing waves are set up in the pipe. These reflections result from the mismatch between the impedance of the tube and that of free space. The shape of the pattern is a function of λ_a/a, where λ_a is the wavelength of the guided wave in free space and a is the radius of

the tube. The free-space patterns are symmetrical about the axis of the tube. Figure 11·2*a* shows the radiation pattern from the open end of a pipe in which E_0 waves have been set up. When H_1 waves are employed, the radiation pattern has a maximum value in front of the tube opening. The H_1 wave thus appears to be best suited for producing a beam type of pattern. The directivity of the H_1 wave is improved as the ratio of guide diameter to wavelength is increased.



(*Barrow, courtesy of IRE*)

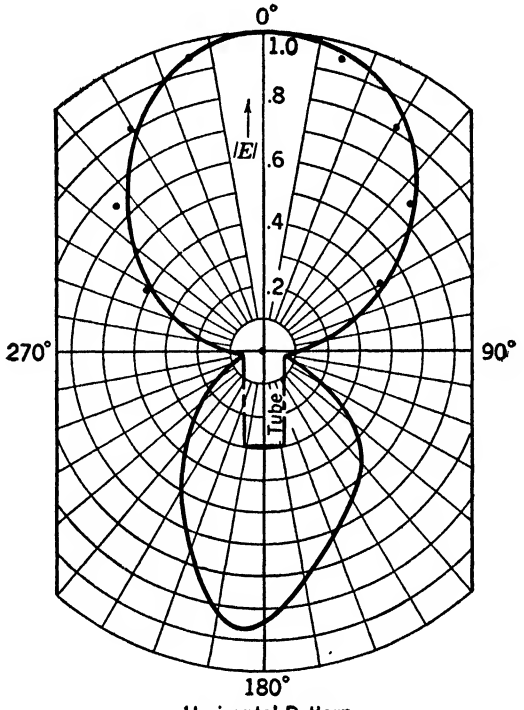
FIG. 11 2 Radiation patterns for the open end of circular pipes.

Radiation patterns from the open end of a rectangular tube are shown in Fig. 11·3 for an H_{01} wave. Although the maximum radiation is produced along the forward axis of the tube, it is seen that a considerable backward radiation also exists, which may be attributed to diffraction around the mouth of the pipe. When the frequency of operation is near the critical frequency of the pipe, the backward and forward radiations are approximately equal. As the frequency is increased, the backward radiation is appreciably reduced, and the forward radiation pattern considerably sharpened. Backward radiation may be thought of as a flow of current out of the tube mouth and back along the outer surface.

11·3 Rectangular Horns

A theoretical analysis of the operation of the electromagnetic horn "antenna" has been developed by Barrow and Chu from Maxwell's

equations. This analysis leads directly to design formulas which have been confirmed experimentally by Barrow and Lewis.† The type of horn used in this work is shown in Fig. 11·4. By appropriate choice



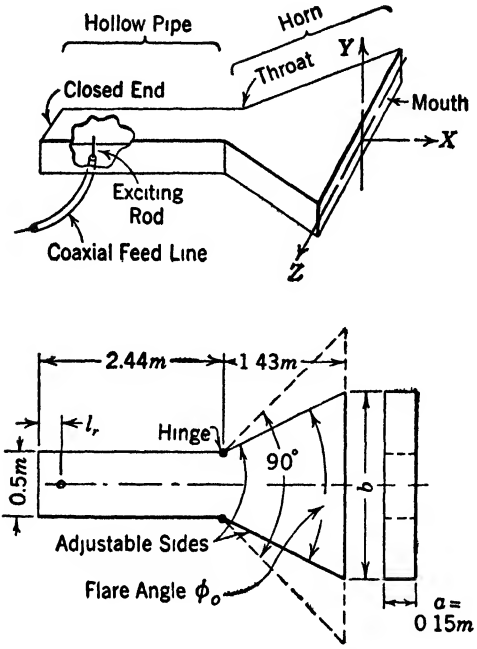
(*Barrow, courtesy of IRE*)

FIG. 11·3 Radiation pattern for H_{01} wave from the open end of a rectangular hollow-pipe radiator.

of the flare angle, ϕ, the impedance of the pipe is very nearly matched to that of free space and hence standing waves are largely suppressed. This results in a great increase in the radiated energy. The approximate distribution of the field in a typical horn is shown in Figs. 11·5 and 11·6 for an H_{01} wave of the two possible orientations.

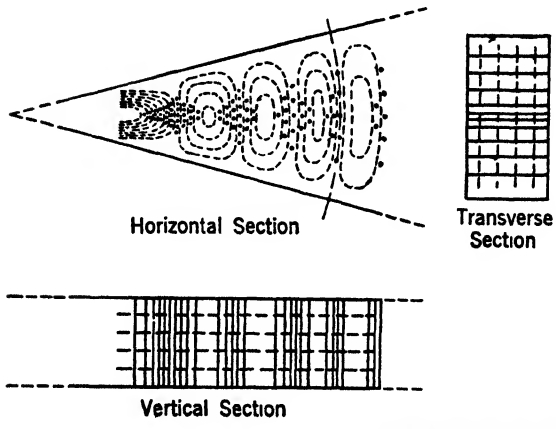
* W. L. Barrow and L. J. Chu, "Theory of the Electromagnetic Horn," *Proc. IRE*, **27**, 51, 1939.

† W. L. Barrow and F. D. Lewis, "The Sectoral Electromagnetic Horn," *Proc. IRE*, **27**, 41, 1939. See also Chu and Barrow, *Trans. AIEE*, **58**, 333, 1939.



(Barrow and Lewis, courtesy of IRE)

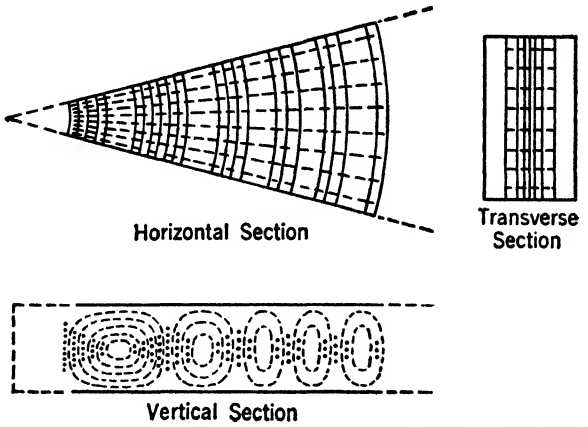
FIG. 11·4 Experimental horn used by Barrow and Lewis.



(Barrow and Lewis, courtesy of IRE)

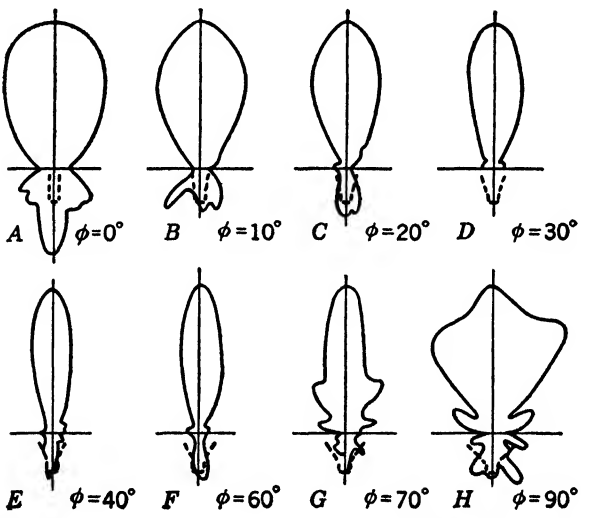
FIG. 11·5 Approximate field distribution for an H_{01} wave in the horn of Fig. 11·4 when the magnetic intensity is parallel to the base of the horn.

As the flare angle of the horn is changed, the radiation patterns that result take the various forms shown in Fig. 11·7. At smaller flare angles the radiant energy is in a rather broad beam along the axis of the tube,



(Barrow and Lewis, courtesy of IRE)

FIG. 11·6 Approximate field distribution for an H_{01} wave in the horn of Fig. 11·4 when the electric intensity is parallel to the base of the horn.



(Barrow and Lewis, courtesy of IRE)

FIG. 11·7 Radiation pattern for a rectangular horn having various flare angles.

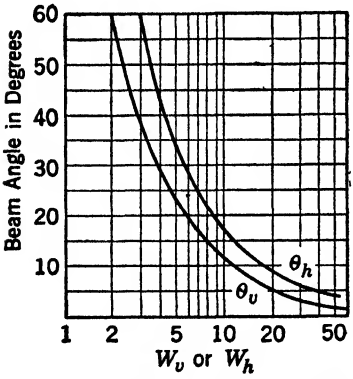
and a marked back-radiation pattern exists. The back radiation becomes smaller for larger values of ϕ and increases again as ϕ approaches 90°. In the interval between 40° and 60°, the radiation pattern is quite

sharp, increasing in breadth for larger values of ϕ. It is seen from these patterns that the electromagnetic horn can produce beams substantially free from secondary lobes in one plane.

It should be noted that the horn under discussion is flared in only one direction. Such a horn gives much the same radiation pattern regardless of whether the electric vector of the wave system is along, or at right angles to, the long dimension of the horn aperture. Neither distribution, however, is symmetrical. The beam is wide in the direction in which the aperture is narrow, and vice versa. This behavior is to be expected since the end of a simple square guide of small dimensions gives a beam that is wide in both directions.

Such an asymmetrical form is potentially useful since it is often desirable to establish a beam that is narrow in one direction and wide in the other.

Unless the flare angle is large the equiphase surface at the mouth of the horn is essentially a plane. Under this condition the radiation is exactly the same as that which would result from a rectangular wave guide having a cross section equal to the aperture of the horn. This is an excellent approximation for flare angles up to 50°, and the error introduced for 60° is not usually serious.



(*Barrow and Lewis, courtesy of IRE*)

FIG. 11·8 Calculated curves giving the angle between two points of zero intensity defining the principal beam.

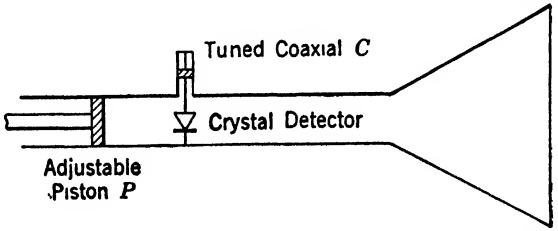
If a beam of the searchlight type, narrow in both directions, is to be produced the horn must be flared in both directions so as to achieve a nearly square aperture. The data of Fig. 11·7 are not exactly applicable here since the patterns are altered somewhat by the polarization of the waves in the guide. In practice, however, the difference is not important, and a square horn of approximately 50° flare is very satisfactory. The length of the horn may be determined from the required beam width by means of Fig. 11·8.

Figure 11·8 is a calculated curve giving the angle between the two points of zero intensity which define the principal beam. For many purposes the beam is effectively much narrower than this value. The actual calculation is based on a long rectangular guide whose aperture is expressed in terms of the ratios $W_v = a/\lambda_a$ and $W_h = b/\lambda_a$, where a is the vertical aperture, b is the horizontal aperture, and λ_a is the wave-

length of the wave in free space. Two curves are given, the one designated θ_h applying when the electric field is parallel to the flared sides as in Fig. 11·5. θ_v applies when the electric field joins the flared sides as in Fig. 11·6. For a wavelength of 10 cm and a beam angle of 5° an aperture about 2.5 meters square is indicated.

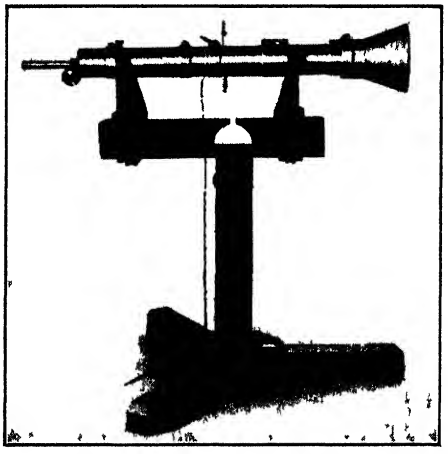
11·4 Circular Horns

A very complete experimental study of the properties of electromagnetic horns of circular cross section has been made by Southworth and King.* Their work was confined to the H_1 wave associated with a



(Southworth and King, courtesy of IRE)

FIG. 11·9 Experimental circular horn used by Southworth and King

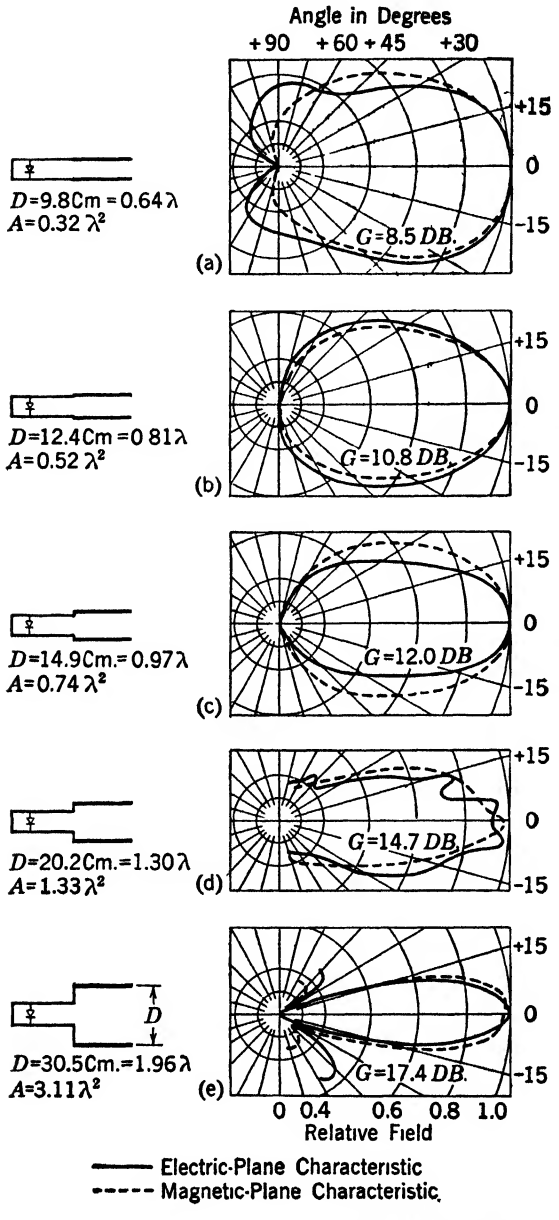


(Southworth and King, courtesy of IRE)

FIG. 11·10 Photograph of the electromagnetic horn used by Southworth and King.

circular guide and to wavelengths between 10 and 20 cm. The study was so conducted that the effects of flare angle, horn length, aperture, and wavelength are readily identified.

* G. C. Southworth and A. P. King, "Metal Horns as Directive Receivers of Ultra-Short Waves," *Proc. IRE*, **27**, 95, February, 1939.



(Southworth and King, courtesy of IRE)

FIG. 11·11 Radiation patterns for a circular wave guide when cylindrical sections are added as indicated by the heavy lines.

The properties of any antenna may be determined equally well by testing it as a transmitter or as a receiver. In their research Southworth and King found it expedient to use the various horns as receivers. A fixed directional transmitter similar to the receiver was used, and the response of the receiving horn as it was turned to various angles about vertical and horizontal axes was determined. The arrangement is shown schematically in Fig. 11·9 and by photograph in Fig. 11·10. By appropriate adjustment of the piston and the coaxial tuner it is possible to match the impedance of the crystal detector and associated meter to that of the wave guide, thus assuring that all the incident energy is absorbed and that no spurious results are obtained.

Results of one series of tests are plotted in Fig. 11·11. The term horn is scarcely applicable here since the unit actually consists of two cylindrical sections joined together. Very considerable directive effects are achieved, and considerable power gain is indicated. However, the irregularity of the pattern, particularly the marked spurious lobes, renders this form undesirable. In these and the following figures the solid curves represent measurements made in the plane containing the electric vector; the dotted curves represent measurements made in the plane containing the magnetic vector. Since the two curves are similar the complete directional pattern may be thought of as an ellipsoid of which these two curves are perpendicular sections.

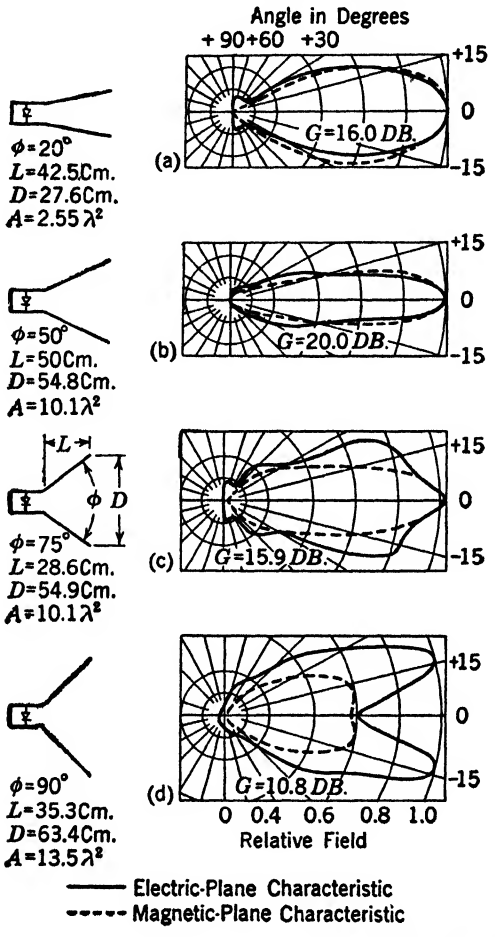
The power gain associated with each horn is most simply expressed in terms of the horn as a radiator. The fundamental although impractical reference of comparison is a point source which radiates equally in all directions. For example, the radiation intensity along the axis of horn of Fig. 11·11*c* is 16 times (12 db)* as high on a power basis as it would be if the given power were radiated equally in all directions. The simple dipole or half-wave antenna gives a gain of 2.15 db on this basis. Similarly the elementary dipole of vanishing length, sometimes used as a standard of reference in radio work, gives a gain of 1.76 db.

Results of a second series of tests are plotted in Fig. 11·12. Here the length of the horn was held essentially constant, the variable being the angle of flare. A maximum power gain of 100 (20 db) is secured with an angle of 50°, and a remarkably clean radiation pattern is secured. Larger angles of flare seem to offer very little advantage, but smaller ones may be useful where beams of appreciable width are required.

A third series of tests is presented in Fig. 11·13. Here a fixed angle of 40° is used and the horn length is the variable. Increase of the horn length shows a steady improvement in the gain and directivity. The material required for the horn also increases very rapidly, and the

* See page 455 for definition of the decibel.

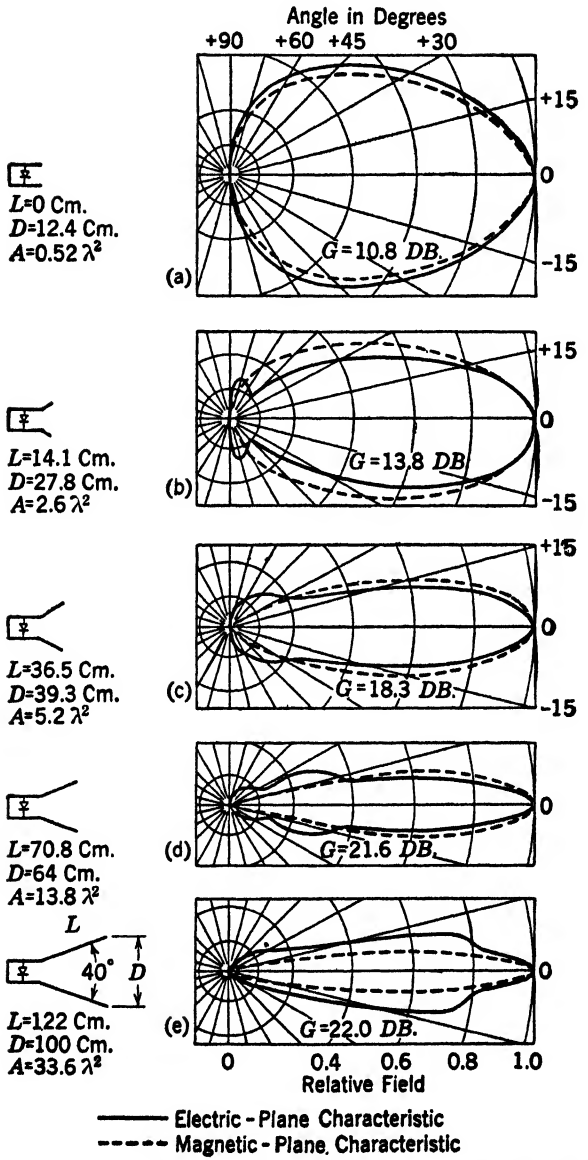
regularity of the figure begins to suffer as the horn is made very long, so that some practical limit exists in this case. The small improvement from *d* to *e* in Fig. 11·13 is ascribed to the fact that the optimum flare angle is different for long and for short horns.



(*Southworth and King, courtesy of IRE*)

FIG. 11·12 Radiation patterns for circular wave guides when circular horns having various flare angles are added. The length of the various horns is held fixed.

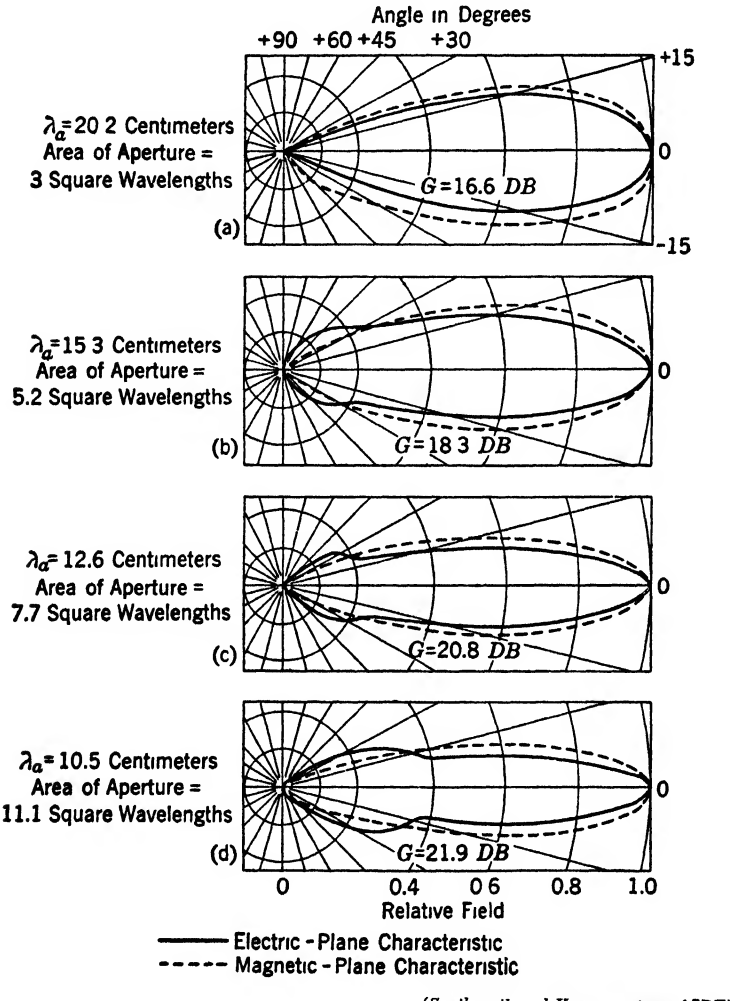
An important property of antennas in general and of horns in particular is that the directional pattern and power gain are unaffected if the wavelength and all dimensions of the horn are modified by some given factor. This is true since all dimensions, when expressed in terms of the wavelength, are constant.



(Southworth and King, courtesy of IRE)

FIG. 11·13 Radiation patterns for circular wave guides when circular horns having a fixed flare angle of 40 degrees are used. The length of the horn is varied.

Important experimental results are presented in Fig. 11·14. A given horn was tested at four different frequencies involving virtually a



(*Southworth and King, courtesy of IRE*)

FIG. 11 14 Radiation patterns for a given circular horn as a function of the wavelength.

2 : 1 frequency range and an actual frequency band that is 1400 megacycles wide. The gain and directivity are steadily improved as the frequency is raised, but the change is regular and not particularly rapid.

11·5 Summary on Single Horns

Square or conical metallic horns give strictly comparable results, the choice depending primarily on practical construction considerations. The optimum angle is approximately 50°, although somewhat smaller angles are better if very large horns are to be built. The adjustment is not critical, in marked contrast to high-frequency antenna array systems.

The beam produced may be made as narrow as necessary, and it is very free from spurious lobes. This is believed to be the most outstand-

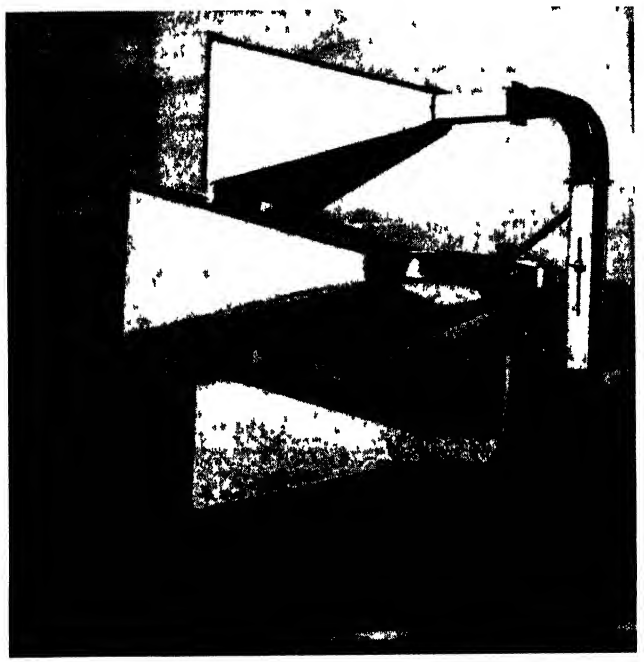


FIG. 11·15 Array of electromagnetic horns used by Barrow and Shulman.

ing of all the features of the horn. The fact that the horn works almost equally well over a wide band of frequencies is again markedly in its favor.

Since horns are seldom of great length the attenuation constant is unimportant and materials of relatively low conductivity may be used. Galvanized iron has been widely used for experimental units.

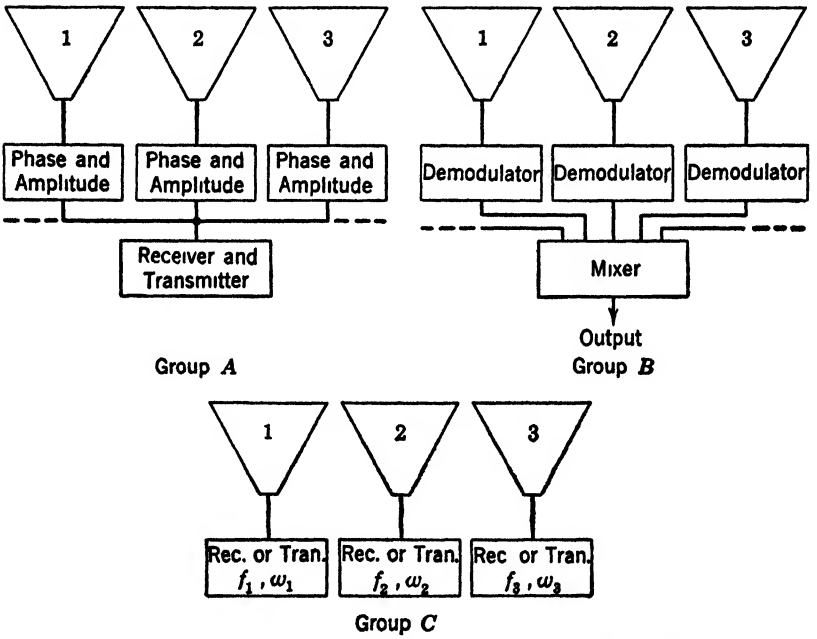
Experimental data shows that the efficiency and directivity of a horn are improved if the edges are somewhat curved rather than straight. The exponential curve, widely used for the flaring of acoustic horns, does not appear to be the best solution, but some curve which reduces

the discontinuity at the throat is desirable. The construction of such horns, particularly in circular sections, is relatively difficult, however, and the advantage to be gained by the curvature may hardly be worth the added expense.

11·6 Multiple Horns

Arrays of electromagnetic horns have been studied by Barrow and Shulman.* A photograph of one of their arrays is shown in Fig. 11·15. A classification of arrays in terms of application is given in Fig. 11·16.

The outstanding conclusion derived from the theoretical work and



(Barrow and Shulman, courtesy of IRE)

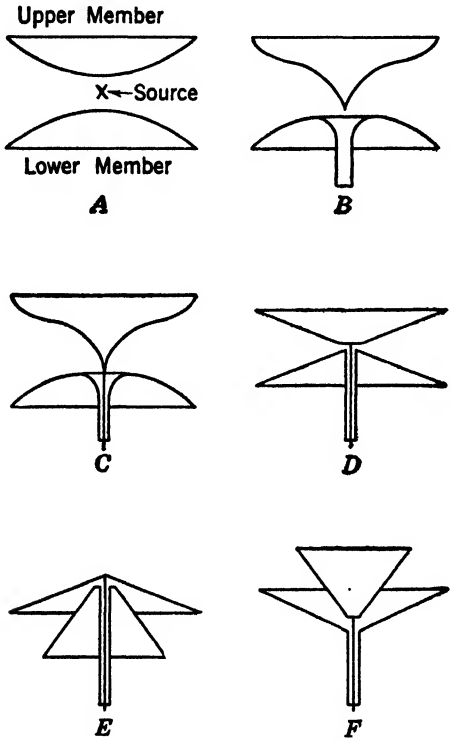
FIG. 11 16 Electromagnetic horn arrays studied by Barrow and Shulman.

verified by experiment is that adjacent horns do not interact with each other to any appreciable extent. Only when they are so rotated as actually to point toward each other do they show appreciable coupling. This is significant in that the coupling between the horns and the driving unit may be designed purely on the basis of the characteristic impedance of each. The overall results are then given by a linear superposi-

* W. L. Barrow and Carl Shulman, "Multiunit Electromagnetic Horns," *Proc. IRE*, **28**, 130, 1940.

tion of the individual effects. This is true even though the outer edges of the horns actually touch.

If four identical square horns are arranged so that their apertures form a larger single square, the directional characteristic and power gain are those of a single longer horn having that value of aperture. The bulk and weight of the four horns are, however, much less than those corre-



(Barrow, Chu, and Jansen, courtesy of IRE)

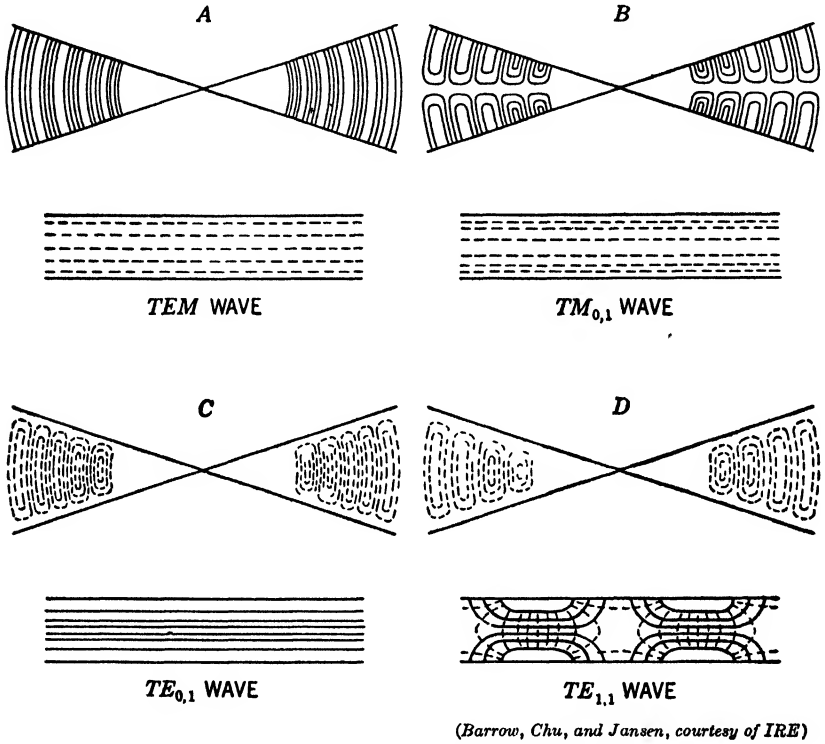
FIG. 11·17 Biconical horns used by Barrow, Chu, and Jansen.

sponding to the single large horn. Arrays in which the edges of adjacent horns do not coincide are subject to marked spurious lobes but show extremely sharp principal characteristics.

The directional characteristic of a horn array may be controlled or steered over a considerable angle by proper adjustment of the phase relation between the individual horns of the array and the common circuit. Techniques of phase control which have been worked out for antenna arrays at lower frequencies have proved very useful.

The independence of adjacent horns is of great importance in that a

transmitter and receiver may be located together and may operate on the same frequency band without excessive interference. Other applications of this principle will undoubtedly be developed.



(Barrow, Chu, and Jansen, courtesy of IRE)

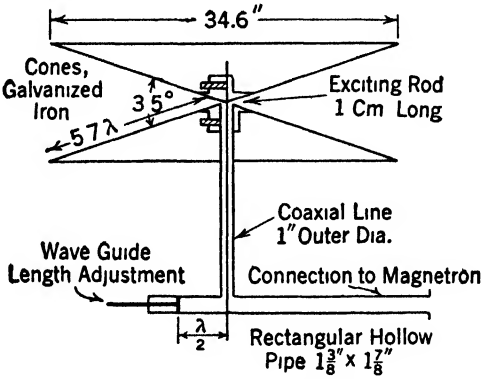
FIG. 11·18 Typical field configurations within the biconical horn shown in Fig. 11·17*D*.

11·7 The Biconical Horn

The biconical horn is an electromagnetic radiating system which concentrates all its power in or near one plane but which radiates equally in all directions in that plane. It is ideally suited for broadcasting since it loses very little energy to the ground or in a useless sky wave. Several basic forms of this horn are shown in Fig. 11·17.

Several types of field configurations which may be obtained in biconical horns are shown in Fig. 11·18. The simplest field configuration is a portion of a spherical transverse-electromagnetic wave. This is shown in Fig. 11·18*a*. The spherical configuration is probably of the greatest practical importance since the field which results at a distance from the antenna is simply a vertically polarized transverse electromagnetic wave.

The schematic construction of such a horn is illustrated in Fig. 11·19. The corresponding physical apparatus is shown in the photograph of



(Barrow, Chu, and Jansen, courtesy of IRE)

FIG. 11 19 Constructional data for the biconical horn shown in Fig 11·17*D*

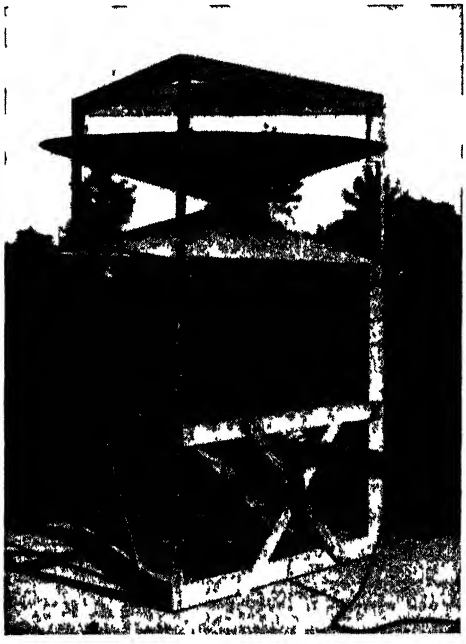
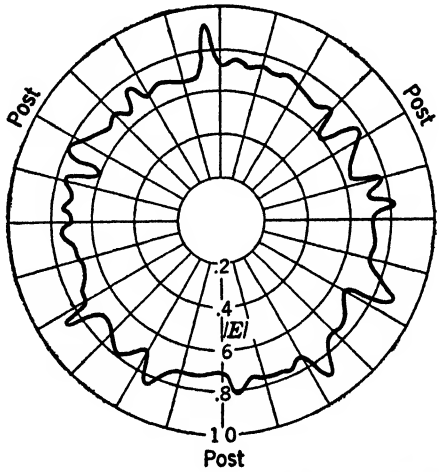


FIG. 11·20 Biconical horn used by Barrow, Chu, and Jansen.

Fig. 11·20. The radiation pattern at a considerable distance from this structure is illustrated in Figs. 11·21 and 11·22. The vertical directivity

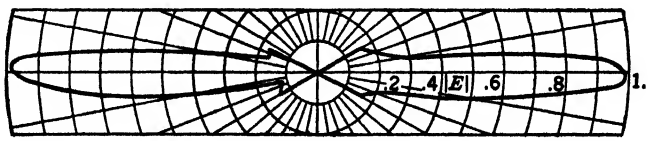
as indicated in Fig. 11·22 is excellent, nearly all the energy being confined to the horizontal plane. The uniformity in this horizontal plane is good, observed ripples being traceable to the fact that the upper horn



(Barrow, Chu, and Jansen, courtesy of IRE)

FIG. 11·21 Radiation pattern in the horizontal direction for the biconical horn of Fig. 11·20.

element is supported on three wooden posts. An extremely uniform characteristic may be achieved by replacing these three supports by a uniform cylindrical structure of dielectric.



(Barrow, Chu, and Jansen, courtesy of IRE)

FIG. 11·22 Radiation pattern in the vertical direction for the biconical horn of Fig. 11·20.

11·8 The Parabolic Reflector

If a half-wave antenna is placed at the focal point of a parabolic metal sheet which is large in comparison with the length of the antenna, the waves radiated from the antenna are reflected from the parabola according to the simple laws of geometrical optics. That is, the reflected radiation has the characteristics of a parallel beam similar to that familiar in light. Since a straight-line antenna can never have zero length it is

evidently impossible for all the radiation to come from exactly the focal point of the parabola. Therefore, a perfect beam may not be obtained in this way, and it is necessary that the reflector be many times larger than the antenna if the beam is to be at all narrow.

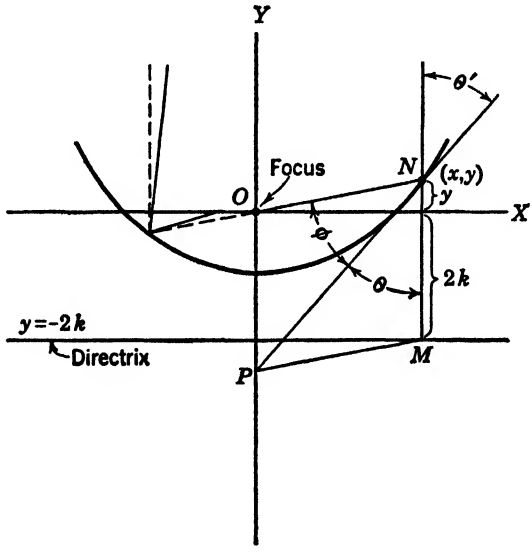


FIG. 11·23 Parabolic reflector.

Let us review the properties of the parabola to see how it is able to produce a parallel beam by reflection from a point source. The parabola is defined as the locus of points equidistant from a point called the focus and a straight line called the directrix. Let the focus be the origin, O, and the directrix the line $y = -2k$. Expressing the geometrical definition of the parabola in algebraic form we write for Fig. 11·23

$$\sqrt{x^2 + y^2} = y + 2k \qquad [11\cdot1]$$

or squaring

$$x^2 = 4ky + 4k^2 \qquad [11\cdot2]$$

which gives

$$y = \frac{x^2 - 4k^2}{4k} \qquad [11\cdot3]$$

That is, the parabola is the locus of all points which satisfy equation 11·3. The slope of the straight line which is tangent to the parabola at any point N, having coordinates (x,y), is found by taking the deriva-

tive of 11·3 with respect to x. Thus

$$\frac{dy}{dx} = \frac{x}{2k} \qquad [11\cdot4]$$

We assume that radiation emanates from the focal point, O, and is reflected from the metal surface according to the familiar law of equal angles. The beam reflected from every point on the parabola will be parallel to the Y axis provided that

$$\theta = \phi = \theta' \qquad [11\cdot5]$$

that is, if the line ON and the vertical line MN make equal angles with the tangent. The lengths ON and MN are equal by the defining equation 11·1. We shall now show that the tangent to the curve at point N intersects the Y axis at a point P such that

$$OP = ON = NM = MP \qquad [11\ 6]$$

The slope of the tangent line from equation 11·4 is

$$\frac{dy}{dx} = \frac{x}{2k} \qquad [11\cdot4]$$

The horizontal distance between N and the Y axis is equal to x, and the corresponding vertical displacement between N and P is

$$\frac{dy}{dx}x = \frac{x^2}{2k} \qquad [11\cdot7]$$

At point N

$$y = \frac{x^2 - 4k^2}{4k} \qquad [11\cdot3]$$

Accordingly, the length of the line OP is

$$\frac{x^2}{2k} - \frac{x^2 - 4k^2}{4k} = \frac{x^2 + 4k^2}{4k} \qquad [11\cdot8]$$

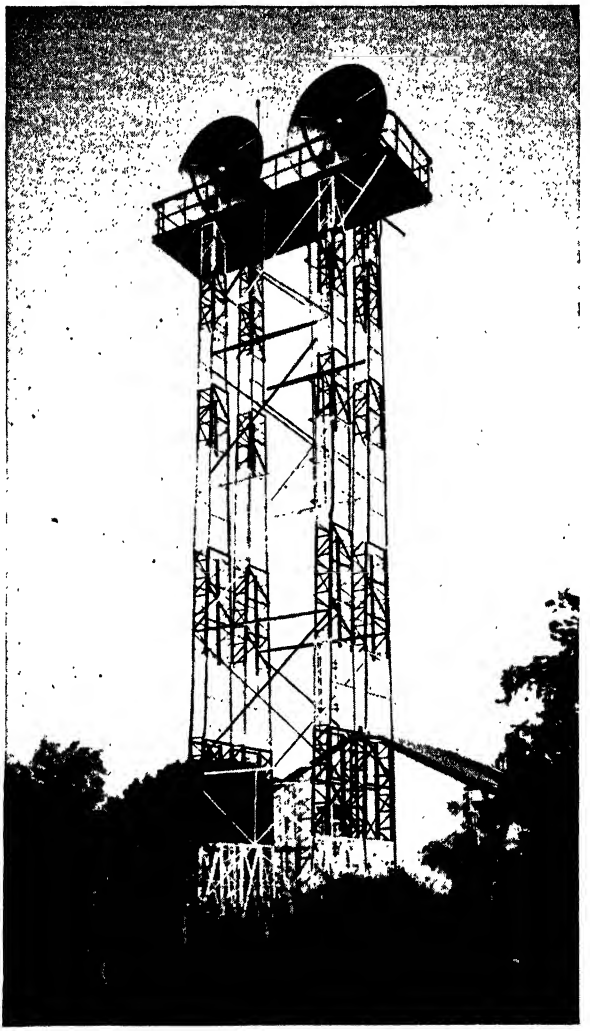
The length of the line MN is given by the expression

$$2k + y = 2k + \frac{x^2 - 4k^2}{4k} = \frac{x^2 + 4k^2}{4k} \qquad [11\ 9]$$

It is seen from equations 11·8 and 11·9 that lengths MN and OP are numerically equal. Since the figure $MNOP$ is a parallelogram and since $MN = NO = OP$ it is also necessary that $MP = MN$.

Thus we have shown that the tangent to the parabola at any point is the angle bisector of the equilateral trapezoid and, therefore, makes equal angles with the incident and vertical rays. From this fact we

conclude that all radiation which leaves the origin and meets the parabolic reflecting surface is reflected parallel to the Y axis.



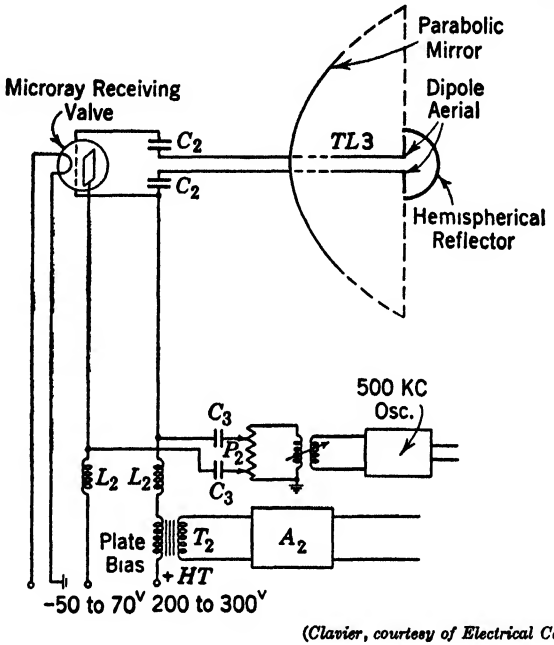
(Clavier, courtesy of Electrical Communication)

FIG. 11·24 French terminal of the hyper-frequency link across the English Channel.

Let us now suppose that an antenna lies along the X axis and extends a short distance to each side of the origin. Radiation from the ends of such an antenna does not strike the parabolic reflector in such a direction as to be reflected in the Y direction. Part of the reflected radiation

crosses the Y axis, suggesting a converging beam, but the general effect is one of divergence.

Fortunately most of the radiation from a dipole is near its midpoint and is propagated at right angles from the middle. Thus a dipole antenna produces a beam reflected from a paraboloid which is considerably sharper than would be expected from the above reasoning.



(Clavier, courtesy of Electrical Communication)

FIG. 11·25 Schematic arrangement of the apparatus used in the reception of hyper-frequency waves.

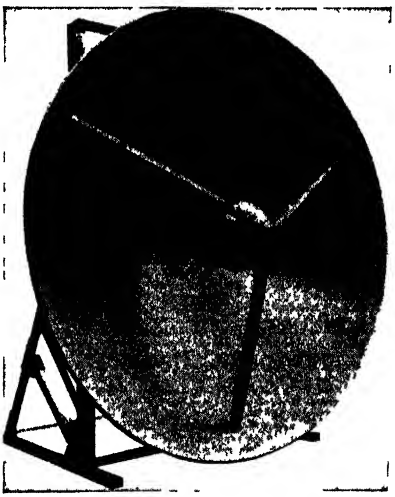
11·9 Application of Parabolic Reflector

One of the most interesting applications of the parabolic reflector is in connection with the hyper-frequency link across the English Channel from St. Inglevert in France to Lymphne, England.* This system was put into commercial use in 1934. The distance covered is 56 kilometers; the frequency, 1600 megacycles; the transmitted power, $\frac{1}{2}$ watt. The terminal sites were so chosen that a line-of-sight operation is assured and the transmitting and receiving parabolas were placed on high towers well out of the way of normal obstructions.

The French terminal is shown in Fig. 11·24. The schematic arrange-

* A. G. Clavier, *Electrical Communication*, **12**, 3, 1933; A. G. Clavier, and L. C. Gallent, *Electrical Communication*, **12**, 222, 1934.

ment is shown in Fig. 11·25, and the actual unit is shown in Fig. 11·26. The reflector is approximately 8 feet in diameter, very large in comparison to the antenna, which is only a few inches long. The parabola gives a power gain of approximately 28 db above that of the dipole antenna alone. This figure must be raised to 30 db for comparison with figures given for horns. The hemispherical reflector with center at the antenna was added to return to the parabola radiation otherwise



(Clavur, courtesy of Electrical Communication)

FIG. 11·26 A parabolic reflector used for the creation of a hyper-frequency beam.

lost from the front of the antenna. The returned radiation passes by the antenna wire and strikes the parabola as if it came directly from the wire. An additional 5 db was achieved in this way, bringing the total gain to 35 db.

It should be mentioned, however, that these results were obtained with large reflectors quite carefully adjusted. A horn of comparable size would give comparable results and would probably require much less precision in the final adjustment. The horn was not generally known at the time of this installation.

PROBLEMS

11·1 A rectangular horn with an aperture 30 by 50 cm is fitted to a wave guide 10 cm square. The tapered section is 40 cm long measured in the direction of the axis. Calculate the vertical and horizontal directivity of this horn for the H_{01} wave with electric field parallel to the long side of the aperture. The frequency is 3000 megacycles.

11·2 Repeat problem 11·1 with the electric field parallel to the short side of the aperture.

11·3 Repeat problem 11·1 if the aperture is increased to 60 by 100 cm and the tapered section is 80 cm long.

11·4 Repeat problem 11·3 using the electric field parallel to the short side of the aperture.

11·5 A circular horn has an aperture 60 cm in diameter. It is fitted to a circular wave guide 8 cm in diameter. The axial length of the tapered section is 80 cm. From the data of Figs. 11·13 and 11·14 estimate the gain of this horn over a non-directive radiator.

11·6 Sketch the expected directional characteristics of the horn of problem 11·5.

11·7 A certain biconical horn such as that of Fig. 11·19 radiates uniformly in all horizontal directions. The radiation is essentially uniform within $\pm 5°$ from the horizontal plane and is negligible elsewhere. Calculate the power gain.

11·8 A particular circular horn has a radiation characteristic which is essentially uniform within 5° of the axis and is negligible elsewhere. Calculate the power gain.

11·9 Discuss the statement that the directional pattern of a horn as a transmitter is the same as that of the same horn as a receiver. Why is it more convenient to test a horn as a receiver than as a transmitter?

11·10 Discuss the relative advantages of the horn and the parabolic antenna. Consider area of aperture, ease of adjustment, weight, etc.

11·11 A parabolic reflector is 1 meter in diameter and is excited by a half-wave dipole antenna 5 cm long. The center of the antenna is placed at the focal point of the parabola, which lies in the plane of the rim. Deduce the greatest possible width of the beam produced.

11·12 Estimate the directional characteristic of the system of problem 11·11. Assume that the antenna is vertical, and deduce the radiation pattern in both vertical and horizontal planes.

11·13 A rectangular horn radiator is flared in only one direction. It is excited by waves of the H_{02} mode. Sketch the wave distribution corresponding to Fig. 11·5.

11·14 Repeat problem 11·13 using the reversed orientation corresponding to Fig. 11·6.

11·15 Radiation is to be produced in the form of a thin sheet having the shape of the surface of a cone. Sketch a suitable horn arrangement.

11·16 Estimate the power gain of the horn of problem 11·1 at a frequency of 3000 megacycles.

11·17 Estimate the power gain of the horn of problem 11·3 at a frequency of 3000 megacycles.

11·18 Repeat problem 11·16 for a frequency of 5000 megacycles.

11·19 Repeat problem 11·17 for a frequency of 5000 megacycles.

CHAPTER 12

THE BEHAVIOR OF VACUUM TUBES AT HIGH FREQUENCIES

12·1 Introduction

The history of physical science makes it clear that significant progress results only from a combination of theoretical and experimental development. Throughout the middle ages the experimental approach was almost completely neglected and progress was slow. Since that time, experimental work has assumed an ever-increasing importance. Electronics is fairly typical of the physical sciences. The high-vacuum diode, a product of the experimental genius of Edison, found neither application nor explanation until about 1900 when J. J. Thomson stated the so-called electron theory. Shortly thereafter Fleming used the diode as a rectifier of radio signals and DeForest created the triode by the addition of the grid. The triode found immediate acceptance as an improved detector of radio signals. Application as an amplifier, oscillator, and control element followed in rapid succession.

Although many of the early workers had an excellent physical picture of the operation of the triode and although much valuable work was done, the scientific application of thermionic tubes in reliable circuits dates from the statement of the equivalent plate circuit theorem by van der Bijl* and others. Later this theorem was extended to multielement tubes, permitting accurate solution of a variety of problems. Innumerable engineering designs are based on results derived from this theorem. The importance of this concept, even at the present time, can hardly be overemphasized.

For some time after its introduction, the simple equivalent plate circuit theorem served to explain all the observed phenomena. As the technology of radio advanced, however, and higher and higher frequencies were used, it became apparent that this explanation of tube behavior was inadequate. The difficulty was found to lie in the interelectrode capacitances, which, although negligible at low frequencies, play an important part in the performance of vacuum tubes at frequencies in the order of a megacycle. It was felt that the analysis of the vacuum tube should be kept distinct from that of the associated net-

* H. L. van der Bijl, *Phys. Rev.*, 12, 180, 1918.

work, and modified equivalent plate circuit theorems were developed on this basis.* Unfortunately, the coefficients developed in this way are extremely complicated.

As more complex circuits were used and a better understanding of the action of vacuum tubes was gained it became apparent that this complex equivalent plate circuit theorem was impractical. The interelectrode capacitances within the tube are usually shunted by comparable capacitances due to the external circuit, which are not accounted for by the theorem. The mathematical treatment of the entire circuit is simplified by taking the sum of all the capacitances which are in parallel in each branch before setting up the equations. The procedure which is now most widely used is to lump the interelectrode capacitances with those of the corresponding external circuit and to express only the thermionic action by the equivalent plate circuit theorem. With typical tubes this method of analysis gives results which are correct for all frequencies below about 30 megacycles.

Two distinctly different effects manifest themselves as the frequency rises still further. When the time required for an electron to travel from cathode to anode becomes comparable to the period of a cycle the behavior of the tube is modified in a fundamental way and the entire performance must be re-examined from basic considerations. At these frequencies the effective input and output impedances are also greatly affected by the self and mutual inductance of the leads which join the tube to the external circuit. The following pages will be devoted to a consideration of these effects.

12·2 Review of Equivalent Plate Circuit Theorem

The equivalent plate circuit theorem results from the separation of total voltages and currents into components. For example, the total plate current is divided into a constant or direct current and an alternating current. Plate and grid voltages and grid current, if one exists, are similarly resolved into constant and alternating components. The constant or direct voltages are known from the applied potentials, or are readily deduced by a simple application of Ohm's law. Accordingly we need spend only a little time in determination of the direct currents and voltages and may reserve most of our effort for solution of the alternating currents and voltages.

Figure 12·1 shows a triode and the equivalent plate circuit theorem which applies at low frequencies. It is seen that this theorem considers only the alternating components of current and voltage. The direct currents and voltages are important only so far as they affect the incre-

* E. L. Chaffee, *Theory of Thermionic Vacuum Tubes*, page 274 ff., for example.

mental dynamic coefficients which are defined by the partial derivatives,

$$\mu = \frac{\partial e_p}{\partial e_g} \qquad r_p = \frac{\partial e_p}{\partial i_p} \qquad g_m = \frac{\partial i_p}{\partial e_g} \qquad [12\cdot1]$$

where μ is the dynamic amplification factor.

r_p is the dynamic plate resistance.

g_m is the dynamic transconductance.

i_p is the instantaneous plate current.

i_g is the instantaneous grid current.

e_p is the instantaneous plate voltage.

e_g is the instantaneous grid voltage.

The validity of the equivalent plate circuit theorem depends upon a more general network theorem due to Thévenin. The theorem proved by Thévenin states that a complex linear electrical network may be

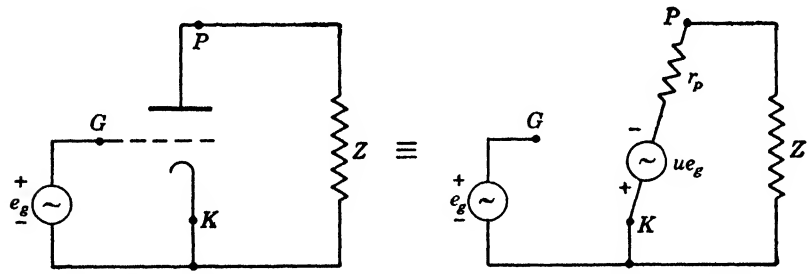


FIG. 12·1 Equivalent plate circuit theorem in constant-voltage form.

replaced by a single generator in series with a single impedance. The correctness of the equality indicated in Fig. 12·1 may be shown by allowing the load impedance Z to assume certain special values. If $Z = \infty$ the voltage developed across Z is equal to μe_g, which is equivalent to the definition of μ in equation 12·1. If $Z = 0$ the current which flows through Z is equal to $\mu e_g / r_p = g_m e_g$. Accordingly the current delivered to an impedance $Z = 0$ is $g_m e_g$ as required by the last definition of equation 12·1. The equality between the two arrangements of Fig. 12·1 is valid for other values of Z. This follows directly from the assumption of small applied voltages and of linear behavior throughout.

A modified form of the equivalent plate circuit theorem, frequently convenient, is shown in Fig. 12·2. Its validity may be proved by application of a general network theorem due to Norton or directly by an argument similar to the one presented for the series form of the theorem. This form of the theorem is often superior to that of Fig. 12·1 because the dynamic plate resistance is directly in parallel with the load impedance and with any plate-to-cathode capacitance which may exist. It

requires the concept of a constant-current generator whose output voltage is directly proportional to the total impedance connected to it. This concept is just as basic as that of the constant-voltage generator but is rather less familiar.

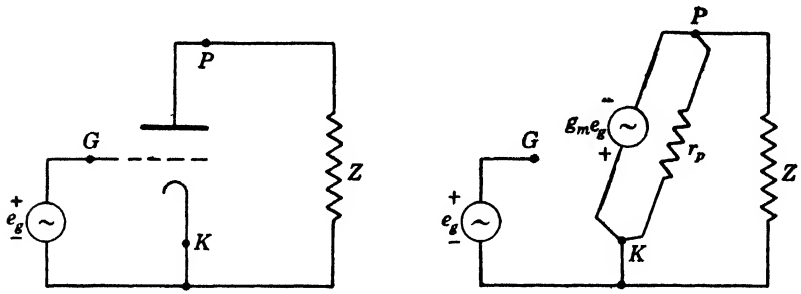


FIG. 12 2 Equivalent plate circuit theorem in constant-current form

The theorems as illustrated in Figs. 12·1 and 12·2 are special in the sense that they neglect the effect of any inductance or capacitance within the tube and that the reaction of the plate circuit upon the grid circuit is not considered. When the grid is biased negative with respect to the cathode there is no electronic current to the grid and the direct action of the plate circuit upon the grid circuit is insignificant. The effects of internal capacitance and of internal inductance will now be considered.

12·3 Effect of Lead Inductances and Internal Capacitances

It must be realized from the outset that both inductance and capacitance in a vacuum tube are essentially distributed elements. At times it is necessary to use the theory of distributed circuits in the solution of the tube behavior. In the present section, however, we shall consider the first approximation case, in which both inductance and capacitance are assumed lumped. A triode may conveniently be thought of as having lumped capacitances between each pair of elements and a lumped inductance in series with each lead. This inductance corresponds to that which exists between the region of thermionic action and the available external terminal. This approximation is illustrated in Fig. 12·3.

In Fig. 12·3 the capacitances C_g, C_k, and C_p are the capacitances which exist within the tube between grid and plate, grid and cathode, and plate and cathode, respectively. These capacitances are in the order of 5 micro-microfarads in typical receiving triodes and are mainly concentrated at the actual region where electrons move. The inductances L_k, L_g, and L_p are the effective inductances between the internal electrodes and the closest available external terminals.

A general equivalent plate circuit theorem consistent with these assumptions may be developed by means of the simple equivalent plate circuit theorem and appropriate network equations. Unfortunately the complicated nature of the equations which result tends to confuse rather than clarify the situation.

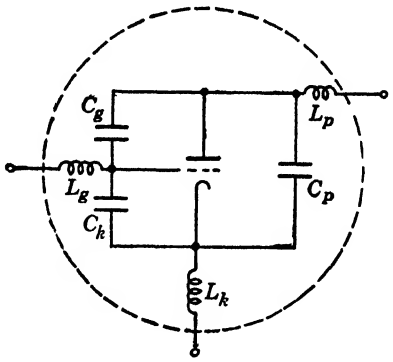


FIG. 12·3 Approximate network for triode at moderately high frequencies.

Input conductance due to grid-plate capacitance

Of the various parasitic elements shown in Fig. 12·3 the grid-plate capacitance C_g is the most important at low frequencies. Let us determine the effective input impedance of a triode due to this capacitance. The actual circuit and the equivalent circuit developed from the constant-current form of the equivalent plate circuit theorem are shown in Fig. 12·4. It is assumed that C_g is a lumped capacitance and that the inductances shown in Fig. 12·3 are negligible. The plate load is replaced by the admittance Y_3 and C_g is replaced by the admittance Y_4.

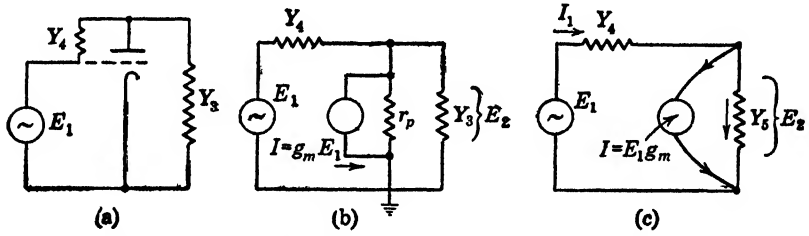


FIG. 12·4 The actual circuit and the equivalent circuits developed from the constant current form of the equivalent plate circuit theorem. It is assumed that the inductances of Fig. 12·3 are negligible and that $Y_4 = jwC_g$.

Let us introduce the substitution

$$Y_5 = Y_3 + \frac{1}{r_p} \qquad [12·2]$$

The equations which apply to the resulting network are

$$I_1 = (E_1 - E_2)Y_4 \qquad [12·3]$$

and

$$I_1 = E_2 Y_5 + E_1 g_m \qquad [12·3a]$$

Eliminating E_2,

$$I_1 = \left(E_1 - \frac{I_1}{Y_4}\right) Y_5 + E_1 g_m \qquad [12\cdot4]$$

which reduces to

$$I_1 \left(1 + \frac{Y_5}{Y_4}\right) = E_1(Y_5 + g_m) \qquad [12\cdot5]$$

The input admittance Y becomes

$$Y = \frac{I_1}{E_1} = Y_4 \frac{Y_5 + g_m}{Y_5 + Y_4} \qquad [12\cdot6]$$

Substituting,

$$Y_4 = j\omega C_g \qquad [12\cdot7]$$

$$Y = j\omega C_g \frac{Y_5 + g_m}{Y_5 + j\omega C_g} \qquad [12\cdot8]$$

Evidently the input admittance is infinite if $Y_5 = -j\omega C_g$. For values near this one the input conductance is large and positive or negative, depending upon whether

$$Y_5 < Y_4 \qquad \text{or} \qquad Y_5 > Y_4$$

In the special case $Y_5 = \infty$, corresponding to a short circuit in the plate circuit, the admittance reduces correctly to $Y = j\omega C_g$.

Thus we have seen that a grid-to-plate capacitance may produce an input conductance that is relatively large and either positive or negative. Oscillations are likely to occur if the conductance is negative, depending upon the magnitude of associated positive conductances. Moreover, this effect is quite pronounced even at low frequencies. It was the primary factor in the development of tetrode and pentode types of tubes. The electrostatic shielding produced by these additional grids greatly reduces the direct grid-plate capacitance.

Input conductance due to cathode lead inductance

In modern tetrodes and pentodes the effective grid-plate capacitance is negligibly small, even at the highest frequencies. An input conductance which is positive and which may be relatively large results from the combined action of the grid-cathode capacitance and cathode inductance. The behavior, which is quite different from that just investigated, is studied with reference to Fig. 12·5. The elements L_k and

C_k are replaced by the general impedances Z_1 and Z_2 for simplicity. The effective input impedance of the triode as affected by these impedances will be developed under the assumption that transit time is still negligibly short in comparison to the period of the applied signal. Either of the two standard forms of the normal equivalent plate circuit theorem may be applied, but the series form is more convenient here. The plate load is represented by the impedance Z_3.

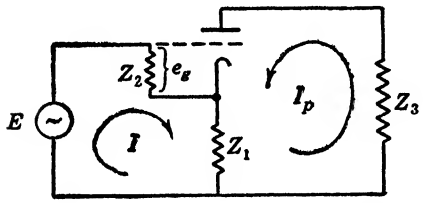


FIG. 12·5 Arrangement for calculation of the input impedance of a triode.

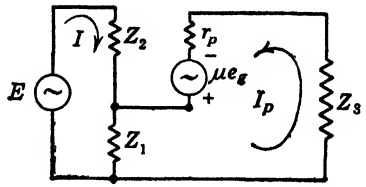


FIG. 12·6 Circuit equivalent to Fig. 12·5.

The actual circuit of Fig. 12·5 is replaced by the equivalent circuit of Fig. 12·6 for calculation. Consistent assumptions of current and voltage are indicated, and the familiar methods of mesh analysis are applied. The effective input impedance is $Z_{in} = E/I$. The value of this impedance may be determined by a relatively simple solution. The equations required are:

$$e_g \qquad -Z_2 I \qquad\qquad = 0 \qquad [12\cdot9]$$

$$\mu e_g \qquad -Z_1 I - (r_p + Z_3 + Z_1) I_p = 0 \qquad [12\cdot10]$$

$$(Z_2 + Z_1) I \qquad\qquad + Z_1 I_p = E \qquad [12\cdot11]$$

These equations have been so arranged as to provide a convenient determinant expression for I in terms of E:

$$I = \frac{\begin{vmatrix} 1 & 0 & 0 \\ \mu & 0 & -(r_p + Z_3 + Z_1) \\ 0 & E & Z_1 \end{vmatrix}}{\begin{vmatrix} 1 & -Z_2 & 0 \\ \mu & -Z_1 & -(r_p + Z_3 + Z_1) \\ 0 & (Z_2 + Z_1) & Z_1 \end{vmatrix}} \qquad [12\cdot12]$$

Minoring the numerator along the second column and the denominator

along the third column

$$I = \frac{-E\begin{vmatrix} 1 & 0 \\ \mu & -(r_p + Z_3 + Z_1) \end{vmatrix}}{Z_1\begin{vmatrix} 1 & -Z_2 \\ \mu & -Z_1 \end{vmatrix} + (r_p + Z_3 + Z_1)\begin{vmatrix} 1 & -Z_2 \\ 0 & (Z_2 + Z_1) \end{vmatrix}} \qquad [12\cdot13]$$

Thus, the input admittance becomes

$$\frac{I}{E} = \frac{(r_p + Z_3 + Z_1)}{\mu Z_1 Z_2 + Z_2 r_p + Z_2 Z_3 + Z_1 Z_2 + Z_1 r_p + Z_1 Z_3} \qquad [12\cdot14]$$

giving for the input impedance

$$Z_{in} = \frac{E}{I} = \frac{Z_1(\mu Z_2 + Z_3 + r_p) + Z_2(r_p + Z_3 + Z_1)}{r_p + Z_3 + Z_1} \qquad [12\cdot15]$$

In order to interpret 12 15, it is necessary to substitute the values $Z_1 = j\omega L$ and $Z_2 = \frac{1}{j\omega C}$, whereupon*

$$Z_{in} = \frac{j\omega L\left(\frac{\mu}{j\omega C} + Z_3 + r_p\right) + \frac{1}{j\omega C}(r_p + Z_3 + j\omega L)}{r_p + Z_3 + j\omega L} \qquad [12\cdot16]$$

Let us set $Z_3 + r_p = R$, implying a pure resistance load. The resulting equation is simplified to

$$Z_{in} = \frac{j\omega L\left(\frac{\mu}{j\omega C} + R\right) + \frac{1}{j\omega C}(R + j\omega L)}{R + j\omega L} \qquad [12\cdot17]$$

Rationalizing:

$$Z_{in} = \frac{j\omega LR\left(\frac{\mu}{j\omega C} + R\right) + \frac{R}{j\omega C}(R + j\omega L) + \omega^2 L^2\left(\frac{\mu}{j\omega C} + R\right) - \frac{L}{C}(R + j\omega L)}{R^2 + \omega^2 L^2} \qquad [12\cdot18]$$

* The subscripts previously used are dropped for simplicity in writing the equations No ambiguity arises because only one capacitance and one inductance are involved.

Separating real and imaginary components in the numerator,

$$Z_{in} = \frac{\frac{\mu LR}{C} + \omega^2 L^2 R + j\omega LR^2 + \frac{R^2}{j\omega C} + \frac{\omega L^2 \mu}{jC} - j\omega \frac{L^2}{C}}{R^2 + \omega^2 L^2} \qquad [12\cdot19]$$

Equation 12·19 shows two pure real terms. That is, two components of the expression for the input impedance are pure resistance. The expression for the input resistance is:

$$R_{in} = \frac{\frac{\mu LR}{C} + \omega^2 L^2 R}{R^2 + \omega^2 L^2} = \frac{R}{C} \frac{\mu L + \omega^2 L^2 C}{R^2 + \omega^2 L^2} \qquad [12\cdot20]$$

As ω approaches zero the input resistance approaches the constant value

$$R_{in} = \frac{\mu L}{CR} \qquad [12\cdot21]$$

As ω approaches infinity, the input resistance approaches the constant and relatively large value

$$R_{in} = R = r_p + Z_3 \qquad [12\cdot21a]$$

Let us examine equations 12·21 and 12·21*a* in terms of a practical ultra-high-frequency tube, the 955. The constants of this tube are closely, in our notation,

$$\begin{aligned} \mu &= 25 \\ r_p &= 12{,}500 \text{ ohms} \\ C &= 1.0\ \mu\mu\text{f} \\ L &= 0.01\ \mu\text{h} \end{aligned} \qquad [12\cdot22]^*$$

Assuming a load impedance of 7500 ohms plus r_p = 12,500 ohms, we have R = 20,000 ohms. At low frequencies, from equation 12·21

$$R_{in} = \frac{25 \times 10^{-8}}{1 \times 10^{-12} \times 2 \times 10^4} = 12.5 \text{ ohms} \qquad [12\cdot23]$$

At high frequencies, the input resistance would be 20,000 ohms if no other limiting effects occurred. At a frequency somewhat above 600 megacycles, $\omega = 4 \times 10^9$. For this frequency equation 12·20 yields

* Based upon the inductance of a straight wire 1.5 cm long and 0.1 mm in diameter. Bureau of Standards Handbook 274, p. 243.

$$R_{in} = \frac{20{,}000}{10^{-12}} \times \frac{25 \times 10^{-8} + 16 \times 100 \times 10^{-12}}{4 \times 10^{8} + 1600} \simeq \frac{2 \times 10^{4} \times 25 \times 10^{-8}}{4 \times 10^{-4}}$$

$$= 12.5 \text{ ohms}$$

The input reactance defined by the imaginary components of equation 12·19 is

$$X_{in} = \frac{\omega L R^2 - \dfrac{R^2}{\omega C} - \dfrac{\omega L^2 \mu}{C} - \omega \dfrac{L^2}{C}}{R^2 + \omega^2 L^2} \qquad [12\cdot24]$$

At $\omega = 4 \times 10^9$, this reactance becomes

$$X_{in} =$$

$$\frac{40 \times 4 \times 10^{8} - \dfrac{4 \times 10^{8}}{4 \times 10^{-3}} - \dfrac{4 \times 10^{9} \times 10^{-16} \times 25}{10^{-12}} - \dfrac{4 \times 10^{9} \times 10^{-16}}{10^{-12}}}{4 \times 10^{8} + 1600}$$

$$X_{in} = \frac{16 \times 10^{9} - 10^{11} - 10^{7} - 4 \times 10^{5}}{4 \times 10^{8}} = \frac{-10^{11}}{4 \times 10^{8}} = -250 \text{ ohms}$$

The effective input impedance defined by equation 12·19 is thus 12.5 ohms resistance in series with 250 ohms of capacitive reactance at approximately 600 megacycles. This is equivalent to a resistance of 5000 ohms shunting a capacitive reactance of 250 ohms.

Although this is not an extremely low shunting resistance, it must be remembered that the tube in question was especially designed to have minimum values of inductance and capacitance.

Examination of equations 12·20 and 12·24 shows that for the lower values of ω the input resistance is directly proportional to L and μ and is inversely proportional to R and C. The input reactance is closely equal to $1/\omega C$. Combining these two equations the input impedance is given approximately by

$$Z_{in} \simeq \frac{\mu L}{CR} + \frac{1}{j\omega C} \qquad [12\cdot25]$$

As long as the resistance is low compared to the reactance, we may write

$$\frac{R}{X} = \frac{G}{B}, \quad G = \frac{BR}{X}, \quad \text{and} \quad B = \frac{1}{X} \qquad [12\cdot26]$$

where G is equivalent conductance, and B is the susceptance. Substituting the values of R and X from equation 12·25

$$G = \frac{R}{X^2} = \frac{\omega^2 C^2 \mu L}{CR} = \frac{\omega^2 \mu LC}{R} \qquad [12\cdot27]$$

In the limiting case of no external load impedance or in most cases involving pentodes $g_m = \mu/R$, and the input conductance becomes

$$G_{in} = \omega^2 g_m LC \qquad [12\cdot28]$$

Since 12·28 varies as the square of the frequency, it is readily seen that moderate extensions of operating frequencies may greatly increase the input conductance. Experimental data on the input conductance of high-frequency pentodes due to cathode inductance are shown in Fig. 12·7. In this case, transit time effects are negligible. The term input loading is appropriate because the conductance from grid to cathode serves to load the circuit which feeds the grid.

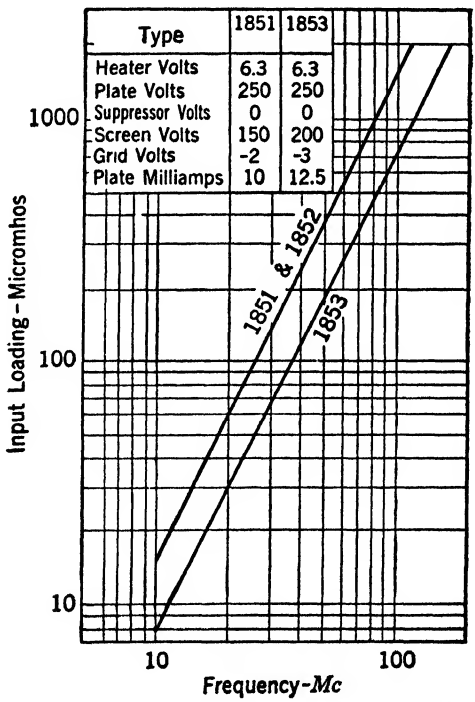


FIG. 12·7 Experimental data showing the effect of cathode inductance on the input conductance of several high-frequency pentodes.

Output conductance

By a process similar to that employed in the development of the input conductance it may be shown that the output conductance of triodes

and pentodes is increased by the effects of lead inductance. Usually this effect is much less important than that of input conductance, and often it is negligible. Strutt and van der Ziel* have made a very thorough study of the effects of lead inductance and of interelectrode capacitance. They present experimental data and a very complete set of equations for calculating the performance of screen-grid and more complex tube structures. In particular, they conclude that multielement tubes are, in general, more subject to these effects than triodes. The upper frequency limit for satisfactory operation of the multielectrode tubes is generally set by lead effects, whereas the upper limit for triodes is set by transit time.

12·4 Transit Time Effects in Diodes

An electron emitted from the cathode of a radio tube requires a finite length of time to arrive at the plate. In tubes of ordinary size and operating at normal potentials, this is a very short time, in the order of 10^{-9} second. Such a short time would ordinarily be regarded as entirely negligible, and in very many applications it is truly insignificant. When we deal with very high frequencies, however, we find it impossible to neglect the time of flight of the electrons. In fact, most of the important generators of high frequencies base their entire operation upon transit time effects. We must, therefore, carefully investigate such effects to see how they limit the operation of ordinary tubes and how they may be used to advantage in special tubes.

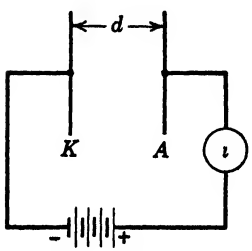


FIG. 12·8 Plane parallel diode.

Two important considerations pertaining to transit time may be developed by reference to Fig. 12 8. A plane parallel structure is assumed in which the left-hand plate is a cathode emitting electrons and the right-hand plate is the anode. As in the familiar calculations of the parallel plate condenser, it will be assumed that the plates are large in comparison to the spacing so that the electric intensity is uniform between the plates.

Calculation of transit time

From the arrangement of Fig. 12·8, we may write $E = V/d$, where E is the electric intensity, d is the distance, and V is the applied voltage. If emission velocity and contact potential are negligible the electrons which

*M. J. O. Strutt and A. van der Ziel, "The Causes for the Increase of the Admittances of Modern High-Frequency Amplifier Tubes on Short Waves," *Proc. IRE*, **24**, 1011, August, 1938.

reach the anode have a velocity given by

$$eV = \tfrac{1}{2}mv^2 \qquad [12\cdot29]$$

where e is the electron charge, m is the electron mass, and v is the velocity with which the electrons reach the plate. Equation 12·29 when solved for v gives

$$v = \sqrt{2\frac{e}{m}V} \qquad [12\cdot30]$$

Using e/m and V in practical units we obtain

$$v = 6 \times 10^5\sqrt{V} \quad \text{meters/sec} \qquad [12\cdot31]$$

Since the electric intensity E between the plates is constant, the force accelerating the electron is constant and the velocity is directly proportional to the length of time that the electron has experienced the force. This is one form of uniformly accelerated motion. For these conditions the average velocity over the interval is half of the final velocity.

Accordingly we may write for the average velocity

$$\bar{v} = 3 \times 10^5\sqrt{V} \quad \text{meters/sec} \qquad [12\cdot32]$$

The total transit time is calculated from the basic relation $d = \bar{v}t$ or $t = d/\bar{v}$. Thus

$$t = \frac{d}{3 \times 10^5\sqrt{V}} = \frac{3.3 \times 10^{-6}}{\sqrt{V}}d \quad \text{second} \qquad [12\cdot33]$$

Assuming values leading to a relatively short transit time, namely, $V = 100$ volts, $d = 0.1$ cm (10^{-3} meter),

$$t = \frac{3.3 \times 10^{-6} \times 10^{-3}}{10} = 3.3 \times 10^{-10} \quad \text{second} \qquad [12\cdot34]$$

corresponding to one full cycle of a wave of 3000-megacycle frequency.

Condition of current flow

When we deal with low-frequency systems we seldom need to know the exact interval during which an electron is active in carrying current. A very large number of electrons pass from cathode to anode during the period of the alternating current, and it is therefore necessary only to multiply this number by the electronic charge in order to determine the electric charge transferred and hence the current.

When the period of the wave is comparable to the transit time, however, the situation is more complex. It is necessary to know over what

interval the current flow associated with an individual electron motion exists. The result may be developed directly by an appropriate interpretation of Maxwell's equations. A development based on the conservation of energy is presented here.

Referring to Fig. 12·8 the instantaneous power supplied by the external circuit is equal to iV where V is the constant potential difference and i is the instantaneous current. A single electron existing in the cathode-anode space is accelerated by the electric field, and the power supplied by the external circuit is converted into kinetic energy by this process. The mechanical power is given by the product Fv, where F is the mechanical force acting and v is the electron velocity. In the electric field the force is equal to eE, and we may write

$$Vi = eEv \qquad [12\cdot35]$$

or

$$i = \frac{eEv}{V} = \frac{ev}{d} \qquad [12\cdot36]$$

Equation 12·36 indicates that a current flows during the entire interval that the electron is in motion and that the current is proportional to the instantaneous velocity of the electron. That the current is also proportional to the electronic charge e is clear from basic definitions of current. The inverse proportionality to d may be argued from the fact that a definite charge is transferred in less and less time as d approaches zero. Figure 12·9 shows the current flow in the external circuit which results if electrons are freed from the cathode at two different regular intervals. In Fig. 12·9*a* the emission rate is slow enough so that one electron reaches the plate before the next electron leaves the cathode. In Fig. 12·9*b* the rate is such that two electrons exist in the cathode-anode space at one time. The sharp corners on these wave forms are of great interest because they constitute the basis of tube or shot noise.

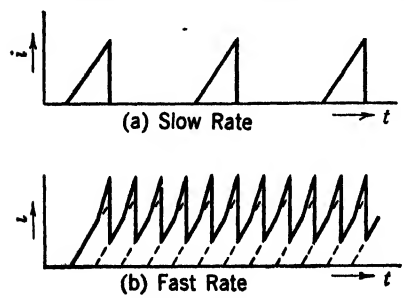


FIG. 12·9 Currents due to regular electron emission in Fig. 12·8.

Although the preceding development is based on the assumption of a plane parallel diode the results are correct for cylindrical and other structures. Accordingly, we may accept them as generally applicable. In actual tubes large numbers of electrons are always present and the current never falls to zero. The actual electronic current is therefore

the sum of the currents due to the individual electrons, as defined by equation 12·36. To this must be added the displacement current. The total current is therefore the familiar expression

$$I = \left(\overline{\rho v} + \frac{\partial \bar{D}}{\partial t}\right) A \qquad [12\cdot37]$$

where $\overline{\rho v}$ is the average of the product of charge density by velocity in the space, $\partial \bar{D}/\partial t$ is the average displacement current density, and A is the area of the plates.

Certain assumptions, not immediately apparent, were made in the foregoing development. It is necessary to examine all such assumptions with great care if we are to assure ourselves that the resulting concept is correct. First we assumed that the current i was the same everywhere in the circuit external to the tube. Second, we assumed that the electric intensity was uniform throughout the space between the plates. Actually these two assumptions are very closely linked, the second necessarily being true if the first is true, but not conversely. The first assumption cannot be satisfied if the external circuit is allowed to expand without limit. For the purposes of the development, however, the physical length of the external circuit need be only enough to connect the two plates. In this case the two assumptions become practically synonymous.

We found the time of electron flight to be 3.3×10^{-10} second for the condition chosen above. Let us now calculate the time required for an electromagnetic wave to be propagated from one plate to the other. Using the basic equation $d = ct$, where c is the velocity of light, we have $10^{-3} = 3 \times 10^{8}t$ or $t = 3.3 \times 10^{-12}$ second. That is, the wave travels the distance d in 1 per cent of the time required for the electron to travel the same distance. Thus the current in the two leads is indeed practically equal and is equal to the current defined by 12·37. The following sections will be devoted to a study of conditions in which this situation exists.

Graphical development of current flow

Important qualitative concepts with respect to transit time effects may be gained by reference to Fig. 12·10. A plane diode such as that shown in Fig. 12·8, is subjected to a square voltage wave. The assumption of a square wave simplifies the analysis and gives results that are qualitatively correct. Again the motion of single electrons is studied; that is, the effects of space charge are neglected.

It will be recalled that the distance traveled by a body under uniform acceleration is proportional to the square of the elapsed time. Accord-

ingly the distance curves of Fig. 12·10 are parabolic. Likewise the kinetic energy of each electron is directly proportional to the distance traveled. It is readily seen that electrons which leave the cathode just before the reversal of voltage are unable to reach the plate, but must return to the cathode. Since the accelerating force is equal in both

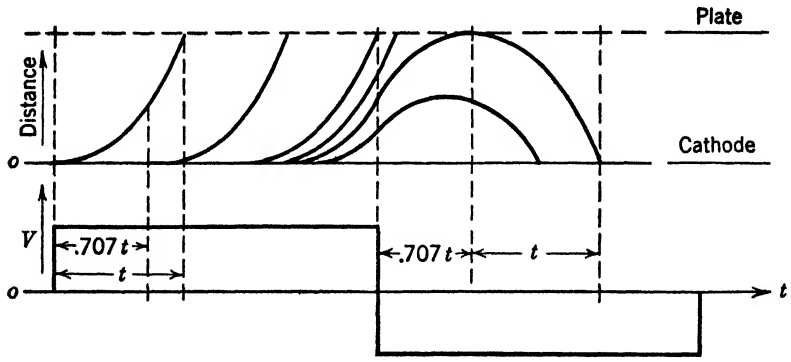


FIG. 12·10 Time-distance curves for individual electrons in a diode. The effect of space charge is neglected.

halves of the wave but is opposite in direction it is evident that the distance the electron has traveled from the cathode at the time of the reversal is exactly doubled before the electron turns back to the cathode. Therefore, electrons which are already more than half way to the plate at the time of the reversal of the voltage reach the plate. Those which are less than half way to the plate are reversed and return to the cathode.

A current wave corresponding to Fig. 12·10 may be deduced if we assume that a relatively large number of electrons are emitted at a uniform time rate but that space charge is still negligible. We must now consider the number of electrons in motion and their average velocity. The development is shown in Fig. 12·11. To this electronic current is added the displacement current. Because of our assumption of a square voltage wave the displacement current takes the form of sharp impulses flowing only at the time of voltage reversal.

The negative electronic current which flows during part of each cycle is due to the electrons which nearly reach the plate and return to the cathode with considerable velocity only after a relatively long time interval.

Figure 12·12 shows a similar condition in which the transit time is a considerably larger portion of a cycle. Curves of distance, voltage, and current are developed by the methods used in Figs. 12·10 and 12·11.

The above development shows that the current wave form need not

resemble the voltage wave form. It suffers, however, from three rather important limitations. The square wave form assumed is neither

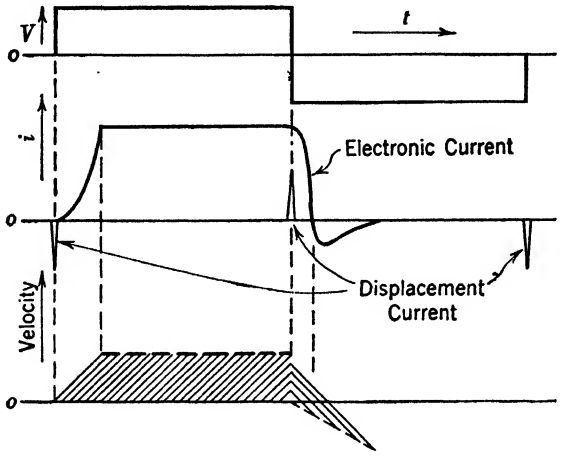


FIG. 12·11 Development of current in a diode having finite transit time, when a large number of electrons is considered. Again the effect of space charge is neglected.

desirable nor readily realized in practice. The effect of space charge is neglected, and no direct-current bias is assumed. That is, the present

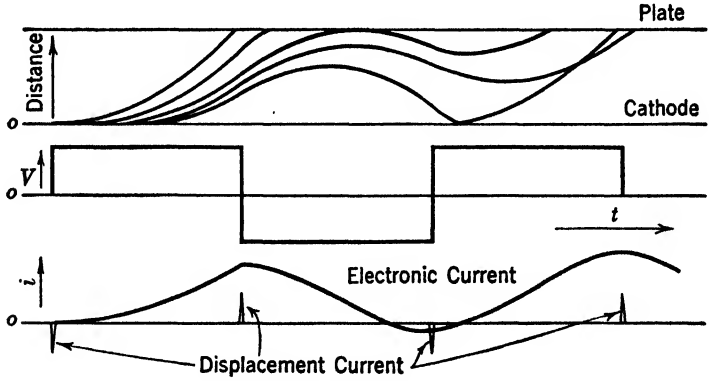


FIG. 12·12 Development of electron current at still higher frequency than that of Fig. 12·11.

development is based upon the so-called large signal theory. We shall now consider mathematically a case in which these limitations are removed.

12·5 Transit Time Effects in Triodes

We have already shown that the current density in the space between two parallel plates is uniform and equal to the Maxwellian form

$$i = \overline{\rho v} + \frac{\partial \bar{D}}{\partial t} \qquad [12\cdot37]$$

Under direct-current conditions of a constant applied voltage the current i is constant and $\partial \bar{D}/\partial t$ is equal to zero. The product $\overline{\rho v}$ is uniform over the volume, and regions of high charge density correspond to regions of low velocity, and conversely. That is, in the neighborhood of the plate where the velocity is high the charge density is low, and in the neighborhood of the cathode the velocity is low and the charge density high.

It is next appropriate to consider a plane parallel triode such as is shown in Fig. 12·13. The grid mesh is assumed to form an effective

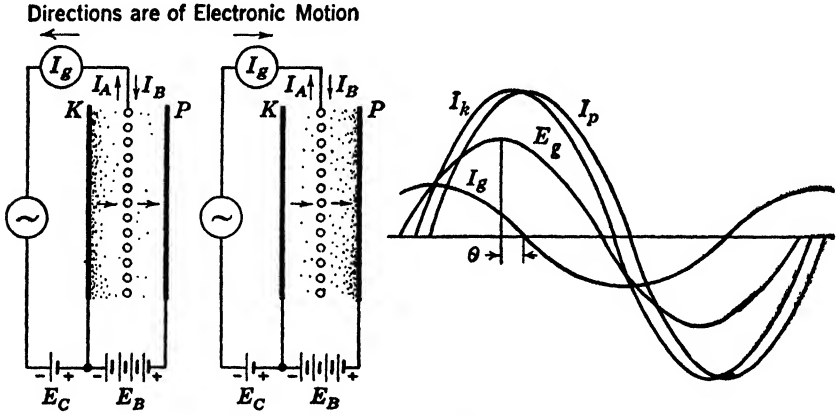


FIG. 12·13 Negative grid triode at high frequency.

equipotential plane, and the current from cathode to plate must be considered in two components. One current equal to $\overline{\rho v} + (\partial \bar{D}/\partial t)$ flows in the grid-cathode space, and a current of the same form flows in the grid-plate space. Under direct-current conditions $\partial \bar{D}/\partial t$ is equal to zero, and if the grid is biased negatively the net current to the grid is zero. Accordingly the product ρv is the same in the grid-plate region as it is in the grid-cathode region.

Now let us consider such a triode in which no alternating voltage exists between plate and cathode, but a small voltage of very high frequency is applied to the grid. The bias voltages are within the normal linear operating range of the tube so that all conditions are typical of class A amplification except that the frequency is very high.

Again it is necessary to consider the grid current as the sum of the grid-cathode and the grid-plate currents. Thus the effective grid current is zero only if these two currents are equal and opposite in phase. The displacement current obeys the ordinary relationship and so may be calculated from the low-frequency capacitance. Our primary interest must be focused upon the electronic or conduction current which may have components both in phase and in quadrature with the voltage.

During the time that the grid voltage is swinging in the positive direction the current leaving the cathode is increasing. The additional electron flow which comprises the grid-cathode current I_k does not reach its maximum value until somewhat later than the maximum value of E_g This is similar to the cases of Figs. 12·11 and 12·12. Moreover, the grid-plate current does not reach its maximum until a still later time, owing to the finite velocity of the electron cloud. The grid current, which is the difference between the grid-cathode and gride-plate currents, is shown as I_g in Fig. 12·13. It is seen to have a considerable component in phase with the voltage. The main component of the current is such as would flow in a capacitive reactance.

The physical situation may be condensed into this simplified statement. During the time the grid is more positive than normal, the number of electrons approaching the grid is greater than the number leaving, and work is required. During the next half cycle when the grid is more negative than normal, the number of electrons leaving the grid toward the plate is greater than the number approaching from the cathode, and again work is required. It can be shown that the power absorbed from the grid circuit appears as heat at the plate owing to the fact that the electron velocities are increased by the action of the grid.

Ferris* presents a very simple and ingenious method for developing the magnitude of the input conductance. He argues that the total grid current is proportional to the frequency f and to the transit time t since these things determine the existence of a grid current. Further, the grid current must be proportional to the grid voltage E_g and to the transconductance g_m. Using an undetermined coefficient K_1 to express the magnitude of the effect we write

$$I_g = K_1 E_g g_m f t \qquad [12·38]$$

and

$$Y_g = \frac{I_g}{E_g} = K_1 g_m f t \qquad [12·39]$$

* W. R. Ferris, "Input Resistance of Vacuum Tubes as Ultra-High-Frequency Amplifiers," *Proc. IRE*, **24**, 81, January, 1936.

From Fig. 12·13 it is seen that the grid current arises as a difference between two sinusoidal waves which are almost but not quite in time phase. As the frequency becomes higher this deviation from phase opposition becomes greater and the effective current is proportional to frequency as shown by equation 12·39. As the two components of current become more and more out of phase with the applied voltage the resultant current i_g deviates more and more from its quadrature relation with the voltage. This angle of deviation, designated θ in Fig. 12·13, is also proportional to frequency. Because θ is small we are justified in the approximation

$$\theta \simeq \sin \theta \qquad [12\cdot40]$$

We may therefore write $G = Y_g \sin \theta \simeq Y_g\theta = Y_gK_2f$ or finally

$$G = Kg_mf^2t^2 \qquad [12\cdot41]$$

where $K = K_1K_2$ and G is the effective input conductance.

In the development of equation 12·41 no definite assumption was made as to the division of the transit time in the grid-cathode and grid-anode spaces. Accordingly K may depend upon this factor as well as upon the geometry and size of the tube, and the applied voltage. In common with equation 12·28 it gives the input conductance proportional to the square of the frequency. Thus it is impossible to segregate conductance due to transit time from that due to lead inductances by any simple experimental method.

North* develops an equation of the same form as 12·41 with numerical coefficients by a comprehensive mathematical treatment.

The mathematical difficulties which arise in a complete solution of problems involving transit time are very great. Llewellyn† has published solutions to a number of problems concerning vacuum tubes at high frequencies, but the results achieved are extremely complicated. He finds it necessary to resolve the space charge, electron velocity, electric intensity, and current density into direct and alternating components as was previously done for the total current and voltage.

A number of interesting conclusions result from these solutions. One of the most startling is that a plane parallel diode shows a negative

* D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance," *Proc. IRE*, **24**, 108, January, 1936.

† "Vacuum Tubes Electronics at Ultra-High Frequencies," *IRE*, **21**, 1532; "Phase Angle of Vacuum Tube Transconductance at Very High Frequencies, *IRE*, **22**, 947; "Note on Vacuum Tube Electronics at Ultra-High Frequencies," *IRE*, **23**, 112; "Operation of Ultra-High-Frequency Vacuum Tubes," *Bell System Tech. J.*, **14**, 632; "Equivalent Networks of Negative Grid Vacuum Tubes at Ultra-High Frequencies," *ibid.*, **15**, 575.

conductance or resistance for certain values of frequency. The most favorable is that for which the transit angle is approximately 450°. That is, the time of flight is equal to the time required for $1\frac{1}{4}$ full cycles of the wave. Negative resistances are also achieved for higher frequencies corresponding to $2\frac{1}{4}$, $3\frac{1}{4}$, etc., cycles, but conditions for oscillation are less favorable. An oscillator applying this principle is described in Chapter 15.

Another rather surprising result of these analyses is that the alternating plate current is non-sinusoidal even though the grid voltage is sinusoidal; no electrons are captured by the grid, and all conditions are favorable for linear operation at low frequency. This result comes about because of the variable velocity with which electrons reach the plate. The square-wave analysis of a diode previously presented leads to a similar conclusion.

Benham* has carried out studies similar to those of Llewellyn. In addition to the equation for the resistance of the plane parallel, space-charge-limited diode, he presents formulas for the input and output impedance of triodes at high frequencies. He also develops precise coefficients for the transconductance and amplification factor of cylindrical triodes at low frequencies. These results are deduced by a direct application of equations published by Maxwell.†

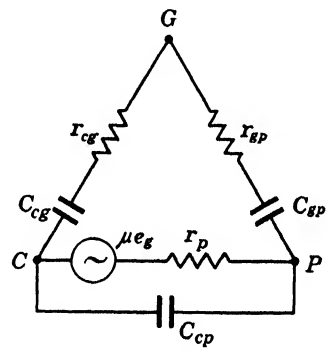


FIG. 12·14 Circuit equivalent to a vacuum tube for moderately high frequencies.

The results obtained by Llewellyn and Benham are complicated and difficult to interpret. They are, however, of fundamental importance, and it is expected that simpler approximation forms will be developed which will find wide application. The results are important in that they establish definite limits of low-frequency technique.

A step toward the simplification of the results of transit time calculations has been taken by Llewellyn.‡ He presents a modified equivalent plate circuit theorem which includes the effects of interelectrode capacitances and of transit angles not exceeding 30°. The effects of lead inductance are neglected. The circuit diagram is shown in Fig. 12·14,

* W. E. Benham, "A Contribution to Tube and Amplifier Theory," *Proc. IRE*, 26, 1093, September, 1938.

† Clerk Maxwell, *Electricity and Magnetism*, Oxford University Press, 1891.

‡ F. B. Llewellyn, "Equivalent Networks of Negative Grid Vacuum Tubes at Ultra High Frequencies," *Bell System Tech. J.*, 15, 575.

and the coefficients which apply are presented in Table 12·1. These expressions are considerably simplified if the ratio of grid-plate to cathode-grid transit time approaches zero. The coefficients which apply in this case are given in Table 12·2. Usually this approximation is sufficiently accurate in practice.

TABLE 12·1

Let $C_{cg'}$, $C_{gp'}$,$C_{cp'}$ = capacitances of cold tube.

$$y = \frac{x_p}{x_c} = \text{ratio of grid-plate to grid-cathode spacing.}$$

$$h = \frac{T_p}{T_c} = \text{ratio of grid-plate to grid-cathode transit time.}$$

$$M = \tfrac{4}{3}(y - h^3)(1 + h) - 2h^2 + h^4.$$

$$N = (y - h^3)(9 + 44h + 45h^2) - 51h^2 - 105h^3 - 27h^4 + 27h^5.$$

Then

$$C_{cp} = \tfrac{4}{3}C_{cp'}\left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0}\right]\left[1 + h - \frac{3}{2}\frac{h^2}{y - h^3}\right].$$

r_p = same as at low frequencies.

$$C_{cg} = C_{cg'}\left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0}\right]\frac{M}{y}.$$

$$r_{cg} = \left[\frac{r_p(y - h^3)}{45\mu_0(1 + M + \mu_0)M^2}\right]\left[(\mu_0 + 1)N + \frac{45\mu_0 h^4}{(y - h^3)}M\right].$$

$$C_{gp} = C_{gp'}\left[\frac{1 + y + \dfrac{\mu_0 y}{y - h^3}}{1 + M + \mu_0}\right].$$

$$r_{gp} = -\left[\frac{r_p(y - h^3)}{45\mu_0(1 + M + \mu_0)}\right]\left[N - \frac{45\mu_0 h^4}{y - h^3}\right].$$

(Llewellyn, courtesy of Bell System Tech. J.)

12·6 Noise in Vacuum Tubes

Maxwell's equations are derived on the assumption that electricity is infinitely divisible. They apply to a majority of the problems which are important to electrical engineers but fail where dimensions of atomic order are of importance. Specifically, results involving very small distances or very small electrical quantities are not generally valid.

Faraday's experiments with electrolysis led him to the conviction that there is a smallest possible electrical quantity. Later Sir J. J. Thomson identified this quantity with the charge of the electron and determined

TABLE 12·2

Let $C_{cg'}$, $C_{gp'}$, $C_{cp'}$ = capacitances of cold tube.

$$y = \frac{x_p}{x_c} = \text{ratio of } g - p \text{ to } c - g \text{ spacing.}$$

$$h = \frac{T_p}{T_c} = \text{ratio of } g - p \text{ to } c - g \text{ transit time.}$$

Then, when $h \to 0$:

$$C_{cp} = \tfrac{4}{3} C_{cp'} \left[\frac{1 + y + \mu_0}{1 + \tfrac{4}{3} y + \mu_0} \right].$$

r_p = same as at low frequencies.

$$C_{cg} = \tfrac{4}{3} C_{cg'} \left[\frac{1 + y + \mu_0}{1 + \tfrac{4}{3} y + \mu_0} \right].$$

$$r_{cg} = \frac{9}{80} \frac{r_p}{\mu_0} \left[\frac{1 + \mu_0}{1 + \tfrac{4}{3} y + \mu_0} \right].$$

$$C_{gp} = C_{gp'} \left[\frac{1 + y + \mu_0}{1 + \tfrac{4}{3} y + \mu_0} \right].$$

$$r_{gp} = -\frac{1}{5} \frac{r_p}{\mu_0} \left[\frac{y^2}{1 + \tfrac{4}{3} y + \mu_0} \right].$$

(Llewellyn, courtesy of Bell System Tech. J.)

the ratio of the electronic charge to mass. The electronic charge was determined by Millikan and others and is now regarded as one of the basic physical constants.

We have seen in Fig. 12·9 that a uniform stream of electrons from cathode to plate in a diode does not constitute a steady current. Each individual electron constitutes a component of current which increases with time as the electron travels from cathode to plate and suddenly falls to zero when the electron strikes the plate. Accordingly we are not surprised to discover that vacuum tubes which have no input signal give a small output consisting of equal voltages over a wide band of frequencies. The magnitude of this output may be calculated from fundamental considerations and is accurately verified experimentally.

Closely associated with this unwanted or noise output from vacuum tubes is the noise produced by all passive elements, such as wire coils, which have resistance. The electrons within such conductors are agitated by the Brownian or thermal agitation of the molecules. This random vibration of the electrons within the metal is closely analogous to the vibration of gas molecules within a closed chamber. The force acting on a unit area of such a container is not constant but varies in a perfectly haphazard fashion with time, depending upon the number of

molecules which chance to strike it in some short interval of time. Similarly the potential difference between the ends of a resistor is not zero but varies from instant to instant depending upon the motion of the electrons within.

The magnitude of this effect due to thermal agitation of the electrons may be calculated by at least two basic methods which yield the same result. One development conceives a resistor connected to an ideal capacitance microphone which is subjected to the molecular bombardment of the air. This bombardment of the microphone produces a voltage which dissipates power in the resistor. Since the laws of thermodynamics forbid one portion of a passive system to be heated at the expense of another it is necessary that the resistor generate a voltage equivalent to that of the microphone. This may be thought of as a force which actuates the microphone as a loud speaker and so dissipates power.

The second development employs a long dissipationless transmission line terminated at either end in its characteristic impedance. Standing waves of the familiar form are postulated, and the limiting case as the line approaches infinite length is taken.

In either event a definite power output is deduced. This may be expressed in a variety of ways, but one of the most convenient is the statement that any resistor delivers to an equal resistor a power of

$$P = 1.36T \times 10^{-23} \quad \text{watt per degree Kelvin per cycle of band width} \qquad [12\cdot42]$$

For a temperature of 300° K (27° C), approximately normal room temperature, this expression becomes

$$P = 4.1 \times 10^{-21} \quad \text{watt per cycle of band width} \qquad [12\cdot43]$$

The effects of thermal noise may also be expressed in terms of the equivalent rms voltage V_t produced by a single resistor. This may be written for a temperature of 27° C

$$\overline{V_t^2} = 1.64FR \times 10^{-20} \quad \text{volt}^2 \qquad [12\cdot44]$$

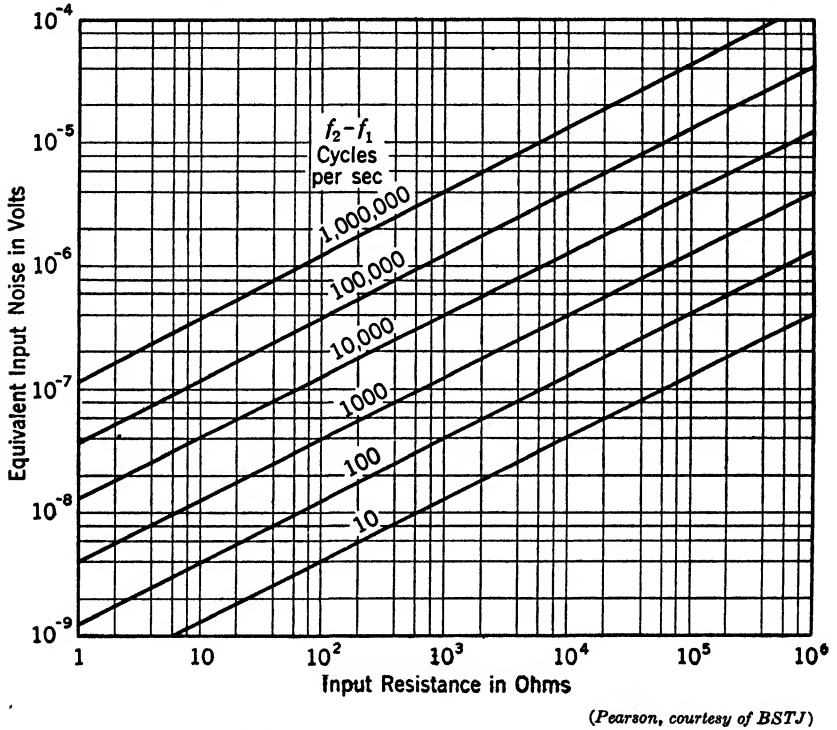
where R is the resistance over a frequency band F cycles wide.

The factor of 4 which seems to have disappeared between equations 12·43 and 12·44 is explained by the fact that a voltage in series with a resistor R delivers a maximum power to another resistor also equal to R, and the power is $e^2/4R$.

Equation 12·44 is plotted in Fig. 12·15 for the ordinary condition of 300° K or 27° C. In practice deviations of temperature from this

average figure are seldom large enough to warrant the use of a correction factor with these curves.

It is convenient to express the noise produced by vacuum tubes in terms of the resistor in its grid circuit which would give the same noise output power. Since noise consists of a random distribution of all frequencies, addition is always carried out on a power basis. The above definition is equivalent to saying that tube noise is expressed in terms of the resistor in the grid circuit which will double the noise output power.



(*Pearson, courtesy of BSTJ*)

FIG. 12·15 Thermal noise level as a function of input resistance and frequency range.

It is well known that tubes which operate with complete space-charge limitation of plate current are less noisy, that is, produce less random output power, than tubes which operate with less copious emission. It is also true that the equivalent noise resistance of a tube is reduced as the ratio of transconductance to plate current is increased. This last fact is explained on the basis that the noise due to a given resistor in the grid circuit is amplified more as the transconductance is increased and that the noise due to the plate current in the tube is therefore less important.

In low-frequency apparatus the input signal may be presented to the grid of the first tube of the system by means of a step-up transformer. The impedance presented to the grid of the first tube is thus made high across the band of frequencies in question, which is another way of saying that the useful signal or its associated thermal noise is made to override the effects of tube noise. Any signal which is appreciably more powerful than thermal noise is then capable of reception by the application of a sufficient amount of amplification. A signal which is weaker than thermal noise is useless.

At higher frequencies the required band widths are often large. In the following chapter it will be shown that low impedances are unavoidably required if wide frequency bands are to be transmitted. It is therefore difficult if not impossible to make the effects of tube noise negligible in comparison to those of thermal noise. Tubes having high transconductance and especially designed to reduce the effects of noise must be used if the best possible results are to be obtained.

The Klystron, discussed in Chapter 17, presents problems of tube noise that are quite different from those associated with ordinary tube structures. At the present time the Klystron is appreciably more noisy than a low-frequency tube of comparable performance. Accordingly the Klystron, in its present form, is not desirable as an amplifier of very small signals.

PROBLEMS

12·1 Calculate the equivalent rms terminal voltage due to thermal noise of a 600-ohm resistor at 27° C over a frequency band 3500 cycles wide.

12·2 Calculate the equivalent rms terminal voltage due to thermal noise of a 600-ohm resistor at −73° C over a frequency band 7000 cycles wide.

12·3 Calculate directly the power delivered by the resistor of problem 12·1 to a similar resistor at the same temperature. Verify by application of equation 12·43.

12·4 Calculate the power delivered by the resistor of problem 12·1 to a 200-ohm resistor at the same temperature. Verify the result by calculating the power delivered to the 600-ohm resistor by the 200-ohm one.

12·5 A vacuum-tube amplifier transmits a band 100,000 cycles wide. It delivers a power output of 0.001 watt with the grid of the first tube short-circuited, 0.0015 watt with 1000 ohms in the first grid circuit, and 0.0025 watt with 3000 ohms in the first grid circuit. Express the noise rating of the first tube.

12·6 A particular vacuum tube operates with a fixed plate potential which is 500 volts more positive than the cathode. Calculate the velocity in meters per second with which an electron reaches the plate.

12·7 How large a discontinuity in the plate current occurs at the instant an electron reaches the plate in problem 12·6?

12·8 A high gain broad-band amplifier having considerable noise output is connected to the vertical plates of a cathode-ray oscilloscope. The sweep circuit which

provides the horizontal deflection is adjusted to a very low frequency so that only one trace at a time is visible on the screen. What sort of trace is to be expected? Sketch roughly.

12·9 Repeat problem 12·8 with the exception that the amplifier is sharply tuned.

12·10 Repeat problem 12·8 except that the horizontal deflecting sweep is fairly rapid so that many traces are superimposed upon the screen at one time.

12·11 Assume a diode subjected to a square voltage wave such as is shown in Fig. 12·12. Let the electron transit time be $1\frac{1}{4}$ cycles. Assume a uniform rate of electron emission, and sketch a curve covering several complete cycles so as to approximate the steady state of current flow. Estimate the phase relation of the fundamental components of voltage and current waves. Is the equivalent resistance positive or negative?

12·12 Show that the two forms of the equivalent plate circuit theorem shown in Figs. 12·1 and 12·2 deliver the same current to the load impedance Z provided that the normal identity $\mu = g_m r_p$ is valid.

12·13 A certain triode has a grid-to-plate capacitance C_g and an impedance Z between grid and cathode terminals. Develop the effective plate-to-cathode impedance of this tube in terms of these constants and the vacuum-tube parameters.

12·14 A coil and condenser having negligible loss form a parallel resonant circuit at 50 megacycles. They serve to feed the grid of a pentode 1851 (see Fig. 12·7). If the Q of the combination is to be 50, determine the constants of the coil and condenser. (In such a parallel combination $Q = R/\omega L = R\omega C$).

12·15 Two radio frequency tetrodes are used as small signal amplifiers. They differ only in the distance from the screen grid to the plate, all other dimensions being the same. The plate and screen grid voltages are equal to each other and are the same for both tubes. How will transit time affect the comparative performances of these two tubes?

12·16 A triode operates with no alternating plate voltage and with a voltage of very high frequency applied to the grid. What becomes of the power transferred to the electron stream by the grid? Why?

CHAPTER 13

AMPLIFIERS

13·1 Introduction

One of the first and probably still the most important application of vacuum tubes is amplification. In spite of innumerable other applications, more tubes are used in amplifiers than in any other way. The theory and design of amplifiers, particularly those for low- and intermediate-frequency ranges, is better understood than the high-frequency applications. It is, therefore, important that we consider the subject carefully.

It will be assumed that the reader has an elementary knowledge of the general subject since space does not permit a complete review of the topic here. We shall review the operation of the audiofrequency amplifier first, since this analysis is basic to a discussion of high-frequency amplifiers.

The subject of amplification may be approached advantageously from the standpoint of the requirements to be met. Typical design requirements include:

Efficiency
Output voltage
Output current
Input impedance
Output impedance
Highest frequency to be amplified
Lowest frequency to be amplified
Allowable harmonic distortion
Allowable frequency distortion
Allowable phase distortion
Allowable noise output
Necessary constancy of gain
Method of gain control
Stability with respect to mechanical shock
Stability with respect to temperature
Power consumption
Size
Weight
Cost

From the above list it is clear that the design of an amplifier is not simply a matter of substituting in a formula. Even when an electrical

design is complete and satisfactory, the engineer has only begun his problem. When the circuit elements and tubes have been chosen and a mechanical design is evolved, then the job may be considered complete.

In amplifiers for low-frequency application it is usually possible to make the mechanical design with little consideration for the schematic diagram, and vice versa. At these frequencies the effect of leads is not vital, and the mechanical design should be for maximum convenience and ease of service. In high-frequency amplifiers, however, such is not the case. The success or failure of a unit depends at least as much upon the mechanical arrangement as upon circuit design. In particular, problems of shielding and of connections to the chassis or panel are of paramount importance. Although no method is known at the present time which will guarantee the success of an amplifier, there are a number of precautions which greatly reduce the hazard of failure. By success is meant an amplifier which meets the design requirements. An amplifier is unsuccessful if it oscillates or is otherwise unable to satisfy the necessary requirements.

13·2 Audiofrequency Voltage Amplifier

Sound waves having frequencies between 20 and 20,000 cycles per second are audible to a majority of human ears. Accordingly, this frequency range has come to be known as the audiofrequency band, and apparatus for transmission of electrical currents of these frequencies is referred to as audiofrequency apparatus. The terms voice frequency and speech frequency are used to designate the same or a somewhat narrower frequency band.

The requirements placed upon an audiofrequency amplifier are seldom severe. Ordinarily the harmonic distortion introduced may be a few per cent of the fundamental signal. Frequency distortion must be controlled but usually may be as much as 10 per cent in the total band, and near the edges of the band is seldom important. Phase distortion may be audibly detected only if present in large amounts and is consequently seldom written as a design requirement. An amplifier satisfying the normal requirements on frequency distortion produces a phase distortion that is negligible from an audio standpoint.

Almost any type of tube will serve as an audiofrequency amplifier. Triodes, tetrodes, and pentodes are all in wide use, although the tetrode is now assuming a relatively insignificant position. The dual tubes consisting of two independent high-mu triodes in a single envelope are extremely convenient for many applications, since they permit the assembly of a high-gain amplifier in a minimum of space.

13·3 Coupling Methods for Audiofrequencies

It is possible by means of rather elaborate voltage supply arrangements to connect the plate of each tube to the grid of the following tube. The resulting system responds to signals of zero frequency. That is, the direct-coupled amplifier is also a direct-current amplifier. Aside from the difficulty of achieving this result it is often an undesirable one. In a high-gain amplifier some small change in the cathode conditions of the first tube may result in such a large effective voltage in that grid circuit as to overload the last tube.

The coupling methods in ordinary use are the transformer, the resistance-capacitance combination, and the choke-capacitance combination. Of these the resistance-capacitance combination is by far the most common, with the transformer coupling still of considerable importance. The so-called impedance coupling using a choke and condenser is largely obsolete.

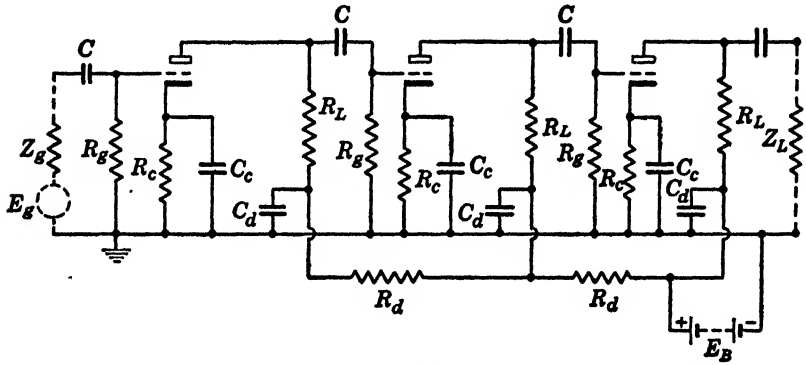


FIG. 13·1 Conventional three-stage resistance coupled amplifier.

C is the plate-grid coupling condenser.
C_c is the cathode by-pass condenser.
C_d is the plate by-pass condenser.
R_c is the cathode bias resistance.
R_d is the plate voltage dropping resistance.
R_g is the grid leak resistance.
R_L is the plate load resistance.

Figure 13·1 shows a conventional three-stage resistance-capacitance coupled audiofrequency amplifier. The operation of the various components of this circuit will be discussed in connection with this schematic diagram. At the frequencies involved the effects of inductance and of interelectrode capacitance are essentially negligible. The plate current in flowing through the cathode resistor R_c makes the cathode more positive than the grid with the result that no electrons

reach the grid. Accordingly, we are justified in calculating the performance of the circuit on the basis of zero grid current. The plate current in turn is controlled by the voltage which exists between grid and cathode.

Over most of the band of frequencies to be amplified, i.e., the midband range, the reactances of all condensers are negligible. That is, from the a-c standpoint all cathodes are grounded, all plate load resistors return to ground, and each plate is connected directly to the following grid.

At high frequencies the amplification falls off from the value at the middle of the band because of the effect of stray capacitances not shown in Fig. 13·1. A capacitance in the order of 50 $\mu\mu$f due to wiring, sockets, tube capacitances, etc., is effectively in shunt with the load resistor R_L. At these higher frequencies the effective load impedance is the parallel combination of these two impedances and falls off approximately inversely with frequency.

At low frequencies the reactance of the coupling condenser is no longer negligible in comparison to the grid leak resistance R_g. Thus only a fraction of the voltage developed at the plate of one tube appears across the grid leak of the following tube. At lower and lower frequencies this voltage-divider action becomes more and more pronounced. Two additional effects are observed. The cathode by-pass condenser C_c no longer has negligible impedance so that the grid-cathode voltage no longer equals the voltage across the grid leak. This corresponds to a reduction in amplification. However, the plate decoupling filter C_d acting in the plate circuit may offer appreciable impedance tending to compensate the other two effects which produced a reduction in gain.

13·4 The Decibel

The performance of amplifiers is most readily stated in terms of the decibel.* Originally the term was reserved as a measure of power ratios. The power at two points in a system is defined to differ by 10 db if the ratio of the two powers is 10. Specifically

$$10 \log_{10} \frac{P_2}{P_1} = N \qquad \text{(db)} \qquad [13\cdot1]$$

The decibel was originally evolved in connection with lines and other constant-resistance networks. In such networks the resistance across which voltage is measured or through which current flows is everywhere

* Pronounced dĕs′ ĭ bĕll, abbreviated db. The decibel is related to the neper by the equation N(db) = 8.7 n (nepers).

the same. In a transmission line, for example, a certain percentage reduction in the current is necessarily associated with an equal percentage reduction in the voltage, and vice versa. Accordingly the power varies as the square of either current or voltage. Under these conditions it is seen that equation 13·1 is equivalent to

$$20 \log_{10} \frac{V_2}{V_1} = N \quad \text{db} \qquad [13\cdot2]$$

and

$$20 \log_{10} \frac{I_2}{I_1} = N \quad \text{db} \qquad [13\cdot3]$$

In the voltage amplifier we are no longer interested in power, and the impedances involved are all high and more or less indefinite. It is convenient, therefore, to use the definition of equation 13·2 without any qualification as to the impedance involved. This definition will be used throughout the remainder of the chapter.

For convenience a brief table of equivalents between voltage ratio and decibels is presented below.

TABLE 13·1

Voltage Ratio	Decibels
1.12	1.0
1.41	3.0
2.0	6.0
5.0	14.0
10.0	20.0

From this table the voltage ratio corresponding to any number of decibels may easily be calculated mentally. If two voltage ratios are to be multiplied, as is necessary to determine the overall amplification of two stages, the decibel figures corresponding are added. Division is accomplished by subtraction of the decibel equivalents. For example, a voltage ratio of 5 is equal to 10/2 and, therefore, corresponds to 20 minus 6 or 14 db. The voltage ratio corresponding to 31 db is evaluated by setting $31 = 20 + 20 - 6 - 3$ db. The corresponding voltage ratio is $\frac{10 \times 10}{2 \times 1.41} = 35$.

It is thus unnecessary to use tables or slide rule for any except the most exact computations. A little practice with the above method makes possible rapid mental calculations which are accurate to $\frac{1}{2}$ db.

The term gain has been used up to this point as being practically synonymous with amplification. Hereafter it will be used to represent the ratio of output to input voltage when expressed in decibels. The

gain is thus zero if the output is equal to the input voltage. If the output is less than the input, the gain is negative; if greater, the gain is positive.

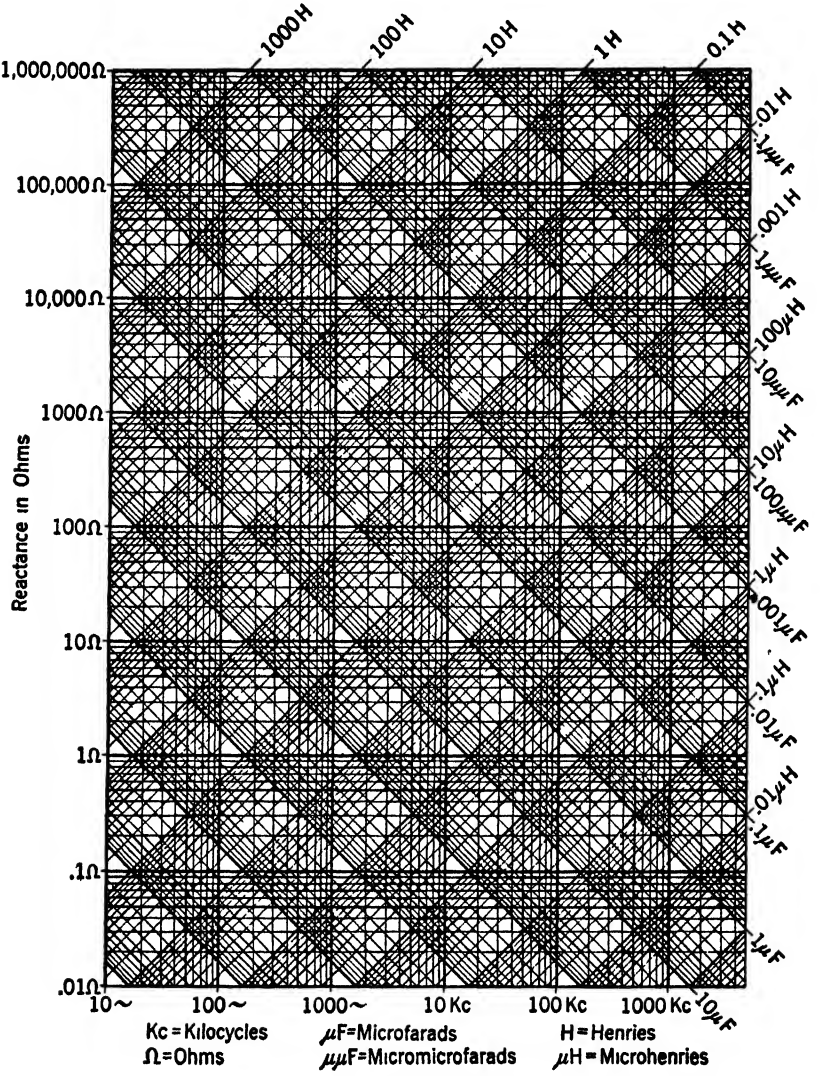


FIG. 13·2 A reactance-frequency chart.

13·5 Reactance Charts

Logarithmic charts of reactance versus frequency have been available for some time. Their use, however, is by no means so general as it

should be. Figure 13·2 shows a typical chart of this type. The abscissa is universally scaled in frequency. The ordinate is usually scaled in ohms reactance, although it is sometimes scaled in micromhos susceptance. The equation $X_L = \omega L$ appears as a straight line on a graph of this kind, the location of the line depending upon the value of inductance L. Similarly, the equation $X_C = 1/\omega C$ appears as a straight line. Since the exponent of ω is $+1$ in the first case and -1 in the second, it follows that the slope of the lines for inductance is $+1$ and the slope of the capacitance lines is -1.

The design of amplifiers is greatly facilitated by the use of such charts. The coupling condenser and grid leak combinations necessary to obtain a prescribed low-frequency response are readily obtained directly from the chart. Similarly, the lowest frequency at which the shunting capacitance seriously decreases the amplification is easily found. Calculation of the other elements C_c and C_d is similarly facilitated.

13·6 Consideration of a Single Stage

Figure 13 3 shows a single-stage amplifier isolated from Fig. 13·1 with the added simplification that C_d is indefinitely large. In practice this result is often achieved by replacing the condenser C_d by a glow discharge tube such as the VR–150. Such tubes act substantially as infinite condensers and so satisfy our requirement. The value of C_c is also assumed to be very large.

Let us develop the gain and phase shift characteristic of the amplifier of Fig. 13 3 at various frequencies.* We shall find it enlightening to extend the results of certain calculations to frequency regions in which they are incorrect. That is, we shall investigate first the several asymptotic relations that exist and shall fill in the details of this characteristic later.

Consider first Fig. 13·3*d*, which holds for the mid-band frequency range. The output voltage is

$$V_2 = I \frac{R_0 R_g}{R_0 + R_g} \qquad [13\text{·}4]$$

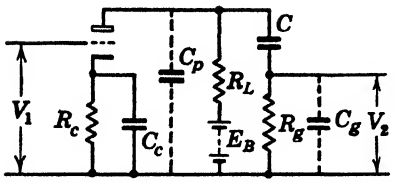
and

$$\frac{V_2}{V_1} = g_m \frac{R_0 R_g}{R_0 + R_g} \qquad [13\text{·}5]$$

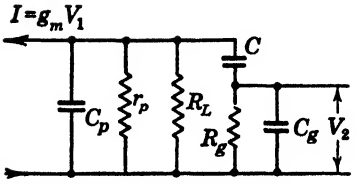
but

$$R_0 = \frac{R_L r_p}{R_L + r_p}$$

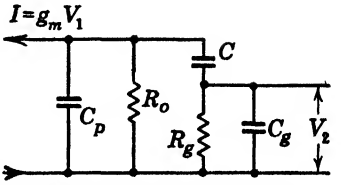
* The equivalent plate circuit theorem as used here is in the form of a constant-current generator $g_m V_1$ in shunt with the dynamic plate resistance r_p.



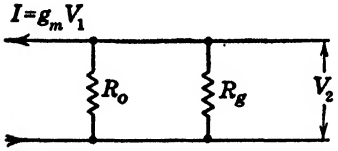
(a) One Stage of Resistance Coupled Amplifier Using a Triode



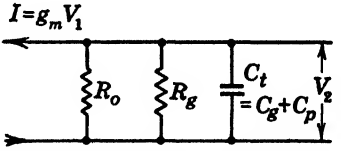
(b) Exact Equivalent Circuit



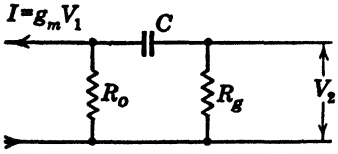
(c) Modified Exact Equivalent Circuit



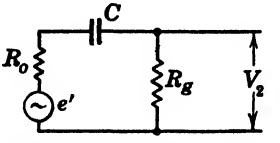
(d) Circuit Correct for Mid-Band Frequencies



(e) Circuit Correct for Upper Edge of Band



(f) Circuit Correct for Lower Edge of Band



(g) Modified Circuit for Low Frequency Calculation

FIG. 13·3 Development of performance of a one-stage resistance-coupled amplifier.

V_1 is input voltage.
V_2 is output voltage.
C is coupling capacitance.
R_g is grid leak resistance.
r_p is plate resistance of the tube.
g_m is transconductance of the tube.
R_L is the plate load resistance.

$$R_0 = \frac{R_L r_p}{R_L + r_p}.$$

C_p is the effective capacitance to ground of plate plus socket and wiring.
C_g is the effective capacitance to ground of grid plus socket and wiring.

so

$$\frac{V_2}{V_1} = g_m \frac{\dfrac{R_L r_p R_g}{R_L + r_p}}{\dfrac{R_L r_p}{R_L + r_p} + R_g} = \frac{g_m R_L r_p R_g}{R_L r_p + R_g r_p + R_L R_g} \qquad [13\text{-}6]$$

Equation 13·6 indicates a gain that is independent of frequency and depends only upon the transconductance of the tube and the various resistances present. Thus, on a gain vs. frequency diagram such as that shown in Fig. 13·4, equation 13·6 appears as a straight line. Gain at higher or lower frequencies will be less than that indicated by equation 13·4. It is convenient to consider the gain at mid-band frequencies

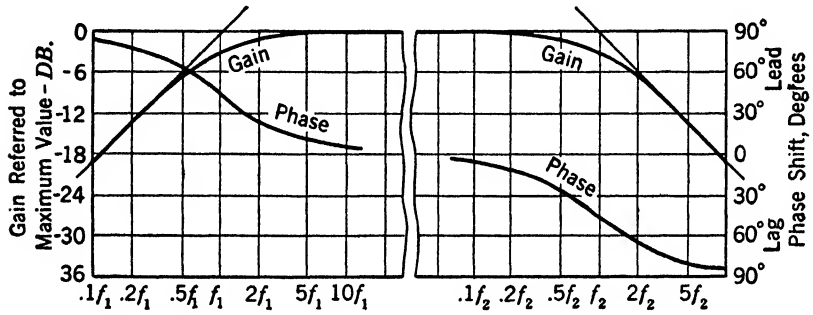


FIG. 13·4 Universal gain and phase characteristic for a one-stage resistance-coupled amplifier.

At the frequency f_1

$$\frac{1}{2\pi f_1 C} = R_0 + R_g$$

At the frequency f_2

$$\frac{1}{2\pi f_2 C_t} = \frac{R_0 R_g}{R_0 + R_g}$$

where $C_t = C_g + C_p$.

as a reference level of zero decibels. Hence equation 13·6 is the horizontal line drawn through zero decibels in Fig. 13·4. The gain at other frequencies outside the mid-band range will then be below this line on the diagram.

Consider now Fig. 13·3*e*, which represents the equivalent circuit which applies at high frequencies. Let

$$R = \frac{R_0 R_g}{R_0 + R_g} \qquad \text{and} \qquad Z = \frac{1}{j\omega C_t}$$

Then

$$V_2 = I\left(\frac{RZ}{R+Z}\right)$$

or

$$\frac{V_2}{V_1} = g_m \frac{RZ}{R+Z} \qquad [13\cdot7]$$

As the frequency is indefinitely increased, the impedance Z of the shunting condenser C_t becomes very low in comparison to the resistance R. Thus, at extremely high frequencies,

$$Z \ll R$$

and equation 13·7 becomes

$$\frac{V_2}{V_1} = g_m Z = \frac{g_m}{j\omega C_t} \qquad [13\cdot7a]$$

Accordingly, the output voltage V_2 varies inversely with frequency. This condition is represented by the line of 45° slope at the right-hand side of Fig. 13·4.

The correctness of this construction is best established by reference to Fig. 13·2. It is seen that the impedance vs. frequency curve of a condenser, plotted on logarithmic paper, is a straight line of slope -1. In Fig. 13·4 a logarithmic frequency scale is used. The ordinate is also logarithmic as shown by equation 13·2. The actual slope of the line in Fig. 13·4 depends upon the scale chosen for ordinates. In the present case 20 db corresponds to a frequency interval of 10: 1. From Table 13·1 it is seen that 20 db corresponds to a 10: 1 ratio of voltage.

At some particular frequency designated f_2 the reactance of the shunting capacitance is equal to the impedance of the various resistances in parallel. This is indicated by the intersection of the line of 45° slope and the horizontal line of equation 13·6. Under this condition we know that the gain is reduced by 3 db from the mid-band value and that the phase shift is 45°.* At higher frequencies the gain curve rapidly approaches the asymptote just drawn, and the phase shift approaches 90°. At lower frequencies the gain curve rapidly approaches the horizontal asymptote and the phase shift approaches 0°.

* If a resistance R and a reactance $X = R$ are connected in parallel, the current produced by a given voltage is $1 + j1$ times as great as that which would flow in the resistance alone. Thus the total current is 45° out of phase with, and is $\sqrt{2}$ times as great as, the current which would flow in the absence of the reactance. But in this case the voltage gain is proportional to the total impedance and therefore is $1/\sqrt{2}$ of its mid-band value. From Table 13·1 it is seen that a gain reduction of 3 db must result.

Consider next Fig. 13·3*f*, which represents the equivalent circuit at low frequencies. Here

$$V_2 = \frac{R_0 R_g}{R_0 + R_g + \dfrac{1}{j\omega C}} I$$

or

$$\frac{V_2}{V_1} = g_m \frac{R_0 R_g}{R_0 + R_g + \dfrac{1}{j\omega C}} \qquad [13\cdot8]$$

As the frequency approaches zero the reactance of the coupling condenser C becomes very high compared to the resistances R_0 and R_g. Thus at very low frequencies $R_0 + R_g \ll \dfrac{1}{j\omega C}$ and equation 13·8 becomes

$$\frac{V_2}{V_1} = g_m R_0 R_g (j\omega C) \qquad [13.8a]$$

These same results may be obtained directly from Fig. 13·3*g* which is derived from Fig. 13·3*f* by application of Thévenin's theorem. This condition is represented by the straight line of 45° slope at the left side of the diagram. A rise of 20 db corresponding to 10: 1 voltage ratio is drawn for a frequency ratio of 10: 1.

At some particular frequency f_1 the reactance of the coupling condenser C is equal to the sum of the resistances R_0 and R_g. It is seen that this is the point at which the slanting asymptote meets the horizontal line. Again the gain is 3 db less than the maximum value, and again the phase shift is 45°. The behavior of the phase curve and the manner in which the gain curve approaches the asymptotes is identical with that at the upper frequency f_2.

The band width of an amplifier or other network is usually stated in terms of the frequencies f_1 and f_2 at which the gain falls or loss rises by 3 db. A familiar and notable example of this is the series-resonant circuit in which the band width is defined as f_0/Q on the basis of a 3 db reduction of current for a fixed value of voltage.

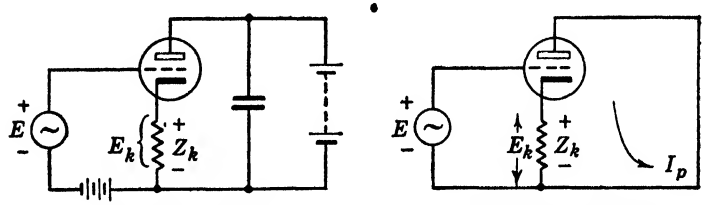
The merit of the construction of Fig. 13·4 is that it is universally applicable to amplifiers in which the upper frequency f_2 exceeds the lower frequency f_1 by a ratio of 100 or more. Even if the ratio is only 10 : 1, the construction applies with very slight correction.

The upper and lower cut-off frequencies are readily calculated by the relations already stated. That is, the upper cut-off frequency f_2 is that one for which the reactance of the total shunting capacitance C_t is

equal to the combined resistance of R_0 and R_g in parallel. The lower cut-off frequency f_1 is that one for which the reactance of the coupling condenser C is equal to the combined resistance of R_0 and R_g in series.

13·7 Effect of the Cathode Condenser

The situation which exists when the reactance of C_c is not negligible is best regarded as simple negative feedback. Consider the circuit of



(a) Actual Circuit (b) Equivalent Circuit

FIG. 13·5 Circuit for analyzing the effective transconductance of a triode. An a-c voltage E is introduced in the grid circuit, and the plate is by-passed to ground.

Fig. 13·5*a*. A triode in which the plate is by-passed to ground is analyzed for its effective transconductance. From Fig. 13·5*b* we may write

$$I_p = g_m(E - E_k)$$

and

$$E_k = I_p Z_k$$

which give

$$I_p = g_m E - g_m I_p Z_k \qquad [13·9]$$

or

$$\frac{I_p}{E} = \frac{g_m}{(1 + g_m Z_k)} = g'_m \qquad [13·10]$$

Thus we see that the effective transconductance of the tube is reduced from its normal value by the ratio $1/(1 + g_m Z_k)$, where g_m is the transconductance and Z_k is the impedance in the cathode circuit. If no by-pass condenser is used so that Z_k is simply the resistance giving suitable self-bias for linear operation, $g_m Z_k$ is often equal to unity. This is true for a large number of tubes, but is not a fundamental relation, and many pronounced exceptions exist. It is, however, a convenient approximation, for under these circumstances equation 13·10 tells us that the effective transconductance is half its normal value when no cathode by-pass condenser is used.

The effective plate resistance is deduced in a similar fashion by reference to Fig. 13·6, in which the grid is assumed short-circuited. Again it is seen that plate current flowing in the cathode impedance results in

a voltage in the grid circuit and so affects the result. The equations applying to Fig. 13·6*b* are

$$I_p = \frac{E}{r_p} - E_k g_m \qquad [13\cdot11]$$

and

$$E_k = I_p Z_k \qquad [13\cdot12]$$

which gives,

$$I_p = \frac{E}{r_p} - I_p Z_k g_m \qquad [13\cdot13]$$

$$\frac{E}{I_p} = r_p(1 + Z_k g_m) = r_p' \qquad [13\cdot14]$$

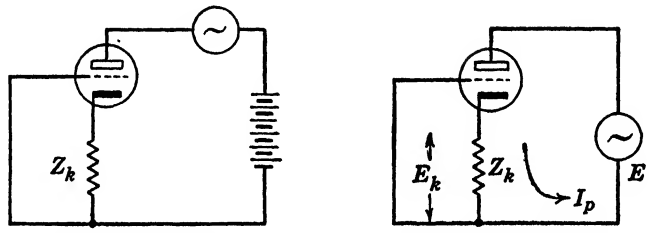


FIG. 13·6 Circuit for analyzing the effective plate resistance of a triode. The grid is short-circuited, and an a-c voltage E is introduced in the plate circuit.

That is, the effective plate resistance is increased in exactly the same ratio that the transconductance is reduced. This leads us to a very interesting general proposition. If the resistances R_L and R_g are large compared to the normal plate resistance r_p of the tube, no significant loss of amplication results from the feedback of the ordinary self-biasing resistance. This is verified by reference to equation 13·6. When $R_L \gg r_p$ and $R_g \gg r_p$, equation 13·6 takes the form

$$\frac{V_2}{V_1} \simeq g_m r_p = g_m' r_p' = \mu$$

In the event that the load resistance R_L is not large enough to satisfy the above condition the gain of the amplifier is reduced by the feedback action. A gain reduction anywhere between 0 and 10 db may result, depending upon the particular conditions that exist.

Since R_c is a relatively low resistance, usually in the order of 1000 ohms, a very large capacitance C_c is required if the reactance is to be small compared to the resistance. Paper condensers are hardly

practical for this application, but modern electrolytic condensers provide the required capacitance at moderate expense and in small size.

The small amount of negative feedback which results from omission of the cathode condenser serves somewhat to stabilize the amplifier and to improve the linearity. When it is permissible to tolerate the relatively small reduction of gain, the omission is to be recommended. The weight, bulk, and expense of the amplifier are reduced and the possibility of failure of the condenser is removed.

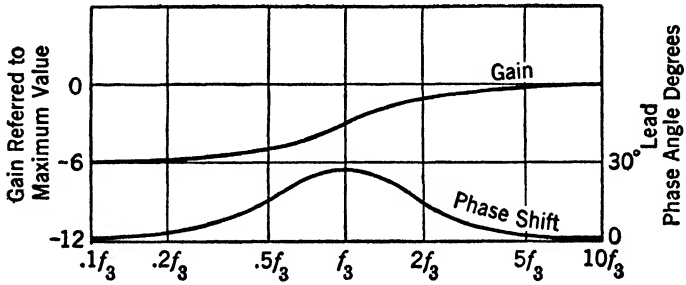


FIG. 13·7 Gain and phase characteristics due to a cathode by-pass condenser.

Universal characteristic applying to feedback produced by the cathode resistor R_c and the condenser C_c. This characteristic is directly additive to Fig. 13·4. At the frequency f_3,

$$\frac{1}{2\pi f_3 C_c} \cong R_c$$

When a cathode by-pass condenser is used there is, over some band of frequencies, a transition region in which the amplification is reduced and a phase shift is introduced. At higher frequencies the reactance of C_c is negligible, and the full transconductance of the tube is effective. At lower frequencies the admittance of C_c is negligible and the gain is constant at the reduced value, which depends upon the cathode resistor. In many cases this transition region may be made lower in frequency than the low-frequency cut-off, f_1, shown in Fig. 13·4. In any event the effect illustrated in Fig. 13·7 is applicable to Fig. 13·4 as a direct addition.

The magnitude of the phase shift is proportional to the magnitude of the gain change. The values shown in Fig. 13·7 are correct for the special case of 6 db change. The frequency f_3 is practically equal to that at which the reactance of the condenser C_c is equal to the associated resistance R_c.

13·8 Effect of the Plate Condenser C_d

The plate-circuit condenser C_d serves two functions which are distinct although related. Its primary function is to return the alternating component of the plate current to the common ground through a short and direct path. In this application it need only have an impedance low compared to that of the B supply, for all frequencies to be amplified. If the impedance of C_d is not low, appreciable currents flow in the B supply and other associated circuits. See Fig. 13·1.

The second function of the plate by-pass condenser is as a decoupling element. That is, the resistors R_d and the condensers C_d serve as a kind of low-pass filter to prevent coupling of the output and input stages through the B battery supply. It is common to specify these elements purely by rule of thumb. This procedure is unwise and entirely unnecessary since reasonably accurate calculations of the required values may be made without difficulty.

Let us examine the circuit of Fig. 13·1. There is a tendency for part of the alternating current in the plate circuit of the last tube to flow through the two decoupling resistors R_d, producing a voltage in the grid circuit of the second tube. This voltage is amplified and, if it is of appreciable magnitude, produces in the last tube a plate current comparable to that originally present. If the effect is large enough the last two tubes may act together as an oscillator. Even if the coupling is reduced considerably from the value which causes oscillation it may contribute undesirable undulations or ripples in the overall curve of gain *vs.* frequency.

In Fig. 13·1 let us assume that the cathode by-pass condensers are fully effective and that the B supply has a high internal impedance. Let each stage have a voltage amplification ratio equal to N. Accordingly some voltage v impressed upon the grid of the second tube results in a voltage Nv upon the grid of the last tube and a voltage N^2v across the final load resistor R_L. The resulting current in the final load resistor is therefore N^2v/R_L. This is the current delivered to the decoupling filter of Fig. 13·1 which is drawn more simply in Fig. 13·8.

To avoid the possibility of oscillation it is necessary that v', the output of this equivalent filter, be less than v, the assumed voltage. Safe design calls for a value of v' some ten times smaller than v, and conservative design calls for a reduction in the order of 100 : 1.

Let us calculate the voltage attenuation of this network on the assumption that X, the reactance of each condenser, is small in comparison with the associated resistance R_d. This is a necessary condition if the loss is to be sufficiently high, and it greatly facilitates the calculation which may be made on the basis of a sequence of voltage dividers.

Here, as often happens, it is simplest to develop the input voltage required. We find

$$V_1 = \frac{R_L + r_p}{r_p} v' \qquad [13\cdot15]$$

Since $R_d \gg X$

$$V_2 = \frac{R_d}{X} V_1 \qquad [13\cdot16]$$

$$V_3 = \frac{R_d}{X} V_2 \qquad [13\cdot17]$$

$$V_3 = \frac{vN^2X}{R_L} \qquad [13\cdot18]$$

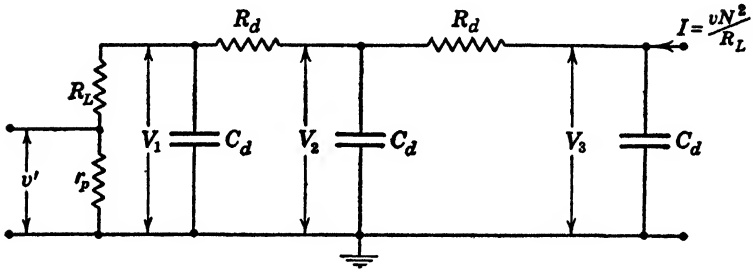


FIG. 13·8 Elements effective in the decoupling network of the B supply in Fig. 13·1. It is assumed that the cathode by-pass condensers are fully effective and that the internal resistance of the B supply is infinite. A current $\frac{vN^2}{R_L}$ flows through the plate load resistance R_L of the last tube of Fig. 13·1 and delivers a voltage to the plate load resistance R_L of the first tube through the decoupling filter above. The effective admittance of the second tube in comparison to that of C_d is neglected.

Combining the first three and equating to the fourth,

$$\frac{R_d^2}{X^2} \cdot \frac{R_L + r_p}{r_p} v' = \frac{vN^2X}{R_L} \qquad [13\cdot19]$$

Solving

$$v = \frac{1}{N^2} \cdot \frac{R_d^2 R_L}{X^3} \cdot \frac{R_L + r_p}{r_p} v' \qquad [13\cdot20]$$

For conservative design we set $v' = v/100$, or

$$v = 100v' \qquad [13\cdot21]$$

giving

$$100 = \frac{1}{N^2} \cdot \frac{R_d^2 R_L}{X^3} \cdot \frac{R_L + r_p}{r_p} \qquad [13\cdot22]$$

For a reasonable practical case using the triode 6F5 we have

$$N = 50, \quad R_L = 250{,}000 \text{ ohms}, \quad r_p = 50{,}000 \text{ ohms}$$

Let us set $R_d = 25{,}000$ ohms, a reasonable value here. Then equation 13·22 becomes

$$100 = \frac{1}{2500} \cdot \frac{25^2 \cdot 10^6 \cdot 25 \cdot 10^4}{X^3} \cdot \frac{3 \cdot 10^5}{5 \cdot 10^4} \qquad [13\cdot23]$$

$$X^3 = 3.75 \times 10^9 \qquad [13\cdot24]$$

$$X = 1.6 \times 10^3 \qquad [13\cdot25]$$

That is, if the reactance of the condensers C_d is less than 1600 ohms, the loss of the filter is such that the voltage v' returned to the grid of the second tube is less than 1 per cent of the voltage assumed at that point. This condition is readily maintained to quite low frequencies. At such frequencies the situation is relieved by the action of the coupling condensers C which reduces the gain N. The magnitude of this gain reduction may be calculated by reference to Figs. 13·4 and 13·7.

In amplifiers employing many stages, the filter network is sometimes more complicated, but the manner of making the calculations is the same. The importance of such calculations in amplifier design can scarcely be overemphasized.

13·9 Variation of the Shunting Capacitance

The total shunting capacitance C_t used in Fig. 13·3*e* is composed of the plate-cathode capacitance of one tube plus the effective grid-cathode capacitance of the following tube. The effective plate-cathode capacitance is usually small and is closely equal to the value measured with the cathode cold. The effective grid-cathode capacitance, as shown in Chapter 12, is considerably greater than that observed on the cold tube. At low frequencies the grid-cathode capacitance is modified by the action of the grid-plate capacitance. This action may readily be explained in terms of the example just discussed. If the grid of the last tube in Fig. 13·1 is made more positive by 1 volt the plate is driven negative by 50 volts. Thus the potential difference of 50 + 1 = 51 volts exists across the grid-plate capacitance. A charging current $N + 1$ times as great as that required in the cold tube is thus required when the tube operates as an amplifier.. Thus the effective input capacitance = $(N + 1)C_{gf} \simeq NC_{gf}$ for large values of N. In the 6F5 tube the gride-plate capacitance C_{gf} is 2.5 $\mu\mu f$. With $N = 50$ the effective input capacitance is approximately $NC_{gf} = 125$ $\mu\mu$f.

Near the upper edge of the frequency band this effect is complicated

by the fact that the plate load on each tube becomes reactive and the input impedance of each tube becomes complex.* In general no tendency to oscillate exists, and the frequency at which this effect occurs is high enough so as to be well above the audio band. In amplifiers designed to amplify wider frequency bands it is the universal practice to employ tetrodes or pentodes. In these tubes the grid-plate capacitance is very low, and the input capacitance is therefore not affected in this way.

13·10 The Video Amplifier

In current television practice the so-called frame-repetition rate is 30 cycles per second. That is, the entire scene is scanned 30 complete times in each second. Accordingly 30 cycles per second is the lowest frequency of importance in the video signal. The resolution of the system continues to improve as the upper frequency of the signal is increased without limit. Very satisfactory images are obtained, however, if frequencies up to 3 or 4 megacycles are transmitted. Thus we require that the entire band of frequencies from 30 cycles to 4 megacycles per second be transmitted equally well. In particular, the gain frequency characteristic of the video amplifier must be flat over this range.

In practical amplifiers consisting of tubes coupled by complex networks of resistances, inductances, and capacitances there is a finite time interval between application of a signal to the input and its appearance at the output. A complex wave such as those typical of television systems becomes distorted if the various frequency components which constitute the wave are delayed by unequal amounts. This phenomenon is well known as phase distortion and has already been discussed as trivial in audiofrequency amplifiers. In video amplifiers, however, it is of the greatest importance. The effect is readily shown by inspection of Fig. 13·9.

The square wave of input voltage V is typical of television and certain other signals. In Fig. 13·9*b* this square wave is analyzed into its fundamental and a few harmonic components. It may be shown that only odd harmonics exist, that all cross the axis at a common point, and that the amplitude of each harmonic is inversely proportional to its frequency.

Let us determine the effect of the circuit of Fig. 13·9*a* upon the square wave. This is most conveniently accomplished by determining the transmission of each separate component of the input wave, and then

* For a very complete discussion of this topic see E. L. Chaffee, *Theory of Thermionic Tubes*, p. 261 ff., 1933, McGraw-Hill Book Company, New York.

adding the resultant components to secure the total output wave. Significant results are obtained if the fundamental frequency is such that the reactance of C is approximately 20 per cent of R. The reactance of C to the third harmonic wave is then only 7 per cent of R. The reactance of C to higher harmonics is negligible. Harmonics higher than the fifth are omitted from the figure for clarity. The gain reduction and phase shift applying to each component of the wave are determined by reference to Fig. 13·4. It is seen that appreciable phase shifts exist at frequencies where the gain reduction is negligible.

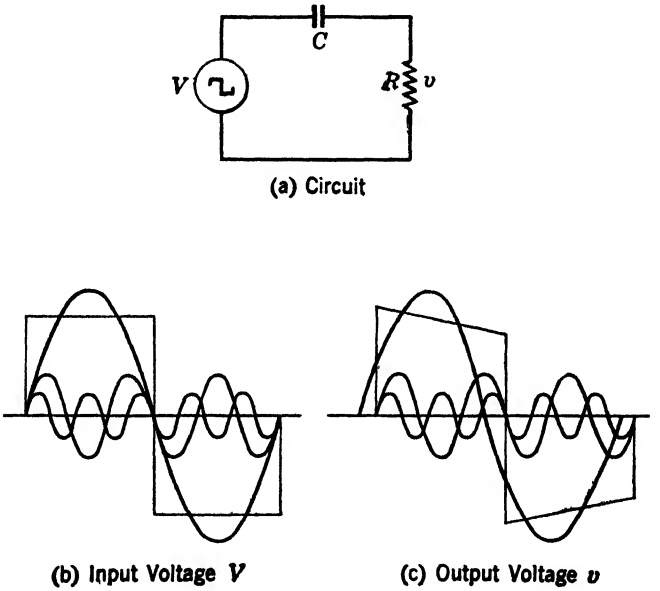


FIG. 13·9 Phase distortion resulting from shift of fundamental with respect to harmonic frequencies. Seventh and higher order harmonics are not shown.

In Fig. 13·9c the various component waves are plotted in approximately their correct relative position, and the resultant wave is derived by summation. It is seen that the primary effect of a small amount of phase shift is to tip the top of the wave. In many applications the relatively small amount of distortion indicated is intolerable.

It is clear that no phase distortion will result if all components of the wave are delayed by an equal amount, since this merely displaces the wave by a finite lateral distance along the time scale. A time delay that is independent of frequency requires a phase shift that is directly proportional to frequency. This proposition may be proved rigorously but is most easily visualized by inspection of Fig. 13·9*b*. If the funda-

mental is shifted by one full cycle, the third harmonic by three cycles, the fifth by five, etc., it is seen that we merely displace the wave by the time required for a cycle of the square wave. Consequently we add the requirement that the phase shift must be proportional to frequency over the same range of 30 cycles to 4 megacycles, in which we have already specified that the gain must be constant. Of course no practical amplifier meets either requirement perfectly but in many amplifiers the deviations from this ideal are quite small.

Power output, as such, is not required from video amplifiers. The output stage delivers a voltage to the intensity-controlling electrode of the kineoscope, a special form of cathode-ray tube. Amplifiers which differ from video amplifiers in no important detail of design or construction are now being widely used to amplify signals to the deflection plates of cathode-ray tubes for measurement and research applications. In such amplifiers it is necessary to deliver outputs in the order of 100 volts. At the high frequencies involved the impedance of irreducible shunting capacitances is quite low, and it is therefore necessary to employ tubes having large current ratings in order to develop the required voltage.

The requirements on harmonic distortion are not severe. In general the output stage is most likely to cause distortion although the stage immediately preceding the output may give trouble if the output tube requires a relatively large voltage to excite its grid. In general a distortion of less than 2 per cent is a desirable goal.

13·11 General Considerations

The requirement that the gain be constant down to 30 cycles per second is readily met by most audiofrequency amplifiers. The accompanying requirement that the phase shift be linear with frequency is not so easily attained. In general some form of phase compensation is employed to reduce the phase distortion in the low-frequency region. In amplifiers for oscilloscope applications this requirement is often even more severe than in video amplifiers and carefully designed circuits are sometimes necessary.

The requirement that the amplification be constant up to frequencies in the region of 4 megacycles is not approached by any audio amplifier. Indeed, this extension of the upper frequency limit by a factor of 100 seems almost insurmountable at first glance. Fortunately the combination of improved tubes and improved circuit design has rendered the goal relatively easily reached and has made possible the construction of special amplifiers having much wider frequency bands.

Since triodes are unsuitable for this sort of work owing to the effect

of the grid-plate capacitance, and since tetrodes are becoming relatively obsolete, the following discussion will be confined to pentode types.

13·12 Consideration of Typical Circuit

The relatively complex circuit of Fig. 13·10 may be arrived at as a logical result of the requirements placed upon the video amplifier. Careful comparison will show that it differs from Fig. 13·1 in only a few important details. The suppressor grid, characteristic of the pentode, is

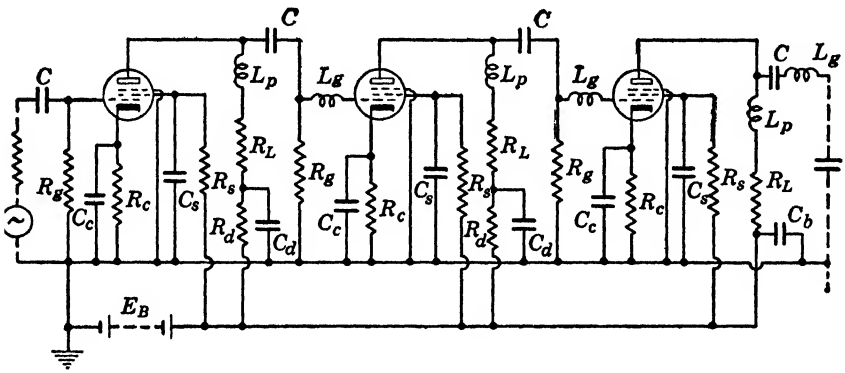


FIG. 13·10 Typical three-stage video amplifier employing series and shunt compensation for high frequencies and conventional low-frequency compensation.

grounded in all tubes. Essentially the same results are obtained if the suppressor is connected to the cathode, but experience has shown that it is more convenient to use the direct ground connection and that certain parasitic effects, sometimes troublesome, are thus avoided. The dropping resistor R_s reduces the screen-grid voltage from that supplied to a value suitable for the tube in question. In some applications it is practical to make the supply voltage appropriate for direct application to the screen grids, but this is seldom desirable since the presence of R_s tends to stabilize the current drawn by each tube. The condenser C_s serves to return the alternating component of the screen current to the common ground. In conjunction with the dropping resistors R_s the screen condensers form an auxiliary decoupling filter network which prevents undesired coupling between stages through the B supply and the screen grids. This is an important function, for otherwise the screen grid of the first tube may readily serve as a control grid causing oscillation or undesired feedback from the plate circuits of the following tubes.

The elements so far discussed are characteristic of pentode amplifier

circuits and are not identified explicitly with the video amplifier. The elements which are truly characteristic of the video amplifier are the added coils L_p and L_g. These coils are approximately in resonance with the shunt capacitances C_p and C_g at the upper frequency limit of the amplifier. By careful control of these resonances it is possible to hold the amplification constant to a considerably higher frequency than would be obtained otherwise. The elements R_d and C_d serve again as a decoupling filter. In this case they must also serve to improve the constancy of gain at low frequencies. Because each decoupling combination serves to compensate for the effect of the following grid leak the compensating network is omitted in the plate circuit of the last tube. Instead a relatively large condenser C_b is used, which may serve to satisfy the requirements on overall decoupling even if R_d and C_d are modified by the requirements for low-frequency compensation.

13·13 Analysis of a Typical Circuit

The behavior of the circuit of Fig. 13·10 is analyzed on the basis of a single stage exactly as was that of the audio amplifier. Again it is practical to divide the analysis into three distinct parts. A very wide band of frequencies exists over which the gain is constant at the value determined by the transconductance and the various resistances of the circuit. At low frequencies the several condensers begin to offer appreciable reactances tending to reduce the amplification. At high frequencies the shunting capacitances become effective and the compensating inductances operate to compensate their effect over a considerable region. These various conditions are shown in Fig. 13·11, analogous to Fig. 13·3.

The gain over the broad mid-band region is readily calculated. To achieve the desired low-frequency behavior R_g is always high, usually greater than 500,000 ohms. The plate resistance of a typical pentode is also high, in the order of a megohm. To achieve the desired constancy of gain to high frequencies the value of R_L is always held low, in the order of 3000 ohms. Accordingly, it is seen that a negligible error results if the mid-band amplification is calculated only on the basis of the load resistor itself. Thus from Fig. 13·11*b*, neglecting r_p and R_g, we have

$$I = g_m V_1 \qquad [13·26]$$

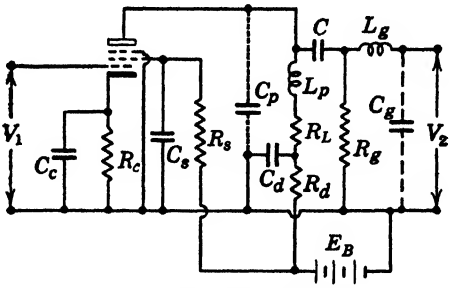
and

$$V_2 = IR_L \qquad [13·27]$$

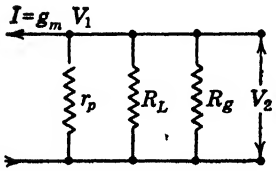
giving

$$\frac{V_2}{V_1} = g_m R_L \qquad [13·28]$$

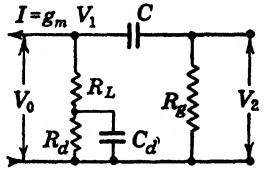
Assuming that R_L may not be raised, it is seen that a high value of transconductance g_m is desirable in order to achieve a high gain per stage.



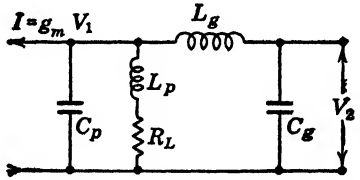
(a) One Stage Video Amplifier



(b) Equivalent Circuit for Mid-Band Frequencies



(c) Equivalent Circuit for Low Frequencies



(d) Equivalent Circuit for High Frequencies

FIG. 13·11 Analysis of one-stage video amplifier neglecting effects of cathode and screen by-pass condensers.

13·14 Low-Frequency Performance

The behavior at low frequencies may be explained on the approximate basis that $R_d \gg R_L$. As the frequency is made lower and lower the reactance of C becomes appreciable with respect to R_g. Accordingly, only a portion of the voltage developed in the plate circuit appears as output voltage V_2. Now if C_d has the proper size its impedance, although very small in comparison to R_d, is appreciable compared to R_L, and the voltage developed in the plate circuit is increased. Exact compensation occurs if $R_L C_d = R_g C$. In practice, of course, R_d is not infinite and the amplification does finally fall off. It is quite practicable,

however, to extend the low-frequency cut-off by a factor of 10 in this way.

The mathematical proof is greatly facilitated by the assumption that $R_d \gg R_L$, which is a legitimate approximation. Moreover, we know that $R_g \gg R_L$ so we may write in terms of Fig. 13·11*c*

$$V_0 = g_m V_1 \left(R_L + \frac{1}{j\omega C_d} \right) \qquad [13\cdot29]$$

and

$$V_2 = \frac{R_g}{R_g + \dfrac{1}{j\omega C}} V_0 \qquad [13\cdot30]$$

giving

$$V_2 = \frac{j\omega C R_g}{1 + j\omega C R_g} \cdot g_m V_1 \cdot \frac{1 + j\omega C_d R_L}{j\omega C_d} \qquad [13\cdot31]$$

Introducing

$$R_g C = R_L C_d \qquad [13\cdot32]$$

for equality of the time constants in the plate and grid circuits

$$\frac{V_2}{V_1} = g_m \frac{j\omega R_L C_d}{1 + j\omega C_d R_L} \cdot \frac{1 + j\omega C_d R_L}{j\omega C_d} = g_m R_L \qquad [13\cdot33]$$

That is, the gain and phase are absolutely independent of frequency under the assumption that R_d is very large. In practice the gain and phase behave exactly in the fashion shown in Fig. 13·4, where f_1 is that frequency for which the reactance of C_d is equal to R_d. Thus the dropping resistor R_d provides a nearly perfect compensation for the loss due to the coupling condenser. If the condenser C_d is made larger or smaller than the value defined above, the compensation is imperfect. If C_d is larger than $(R_g/R_L)\ C$, the low-frequency amplification is somewhat less than that existing across the main band. The behavior is very similar to, but not identical with, that shown in Fig. 13·7. Similarly, if C_d is smaller than $(R_g/R_L)\ C$, the low-frequency gain is above the mid-band value. This suggests the possibility of compensating the effect of the cathode by-pass condenser as well as that of the coupling condenser by a proper choice of C_d and C_c. The necessary conditions are shown in Fig. 13·12.

The compensation illustrated in Fig. 13·12 is best explained in terms of a design procedure. The following is one of the many possible procedures. Choose R_g as the maximum resistance recommended by the tube manufacturer. Choose C as large as is consistent with physical

considerations or d-c leakage requirements. These values determine f_1 as explained in section 13·6. The magnitude of the loss due to the action of R_c at very low frequencies is calculated from equation 13·10 which applies accurately because $R_L \ll r_p$. The decoupling condenser is next chosen from the relationship

$$R_g C = (1 + g_m R_c) R_L C_d$$

in terms of Fig. 13·10 where it is assumed that R_c and R_L have already been fixed by other requirements. This relation is such that the sloping asymptotes to "gain due to action of C_d" and to "loss due to action of

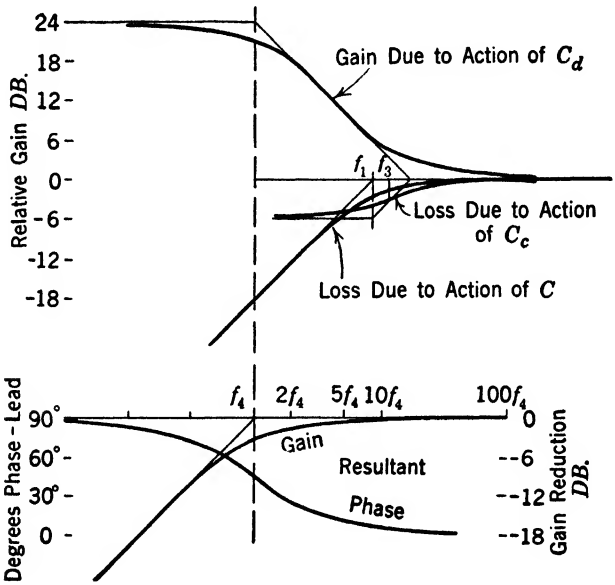


FIG. 13·12 Compensation in plate circuit for effects of coupling and of cathode condensers.

Summation of gain characteristics to produce characteristic shown in lower part of figure. Frequency f_1 is determined by the grid leak, R_g, and coupling condenser C. Frequency f_3 is determined by the cathode resistor R_c and condenser C_c. Frequency f_4 is determined by the decoupling resistor R_d and condenser C_d.

C" differ by the amount of the horizontal asymptote to "loss due to action of C_c," all referring to Fig. 13·12. Finally C_c is so chosen that $\frac{1}{2\pi f_3 C_c} = R_c$ where f_3 has the mean value indicated in the figure.

Under the above conditions the gain is independent of frequency to some lower frequency at which the reactance of C_d becomes comparable to R_d. The curve of "gain due to the action of C_d" then

ceases to rise and the resultant gain falls because of the continued action of the coupling condenser C. The frequency f_4 at which this final drop in gain occurs is defined approximately by the relation $\frac{1}{2\pi f_4 C_d} = R_d$ if $R_d \gg R_L$.

It is still necessary to examine the behavior of the screen by-pass condenser C_s. If C_s is removed it will be found that a rather large loss, in the order of 10 db, results. The exact value depends upon the value of R_s and upon the properties of the tube. The effect may be explained in two steps. The alternating screen current in flowing through the resistor R_s produces an alternating voltage at the screen grid. This alternating voltage acts through the effective screen grid-to-plate transconductance to decrease the alternating plate current.

Because the impedance of a typical screen-grid circuit is relatively high, a condenser of moderate size serves as an adequate by-pass at the lowest frequencies which are ordinarily of interest. In view of the difficulty in compensating the loss due to imperfect by-passing in the screen circuit and the fact that adequate by-pass condensers are not unduly large or expensive it is usually expedient to provide an adequate capacitance.

13·15 High-Frequency Performance

Inspection of Fig. 13·11*d* reveals that the behavior is not simple. The circuit shown may be explained as a parallel resonant circuit of low Q, shunted by a series-resonant voltage divider. Since the circuit shown is complex it will be most convenient to analyze several simpler circuits, thereby developing the one shown in Figs. 13·10 and 13·11.

Let us refer back to Fig. 13·4, in which the upper cut-off frequency was defined as that for which the reactance of the total shunting capacitance equals the load resistance. Evidently this frequency may be raised either by decreasing the shunting capacitance or by decreasing the load resistance. In practice both of these steps are taken. The capacitance is reduced as far as possible by the choice of tubes having low input and output capacitances and by careful wiring and assembly. The plate load resistance is then lowered as far as is necessary to achieve the desired cut-off frequency.

Everything else being fixed the amplification of each stage is directly proportional to the transconductance of the tube. Since we have already shown that the allowable load resistance is inversely proportional to the total capacitance it follows that the most desirable tube for such circuits is one having the highest possible ratio of transconductance to capacitance. If, however, the tube capacitances are small, the effect

of socket and wiring capacitances must be considered in order to obtain a true comparison of the merits of two given tubes.

The input (grid-to-cathode) capacitance of vacuum tubes is usually greater when the cathode is normally heated than when it is cold. This increase is due to the action of space charge which reduces the effective spacing from grid to cathode. The effective capacitance, although affected by the conditions of space charge, is essentially independent of frequency.

It should be emphasized that the only practical limitation on the upper frequency which may be amplified by a video amplifier is set by the shunting capacitances. If it were somehow possible either to remove or to annul these capacitances it would be possible to obtain a very large amplification per stage and to extend the amplification to frequencies at which lead inductance or transit time effects become important. The following material deals entirely with ways of minimizing the effect of this shunting capacitance and concludes in a statement of the best results that may be obtained.

13·16 Shunt Compensation

The simplest form of compensation, and one that has many good features, is illustrated in Fig. 13·13. It is seen that the interstage network differs from the conventional resistance-coupled interstage only in that a coil L has been added in series with the plate load resistance. From the equivalent circuit of Fig. 13·13*b* it is seen that the output voltage and therefore the amplification of the stage is directly proportional to the impedance presented by the combination shown. Figure 13·14 shows the impedance characteristics for several values of R when $L = 0$ and $C = 40\mu\mu f$, a reasonable value. Figure 13·15 shows the impedance characteristics for several values of L where $R = 2000$ ohms and $C = 40\ \mu\mu f$.

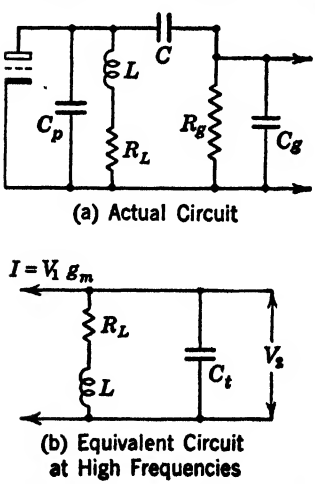


FIG. 13·13 Shunt compensated video interstage network. $C_t = C_g + C_p$.

It is seen that a very great improvement in the constancy of Z and therefore in the constancy of gain results from the addition of a moderate inductance. Too small an inductance gives much the same result as if none were used. Too large an inductance produces a pronounced peak in the characteristic. The curves of Fig. 13·15, reduced to a universal

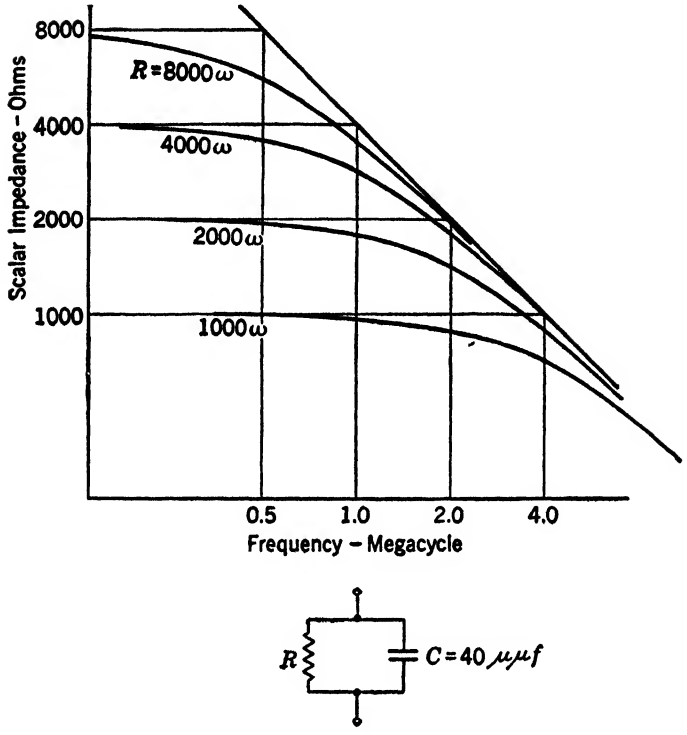


FIG. 13·14 Impedance of a parallel circuit of resistance and capacitance.

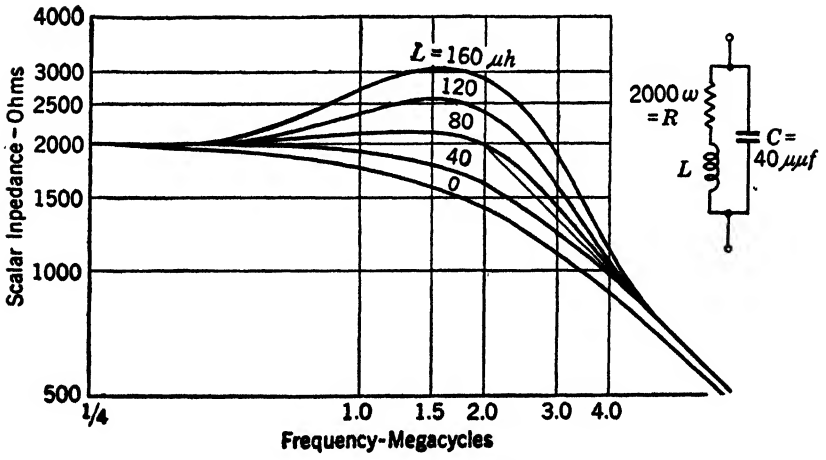


FIG. 13·15 Impedance of a parallel circuit.

basis by an appropriate choice of coefficients, are plotted in Fig. 13·16. The phase shift associated is plotted in terms of the same coefficients so that design is reduced to a relatively simple basis.

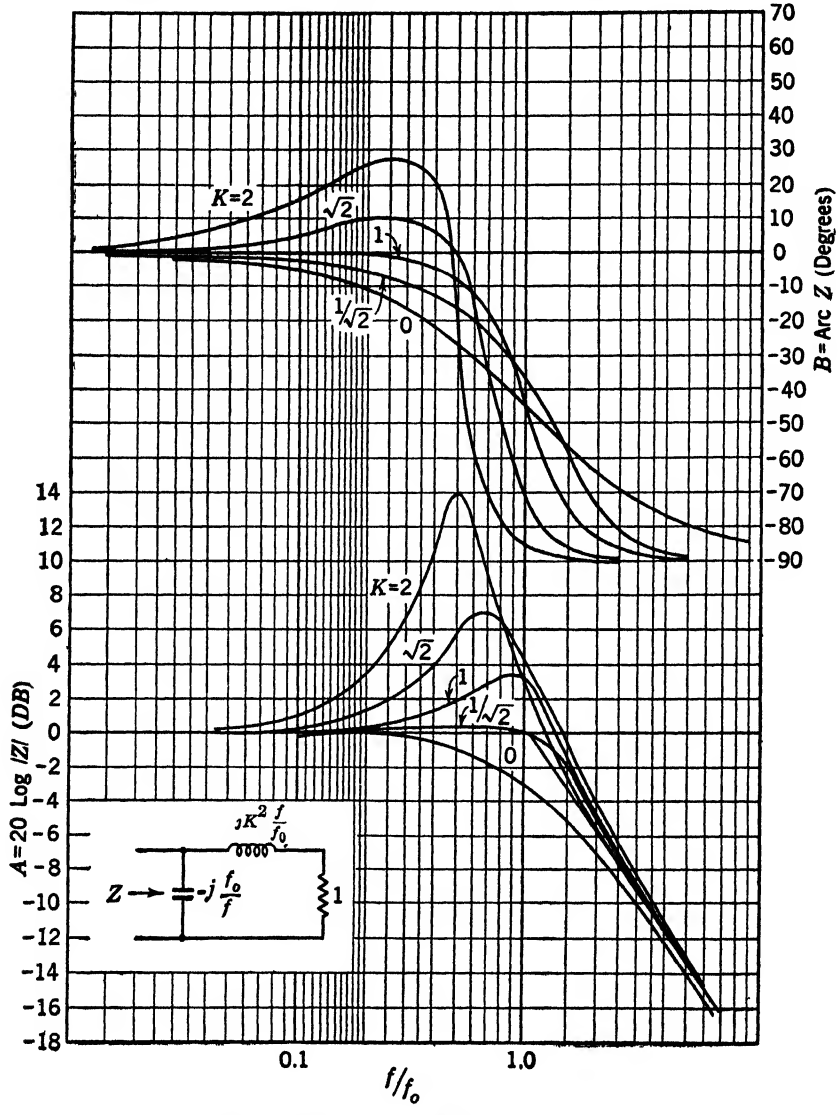


FIG. 13·16 Universal curves for shunt compensation.

The curves of Fig. 13·16 are very convenient for the rapid design of simple video interstage networks. They are universal in the sense that

all the parameters are expressed in terms of a critical frequency f_0, and unit resistance. It should be noted that f_0, as used here, is equivalent to f_2 in Fig. 13·4 and associated discussion.

If each impedance in Fig. 13·16 is multiplied by some factor R, the impedance of the complete network will change by the same ratio, but the phase angle will not change, and the variation of impedance with frequency will not change. Using this fact, and comparing Figs. 13·16 and 13·13*b*, it is seen that

$$R_L = R$$

$$j\omega L = jRK^2\frac{f}{f_0}$$

$$-\frac{j}{\omega C_t} = -jR\frac{f}{f_0}$$

These equations when rearranged give

$$L = R^2K^2C_t$$

and

$$C_t = \frac{1}{2\pi f_0 R}$$

The critical frequency, f_0, is defined by the highest frequency which must be transmitted and the shape of the transmission curve desired. The shunting capacity C_t is fixed by the tubes and assembly. The value of K is chosen to produce the desired shape of transmission characteristic. Finally the load resistance R_L and the compensating inductance L are successively chosen.

13·17 Modified Shunt Compensation

Careful inspection of Fig. 13·16 shows that the value of inductance which gives best compensation below f_0 is considerably smaller than the one which gives best compensation above f_0. Therefore, we could improve the performance still more if we had an inductance whose value increased with increase of frequency. Such inductances are not readily obtained. However, a parallel resonant circuit of high Q behaves at frequencies below its resonance in much this way. That is, the reactance of a parallel LC combination increases with frequency at a rate that is higher than the first power.

Figure 13·17 shows an interstage network using this principle and a typical response curve. It is seen that the gain curve is very flat to a frequency about 1.4 f_0. At higher frequencies the gain falls off rapidly to the value given by the capacitance alone.

We have seen that the performance of the interstage network has been improved by the successive addition of two compensating elements. The question immediately arises whether or not this process may be continued indefinitely. Wheeler* has considered the problem from the

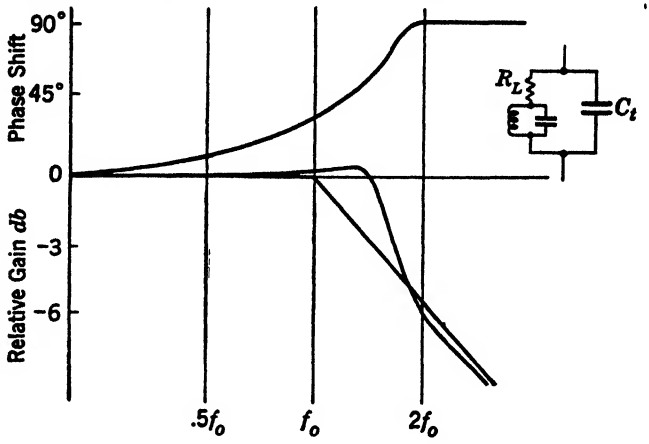


FIG. 13·17 Modified shunt compensation.

standpoint of normal filter theory; he concludes that a very definite limit does exist and that the network of Fig. 13·17 approaches it fairly closely. The limiting curve is shown to be exactly the attenuation characteristic of a prototype filter and is associated with a phase shift that is 90° to all frequencies outside the trasmission band. The situation is best explained by the comparative curves of Fig. 13·18. The values of the parameters plotted were necessarily chosen somewhat arbitrarily but are satisfactory in that they show typical behavior. It is seen that the relatively simple networks III and IV approach the ideal curve very closely indeed. Since the ideal curve requires an infinite number of elements for its construction it is clear that the principle of diminishing returns has begun to apply. The values in Fig. 13·18 refer to the impedance of the element in question in comparison to the load resistance R_L for the particular frequency f_0. From these constants it is relatively easy to deduce numerical designs as the occasion demands.

The circuits of Fig. 13·18 are drawn in such a way as to be recognizable as half sections of several low-pass filters of varying degrees of complexity.

The interstage networks so far discussed are referred to as two-terminal networks because the entire performance of the stage is charac-

* H. A. Wheeler, "Wide Band Amplifiers for Television," *Proc. IRE*, 27, 429, July, 1939.

terized by the impedance as measured between two points. Such networks are characterized by the fact that the phase shift cannot exceed 90° and that the impedance always approaches the impedance of the shunting condenser alone as the frequency is raised.

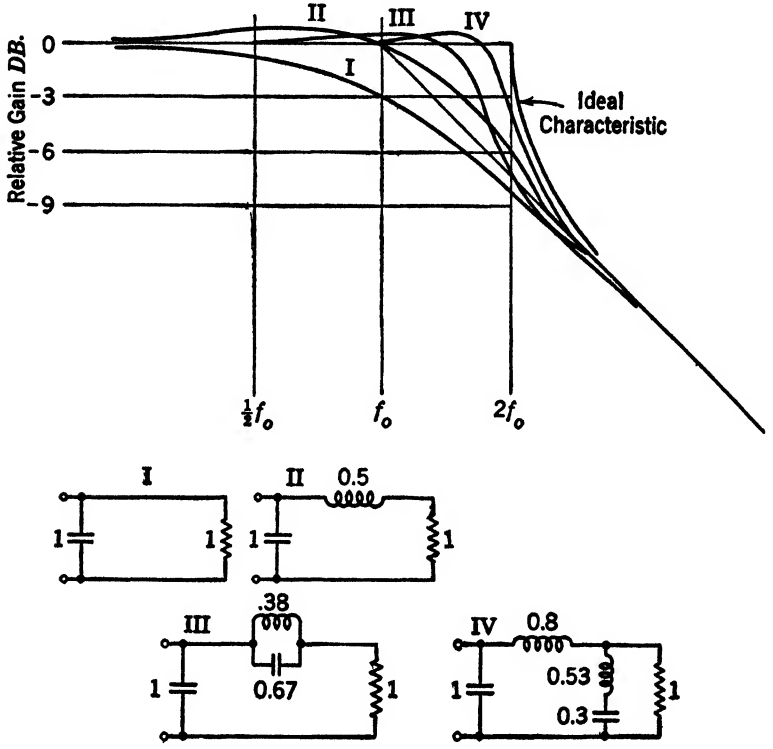


FIG. 13·18 Comparison of several two-terminal interstage networks.

13·18 Four-Terminal Interstage Networks

It will be recalled that the interstage network represented in Fig. 13·11*d* is of such a nature that the capacitances C_p and C_g are separated by the inductance L_g. Such a design is the logical outcome of a consideration of the two-terminal network. It was observed that the frequency band could be doubled if the shunting capacitance were cut in half. Since the capacitances C_p and C_g are comparable if not actually equal, it appears possible to deal with them separately by use of the coil L_g. This procedure is sometimes referred to as *splitting* the capacitances.

The simplest possible four-terminal interstage network is shown in Fig. 13·19. It is seen to resemble Fig. 13·13 but to differ in that the coil is in series with the blocking condenser rather than with the load

resistance. The calculation of the performance is by no means so simple as is that for Fig. 13·13 because the input and output terminals are separate.

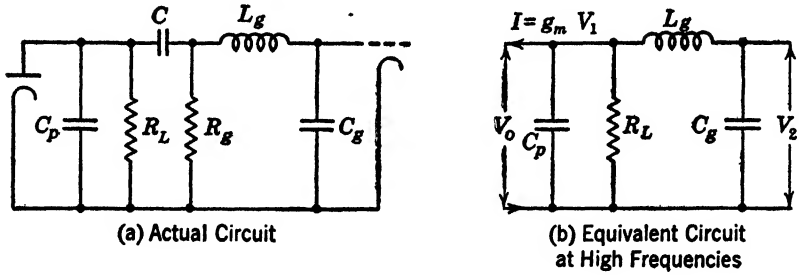


FIG. 13·19 Series compensated video amplifier stage.

The equations that apply are

$$I = V_0\left(j\omega C_p + \frac{1}{R_L} + \frac{1}{j\omega L_g}\right) - \frac{V_2}{j\omega L_g} \qquad [13\cdot34]$$

and

$$0 = -\frac{V_0}{j\omega L_g} + V_2\left(j\omega C_g + \frac{1}{j\omega L_g}\right) \qquad [13\cdot35]$$

Eliminating V_0 and introducing $I = V_1 g_m$

$$V_1 g_m = V_2 j\omega L_g\left(j\omega C_g + \frac{1}{j\omega L_g}\right)\left(j\omega C_p + \frac{1}{R_L} + \frac{1}{j\omega L_g}\right) - V_2\frac{1}{j\omega L_g} \qquad [13\cdot36]$$

or

$$\frac{V_2}{V_1} = \frac{g_m}{(1 - \omega^2 L_g C_g)\left(j\omega C_p + \frac{1}{R_L} + \frac{1}{j\omega L_g}\right) - \frac{1}{j\omega L_g}} \qquad [13\cdot37]$$

$$\frac{V_2}{V_1} = \frac{j\omega L_g g_m R_L}{(1 - \omega^2 L_g C_g)(R_L + j\omega L_g - \omega^2 R_L C_p L_g) - R_L} \qquad [13\cdot38]$$

The calculation of equation 13·38 is difficult but perfectly possible. However, it is impossible to tabulate the results on a single sheet like Fig. 13·16 since both L_g and C_g may be varied for any chosen values of R_L and C_p.

The interstage network of Fig. 13·19 is commonly referred to as *series compensated* or *series peaked*. The performance of a typical interstage of this sort is plotted in Fig. 13·20. Here as in Fig. 13·16 the resistance value is made equal to unity. The two-to-one ratio of the shunting

capacitances is typical of modern video amplifier tubes. The ordinate is equivalent to equation 13·38 because $I = g_m V_1$. The critical frequency is that at which the sum of the shunting capacitances would have a reactance equal to the load resistance.

The series-compensated interstage of Fig. 13·19 produces a somewhat greater band width than the shunt circuit of Fig. 13·13, and the phase characteristic is more linear. However, the ultimate phase shift reached at very high frequencies is 270° instead of 90°, a point of great importance if feedback is contemplated, and the behavior under transient conditions is hardly as favorable.

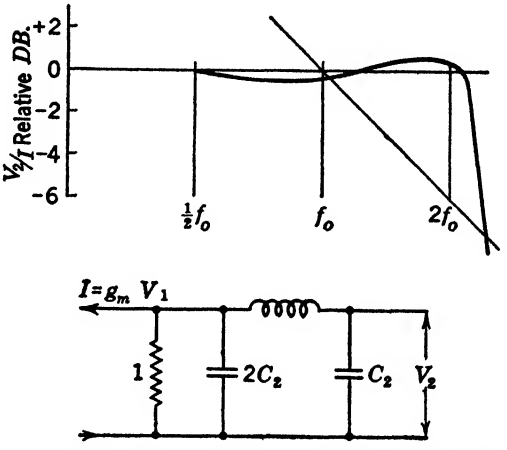


FIG. 13·20 Performance of a series compensated video interstage.

The doubly compensated interstage shown in Figs. 13·10 and 13·11 combines the features of the series and shunt circuits. With proper

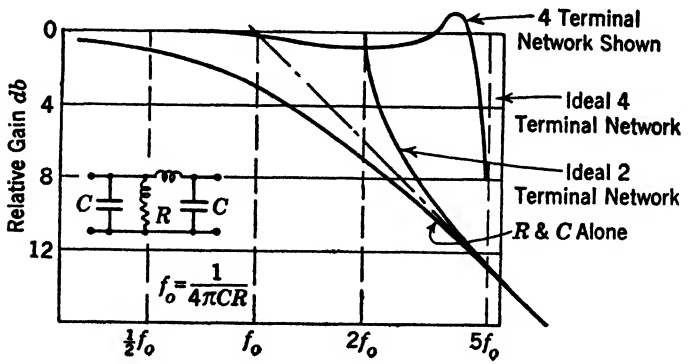


FIG. 13·21 Comparison between various ideal and physical interstage networks. Based upon a definite resistance and two equal capacitances. $C = C_p = C_g = C_t/2$.

design the amplification is nearly constant to a frequency $4f_0$ where $f_0 = \dfrac{R}{2\pi(C_p + C_g)}$. The theoretical performance of this network is still further improved by addition of a suitable condenser in shunt with

L_p, but the practical difficulties of adjustment make this addition of doubtful importance. The calculated performance of a particular doubly compensated four-terminal interstage network is shown in Fig. 13·21. The calculations, which are relatively lengthy and laborious, are omitted.

We have observed that the extension from two-terminal to four-terminal interstage networks allows us to extend the band of constant amplification to approximately twice its former value. Since it appears highly improbable that this process can be continued indefinitely we again seek some ideal limiting value by which the performance of actual networks may be judged. Wheeler* considers this case also, applying the methods of standard filter theory. He concludes that the ideal four-terminal interstage gives, for a constant total shunting capacitance, a band of given constant amplification which is exactly twice as wide as that for the ideal two-terminal interstage with the same total capacitance. This conclusion is derived by a consideration of networks from the standpoint of ordinary filter theory.

Unpublished work of more recent date approaches the problem through contour integration in the complex plane. This work, which is believed to be absolutely general, gives a figure equal to 2.5 for the advantage of the ideal four-terminal network over the ideal two-terminal network. Again the number of elements required is infinite in order to achieve the ideal response. Even worse, the phase shift at the upper edge of the band must be infinite. The situation is illustrated in Fig. 13·21.

The network shown in Fig. 13·21 may be designed on the basis of standard filter theory. This derivation requires that the capacitance associated with the resistance R be twice as large as the other capacitance. In order to meet the two-to-one ratio of capacitance a number of experimenters have recommended the addition of a suitable shunt condenser. Actually the frequency f_0 is always lowered by this procedure, sometimes seriously. Even more important, the concept of separating the total capacitance into two parts suggests that an equality of the two capacitances is an optimum situation. Careful calculations of a variety of conditions show that the ratio is not critical, curves similar to Fig. 13·21 being obtained over a considerable range of the variables. The addition of shunt capacitance is therefore to be resorted to only after a very careful consideration.

In multistage video amplifiers it is common practice to make all the interstage networks as similar as possible. This practice is admirable from the standpoint of manufacturing economy but leaves much to be desired in the way of performance. If several identical stages are used

* *Loc. cit.*

in tandem the distortions of phase and of amplitude both add up directly so that the individual stages must be very good indeed if the sum is to be tolerable.

It is desirable whenever possible to make some of the interstage designs rather different from the others. By a careful choice of the combination of designs it is possible to reduce the overall distortions to somewhat less than that of any one stage. By this procedure it is often possible to increase the gain per stage by a considerable amount without sacrificing overall performance.

13·19 Principle of Conservation of Band Width

The results of the foregoing work may now be summarized in a compact and useful way that is easily remembered. The voltage amplification of each stage is directly proportional to the load resistance R_L as shown in equation 13·33. The critical frequency f_0 is inversely proportional to R_L, according to the relation $1/2\pi f_0 C_t = R_L$, and the greatest band over which uniform constant amplification may be maintained is $5f_0$. Accordingly the product of voltage amplification in numerical ratio, and band width in cycles per second, is a constant independent of

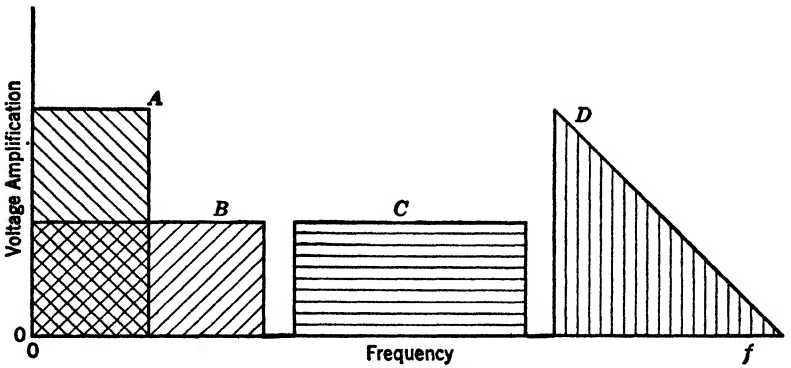


FIG. 13·22 Principle of conservation of band width. The four areas are equal as required by the fact that the product of voltage amplification and band width is constant.

R_L. This constant is proportional to the transconductance of the tube and is inversely proportional to the total shunting capacitance. The ratio of the transconductance to the capacitance of the tube is referred to as the figure of merit. This is expressed in micromhos per micromicrofarad.

For a given tube with fixed transconductance and capacitance the plot of voltage amplification against frequency has a definite area as seen in Fig. 13·22. For example, half as large an amplification is chosen in B as in A, but the frequency over which it is constant is

doubled. In the contour integration which establishes the fact that the ideal four-terminal network produces a gain which is constant up to a frequency $f = 5f_0$, it is also proved that the same product of amplification by band width holds when the band of frequencies does not extend to zero, as in C. This is conveniently expressed by the statement that the gain area is constant. It is also proved that the area concept applies unchanged in cases such as D where the amplification is not held constant. The figure of merit of a tube g_m/C_t' may therefore be as well expressed in terms of the gain area represented. Gain area, the product of a ratio by a frequency, is a most convenient parameter by which to compare tubes for video-amplifier applications. Expressed in terms of voltage amplification and frequency in megacycles the approximate gain areas of some tubes are shown in Table 13·2. The gain area of a tube is numerically equal to the highest frequency to which a video-amplifier stage employing that tube can be made to produce a constant amplification of unity.

TABLE 13·2

GAIN AREAS OF VARIOUS TUBES

Type	g_m Micromhos	C_g $\mu\mu$f	C_p $\mu\mu$f	Gain Area Amplification × Frequency in Megacycles
6AC7–1852	9000	11	5	440
6AG7	8000	13	7.5	300
6AB7–1853	5000	8	5	300
6SH7	5000	8.5	7.0	250
6SG7	4500	8.5	7.0	225
954	1400	3	3	180
6SJ7	1650	6.0	7.0	100

13·20 The Intermediate-Frequency Amplifier

The principle of conservation of band width is of great value in the design of amplifiers for a definite band of frequencies, such as the so-called intermediate-frequency or band-pass amplifier. It is convenient to use the term band-pass amplifier because the interstage networks reduce to more or less complex forms of band-pass filters. Similarly, it is common to refer to video or audio amplifiers as low-pass amplifiers, since their interstage networks are essentially low-pass filters.

Several practical features combine to make the actual construction of the band-pass amplifier more difficult than that of the comparable low-pass unit. First of these is the fact that a band-pass filter requires twice as many elements for a given type of performance as the comparable low-pass filter. Typically a series-resonant circuit in the band-

pass filter replaces a coil in the low-pass equivalent, and a parallel or anti-resonant circuit replaces each condenser. Thus there is a band-pass network which exactly corresponds to each low-pass structure, but the converse is not true. In Chart 13·1 the equivalence of a number of networks is illustrated.

Inspection of Chart 13·1 shows that even relatively simple low-pass structures require relatively complex structures for the equivalent band-pass unit. In view of the difficulty of adjusting elements, particularly the small elements required in high-frequency units, it is seldom practicable to use any but the simplest of the band-pass designs.

The second great difficulty in construction of band-pass units is the precision required of the elements. Let us suppose that elements accurate to 5 per cent are acceptable in a certain low-pass amplifier covering the range of 0–2 megacycles. From the conservation of band width we expect to be able to use the same tubes to cover the band of 20–22 megacycles with a comparable gain per stage. Now, however, a coil and condenser both 5 per cent high would shift their resonance from mid band to the edge of the band. Similarly a pair of elements whose values were low would shift the resonance to the other edge of the band. In this case the elements would have to be held to 0.5 per cent in order to obtain as consistent results as those obtained in the low-pass case.

This is an example of the general proposition that the accuracy required of the elements increases directly with the ratio of mid-band frequency to band width. More explicitly, the allowable tolerance of the elements is proportional to the band width over the mid-band frequency. In commercial radio receivers this fact has led to the universal use of individually adjusted units. At higher frequencies where wide bands are often necessary the problems resulting from the necessity of holding capacitances to a minimum as well as providing adjustment of the resonances are very difficult to solve.

In view of the difficulties just outlined it is universal practice to use only the simplest of the networks shown in Chart I and to provide manual adjustment for obtaining the exact characteristic desired. Circuit *a* is a two-terminal network and obeys the familiar laws of the resonant circuit. Dissipation in the coil and condenser is lumped with any physical load resistor to form the single shunt conductance. Provided that the effective conductance is essentially independent of frequency the gain characteristic is symmetrical against frequency plotted on a logarithmic scale. For values of Q above 25 the significant portion of the curve occurs in such a narrow frequency interval that the logarithmic scale is practically equivalent to a linear scale. A universal

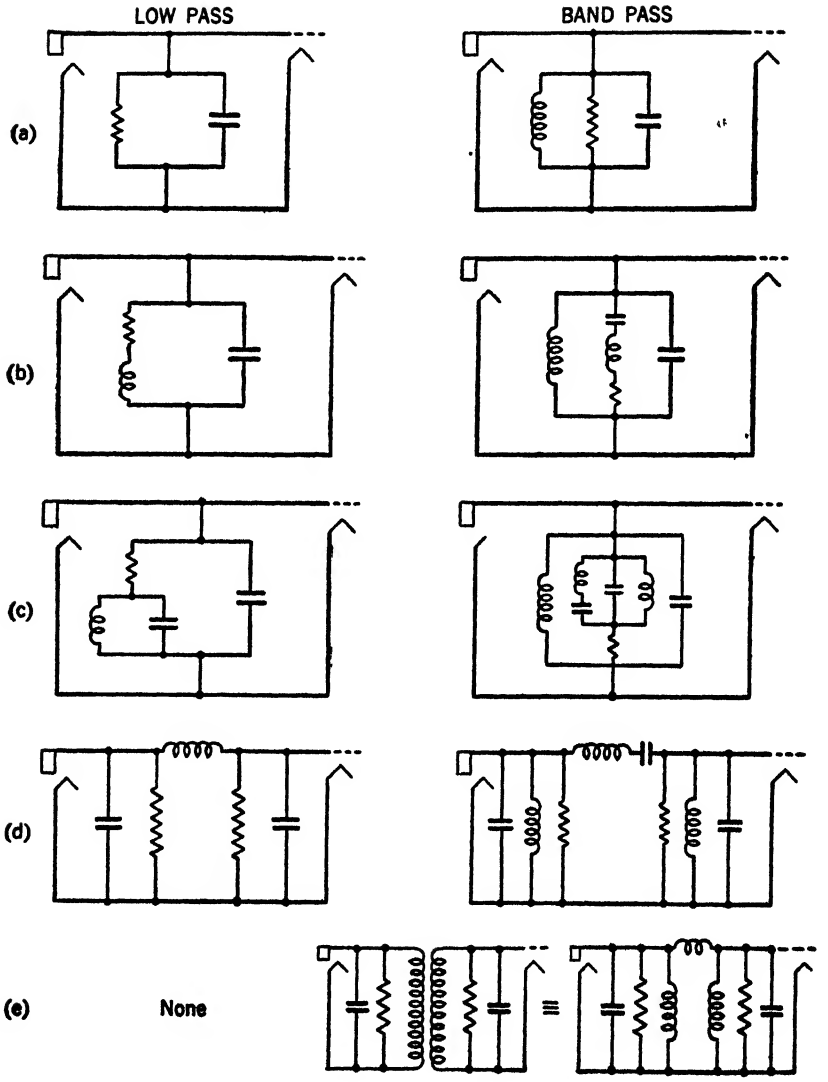
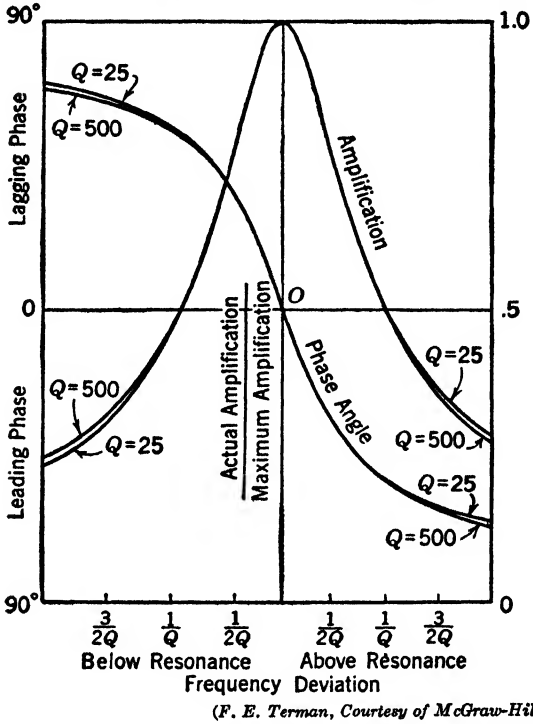


CHART 13·1. Low-pass and band-pass interstage networks.

plot, based on a linear scale, is presented in Fig. 13·23. A phase shift of 45° and a 3-db reduction in gain are associated with the two



(F. E. Terman, Courtesy of McGraw-Hill Book Company)

FIG. 13·23 Characteristics of a tuned band-pass amplifier.

frequencies f_1 and f_2 at which the reactance of the coil-condenser combination equals the resistance. The difference between these two frequencies is given by the relation

$$\frac{f_2 - f_1}{f_2 + f_1} = \frac{1}{2Q} \qquad [13\cdot39]$$

where

$$Q = \frac{R}{2\pi\sqrt{f_1 f_2}L} \qquad [13\cdot40]$$

a definition slightly more general than that of equation 13·43. If the band is relatively narrow we may write

$$f_0 = \sqrt{f_1 f_2} \simeq \tfrac{1}{2}(f_1 + f_2) \qquad [13\cdot41]$$

so that

$$\frac{f_2 - f_1}{f_0} \simeq \frac{1}{Q} \qquad [13\cdot42]$$

and

$$Q \simeq \frac{R}{2\pi f_0 L} \qquad [13\cdot43]$$

These are seen to be the normal equations of a resonant system except for the change in the definition of Q necessitated by the fact that the dissipation is assumed in parallel with the coil rather than in series with it.

The network designated e in Chart 13·1 is probably the most used of all the band-pass amplifier interstage circuits. It is a four-terminal network having no low-pass equivalent. It is seen to differ only slightly from network d above it, however, and this one does have a low-pass equivalent in the form of the series-compensated circuit. We have already shown* that d gives about half the useful gain area of the ideal

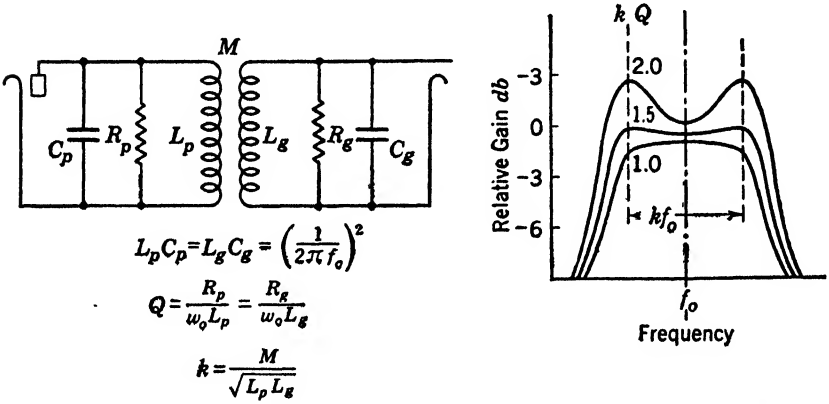


FIG. 13·24 Design and behavior of double tuned band-pass amplifier stage.

four-terminal network, and it can be shown that e is slightly superior. Thus we see that a sacrifice of approximately 50 per cent of the available gain area is ordinarily made in the interest of simplicity.

Development of design constants for network e may be carried out from either of two viewpoints. Standard filter theory may be used throughout, or the problem may be approached from the standpoint of coupled circuits. The results, of course, are identical. The choice will depend primarily upon the experience of the individual and the reference material at hand.

In the limiting case where the maximum possible gain area is to be obtained the two shunting capacitances may be reduced very nearly to the internal capacitances of the tube. In order to bring the two

* By comparison of Figs. 13·20 and 13·21.

circuits into resonance at the same frequency it is necessary, therefore, to use unequal inductances. If the two circuits are to have the same value of Q, the two resistors must likewise be unequal.

It should be noted that the two peaks of gain indicated in Fig. 13·24 are equally high only if the resistors R_p and R_g are independent of frequency. If the band is relatively narrow the approximation achieved is usually satisfactory, but if the band is relatively wide, difficulty may be experienced. This may occur at relatively high frequencies if part of the conductance in the grid circuit is due to the action of cathode inductance or transit time effects, since these conductances vary as the square of the frequency. A similar situation arises if a fixed resistor is used in series with either coil since the equivalent shunt resistance varies inversely with frequency. Explicitly, the total effective shunting conductance must be independent of frequency if the gain peaks are to be equally high.

13·21 Decoupling Circuits

The by-pass condensers shown in the video amplifier of Fig. 13·10 should all be used in the band-pass amplifier. Since band-pass amplifiers seldom operate at a frequency lower than 100 kilocycles, however, there is no difficulty in obtaining the desired low reactance in these condensers. At the higher intermediate frequencies it happens that the most compact available condensers present a reactance which is inductive rather than capacitive. This difficulty is relieved by extreme care in mechanical design to achieve the shortest possible leads. Occasionally it is practical to choose a relatively small capacitance so that the capacitative reactance is equal to that of the unavoidable inductive reactance of the condenser and the combination is series resonant.

At intermediate frequencies it is relatively easy to avoid undesired coupling of input and output circuits through the B supply. The reactance of available condensers is quite low, and the impedance of ordinary resistors does not fall seriously at these frequencies.* Serious problems do exist, however, in the form of electromagnetic and electrostatic coupling of circuits through the air and in conduction currents through the chassis or panel.

Placing closed metallic shield cans around each stage or in conjunction

* The resistors in common use are either of a carbon composition or are wire wound. Both types have some direct capacitance between the terminals, and the wire-wound resistors show an inductive effect as well. At a sufficiently high frequency both types of resistor become a low capacitive impedance, whose value depends upon the exact construction and material of the resistor. Wire-wound resistors are generally unreliable at frequencies above 10 megacycles. Many carbon resistors fail at frequencies in the order of 100 megacycles.

with groups of stages serves to solve the problem of coupling through the air. Awkward mechanical problems sometimes arise, and under these circumstances the designer must act primarily as a mechanical rather than an electrical engineer.

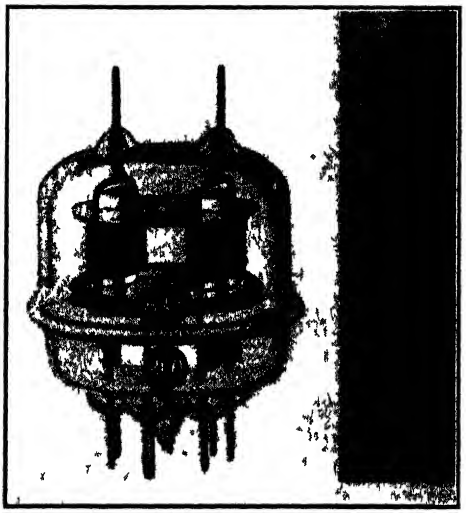
The problem of conductive coupling between output and input circuits is not so well understood. It is always true, however, that the highest possible conductivity in the panel or chassis is to be desired. In general the greatest difficulty is caused by currents in the plate circuit of the last tube inducing voltages in the grid circuit of the first tube.

If the plate, screen, and cathode condensers of the last tube are all returned to the same point on the panel where the suppressor grid is grounded, the difficulty is greatly relieved since the tendency to cause induction in the panel is thereby minimized.

Similarly, returning all the elements of the first grid circuit to a single point on the panel is helpful in that a given current flow in the panel induces a minimum voltage in such a circuit.

13·22 Power Amplifiers for High Frequencies

The foregoing discussion has applied primarily to low-power amplifiers for reception, although many of the remarks apply equally to higher-

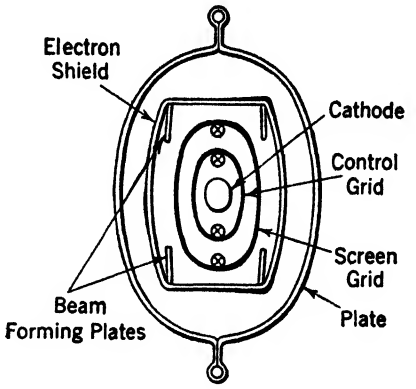


(Courtesy of RCA Radiotron Company.)

FIG. 13 25 High-frequency push-pull power tube.

power devices. The technique of high-power amplifiers for transmission at high frequencies is somewhat behind that of low-power amplifiers because of the difficulty of producing suitable tubes. The

medium- and high-power tubes that have served so well in the past become unsatisfactory at frequencies in the order of 20 megacycles. The same effects are operative here that set the limit on the performance of smaller tubes at somewhat higher frequencies. Interelectrode capacitances limit the attainable impedances. A relatively large part of the circuit capacitance resides inside the tube envelope, and the resulting capacitative currents in the leads are heavy and cause heating. Internal inductances make it difficult to control the constants of the circuit.



In receiving tubes it has been found possible to reduce the magnitude of these difficulties by reducing the size of the structures. For high-power transmitting tubes this technique is seriously limited by difficulties with dissipation of heat. Refined triode designs using air or water cooling and carefully chosen materials now give power outputs in the order of 100 watts at frequencies between 100 and 200 megacycles. In such tubes transit time effects are not ordinarily of importance because of the relatively high voltages and because the tubes are used in class C circuits. In such circuits the grid is highly positive during the time that the major portion of the plate current flows.

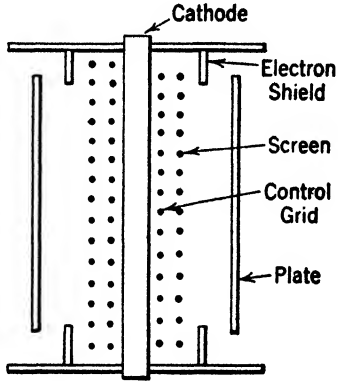


FIG. 13·26 Arrangement of electrodes in RCA 832.

The beam tetrode design has proved to be extremely desirable at moderate powers and frequencies. The 832 type illustrated in Fig. 13·25 is an excellent example of such a tube. This tube consists of two identical units in the same envelope. The two plate leads are brought out at one end of the tube while the grid and associated leads are brought out at the other. By using the two units in a balanced or push-pull fashion, it is possible to balance out all high-frequency currents in

cathode, heater, and screen-grid leads. Thus the problems of by-passing and of lead inductance are reduced to a minimum. This tube operates with practically no loss of efficiency up to frequencies of 150 megacycles and is still usable at 300 megacycles. Figure 13·26 shows the internal structure, and Fig. 13·27 presents typical operating characteristics. The power output is in the order of 20 watts.

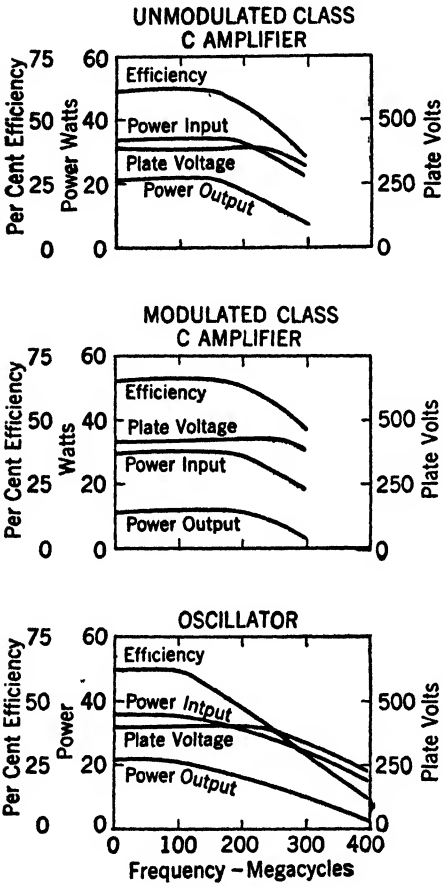


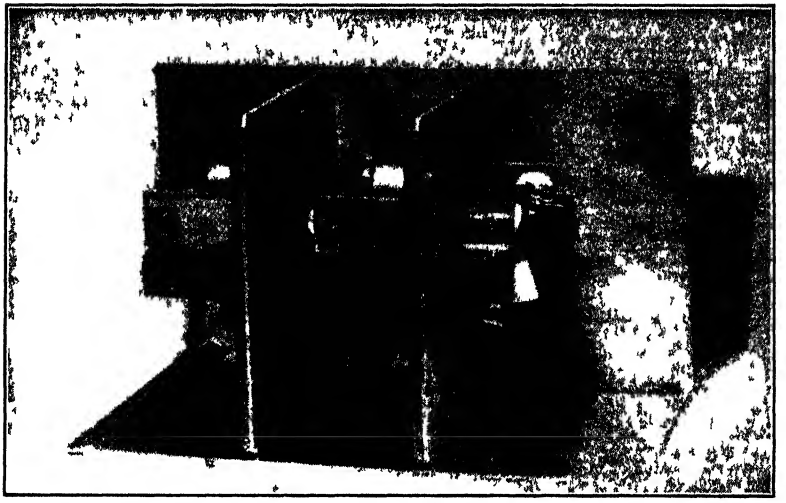
FIG. 13·27 Variations of efficiency and power output of 832 with frequency.

Figure 13·28 is a photograph of an amplifier for tubes of this sort operating at 300 megacycles as constructed by Samuel and Sowers.* The tube they use is similar to the 832 just described.

For larger power outputs at somewhat lower frequencies the 888 triode is satisfactory. The construction, shown in Figs. 13·29*a* and *b*, is characterized by short leads of large cross section, relatively small clearance, and large power dissipation per unit area. The relatively large value of plate dissipation is achieved by circulating cooling water through the plate structure. The inlet and outlet passage are shown in Fig. 13.29*a*. Characteristics are shown in Fig. 13·30.

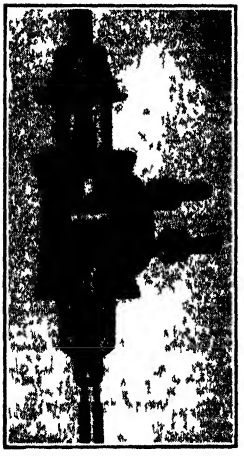
The techniques required for neutralization, shielding, etc., of these special triodes differ only in degree from those required at lower frequencies. As in all high-frequency work it is necessary to avoid any unnecessary lead inductance and to use extreme care in confining conduction currents in the shields, etc., to regions where they do no harm.

* A. L. Samuel and N. E. Sowers, "A Power Amplifier for Ultra High Frequencies," *Proc. IRE*, **24**, 1464, November, 1936.



(Samuels and Sowers, courtesy of IRE)

FIG. 13 28 Two-stage amplifier for wavelength of 1 meter.



(Courtesy of RCA Radiotron Company)

FIG. 13·29*a* A high-power high-frequency water-cooled triode.

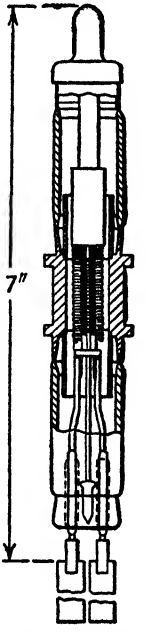


FIG. 13·29*b* Cross section of RCA type 888.

13·23 Summary

The design of audio- and video-frequency amplifiers is well understood. The various basic limitations are known, and methods are available for producing amplifiers which approach these limitations very closely. The design of intermediate-frequency amplifiers is also well understood, and again the basic limitations are known. Here, however, the prac-

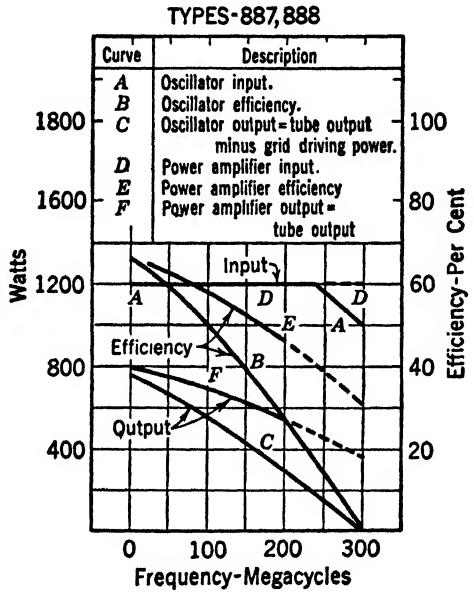


FIG. 13·30 High-frequency performance and input ratings of RCA type 888.

tical difficulties of approaching the ideal performance are so great that inferior results are ordinarily accepted. In intermediate-frequency amplifiers in the region from 20 to 100 megacycles it is necessary to exercise the greatest care in details of grounding, shielding, and by-passing, if large stable gains are to be obtained. Input conductance of the tubes produces serious loading, and the problem is a difficult one.

Whereas the principal difficulty with low-power amplifiers lies in circuit design, the greatest difficulty with high-power amplifiers lies in the construction of suitable tubes. At the present time no negative grid tube giving as much as 100 watts output at frequencies above 300 megacycles is commercially available. We shall see in a later chapter how a compromise solution of this problem is achieved by other means.

PROBLEMS

13·1 Calculate by the exact and by the approximate method the decibel equivalents of the following voltage ratios: 4.0; 8.0; 11.0; 20.0; 28.0; 50.0; 100; 500; 1400; 1,000,000.

13·2 Calculate by the exact and approximate method the voltage ratios corresponding to the following numbers of decibels: 2.0; 4.0; 5.0; 8.0; 12.0; 21.0; 37.0; 66.0; 94.0; 128.

13·3 Draw the circuit of a two-stage direct-coupled voltage amplifier using triodes with separately heated cathodes. Show the necessary battery connections, and indicate reasonable values for the voltages used.

13·4 A particular triode has the following parameters at a normal operating point: $g_m = 2000$ micromhos; $r_p = 10{,}000$ ohms; $R_C = 600$ ohms. This tube operates as an amplifier, and no cathode-bypass condenser C_c is provided. What are the equivalent values of transconductance, plate resistance, and amplification factor which exist under these conditions?

13.5 A cathode by-pass condenser is added to the system of problem 13·4. What are the equivalent values of transconductance, plate resistance, and amplification factor at the frequency at which the reactance of the condenser is 600 ohms?

13·6 In the amplifier circuit of Fig. 13·3 let $g_m = 2000$ micromhos, $r_p = 10{,}000$ ohms, $R_L = 100{,}000$ ohms, $C = 0.01$ μf, $R_g = 1$ megohm, $C_t = 100$ $\mu\mu$f. Calculate the midband (maximum) amplification per stage and the two frequencies at which the amplification is reduced to 0.707 of its maximum value.

13·7 In the amplifier of problem 13·6 the actual plate-to-cathode direct voltage of the first tube is 100 volts. The coupling condenser is known to have an appreciable leakage of direct current. How high must the resistance of this condenser be if the bias of the second tube is not to be shifted by more than 1 volt?

13·8 In the three-stage amplifier shown in Fig. 13·1 assume the constants of problem 13·6. Determine the smallest value for the decoupling condenser C_d which gives a net loss of 100 at 10 cycles per second. See Fig. 13·8.

13·9 Discuss the effects upon a video amplifier which result if the screen-grid by-pass condenser is too small; if the cathode by-pass condenser is too small.

13·10 Develop the exact formula for the impedance of the network of Fig. 13·13. Rationalize and separate real and imaginary components. Sketch each against the variable ω.

13·11 Develop the exact formula for the impedance of the network of Fig. 13·17. Outline a method and tabular form for calculating the impedance curves shown in the same figure.

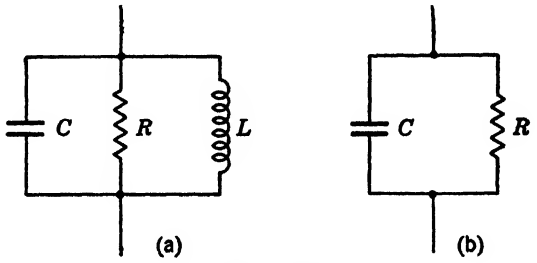
13·12 Given a tube having a transconductance of 6000 micromhos and capacitances such that $C_t = 30$ $\mu\mu$f, design an interstage network having an amplification constant to $\pm$ 1 db up to 4 megacycles. Use the curves of Fig. 13·16 and the network associated. (Consider only high-frequency performance.)

13·13 Using $R_g = 500{,}000$ ohms and $R_d = 20{,}000$ ω, choose values of C_d and C so that the amplification is at least 0.707 of its midband value for all frequencies above 2 cycles per second. (Refer to problem 13·12.)

13·14 Develop the transfer impedance (ratio of output voltage to input current) of the network of Fig. 13·21.

13·15 Repeat the development of problem 13·14 but interchange input and output terminals. (The identity of these results checks a general network theorem which is a form of the reciprocity theorem.)

13·16 Show that the difference between the two frequencies at which the absolute impedance of network (*a*) is $0.707R$ is equal to the frequency at which the impedance of network (*b*) is $0.707R$. Interpret this result in terms of the principle of conservation of band width.



PROB. 13·16

13·17 By use of an ideal four-terminal interstage network a particular tube gives a constant amplification of 40 up to 3 megacycles. What amplification will it give if the band of uniform amplification must be extended to 10 megacycles? What constant amplification will it give across the band 20 to 25 megacycles?

13·18 The tube of problem 13·17 is to be used in a special amplifier in which the voltage amplification is directly proportional to the frequency and is 10 at 10 megacycles. Again assuming an ideal four-terminal network, what is the highest frequency to which this procedure may be extended?

13·19 Referring to Fig. 13·10, make a sketch showing how input and output stages of an amplifier may be coupled together by a chassis of finite conductivity if suitable grounding precautions are not observed.

13·20 Consider a voltage amplifier in which the effective impedance of the first grid circuit is 500,000 ohms and the voltage amplification is 2 million. At a frequency of 10 kc, how large a direct capacitance between the output terminal and the first grid terminal may exist without causing spontaneous oscillation? (Assume most unfavorable phase relations.)

13·21 In a particular high-frequency tuned amplifier the input and output circuits have the same impedance. If the gain of the amplifier is 100 db, how large a coefficient of coupling may the input and output transformers have? Express this as a mutual inductance, given the fact that the primary and secondary of each transformer has an inductance of 1 millihenry.

13·22 Sketch an arrangement employing resonant lines as a tuned interstage network for a high-frequency amplifier. Show the equivalent low-frequency circuit.

13·23 The push-pull arrangement of two identical tubes in one envelope is quite advantageous for high-frequency amplifiers. Are comparable advantages to be expected of push-pull tubes as high-frequency oscillators?

CHAPTER 14

THE NEGATIVE GRID OSCILLATOR

14·1 Introduction

The term negative grid oscillator is used to denote the ordinary vacuum-tube oscillator when modified for ultra-high-frequency operation. The term negative is necessary to distinguish it from the positive grid or Barkhausen oscillator. The mechanism of oscillation at low frequencies is now well understood, and circuits which operate as virtually perfect oscillators are in use. That is, existing oscillators give an output approaching a perfect sinusoid which is constant in frequency and amplitude.*

At frequencies of a few megacycles or less it is common to use tetrodes or pentodes as oscillators. The large values of plate resistance and the low grid-plate capacitance which are characteristic of these tubes are favorable to the production of stable high-gain amplifiers. Because amplification is essential to the ordinary forms of oscillators the same features which make a tube desirable as an amplifier make it desirable as an oscillator.

At higher frequencies the triode is favored above all other types of oscillator tubes. The grid-plate capacitance is always undesirable in amplifiers but may be made to serve as part of the resonant circuit of an oscillator. The extra grids of the pentode structure decrease the effective grid-plate capacitance only at the expense of increasing both grid-cathode and plate-cathode capacitances. This modification serves only to decrease the maximum possible frequency of oscillation. Moreover, difficulties of by-passing, etc., are presented by these grids.

14·2 Basic Oscillator Theory

The most characteristic feature of an amplifier is that the power output, which is delivered to some sort of load impedance, is greater than the input power required. Because of this fact it is possible to return a fraction of the output power to the input, and, if the conditions are favorable, stable oscillations are produced. The output which for-

* L. A. Meacham, " The Bridge Stabilized Oscillator," *Bell System Tech. J.*, **17**, 574, 1938.

merly resulted from the action of a separate input now is generated spontaneously. In order for the output to be unaltered when the change from separate excitation to self-excitation is made, it is necessary that the voltage returned to the input be exactly equal to, and in phase with, the voltage which originally existed. If an attempt is made to increase the output by returning a larger fraction of the output power to the input, it will ordinarily be found that the increase produced is small. A limiting action in the vacuum tube has taken place so that the output is not increased appreciably when the input is increased. It may be shown in general for any kind of oscillator that the returned voltage is always equal to the voltage required at the point in question.

We may readily show that three distinct functions must be simultaneously performed in any system which is to produce steady periodic oscillations. A functional block diagram illustrating the situation is shown in Fig. 14·1. The unit designated amplifier ordinarily consists of a single vacuum tube, particularly at high frequencies. At lower

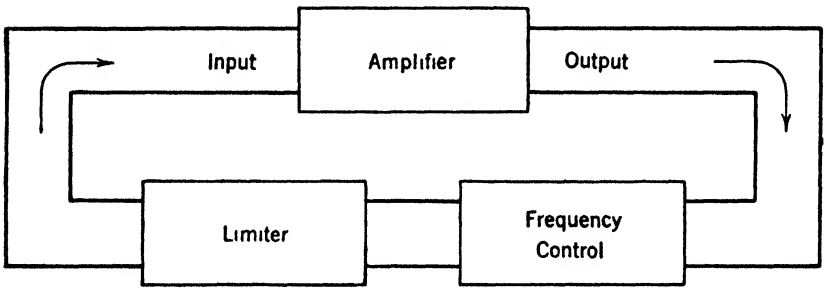


FIG. 14·1 Functional block diagram of an oscillator.

frequencies it is possible, and often advantageous, to use two or more tubes in this unit. In this analysis the amplification is independent of frequency or the magnitude of the signal. It may be a function of battery voltage or of tube properties. The unit designated frequency control consists of some kind of passive network in which the phase shift and the transmission are functions of frequency. Ordinarily the frequency-control unit consists of a single tuned circuit, but resistance-capacitance combinations or bridge circuits may also be used.

We have tacitly assumed in the above that both amplifier and frequency-control units are linear, that is, that the output of each is directly proportional to the input and that a sine wave is not distorted by either. It is not practical to build units in which the amplification or attenuation is absolutely constant. Accordingly, the two units so far mentioned do not satisfy the basic requirement that the voltage

returned to the input after one round trip is equal to the voltage initially assumed. Hence the function of the limiter is necessary.

If the gain of the amplifier is increased by some disturbance, such as a change of battery voltage, the oscillation level tends to rise without bound. The limiter is a device in which an increase of the oscillation voltage increases the loss, thus compensating the assumed increase of amplifier gain. The limiter is assumed to have no phase shift but may have attenuation.

In the operating condition the gain of the amplifier is exactly equal to the sum of the losses of the limiter and of the frequency-control units. This is necessary in order that the voltage returned to the input of the amplifier around the loop be equal to the voltage assumed at the input. Since the limiter is assumed to have no phase shift, and the amplifier phase shift is assumed to be either 0 or 180°, the phase shift in the frequency-control unit must be either 0 or 180°. This condition is necessary, since the returned voltage must be in phase with the input voltage for stable oscillation.

The limiting function may be accomplished in several different ways. Of these, three are relatively important. The most common method is to allow the vacuum tube, which also serves as amplifier, to overload. Under these conditions a decrease of attenuation in the frequency-control unit increases the input voltage to the tube, but the increase of the output voltage is negligible because of overload. This mode of operation always results in the generation of harmonic voltages and currents, which may be shown to affect the operating frequency.* The oscillators discussed in this chapter employ limiting by overload.

A tungsten lamp or other resistive element which has a marked temperature coefficient of resistance may be used as a limiter. As the oscillation level rises, the resistance of the lamp rises. If the lamp is used in a suitable bridge circuit, a considerable increase of attenuation in the limiter may be made to result from a small change in the oscillation voltage. It is obviously necessary to use lamps of very low power rating if the output of an ordinary tube is to produce an appreciable heating.

A portion of the oscillatory voltage may be amplified, rectified, filtered, and fed back as a bias voltage to the oscillating tube. Under favorable circumstances the resulting system, which is referred to as the automatic volume-control oscillator, gives an output that is accurately constant and virtually free from harmonics. This system, however, is not very practical at ultra-high frequencies.

* Groszkowski, "The Interdependence of Frequency Variation and Harmonic Content, and the Problems of Constant Frequency Oscillators," *Proc. IRE*, 21, 958, 1933.

14·3 Requirements

For laboratory purposes, the ideal oscillator is one in which the output frequency may be set at any desired value and will remain fixed at that value. The wave form is perfectly sinusoidal, and the output is equivalent to a known constant voltage in series with a known constant impedance.

The ideal oscillator for communication purposes must meet slightly different requirements. The output frequency must be as constant as possible. The wave form need not be perfect but should be good. The constancy of output voltage is less important, and the output impedance is relatively unimportant. It is important, however, that an oscillator for communication at ultra-high frequencies be capable of direct modulation, since modulated amplifiers are not yet practical at these frequencies. It is very desirable that pure amplitude modulation or pure frequency modulation be achieved separately, but this is seldom possible. Ordinarily both frequency and amplitude modulation occur in significant proportions.

Of all the requirements, however, that of frequency stability is most important. At a frequency of 1000 megacycles a deviation of 1 part in a million, long regarded as an unattainable goal, amounts to a shift of 1000 cycles. Deviations as large as 100 parts per million occur unless extreme precautions are taken, and such deviations lead to great practical difficulties. Because of its importance a large amount of effort has been devoted to improvement of frequency stability.

Frequency deviations may be traced to two primary causes, changes of the applied voltage and changes of the circuit elements. The applied voltages affect the amplification of the vacuum tube and thereby affect the limiting action. The distribution of harmonics is thus disturbed, and the equivalent reactance of the circuit is modified by the intermodulation produced. Moreover, the changes of heat produced in the tube affect the internal inductance and the internal capacitance, as well as modify the temperature of the external elements.

Probably the most serious problem at the present time is the construction of frequency-control (tank) circuits in which the natural frequency is adequately independent of temperature, humidity, vibration, and aging. This is a fundamentally difficult problem in that the conditions of operation are often quite unfavorable, and the stability requirements are high. The effects of vibration are minimized by the choice of stable mechanical designs and by the use of vibration-absorbing supports for critical components. The effects of humidity are often eliminated by sealing the circuit elements into an airtight or evacuated container. The effects of aging are avoided by the use of materials

having great permanence. Ceramic materials and most of the familiar metals have this required permanence; that is, their mechanical properties or dimensions do not change appreciably with the passage of time.

The problem of temperature is best illustrated by an example. Many modern ultra-high-frequency applications require that the apparatus operate successfully over the temperature range $-40°$ to $+60°$ C. Over this range an unrestrained copper wire changes its length by 1620 parts per million and its resistance by nearly 50 per cent. Evidently a very considerable amount of refinement is necessary if a tank circuit is to be produced in which the resonant frequency changes by only a few parts per million for the same temperature range.

The natural frequency of a resonant transmission line varies inversely with its length. A frequency variation of 1620 parts per million for a temperature change of 100° C is to be expected from a copper transmission line used as an oscillator tank. The variation, in common with most temperature coefficients of frequency, is negative. That is, an increase in temperature results in a decrease in frequency.

The inductance of a coil of fixed proportions and a fixed number of turns is directly proportional to its linear dimensions. Accordingly the temperature coefficient of inductance for a coil which expands without constraint is equal to the linear coefficient of thermal expansion. This relation has been tested carefully and holds rather accurately at low frequencies. At high frequencies, however, the observed coefficient is considerably larger than the value predicted above. Groszkowski explains this observation on the basis of skin effect.* At low frequencies the current flows over the entire section of the wire, and the behavior is uniform. At higher frequencies the current flows only in a thin surface skin of the wire, the thickness of the skin decreasing as the frequency and conductivity are increased. Accordingly the inductance observed at high frequencies is lower than that at low frequencies because the magnetic energy stored within the conductor is reduced. At some given frequency the inductance varies with temperature because of the change of conductivity, even though no change of dimensions occurs. A rise of temperature causes a decrease in the conductivity and a corresponding increase in the thickness of the skin which is effective in conducting the current. Therefore more magnetic energy is stored within the conductor and the inductance is increased.

The temperature coefficient of typical copper coils is about twice as large as the linear expansion coefficient. Coils having much lower coefficients may be constructed by using fine wires or ribbon-shaped

* Groszkowski, "The Temperature Coefficient of Inductance," *Proc. IRE*, **25**, 448, 1937.

conductors wound on ceramic forms. The construction of coils and condensers for low temperature coefficient of frequency is a highly specialized and rather difficult problem. The associated references are recommended.*

14·4 The Power Oscillator as a Class C Amplifier

Let us consider the operation of a triode as a low-frequency power amplifier or oscillator. This review is profitable because many design features are the same at all frequencies. The term power oscillator indicates that the tube carries plate current only in short pulses and that the C bias is relatively large. Under these conditions the tube operates essentially as a class C amplifier, and the efficiency of power conversion in the plate circuit is relatively high.

In order to obtain the maximum possible power *output* from a triode as a class C amplifier it is necessary to have a relatively low value of plate load impedance and to use large plate currents. Maximum *efficiency* in the plate circuit necessitates a higher plate load impedance and a reduction of the plate current and power output. In general the frequency stability of an oscillator is improved by increasing the plate load impedance to relatively high values. Thus the conditions for optimum frequency stability lead to a high plate circuit efficiency but a relatively low output power.

In the oscillator, a portion of the output power is required to drive the grid, whereas in the amplifier this power is supplied by a previous tube.. Accordingly it is desirable to have somewhat smaller values of grid bias and of grid excitation in the oscillator than in the amplifier so as to minimize this driving power, which is subtracted from the useful output power.

In the design of class C amplifiers it is common to set the C bias at approximately twice the value which reduces the plate current to zero. Extensive tests of a variety of typical commercial tubes indicate that this value is somewhat too high for best operation. A bias approximately 1.7 times as large as the cut-off value is more nearly optimum for almost all tubes. This relation is expressed by the equation

$$E_{bb} \simeq \frac{\mu}{1.7} E_{cc} \qquad [14\cdot1]$$

where μ is the amplification factor of the tube and E_{bb} and E_{cc} are plate and grid polarizing or bias voltages, respectively.

* H. A. Thomas, "The Stability of Inductance Coils for Radio Frequencies," *Jour. IEE*, **77**, 702, 1935; "The Electrical Stability of Condensers," *ibid.*, **79**, 297, 1936.

In a well-adjusted class C amplifier the maximum value of the alternating voltage in the plate circuit is nearly equal to the plate biasing voltage. The maximum value of the alternating voltage in the grid circuit is somewhat larger than the value of the grid biasing voltage so that the grid goes positive by an appreciable amount during each cycle. This condition may be expressed in terms of equation 14·1 by the approximation

$$E'_p \simeq \frac{\mu}{2.5} E'_g \qquad [14\cdot2]$$

where E'_p and E'_g denote the maximum values of the alternating voltages in plate and grid circuits respectively.

Relations similar to those of equations 14·1 and 14·2 evidently apply to the power oscillator. In this case, however, the grid driving power is

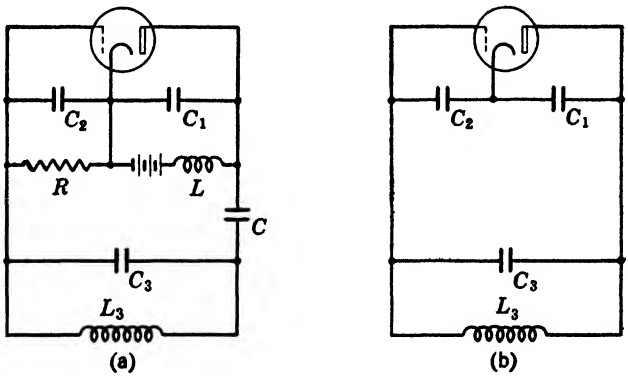


FIG. 14·2 Actual and equivalent circuit of a triode oscillator.

abstracted from the tank circuit, reducing the output, and smaller values of grid bias and grid excitation are desirable. The conditions for optimum operation may be described approximately by the relations

$$E_{bb} \simeq \frac{\mu}{1.5} E_{cc} \qquad [14\cdot3]$$

and

$$E'_p \simeq \frac{\mu}{2} E'_g \qquad [14\cdot4]$$

The ratio E'_p / E'_g, called the excitation ratio of an oscillator, is of considerable importance. It is fixed by the circuit which provides the feedback from plate circuit to grid circuit. In the circuit of Fig. 14·2, the oscillatory current which circulates through the coil L_3 and the con-

densers C_1 and C_2 is large compared to the pulses of plate current which flow only through C_1. Accordingly we may write to an excellent approximation.

$$\frac{E_p'}{E_g'} = \frac{C_2}{C_1} \qquad [14\cdot5]$$

Combining equations 14·4 and 14·5 to eliminate the alternating voltages,

$$\frac{\mu}{2} \simeq \frac{C_2}{C_1} \qquad [14\cdot6]^*$$

This equation must not be regarded as exact or infallible, but it does serve as a general guide in design. Experimental work with a variety of small triodes at low frequencies indicates that this relation is essentially correct. Optimum frequency stability and excellent constancy of output are observed when equation 14·6 is satisfied.

14·5 Frequency Limits of Triodes

It is generally known that there is a highest frequency at which any particular tube may be made to generate spontaneous oscillations. This may not exceed the frequency which results when the external terminals are connected in the most direct fashion possible. Usually, however, tubes fail to oscillate in such a short-circuit condition and require an appreciable external inductance. Two effects contribute to this failure. Electron transit time causes the effective grid-cathode conductance to rise to relatively high values, decreasing the effective Q of the resonant circuit and absorbing power. This conductance varies as the square of frequency and, therefore, changes from a magnitude which is negligible, to one which is of controlling importance, in a relatively small frequency ratio. Transit time also affects the phase angle of the transconductance, the plate current no longer being 180° out of phase with the grid voltage at high frequencies. In some cases this effect limits the band over which oscillations are produced even more severely than does input conductance. It is possible to extend the range of oscillation in these cases by introducing a suitable phase shift in the circuit which couples the grid to the plate. Fortu-

* The approximation expressed in equation 14·1 is accurate within 10 per cent for practically all the tubes now available commercially. The relation of 14·2 is much less exact in that it depends upon a choice between high efficiency and high power output as well as upon the tube. Equation 14·3 is of intermediate accuracy, seldom being in error by as much as 25 per cent. Equation 14·5 is quite exact if the Q of the circuit is moderately high so that 14·4 and 14·6 are of equal accuracy, of the order of 50 per cent.

nately, the phase shift which results from the grid conductance in typical circuits is in such a direction as partially to compensate for the effects of transit time upon the phase of the transconductance.

The validity of this statement is established by reference to Fig. 14·3. The elements shown represent the tuning coil, the capacitances C_1 and C_2, and the grid-cathode conductance. The vector diagram of this

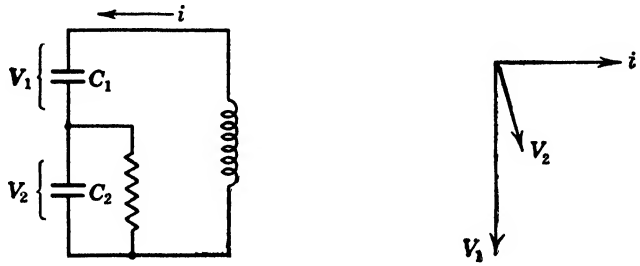


FIG. 14·3 Phase relations which result from grid-cathode conductance.

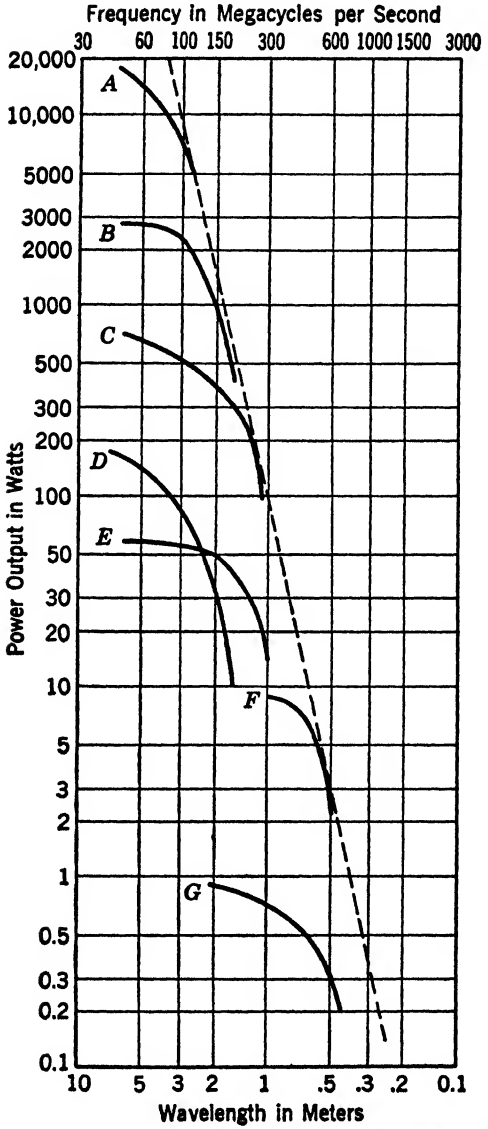
system is drawn, using the circulating current as the reference. The voltage across C_1 then lags the current by 90°, whereas the voltage across C_2 lags by a smaller angle. Accordingly the voltage V_2 reaches its maximum at an earlier time than V_1, and the delay between the time of maximum grid voltage and maximum plate current flow is partially compensated.

14·6 Effects of Tube Geometry

Figure 14·4 shows the variation with wavelength of the output power available from a number of commercial triodes. These curves are quite similar, suggesting that the same mechanism of limitation is active in all. Moreover, a number of these curves are tangent to a common straight line or asymptote shown dotted in the figure. This common asymptote is substantial evidence that some common physical limitation exists. Large tubes produce larger power outputs at low frequencies than small tubes, but the reduction of output begins at lower frequencies for large tubes.

The question immediately arises whether this line is fixed by some unchangeable physical limit. If so, the negative grid triode must be abandoned for wavelengths below about 20 cm because the output power, less than $\frac{1}{10}$ watt, is insignificant. If, on the other hand, the position of the asymptote depends upon the structure of the tube rather than upon its operating principle, we are far better off.

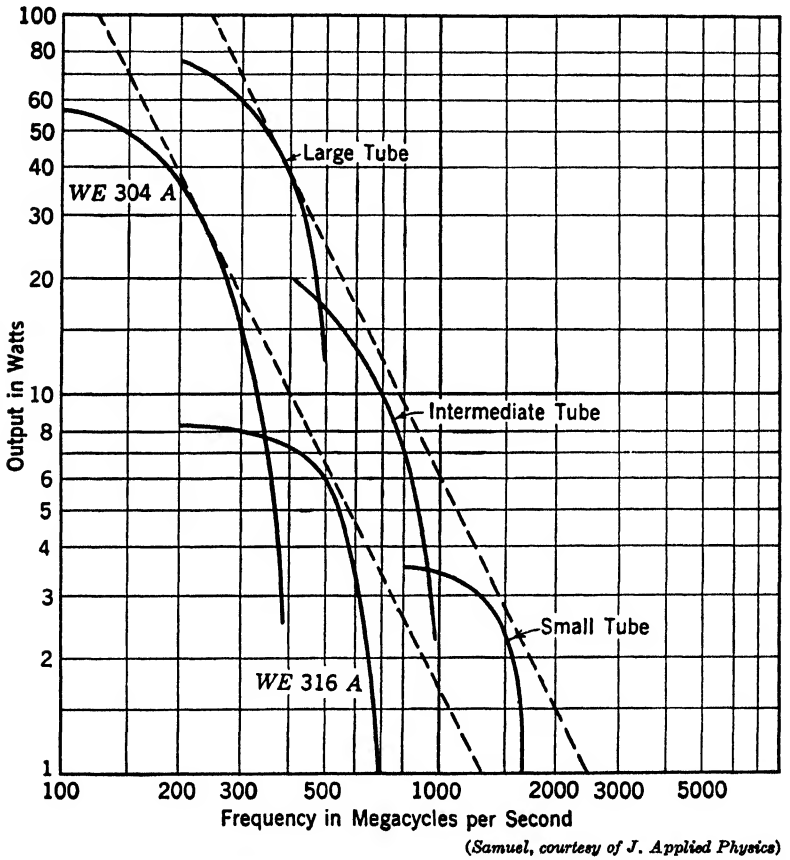
That the position of the dotted line does depend upon the structure and lead arrangement is shown by Fig. 14·5. Three special tubes of



(Samuel, courtesy of J. Applied Physics.)

FIG. 14-4 Output vs. frequency curves for some commercially available three-element tubes operating as oscillators.

similar double-ended construction but different sizes are compared with two similar tubes of single-ended construction. It is seen that the double-ended construction illustrated in Fig. 14·6 leads to approximately twice as high a limiting frequency for any particular power output as does the single-ended construction. Also, the dotted line associated with



(*Samuel, courtesy of J. Applied Physics*)

FIG. 14·5 Output characteristics of three experimental double-ended tubes compared with two modern single-ended tubes.

304A and 316A tubes is more favorable than is that of Fig. 14·4. The single-ended tubes of Fig. 14·5 are illustrated in Figs. 14·7 and 14·8. The advantage of double-ended operation is made clear by the characteristics of Fig. 14·9. Here a particular tube was used first in single-end and then in double-end connection.

The tube illustrated in Fig. 14·6*a* is an experimental hand-made

model. Figure 14·6*b* is the Western Elcctric 368A which is the commercial form of the same tube. The unusual construction of this tube deserves attention because of the excellent performance which results. In an appropriate circuit the 368A will oscillate to a maximum frequency of 1700 megacycles and it gives a power output of the order of

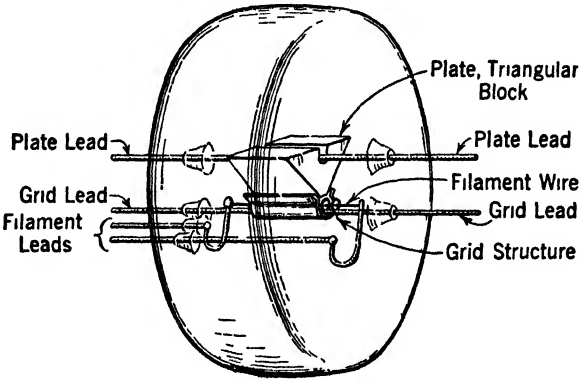
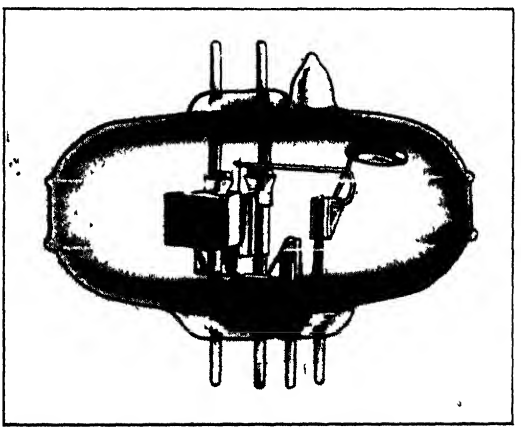


FIG. 14·6*a* Experimental model of double-ended high-frequency oscillator tube.



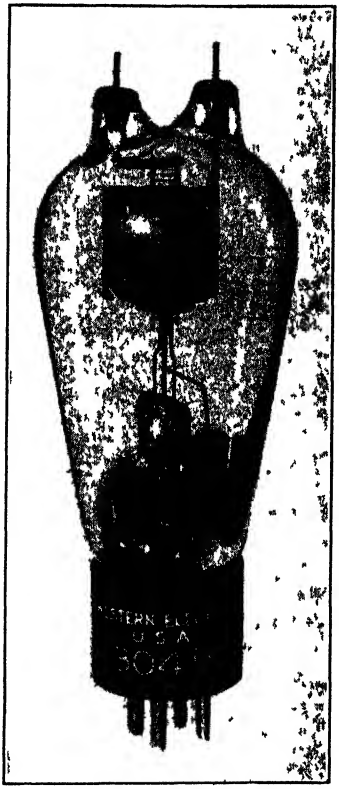
(*Courtesy of Bell System Tech J.*)

FIG. 14·6*b* Photograph of the W. E. 368A double-ended tube.

1 watt at 1500 megacycles. By the use of a suitable cavity resonator a considerable power output at the second harmonic frequency may be secured. No triode commercially available at this time gives comparable performance.

From the preceding work we conclude that the performance of nega-

tive grid triodes as oscillators is limited at least in part by the nature of the construction. Accordingly, we may expect to produce larger amounts of power and at higher frequencies by continued refinement of triode design.



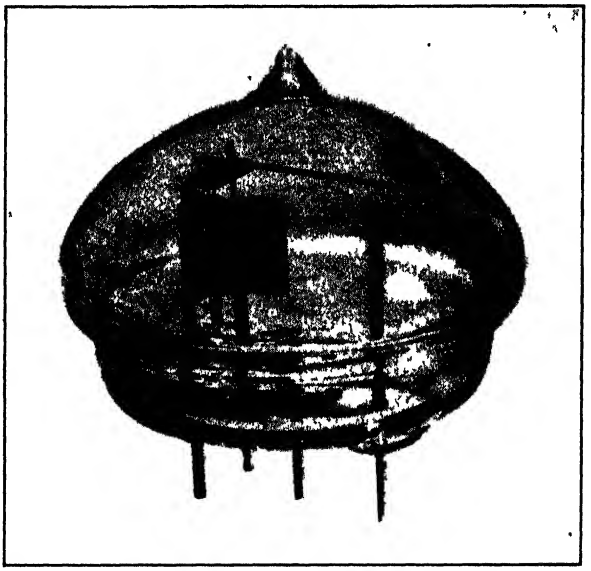
(*Courtesy of Bell System Tech J.*)

FIG 14 7 A high-frequency triode.

14·7 Practical Oscillator Circuits

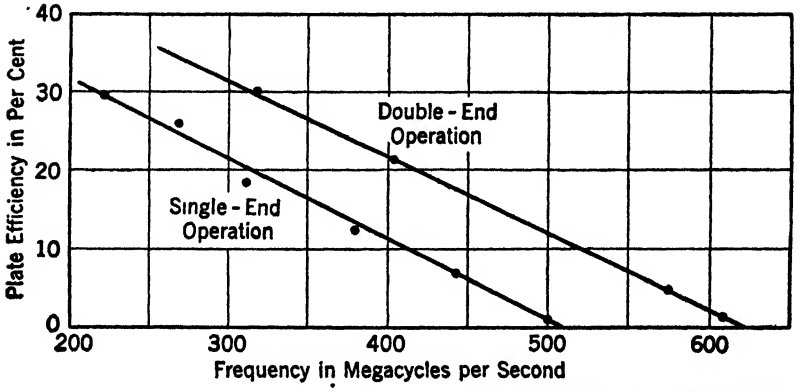
Let us consider the circuit of a practical ultra-high-frequency oscillator as shown in Fig. 14·2. The simplified equivalent diagram of Fig. 14·2*b* is a conventional Colpitts circuit except for the addition of the capacitance C_3. In the limiting case C_1, C_2, and C_3 are the plate-cathode, grid-cathode, and grid-plate capacitances of the tube itself. The effective capacitance across the coil is thus $C_3 + \dfrac{C_1C_2}{C_1 + C_2}$, and the oscillatory frequency is that for which the reactance of this effective

capacitance is equal to that of the inductance. The inductance in this case is the total inductance between the active portions of grid and plate, including the inductance of the internal leads as well as that of the con-



(Courtesy of Bell System Tech. J.)

FIG. 14·8 The Western Electric 316A single-ended "doorknob" tube.



(Samuel, courtesy of J. Applied Physics)

FIG. 14·9 Comparison plot showing effect of double-end operation.

ductors external to the tube. If the limiting frequency is to be high, it is necessary that both inductance and capacitance be held to minimum practical values. The short leads required are achieved only by placing

the internal elements close to the envelope of the tube. The low capacitances required are achieved by making the elements of the smallest practical size and by careful proportioning of the structure. These steps tend toward a marked decrease in the ability of the tube to dissipate heat and, therefore, tend to decrease the available power output. It is only by the use of materials capable of operating at very high temperatures and by refined design procedure that the power outputs now available at these high frequencies have been achieved. At best these power outputs are objectionably small.

When a tube is operating at its highest possible frequency the capacitances C_1 and C_2 are the plate-cathode and grid-cathode capacitances of the tube itself. The user of the tube is thus able to satisfy equation 14·6 only by adding capacitance in parallel with C_1 or C_2 as required. This is not a desirable procedure in that it increases the total shunting capacitance. Moreover, the unavoidable stray inductances make it difficult, if not impossible, to add a capacitance which is effectively in parallel with either grid-cathode or plate-cathode capacitance.

The magnitude and phase relations of the grid and plate voltages may be controlled by suitable coupled transmission lines. The necessary adjustments are difficult to make, and their relation depends upon the frequency required. Accordingly the resulting complication is seldom justified in practice.

Equation 14·6 should, whenever possible, be satisfied by the tube designer. A given grid mesh produces a greater grid-cathode capacitance and a greater transconductance as the mesh is moved closer and closer to the cathode. The plate-cathode capacitance is not greatly changed and the μ is somewhat reduced by this modification. Accordingly it is generally possible to approach or satisfy the desired relation, the principal difficulty being in the control of the small grid-cathode spacing required.

14·8 Practical Oscillator Construction

Ordinarily ultra-high-frequency oscillators employ resonant lines rather than lumped circuits as frequency-stabilizing elements. As previously mentioned, lines have the advantage of high Q and high intrinsic mechanical stability. At frequencies above 200 megacycles they are almost exclusively used.

A practical oscillator must utilize a suitable tube in conjunction with a resonant line system in such a way as to satisfy the conditions of the schematic circuit. Ordinarily the grid and plate of the tube are connected to the ends of the resonant line system, the other end of the line being open- or short-circuited as the conditions dictate. In the circuit

of Fig. 14·10 a Western Electric 316A is tuned by an open-circuited line one-half wavelength long. The filament is isolated by means of small high-impedance choke coils so that the high-frequency currents are confined to the resonant line and the interelectrode capacitances.

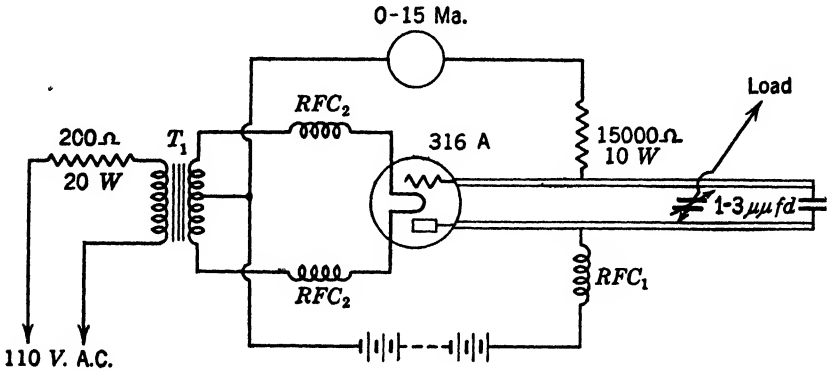
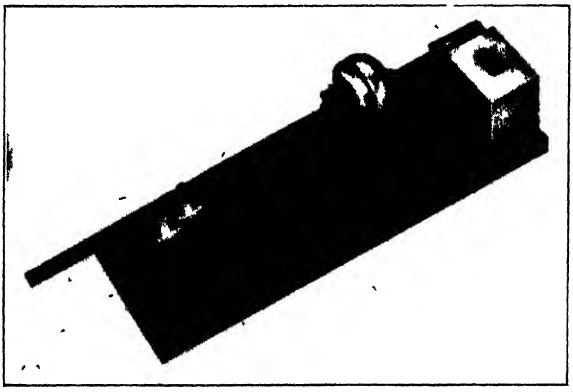


FIG. 14·10 A 400-mc negative grid oscillator with WE-316-A tube.

The physical arrangement of the 316A oscillator is shown in Fig. 14·11. The resonant line consists of two copper rods 6 mm in diameter and 12 mm apart on centers. The filament chokes, of relatively heavy wire as dictated by the large filament current, consist of approximately 30

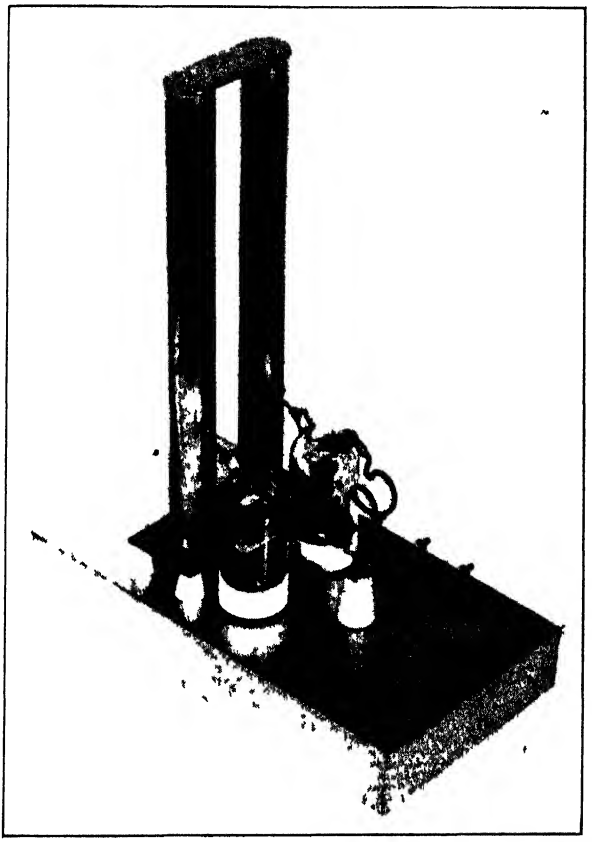


(Courtesy of Radio)

FIG. 14·11 View of an experimental 400-mc oscillator.

turns. This particular unit gives a power output of 5 watts at 400 megacycles. The plate circuit efficiency exceeds 50 per cent, and the frequency stability is good.

An interesting oscillator construction is illustrated in Fig. 14·12. The tubes employed have both grid and plate leads brought out at the top of the tube. Only the filaments are connected to the conventional tube base pins. The plates are connected by a small coil which is tuned



(Courtesy of Radio)

FIG 14 12 A push-pull oscillator for 224 mc employing HY-75 tubes.

to the operating frequency by the interelectrode capacitances. The grid circuit provides the principal frequency-stabilizing element in the form of a short-circuited half-wavelength line. The grids are tapped at points near but not at the geometrical centers of the bars, thereby decreasing the effect of the conductance added by the grids. Figure 14·13 is a semi-schematic diagram illustrating the connections. The circuit is recognized as a symmetrical balanced, or push-pull, oscillator. Accordingly the grid-cathode impedances of the two tubes are in series

and the effective impedance is thus doubled. The necessary coupling between grid and plate circuits is provided through the grid-plate capacitances of the two tubes. Accordingly this circuit is merely a symmetrical form of the tuned-grid tuned-plate or Meissner oscillator. This

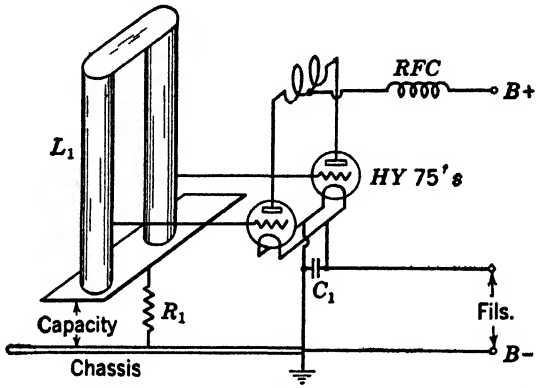


FIG. 14·13 Semi-schematic of a push pull 224-mc. oscillator.

C_1—0.003 μf midget mica.
R_1—5000 Ω 10 w.
L_1—Half-wave parallel rod line.
L_2—2 turns $\frac{5}{8}$ in. dia., 1 in. long.
RFC—6 turns $\frac{1}{4}$ in. dia.

particular unit is designed to operate at 224 megacycles, but by proper choice of tubes considerably higher frequencies may be produced. At higher frequencies it is usually desirable to use a resonant line for the plate circuit as well as for the grid circuit.

(*Samuel, courtesy of J. Applied Physics*)

FIG. 14·14 Photo of oscillator using a double-ended tube.

An oscillator for use with the Western Electric 368A or similar double-ended tubes is shown in Fig. 14·14. Here the tube is located at the middle of a half-wavelength line, the tube leads themselves forming a significant portion of the total length of line. In this assembly the

filament leads are contained within metal sleeves, the structure being so adjusted that each operates as a line one-fourth wavelength long and so offers a high impedance to the oscillatory frequency. This particular unit may be made to oscillate at frequencies as high as 1700 megacycles, the limitation being due to the tube and not to the oscillator structure itself.

At frequencies above about 500 megacycles the radiation of power from open resonant line systems is so great that some refinement is necessary. The use of large plates as short-circuiting elements is helpful but mechanically awkward. The most satisfactory solution to this difficulty is to enclose the entire oscillator and resonant line system within a metal container. The resonant lines now operate as a shielded pair rather than as a coaxial or simple parallel wire pair. Radiation is thus entirely eliminated.

Several difficulties arise when an oscillator is completely enclosed in this fashion. The heat liberated by the tube is not readily dissipated, and the power input must, therefore, be reduced. The mechanical construction and assembly are necessarily complicated by the addition of the enclosure. Sometimes the mechanical problems which arise are very serious. Finally, the enclosure must not be too large, or it may function as a cavity resonator, greatly complicating the operation of the circuit. In view of these difficulties the design of such enclosures must be carried out with considerable caution.

PROBLEMS

14·1 The receiving triode 7A4 may be made to oscillate at relatively high frequencies. Compare the constants of this tube with those called for by equation 14·6. What change in the value of C_2 is indicated if the other two values are regarded as fixed?

	7A4	6SF5
Transconductance	2500 micromhos	1500 micromhos
Plate resistance	8000 ohms	67,000 ohms
Amplification factor	20	100
Capacitances		
Grid to plate	4·0 $\mu\mu$f	2·4 $\mu\mu$f
Grid to cathode	3·4 $\mu\mu$f	4·0 $\mu\mu$f
Plate to cathode	3·0 $\mu\mu$f	3·6 $\mu\mu$f

14·2 Repeat problem 14·1 for the high-mu triode 6SF5.

14·3 If the triode of problem 14·1 has an effective grid-cathode conductance of 300 micromhos due to electron transit time at a frequency of 200 megacycles, calculate the phase angle between grid-cathode and plate-cathode voltages. Assume that transit time effects do not appreciably affect the plate resistance as given.

14·4 Repeat problem 14·3 for the 6SF5 of problem 14·2.

14·5 Calculate the inductance required to tune the interelectrode capacitances of the 7A4 to a frequency of 500 megacycles. Using standard formulas,

calculate the approximate size of a one-turn coil of No. 20 wire which produces this inductance.

14·6 Repeat problem 14·5 for the 6SF5.

14·7 For a particular lamp the resistance is given approximately by the relation $r = 100 + 20I$, where I is the current in milliamperes. A Wheatstone bridge uses two of these lamps as diagonally opposite arms, the other two arms being fixed resistors of 200 ohms. A galvanometer of negligible impedance is used with this bridge. Calculate and plot the galvanometer current as a function of the voltage applied to the bridge. (Consider such a bridge as a limiter in an oscillator.)

14·8 Repeat problem 14·7 using a galvanometer of very high impedance. Plot galvanometer voltage versus applied voltage.

14·9 In a dynatron oscillator the only tuning elements are the coil and condenser of a parallel resonant circuit. These are connected to the tube, which acts as a non-linear negative resistance. Explain the observed fact that the frequency of oscillation is not exactly the natural frequency of the tank circuit.

14·10 In problem 14·9 the frequency stability may be improved by shunting across the tank series-resonant LC circuits tuned to the harmonics of the oscillating frequency. Explain.

14·11 It has been suggested that a cylindrical triode be built as part of a coaxial line resonator. Sketch such an arrangement, and mention some of the practical difficulties that would have to be overcome. Consider the advantages of this arrangement in terms of transit time, internal capacitances and inductances, and phase shifts.

14·12 Consider the use of a screen-grid tetrode as a high-frequency oscillator. Will transit time of the electrons from screen grid to plate affect the operation? How may the required feedback be secured in such an oscillator?

14·13 A particular cylindrical triode has a straight filamentary cathode, a small circular grid, and a plate whose diameter is several times that of the grid. A screen grid of relatively fine mesh is to be inserted between grid and plate. What effect upon this tube as an oscillator will the addition of such a grid have if it is thoroughly by-passed to the cathode?

14·14 A tetrode is built in which both grids are of the conventional helical form, one being slightly larger in diameter than the other. If the inner grid is biased somewhat positive it is found that the outer grid has a relatively large value of transconductance. Consider transit time and other features of this tube as a high-frequency oscillator.

14·15 If the various grids of a normal pentode form true equipotential planes, a single electron in its motion from cathode to plate constitutes, successively, currents between cathode and control grid, control grid and screen grid, screen grid and suppressor grid, and suppressor grid and plate. Which of these several transit times are important in affecting the input conductance of the tube? Which affects the total phase shift between input and output?

14·16 High frequencies are sometimes generated by using a harmonic of an oscillator. A particular oscillator of the class C variety operates with the grid excitation voltage of frequency f and the plate a–c voltage largely of frequency $2f$. Sketch curves of plate and grid voltage and current. What are the advantages of such operation?

14·17 Repeat problem 14·16 for third harmonic operation. Suggest a network suitable for producing the operation shown.

CHAPTER 15

THE POSITIVE-GRID OR RETARDING-FIELD OSCILLATOR

15·1 Introduction

Perhaps the most commonly available generator of ultra-high-frequency oscillations is the positive grid triode. Many commercial triodes will generate these oscillations provided that the electrodes are cylindrical and that suitable voltages are applied. Satisfactory oscillations may also be produced with other electrode arrangements, but suitable proportions are seldom found in commercial tubes. Unfortunately the efficiency is never high, and the design features which lead to high frequencies also lead to low power outputs. Nevertheless this generator is one of considerable practical importance. When used in this connection a triode is referred to as a retarding field generator or oscillator.

The technology of positive grid tubes dates from 1920 when Barkhausen and Kurz* published the results of certain experiments with high-power triodes. They discovered the existence of strong oscillations of very high frequency which are maintained so long as the grid voltage is relatively large and positive and the plate voltage zero or negative. Their work indicated that the frequency was dependent only on the transit time of the electrons, as determined by the geometry of the tube and the applied voltages. The external circuit seemed to produce no effect upon the frequency.

Two years later Gill and Morrell† published the results of further work on positive grid or retarding field oscillators. Their work indicated that the frequency was dependent primarily upon the constants of the external circuit and was essentially independent of the electron transit time. It was therefore assumed that an entirely different phenomenon had been discovered. The distinction in terminology continues to the present time, in spite of conclusive evidence that the two forms of oscillation are reducible to the same cause.

* H. Barkhausen and K. Kurz, "The Shortest Waves Producible by Means of Vacuum Tubes," *Physik. Zeit.*, **21**, 1, 1920.

† E. W. B. Gill and J. H. Morrell, "Short Electric Waves Obtained by Valves," *Phil. Mag.*, **44**, 161, 1922.

15·2 Electron Transit Time in a Plane Triode

Let us first examine the motion of a single electron under the action of the electrostatic fields present in a vacuum tube. The distances and time intervals will be determined for a typical case. In a later section it will be shown how electrons form themselves into groups that oscillate together and how they transfer power to an external circuit.

Consider the plane-parallel arrangement shown in Fig. 15·1. Assume that no appreciable space charge exists, that electrons are emitted from the cathode with zero velocity, that no contact potentials are present, and that the grid mesh defines an effective equipotential plane. Elec-

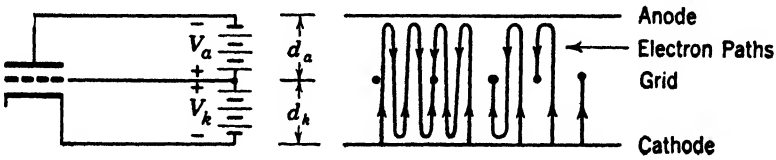


FIG. 15·1 Arrangement of a retarding field oscillator.

trons emitted from the cathode find themselves in a relatively strong electric field and are accelerated toward the positive grid. Some of the electrons strike the grid wires and give up their kinetic energy as heat. In general, however, electrons pass through openings of the grid structure with the velocity acquired in the grid-cathode transit. In the grid-plate region the electric field is in the opposite direction and tends to slow down the electrons, hence the term *retarding field.*

If the voltage V_a is greater than V_k the electrons come to rest somewhere between the grid and plate and return toward the grid. Those which again miss the grid structure are once more slowed down in the retarding field of the cathode and come to rest at the cathode surface, returning toward the grid again. If V_a is less than V_k the electrons are not returned to the grid but strike the plate and are lost. Therefore we must devote our attention to values of $V_a \geqq V_k$ if we are to produce effective electronic oscillations.

Figure 15·2 illustrates the potential distribution existing in the tube of Fig. 15·1. Unequal spacings and voltages are taken in the interest of generality. An electron freed from the cathode arrives at the grid with a velocity v as deduced by equating the electrical work done to the kinetic energy,*

$$\tfrac{1}{2}mv^2 = V_k e \qquad [15·1]$$

* It can be shown that the increase of mass predicted by the theory of relativity is negligible for the cases in question here.

which gives for v

$$v = \sqrt{\frac{2e}{m} V_k} \qquad [15\cdot2]$$

where m and e are the electronic mass and charge, respectively. Using

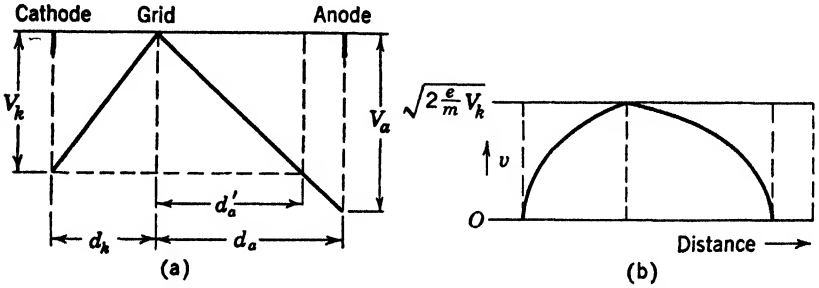


FIG. 15·2 Potential and velocity distribution in a plane retarding field oscillator.

the value $e/m = 5.30 \times 10^{17}$ esu per gram, and V_k in esu, equation 15·2 becomes

$$v = 1.03 \times 10^8 \sqrt{V_k} \text{ cm/sec} \qquad [15\cdot3]$$

which when converted to practical units gives

$$v = 6.0 \times 10^5 \sqrt{V_k} \text{ meters/sec} \qquad [15.4]$$

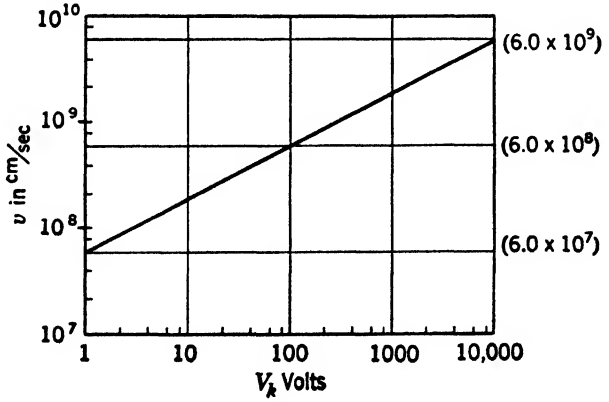


FIG. 15·3 Relation between the final electron velocity v and the applied voltage V_k.

where v is the velocity with which electrons reach the grid and V_k is in volts. This relation is plotted in Fig. 15·3.

Since the potential distribution curve of Fig. 15·2 is linear it is possible

to deduce the electron velocity at any point by means of equation 15·4. The resulting parabolic curve is shown in Fig. 15·2*b*.

Because the electric field is uniform the force on the electron is constant and the motion that results is one of uniform acceleration. In such a case we may write

$$\bar{v} = \frac{v}{2} \qquad [15\cdot5]$$

where $\bar{v}$ is the average velocity from cathode to grid and v is the final velocity as given by equation 15·4. The time required for an electron to pass from cathode to grid is expressed by the equality

$$d_k = \bar{v}t \qquad \text{or} \qquad t = \frac{d_k}{\bar{v}} \qquad [15\cdot6]$$

Using equations 15·4, 15·5, and 15·6 in combination we have

$$t = \frac{d_k}{3 \times 10^5 \sqrt{V_k}} = 3.33 \frac{d_k}{\sqrt{V_k}} \times 10^{-6} \text{ sec} \qquad [15\cdot7]$$

where t is the time required for an electron to pass from cathode to grid, d_k is the grid cathode spacing in meters, and V_k is the grid-cathode potential in volts.

Evidently this value applies equally well to the time required for an electron to be accelerated from cathode to grid or to be retarded to rest at the cathode after returning through the grid.

Conditions in the anode space are not appreciably more complicated. An electron which passes through the grid is retarded to rest at a point d_a' distant from the grid where the potential is equal to that at the cathode. The linear potential distribution permits us to write the equation

$$\frac{d_a'}{d_a} = \frac{V_k}{V_a} \qquad [15\cdot8]$$

where d_a and d_a' are the distances shown in Fig. 15·2. This relation may be used only if $V_k < V_a$.

The time required for an electron to leave the grid and come to rest at a point d_a' distant is given from the above relations

$$t = 3.33 \frac{d_a'}{\sqrt{V_k}} \times 10^{-6} = 3.33 \frac{d_a \sqrt{V_k}}{V_a} \times 10^{-6} \text{ sec} \qquad [15\cdot9]$$

We are now in a position to tabulate three time intervals which are of

fundamental importance:

$$t_a = 6.67 \frac{d_a\sqrt{V_k}}{V_a} \times 10^{-6} \text{ sec} \qquad [15\cdot10]$$

$$t_k = 6.67 \frac{d_k}{\sqrt{V_k}} \times 10^{-6} \text{ sec} \qquad [15\cdot11]$$

$$t_t = 6.67 \frac{V_a d_k + V_k d_a}{V_a\sqrt{V_k}} \times 10^{-6} \text{ sec} \qquad [15\cdot12]$$

where t_a is the time required for an electron to pass from the grid to the zero potential plane near the plate and return to the grid; t_k is the time required for an electron to pass from the grid to the cathode surface and return to the grid; and $t_t = t_a + t_k$ is the time required for one entire transit from cathode to plate and back. The distances are expressed in meters and the potentials in volts. Under the conditions assumed the above equations are exact, and these assumptions are generally met in practice.

15·3 Electron Transit Time in a Cylindrical Triode

In the cylindrical structure of Fig. 15·4 the potential with respect to the grid at any point r in the grid-cathode region is a logarithmic function given by the expression*

$$V = V_k \frac{\log_e \frac{r}{r_g}}{\log_e \frac{r_k}{r_g}} \qquad [15\cdot13]$$

In the grid-anode region the potential at any point r is given by the expression

$$V = V_a \frac{\log_e \frac{r}{r_g}}{\log_e \frac{r_a}{r_g}} \qquad [15\cdot14]$$

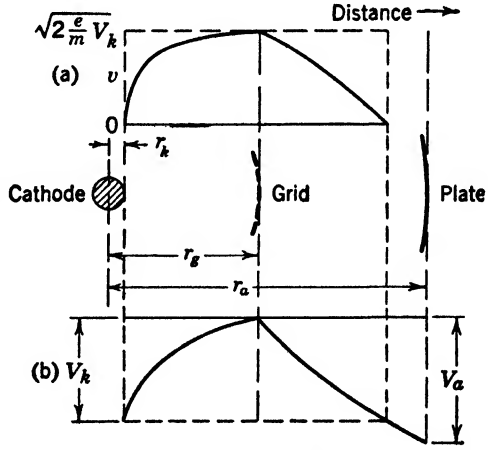


FIG. 15·4 Potential and velocity distribution in a cylindrical retarding field oscillator.

where r_k, r_g, and r_a are the radii of cathode, grid, and plate, respectively. Because of the high field strength in the region of the cathode an electron

* This expression is readily derived by assuming a fixed charge on a cylindrical condenser and integrating the resulting field, which varies inversely with r.

is strongly accelerated, and the average velocity of the transit from cathode to grid is greater than half the final velocity. In the grid-anode region the potential distribution is less distorted, but an electron returning from the plate region to the grid has an average velocity which is less than half the final velocity. The computation of the transit times involved is a rather lengthy process. It has been carried out by Scheibe,* who has obtained the following relations:

$$t_a = \frac{4r_g}{\sqrt{\frac{2e}{m}V_k}} \cdot G(y) \quad \text{sec} \qquad [15\cdot15]$$

$$t_k = \frac{4r_g}{\sqrt{\frac{2e}{m}V_k}} \cdot F(x) \quad \text{sec} \qquad [15\cdot16]$$

$$t_t = \frac{4r_g}{\sqrt{\frac{2e}{m}V_k}} [F(x) + G(y)] \quad \text{sec} \qquad [15\cdot17]$$

where t_a is the time required for an electron to pass from the grid to the zero potential region near the plate and return to the grid; t_k is the time required for an electron to pass from the grid to the cathode surface and return to the grid; and $t_t = t_a + t_k$ is the time required for one complete transit from cathode to plate and back. The quantities may be expressed in any system of units. The functions $F(x)$ and $G(y)$ are defined by the relations

$$x = \sqrt{\log_e \frac{r_g}{r_k}} \qquad [15\cdot18]$$

$$y = \sqrt{\frac{V_k}{V_a} \log_e \frac{r_a}{r_g}} \qquad [15\cdot19]$$

$$F(x) = x \cdot \varepsilon^{-x^2} \int_0^x e^{u^2}\, du \qquad [15\cdot20]$$

$$G(y) = y \cdot \varepsilon^{+y^2} \int_0^y e^{-u^2}\, du \qquad [15\cdot21]$$

where u is any variable whatever, since it disappears in taking the definite integral.

* A. Scheibe, "The Generation of Ultra Short Waves with Hot Cathode Tubes," *Ann. Physik.*, 73, 54, 1924.

These functions are not readily calculated directly but may be derived from the curves of Fig. 15·5 provided that $V_a = V_k$.

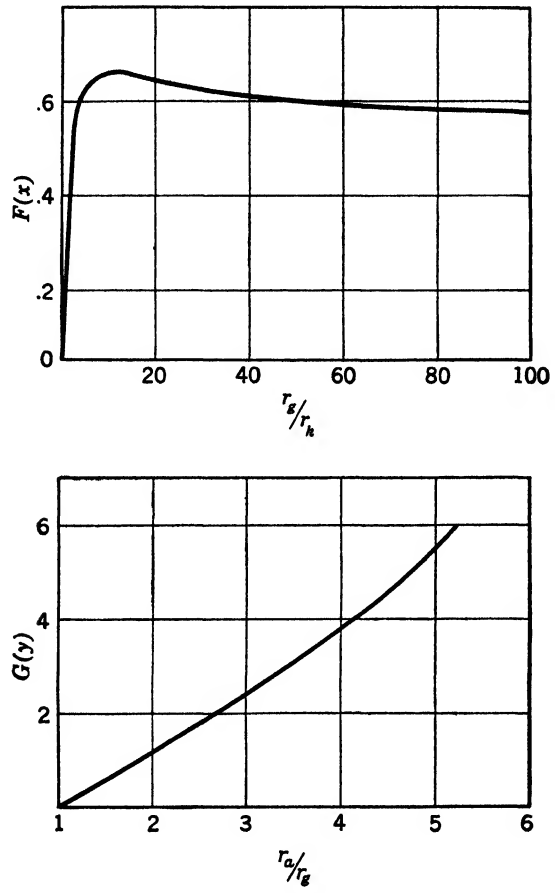


FIG. 15·5 The Scheibe functions $F(x)$ and $G(y)$.

Converted to practical units, equations 15·15, 15·16 and 15·17 become

$$t_a = 6.67 r_g \frac{G(y)}{\sqrt{V_k}} \times 10^{-6} \quad \text{sec} \qquad [15\cdot22]$$

$$t_k = 6.67 r_g \frac{F(x)}{\sqrt{V_k}} \times 10^{-6} \quad \text{sec} \qquad [15\cdot23]$$

$$t_t = 6.67 r_g \frac{G(y) + F(x)}{\sqrt{V_k}} \times 10^{-6} \quad \text{sec} \qquad [15\cdot24]$$

where r_g is the radius of the grid in meters, $V_a = V_k$ is the applied voltage, and $F(x)$ and $G(y)$ have the values shown above.

The strongest oscillations result if $V_a = V_k$ and $t_a = t_k$. This requires that $F(x) = G(y)$. Reasonable conditions which satisfy this requirement are

$$\frac{r_a}{r_g} = 1.6 \qquad \text{and} \qquad \frac{r_g}{r_k} = 50 \tag{15·25}$$

Under these conditions

$$F(x) = G(y) = 0.6 \tag{15·26}$$

and

$$t_a = t_k = \frac{4r_g}{\sqrt{V_k}} \times 10^{-6} \quad \text{sec.} \tag{15·27}$$

If, for example, we choose $r_g = 0.001$ meters (0.1 cm), a reasonable figure, and if $V_k = 100$ volts, we find

$$t_a = t_k = \frac{0.004}{10} \times 10^{-6} = 0.4 \times 10^{-9} \quad \text{sec} \tag{15·28}$$

For one particular mode of oscillation we shall show later that the corresponding frequency is 2500 megacycles.

15·4 Frequency and Wavelength of Oscillation

The frequency of an alternating current is the number of complete cycles described per second and is therefore equal to the reciprocal of the time required for one cycle. This time, called the period, is usually designated T. Thus

$$f = \frac{1}{T} \tag{15·29}$$

The length of the wave produced in free space by a given frequency is often useful. It is understood in the following that the wavelength in free space is meant whenever the term wavelength is used. The wave length is defined by the identity.

$$c = f\lambda \tag{15·30}$$

where c is the velocity of light in meters per second and λ is the free-space wavelength in meters. Combining the two we have

$$\lambda = \frac{c}{f} = cT \tag{15·31}$$

It will be shown that oscillations may exist corresponding to any

of the three transit times previously discussed. Accordingly it is useful to tabulate the frequencies and wavelengths which correspond to these three modes for the plane and cylindrical structures. The values presented in Table 15·1 are deduced from equations 15·10, 15·11, 15·12, 15·22, 15·23, and 15·24 by means of equations 15·2 and 15·3. Since $G(y)$ and $F(x)$ are less than 1 for typical cases it is seen that the wavelengths for cylindrical electrodes are typically less than those for plane electrodes of comparable spacing.

TABLE 15·1

FREQUENCIES AND WAVELENGTHS OF BARKHAUSEN OSCILLATORS

Plane Structure	*Cylindrical Structure*
$\lambda_a = 2000 \dfrac{d_a\sqrt{V_k}}{V_a}$ meters	$\lambda_a = 2000 \dfrac{r_g}{\sqrt{V_k}} G(y)$ meters
$\lambda_k = 2000 \dfrac{d_k}{\sqrt{V_k}}$ meters	$\lambda_k = 2000 \dfrac{r_g}{\sqrt{V_k}} F(x)$ meters
$\lambda_t = 2000 \dfrac{V_a d_k + V_k d_a}{V_a\sqrt{V_k}}$ meters	$\lambda_t = 2000 \dfrac{r_g}{\sqrt{V_k}} [G(y) + F(x)]$ meters
$f_a = 0.15 \dfrac{V_a}{d_a\sqrt{V_k}}$ megacycles/sec	$f_a = 0.15 \dfrac{\sqrt{V_k}}{r_g} \dfrac{1}{G(y)}$ megacycles/sec
$f_k = 0.15 \dfrac{\sqrt{V_k}}{d_k}$ megacycles/sec	$f_k = 0.15 \dfrac{\sqrt{V_k}}{r_g} \dfrac{1}{F(x)}$ megacycles/sec
$f_t = 0.15 \dfrac{V_a\sqrt{V_k}}{V_a d_k + V_k d_a}$ megacycles/sec	$f_t = 0.15 \dfrac{\sqrt{V_k}}{r_g} \dfrac{1}{G(y) + F(x)}$ megacycles/sec
where V_k = grid-cathode potential in volts.	V_k = grid-cathode potential in volts.
V_a = grid-anode potential in volts.	r_g = grid radius in meters.
d_k = grid-cathode distance in meters.	$G(y)$ = given in equation 15·21.
d_a = grid-anode distance in meters.	$F(x)$ = given in equation 15·20.

15·5 Requirements for Sustained Oscillations

Retarding field oscillators usually operate with negligible space charge. Consequently, the emission of electrons from the cathode is random with respect to time and does not vary in synchronism with the period of electronic oscillation about the grid. Thus, though it is true that electronic oscillations occur in any tube with a positive grid and negative plate, it is also true that no effect on an outside circuit is produced unless the random electronic oscillation is somehow turned

into a synchronous oscillation. The mechanism for rendering these oscillations synchronous and thus supplying power to an outside circuit will now be discussed.

Consider the circuit of Fig. 15·6. A physical tuned circuit having a positive value of resistance and a definite value of Q is connected to some device which acts as a negative resistance. It is convenient here

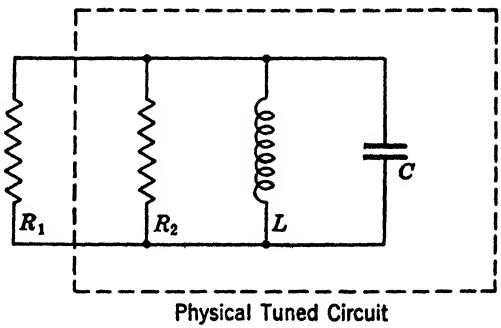


FIG. 15·6 Oscillatory circuit. The resistance R_2 represents all sources of loss in the coil and condenser. The resistance R_1 is negative and represents the action of the vacuum tube. Oscillation builds up if $|R_1| < |R_2|$.

to assume that the combined power dissipation of the coil and condenser is lumped in a single parallel resistance. We are interested in the behavior of the circuit only near the resonant frequency, and therefore no appreciable error need result from this assumption.

The effective resistance R_e in parallel with the coil and condenser is given by the familiar relation

$$R_e = \frac{R_1 R_2}{R_1 + R_2} \qquad [15\cdot32]$$

The denominator of this equation becomes zero and R_e approaches infinity if $R_1 = -R_2$. The effective resistance is positive if $|R_1| > |R_2|$ and is negative if $|R_1| < |R_2|$, R_1 being negative in both cases.

A small transient disturbance disappears as an exponentially decaying sine wave if R_e is positive but increases exponentially if R_e is negative. If all the elements of Fig. 15·6 were linear and R_e negative the oscillatory voltage resulting from an initial transient disturbance would continue to increase exponentially without limit.. Linear negative resistances do not exist, however, and R_e ceases to be negative for some particular level of oscillation. This is the normal oscillating condition.

A negative resistance is a device in which the voltage and current are 180° out of phase. Power is delivered from, rather than received by,

such a device. Thus a negative resistance is able to deliver power to the oscillatory circuit. Accordingly we may expect oscillations to result if a suitable tuned circuit is connected to it.

We shall show that at a suitable frequency the current which flows to the electrodes of a positive grid tube has a component which is 180° out of phase with the voltage. That is, the positive grid tube functions as a negative resistance at this frequency and is therefore capable of producing oscillations.

15·6 Oscillations in the Grid Circuit

Refer to Fig. 15·1, in which the distances d_a and d_k are equal, and assume that no potential difference, alternating or direct, exists between anode and cathode. Let us assume that an alternating voltage of perhaps a few volts is impressed upon the grid (with respect to anode and cathode) and that the frequency is such that one full cycle is described in the time required for an electron to travel from cathode to anode.

Consider an electron which leaves the cathode when the alternating component of the grid voltage is zero and is going from negative to positive. During the entire time that the electron is traveling from cathode to grid the accelerating field is stronger than normal so that the electron arrives at the grid with a velocity higher than that corresponding to the steady voltage V_k. Assuming that the electron does not strike a grid wire it proceeds into the retarding field of the grid-anode space. At the same time the grid voltage reverses so that the retarding field is now of less than normal intensity. As a result this electron will not come to rest before reaching the plate but will strike the plate with some velocity depending upon the alternating voltage applied to the grid. Evidently this energy, which was supplied by the alternating voltage, is dissipated as heat. This electron has then absorbed power from the applied voltage.

Now consider an electron which leaves the cathode one-half cycle later. Because the accelerating field is weaker than normal, it arrives at the grid with less energy than that corresponding to V_k. Again the alternating voltage reverses as the electron passes through the grid plane, so that the retarding field is stronger than normal during the transit toward the plate. The electron is then brought to rest somewhat in front of the plate and returns once more toward the grid. Again, however, the voltage has reversed, so that the field accelerating the electron is less than normal, and the electron returns to the grid with less energy than it had upon leaving it.

With each successive oscillation the electron returns to the grid with a lower energy and moves a shorter distance away from the grid. The

kinetic energy originally derived from the direct voltage V_k is converted into power at high frequency and delivered to the external circuit. Ultimately, of course, the electron strikes a grid wire and is lost. But on the average those electrons which deliver power to the external circuit execute several oscillations while those which absorb power travel only once from cathode to plate. Thus a net positive power output is available. Refer to Fig. 15·7.

This mechanism is known as amplitude selection. Those electrons with unsuitable phase are rejected from the system by the excess of velocity which they acquire in one transit. Those electrons with

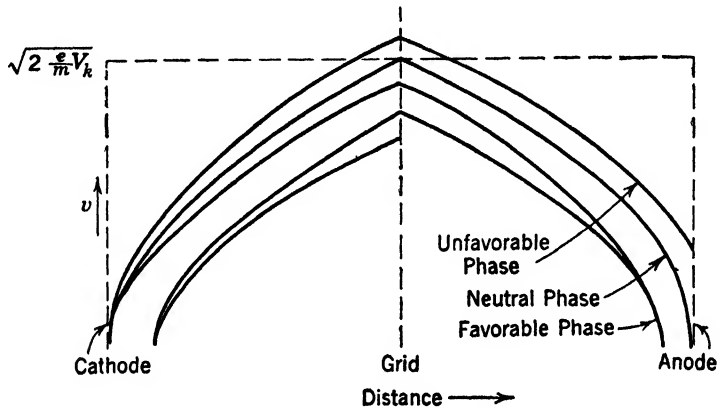


FIG. 15·7 Velocities and transit distances of electrons emitted at various phase angles in a retarding-field tube with a small alternating voltage superimposed upon the steady direct grid voltage.

suitable phase are retained in the system for several oscillations. So far only two specific phase angles have been considered, but it is easy to show that all electrons emitted over one half-cycle produce, to varying degrees, loss of power, and that they are ejected after one transit. Likewise all electrons emitted during the other half-cycle deliver useful output and are retained in the system until captured by the grid.

The frequency specified in this development is such that one full cycle corresponds to the transit time from cathode to plate or from grid to plate to grid. This corresponds to

$$\lambda_k = \lambda_a = 2000\frac{d_k}{\sqrt{V_k}} \quad \text{meters} \quad . \qquad [15\cdot33]$$

from Table 15·1. If, for example, $d_k = 0.001$ meter (0.1 cm) and $V_k = 100$ volts, the generated wavelength is approximately 0.2 meters (20 cm), a very short wave.

15·7 Mechanical Analogy

Figure 15·8 shows a mechanical device which illustrates many of the characteristics of the positive grid oscillations just described. Let us assume that marbles are discharged at a uniform rate from the hopper and roll down the inclined side, across the flexible section, and up the other side. Suppose that the unit is driven as shown so that the driving

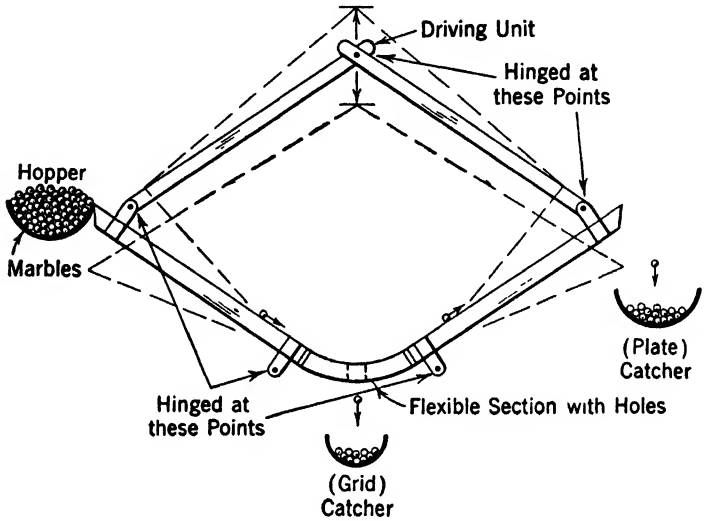


FIG. 15·8 Model illustrating the operation of a positive-grid oscillator.

bar makes one full cycle up and down in the time required for a marble to roll down from the hopper and up the other side. Marbles may be lost from this system either by rolling over the edge into the catcher trough or by falling through a hole in the flexible section.

A marble which emerges from the hopper when the system is as shown by the solid lines, and when the edges are rising, rolls down an incline that is always steeper than normal. During the time that it crosses the flexible section the displacement is reversed and the marble finds the second incline less steep than normal. It is thus able to roll over the edge and escape to the catcher.

A marble which enters the system half a mechanical cycle later finds its original incline always less steep than normal and arrives at the flexible section with a velocity below normal. The second incline is now steeper than normal and the marble is thus unable to reach the far edge. The displacement of the system from normal again passes through zero as the marble reverses and the velocity with which it

again reaches the bottom is still further reduced. This marble oscillates back and forth in the system, losing more and more energy with each cycle. It ultimately is lost through a hole in the flexible section.

The similarity of this device to the electronic one just described is very great. Marbles or electrons which enter the system at times of unfavorable phase are ejected after one full transit. Marbles or electrons which enter the system at times of favorable phase give up energy to the system and remain for several cycles of oscillation. Since as many electrons or marbles enter in favorable as in unfavorable phase the power contributed by those in favorable phase outweighs that absorbed by those of unfavorable phase.

15·8 Phase Selection

The mechanism sometimes called phase selection remains to be considered. Figure 15·7 shows that electrons of favorable phase continually lose velocity because they do work on the external circuit. Accordingly they are able to reach only shorter and shorter distances from the grid. The time required for each oscillation is thereby reduced according to equation 15·11, which may be written

$$t = 6.67 \frac{d}{\sqrt{V}} \times 10^{-6} \quad \text{sec} \qquad [15\cdot34]$$

This relation is now only approximate since it does not include the effect of the alternating voltage. It is a good approximation, however. The potential gradient is still uniform, and therefore d and V in equation 15·34 are reduced by the same factor, as the amplitude of the motion decreases. Accordingly the transit time is proportional to the square root of d.

This shortening of the transit time with each successive oscillation is a serious limitation of the operation of the positive grid oscillator. Electrons making their first oscillation require a lower frequency for optimum performance than is suitable for electrons which have executed several oscillations. The actual output frequency is at best a compromise. Electrons in their first transit lag in phase; electrons which have made several vibrations lead in phase.

The phase shift which results if the alternating component of the grid voltage is large is so great that electrons are transferred from favorable to unfavorable phase in only one or two cycles. This difficulty, fundamental to positive grid oscillators, is an important one. A high efficiency is possible only if the electrons have come almost to rest before they reach the grid and so dissipate very little energy as

heat. Unfortunately the effects of phase selection are such that this condition can never be approached in practice.

The increase of oscillation frequency incident to delivery of external power as indicated by the progressive phase shift is of the greatest interest, both practically and theoretically: practically, because we usually wish to abstract the greatest possible power output from the oscillator; theoretically because it explains the Barkhausen-Kurz, and Gill-Morrell modes of oscillation previously discussed and shows why many of the observed phenomena occur.

The optimum output voltage and therefore the optimum load impedance is something of a compromise. Too high a voltage shifts the oscillation amplitude and therefore the electronic phase too rapidly for best results. Too low a voltage does not allow a sufficient abstraction of energy from each electron before it strikes a grid wire and is lost.

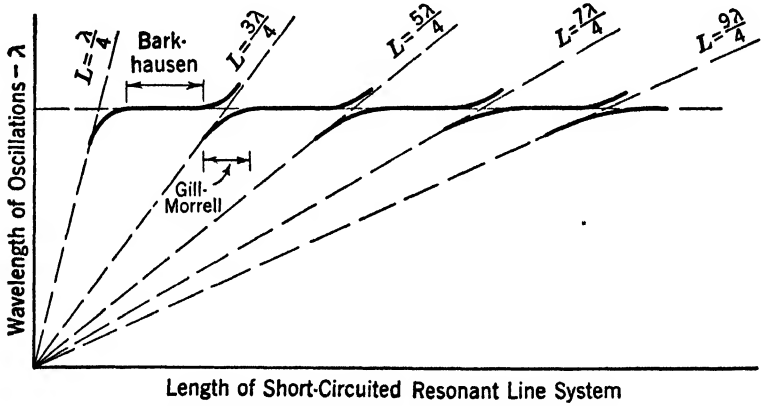


FIG. 15·9 Effect of the load circuit in determining the mode of oscillation of a positive grid tube.

It is worthy of mention that the mechanical model described in section 15·7 shows the effect of phase selection. Marbles of favorable phase remain in the system for several cycles and roll through successively smaller distances. As the distances traveled becomes less, the time required for each round trip is also decreased.

In the preceding discussion it was assumed only that an alternating voltage exists in the grid circuit, and it was shown that the alternating current which results is more than 90° out of phase with the voltage. The external conditions are satisfied by any form of impedance since the reference voltage is reversed and the phase shift is therefore less than 90°. Accordingly, power is delivered to any external impedance which is connected to the tube. Oscillations whose frequency is independent

of the load are then produced as originally described by Barkhausen and Kurz.

If, however, the external load is a tuned circuit of suitable impedance the power delivered is considerably increased. The frequency tends to be increased by the process of phase selection and depends upon the tuning of the external circuit. These are the oscillations described by Gill and Morrell.

Figure 15·9 is characteristic of the performance of a practical positive grid tube. When the lines are of such a length as to simulate a parallel resonant circuit the impedance is high and a considerable output voltage and power result. The wavelength is modified as shown, and the oscillations are of the Gill-Morrell mode. When the lines are detuned the impedance and voltage are low and Barkhausen-Kurz oscillations result.

15·9 Oscillations in the Plate-Cathode Circuit

In section 15·6 it was shown that oscillations can exist in a plane parallel triode having equal spacings and no cathode-anode voltage. The oscillations appear from grid to cathode (or anode) and are of quite high frequency. It is clear in reviewing the performance that the success of the arrangement depends upon the fact that cathode and anode are at the same potential and that the transit times from grid to cathode and from grid to anode are equal. Serious failure of the equality of transit times is fatal to this mode of oscillation since the impulses cannot become synchronized. Equality of plate and cathode voltage, although very desirable, is not essential since the amplitude selection mentioned before can take place at the cathode after one full cycle of electron oscillation. Electrons which leave the cathode in unfavorable phase are accelerated during one full excursion from cathode to plate and back, and are then lost at the cathode. Electrons which leave the cathode in favorable phase continue in their oscillation, as shown above, until they strike a grid wire and are lost.

Reference to Fig. 15·2*a* shows that the grid-cathode and grid-anode transit times are equal in the plane parallel case if $d_a/V_a = d_k/V_k$. Many experimenters have reported strong oscillations with large negative plate voltages, probably in fulfillment of essentially this condition. Such oscillations exist provided only that some impedance is present in the grid lead to cause a voltage drop and that the transit times are suitable. Power is delivered to the external circuit if an appropriate resistive load such as a tuned circuit is applied.

A different sort of oscillation results if the load is applied between cathode and plate. An electron moving from cathode towards the

plate constitutes an elementary electric current whose direction is opposite to the motion because of our convention as to current flow. Evidently the motion from cathode toward the plate causes the plate to become momentarily negative and the cathode positive with respect to the bias values. This effect is greatest when the electron is in mid flight, that is, in the grid plane.*

Let us assume that the grid is grounded; that is, we use it as our reference of potential. Further, let the load impedance be connected directly between cathode and plate so that an increment of grid-cathode voltage is matched by an equal and opposite increment of grid-plate voltage.

This oscillation is explained in much the same way as the other. Assume equal alternating voltages 180° out of phase to be connected from grid to cathode and from grid to plate as shown in Fig. 15·10*a*. Let the frequency be such that one cycle corresponds to a full round trip of the electron, half as high as the frequency previously assumed. Consider an electron which leaves the cathode at an instant when the applied alternating voltage is zero and when the cathode potential is changing in a positive direction. This electron will reach the grid 90° later when the cathode potential is a maximum (grid to cathode potential a minimum), and its velocity will be correspondingly reduced. During the following quarter cycle the electron will be decclerated by the retarding field of the plate, which is by our assumption at a maximum, because the plate voltage is now a minimum or the grid-plate voltage a maximum. The third quarter cycle finds the electron returning toward the grid with the plate at its positive maximum so that the electron is again slowed down. During the last quarter cycle the electron approaches the cathode against a retarding field that is again above its normal value. This electron does work upon the external system and will continue to do so through several oscillations. This is summarized graphically in Figs. 15·10*b*, *c*, *d*.

Now consider an electron which leaves the cathode 180° later. This electron arrives at the grid with greater than normal velocity. Since the plate is now swinging in a positive direction the additional velocity will help the electron to follow through to the plate and remove itself from the system.

Again we generalize to say that all electrons emitted over one half cycle are able to do work upon the external circuit until they are captured by the grid. The electrons emitted during the other half cycle are quickly removed from the system at the grid or plate.

It is important to note that the frequency generated in this case

* Refer to Chapter 12.

is only half as great as that developed when the grid voltage is allowed to oscillate. This disadvantage is often compensated for by superior stability and power output.

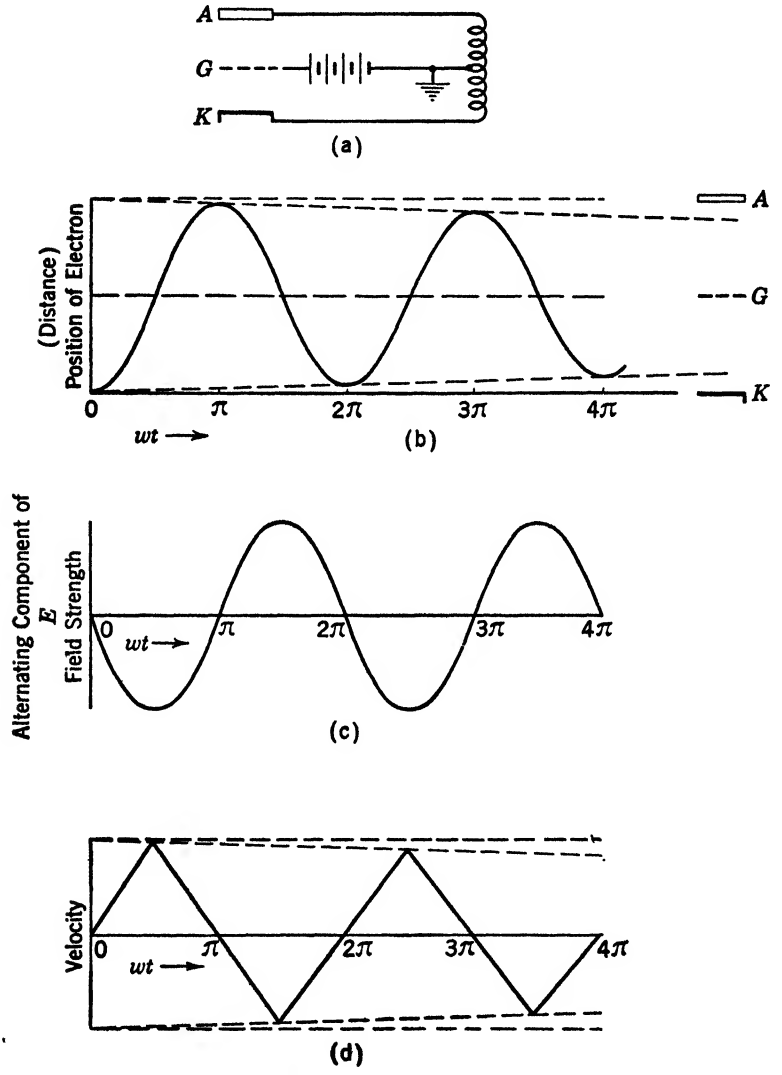


FIG. 15·10 Behavior of an electron of favorable phase in a retarding-field oscillator. Load circuit connected between plate and cathode.

The mechanical model previously described is easily modified to suit this condition. The two inclined sides now bear a fixed angle to each

other, and the trough so formed is merely tipped from side to side at half the frequency previously used.

Marbles which emerge from the hopper when the trough is level and the hopper is moving upward are in unfavorable phase and arrive at the bottom of the trough with relatively large speeds by means of which they overrun the far edge and escape to the catcher.

Marbles which emerge from the hopper a half cycle later give up work to the system and remain for several cycles.

It is interesting to note that both mechanical models show the effect of phase selection. As the amplitude of the mechanical oscillation becomes smaller the natural period becomes shorter and the marbles originally in favorable phase have a tendency to advance into an unfavorable phase. This behavior is in marked distinction to that of the pendulum and most electrical oscillations, where the very nature of the system makes the period essentially independent of the amplitude.

15·10 Oscillations in Plate Circuit or Cathode Circuit

Oscillations may also be produced if the grid and cathode have no alternating voltage and the anode is allowed to swing. The demonstration that such oscillations are self-sustaining is similar to the previous ones. All electrons which leave the cathode and pass through the grid have the same velocity at the grid plane since no alternating field exists in the grid-cathode space. Let us assume that the plate voltage oscillates at a frequency such that one full cycle is executed in the time required for the electron to travel from grid to plate and return.

An electron which leaves the grid mesh at a time when the alternating plate voltage is zero and going positive is retarded by a less than normal field during its transit to the plate, and therefore strikes the plate and is lost. An electron which leaves the grid mesh a half cycle later is retarded by a more than normal force and therefore is unable to reach the plate and comes to rest somewhere near the plate. The alternating component of the plate voltage now reverses so that the electron is less than normally accelerated in the return to the grid. The operation outlined above is adequate to support stable oscillation, provided that virtually all the electrons returned to the grid are lost, for the electrons of favorable phase react upon the plate twice as long as those of unfavorable phase.

Actually, however, a considerable number of the electrons returned from plate to grid pass into the grid-cathode region and again return through the grid mesh toward the plate. These will be in a phase to aid the oscillation if the grid-cathode transit time is equal to, or is an integral multiple of, the grid-anode transit time. If this relation is not

approximately correct the number of electrons in unfavorable phase at the plate is increased and the chance of oscillation is greatly diminished.

Here, as before, it is desirable but not absolutely necessary that no steady voltage difference shall exist from cathode to anode. The selection of favorable *vs.* unfavorable phase is much better effected at the plate than at the cathode. The generated wavelength is given by the equation of Table 15·1, when $V_k = V_a$

$$\lambda_a = 2000 \frac{d_a}{\sqrt{V_k}} \quad \text{meters} \qquad [15\cdot35]$$

with the provision that

$$\lambda_k = 2000 \frac{d_k}{\sqrt{V_k}} = n\lambda_a \qquad [15\cdot36]$$

where n is some small integer.

An oscillation extremely similar to the one just described may be produced by allowing only the cathode voltage to oscillate. Again electrons of unfavorable phase are best eliminated at the plate. The wavelength is now given by the relation

$$\lambda_k = 2000 \frac{d_k}{\sqrt{V_k}} \quad \text{meters} \qquad [15\cdot37]$$

with the provision that

$$\lambda_a = 2000 \frac{d_a}{\sqrt{V_k}} = n\lambda_k \qquad [15\cdot38]$$

where n is again a small integer and $V_k = V_a$.

It is seen that all these different modes of oscillation may exist in either cylindrical or planar tubes. The processes of amplitude selection and phase selection are not essentially affected, and the wavelengths are not greatly different. In general the cylindrical arrangement is more convenient in manufacture, and the fact that it yields somewhat higher frequencies than the plane structure for given spacings is in its favor.

The mode of oscillation in which only the plate voltage swings is attractive in that a small grid plate spacing may be achieved without severe restriction of the grid area. A very high resulting frequency is indicated. For $d = 0.0005$ meter (0.5 mm) and $V = 100$ volts the value of λ is approximately 0.1 meter (10 cm).

The mechanical model corresponding to these last two modes of oscillation has one side permanently fixed. The swinging side oscillates

at the frequency given in the initial description of the model. The process of eliminating the marbles which are introduced in unfavorable phase is accomplished as before, and the entire operation is very similar to that previously described.

15·11 Diode Oscillations

It has been known for some time that an electronic diode, under appropriate conditions, can generate oscillations. Probably the first to point out this possibility was W. E. Benham.* As with triodes the oscillation may exist in either cylindrical or plane structures, and again, at least two modes of oscillation can exist.

Llewellyn and Bowen† in a very interesting article describe a plane-parallel diode which generates approximately 300 milliwatts of power at $\lambda = 10$ cm. The maximum efficiency is 0.23 per cent. The possibility of such oscillations in a space-charge-saturated diode is of the greatest interest. Unfortunately the mechanical structure is relatively difficult, and the low efficiency achieved does not make this an attractive generator for practical applications.

The diode of Llewellyn and Bowen employs an oxide-coated plane cathode of the separately heated variety. The average current is limited by anode voltage rather than by cathode temperature, and the oscillation comes about purely as the result of electron transit time. There is, however, no such vibration of free electrons as constitutes the "electron dance" of the Barkhausen oscillator. Each electron passes directly from cathode to anode, and the negative resistance which causes the oscillation arises from the fact that any superimposed alternating voltage results in an alternating component of current which is more than 90° out of phase with the voltage. The greatest effective series negative resistance exists when the frequency is such as to cause a transit angle of approximately $1\frac{1}{4}$ cycles, although negative resistances also occur for $2\frac{1}{4}$, $3\frac{1}{4}$, etc., cycles.

Mr. J. S. McPetrie‡ describes a very different kind of diode in which the anode is a relatively small rod concentric with a larger cylindrical cathode. This closely resembles a Barkhausen oscillator since electrons leaving the cathode may easily miss the slender anode and oscillate about it. This construction lends itself well to a shielded structure hav-

* W. E. Benham, "Theory of the Internal Action of Thermionic System at Moderately High Frequencies," *Phil. Mag.*, March, 1928, and *Phil. Mag. Supplement*, Vol. II, February, 1931.

† F. B. Llewellyn and A. E. Bowen, "The Production of Ultra-High-Frequency Oscillations by Means of Diodes," *Bell System Tech. J.*, **18**, 280, April, 1939.

‡ J. S. McPetrie, "Experiments with Inverted Diodes Having Various Filament Cathodes," *Phil. Mag.*, **19**, 501, 1935.

ing low radiation loss, a feature of the greatest importance at these frequencies. It seems probable that some such generator could be built as part of a coaxial transmission line, a very desirable arrangement.

15·12 The Spiral Grid Tube

The Barkhausen oscillator as ordinarily constructed is not satisfactory for frequencies much above 600 megacycles per second. In order to obtain higher frequencies the spacings must be reduced and the voltages increased. Both steps are disadvantageous in that a smaller structure is called upon to dissipate a larger amount of power. The spiral-grid tube as described by A. G. Clavier* is useful to much higher frequencies, in the order of 5000 megacycles.

In this tube the grid takes the form of an open (unshorted) helix surrounding an axial filament. The plate, a concentric cylinder, is ordinarily supported at the middle and is maintained at a high negative voltage with respect to the filament. The output is taken from the two ends of the grid, which is maintained at the usual high positive potential. The generated frequency is much higher than that predicted by the simple theory of free electron oscillations. Experiment shows that it is dependent upon the resonant properties of the grid and associated load, as well as upon the applied voltages. Apparently the exciting mechanism is dependent upon transit time as in the diode proposed by Benham and described by Llewellyn and Bowen.

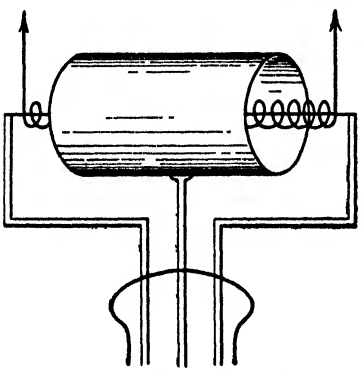


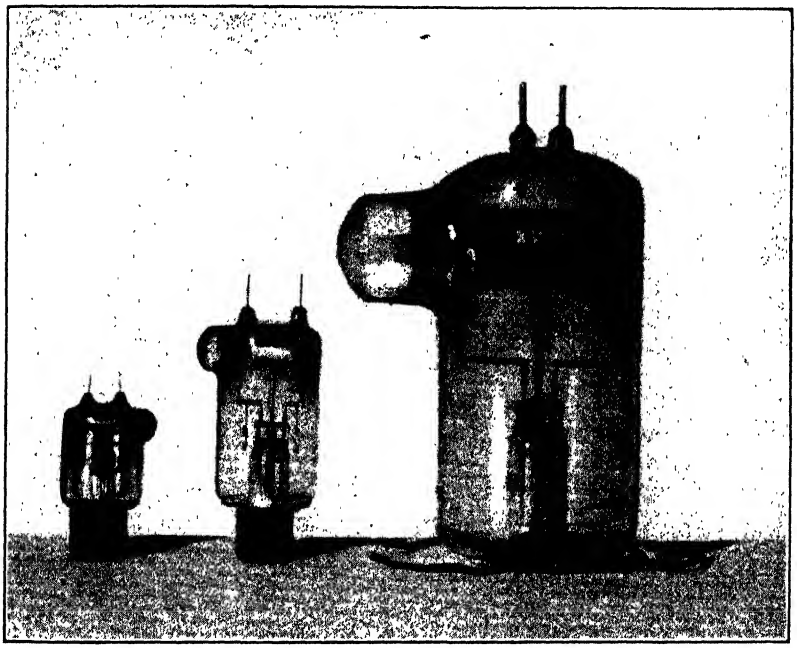
FIG. 15·11 Electrode configuration of a spiral grid retarding field oscillator tube.

Figure 15·11 shows the construction of one of these tubes. The filament, usually of pure tungsten, is supported in tension between two end supports. The grid, formed of tantalum or some other highly refractory metal, is an open spiral concentric with the filament. The ends of the grid are supported by two relatively heavy rods which form the output terminals. They pass directly through the hard-glass envelope to the load circuit, which is usually a parallel-wire or Lecher system. These rods, then, are themselves part of the output line. The plate, which forms no part of the high-frequency circuit, is supported at the middle by a single rod or bar. Thus the entire structure is very

* A. G. Clavier, "The Production and Utilization of Micro-Rays," *Electrical Communication*, 12, 3, 1933.

symmetrical and simple in comparison with conventional receiving tubes.

Fay and Samuel* describe extensive tests on some seventy different experimental tubes of this type. The frequency range from 500 to 2000 megacycles was covered. The most significant result of this work is a relation between the expanded length of the grid coil and the frequency



(Fay and Samuel, courtesy of IRE)

FIG. 15·12 Spiral grid retarding field oscillator tubes.

at which optimum oscillating conditions exist. The ratio of expanded length of the grid to the wavelength of optimum performance was found to be closely equal to 1.25 with few significant deviations. The ratio of plate diameter to grid diameter was approximately 2.5. In all tests a rather high negative voltage was applied to the plate. And in all cases the efficiency was low, in the order of 0.5 to 1 per cent. Three of their tubes are shown in Fig. 15·12.

Spiral-grid Barkhausen oscillators are subject to a certain amount of trouble due to parasitic oscillations. That is, undesired oscillations of a frequency far removed from that of the desired oscillation may result

* C. E. Fay and A. L. Samuel, "Vacuum Tubes for Generating Frequencies above One Hundred Megacycles," *Proc. IRE*, 23, 199, 1935.

if the load tuning or applied voltages are disturbed by a small amount. Data illustrating this behavior in a typical spiral-grid tube are presented in Table 15·2. Here V_g is the voltage between grid and cathode. The voltage between cathode and plate is designated V_p. The wavelength of oscillation is designated λ, and the current to the grid I_g. Because the plate is biased negative with respect to the cathode it draws no current.

TABLE 15·2

λ(cm)	V_g (volts)	I_g (ma)	V_p (volts)	$\lambda^2 V_g \times 10^{-4}$
22.6	470	70	−75	24 0
22.8	450	70	−75	23.4
32.1	320	100	−15	32.9
34.5	280	100	−20	33.3
36.0	220	65	−17	28.5
88.1	155	60	−10	120.5
94.7	135	50	−2.5	121.0
97.0	125	50	−2.5	117.5

If operation occurred in a single mode the numbers in the last column would all be equal, as shown by the wavelength equations of Table 15·1. Actually the numbers group around 24, 32, and 120, indicating at least three modes of operation. The particular mode at which oscillation occurs depends upon the applied voltages and the circuit tuning.

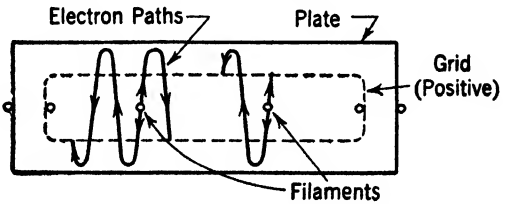


FIG. 15·13 Section through an ordinary triode having a V type wire filament. Approximate electron paths are indicated.

15·13 The Backing Plate Tube

Many workers have noted that electronic oscillations are not produced in filamentary tubes of the ordinary "flat-type" construction. The behavior of such tubes is readily investigated by means of Fig. 15·13. It is seen that electrons which succeed in passing through the grid mesh and are returned by a slightly negative plate potential must pass twice more through the grid before they again experience a strong retarding field. This is true because the filamentary cathode is of such small area that it cannot produce a zero potential region of substantial size in the center of the grid rectangle. Consequently the times of flight

are rather different from any that have been computed. It is clear that oscillations in such a structure could exist only if the plate were run at a slightly negative bias so as to capture only electrons of unfavorable phase, and if the spacings of the tube were rather critically chosen. The required transit through both sides of the grid mesh is particularly unfavorable to oscillation of electrons since it more than doubles the chance of their capture by the grid.

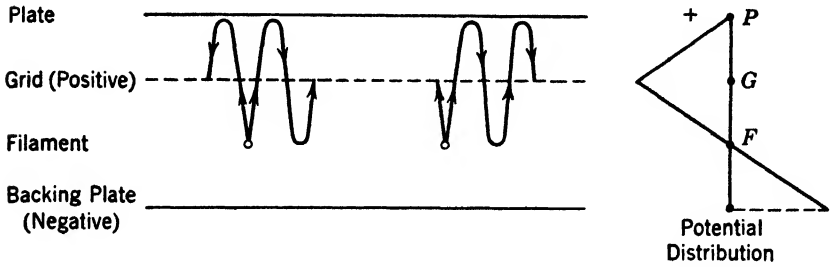


FIG. 15·14 Arrangement of a retarding field oscillator employing directly heated filaments and a backing plate.

Thompson and Zottu* describe retarding field oscillations in a plane tube configuration shown in Fig. 15·14. This construction offers several interesting features. One of the most notable of these is the possibility of operation with normal filament temperature. This is accomplished by applying a somewhat larger negative potential to the backing plate than is indicated in Fig. 15·14. The potential in the plane of the filament is now negative so that a field exists at each filament wire which opposes the emission of electrons. The filament is operated at normal temperature, exactly as in conventional negative grid tubes, and the total current flow is controlled by the potential of the backing plate. The normal plate is operated at zero or a slightly negative potential and serves to capture electrons which leave the filament in a phase unfavorable to oscillation.

An experimental tube of this design consisted of backing plate, two-strand thoriated tungsten filament, parallel-wire grid, and plate. Each of the three electrode spacings was 1.3 mm. The oscillation occurred at a wavelength of 45 to 49 cm for a positive grid voltage of 250 volts and a negative bias of 10 volts on the two plates. The output was taken from the two flat plates, the grid and filament electrodes being fed through high-frequency choke coils. The wavelength observed for this mode of oscillation is in excellent accord with the equation for λ_t in

* B. J. Thompson and P. D. Zottu, "An Electron Oscillator with Plane Electrodes," *Proc. IRE*, 22, 1374, 1934.

Table 15·1. Under the above conditions the output power was 0.08 watt, the input current 120 milliamperes, the efficiency 0.37 per cent.

With 10 volts negative bias applied to the backing plate it was found that the operation was virtually independent of the filament voltage, indicating a condition of complete space-charge saturation.

Another tube of this general type is illustrated in Fig. 15.15. In this structure the backing plate and filament lie in the same plane and the grid and plate appear on both sides. This construction has certain mechanical advantages but does not lend itself quite so well to simple calculation.

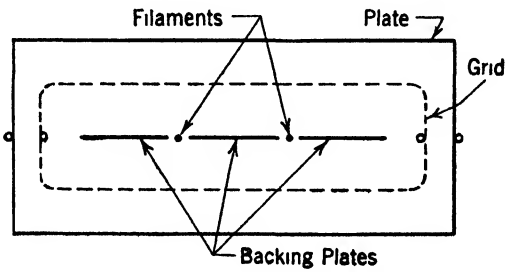


FIG. 15·15 Section through a triode having directly heated filament and backing plates.

It is clear that these plane-parallel tubes are capable of oscillating in any of the modes that have been discussed. The connection used in the experiment described (backing plate to plate) gives rise to only one frequency as was verified by Thompson and Zottu. No parasitic or harmonic modes were discovered in their work with these tubes. Thompson and Zottu report a tube which oscillates at a wavelength of 9.5 cm with 150 volts applied. The total spacing from plate to backing plate in this tube is less than 1 mm. Because of the simplicity of the operation and because of the relatively large grid surface offered by these plane tubes they appear quite suitable for hyper-frequency work.

15·14 Multigrid Tubes

The dynatron oscillator first became a practical engineering tool with the introduction of the screen-grid tube, in spite of the fact that the basic principle had been known for years. The addition of the second grid made it possible to control the total cathode current by means of a control bias rather than by the filament temperature.

The retarding field oscillator bears a considerable resemblance to the dynatron in that both tubes operate with high positive grid bias and

current. The desirability of a grid to control the space current is at least as great in this tube as in the dynatron. A cylindrical tetrode offers several attractive possibilities as a positive-grid oscillator. The inner grid readily controls the total space current, thereby offering easy means of modulation. The cathode structure may be conventional since operation is in the normal space-charge condition. Close electrode spacing may be achieved without making the grid excessively small. Therefore, considerable power output may be expected at high fre-

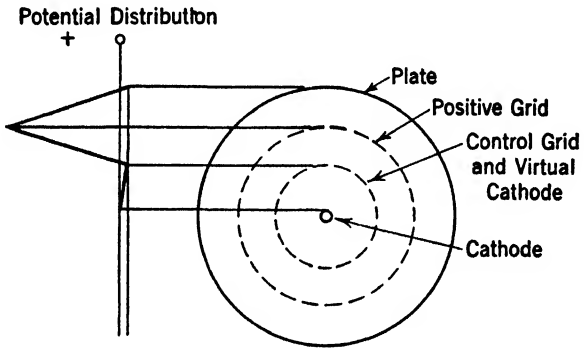


FIG. 15·16 Cylindrical tetrode used as a retarding-field oscillator.

quencies. Figure 15·16 shows the cross section of such a tube together with a potential distribution curve. Because the radius of curvature is large the field distributions are almost uniform and the electrode spacings may be equal.

Evidently the tetrode may function as a spiral-grid oscillator as well as in the conventional manner. In this case the inner or control grid should be carefully shorted so that it will not contribute a tendency to operate at other, undesired frequencies. The divided plate arrangement, described below, is likewise possible.

Hamburger* describes experiments performed upon a commercial triple-grid power tube of cylindrical structure. He used a type 89 tube and produced wavelengths in the order of 140 cm. Relatively good efficiencies and freedom from parasitic modes of oscillation were observed. Unfortunately the nature of his tuning elements and of the applied voltages makes it difficult to interpret the exact mode of oscillation achieved. It is seen, however, that oscillations of Barkhausen-Kurz type are not confined to triodes.

* F. L. Hamburger, "Electron Oscillations with a Triple Grid Tube," *Proc. IRE*, 22, 79, 1934.

15·15 The Divided Plate Oscillator

An interesting and important variant of the spiral-grid tube is presented by W. D. Hershberger.* This tube, which is illustrated in Fig. 15·17, differs from the normal spiral-grid tube only in that the plate is divided into three equal sections by transverse cuts. Two important advantages are made possible by this division. The various sections of the plate may be operated at different direct potentials, thereby improving the efficiency. Also resonant circuits may be connected between them, improving the frequency stability.

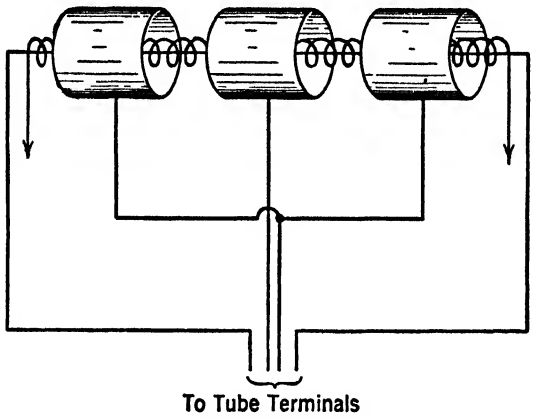


FIG. 15·17 Spiral grid oscillator tube with divided plate. The spacing between the plate sections is shown magnified for the sake of clarity.

The operation of this tube is not essentially different from that of the normal spiral-grid tube; in fact, it is essentially an idealization of that behavior. The central portion of the grid and plate contribute little to the operation, while the end sections oscillate in a pure push-pull or symmetrical fashion at or very near the natural frequency of the grid spiral. Since the central section of the structure does not contribute to the useful operation it should be rendered inactive in order to avoid unnecessary power dissipation. This is accomplished by applying a suitable negative bias to the central portion of the plate. When this voltage is properly chosen the central grid section draws little or no current. The electrostatic field pattern which results tends to crowd the current toward the outer portions of the end sections where it is most effective.

* W. D. Hershberger, "Modes of Oscillation in Barkhausen-Kurz Tubes," *Proc. IRE*, **24**, 964, 1936.

A very marked improvement in performance results if suitable resonant lines are connected to the two plate sections as well as to the grid ends and if these lines are coupled together. It is found that the plate line is several times as effective as the grid line in setting the frequency of oscillation and that the frequency is virtually independent of the grid voltage. In one test, the frequency was observed to change by only 0.1 per cent for a 10 per cent change of grid voltage. Since high-frequency stability is very important and is generally difficult to achieve at these frequencies this feature is of great interest.

Another feature of this arrangement is the freedom from parasitic or spurious modes of oscillation. With the grid and plate circuits properly tuned only one frequency of oscillation is produced. Wide deviations of grid, plate, or filament voltages affect the output but do not introduce new frequencies or seriously affect the frequency in question. This again is in sharp contrast to ordinary Barkhausen oscillators, which are relatively sensitive to the electrode voltages.

The efficiency to be expected from this arrangement is somewhat better than that characteristic of the standard spiral-grid tube. The improved efficiency results from the large negative bias applied to the central portion of the plate. Thus the electron current is concentrated at the ends of the tubes where the maximum voltage swings exist. An efficiency of 1 to 2 per cent is characteristic of well-built tubes of this design.

PROBLEMS

15·1 A plane parallel triode is to be used as a positive grid oscillator. The grid-cathode spacing is equal to the grid-plate spacing and is 8 mm. The cathode and plate are at the same potential, and the grid is biased positive by 150 volts. Calculate the frequencies which may be generated by this tube oscillating in the several modes described.

15·2 A plane parallel triode is to be used as a positive grid oscillator. The grid-cathode spacing is 8 mm; the grid-plate spacing is 16 mm. The cathode and plate are at the same potential, and the grid is biased positive by 150 volts. Calculate the frequencies which may be generated by this tube oscillating in the several modes in which oscillation is possible.

15·3 The tube of problem 15·2 must be used in the mode described in section 15·6. If the grid-cathode voltage is 150 volts, what grid-plate voltage is required? Calculate the frequency to be expected, and explain the mechanism of amplitude selection.

15·4 A cylindrical triode is to serve as a positive grid oscillator. The cathode radius is 0.5 mm, the grid radius is 2 mm, and the plate radius is 3 mm. The plate and cathode are at the same potential, which is 200 volts negative with respect to the grid. Calculate the frequencies which may be produced in the various modes of oscillation which are possible.

15·5 How large must the radius of the plate be made if the grid-cathode and grid-plate transit times of the tube of problem 15·4 are to be equal?

15·6 In the so-called beam power tube the wires of the screen grid are aligned with those of the control grid so that electrons tend to pass between the screen-grid wires rather than striking them. In some cases more than 90 per cent of the electrons pass through the screen grid, less than 10 per cent being captured. Discuss a tetrode built in this way as a positive grid oscillator.

15·7 Two similar positive grid oscillators are to be connected in parallel so as to secure additional power output. Is such a mode of operation likely to be successful, and if not why not? What if any precautions as to phase relation need be taken?

15·8 The grid of a spiral grid oscillator consists of 12 turns of wire. The helix so formed has a mean radius of 3 mm and a pitch of 2 mm per turn. At what frequency is optimum performance to be expected?

15·9 Discuss the production of oscillations in a diode due to transit time effects. Does it seem probable that such oscillators will be of commercial importance?

15·10 A plane parallel backing plate oscillator tube has a plate 3 mm from the grid. The grid is 2 mm from the plane in which the filament wires lie, and the backing plate is 1 mm from the filament plane. The filament and normal plate are directly connected, and the grid is 100 volts positive with respect to these. At what potential should the backing plate be operated? What frequencies may be generated by the tube with these potentials applied?

15·11 It is desired to bias the plate of the tube of problem 15·10 negatively so as to secure equality of grid-cathode and grid-anode transit times. Discuss the problem of amplitude selection under these conditions. Are oscillations likely to occur?

15·12 It is possible, within limits, to produce amplitude modulation of a positive grid oscillator by applying the signal voltage in series with the anode. Sketch a curve of output versus plate voltage which corresponds to linear modulation. Explain how this curve might occur in terms of amplitude selection.

15·13 There is a considerable tendency for the amplitude modulation produced as described in problem 15·12 to be accompanied by undesired frequency modulation. Discuss the cause of this effect in terms of transit time. How may this effect be eliminated or balanced out?

15·14 In the tube of problem 15·1 plot the electron velocity as a function of time and of distance, assuming that no oscillation exists. Also consider an electron of most unfavorable phase and a tank circuit in series with the grid across which an oscillation of 5 volts peak exists. How much energy does such an electron absorb from the tank in its passage from cathode to anode?

15·15 In problem 15·14 assume that each electron of favorable phase makes $3\frac{1}{2}$ complete transits from cathode to plate while those of unfavorable phase make only one. If the total emission current is 5 milliamperes, calculate the power output, assuming that electrons emitted in phase intermediate between most favorable and least favorable give or absorb power according to the cosine of the angle of deviation from the angle of most favorable phase.

15·16 Using the result of problem 15·14, calculate the amplitude of and time required for each of the first four complete cathode-plate or plate-cathode transits. Interpret in terms of frequency.

15·17 Explain in terms of phase selection how a given tube with fixed voltages and tuning may oscillate in either a Barkhausen-Kurz or Gill-Morrell mode. Which is associated with the higher frequency? The higher output? The natural period of the tank? The natural period of the free electronic oscillation?

15·18 Is it possible to secure the advantages of the divided-plate, spiral-grid oscillator without using the spiral-grid structure? Discuss.

15·19 The positive grid oscillator has been employed as a special kind of high-frequency detector. Explain this operation in general terms and in terms of actual voltages and currents. Assume that the tank circuit is in series with the incoming modulated voltage and the bias battery.

CHAPTER 16

THE MAGNETRON

16·1 Introduction

A vacuum tube in which the behavior is dependent upon a magnetic as well as an electric field is known as a magnetron.* The magnetron as usually constructed is a diode consisting of a cylindrical anode and an axial filamentary cathode. The magnetic field is relatively strong and is directed parallel or nearly parallel to the filament. The electric field is radial and is therefore everywhere normal to the magnetic field. If the plate is divided by planes parallel to the filament, the resulting tube is referred to as a split-anode magnetron. See Fig. 16·1.

(*Kilgore, Courtesy of J. Applied Physics*)

FIG. 16·1 Four-segment split anode magnetron.

Hull in 1921 first published a mathematical analysis of such an arrangement, presenting experimental curves in confirmation of his results.† Since that time a vast store of experimental data has been accumulated, and much discussion of the theory has ensued. Although certain details of the performance remain obscure, the general operation of the magnetron is now well understood.

The magnetron is important chiefly as a generator of powerful oscillations at hyper and ultra-high frequencies. The tube itself is of simple and compact structure, and the overall efficiency is better than that of comparable generators. The shortest continuous waves yet recorded, in the order of several millimeters, have been produced with a magnetron of the split-anode type.

16·2 The Magnetic Field

A relatively strong magnetic field is necessary for proper operation of the magnetron, particularly if very high frequencies are to be produced.

* Actually the electron motion is controlled by the magnetic induction rather than the magnetic field. Since the permeability is necessarily μ_0, that of free space, no numerical errors are likely to arise.

† A. W. Hull, *Phys. Rev.*, **18**, 31, July, 1921.

Approximately 100,000 ampere-turns per meter (1200 gauss) is appropriate for a wavelength of 0.1 meter (10 cm). For field strengths up to about 10,000 ampere-turns per meter it is practical to use simple air-core solenoids of one or several layers of wire. A coil 2 to 3 inches in diameter and 6 to 10 inches long creates a very uniform magnetic field at its central portion and is light and easy to construct. The heat which may be dissipated by such a structure sets an upper limit to the field that can be obtained in this way.

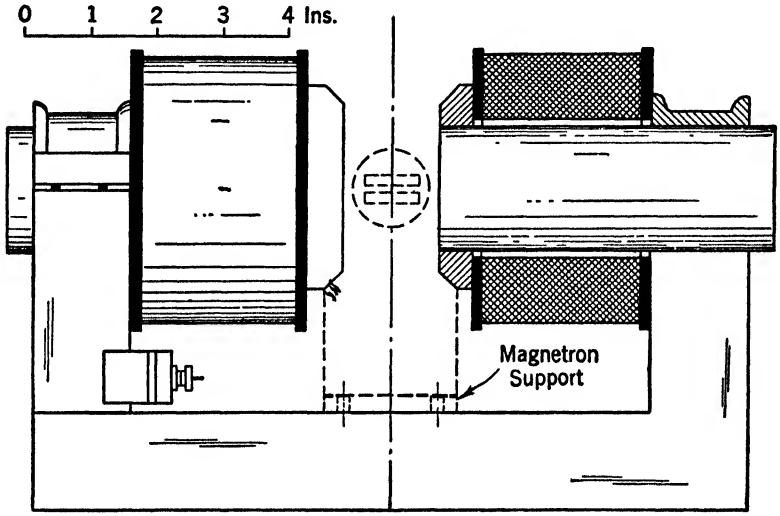


FIG. 16·2 Electromagnet for use with a magnetron.

For small tubes in applications requiring fixed tuning it is convenient and practical to use permanent magnets. Some modern magnetic materials such as Alnico have a high coercive force so that a strong field may be maintained across a relatively long gap without an excessively heavy or large magnet.

For the highest field strengths or for convenience in experimental work it is usually best to use some form of electromagnet. An electromagnet having a cross section of approximately 4 square inches, an air gap of 2 inches, and a total winding of 2000 turns is appropriate. With 5 amperes current, such a magnet (Fig. 16·2) produces a field of approximately 200,000 ampere-turns per meter in the air gap.

16·3 The Electron Motion with Steady Fields

Consider the structure of Fig. 16·3 in which the magnetic induction B is directed out of the paper. Let V_a be the potential difference between

cathode and anode and V the potential existing at some point a distance r from the common axis. It is assumed that the electrons leave the cathode with zero velocity.

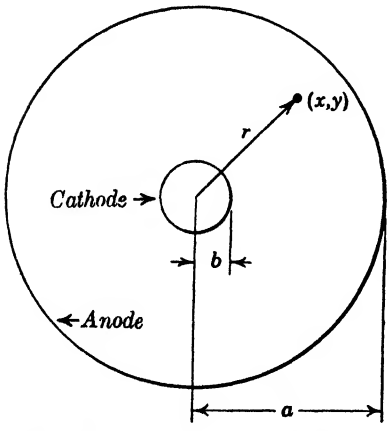


FIG. 16·3 Section of a cylindrical magnetron.

An electron emitted from the cathode finds itself in a strong electric field and is accelerated radially toward the plate. A chain of such electrons, however, is equivalent to an electric current flowing inward from the plate and, as such, experiences a force in the magnetic field. It may be shown that each individual electron experiences a force equal to veB, where v is the velocity, e is the electronic charge, and B is the magnetic induction.

Because of the symmetry of the structure the space-charge density and the potential V are functions only of the radius r. The electron velocity is a function only of r and t. An electron in flight is acted upon by three separate forces. The electric field $E = -\partial V/\partial r$ exerts a

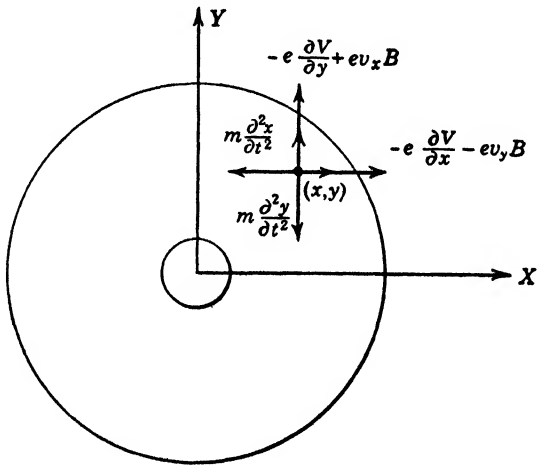


FIG. 16·4 Forces acting upon an electron in a cylindrical magnetron.

radial force outward toward the plate. The magnetic induction exerts a force veB at right angles to the direction of motion. Inertia exerts a

force opposite to the instantaneous acceleration produced by the other two forces which is equal to $m(\partial v/\partial t)$, where m is the mass of the electron. The equations which express the equilibrium of these forces in terms of rectangular coordinates are, from Fig. 16·4,

$$m\frac{\partial^2 x}{\partial t^2} = -e\frac{\partial V}{\partial x} - ev_y B \qquad [16\cdot1]$$

$$m\frac{\partial^2 y}{\partial t^2} = -e\frac{\partial V}{\partial y} + ev_x B \qquad [16\cdot2]$$

The left member of equation 16·1 represents a force of inertia directed to the left which is matched by the sum of two forces directed to the right. The negative sign is due to the fact that the electronic charge e is intrinsically negative.

Similarly the left member of equation 16·2 is directed downward and is matched by two upward forces. These equations are readily converted to the form

$$m\frac{\partial^2 x}{\partial t^2} = -\frac{ex}{r}\frac{\partial V}{\partial r} - e\frac{\partial y}{\partial t}B \qquad [16\cdot1a]$$

$$m\frac{\partial^2 y}{\partial t^2} = -\frac{ey}{r}\frac{\partial V}{\partial r} + e\frac{\partial x}{\partial t}B \qquad [16\cdot2a]$$

by use of the relations

$$\frac{\partial y}{\partial t} = v_y, \qquad \frac{\partial x}{\partial t} = v_x \qquad \text{velocities}$$

and

$$\frac{\partial V}{\partial x} = \frac{x}{r}\frac{\partial V}{\partial r}, \qquad \frac{\partial V}{\partial y} = \frac{y}{r}\frac{\partial V}{\partial r} \qquad \text{components of electric intensity.}$$

Multiplying equations 16·1a by y and 16·2a by x, and subtracting the upper from the lower, we have

$$mx\frac{\partial^2 y}{\partial t^2} - my\frac{\partial^2 x}{\partial t^2} = Bex\frac{\partial x}{\partial t} + Bey\frac{\partial y}{\partial t} \qquad [16\cdot3]$$

which upon integration yields the relation

$$mx\frac{\partial y}{\partial t} - my\frac{\partial x}{\partial t} = \frac{Be}{2}y^2 + \frac{Be}{2}x^2 + Cm \qquad [16\cdot4]$$

This integration may be readily verified by differentiating equation 16·4

with respect to t. Dividing both sides of equation 16·4 by m gives

$$x\frac{\partial y}{\partial t} - y\frac{\partial x}{\partial t} = \frac{Be}{2m}(x^2 + y^2) + C \qquad [16\cdot5]$$

where C is a constant of integration yet to be determined. Substituting the following trigonometric identities

$$x = r\cos\theta, \quad y = r\sin\theta, \quad \text{and} \quad \sin^2\theta + \cos^2\theta = 1 \qquad [16\cdot6]$$

into the left side of equation 16·5 we obtain

$$x\frac{\partial y}{\partial t} - y\frac{\partial x}{\partial t} = \left\{\begin{array}{l} r\cos\theta\sin\theta\dfrac{\partial r}{\partial t} + r^2\cos^2\theta\dfrac{\partial\theta}{\partial t} \\ \\ -r\sin\theta\cos\theta\dfrac{\partial r}{\partial t} + r^2\sin^2\theta\dfrac{\partial\theta}{\partial t}\end{array}\right\} = r^2\frac{\partial\theta}{\partial t} \qquad [16\cdot7]$$

Equating this to the right side of equation 16·5, and using the relation $x^2 + y^2 = r^2$, we have

$$r^2\frac{\partial\theta}{\partial t} = \frac{Be}{2m}r^2 + C \qquad [16\cdot8]$$

which is a fundamental differential equation of motion.

The identities of eq. 16·6 define θ as the angle between the X axis and the radius vector r to the point x, y at which the electron is located. The derivatives $\partial r/\partial t$ and $\partial\theta/\partial t$ express the velocity of the electron in polar coördinates. At the surface of the cathode $r = b$ and the electron velocity is zero. We may therefore evaluate the constant C by setting $r = b$ and $\partial\theta/\partial t = 0$ in equation 16·8. Using this value, we have

$$r^2\frac{\partial\theta}{\partial t} - \frac{Ber^2}{2m} + \frac{Beb^2}{2m} = 0 \qquad [16\cdot9]$$

or

$$\frac{\partial\theta}{\partial t} = \frac{Be}{2m}\left(1 - \frac{b^2}{r^2}\right) \qquad [16\cdot10]$$

Conservation of energy permits us to write the relation:

$$\tfrac{1}{2}mv^2 = eV \qquad [16\cdot11]$$

But

$$v^2 = \left(\frac{\partial r}{\partial t}\right)^2 + r^2\left(\frac{\partial\theta}{\partial t}\right)^2 \qquad [16\ 12]$$

which with 16·11 gives

$$\left(\frac{\partial r}{\partial t}\right)^2 + r^2\left(\frac{\partial\theta}{\partial t}\right)^2 = 2\frac{eV}{m} \qquad [16\cdot13]$$

Combining 16·13 and 16·10 so as to eliminate $\partial\theta/\partial t$, we have

$$\left(\frac{\partial r}{\partial t}\right)^2 + r^2\left(\frac{Be}{2m}\right)^2\left(1 - \frac{b^2}{r^2}\right)^2 = 2\frac{eV}{m} \qquad [16\cdot14]$$

Equation 16·14 is a completely general equation describing the motion of an electron under any condition of space charge for a fixed applied voltage. In view of the difficulty of a general solution we shall consider only a special case.

16·4 Cut-off

If $\partial r/\partial t$, the radial velocity, is set equal to zero for $r = a$, or more properly r just less than a, the plate current becomes zero. The motion of each electron becomes tangential to the plate, and no electrons are collected. This condition is referred to as cut-off and is prescribed by the equation

$$a^2\left(\frac{Be}{2m}\right)^2\left(1 - \frac{b^2}{a^2}\right)^2 = 2\frac{eV}{m} \qquad [16\cdot15]$$

or explicitly

$$V = \frac{a^2B^2e}{8m}\left(1 - \frac{b^2}{a^2}\right)^2 \qquad [16\cdot16]$$

defining the cut-off voltage V.

Since $b \ll a$ in most practical cases

$$V = \frac{a^2B^2e}{8m} \qquad [16\cdot17]$$

or

$$B = \frac{2}{a}\sqrt{\frac{2mV}{e}} \qquad [16\cdot18]$$

This expression was first derived by Hull, and his derivation has been checked by other investigators. Unfortunately the experimental verification is not so gratifying. Many painstaking tests by Dr. A. F. Harvey* indicate that a field nearly 10 per cent stronger than that calculated is required for cut-off. Moreover, the cut-off curve is not sharp but rounded, and a relatively large increase in the magnetic field is necessary to reduce the plate current to zero. A typical characteristic is illustrated in Fig. 16·5.

A large number of factors influence the cut-off and might be expected to explain the observed curves. The behavior of several of these factors will be discussed.

* A. F. Harvey, "High Frequency Thermionic Tubes," John Wiley & Sons, 1943.

16·5 Voltage Drop along the Filament

Ordinarily the magnetron employs a filamentary cathode heated by a direct current. There is a potential drop of several volts in the length of the filament and accordingly several volts' difference in the plate voltage effective at the two ends of the tube. By means of a mechanical

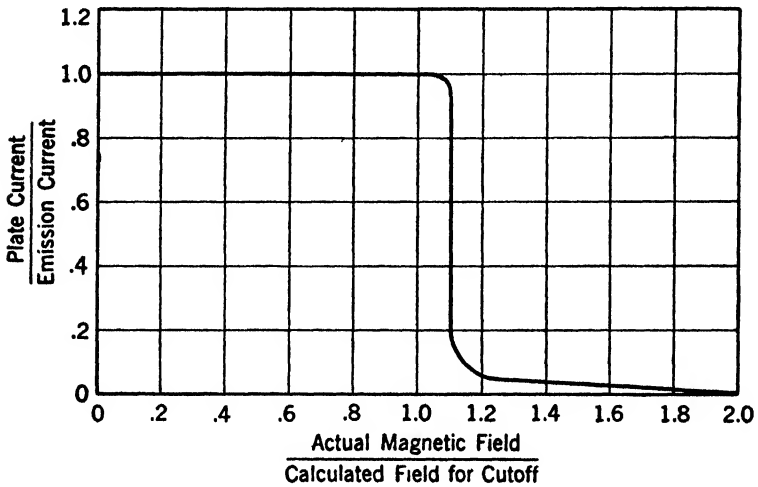


FIG. 16·5 Cut-off characteristic of a magnetron having limited emission.

switching arrangement it is possible to heat the filament by intermittent pulses of direct current and to measure the plate current during the intervals when the filament current is zero. The results of such tests indicate that the rounded corners of Fig. 16·5 do not result from voltage drop along the filament.

16·6 Mechanical Dissymmetry

In a device so small and delicate as a vacuum tube it is difficult, if not impossible, to achieve a mechanical structure that is accurately round, and coaxial. Any deviation from the geometry assumed in the mathematical development should tend to round off the sharp corners of the theoretical curve. Careful tests indicate that the accuracy ordinarily achieved is such that further refinement would not appreciably alter the shape of the cut-off curve.

16·7 Tilt of Magnetic Field

It is impossible to produce a magnetic field that is perfectly uniform or parallel to the axis of the tube elements. However, a long solenoid produces a field near its center which is parallel and uniform to within

0.1 per cent. Experiments involving deliberate introduction of an angle between field and tube axis indicate that a tilt of one or two degrees is not adequate to explain the observed phenomena.

16·8 End Effects

In practical tubes the ratio of axial length to plate diameter is never large. Accordingly, it is proper to consider the effect upon the cut-off produced by divergence or fringing of the field at the ends of the structure. Evidently electrons emitted from the filament outside the ends of the plate will have an axial component of velocity toward the center of symmetry. It can be shown, however, that they should obey the same cut-off relation as electrons emitted near the center of the tube, and experiments with tubes employing guard rings verify the fact that end effects do not appreciably change the cut-off characteristic.

16·9 Space Charge

Since the equation of cut-off was derived in such a way that space charge and potential distribution do not appear, it is clear that they should not affect the observed cut-off. Careful experiments indicate, however, that the space charge does affect the results, at least to some extent.

16·10 Emission Velocity

In ordinary tubes an average emission velocity of about $\frac{1}{2}$ volt is predicted from theoretical considerations and is verified by experiment. That is, the total kinetic energy of a large number of emitted electrons is such that, if it were divided equally, each electron would receive as much as if it had been accelerated through a potential difference of $\frac{1}{2}$ volt. The kinetic theory of gases, as modified to suit the present conditions, indicates that a few electrons of much higher velocity are present as well as some of very low velocity. But the number of electrons having velocities in excess of 5 volts should be negligible.

An interesting experiment to test this point is performed with negative plate potentials. For a particular magnetron, a plate current of 60 microamperes flowed with zero applied potential and was reduced to 1 microampere with -3 volts applied. The magnetic field was zero. In order to reduce the plate current to 1 microampere by means of the magnetic field with zero plate voltage it was necessary to set the magnetic field at a value equivalent to cut-off for an applied potential of 15 volts. That is, the equivalent velocity of many of the electrons is now comparable to 15 volts. Thus experiment indicates that the number of high-velocity electrons is greatly increased by the application of the magnetic field. This effect may be explained upon the basis of

collisions between electrons, an explanation originally advanced by Hull but largely neglected in recent years. We shall use the term collision to designate the close approach of two electrons and shall consider the action by analogy with elastic spheres. Actually, of course, no true collision occurs, but the strong forces of electrostatic repulsion at short distances give much the same effect.

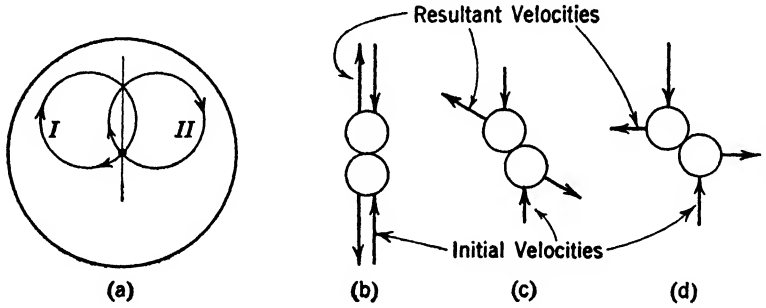


FIG. 16·6 Possible electron paths and collisions in a magnetron.

Let us consider the electron paths indicated in Fig. 16·6 for a cylindrical magnetron with the magnetic field intensity somewhat above the cut-off value. Electrons which are emitted from one side of the cathode follow an approximately circular path and are returned to the opposite side of the cathode. Electrons emitted from a near-by point on the cathode follow a similar path which may cross the path of the others as shown in Fig. 16·6*a*. From equation 16·11 it follows that the speeds of the two electrons are the same at the point where the paths cross. Let us assume that the electrons do collide and that the electron paths cross at right angles to each other.

For convenience the velocities are resolved into X and Y components. Each path crosses the Y axis at an angle of 45°, and accordingly the x and y velocities of each electron are numerically equal. The x velocities of the two electrons in question are therefore equal and in the same direction, whereas the y velocities are equal in magnitude but opposite in direction. The equivalent collision is therefore entirely in terms of y velocity, the x velocity serving merely to translate the motion with respect to a stationary observer.

Figure 16·6*b*, *c*, *d* shows three typical collisions between elastic spheres having equal velocities. It is seen that an exact return along the original direction is possible but that some angular deviation is more probable. Part *d* is particularly interesting. Here the vertical components of velocity are converted entirely into horizontal velocity.

Because the action is entirely elastic there is no change in the magnitude of the velocity. We find by superposition of velocities in the x direction that one electron is brought to rest and that the other electron proceeds in the x direction with a velocity twice its original x velocity ($\sqrt{2}$ times its total original velocity). That is, all the kinetic energy of one electron is transferred to the other.

The electron which was instantaneously brought to rest is again accelerated toward the plate, reaching it easily since the magnetic field has small effect upon its motion owing to the low velocity acquired. The electron with the high angular velocity describes a more or less circular orbit around the cathode until another collision alters its path.

Although the idealized form of collision just described is extremely improbable it is clear that comparable effects result from any similar type of collision. In the example taken it was assumed that space charge exerts only a small effect, and accordingly few collisions are to be expected. In practical tubes the emission of electrons is large and space charge is important. Actually the statement that space charge is significant automatically implies that many collisions or close encounters between electrons occur.

It should be noted that the magnetron is the only familiar form of vacuum tube in which two electrons which are near each other may be going in quite different directions. In other tubes, electrons move in essentially parallel paths and nothing resembling collision occurs. It is therefore reasonable to expect that the behavior of the magnetron is quite different from that of other types of tubes.

From the above considerations we are led to look with suspicion upon any equations which treat the magnetron solely from the standpoint of potential and space charge. Here we are dealing with distances of atomic order and the field equations of Maxwell are known to be inapplicable. Some statistical method such as that used in the study of gases is necessary. We have, however, given a qualitative explanation of the rounded cut-off characteristic. Also we have shown the possibility of electrons moving in circular orbits concentric with the plate.

16·11 Modes of Oscillation

At least three distinct modes of oscillation in the magnetron are well known, and several significant subdivisions of these exist. Oscillations in the audio frequency region may be produced by coupling the winding of the magnet to the plate circuit of the tube. An increasing magnetic field decreases the plate current, much as an increasing negative grid

bias in an ordinary triode decreases its plate current. Oscillations of this type are called feedback oscillations.

Oscillations over a very wide range of frequencies may be produced in a magnetron having the anode split into similar half-cylinders. The term *dynatron* is used to refer to this type of oscillation. Secondary emission plays no part in the performance. Habann* was probably the first to discuss this sort of operation, and this oscillator is sometimes called the Habann oscillator.

Very high-frequency oscillations may be produced in single-anode or split-anode magnetrons by a mechanism similar to that of the positive grid or retarding field oscillator. The operation depends primarily upon electron transit time, and the frequency is, therefore, relatively independent of the external circuit elements.

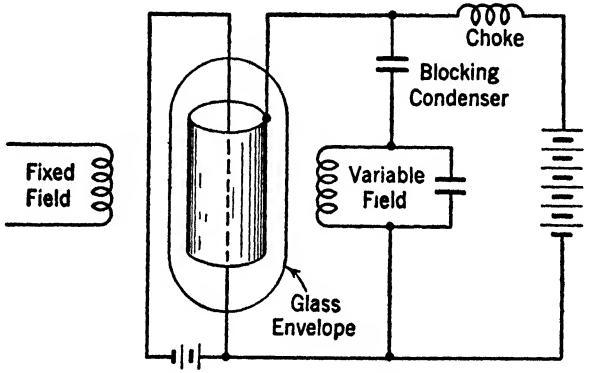


FIG. 16·7 Circuit of feedback magnetron oscillator. Suitable only for low-frequency operation.

16·12 Feedback Oscillations

A feedback circuit for the production of oscillations with a simple magnetron is shown in Fig. 16·7. A relatively small change in the magnetic field produces a large change in plate current. It is necessary only that the connections are such as to deliver a maximum plate current at a minimum plate voltage. This is possible only if the magnetic field is reduced by the flow of plate current. Here, as in triode circuits, best results are achieved if a fixed bias, in this case a magnetic bias, is applied.

With this circuit and a relatively large tube, a power output of 8 kw at an efficiency of 69 per cent was reported by Elder.† The operating

* E. Habann, *Zeitschrift für Hochfrequenztechnik*, 24, 115, 1924.

† F. R. Elder, "The Magnetron Amplifier and Power Oscillator," *Proc. IRE*, 13, 159, 1925.

frequency was 30 kc. With a smaller tube of anode diameter $\frac{3}{4}$ inch and length $1\frac{1}{2}$ inches, he produced a few watts output at 1500 kc.

This circuit finds little application at high frequencies because of the relatively large size of the field coils. The large inductance which results prevents the generation of very high frequencies. Triode oscillators give superior performance with less size, weight, and expense.

16·13 Dynatron Oscillations

The static characteristic of a split-anode magnetron has a negative slope in one region; that is, the device acts as a simple negative resistance. Consequently, oscillations are produced if a suitable tuned circuit is connected to it. Here, as in conventional triode oscillators, the opera-

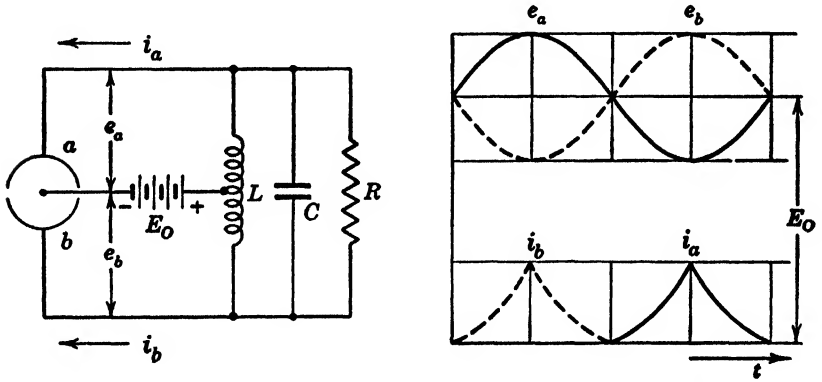


FIG. 16·8 Circuit and waveforms of a split anode magnetron oscillator. (Dynatron mode.)

tion is independent of frequency up to the region in which the electron transit time becomes comparable to the period of a cycle. At somewhat higher frequencies the device continues to operate, but at reduced efficiency. At still higher frequencies satisfactory oscillations are produced only by taking advantage of the electron inertia. The fundamental circuit of this oscillator and certain typical wave forms are shown in Fig. 16·8.

From the static characteristics, Fig. 16·9, it is seen that the anode at the lower potential draws the larger current and that the anode at the higher potential draws no current until Δ becomes approximately $\frac{1}{3}$ as large as E. For $\Delta = E$ the current i_b necessarily falls to zero.

The static characteristic is readily determined, and the oscillator is easy to assemble and to operate. It is not obvious, however, why the static characteristic takes this particular form. Were it not for the

excellent pictures of true electron paths due to Kilgore* there might well remain some uncertainty as to the operation of this device.

Figure 16·10 shows the potential distribution in a split-anode magnetron when the field is somewhat above the cut-off value and the two plates are at equal potential. The emitted electrons regardless of their direction describe essentially circular orbits and return to the cathode. The current to each portion of the anode is zero.

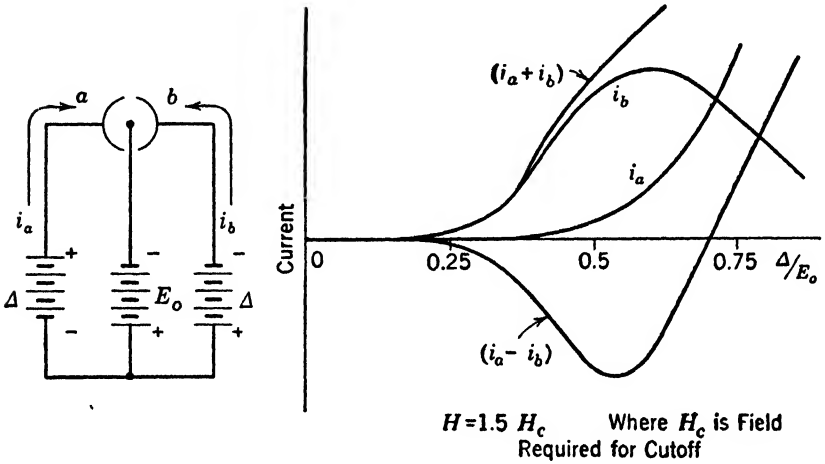


FIG. 16·9 Static characteristics of a split-anode magnetron. (Dynatron mode.)

Figure 16·11 shows the potential distribution when equal and opposite potentials are applied to the two anodes.

Figure 16·12 shows the potential existing in a split-anode magnetron in an operating condition with equal increments of voltage superimposed upon a fixed bias voltage. The potential at any point in this figure is the sum of the potentials corresponding to that point in the two previous figures.

Figures 16·13 and 16·14 show the paths of two electrons emitted in opposite directions under the bias condition of Fig. 16·12. It is seen that both electrons are drawn to the plate having the lower potential.

Figure 16·15, a photograph of an actual magnetron, shows clearly the electron path. A somewhat enlarged model of a normal tube was fitted with a cathode which emitted electrons at only one spot and was capable of being rotated. Argon gas at a pressure of a few microns made the electron trace visible without seriously affecting the operation. The

* G. R. Kilgore, "Magnetron Oscillators for the Generation of Frequencies between 300 and 600 Megacycles," *Proc. IRE*, 24, 1140, 1936.

actual path followed by electrons was then studied for various directions of emission and for a variety of applied potentials. As many as ten loops in the path from cathode to anode were sometimes observed.

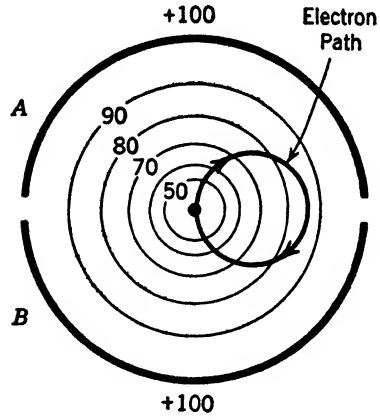


FIG. 16·10 Electron path and electric potential distribution when the two anodes are at equal potential and the magnetic field is somewhat above the cut-off value.

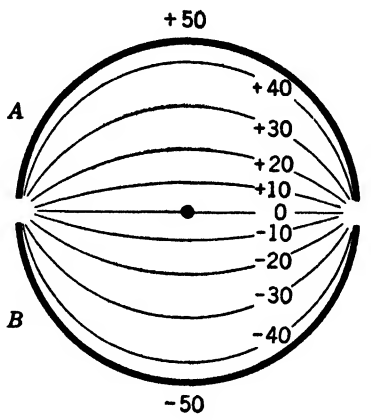


FIG. 16·11 Electric potential distribution when equal and opposite voltages are applied to the two anodes.

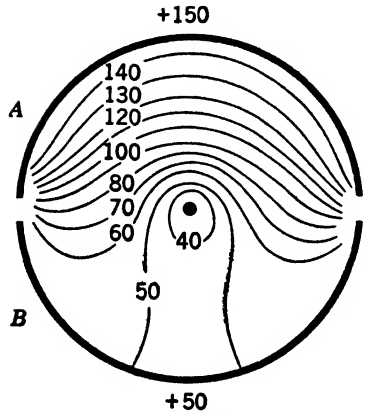


FIG. 16·12 Electric potential distribution when equal increments of voltage are superimposed upon a fixed bias voltage.

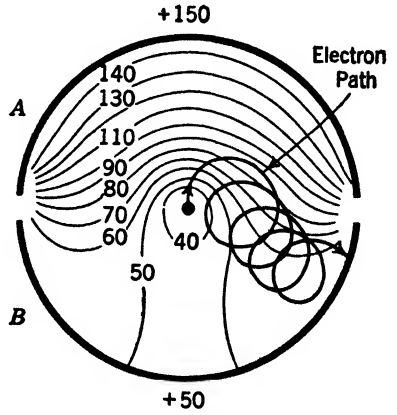


FIG. 16·13 Path taken by an electron when emitted from the filament toward the more positive anode.

The static characteristics of Fig. 16·9 may be explained qualitatively in terms of the electron motion just described. For small values of Δ the electric field distribution within the tube is not greatly affected and

no electrons reach either anode. For $\Delta = E_o/4$ the field distribution is considerably distorted from that of Fig. 16·10, and electrons are able to reach the anode having the lower potential. It is probable that electrons from only certain regions of the cathode reach the plate, all others

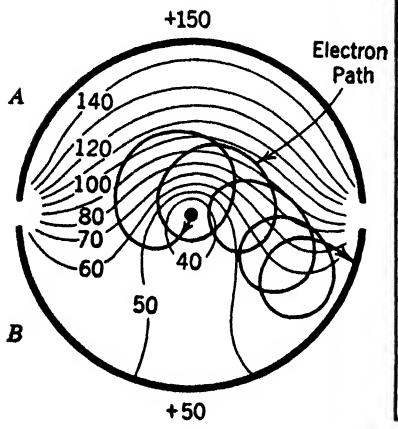
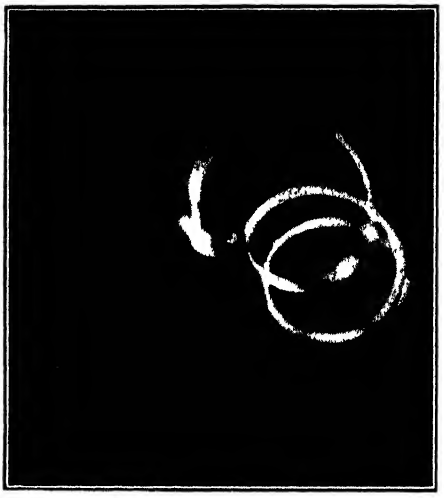


FIG. 16·14 Path taken by an electron when emitted from the filament toward the less positive anode.



(*Kilgore, courtesy of IRE*)

FIG. 16·15 Photograph of an electron beam in an experimental magnetron.

being returned to the cathode. For a rather larger value of Δ most of the electrons from all parts of the cathode reach the anode of lower potential. However, a few electrons reach the other anode. When Δ becomes still larger the two anodes draw comparable currents and the negative resistance property disappears.

The dynatron type of magnetron oscillator is capable of excellent performance for frequencies up to about 600 megacycles. The frequency stability is good because the frequency is controlled almost entirely by the properties of the resonant circuit. At low frequencies the coil and condenser combination of Fig 16·8 is very satisfactory, and the load is readily coupled to the coil by simple mutual induction. At higher frequencies the use of a resonant transmission line for the tank is preferred. The relatively low capacitance present between the plates of the tube guarantees that the resonance of the line is not greatly affected by the presence of the tube.

Generation of harmonics is minimized in the circuit of Fig. 16·8, and frequency stability is improved by decreasing the value of inductance

and increasing the value of capacitance as far as possible. The performance and calculation of this circuit are very similar to those of a push-pull class C triode amplifier.

16·14 Motion of an Electron in Plane-Parallel Magnetron

Consider a pair of parallel conducting plates between which a difference of potential exists. A uniform magnetic field is parallel to the plates and is, therefore, perpendicular to the electric field which results from the assumed potential difference. A single electron moving perpendicular to the magnetic field experiences a force Bev due to the magnetic field. The force is perpendicular to the motion and to the magnetic field. If the electron motion is perpendicular to the electric field, the force eE acting upon the electron due to the electric field is also perpendicular to the motion.

If an electron moves with a suitable velocity and direction between the parallel plates, the two forces just described are equal and opposite.

$$eE = Bev \qquad \text{or} \qquad E = Bv \tag{16·19}$$

An electron whose motion satisfies equation 16·19 moves perpendicularly to both fields and, therefore, parallel to the two bounding planes. Such an electron experiences no net accelerating force and accordingly moves with uniform velocity.

An electron which is freed in a region of zero electric field with some definite velocity v at right angles to a uniform magnetic field describes a circular path. The centrifugal force is exactly matched by the magnetic force according to the equation $Bev = mv^2/r$. Hence the radius of the path is given by

$$r = \frac{mv}{Be} \tag{16·20}$$

and the uniform angular velocity ω_m by

$$\omega_m = \frac{v}{r} = \frac{Be}{m} \tag{16·20a}$$

Thus we see that the radius of the circular path described is directly proportional to the initial velocity, which is maintained unchanged as long as no additional forces come into play. The time required for one revolution is inversely proportional to B and is independent of r or v.

Consider an electron having a velocity component perpendicular to both fields, which satisfies equation 16·19, and an additional small component of velocity parallel to the electric field. The velocity component satisfying equation 16·19 completely annuls the effect of the

electric field. The second component of velocity results in a circular motion superimposed upon the linear one. The path in space which results from such a superposition of velocities is a cycloid.

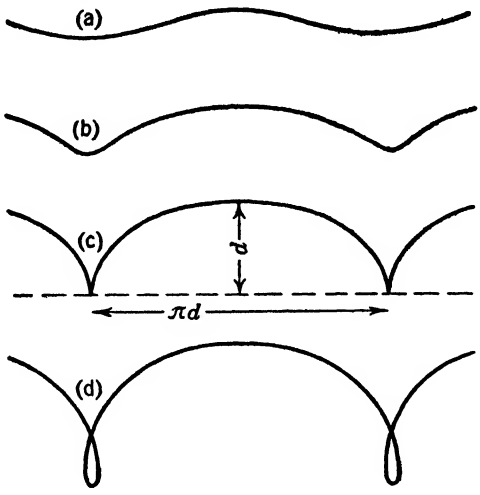


FIG. 16·16 Cycloidal curves.

Several typical cycloidal curves are shown in Fig. 16·16. The path designated *c* is that traced by a point on the periphery of a rolling wheel. It is a particular case which results when the velocity components of equations 16·19 and 16·20 are equal. In this event the electron comes

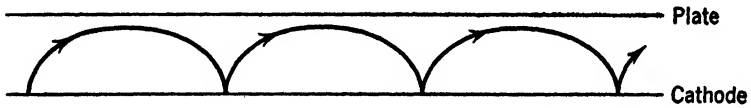


FIG. 16·17 Cycloidal path of an electron in a plane-parallel magnetron. The magnetic field is directed into the page.

to rest at one instant during each revolution of the circular motion. This case is of importance in that electrons emitted from the cathode are at rest, a boundary or initial condition which must be satisfied by any proposed path of motion.* In view of the above we may conclude that electrons in a plane-parallel magnetron describe cycloidal paths as shown in Fig. 16·17. The dimensions of the cycloid depend upon the strength

* This statement is based upon the assumption that electrons are emitted from the cathode with zero velocity and that contact potentials are negligible. These assumptions are justified in ordinary cases, where the applied voltages are large in comparison to the contact potential.

of the electric and magnetic fields, a stronger magnetic field or a weaker electric field resulting in a path of smaller dimensions.

The previous development has assumed that no space-charge effects are present and that no collisions between electrons occur. In an actual tube the emission of electrons takes place over the cathode surface, and accordingly the resulting cycloidal paths cross at all angles. The collisions or near-collisions which necessarily result appreciably affect the performance and allow a considerable number of electrons, which would otherwise return to the filament, to reach the plate.

16·15 Electronic Oscillations in Plane-Parallel Magnetron

The electronic motions just described are very favorable to the production of high-frequency oscillations. An individual electron oscillates back and forth between cathode and plate with a velocity and frequency determined by the dimensions of the structure and the electric and magnetic fields applied. This oscillation is comparable to that which exists in the positive grid oscillator studied in the previous chapter. In the positive grid tube, however, oscillations are produced by the action of electric fields alone, and the grid necessarily collects a large fraction of the emitted electrons. In the magnetron no such grid is required, and relatively high efficiencies are possible.

In the magnetron, as in the positive grid tube, electrons are emitted from the cathode at a uniform time rate. A useful power output may be derived from the tube only if the electronic oscillations are somehow brought into synchronism. This is accomplished in both devices by removing electrons of unfavorable phase from the system after only one oscillation, allowing the electrons of favorable phase to remain for many cycles or oscillations. This amplitude selection is accomplished in the magnetron exactly as it is in the positive grid tube. Refer to Fig. 16·18. A plane-parallel magnetron is subjected to a small alternating voltage in series with the direct plate voltage. The magnetic field is adjusted to a value somewhat above that for cut-off. Let the frequency of the alternating voltage be such that one full cycle is described in the time required for an electron to travel one full loop of the cycloidal curve.

Let us first consider an electron which leaves the cathode at an instant when the alternating voltage is zero and is swinging in such a phase as to aid the direct voltage. During the entire interval that the electron is moving toward the plate the accelerating voltage is above normal and the velocity acquired is above normal. During the following half cycle the electric field is below normal and the large velocity acquired by the electron permits it to strike the cathode with a finite velocity, removing itself from the system. Such an electron is of unfavorable phase

because it moves in the direction of the alternating component of the electric field.

Consider now an electron which leaves the cathode 180° later, that is, when the alternating voltage is zero and is swinging in such a phase as to oppose the direct voltage. During the interval that the electron is moving toward the plate the accelerating field is below normal, and the velocity acquired is less than normal. During the following half cycle

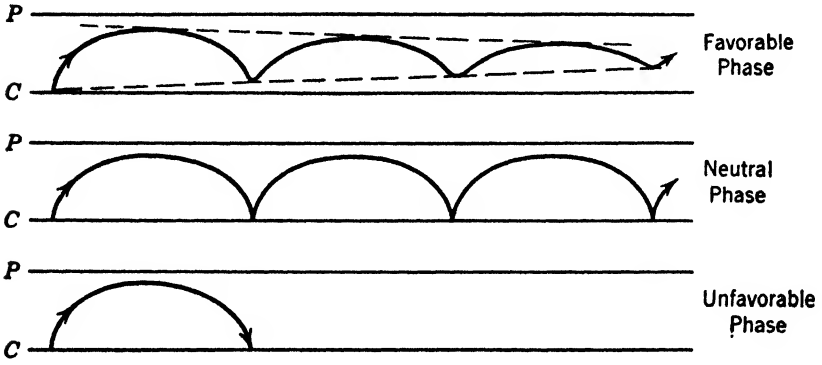


FIG. 16·18 Amplitude selection in a plane-parallel magnetron.

the electric field is above normal and the electron is unable to reach the cathode in its return motion. As a result, it makes several oscillations, the motion being always in the opposite direction to the alternating component of the electric field. Such an electron is of favorable phase.

Here, as in the positive grid tube, we have considered only the most favorable and most unfavorable phase angles. Actually all electrons which are emitted over half the cycle are more or less favorable, whereas those emitted during the other half cycle are more or less unfavorable.

The electronic oscillations in the magnetron differ from those in the positive grid tube in one important respect. In the positive grid tube the time required for the oscillation varies with the amplitude of the motion; the smaller the motion the higher the frequency. In the magnetron the time required for the oscillation is independent of the amplitude of the motion. This fact is verified by application of equation 16·20 to the equality $T = 2\pi r/v$, where T is the period of the wave. Hence

$$T = \frac{2\pi r}{v} = \frac{2\pi m}{Be} \qquad [16\cdot21]$$

Thus it is possible to achieve relatively high efficiency in magnetrons employing electronic oscillations, but not in positive grid oscillators.

The frequency of oscillation is readily calculated from equation 16·21 by means of the equality $f = 1/T$. Thus

$$f = \frac{1}{T} = \frac{Be}{2\pi m} \qquad [16\cdot22]$$

It is seen to depend only upon the magnetic induction and certain basic physical constants. Reduced to a numerical form

$$f = 2.8 \times 10^6 B \qquad [16\cdot23]$$

where B is measured in gauss and f is in cycles per second. Or

$$f = 2.23 \times 10^8 H \qquad [16\cdot24]$$

where H is measured in ampere-turns per meter and f is in cycles per second.

Thus we have shown that the magnetron acts as a two-terminal negative resistance at a frequency depending only upon the strength of the applied magnetic field. It is necessary, then, only to provide a suitable resistive load impedance in order to derive a useful power output. In practice it is universal to employ some form of tuned transmission line as this load. The tuning of the line somewhat affects the frequency as well as the load impedance. By the use of a suitable line the frequency stability of a magnetron may be made somewhat better than the stability of the magnetic field.

16·16 Practical Magnetron Oscillators

The plane-parallel structure is undesirable for various practical reasons. It is universal practice to construct tubes of this sort in a cylindrical form. The operation is not greatly different, although the electron paths are now approximately circular, as shown in Fig. 16·6, rather than cycloidal. Electrons oscillate between cathode and anode at a frequency comparable to that defined by equation 16·24, and the process of amplitude selection operates as it does in the plane-parallel tube. The operation is essentially similar to that just outlined. This cylindrical construction, although perfectly practical, is not ordinarily used. The dissymmetry between filament and plate makes it difficult to achieve the operation desired.

The split-anode construction shown in Fig. 16·19 is used in practically all high-frequency magnetrons, the load being a symmetrical Lecher wire system connected to the two halves of the plate. The apparent short circuit within the tube is actually a quarter-wave resonant line. In the split-anode tube the high-frequency currents are balanced out in the filament circuit and complete symmetry prevails.

The electronic oscillations are essentially the same as before. It remains only to show that the phase relations which are favorable for one half of the tube are also favorable for the other half.

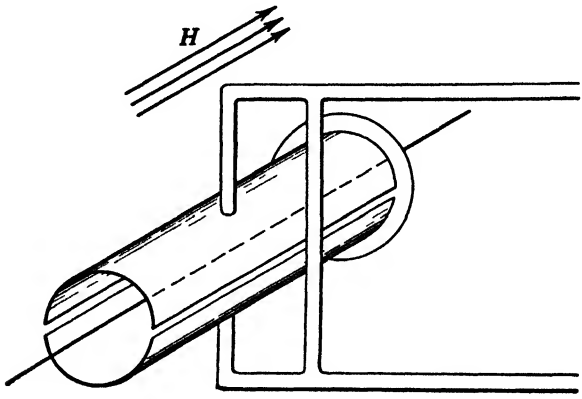


FIG. 16·19 Split-anode structure for a magnetron.

The behavior of the cylindrical split-anode magnetron may be explained directly by reference to Fig. 16·20. Electrons which pass the slit between the anodes when the voltage is a maximum, in such a way as to accelerate their motion, quickly strike the plate and are removed from the system. Electrons which pass the slit when the voltage is in the opposite polarity give up energy to the external circuit and are decelerated. These electrons are of favorable phase and make a number of complete oscillatory motions about the system. On this basis the cathode plays very little part in the operation, the field between the edges of the anode being the primary factor.

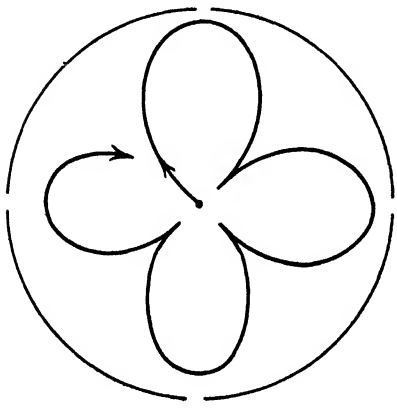


FIG. 16·20 Electron orbit in a four-segment split-anode magnetron. The decreasing vibration of an electron of favorable phase is shown.

It is found experimentally that the operation of such a magnetron is improved by tilting the magnetic field approximately 5° from the axis of the structure. The electric and magnetic fields are no longer perpendicular, and the electrons proceed lengthwise along the axis of the tube in a sort of helical path. Electrons which have given up much

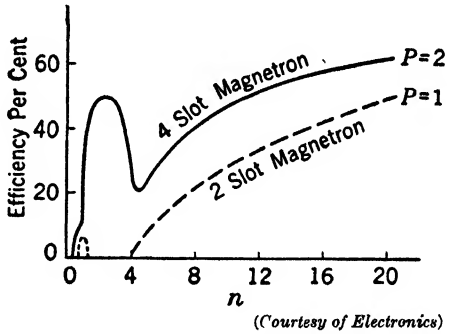
of their initial energy in useful output are thus enabled to reach the plate and reduce the space charge which tends to collect. Somewhat the same effect is achieved by the use of end plates which are at approximately the same potential as the anode. These end plates, one of which is shown in Fig. 16·19, contribute an axial component to the electric field, again causing electrons to drift along the length of the tube and remove themselves from the system.

16·17 Electronic Oscillations of Higher Order

The mode of oscillation just described is usually referred to as that of the first order, where the order of oscillation is defined by the relation

$$n = \frac{\omega_m}{\omega}$$

In this expression n is not necessarily an integer and ω_m is defined by equation 16·20a; $\omega/2\pi$ is the frequency of the oscillations produced. The magnetic field is greater than the cut-off value for the voltage applied, and the frequency is directly proportional to the magnetic field. Other modes, referred to as electronic oscillations of higher order, $n > 1$, may be produced in the same tube structure. These oscillations are of a still higher frequency, which is not simply related to the tube structure or field intensity. In general the efficiency is relatively high, as indicated in Fig. 16·21.



(*Courtesy of Electronics*)

FIG. 16·21 Efficiency of two magnetrons as a function of the order n of oscillation.

The operation may be thought of as analogous to that existing in the class C frequency multiplier. Pulses of current flow when the alternating component of the voltage opposes the flow. The voltage may describe two, three, or even more complete cycles between the pulses of current. It is necessary only that the voltage oppose the passage of each successive pulse of current. In the magnetron, electrons of unfavorable phase are accelerated in their first passage across the slit between the anode segments. They accordingly describe a circle of increased radius and are lost to the anode after one revolution. Electrons of favorable phase are decelerated at each successive slit they pass, the voltage alternating two or more full cycles in the time required for an electron to travel from one slit to the next.

These higher-order modes of oscillation are largely used in the commercial production of hyper frequencies. Frequencies between 1000 and 10,000 megacycles are readily produced in this way, and higher as well as lower frequencies may be generated without great difficulty. The wavelength produced is given approximately by the empirical relation

$$\lambda = \frac{940r^2H}{PV} \qquad [16\ 25]$$

where r is the anode radius in centimeters, H is the field strength in gauss, P is the number of pairs of anode segments, V is the applied potential in volts, and λ is the wavelength in centimeters. For best operation the field must have the value given approximately by the equation

$$H = \frac{nP}{2} \frac{1}{\sqrt{nP - 1}} H_c \qquad [16\cdot26]$$

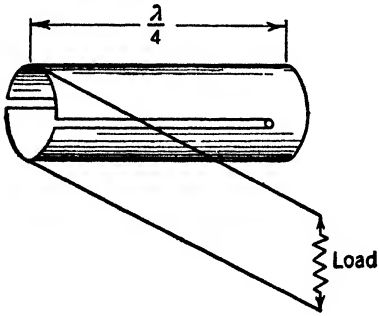


FIG. 16·22 An anode-tank-circuit for magnetron. The split-plate structure, which is closed at one end, forms a parallel line system.

(*Linder, courtesy of IRE*)

FIG. 16·23 An experimental anode-tank-circuit magnetron.

where n is the order of the oscillation, P is the number of pairs of anode segments, and H_c is the magnetic field for cut-off with the applied voltage V. Figure 16·21 shows an experimental curve of efficiency for magnetrons of two- and four-segment construction and for various orders of oscillation. It is seen that good efficiency is to be expected if proper precautions are taken.

In the magnetron as in other high-frequency generators it is often advantageous to construct a resonant line as part of the electrode system. In the magnetron this is very readily done, the split anodes themselves forming the line or tank circuit.* In Fig. 16·22 the tank circuit of such a magnetron is drawn. The anode is split except for a small supporting ring at one end, and the load is connected by a pair of wires attached near the other end. The complete structure is shown in Fig. 16·23.

A particular magnetron of this sort had the following characteristics.

Plate potential	3300 volts
Magnetic field	1500 gauss
Plate current	20 ma
Tilt of magnetic field	0–10 degrees
Wavelength	8–9 cm
Load resistance	80–140 ohms
Output power	13 watts
Efficiency	20 per cent
Anode diameter	0.7 cm
Length of anode slot	2.3 cm
Width of anode slot	0.063 cm

The advantages of joining the tank circuit with the active electrodes is equally pronounced at lower frequencies. Kilgore† describes a

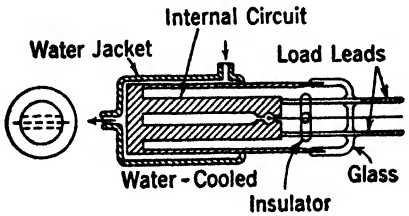
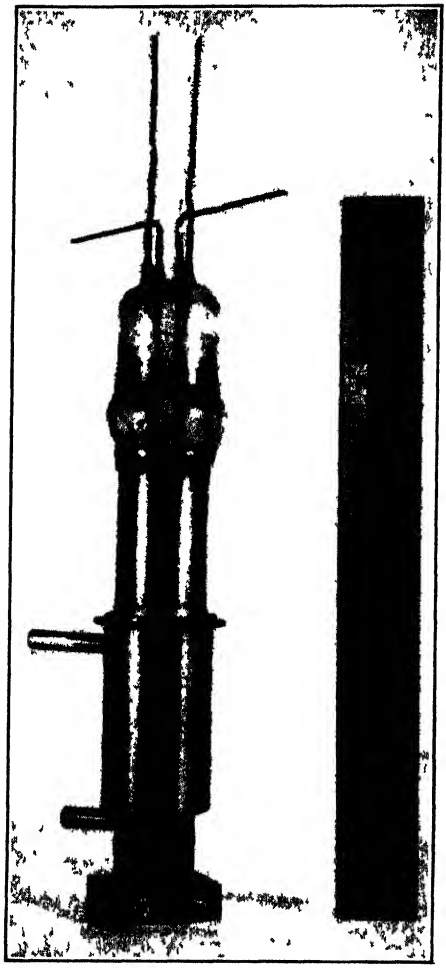


FIG. 16·24 Cross section of a water-cooled magnetron having tank circuit integral with the anodes.

magnetron for operation at 600 megacycles in the dynatron mode which produces an output of 100 watts at an efficiency of about 25 per cent. The large power output is achieved by means of water cooling. A relatively large metal "hairpin" forms the resonant circuit as well as the active electrodes. The general arrangement is shown in Fig. 16·24. The physical appearance of the working tube is shown by Fig. 16·25.

* E. G. Linder, "Anode-Tank-Circuit Magnetron," *Proc. IRE*, 27, 732, 1939.

† See footnote on page 564.



(*Kilgore, courtesy of IRE*)

FIG. 16 25 Experimental water-cooled magnetron for operation in the dynatron mode at 600 megacycles.

PROBLEMS

16·1 A plane-parallel magnetron employs a uniform magnetic field parallel to the plates. A single electron describes a circular orbit between the plates, generating a current in the circuit connected to the plates. (The effect of any fixed electric field is neglected.) If the frequency is 1000 megacycles, how strong is the magnetic field?

16·2 In problem 16·1 is the current wave produced by the electron motion sinusoidal? Why? Assuming the plates are 5 mm apart, how great a current may be produced in this way?

16·3 An electron describes a circular orbit concentric with the axis of an ordinary two-section split-anode magnetron. Sketch the current wave which results in the external circuit.

16·4 Discuss the use of a simple magnetron as a class C power amplifier at relatively low frequencies. What sort of efficiency is to be expected, and what factors are likely to limit the efficiency? What factors tend to limit the frequency range of the device? Consider the magnetic "bias" and the exciting field. Is the adjustment likely to be critical?

16·5 Discuss the split-anode magnetron as a dynatron oscillator in terms of the general theory presented in Chapter 14. What elements contribute the limiter and frequency control functions? Is the amplifier function the same here as it is in other oscillators?

16·6 A split-anode magnetron serves as a transit-time oscillator. Discuss this operation in terms of the general oscillator theory presented in Chapter 14.

16·7 It is sometimes observed in a magnetron operating as a transit-time oscillator that the filament is much hotter than it would be if the tube were not oscillating. Discuss this observation in terms of the collision of electrons.

16·8 Write the differential equations which apply to the motion of an electron in a plane-parallel magnetron. Let the electron motion be always perpendicular to the magnetic field which itself is parallel to the plates.

16·9 Solve the equations of problem 16·8. Look for a parametric equation in which the rectangular coordinates, x, y, z, are expressed in terms of some separate angular variable such as θ.

16·10 A certain split-anode magnetron oscillator is unable to supply sufficient power for the desired application. Will it be possible to utilize two or more identical tubes in parallel for this purpose? What precautions are likely to be necessary? Consider a relatively low-frequency case and one in which the conducting leads are of a length comparable to the wavelength.

16·11 Two four-segment split-anode magnetrons are available. Under what conditions will it be possible to connect both tubes to a common load so as to double the power output produced by a single tube? Consider the case of high-frequency electronic oscillations.

16·12 A magnetron consists of a cylindrical anode concentric with a slender straight wire filament. The magnetic field is above the cut-off value. Show how electrons may reach the anode if the magnetic field is tilted slightly with respect to the axis.

16·13 Assuming negligible space charge and no alternating voltages in the system of problem 16·12, consider the energy of the electron in terms of potential and kinetic energy. With what velocity does the electron strike the plate?

16·14 In the magnetron of problem 16·12 an alternating voltage synchronous with the rotational period of the electron is superimposed upon the direct voltage. With what velocity will electrons now reach the plate? Consider several values of phase angle. Show how the paths of problem 16·12 are modified.

16·15 A four-segment split-anode magnetron is available in which all four segments are brought out to separate terminals. It is desired to produce oscillations of approximately 100-megacycle frequency in the dynatron mode. Show a possible circuit arrangement.

16·16 For a special application it is desired to produce a dynatron oscillation of a two-phase nature. Can the tube of problem 16·15 be used for this purpose? How?

16·17 In a magnetron the useful output power comes from the direct-current supply. Explain in detail how this conversion comes about in terms of individual electrons.

16·18 In terms of the conversion discussed in problem 16·17, what features of magnetron design are necessary to secure high operating efficiency? Suggest a mechanical design which satisfies these requirements.

16·19 It has been suggested that a cylindrical magnetron be modified by the addition of some kind of grid structure. Consider the possibilities of such an arrangement using grids formed of circular wire. The grid may take the form of wires parallel to the filament or the usual helical form.

16·20 Extend the consideration of problem 16·19 to a grid made of thin slats of metal formed so as to correspond to the curved paths of the electrons. Consider such a tube as a low-frequency amplifier employing either electric or magnetic control. What sort of characteristic curves might such a tube have?

CHAPTER 17

TUBES EMPLOYING VELOCITY MODULATION*

17·1 Introduction

It was shown in Chapter 12 how the transit time of the electron in the interelectrode space affects and limits the performance of conventional vacuum-tube structures at high frequencies. In particular we have found that the power required to drive the control grid of ordinary tubes becomes comparable to the output power and that amplification is, therefore, impossible. We shall now describe several devices in which these difficulties are more or less completely overcome.

Although the transconductance of ordinary vacuum tubes becomes complex, and its magnitude decreases at very high frequencies, this effect seldom proves to be the real limit of operation. At considerably lower frequencies the effective conductance of the grid circuit becomes so high that it is impractical to operate the tube. The power supplied at radio frequencies by the grid circuit serves to increase the average velocity of the electrons which reach the plate and so is lost as heat at the plate.

In the ordinary vacuum tubes the grid is successful in controlling the plate current only if a considerable space charge exists in the grid-cathode region. The variation of electronic current must therefore exist in the entire cathode-anode space. In Chapter 12 it was shown that the cathode-grid current must properly be considered to be the sum or difference of a grid-cathode current and a grid-anode current. At low frequencies in negative grid tubes these two currents are identical and the effective grid impedance is infinite. At high frequencies, however, the two currents are out of phase with each other and with the grid voltage. For this reason both a grid conductance and a grid susceptance are developed. The losses that result may be reduced by reduction of the transit time since they vary as the square of this time, but this process has rather definite limitations. A more profitable approach is through utilization of velocity modulation which is a by-product of our efforts to modulate the density or strength of the electron current. It has been found possible to produce satisfactory velocity

* The term " to modulate " as used in communication engineering means *to vary*. Thus velocity modulation means velocity variation.

modulation in practical vacuum tubes, without reducing the effective grid impedance below 50,000 ohms, even at frequencies of 5000 megacycles. Several methods of producing and utilizing this type of modulation will now be described.

17·2 Velocity Modulation

The operation of velocity-modulated tubes is probably no more complicated than that of some forms of low-frequency tubes. However, the principles involved in velocity modulation tubes are relatively unfamiliar to most engineers and we shall consider them in some detail. Just as the operation of triodes is greatly clarified by a separate consideration of the direct and alternating components of voltage, current, etc., so is the operation of the velocity-modulated tubes clarified in the same way. The most characteristic property of velocity-modulated tubes is that the current leaving the cathode is constant. Space charge is therefore unnecessary to the successful operation of such devices, though it usually exists for reasons of cathode design.

Let us consider for a moment the motion of electrons at some small region in a conventional vacuum tube. Provided that all the electrons in this region have the same velocity at a given instant we can resolve the velocity into two components, one constant, the other alternating. The steady velocity component is due to some direct accelerating voltage; the alternating component is due to some alternating or modulating voltage. Such a condition exists in the region of the electrode system which produces the velocity modulation.

At other points in the tube the number of electrons passing in a given unit of time may vary and the electrons may have different velocities. Such a condition exists if a velocity-modulated stream of electrons is allowed to "drift" for an appreciable distance. Here it may be more appropriate to consider only the variation in number of electrons with respect to time, neglecting the effect of velocity.

Modulation of the current density as carried out in ordinary tubes is readily expressed in terms of the total current since it is the same at all points in the system. The degree of velocity modulation, however, is not so simply expressed, for, although it may be expressed in percentage of the steady velocity at a given point, a different result obtains if the observation is made at a point where the steady velocity is different. It is commonly expedient to express the velocity modulation directly in terms of volts, a figure which is constant at all points in the tube.

17·3 Production of Velocity Modulation

Two adjacent grid structures form the simplest device for producing velocity modulation. Such a device in conjunction with other elec-

trodes is shown in Fig. 17·1. Let us assume that the direct accelerating voltage V_1 is large in comparison to the alternating voltage V. Also let the velocity of the electrons reaching the first grid, G_1, be high enough, owing to the action of V_1, so that the interval of time spent in traveling between the grids is short compared to the period of the applied alternating-voltage wave. Under these conditions the electrons which pass through the second grid have a velocity which depends upon the instantaneous sum of V_1 and V. All the electrons which pass through the grid structure continue on to the plate or collector. Since V_2 is lower than V_1 the electrons are decelerated before reaching the plate and dissipate only a relatively small amount of power there.

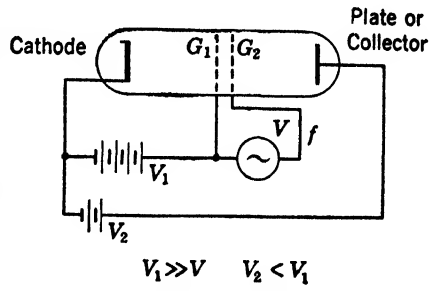


FIG. 17·1 Structure for producing velocity modulation. The maximum value of the alternating voltage of frequency f is represented by V.

The electric field in the space between the cathode and G_1 is fixed because V_1 is constant, and accordingly the velocity and number of electrons which reach G_1 per unit time are fixed.* Since $V_1 \gg V$ and since the transit time between the grids is short it follows that the number emerging from G_2 per second is also virtually constant. This statement is of great importance since it permits us to estimate the power required of the generator V. That is, we may deduce the effective input grid conductance in this way.

It was shown in Chapter 12 that the effective current density due to the motion of the electrons between G_1 and G_2 is equal to ρv, where ρ is the volume density of electronic charge and v is the electron velocity. During the time that V is positive, therefore, the electron velocity is in the same direction as the electric field and energy is drawn from the source V. During the next half cycle the voltage is opposite to the electron velocity and an equal amount of energy is returned to the source. The situation is equivalent to a low-frequency system in which direct current flows through an alternator. The power flow is reversed with each half cycle, and the net power integrated over a cycle is zero.

Actually the input conductance is not identically zero as indicated above. An analysis too lengthy to include here is necessary in order to obtain the exact result. Such an analysis shows that the conductance, although finite, is ordinarily too small to be of practical importance.

* Except for small variations due to the random character of emission.

The resonators or other circuits which supply the excitation V ordinarily have inherent losses large in comparison to this active loss.

Thus it is seen that velocity modulation may be produced by a pair of grids similar to those employed at low frequencies. The success of the grids in producing velocity modulation is independent of space charge near the cathode so that any form of cathode may be used. The power required for the production of velocity modulation is intrinsically low and decreases as the accelerating voltage V_1 is increased and as the spacing between the grids is decreased.

17·4 Velocity Modulation Produced in Two Steps

The structure of Fig. 17·2 may be used for the production of velocity modulation. It is twice as effective as that of Fig. 17·1 in that the same voltage acts twice upon each electron. Let us first consider the action of this device when the alternating voltage is zero. Electrons which leave the cathode are accelerated to a relatively high velocity as they approach G_1. In the space between G_1 and the sleeve electrode they are somewhat decelerated so that they travel the length of the sleeve at the uniform velocity corresponding to V_3. In the region between the sleeve and G_2 they are again accelerated to the velocity corresponding to V_1, and finally they are decelerated toward the plate to a velocity corresponding to V_2.

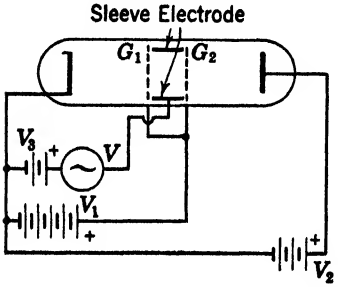


FIG. 17·2 Alternative structure for producing velocity modulation. The maximum value of the alternating voltage of frequency f is represented by V.

By an appropriate choice of V_1 and V_3 and the dimensions, the time required for an electron to travel from G_1 to G_2 can be made equal to that of a half cycle of a given alternating voltage V. If this voltage is applied as is indicated in Fig. 17·2, the same electrons which experience a minimum of deceleration in passing from G_1 to the sleeve experience a maximum of re-acceleration between the sleeve and G_2 and thus leave G_2 with a velocity corresponding to the voltage $V_1 + 2V$. Electrons which come through the system a half cycle later experience a maximum deceleration and a minimum re-acceleration and leave G_2 with a velocity corresponding to $V_1 - 2V$. Electrons arriving at intermediate times experience intermediate values of acceleration or deceleration relative to the velocity corresponding to $V = 0$.

Since $V \ll V_1$ it is still true that the stream of electrons leaving

the second grid G_2 is practically uniform. Actually, of course, the electrons which are accelerated at the first grid-sleeve transit tend to come out of the second grid a little ahead of their normal time. Similarly, electrons which are retarded at the first grid-sleeve transit tend to come out of the second grid somewhat behind normal. So far as these effects are small the power required to produce velocity modulation in this device also is negligible.

17·5 Utilization of Velocity Modulation

The electrode arrangements so far described have been useless in the sense that no output signal is derived from the electron beam even though velocity modulation has been achieved. In order to be of practical use, the velocity-modulated beam must first be converted into an intensity-modulated beam.* There are at least three ways to accomplish this:

(1) by deflection method; (2) by the use of a retarding field; and (3) by the drift tube.

Conversion by deflection

If a beam of electrons is deflected by means of a transverse field, either electric or magnetic, the path described depends upon the velocity of the electrons. Accordingly, by the use of an appropriate field it is possible to separate a velocity-modulated beam into two beams which are modulated in intensity. The two beam currents are necessarily out of time phase by 180° so that a natural push-pull system results. One possible arrangement for utilizing this effect is shown in Fig. 17·3. Electrons which leave the velocity-modulating grids with maximum velocity are least deflected and reach the further anode P_1. Electrons with minimum velocity are most deflected and strike P_2.

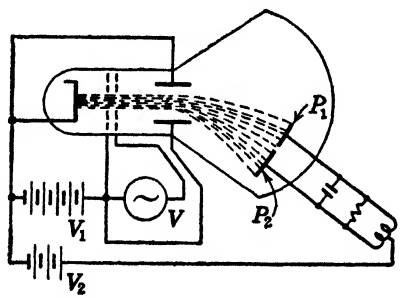


FIG. 17·3 Tube employing fixed electric field to separate velocity modulated beam into two intensity-modulated beams.

Actually the arrangement shown in Fig. 17·3 is not particularly practical. The voltages must be controlled rather accurately, and the operation is quite sensitive to stray electric or magnetic fields. Moreover, it is fundamentally nonlinear, since a finite division between the

* Intensity-modulated beams are those in which the convection current ρv at a given point is a function of time.

plates P_1 and P_2 must exist. Accordingly, very small alternating voltages at V may be unable to excite any finite output in the tank circuit. At high levels of input the beam is completely separated and no further increase of output is possible.

Conversion by retarding field

If in Fig. 17·1 the auxiliary voltage V_2 is made equal to zero we find that electrons which leave the grids with velocity above normal reach the collector or plate but that electrons which leave the grid with lower velocity are turned back and ultimately return toward the grid or other positive structure. Thus a true conduction current exists in the plate lead of the tube. Fairly efficient operation is possible with such a tube provided that suitable steps are taken to prevent the return of electrons to the exciting grids. This precaution is necessary because the distance traveled to and from the plate permits the faster electrons to overtake the slower ones, producing an intensity-modulated beam. Such a beam causes the input impedance to have a large resistive component which may be either positive or negative. If the resistance is positive a large driving power may be required. If it is negative an undesired mode of oscillation may occur.

A rather different process takes place if the voltage of V_2 in Fig. 17·1 is made sufficiently negative so that no electrons reach the plate. Now we find that the faster electrons approach the plate very closely while the slower ones are turned back somewhat farther away. The charge induced in the plate circuit by these moving electrons is thus controlled by the potential of the velocity-modulating electrodes. Again, suitable precautions must be taken if the electrons are not to return to the grid with undesirable consequences.

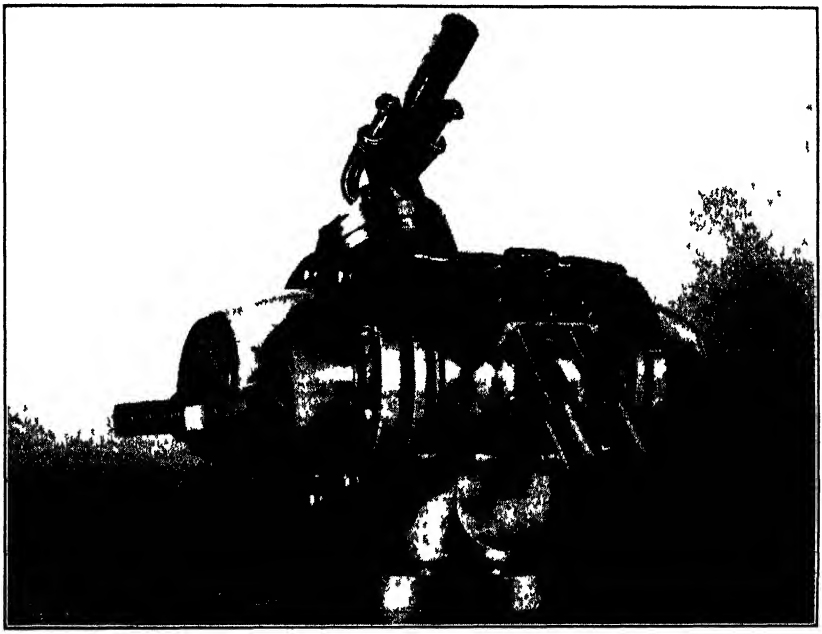
Successful tubes have been constructed upon this principle, and reasonably satisfactory results are obtained. It appears, however, that certain other designs are fundamentally better, and we shall therefore give this one only brief consideration.

Conversion by drift

An electron beam that is velocity modulated will convert itself into one that is intensity modulated if given sufficient time. That is, a uniform velocity-modulated beam will automatically gather itself into clumps if allowed to drift freely in an equipotential region. Such a region is known as a drift space, and the resulting clumped or bunched beams may be partly or completely intensity modulated.

The most familiar device for utilization of this principle is the

Klystron.* Although a number of other workers, notably Hahn and Metcalf of the General Electric Company, produced similar devices at about the same time, the brothers S. F. and R. H. Varian are generally credited with its development.† Figure 17·4 is a photograph of an early experimental model (1939). Figure 17·5 shows the internal arrangement of two typical Klystrons suitable for amplifica-



(R. H. and S. F. Varian, courtesy of J. Applied Physics)

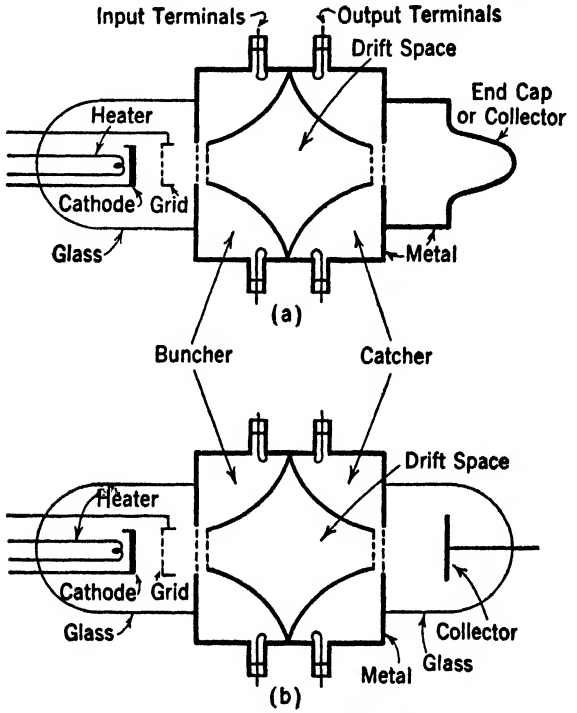
FIG. 17·4 Experimental model of the Klystron.

tion, detection, or oscillation. The cathode is plane and relatively large. The grid indicated is actually a beam-forming electrode which operates at a small positive bias and serves to control the total cathode current. The principal structure is of copper and comprises two cavity resonators coupled mechanically and at the same direct potential. Each cavity resonator includes two gridlike structures close together and operating in much the same way as the double grid of Fig. 17·1.

* The word Klystron is a trademark owned by the Sperry Gyroscope Company, Inc. The term is derived from the Greek and means "waves breaking on a beach."

† R. H. and S. F. Varian, "A High-Frequency Oscillator and Amplifier," *J. Applied Physics*, **10**, 321, May, 1939.

In operation the electrons emitted by the cathode are formed into a relatively narrow circular beam by the grid and are accelerated toward the metallic structure. Some of them strike the grid mesh, but a majority pass through the first cavity resonator into the equipotential drift space. They proceed through this space with essen-



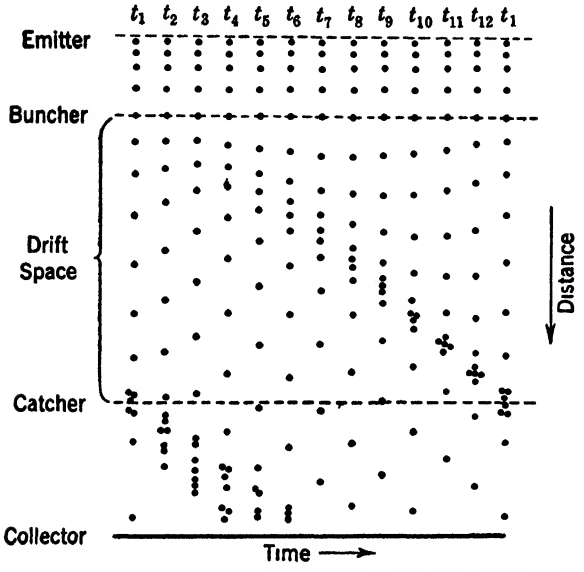
(R. H. and S. F. Varian, courtesy of J. Applied Physics)

FIG. 17·5 Schematic representations of two modern Klystrons.

tially uniform velocity to the grids of the second resonator; again a portion of the electrons strike the mesh but a majority pass through and are decelerated, finally being captured by the plate or collector. By applying a moderate potential to the plate of Fig. 17·5*b*, it is possible to capture most of the electrons without the dissipation of a large amount of power as heat.

The situation is appreciably altered if some alternating voltage exists across the first two grids (the buncher) due to oscillations fed into the resonator. Now the electrons which pass into the equipotential or drift space are velocity modulated. Electrons of higher velocity may overtake slower electrons which preceded them in time. That is, the electrons which arrive at the second pair of grids (the

catcher) are in clumps or bunches. The development of this condition is shown in Fig. 17·6. To simplify the picture it is assumed that the space between the two grids of the buncher is negligible.



(R. H. and S. F. Varian, courtesy of J. Applied Physics)

FIG. 17·6 Conversion of an electron beam from velocity-modulated to intensity-modulated by drift. Sequence of diagrams showing, for 12 times equally spaced throughout one cycle (30° intervals), the positions of chosen typical electrons (dots) in the beam. As to phase, the first vertical row of dots shows the positions of the electrons at a time when the field in the buncher is zero and is coming to be in the direction to increase the speeds of the electrons. Horizontal lines show the positions of the cathod and the buncher Rhumbatrons. A suitable place for the catcher is designated in the figure.

Also, it is assumed that the voltage fed to the buncher is of such a value as to produce maximum intensity modulation of the beam at the catcher, an optimum condition.

17·6 The Applegate Diagram

Further clarification of the action of the buncher may be obtained from a study of the Applegate diagram, Fig. 17·7. The straight lines in this diagram represent the positions of individual electrons after they leave the buncher, as a function of time. The slope of these lines is made proportional to the electron velocities, i.e., proportional to the a-c and d-c components of the buncher voltages. In drawing this

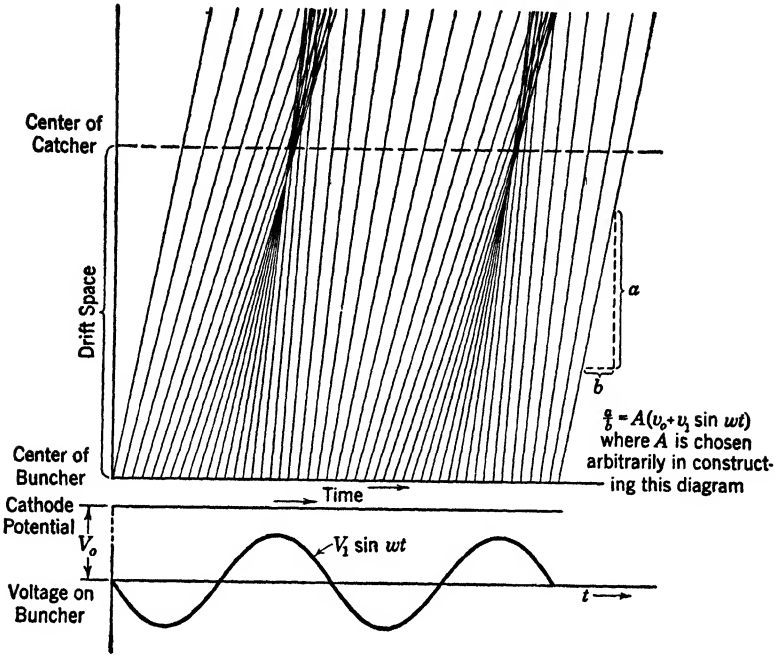


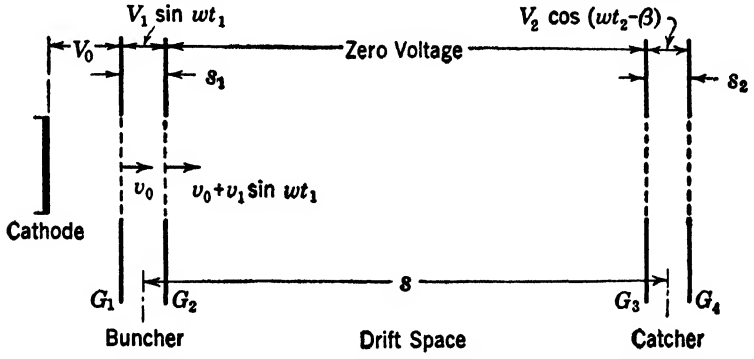
FIG. 17·7 The Applegate diagram of the Klystron.

diagram reference is made to Fig. 17·8 and the following assumptions are made:

1. All electrons have the same velocity before passing through the buncher.
2. Electrons pass through buncher at equal time intervals.
3. The velocity $v_1 \sin \omega t_1$ due to the a-c voltage component $V_1 \sin \omega t_1$ is small compared to the velocity v_0 due to the voltage V_0.
4. The change in velocity of the electron in the buncher is $v_1 \sin \omega t_1$.

All these conditions are met to a fair degree in the practical Klystron. The d-c component of the voltage V_0 is always large in comparison to V_1 so that v_0 is always much greater than v_1. If only a small number of electrons strike the grid mesh of the buncher, it is evident that practically as many electrons leave the buncher as come to it. ' Further, the currents induced in the buncher by the electrons approaching and leaving it are of opposite sign. Hence the current in the external circuit of the buncher is quite small, and power losses in the buncher are likewise small.

The variation of current with respect to time at any particular distance from the buncher is represented in Fig. 17·7 by the intersections of the sloping lines with a horizontal line at that location. At



(Webster, courtesy of J. Applied Physics)

FIG. 17·8 Calculation of electron bunching in a Klystron.

the buncher the intersections are uniform, indicating a uniform or constant current. At the catcher a large number of intersections are grouped together, indicating a pulse of current at one part of the cycle.

17·7 Kinematic Bunching*

Let us now consider in some detail the qualitative relations which apply in the bunching of electrons by the first resonator of a Klystron. This treatment is approximate in that the debunching caused by space charge is neglected. It is also assumed that all electrons leave the cathode with zero velocity and proceed parallel to the axis of the structure. The grids of the buncher are taken to define true equipotential planes perpendicular to the axis of the tube, and the loss of electrons by collision with the metal of the grid structure is ignored.

The result of such an analysis may be affected by the focusing action of actual grid structures, by loss of electrons at the grids, and by electrons which stray from the beam. However, it is found that the results of this analysis describe the observed behavior to a good approximation.

The diagram of Fig. 17·8 is convenient for the calculation of kinematic bunching. Velocities are designated by the small v with suitable sub-

* This treatment follows closely the work of D. L. Webster, " Cathode Ray Bunching," *J. Applied Physics*, **10**, 501, 864, July, 1939. The authors are indebted to Dr. Webster for a very gracious personal letter in which he clarified certain details of the original paper.

The word kinematic is used to signify the fact that the calculation treats the electrons as small solid bodies, projected into the drift space with various velocities. Kinematics is the study of the motions of bodies.

scripts. Applied voltages are designated by the capital V, again with appropriate subscripts. To be explicit we shall assume that no alternating potential difference exists between the first grid, G_1, and the cathode. Further, no potential difference of any kind is assumed in the drift space between G_2 and G_3. The alternating voltage $V_1 \sin \omega t_1$ is supplied from some external source to excite the buncher. The useful output is derived from the voltage $V_2 \cos(\omega t_2 - \beta)$. The constant β is required to account for the time which electrons spend in traversing the drift space. The separate subscripts on the time variable help to simplify the mathematical procedure.

The velocity v_0 with which electrons leave the first grid is defined by the familiar equation

$$\tfrac{1}{2}mv_0^2 = eV_0 \qquad [17\cdot1]$$

where m is the mass and e the charge of the electron. Under our assumptions that s_1 is small compared to the wavelength of the voltages in question and that the electron velocity is small compared to that of light we may state that the number of electrons which leave the second grid is uniform with respect to time and that their velocity is given by the relation

$$v = v_0 + v_1 \sin \omega t_1 \qquad [17\cdot2]$$

where the magnitude of v_1 is determined by a process analogous to that used in equation 17·1. We have:

$$\tfrac{1}{2}m(v_0 + v_1)^2 = e(V_0 + V_1) \qquad [17\cdot3]$$

Subtracting equation 17·1 from 17·3 and neglecting the term in v_1^2,

$$mv_0v_1 = eV_1 \qquad [17\cdot4]$$

This equation is taken to define the velocity v_1 which is the maximum value of the velocity contributed to the electrons by the buncher.

Because the acceleration in the buncher and the deceleration in the catcher are essentially linearly distributed it is appropriate to compute the transit time as if the velocity changes were discontinuous at the centers of these spaces. This approximation is consistent with Fig. 17·8, where s is drawn from the center of the buncher to the center of the catcher. Accordingly we modify our previous statements slightly so that the velocity at the middle of the buncher space s_1 is given by

$$v = v_0 + v_1 \sin \omega t \qquad [17\cdot5]$$

where v_1 and v_0 are determined by equations 17·4 and 17·1, respectively.

17·8 Current Relations in the Klystron

Let us consider an electron which passes the center of the buncher at a time t_1 and arrives at the center of the catcher at some later time t_2. This time is expressed by the relation

$$t_2 = t_1 + \frac{s}{v_0 + v_1 \sin \omega t_1} \qquad [17\cdot6]$$

where $s/(v_0 + v_1 \sin \omega t_1)$ is the time required for the electron to travel the distance s when its velocity is $v_0 + v_1 \sin \omega t_1$. If we introduce the approximation

$$\frac{1}{1+a} \simeq 1 - a \qquad \text{if} \qquad a \ll 1 \qquad [17\cdot7]$$

we have

$$t_2 = t_1 + \frac{s}{v_0}\left\{\frac{1}{1 + \frac{v_1}{v_0}\sin \omega t_1}\right\} = t_1 + \frac{s}{v_0} - \frac{sv_1}{v_0^2}\sin \omega t_1 \qquad [17\cdot8]$$

Differentiating with respect to t_1 we have

$$\frac{dt_2}{dt_1} = 1 - \frac{sv_1}{v_0^2}\omega \cos \omega t_1 \qquad [17\cdot9]$$

The current represented by the modulated beam as it reaches the catcher may be deduced from the continuity of the electron stream. If we designate i_0, i_1, and i_2 as currents at the input of the buncher, middle of buncher, and middle of catcher, respectively, we have $i_0 \simeq i_1$ at all times. The continuity relation thus becomes

$$i_1(t_1)\, dt_1 = i_2(t_2)\, dt_2 \qquad [17\cdot10]$$

where the first term represents i_1 evaluated at t_1 multiplied by the incremental time dt_1. The second term represents i_2 evaluated at t_2 multiplied by the incremental time dt_2. Replacing i_1 by i_0, which is independent of time, we have, upon substitution of equation 17·9

$$i_0 = i_2(t_2)\frac{dt_2}{dt_1} = i_2(t_2)\left\{1 - \frac{sv_1}{v_0^2}\omega \cos \omega t_1\right\} \qquad [17\cdot11]$$

or

$$i_2(t_2) = \frac{i_0}{1 - \frac{sv_1\omega}{v_0^2}\cos \omega t_1} \qquad [17\cdot12]$$

This equation 17·11 defines the current at the catcher, i_2, as a function of the time electrons pass through the buncher. If $sv_1\omega/v_0^2$ is small in

comparison to 1 we may reintroduce the approximation of 17·7 to write

$$i_2(t_2) = i_0\left(1 + \frac{sv_1\omega}{v_0^2}\cos\omega t_1\right) \quad \text{if} \quad \frac{sv_1\omega}{v_0^2} \ll 1 \qquad [17\cdot13]$$

which is a sinusoidal function.

Equation 17·12 requires careful consideration. At first glance it appears that the current is negative over part of each cycle for $sv_1\omega/v_0^2 > 1$. Such a negative current is contrary to fact, because the direction of the motion of electrons through the catcher never reverses. The apparent discrepancy is clarified by reference to Fig. 17·7 or 17·6. It is seen that fast electrons are able to overtake slower ones which preceded them before reaching the catcher. The wave of current which results when i_2 is plotted with respect to time is thus doubly peaked in each cycle.

The situation is further clarified by reference to equation 17·9 which shows that dt_2/dt_1 is also negative over part of each cycle for the conditions in question. A sort of folding back or overlapping process takes place so that electrons which left the buncher over three separate intervals of dt_1 pass through the catcher in the same time interval.

We are therefore correct in using the magnitude of current represented by equation 17·12 but in taking it as always positive. If $sv_1\omega/v_0^2 = 1$, the current value indicated is infinite once per cycle. For $sv_1\omega/v_0^2 > 1$, the current value indicated is infinite twice during each cycle. In the physical tube, of course, these peaks of current are lowered and spread out in time by the action of space charge. The peaks are relatively sharp and the presence of the double peak is particularly interesting in that it suggests efficient operation as a high order frequency multiplier.

17·9 Phase Shift in the Klystron

The Klystron differs from all conventional types of tubes in that it operates with an intrinsically large amount of phase shift. This phase shift is fundamental to the operation of the tube itself and, therefore, may not be removed by any ordinary correction method. For most purposes it is not objectionable, but a number of unusual phenomena result from its presence.

The phase shift with which we are dealing is caused by a time delay and is most readily calculated on that basis. Electrons require an appreciable time to travel from the buncher to the catcher. A signal which is suddenly applied to the buncher is thus able to affect the catcher only when the electrons controlled by the signal arrive at the catcher.

The time required for electrons to travel from buncher to catcher is

readily calculated because the voltage applied to the buncher grid is usually small compared to the applied direct potential. Accordingly the velocity with which electrons travel the drift space may be taken as

$$v_0 = \sqrt{\frac{2eV_0}{m}} \qquad [17\cdot14]$$

from equation 17·1. In practical units (equation 15·4),

$$v_0 = 6.0 \times 10^5 \sqrt{V_0} \text{ meters/sec} \qquad [17\cdot15]$$

The time required for electrons to cross the drift space is, in terms of Fig. 17·12,

$$t = \frac{s}{v_0} = \frac{s}{\sqrt{V_0}} \cdot \frac{1}{6} \times 10^{-5} \text{ second} \qquad [17\cdot16]$$

where t is the time in seconds, s is the distance in meters, and V_0 is the applied potential in volts.

For a typical commercial Klystron, $s = 0.03$ m. Accordingly

$$t = \frac{1}{\sqrt{V_0}} \cdot \frac{3}{6} \times 10^{-7} = \frac{5}{\sqrt{V_0}} 10^{-8} \text{ second} \qquad [17\cdot17]$$

The time delay just calculated is readily converted to a phase shift, θ, by means of the formula

$$\theta = \omega t \qquad [17\cdot18]$$

For a typical Klystron such as we are considering the operating frequency may be taken as $f = 3 \times 10^9$ and $\omega = 6\pi \times 10^9$. Using these relations we write

$$\theta = \frac{300\pi}{\sqrt{V_0}} \text{ radians} \qquad [17\cdot19]$$

For an operating voltage of 900 volts, a reasonable value, $\sqrt{V_0} = 30$ and

$$\theta = 10\pi \text{ radians} \qquad [17\cdot20]$$

This is recognized as 5 complete cycles or 1800°.

An additional phase shift of 90° may be explained in terms of the mechanism of bunching. Let us consider three electrons that pass through the buncher at three successive instants of time corresponding to minimum, zero, and maximum acceleration respectively. The electron which passed through the buncher at zero alternating voltage will arrive at the catcher at the same instant as the decelerated electron which preceded it and the accelerated electron which followed it. That

is, the current peak at the catcher is associated with zero field in the buncher. Since the voltage maximum of the catcher must coincide with the current peak there for best operation, we conclude that a 90° phase shift between the two resonators must exist in addition to the phase shift of delay.

When the Klystron serves as an amplifier the phase shift is relatively unimportant. The output signal is delayed a small fraction of a microsecond with respect to the input, but no other effect is observed. Even if negative feedback is to be applied the situation is not serious because the phase shift is largely a function of the applied direct voltage rather than of frequency.

17·10 The Klystron as an Oscillator

The Klystron oscillator is in some ways considerably more complex than any oscillator so far discussed. A typical circuit is shown in Fig. 17·9. Its operation is best analyzed in terms of the functional

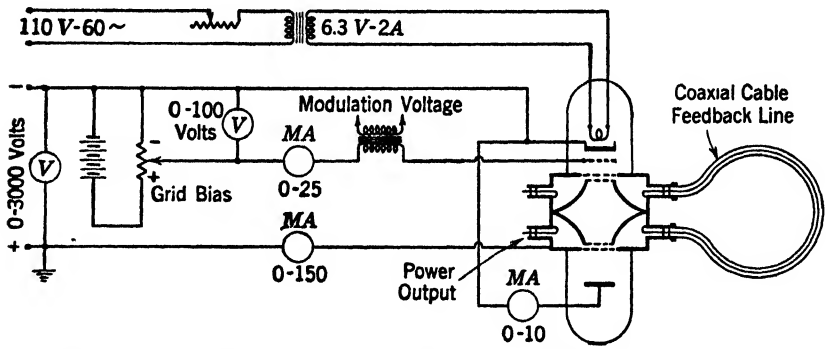


FIG. 17·9 Circuit diagram for Klystron oscillator with plate collector terminal brought out.

block diagram of Fig. 17·10, originally presented in Chapter 14. The amplification is provided by the electron beam as previously discussed. The limiter action takes place as a result of the mechanism of bunching. When the voltage at the buncher exceeds a certain value the current at the catcher forms a double rather than single peak, and the effective transconductance is decreased.*

The frequency-control mechanism is complicated by the fact that there are two separate resonators and that they are coupled relatively tightly together. The behavior is best explained in terms of low-

* Strictly speaking, the best operation occurs when the catcher current just begins to show a double peak. The above statement refers to an excessive broadening of the peaks.

frequency coupled-circuit theory. It will be recalled that the transmission characteristic of two identical tuned circuits shows a double

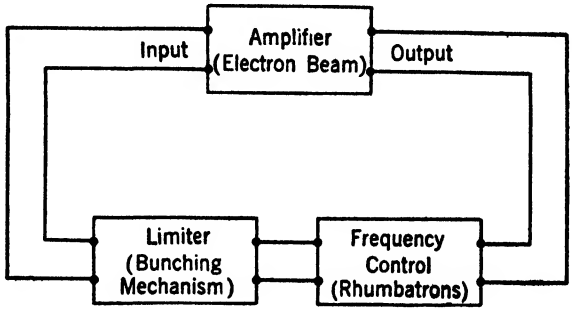


FIG. 17·10 Functional block diagram of a Klystron oscillator.

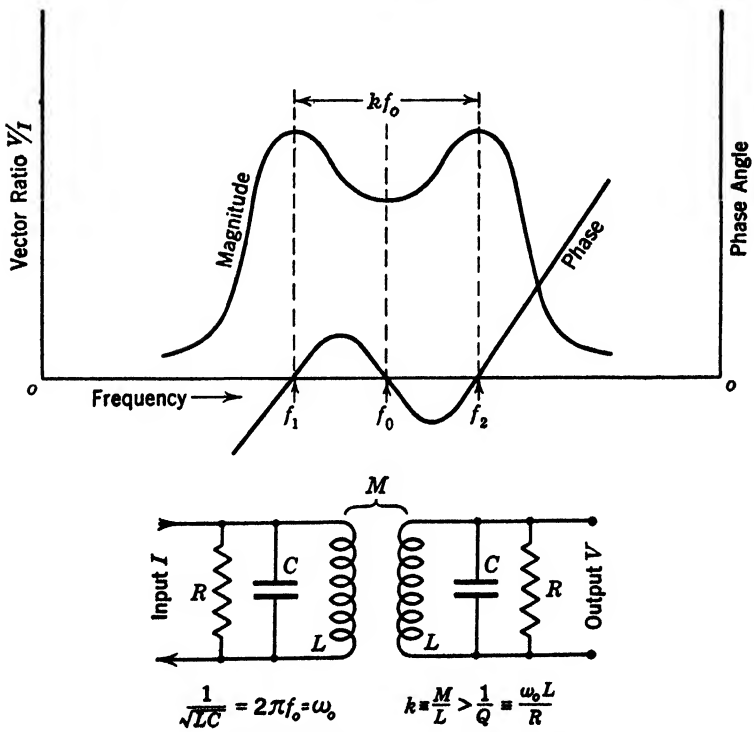


FIG. 17·11 Properties of identical coupled tuned circuits.

hump provided that the coupling exceeds a certain value known as the critical value. See Fig. 17·11. Associated with this transmission characteristic is a curve of phase shift which is zero at three different

frequencies. Oscillations are readily produced at the two frequencies of peak transmission. The reduced transmission associated with the midpoint is relatively unfavorable to oscillation.

In any oscillator it is necessary that the total phase shift around the loop be zero or an integral multiple of 2π at the operating frequency and that the voltage amplification of the entire system be unity. In the ordinary oscillator the phase shift of the system is relatively independent of applied voltages. Accordingly the conditions for oscillation are relatively independent of the applied voltage. The frequency of such an oscillator then adjusts itself until the total loop phase shift is zero.

In the Klystron oscillator the total phase shift is the sum of three components. These are

$$\left.\begin{aligned} \theta_1 &= \frac{\pi}{2} \text{ due to inherent buncher-catcher relationship} \\ \theta_2 &= \frac{\omega s}{6\sqrt{V_0}} \times 10^{-5} \text{ due to transit time in drift tube} \\ \theta_3 &= \frac{l\omega}{v_c} \text{ due to feed back cable} \end{aligned}\right\} \quad [17\cdot21]$$

where ω is the operating angular velocity, s is the drift distance, l is the effective length of the feedback cable, V_0 is the applied direct voltage, and v_c is the velocity of transmission of the feedback cable. Because only one of these terms involves the applied voltage, it is to be expected

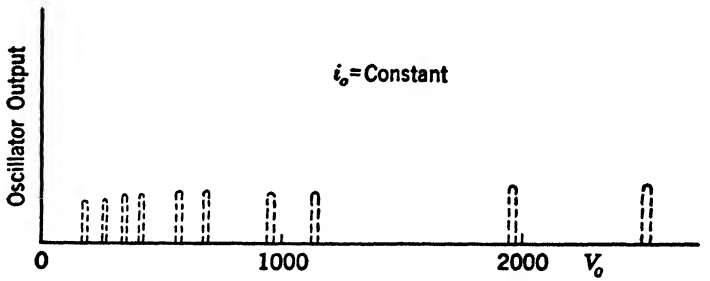


FIG. 17·12 Oscillation characteristic of a Klystron showing critical operating potentials.

that only certain values of this voltage will produce oscillation. Such is actually the case, and it is usually possible to identify two distinct sets of these values. One set of voltages reduces the total net phase shift of the system to a multiple of 2π for the frequency f_1 of Fig. 17·11; the second set reduces the phase shift to a multiple of 2π for the frequency f_2.

A typical plot of oscillation output *vs.* applied voltage is shown in Fig. 17·12. The oscillations are in general somewhat more powerful

as the applied voltage is raised, and the voltage increments between successive points of oscillation are increased.

The above relationships may be put into numerical form. The total phase shift must equal $2n\pi$ where n is an integer.* Accordingly we write

$$2n\pi = \frac{\pi}{2} + \frac{\omega s}{6\sqrt{V_0}} \times 10^{-5} + \frac{l\omega}{v_c} \qquad [17\cdot22]$$

or

$$\frac{\omega s}{6\sqrt{V_0}} \times 10^{-5} = 2n\pi - \frac{\pi}{2} - \frac{l\omega}{v_c} \qquad [17\cdot23]$$

Dividing by 2π we have

$$\frac{fs}{6\sqrt{V_0}} \times 10^{-5} = \left(n - \frac{1}{4} - \frac{lf}{v_c}\right) \qquad [17\cdot24]$$

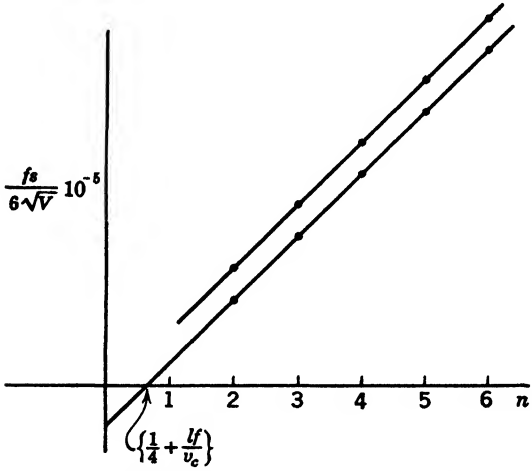


FIG. 17·13 Critical voltages for oscillation of a Klystron.

If the left member of the equation is evaluated for the various voltages at which oscillation occurs and if these values are plotted against a suitable series of integers, a straight line results. In this plot the larger values of voltage are associated with the smaller integers. That two such lines are often observed is evidence that oscillation occurs at two separate frequencies. Such a plot is shown in Fig. 17·13. It is not

* An ambiguity of π radians may easily arise in this work. In order to avoid such an error it is necessary to know the orientation of the coupling loops within the resonators so as to relate the field across buncher and catcher grids to the current flow in the feedback cable. Because the effective length of the feedback cable is seldom known exactly, the error is not of great importance.

always possible to make V_0 large enough to obtain the point $n = 1$. The intercept on the horizontal axis is the value $\left(\frac{1}{4} + \frac{lf}{v_c}\right)$. Accordingly a means is at hand for evaluating the total phase shift of the circuit which couples the two resonators.

17·11 The Cavity Resonator

We have now shown how the process of velocity modulation takes place and how the electrons gather themselves into an intensity-modulated wave of bunches. It remains to show how the output resonator or catcher derives power from the rhythmic passage of these bunches of electrons. From our previous work it is clear that the passage of these clumps of electrons across the space between the two grids is equivalent to the passage of a succession of pulses of electric current. It is necessary only that a potential difference exist across the grids in such a time phase as to oppose the passage of these electrons in order to derive a power output. Such a voltage will automatically result if the second resonator is tuned to the frequency of the input signal.

The situation is remarkably similar to that existing in a class C amplifier at lower frequencies. The grid produces a current in the plate-cathode circuit that flows in periodic short pulses. These pulses act upon the tuned tank circuit to build up a large voltage which opposes the flow of current and so produces the useful power output.

It is apparent that the flexibility of the Klystron is seriously limited by the fact that cavity resonators are permanently attached to the grids and thus form an integral part of the tube itself. By their very nature these resonators are adjustable over only a narrow frequency range, and accordingly the entire tube is limited to this particular narrow band for which the resonators are designed. Ordinary triodes and pentodes, on the other hand, are not so limited and therefore operate over a wide ratio of frequencies as controlled by the external oscillatory circuits or resonators which are attached.

A pair of grids suitable for producing a velocity-modulated beam in a Klystron have an area in the order of 1 sq cm and a spacing of approximately 1 mm. The net capacitance of such a pair is in the order of $1\mu\mu f$ (10^{-12} farad). If the wavelength to be produced is approximately 9 cm the frequency is 3.3×10^9 cycles per second and the natural angular velocity is 2×10^{10}. Using the basic relation

$$\omega_0^2 = \frac{1}{LC} \qquad [17\cdot25]$$

we have

$$4 \times 10^{20} = \frac{1}{10^{-12}L} \qquad [17\cdot26]$$

or

$$L = \tfrac{1}{4} \times 10^{-8} = 0.0025\ \mu\text{h} \qquad [17\cdot27]$$

It is immediately evident that such a low value of inductance is not readily achieved and that the simultaneous achievement of a high Q is not possible by ordinary methods. It is rather illluminating to see how one form of cavity resonator results as a logical extension of a low-frequency oscillatory circuit. This process is illustrated in Fig. 17·14.

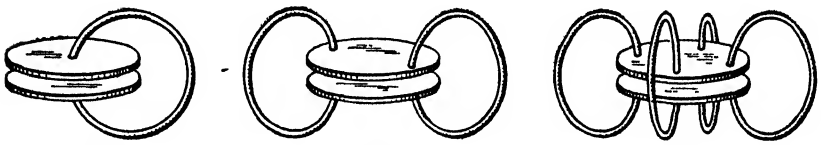


FIG. 17·14 Development of toroidal cavity resonator from parallel-plate condenser and single-loop coil.

The simple circuit of Fig. 17·14*a* is well known as the Hertzian oscillator and is readily recognized as a lumped capacitance associated with a single-turn inductor. At frequencies up to several hundreds of megacycles such a resonator is quite practical. At higher frequencies, however, the coil is necessarily small and the Q is degraded by radiation loss. The arrangement of Fig. 17·14*b* is capable of operation at somewhat higher frequencies because the two coils are essentially in parallel and the effective inductance is halved. Also radiation loss is reduced because each coil tends to cancel the field of the other.

In Fig. 17·14*c* two more loops of wire are added with an additional improvement in Q and reduction of inductance. The limit which is approached as more and more loops of wire are added is the cavity resonator in the form described by the Varians in their original paper. In such a resonator the electric current may still be thought of as flowing along the direction of the wires of the original structure. The magnetic field produced is confined entirely within the resonator, and accordingly no radiation loss exists. Typical field distributions are shown in Fig. 17·15. The Q of such systems is therefore limited only by the conductivity of the metal and is readily made quite high. The Q of a resonator without external loading is in the order of 10,000, and the external load is typically adjusted to such a value as to reduce this to about 1000.

The conclusion to be drawn from the foregoing is that the use of the cavity resonator is necessary in order to achieve a satisfactory tuned circuit at the desired frequency. Moreover, the frequency range over which any particular Klystron is operative depends to some extent upon the spacing between the two pairs of grids. If the frequency is

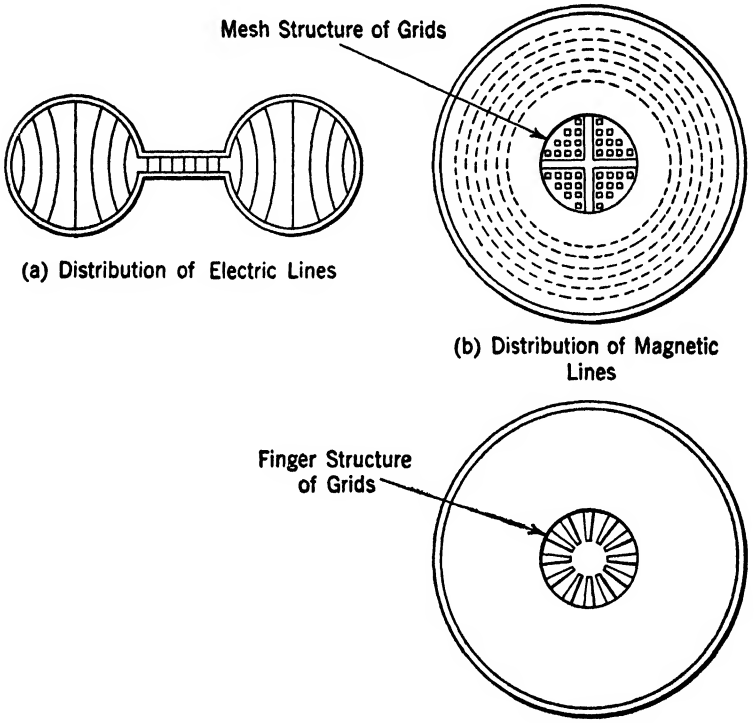


FIG. 17·15 Sections of a form of cavity resonator showing electric and magnetic lines.

relatively low the process of conversion from velocity modulation to intensity modulation requires a length that is excessive. Finally, the use of the cavity resonator is advantageous in that the high values of Q serve to give good frequency stability when the unit is used as an oscillator.

The resonators shown in Fig. 17·5 differ in form from those just shown, and the effective Q is probably reduced by this alteration. The mechanical structure, however, is improved thereby, and the modification of the electrical performance is not serious. Evidently the basic nature of the resonance as illustrated in Fig. 17·15 is not changed.

17·12 Design and Applications of the Klystron

Since the Klystron is probably the most flexible and the most important of the velocity-modulated tubes, it is well to consider its practical operation in more detail. In particular there is no other device available at the present time which successfully replaces the Klystron as an amplifier of weak signals at hyper frequencies. Let us examine this operation with a consideration of the design features involved.

The past discussion dealt primarily with electron beams in which the velocity modulation was relatively large. Such beams upon drifting only a short distance are converted into beams which are fully intensity modulated. This condition corresponds to maximum power output and is desirable in the oscillator or power amplifier.

In the low-power amplifier a very small signal is applied to the grids of the buncher and relatively small velocity modulation results. For any moderate length of drift tube such a condition results in a beam having relatively small, nearly sinusoidal, intensity modulation. As the degree of velocity modulation is increased the corresponding intensity modulation becomes less sinusoidal and of greater amplitude. This deviation from a sinusoidal current wave is of great interest since it leads to the possibility of frequency multiplication or of certain forms of modulation. It may be shown that the Klystron as a frequency doubler suffers somewhat less severe drawbacks than the ordinary class C triode doubler at lower frequency. When the Klystron is operating under ordinary conditions as an amplifier or oscillator, the buncher and catcher rhumbatrons are tuned to the same frequency. In frequency multiplier applications, however, the catcher is tuned to some harmonic of the buncher frequency.

A rigorous mathematical treatment of the action of the Klystron is very difficult since the action becomes non-linear at relatively small signal strengths. Also space charge and a variety of other effects greatly complicate the problem. Space charge is particularly important in two ways. It tends to spread any form of electron beam because each electron repels those beside it. This action causes electrons to deviate from the main path and to be lost on the inner surface of the drift space. The use of a longitudinal uniform magnetic field causes the electrons to move in helical paths and renders this effect unimportant.

Electrons also exert repulsive forces upon those electrons which precede and follow them in the beam. This action cancels to zero in a uniform beam and so tends to reduce an intensity-modulated beam into a uniform one. In a drift tube, therefore, the tendency of a small velocity modulation to produce an intensity-modulated beam is opposed

by the action of space charge. Hahn* treats this problem in great detail, and his work is recommended to the serious student of velocity-modulated devices.

Two major limitations on the performance of these units are set by the effects of space charge. The intensity modulation produced by a given small signal may not be indefinitely increased by increasing the length of the drift tube since loss by space charge presently exceeds the gain by drift. The current density in the electron beams may not be increased above a certain definite value so that a specific maximum power output is obtainable with a structure of given size.

In the Klystron tubes at present commercially available the catcher and buncher Rhumbatrons are made slightly adjustable by corrugating their side walls. Adjustment screws are provided to expand or contract these walls, thus changing the volume and shape of the cavities. The Rhumbatrons are thus exposed and subject to adjustment while operating, and since they are at a high positive potential with respect to the cathode it is customary to ground the positive terminal of the voltage supply. If good frequency stability is desired it is necessary to regulate both the plate and filament voltages. The grid bias may be obtained from a voltage dropping resistor connected across the plate voltage supply when the operating plate voltages are under 1000 volts. At higher operating voltages it is better to employ a separate source.

Before a Klystron oscillator can be set in operation it is necessary to tune both buncher and catcher to the same frequency. In view of the high selectivity of these elements, this may be quite difficult to do, because the correct operating voltage is not known until the system is in oscillation. The problem may be greatly simplified by inserting an alternating voltage in series with the direct voltage plate supply. The amplitude of this voltage should be sufficient to insure that the range of variation of the pulsations covers one of the operating potentials. See Fig. 17·12.

17·13 The Reflex Klystron Oscillator

The reflex Klystron oscillator shown in Fig. 17·16 employs a single Rhumbatron which performs the functions of both buncher and catcher. The electrons emerging from the buncher are turned back into the buncher by the plate which is held at a negative potential. When the system is properly adjusted so that the electrons returning to the cavity

* W. C. Hahn, "Small Signal Theory of Velocity-Modulated Electron Beams," *Gen. Elec. Rev.*, **42**, 258, 1939; "Wave Energy and Transconductance of Velocity Modulated Beams," *Gen. Elec. Rev.*, **42**, 497, 1939.

resonator deliver their energy to it in the correct phase, oscillations are sustained. In many respects the reflex Klystron is similar to the positive grid oscillators discussed in Chapter 12.

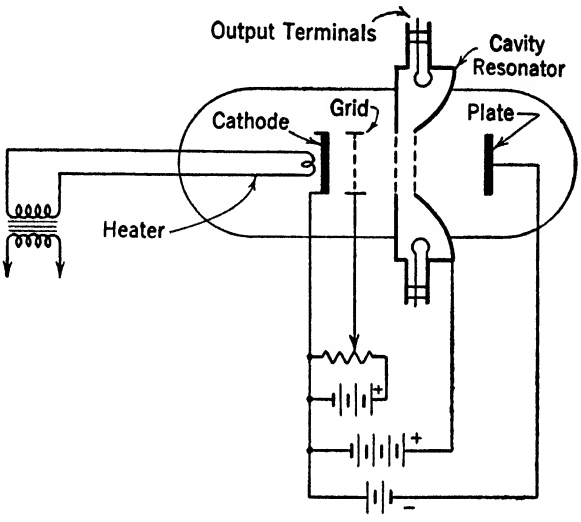


FIG. 17·16 Reflex Klystron oscillator.

17·14 The Inductive Output Amplifier*

We shall now describe another tube which resembles the Klystron in some respects, but in other respects is quite different. It appears to promise considerable commercial importance in the band of frequencies between 100 and 1000 megacycles. It is not a velocity-modulated tube in the same sense as the Klystron because the cathode current is controlled directly by a grid of the ordinary type. An electron beam is used, however, and the output is derived from a cavity resonator in a manner very similar to that in the Klystron. Because of the relatively long path used it has been found necessary to provide a certain amount of magnetic focusing to prevent undue spreading of the electron beam.

Figure 17·17 shows the basic elements of the system, and Fig. 17·18 illustrates the additional features embodied in the practical tube. The number of electrons which leave the cathode is controlled by the grid. After leaving the grid the electrons are formed into a beam and accelerated by the high voltage applied to the electrodes *c* and *d*. This high-velocity beam which is intensity modulated by the control grid moves

* A. V. Haeff and L. S. Neygaard, "A Wide-Band Inductive Output Amplifier," *Proc. IRE*, **28**, 126, March, 1940; A. V. Haeff, "An Ultra-High-Frequency Power Amplifier of Novel Design," *Electronics*, **12**, 30, March, 1939.

past the aperture or slot in the output resonator. This resonator differs somewhat in form from that of the Klystron and in particular lacks the capacitance formed by the parallel-grid structure. Accordingly, it has a higher effective L/C ratio and therefore a higher impedance than that of the Klystron resonator. The mechanism of excitation is the same in

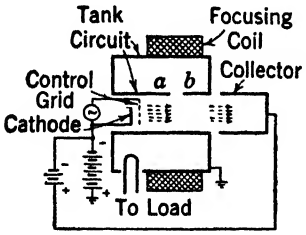


FIG. 17·17 Essential elements of the inductive output tube.

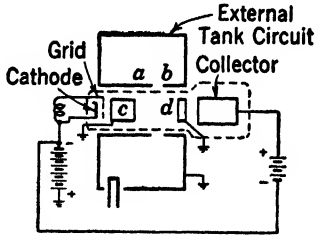


FIG. 17·18 Experimental form of the inductive output tube.

both. The potential difference existing across the gap creates a field which retards the electrons, thus absorbing power from them. Electrons which pass through the resonator structure are somewhat decelerated before being captured by the final anode or collector. Thus the plate dissipation is reduced and the overall efficiency is improved.

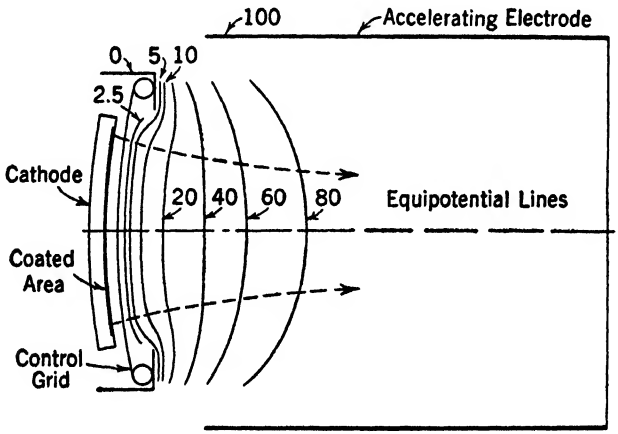
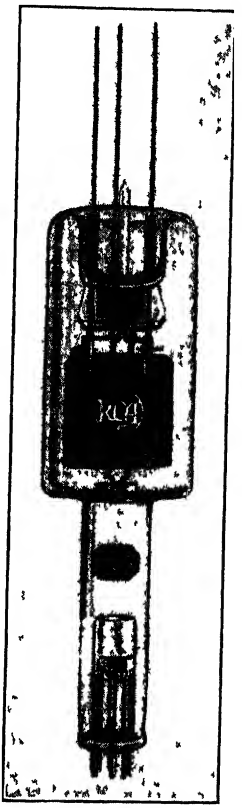


FIG. 17·19 Control grid and cathode structure of the inductive output tube.

Figure 17·19 shows the arrangement of cathode, grid, and accelerating electrodes with equipotential lines expressed in per cent of the potential applied. The control grid is of relatively fine mesh and is placed very close to the cathode so that a large value of transconductance is ob-

tained. The high value of the accelerating field and the small spacings result in a very short transit time and therefore reduce the effective input conductance to a reasonable value. Figure 17·20 shows the commercial model of this tube.

It is appropriate to investigate the exact source of the output power in this tube, for the result is not obvious. Electrons are emitted from



(Courtesy of RCA Radiotron Company)

FIG. 17·20 Commercial form of the inductive output tube.

the cathode in bunches under the action of the control grid. These bunches are accelerated to a high velocity by the action of the accelerating electrodes, but a negligible number reach these electrodes. The electrons are decelerated by the relatively high voltage existing across the gap of the resonator and are finally collected by the plate. If the accelerating voltage is designated V_a and the voltage across the output resonator is V_0, then electrons which leave the resonator have a velocity

corresponding to an applied voltage of $V_a - V_0$. They are, however, still in a region of potential equal to V_a, and therefore a plate or collector voltage at least V_0 above the cathode voltage is necessary if the electrons are to be captured at the plate. Thus we conclude that the output power is taken directly from the plate supply voltage and that a good conversion efficiency is at least theoretically possible.

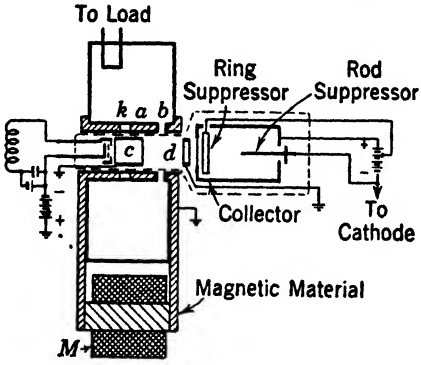


FIG. 17·21 Circuit for use of the 825 inductive output tube.

The arrangement of Fig. 17·21 using the 825 tube shown in Fig. 17·20 gives the following experimental results:

Accelerating potential	3000 volts
Beam current	40 ma
Transconductance	6000 micromhos
Input capacitance	6 $\mu\mu$f
Frequency	500 mc
Effective band width	10 mc
Equivalent shunt capacitance of output resonator	2 $\mu\mu$f
Input of power	1 watt
Output of power	10 watts
Efficiency	25 per cent

Similar tubes under other conditions give larger power outputs and somewhat larger values of high-frequency power amplification.

The low equivalent shunt capacitance of the output resonator mentioned previously has an important practical significance. In the section on video amplifiers it was shown that a low shunting capacitance is necessary if high amplification is to be obtained over wide bands of frequencies. This proposition is a perfectly general one, equally applicable here. At these frequencies, however, it is seldom practical to use the relatively elaborate compensating networks discussed under video amplifiers. Accordingly wide-band amplification is achieved most

directly by reduction of the capacitance. In the present instance a band width of 10 megacycles was achieved by reduction of the effective load impedance. An important property of this particular tube is that the efficiency is not seriously affected by such an adjustment of the impedance.

17·15 Associated Tubes

A variety of tubes more or less similar to the Klystron are described by Hahn and Metcalf.* They use grid structures such as are shown in Figs. 17·1 and 17·2 and utilize drift and retarding field methods of converting velocity modulation into intensity modulation. These tubes are used as oscillators, amplifiers, and modulators at frequencies from 60 to 6000 megacycles. Certain of the tubes give large power outputs, good efficiency, and excellent stability.

One of the first to publish on the subject of velocity modulation tubes was Heil.† His arrangement is not greatly dissimilar to the Klystron. Two pairs of grids are used and a drift space between the two transforms velocity modulation into intensity modulation. Only one resonator rather than two is used, however, and the device serves only as an oscillator and not as an amplifier or modulator.

17·16 Summary

The velocity-modulated tube, in common with the positive-grid and magnetron oscillators, depends for its operation upon electron transit time. The positive-grid tube serves only as an oscillator, no practical arrangements for producing amplification being known. As an oscillator the efficiency is low because the transit time is affected by the existence of an output voltage. The undesired phenomenon of phase selection sets in to limit the output and efficiency. Moreover, the output frequency depends upon the applied voltage and the tube structure as well as upon the tuning of the resonant circuit.

The magnetron as a transit time oscillator has two major advantages over the positive-grid generator. The direct and important loss of electrons to the grid mesh does not exist, and the process of phase selection is absent. Accordingly, the magnetron is ideally capable of 100 per cent efficiency. Although this ideal is not closely approached in practice, the efficiencies of commercial magnetrons are high in comparison to the efficiencies of positive-grid tubes. In the magnetron the

* W. C. Hahn and G. F. Metcalf, "Velocity Modulated Tubes," *Proc. IRE*, **27**, 106, February, 1939.

† O. Heil and H. Arsenjewa-Heil, "A New Method of Producing Short Undamped Waves of Great Intensity," *Z. Physik*, **95**, 752, 1935.

frequency depends primarily upon the tube structure and the magnetic field and only secondarily upon the tuning of the resonant circuit. Relatively good frequency stability is achieved in practice, particularly if permanent magnets are used to produce the required field. No practical arrangement for the amplification of hyper frequencies by means of the magnetron has yet been disclosed.

Thus the Klystron, in common with similar velocity-modulated tubes, has the distinction of being the only known device for the amplification of hyper frequencies. This success may be attributed to the fact that transit time, although essential to the operation of the tube, does not contribute any critical time interval equivalent to a resonance. The two cavity resonators function in a manner which is essentially independent of the electron beam. Because of this independence of the resonators and the electron stream it is possible to excite one resonator with a small signal and to abstract a large signal from the second resonator.

In many respects the Klystron is simpler than other widely accepted types of tubes. No magnetic field is required, and no extremely high voltage is required. Relatively strong oscillations are produced with an applied potential of only a few hundred volts. With improvements comparable to those which have been made in other tubes there seems no reason to doubt that maximum ratings below 1000 volts will soon be available.

At the present time the most severe drawback of the Klystron is the high cost of the tube and the associated tuner. Because the tube itself is not essentially more complex than other available types it seems probable that mass-production methods will reduce the cost to a value well within the reach of all. The tuner design now employed requires machine work of high precision. It is thus intrinsically expensive. Past experience, a relatively trustworthy guide, indicates that modified designs will be produced which accomplish the same results less expensively.

In the Klystron we find a device capable of oscillation, amplification, modulation, and detection. Accordingly the situation at hyper frequencies now is closely comparable to that which existed at lower frequencies some twenty-five years ago when the triode was relatively new. The developments in communication, navigation, aviation, and industry which will result are beyond the power of any prophet to predict. We may be sure, however, that the steadily accelerating advance in the knowledge and use of electronic devices will not halt and that hyper frequencies will play a leading part in this advance.

PROBLEMS

17·1 A certain Klystron operates with an accelerating voltage of 2000. The effective spacing of the grids of the resonator is 2 mm, and the operating frequency is 2500 megacycles. Calculate the velocity with which electrons reach the grid, the ransit time between the grids in seconds, and the transit angle in degrees.

17·2 The grids of a Klystron resonator have an effective spacing of 2 mm and an effective area of 3 cm^2. Calculate the capacitance due to this portion of the resonator. What effective inductance must be contributed by the rest of the resonator if the natural frequency is 2500 megacycles?

17·3 The resonator of problem 17·2 has the annular shape shown in Fig. 17·15. Assuming that the grids are essentially equipotential surfaces and that the maximum voltage across the grids is 100 volts, calculate the peak current in the equivalent parallel resonant circuit. Calculate the total magnetic flux which must exist within the annular portion of the resonator.

17·4 An annular cavity resonator such as that shown in Fig. 17·15 operates with a peak voltage of 250 across the grids. The radius of the circular sections viewed normally in Fig. 17·15*a* is 1 cm. A loop of wire is inserted to couple the resonator to a coaxial line. If the output voltage is to be approximately 5 volts, calculate the radius of the coupling loop.

17·5 A Klystron is to be used as a frequency doubling amplifier. Discuss the tuning of the resonators, current and voltage wave shapes, and magnitudes of voltage for high-efficiency operation.

17·6 Consider the possibility of using a Klystron oscillator in which oscillation occurs at one frequency but the major power output is taken at some harmonic of this frequency. What special adjustments are necessary? Is this likely to be a desirable kind of operation?

17·7 In a normal Klystron the electrons which approach the catcher all have essentially the same velocity as defined by the accelerating voltage. If the bunching action were perfect, how large an efficiency could be produced, and what voltage swing across the catcher would be necessary? Compare with a class C amplifier.

17·8 In the so-called reflex Klystron the bunching and catching functions are performed by the same resonator. Discuss phase relations and potentials necessary for high efficiency. May the efficiency of this device be made as high as that of the normal Klystron?

17·9 A Klystron oscillator may, to a considerable extent, be amplitude-modulated by supplying the signal voltage in series with the grid which controls the average beam current. Consider the operation in terms of efficiency and linearity of modulation.

17·10 A Klystron oscillator may be frequency-modulated by applying the signal voltage in series with the accelerating voltage. The transit time is thus varied and the frequency must vary so as to compensate the resulting phase shift. Is a high or low Q desirable in the resonators for this application? Is amplitude modulation likely to result?

17·11 Discuss the possibility of producing pure amplitude modulation and pure frequency modulation by a combination of the arrangements of problems 17·9 and 17·10.

17·12 The resonators of a certain Klystron have the shape shown in Fig. 17·15. The operating frequency is 2700 megacycles when the grids are separated by 2.0 mm. How widely must the grids be spaced for a frequency of 3000 megacycles? Refer to the simple analysis of Fig. 17·14.

17·13 In the Klystron it is not even theoretically possible to achieve 100 per cent efficiency. Is this statement applicable to the inductive output tube? Why?

17·14 The Klystron is the only known device which serves as a practical amplifier at frequencies above 1000 megacycles. Discuss the use of a Klystron as a modulated amplifier.

17·15 A Klystron is to be used as a master oscillator for an application in which high frequency stability is important. Discuss the construction of the tube and the design of associated circuits to produce this result.

17·16 A longitudinal magnetic field is sometimes used to prevent the spreading of an electron beam. Sketch the motion of electrons which tend to diverge, as influenced by this sort of field.

17·17 Discuss the possibility of applying negative feedback to a Klystron amplifier in order to improve the linearity and the constancy of amplification. What difficulties are likely to be serious?

17·18 At moderate frequencies it has become common practice to combine two electrically distinct tubes in a common envelope. Discuss the possibility of extending this practice to Klystrons.

17·19 At moderate frequencies the push-pull amplifier is often used to eliminate even harmonics and other undesired effects. Discuss the possibility of using two Klystrons in push pull. What difficulties are likely to be encountered?

17·20 Consider the possibility of securing push-pull operation in a single Klystron employing special resonators and divided grids.

APPENDIX I

FUNDAMENTAL CONSTANTS

Electronic charge e

$$e = (1.600 \pm 0.002) \times 10^{-19} \quad \text{coulomb}$$

Electronic mass m

$$m = (9.156 \pm 0.018) \times 10^{-31} \quad \text{kilogram}$$

Electronic charge to mass ratio e/m

$$\frac{e}{m} = (1.7571 \pm 0.0015) \times 10^{11} \quad \text{coulomb per kilogram}$$

The constant ε_0

$$\varepsilon_0 = 8.854 \times 10^{-12} \simeq \frac{1}{36\pi} \times 10^{-9} \quad \text{farad per meter}$$

The constant μ_0

$$\mu_0 = 4\pi \times 10^{-7} \simeq 1.257 \times 10^{-6} \quad \text{henry per meter}$$

The velocity of light c

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 2.998 \times 10^8 \simeq 3 \times 10^8 \quad \text{meter per second}$$

$$\frac{1}{\varepsilon_0} = 1.129 \times 10^{11} \quad \text{meter per farad}$$

$$\frac{1}{\mu_0} = 7.958 \times 10^5 \quad \text{meter per henry}$$

$$\sqrt{\frac{\mu_0}{\varepsilon_0}} = 376.7 \quad \text{ohm}$$

$$\sqrt{\frac{\varepsilon_0}{\mu_0}} = 2.654 \times 10^{-3} \quad \text{mho}$$

APPENDIX II

Conversion Table for Units Unaffected by Rationalization

The mks unit below must be multiplied	by	to obtain the cgs electromagnetic unit of	or by	to obtain the cgs electrostatic unit of	Symbol
Ampere	10^{-1}	Current (abampere)	3×10^{9}	Current (statampere)	I
Ampere per square meter	10^{-5}	Current density (abamp/cm^2)	3×10^{5}	Current density (statamp/cm^2)	ι
Coulomb	10^{-1}	Charge (abcoulomb)	3×10^{9}	Charge (statcoulomb)	Q
Coulomb per cubic meter	10^{-7}	Charge density (abcoulomb/cm^3)	3×10^{3}	Charge density (statcoulomb/cm^3)	ρ
Farad	10^{-9}	Capacitance (abfarad)	9×10^{11}	Capacitance (statfarad)	C
Henry	10^{9}	Inductance (abhenry)	$\frac{1}{9} \times 10^{-11}$	Inductance (stathenry)	L
Joule	10^{7}	Energy (erg)	10^{7}	Energy (erg)	W
Mho per meter	10^{-11}	Conductivity (abmho/cm)	9×10^{9}	Conductivity (statmho/cm)	G
Newton	10^{5}	Force (dyne)	10^{5}	Force (dyne)	F
Ohm	10^{9}	Resistance (abohm)	$\frac{1}{9} \times 10^{-11}$	Resistance (statohm)	R
Volt	10^{8}	Potential (abvolt)	$\frac{1}{3} \times 10^{-2}$	Potential (statvolt)	V
Weber	10^{8}	Magnetic flux (maxwell)	$\frac{1}{3} \times 10^{-2}$	Magnetic flux (statweber)	ϕ
Weber per square meter	10^{4}	Magnetic flux density (gauss)	$\frac{1}{3} \times 10^{-6}$	Magnetic flux density (statweber/cm^2)	B

Conversion Table for Several Units Affected by Rationalization

The rationalized mks unit below must be multiplied	by	to obtain the unrationalized cgs electromagnetic unit of	or by	to obtain the unrationalized cgs electrostatic unit of	Symbol
Ampere-turn	$4\pi \times 10^{-1}$	Magnetomotive force (gilbert)	$12\pi \times 10^{9}$	Magnetomotive force	MMF
Ampere-turn per meter	$4\pi \times 10^{-3}$	Magnetic field intensity (oersted)	$12\pi \times 10^{7}$	Magnetic field intensity	H
Mks unit of electric displacement	$4\pi \times 10^{-5}$	Electric displacement	$12\pi \times 10^{5}$	Electric displacement	D

BIBLIOGRAPHIES

General Communications: Textbooks and Articles

ALBERT, A. L., *Electrical Communication*, John Wiley & Sons, 1940.

ALBERT, A. L., *Electrical Fundamentals of Communication*, McGraw-Hill Book Company, 1942.

ALBERT, A. L., *Fundamental Electronics and Vacuum Tubes*, The Macmillan Company, 1938.

BERGMANN, L., "Elektromagnetische Felder und Schwingungen" (Electromagnetic Fields and Oscillations), *Physica*, vol. 9, No. 1, pp. 1–13, 1941.

BRAINERD, J. G., KOEHLER, G., REICH, H. J., and WOODRUFF, I., *Ultra High Frequency Techniques*, D. Van Nostrand Company, 1942.

CLEMENT, L. M., and others, *Radio at Ultra-High Frequencies*, R.C.A. Institute Technical Press, 1940.

DOW, W. G., *Fundamentals of Engineering Electronics*, John Wiley & Sons, 1937.

EASTMAN, A. V., *Fundamentals of Vacuum Tubes*, McGraw-Hill Book Company, 1941.

Electronics, April, 1942, issue devoted to U.H.F.

ARRL Handbook, American Radio Relay League.

ESHBACH, O., *Handbook of Engineering Fundamentals*, John Wiley & Sons, 1936.

EVERITT, W. L., *Communication Engineering*, McGraw-Hill Book Company, 1937.

FINK, D. G., *Engineering Electronics*, McGraw-Hill Book Company, 1938.

FINK, D. G., *Principles of Television Engineering*, McGraw-Hill Book Company, 1940.

GLAGOLEWA-ARKADIEWA, A., "Short Electromagnetic Waves of a Wavelength up to 82 Microns," *Nature*, vol. 113, p. 640, May, 1924.

GLASGOW, R. S., *Principles of Radio Engineering*, McGraw-Hill Book Company, 1936.

GUILLEMIN, E. A., *Communication Networks*, John Wiley & Sons, vol. 1, 1931; vol. 2, 1935.

HARNWELL, G. P., *Principles of Electricity and Electromagnetism*, McGraw-Hill Book Company, 1938.

HARVEY, A. F., *High Frequency Thermionic Tubes*, John Wiley & Sons, 1943.

HENNEY, K., *Principles of Radio*, John Wiley & Sons, 1942.

HENNEY, K., *Radio Engineering Handbook*, McGraw-Hill Book Company, 1940.

HOAG, J. B., *Basic Radio*, D. Van Nostrand Company, 1942.

HUND, A., *High Frequency Measurements*, McGraw-Hill Book Company, 1933.

HUND, A., *Phenomena in High Frequency Systems*, McGraw-Hill Book Company, 1936.

JORDAN, E. C., and others, *Fundamentals of Radio*, Prentice-Hall, 1942.

LLEWELLYN, F. B., *Electron-Inertia Effects*, Cambridge University Press, 1941.

KELSEY, E., *Reference Guide to Ultra High Frequencies*, Zenith Radio Corporation, 1942.

MARCONI, G., "Radio Micro-Waves," *Electrician*, vol. 110, p. 3, Jan. 6, 1933.

MCILWAIN, K., and BRAINERD, J. G., *High-Frequency Alternating Currents*, John Wiley & Sons, 1939.

MILLMAN, J., and SEELY, S., *Electronics*, McGraw-Hill Book Company, 1941.
M.I.T. STAFF, *Applied Electronics*, John Wiley & Sons, 1943
M.I.T. STAFF, *Electric Circuits*, John Wiley & Sons, 1940.
MORGAN, H. K., *Aircraft Radio and Electrical Equipment*, Pitman, London, 1941.
PENDER, H., and MCILWAIN, K., *Handbook for Electrical Engineers*, John Wiley & Sons, 1936.
PIERCE, G. W., *Electric Oscillations and Electric Waves*, McGraw-Hill Book Company, 1920.
REICH, H. J., *Principles of Electron Tubes*, McGraw-Hill Book Company, 1941.
REICH, H. J., *Theory and Application of Electron Tubes*, McGraw-Hill Book Company, 1939.
SANDRETTO, P. C., *Principles of Aeronautical Radio Engineering*, McGraw-Hill Book Company, 1942.
SCHELKUNOFF, S. A., *Electromagnetic Waves*, D. Van Nostrand Company, 1943.
SCHELKUNOFF, S. A., "Transmission Theory of Plane Electromagnetic Waves," *Proc. IRE*, vol. 25, p. 1457, November, 1937.
SCHELKUNOFF, S. A., "Theory of Antennas of Arbitrary Size and Shape," *Proc. IRE*, vol. 29, p. 493, September, 1941.
SLATER, J. C., *Microwave Transmission*, McGraw-Hill Book Company, 1942.
SMYTHE, W. R., *Static and Dynamic Electricity*, McGraw-Hill Book Company, 1939.
STARLING, S. G., *Electricity and Magnetism*, Longmans, Green and Company, 1939.
STRATTON, J. A., *Electromagnetic Theory*, McGraw-Hill Book Company, 1941.
TERMAN, F. E., *Measurements in Radio Engineering*, McGraw-Hill Book Company, 1939.
TERMAN, F. E., *Radio Engineering*, McGraw-Hill Book Company, 1938.
WARE, L. A., and REED, H. R., *Communication Circuits*, John Wiley & Sons, 1942.
WILSON, J. C., *Television Engineering*, Pitman, London, 1937.
ZWORYKIN, V. K., and MORTON, G. A., *Television*, John Wiley & Sons, 1940.

Bibliography for Chapters 1 to 4

BENNETT, E., and CROTHERS, H. M., *Introductory Electrodynamics for Electrical Engineers*, McGraw-Hill Book Company, 1926.
BROWN, T. B., *Foundations of Modern Physics*, John Wiley & Sons, 1940.
BURINGTON, R. S., and TORRANCE, C. C., *Higher Mathematics*, McGraw-Hill Book Company, 1939.
EHRENHAFT, F., "Photophoresis and Its Interpretation by Electric and Magnetic Ions," *J. Franklin Inst.*, vol. 232, p. 235, 1942.
FRANKLIN, W. S., and TERMAN, F. E., *Transmission Line Theory*, Franklin and Charles, 1926. This book gives a very simple treatment of filter theory.
HARNWELL, G. P., *Principles of Electricity and Electromagnetism*, McGraw-Hill Book Company, 1938.
KELLY, H. C., *Electricity and Magnetism*, John Wiley & Sons, 1941.
MCILWAIN, K., and BRAINERD, J. G., *High-Frequency Alternating Currents*, John Wiley & Sons, 1939.
PAGE, L., and ADAMS, N. I., *Electrodynamics*, D. Van Nostrand Company, 1940.
PAGE, L., and ADAMS, N. I., *Principles of Electricity*, D. Van Nostrand Company, 1934.
PAGE, L., *Introduction to Theoretical Physics*, D. Van Nostrand Company, 1935.
PIERCE, G. W., *Electric Oscillations and Electric Waves*, McGraw-Hill Book Company, 1920.

PLANCK, M. K. E. L., *Theory of Electricity and Magnetism*, The Macmillan Company, 1932.

SKILLING, H. H., *Fundamentals of Electric Waves*, John Wiley & Sons, 1942.

SLATER, J. C., and FRANK, N. H., *Introduction to Theoretical Physics*, McGraw-Hill Book Company, 1933.

SMYTHE, W. R., *Static and Dynamic Electricity*, McGraw-Hill Book Company, 1939.

SOKOLNIKOFF, I. S. and E. S., *Higher Mathematics for Engineers and Physicists*, McGraw-Hill Book Company, 1934. This book gives a simple treatment of conjugate functions.

STARLING, S. G., *Electricity and Magnetism*, Longmans, Green and Company, 1939.

STRATTON, J. A., *Electromagnetic Theory*, McGraw-Hill Book Company, 1941.

Bibliography for Chapters 5 to 8

ADAMS, N. I., and PAGE, L., "Electromagnetic Waves in Conducting Tubes," *Phys. Rev.*, vol. 52, p. 647, 1937.

AWENDER, H., and LANGE, O., "The Propagation of Decimetre and Centimetre Waves along Single Metallic and Dielectric Wires," *Funktech. Monat.*, vol. 1, p. 5, and vol. 2, p. 51, 1938.

BARROW, W. L., "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," *Proc. IRE*, vol. 24, p. 1298, October, 1936.

BARROW, W. L., and SCHAEVITZ, H., "Hollow Pipes of Relatively Small Dimensions," *Elec. Eng. Trans.*, vol. 60, p. 119, March, 1941.

BRILLOUIN, L., "Theoretical Study of Dielectric Cables," *Elec. Com.*, vol. 16, p. 350, April, 1938.

BRILLOUIN, L., "Hyperfrequency Waves and Their Practical Use," *Jour. Franklin Institute*, vol. 229, p. 709, June, 1940.

BRILLOUIN, L., "Propagation d'ondes électromagnetiques dans un tuyau," *Rev. gén. élec.*, August, 1936.

BRILLOUIN, L., "Propagation of Waves along Dielectric Cables," *Bull. soc. des élec.*, vol. 8, p. 899, 1938.

BRILLOUIN, L., and CLAVIER, M., "Propagation in Dielectric Cables and Circular Guide Cables," *Génie civil*, vol. 113, pp. 504, 505, 1938.

BUCHHOLZ, H., "The Influence of the Curvature of Rectangular Hollow Conductors on the Phase Constant," *E.N.T.*, vol. 16, p. 73, 1939.

BUCHHOLZ, H., "The Quasi-optics of Ultra-short Wave Guides," *E.N.T.*, vol. 15, p. 297, 1938.

CARSON, J. R., MEAD, S. P., and SCHELKUNOFF, S. A., "Hyperfrequency Wave Guides — Mathematical Theory," *Bell Sys. Tech. J.*, vol. 15, p. 310, April, 1936.

CHU, L. J., "Electromagnetic Waves in Elliptic Hollow Pipes of Metal," *J. Applied Physics*, vol. 9, p. 583, 1938.

CHU, L. J., and BARROW, W. L., "Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section," *Proc. IRE*, vol. 26, p. 1520, December, 1938.

CLAVIER, A. G., "Theoretical Relationships of Dielectric Guides," *Elec. Com.*, vol. 17, p. 276, January, 1939.

CLAVIER, A. G., "Theory of Cylindrical Dielectric Wave Guides and Coaxial Cables," *Bull. soc. des élec.*, vol. 8, p. 355, 1938.

CLAVIER, A. G., and ALTOVSKY, V., "Experimental Research on the Propagation of Electromagnetic Waves in Cylindrical Guides," *Rev. gén. élec.*, vol. 45, p. 697, 1939.

CLAVIER, A. G., and ALTOVSKY, V., "Experimental Researches on the Propagation of Electromagnetic Waves in Dielectric Guides," *Elec. Comm.*, vol. 18, p. 81, 1940.

CLAVIER, A. G., and ROSTAS, E., "Some Problems of Hyperfrequency Technique," *Elec. Com.*, vol. 16, p. 269, January, 1938.

DEBYE, P., and HONDROS, O., "Electromagnetic Waves in Dielectric Wires," *Ann. Physik*, vol. 32, p. 465, 1920.

DROSTE, H. W., "Ultra High Frequency Transmission along Cylindrical Conductors and Non-Conductors," *T.F.T.*, vol. 27, pp. 199, 273, 310, and 337, 1938.

HARTIG, H. E., and MELLOH, A. W., "The Transmission of Damped Electromagnetic Waves through Small Hollow Metal Tubes," *Phys. Rev.*, vol. 54, p. 646, 1938.

HOWE, G. W. O., "A New Type of Wave Transmission," *Wireless Engineer*, vol. 13, p. 291, 1936.

HOWE, G. W. O., "Electromagnetic Waves in Rectangular Metal Tubes," *Wireless Engineer*, vol. 19, p. 93, March, 1942.

IWAKATA, H., "On the Relation between the Inherent Value (of the Electromagnetic Field) of Hollow Metal Tubes and Their Miscellaneous Constants," *Electrotech. Jour.* (*Tokyo*), vol. 5, pp. 58–59, March, 1941.

KAŠPAR, E., "Experimental Investigation of Electromagnetic Waves on Dielectric Wires," *Ann. Physik*, vol. 32, p. 353, 1938.

KLEMT, A., "Contribution to the Propagation of Electromagnetic Waves along Dielectric Wires," *Funktech. Monat.*, vol. 4, p. 122, 1939.

MALOW, N. N., "Elektromagnetische Wellen in einem Hohlleiter mit veranderlichem Schnitte" (Electromagnetic waves in a hollow conductor of variable cross section), *Jour. Phys.* (*U.S.S.R.*), vol. 1, No. 5, pp. 473–478, 1941.

MELLOH, A. W., "Damped Electromagnetic Waves in Hollow Metal Pipes," *Proc. IRE*, vol. 28, p. 179, 1940.

LORD RAYLEIGH, "On the Passage of Electric Waves through Tubes," *Phil. Mag.*, vol. 43, p. 125, February, 1897.

REIDEL, H., "The Metallic Hollow Conductor as a Transmitter of Electromagnetic Waves," *Hochf. Tech. u. Elek. Akus.*, vol. 53, p. 122, 1939.

RYTOV, S. M., "On the Attenuation of Electromagnetic Waves in Tubes," *J. Phys. U.S.S.R.*, vol. 2, p. 187, 1940.

SAPHORES, J., "General Properties of Dielectric Guides," *Elec. Com.*, vol. 16, p. 346, April, 1938.

SCHELKUNOFF, S. A., "Note on Certain Guided Waves in Slightly Non-Circular Tubes," *J. Applied Physics*, vol. 9, p. 484, 1938.

SCHELKUNOFF, S. A., "Electromagnetic Waves in Conducting Tubes," *Phys. Rev.*, vol. 52, p. 1078, 1937.

SCHRIEVER, O., "Electromagnetic Waves in Dielectric Conductors," *Ann. Physik*, vol. 63, p. 645, 1920.

SCHRIEVER, O., "On the Use of Waves in Tubes as Transmission Channels," *Telefunken. Hausmittel.*, vol. 20, p. 55, 1920.

SONADA, S., "Research on Wave Guides and Electromagnetic Horns," *Electrotech. J. Tokyo*, vol. 4, pp. 35 and 41, 1940.

SONADA, S., MORIMOTO, S., and ITO, M., "Transmission along Circularly Curved Wave Guides," *Electrotech. J. Tokyo*, vol. 3, p. 215, 1939; vol. 4, p. 47, 1940.

SONADA, S., "Electromagnetic Waves Propagating through Metallic Tubes of Sectoral Section," *Electrotech. J. Tokyo*, vol. 1, p. 214, 1937.

SOUTHWORTH, G. C., "Some Fundamental Experiments with Wave Guides," *Proc. IRE*, vol. 25, p. 807, July, 1937.

SOUTHWORTH, G. C., "Hyper Frequency Wave Guides," *Bell Sys. Tech. J.*, vol. 15, p. 284, 1936.

SOUTHWORTH, G. C., "New Experimental Methods Applicable to Ultra Short Waves," *J. Applied Physics*, vol. 8, p. 660, 1937.

SOUTHWORTH, G. C., "Hyperfrequency Wave Guides — General Considerations and Experimental Results," *Bell Sys. Tech. J.*, vol. 15, p. 284, April, 1936.

SOUTHWORTH, G. C., "Wave Guides for Electrical Transmission," *Elec. Eng.*, vol. 57, p. 91, March, 1938.

SOUTHWORTH, G. C., "A Demonstration of Guided Waves," *Bell Lab. Rec.*, vol. 18, p. 194, 1940.

ZAHN, H., "Detection of Electromagnetic Waves in Dielectric Wires," *Ann. Physik*, vol. 49, p. 907, 1916.

Bibliography for Chapter 9

BARROW, W. L., "Measurement of Radio Frequency Impedance with Network Simulating Lines," *Proc. IRE*, vol. 23, p. 807, 1935.

BUCHHOLZ, H., "Ultra Short Waves in Concentric Cables and Hollow Space Resonators," *Hochf. Tech. u. Elek. Akus.*, vol. 54, p. 161, 1939.

BURROWS, C. R., "Exponential Transmission-Line," *Bell Lab. Rec.*, vol. 18, p. 174, February, 1940.

CARTER, P. S., "Charts for Transmission-Line Measurements and Computations," *R.C.A. Rev.*, vol. 3, p. 355, January, 1939.

CONKLIN, J. W., FINCH, J. L., and HANSELL, C. W., "New Methods of Frequency Control Employing Long Lines," *Proc. IRE*, vol. 10, pp. 1918–1930, November, 1931.

DARBORD, R., "Reflectors and Transmission Lines for 18 cm. Waves," *l'Onde électrique*, vol. 11, p. 53, February, 1932.

GREEN, E. I., LEIBE, F. A., and CURTIS, H., "Proportioning of Shielded Circuits for Minimum High Frequency Attenuation," *Bell Sys. Tech. J.*, vol. 15, p. 248, April, 1936.

HANSELL, C. W., "Resonant Lines for Frequency Control," *Elec. Eng.*, vol. 54, p. 852, August, 1935.

HANSELL, C. W., and CARTER, P. S., "Frequency Control by Low Power Factor Line Circuits," *Proc. IRE*, vol. 24, p. 597, April, 1936.

HANSEN, W. W., "On the Resonant Frequency of Closed Concentric Lines," *J. Applied Phys.*, vol. 10, p. 38, January, 1939.

HIKOSABURO, A., "Ellipse Diagram of a Lecher Wire System," *Proc. IRE*, vol. 21, p. 303, February, 1933.

HULL, R. A., "Stabilizing the Ultra-High Frequency Transmitter," *QST*, vol. 19, p. 13, February, 1935.

KING, R., "General Amplitude Relations for Transmission Lines with Unrestricted Line Parameters, Terminal Impedances and Driving Point," *Proc. IRE*, vol. 29, p. 640, December, 1941.

KING, R., "Wavelength Characteristics of Coupled Circuits Having Distributed Constants," *Proc. IRE*, vol. 20, p. 1368, August, 1932.

KING, R., "A General Reciprocity Theorem for Transmission Lines at U.H.F.," *Proc. IRE*, vol. 28, p. 223, May, 1940.

KING, R., "A Generalized Coupling Theorem for Ultra-High Frequency Circuits," *Proc. IRE*, vol. 28, p. 84, February, 1940.

KING, R., "Application of Low Frequency Circuit Analysis to the Problem of Distributed Coupling in u.h.f. Circuits," *Proc. IRE*, vol. 27, p. 715, November, 1939.

KOLSTER, F. A., "High Q Tank Circuits for U.H.F.," *QST*, vol. 18, p. 69, May, 1934.
MOHAMMED, A., and KANTEBET, S. R., "Formation of Standing Waves on Lecher Wires," *Proc. IRE*, vol. 19, p. 1983, November, 1931.
NERGAARD, L. S., and SALZBERG, B., "Resonant Impedance of Transmission Lines," *Proc. IRE*, vol. 27, p. 579, 1939.
PIPES, L. A., "An Operational Treatment of Electromagnetic Waves Along Wires," *J. Applied Phys.*, vol. 12, p. 800, November, 1941.
PIPES, L. A., "Steady-State Analysis of Multiconductor Transmission-Lines," *J. Applied Phys.*, vol. 12, p. 782, November, 1941.
REUKEMA, L.E., "Transmission Lines at Very High Radio Frequencies," *Elec. Eng.*, vol. 56, p. 1002, 1937.
SALZBERG, B., "On the Optimum Length of Transmission Lines Used as Circuit Elements," *Proc. IRE*, vol. 25, p. 1561, December, 1937.
SCHELKUNOFF, S. A., "Coaxial Communication Transmission Lines," *Elec. Eng.*, vol. 53, p. 1592, December, 1934.
SMITH, P. H., "Transmission Line Calculator," *Electronics*, vol. 12, p. 29, January, 1939.
STERBA, E. J., and FELDMAN, C. B., "Transmission Lines for Short Wave Radio Systems," *Proc. IRE*, vol. 20, pp. 1163–1202, July, 1932.
TERMAN, F. E., "Resonant Lines in Radio Circuits," *Elec. Eng.*, vol. 53, p. 1046, July, 1934.
WHITMER, R., "Radiation Resistance of Concentric Conductor Transmission Lines," *Proc. IRE*, vol. 21, p. 1343, September, 1933.

Bibliography for Chapter 10

BARROW, W. L., and MIEHER, W. W., "Natural Oscillations of Electrical Cavity Resonators," *Proc. IRE*, vol. 28, p. 184, April, 1940.
BERG, T. G. O., "Teorien för den Sfäriska Resonatorn" (The theory of spherical resonators), *Teknisk Tidskrift*, vol. 49, pp. 200–204, December 7, 1940.
BORGNIS, F., "The Fundamental Electrical Oscillations of Cylindrical Cavities," *Hochf. Tech. u. Elek. Akus.*, vol. 54, p. 121, 1939.
BORGNIS, F., "Natural Electromagnetic Oscillations of Dielectric Spaces," *Ann. Physik*, vol. 35, p. 359, 1939.
BUNIMOVICH, V. I., "An Oscillating System with Small Losses," *J. Tech. Phys.*, vol. 9, p. 984, 1939.
CONDON, E.U., "Forced Oscillations in Cavity Resonators," *J. Applied Phys.*, vol. 12, p. 129, February, 1941.
HAHN, W. C., "New Method of the Calculation of Cavity Resonators," *J. Applied Phys.*, vol. 12, p. 62, January, 1941.
HANSEN, W. W., and RICHTMYER, R. D., "On Resonators Suitable for Klystron Oscillators," *J. Applied Phys.*, vol. 10, p. 189, March, 1939.
HANSEN, W. W., "On the Resonant Frequency of Closed Concentric Lines," *J. Applied Phys.*, vol. 10, p. 38, 1939.
HANSEN, W. W., "A Type of Electrical Resonator," *J. Applied Phys.*, vol. 9, p. 654, October, 1938.
JOUGET, M., "Natural Electromagnetic Oscillations of a Spherical Cavity," *Compt. rend.*, vol. 209, pp. 25 and 203, 1939.
MORITA, K., "Theory of Frequency Stabilizer for Decimeter Waves Using Metallic Ellipsoid," *Electrotech. J. Tokyo*, vol. 4, pp. 229–230, October, 1940.

MULLER, J., "Investigation of Electromagnetic Hollow Spaces," *Hochf. Tech. u. Elek. Akus.*, vol. 54, p. 157, 1939.

REBER, G., "Electric Resonance Chambers," *Communications*, vol. 18, p. 5, December, 1938.

RICHTMYER, R. D., "Dielectric Resonators," *J. Applied Phys.*, vol. 10, p. 391, 1939.

VON LINDEN, A., and DE VRIES, G., "Resonance Circuits for Very High Frequencies," *Phillips Tech. Rev.*, vol. 6, p. 217, July, 1941.

WATANABE, M., "Über der Eigenschwingung der elektromagnetische Hohlraum" (A note on resonators and waves guides), *Electrotech. J. Tokyo*, vol. 5, pp. 7–10, January, 1941.

Bibliography for Chapter 11

BARROW, W. L., and CHU, L. J., "Electromagnetic Horn Design," *Elec. Engg.*, vol. 58, p. 333, 1939.

BARROW, W. L., CHU, L. J., and JANSEN, J. J., "Biconical Electromagnetic Horns," *Proc. IRE*, vol. 27, p. 769, 1939.

BARROW, W. L., and CHU, L. J., "Theory of the Electromagnetic Horn," *Proc. IRE*, vol. 27, pp. 51–64, January, 1939.

BARROW, W. L., and GREENE, F. M., "Rectangular Hollow-pipe Radiators," *Proc. IRE*, vol. 26, p. 1498, December, 1938.

BARROW, W. L., and LEWIS, F. D., "The Sectoral Electromagnetic Horn," *Proc. IRE*, vol. 27, pp. 41–50, January, 1939.

BARROW, W. L., and SHULMAN, C., "Multi-unit Electromagnetic Horns," *Proc. IRE*, vol. 28, p. 130, 1940.

BUCHHOLZ, H., "The Movement of Electromagnetic Waves in a Cone-Shaped Horn," *Ann. Physik*, vol. 37, p. 173, 1940.

HOWE, G. W. O., "The Reception of Ultra Short Wavelengths by Metal Horns," *Wireless Engineer*, vol. 16, p. 109, 1939.

KRAUS, J. D., "The Corner Reflector Antennas," *Proc. IRE*, vol. 28, p. 513, November, 1940.

REBER, G., "Electromagnetic Horns," *Communications*, vol. 19, p. 13, February, 1939.

SOUTHWORTH, G. C., and KING, A. P., "Metal Horns as Directive Receivers of Ultra Short Waves," *Proc. IRE*, vol. 27, p. 95, February, 1939.

SONADA, S., "Research on Wave Guides and Electromagnetic Horns," *Electrotech. J. Tokyo*, vol. 4, pp. 35 and 41, 1940.

Bibliography for Chapters 12 and 13

BALLANTINE, S., Schrot-effect in High Frequency Circuits," *J. Franklin Inst.*, vol. 206, p. 159, August, 1928.

BENHAM, W. E., "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies," *Phil. Mag.*, vol. 5, p. 641, 1928; vol. 11, p. 457, 1931.

BENHAM, W. E., "The Variation of Resistance and Capacitance of Thermionic Valves with Frequency," *Wireless Engineer*, vol. 8, p. 488, 1931.

BENHAM, W. E., "A Contribution to Tube and Amplifier Theory," *Proc. IRE*, vol. 26, p. 1093, 1938.

BENNER, S., "Alteration of the Dielectric Constant of a Vacuum By Electrons," *Ann. Physik*, vol. 3, p. 111, 1929.

BLACK, H. S., "Stabilized Feed Back Amplifiers," *Bell. Sys. Tech. J.*, vol. 13, p. 1, 1934.

BLEWETT, J. P., and RAMO, S., "High-Frequency Behaviour of a Space Charge Rotating in a Magnetic Field," *Phys. Rev.*, vol. 57, p. 635, 1940.

BRÜCHE, E., and RECKNAGEL, A., "Phase Focussing of Electrons in Motion in Ultra High Frequency Fields," *Z. Physik*, vol. 108, p. 459, 1938.

BURNSIDE, D. G., and SALZBERG, B., "Recent Developments in Miniature Tubes," *Proc. IRE*, vol. 23, p. 1142, 1935.

CANNAS, A., "Acorn Valves for Ultra Short Waves," *Alta Freq.*, vol. 4, p. 115, 1935.

CARRARA, N., "On the Detection of Micro-Waves," *Alta Freq.*, vol. 3, p. 661, 1934.

CLAVIER, A. G., "The Influence of Transit Times of Electrons in Vacuum Tubes," *l'Onde élec.*, vol. 16, p. 145, 1937.

COCKBURN, R., "Variation of Voltage Distribution and Transit Time with Current in the Planar Diode," *Proc. Phys. Soc.*, vol. 47, p. 810, 1935.

COCKBURN, R., "Variation of Voltage Distribution and Transit Time in the Space Charge Limited Planar Diode," *Proc. Phys. Soc.*, vol. 50, p. 298, 1938.

DEWALT, K. C., "Three New Ultra-High-Frequency Triodes," *Proc. IRE*, vol. 29, pp. 475–480, September, 1941.

FERRIS, W. R., "Input Resistance of Vacuum Tubes as Ultra High Frequency Amplifiers," *Proc. IRE*, vol. 24, p. 82, 1936.

FORTESCUE, C. L., "Thermionic Voltmeters for Use at Very High Frequencies," *Jour. I.E.E.*, vol. 77, p. 429, September, 1935.

FRANKS, C. J., "Measured Input Losses of Vacuum Tubes," *Electronics*, vol. 8, p. 222, July, 1935.

FREEMAN, R. L., "Input Conductance Neutralization," *Electronics*, vol. 12, p. 22, October, 1939.

GAVIN, M. R., "Electron Pump Effect at High Frequencies," *Wireless Engineer*, vol. 15, p. 81, 1938.

GAVIN, M. R., "Valves for Use at Ultra Short Wavelengths," *Wireless Engineer*, vol. 16, p. 287, 1939.

GUTTINGER, P., "Behaviour of Space Charges in High Frequency Electric Fields," *Bull. assoc. suisse élec.*, vol. 31, p. 29, 1940.

HAEFF, A. V., and SALZBERG, B., "Effect of Space Charge in the Grid Anode Region of Vacuum Tubes," *R.C.A. Rev.*, vol. 2, p. 336, 1938.

HAEFF, A. V., "Effect of Electron Transit Time on the Efficiency of a Power Amplifier," *R.C.A. Rev.*, vol. 4, p. 114, 1939.

HAEFF, A. V., "U.H.F. Power Amplifier of Novel Design," *Electronics*, vol. 12, p. 30, February, 1939.

HAEFF, A. V., "Space Charge Effects in Electron Beams," *Proc. IRE*, vol. 27, p. 586, 1939.

HALLER, C., "Design and Development of Three New U.H.F. Transmitting Tubes" (RCA: 815, 829, 826), *Proc. IRE*, vol. 30, p. 20, January, 1942.

HARVEY, A. F., "The Impedance of the Magnetron in Different Regions of the Frequency Spectrum," *J. Inst. Elec. Eng.*, vol. 86, p. 297, 1940.

HEROLD, E. W., "An Analysis of Signal to Noise Ratio of U.H.F. Receivers at 300, 500 and 1000 Megacycles," *R.C.A. Rev.*, vol. 6, p. 302, January, 1942.

JANSKY, K. G., "Minimum Noise Levels Obtained on Short-Wave Radio Receiving Systems," *Proc. IRE*, vol. 25, p. 1517, December, 1937.

JEN, C. K., "On the Energy Equation in Electronics at Ultra-High Frequencies," *Proc. IRE*, vol. 29, pp. 464–466, August, 1941.

KOMPFNER, R., " Transit-Time Phenomena in Electronic Tubes " (A graphic method of investigation), *Wireless Engineer*, vol. 19, p. 2, January, 1942.

LLEWELLYN, F. B., " A Rapid Method of Estimating the Signal-to-Noise Ratio of a High Gain Receiver," *Proc. IRE*, vol. 19, p. 416, March, 1931.

LLEWELLYN, F. B., " Vacuum Tube Electronics at Ultra High Frequencies," *Proc. IRE*, vol. 21, p. 1532, 1933.

LLEWELLYN, F. B., " Phase Angle of Vacuum Tube Transconductance at very High Frequencies," *Proc. IRE*, vol. 22, p. 947, 1934.

LLEWELLYN, F. B., " Note on Vacuum Tube Electronics at Ultra High Frequencies," *Proc. IRE*, vol. 23, p. 112, 1935.

LLEWELLYN, F. B., " Operation of Ultra High Frequency Vacuum Tubes," *Bell Sys. Tech. J.*, vol. 14, p. 632, 1935.

LLEWELLYN, F. B., " Equivalent Networks of Negative Grid Vacuum Tubes at Ultra High Frequency," *Bell Sys. Tech. J.*, vol. 15, p. 575, 1936.

LOCKHART, C. E., " The Generation and Amplification of Microwaves " (Performance characteristics of U.H.F. tubes), *Electronic Eng.*, vol. 14, p. 336, August, 1941; vol. 14, p. 384, September, 1941; vol. 14, p. 432, October, 1941; vol. 14, p. 530, December, 1941.

MALTER, L., " Deflection and Impedance of Electron Beams at High Frequencies," *R.C.A. Rev.*, vol. 5, p. 439, April, 1941.

MEGAW, E. C. S., " Voltage Measurement at Very High Frequencies," *Wireless Engineer*, vol. 65, pp. 135 and 201, 1936.

MOULLIN, E.B., " Time of Flight of Electrons," *Wireless Engineer*, vol. 12, p. 371, 1935.

MOULLIN, E. B., " An Analysis of the Diode Rectifier," *Jour. I.E.E.*, vol. 81, p. 667, 1937.

NERGAARD, L. S., " Electrical Measurements at Wavelengths Less than 2 Metres," *Proc. IRE*, vol. 24, p. 1207, 1936.

NORTH, D. O., " The Absolute Sensitivity of Radio Receivers," *R.C.A. Rev.*, vol. 6, p. 332, January, 1942.

NORTH, D. O., " Analysis of the Effects of Space Charge on Grid Impedance," *Proc. IRE*, vol. 24, p. 108, 1936.

NORTH, D. O., and FERRIS, W. R., " Fluctuations Induced in Vacuum Tube Grids at High Frequencies," *Proc. IRE*, vol. 29, pp. 49–50, February, 1941.

NYQUIST, H., " Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, vol. 32 (Second Series), p. 110, July, 1928.

OKABE, K., " The Amplification and Detection of Ultra-Short Electric Waves," *Proc. IRE*, vol. 18, p. 1028, June, 1930.

PETERSON, L. C., " The Impedance Properties of Electron Streams," *Bell Sys. Tech. J.*, vol. 18, p. 465, 1939.

POSTHUMOUS, K., " Water-Cooled Pentodes at Ultra Short Wave-Lengths," *Philips Trans. News.*, vol. 4, p. 1, 1937.

PREISACH, F., and ZAKARIAS, I., " Input Conductance," *Wireless Engineer*, vol. 17, p. 147, 1940.

RACK, A. J., " Effect of Space Charge and Transit Time on the Shot Noise in Diodes," *Bell Sys. Tech. J.*, vol. 17, p. 592, October, 1938.

RAMO, S., " Currents Induced by Electron Motion," *Proc. IRE*, vol. 27, p. 584, 1939.

RATCLIFFE, J. A., and KOWNACKI, S., " A Method of Investigating Electron Inertia Effects in Thermionic Tubes," *Nature*, vol. 141, p. 1009, 1938.

RODWIN, G., and KLENK, L., " High Gain Amplifier for 150 Megacycles," *Proc. IRE*, vol. 28, p. 257, June, 1940.

ROTHE, H., "Input and Output Resistance of Thermionic Tubes at Very High Frequencies," *Telefunken Rohre*, July, 1936, p. 101.

ROTHE, H., and KLEEN, W., "The Importance of Electron Optics in the Technique of Amplifying Valves," *Z. tech. Physik*, vol. 17, p. 635, 1936.

ROTHE, H., "The Operation of Electron Tubes at High Frequencies," *Proc. IRE*, vol. 28, p. 325, July, 1940.

RUNGE, I., "Transit Time Effects in Electronic Tubes," *Z. tech. Physik*, vol. 18, p. 438, 1937.

SABAROFF, S., "An Ultra High Frequency Measuring Assembly," *Proc. IRE*, vol. 27, p. 208, March, 1939.

SALZBERG, B., "Design and Use of Acorn Tubes for U.H.F.," *Electronics*, vol. 7, p. 282, September, 1934.

SALZBERG, B., and BURNSIDE, D. G., "Recent Developments in Miniature Tubes," *Proc. IRE*, vol. 23, p. 1142, October, 1935.

SAMUEL, A. L., FAY, C. E., and SHOCKLEY, W., "The Theory of Space Charge between Parallel Plane Electrodes," *Bell. Sys. Tech. J.*, vol. 17, p. 49, 1938.

SAMUEL, A. L., and SOWERS, N. E., "A Power Amplifier for Ultra High Frequencies," *Bell Sys. Tech. J.*, vol. 16, p. 10, January, 1937.

SIL, B. C., "On the Variation of the Interelectrode Capacity of a Triode at High Frequencies," *Phil. Mag.*, vol. 16, p. 114, 1933.

SLOANE, R. W., and JAMES, E. G., "Transit Time Effects in Diodes in Pictorial Form," *Jour. I.E.E.*, vol. 79, p. 291, 1936.

STRUTT, M. J. O., and VAN DER ZIEL, A., "Causes for the Increase of Admittance of Modern High Frequency Amplifier Tubes on Short Waves," *Proc. IRE*, vol. 26, p. 1011, 1938.

STRUTT, M. J. O., "The Characteristic Admittances of Mixing Tubes for Frequencies up to 70 Megacycles," *E.N.T.*, vol. 15, p. 10, 1938.

STRUTT, M. J. O., "Electron Transit Time Effects in Multigrid Valves," *Wireless Engineer*, vol. 15, p. 315, 1938.

STRUTT, M. J. O., "Characteristic Constants of High Frequency Pentodes," *Wireless Engineer*, vol. 14, p. 478, 1937.

STRUTT, M. J. O., and VAN DER ZIEL, A., "Some Dynamic Measurements of Electron Motion in Multi-Grid Tubes," *Proc. IRE*, vol. 27, p. 218, 1939.

THOMPSON, B. J., "Review of U.H.F. Vacuum Tube Problems," *R.C.A. Rev.*, vol. 3, p. 146, October, 1938.

THOMPSON, B. J., and ROSE, G. M., "Vacuum Tubes of Small Dimensions for Use at Extremely High Frequencies," *Proc. IRE*, vol. 21, p. 1707, 1933.

THOMPSON, B. J., NORTH, D. O., and HARRIS, W. A., "Fluctuations in Space-Charged-Limited Currents at Moderately High Frequencies," *R.C.A. Rev.*, vol. 4, p. 269, January, 1940; vol. 4, p. 441, April, 1940; vol. 5, p. 106, July, 1940; vol. 5, p. 244, October, 1940; vol. 5, p. 371, January, 1941; vol. 5, p. 505, April, 1941; vol. 6, p. 114, July, 1941.

WAGENER, W. G., "Requirements of Performance of a new U.H.F. Power Tube (R.C.A. 888)," *R.C.A. Rev.*, vol. 2, p. 258, October, 1937.

WAGENER, W. G., "Development Problems and Operating Characteristics of Two New Ultra High Frequency Triodes," *Proc. IRE*, vol. 26, p. 401, 1938.

WAGNER, H. M., and FERRIS, W. R., "The Orbital-Beam Secondary-Electron Multiplier for Ultra-High-Frequency Amplification," *Proc. IRE*, vol. 29, pp. 598–602, November, 1941.

WANG, CHAO-CHEN, "Large Signal High Frequency Electronics of Thermionic Vacuum Tubes," *Proc. IRE*, vol. 29, p. 200, April, 1941.

WING, A. K., "A Push-Pull Ultra High Frequency Beam Tetrode," *R.C.A. Rev.*, vol. 4, p. 62, 1939.
"Midget Amplifiers," *Rev. Sci. Instruments*, vol. 12, p. 514, October, 1941.
"Transmitting Triode," *Rev. Sci. Instruments*, vol. 12, p. 233, April, 1941.

Bibliography for Chapter 14

BENHAM, W. E., "Discussion on Ultra-High Frequency Oscillations Generated by Means of a Demountable Thermionic Tube Having Electrodes of Plane Form," *Proc. Phys. Soc.*, vol. 53, p. 490, July, 1941.
ENGLUND, C. R., "The Short Wave Limit of Oscillators," *Proc. IRE*, vol. 15, p. 914, November, 1927.
GAVIN, M. R., "Triode Oscillators for Ultra-Short-Wavelengths," *Wireless Engineer*, vol. 16, p. 287, June, 1939.
HEROLD, E. W., "Negative Resistance and Devices for Obtaining It," *Proc. IRE*, vol. 23, p. 1201, 1935.
HOLLMAN, H. E., "Spherical Tank Ultra High Frequency Oscillator," *Electronics*, vol. 11, p. 26, December, 1938.
KELLY, M. J., and SAMUEL, A. L., "Vacuum Tubes as High Frequency Oscillators," *Elec. Engg.*, vol. 53, p. 1517, 1934.
KING, R., "Beam Tubes as U.H.F. Generators," *J. Applied Phys.*, vol. 10, p. 638, September, 1939.
KOLSTER, F. A., "Generation and Utilization of Ultra-Short Waves in Radio Communication," *Proc. IRE*, vol. 22, p. 1335, December, 1934.
KUHNOLD, W., "The Upper Frequency Limit for Negative Grid Oscillators," *Hochf. Tech. u. Elek. Akus.*, vol. 46, p. 78, 1935.
LEYSHON, W. A., "On Ultra-High Frequency Oscillations Generated by Means of a Demountable Thermionic Tube Having Electrodes of Plane Form," *Proc. Phys. Soc.*, vol. 53, p. 141, March, 1941.
LINDENBLAD, N. E., "Development of Transmitters for Frequencies above 300 Megacycles," *Proc. IRE*, vol. 23, p. 1013, September, 1935.
LOEBER, C. W., "Developments in Ultra-High Frequency Generation," *Electronics*, vol. 3, p. 376, November, 1930.
MCARTHUR, E. D., and SPITZER, E. E., "Vacuum Tubes as High Frequency Oscillators," *Proc. IRE*, vol. 19, p. 1971, November, 1931.
MEACHAM, L. A., "The Bridge-Stabilized Oscillator," *Proc. IRE*, vol. 26, p. 1278, October, 1938.
MOULLIN, E. B., "Effect of Curvature of the Characteristic on the Frequency of the Dynatron Generator," *Jour. I.E.E.*, vol. 73, p. 186, 1933.
MOUROMTSEFF, I.E., "Tuned-Grid Tuned-Plate Oscillator," *Communications*, vol. 20, p. 7, August, 1940.
SAMUEL, A. L., "Extending the Frequency Range of the Negative Grid Tube," *J. Applied Phys.*, vol. 8, p. 677, October, 1937.
SAMUEL, A. L., and SOWERS, N. E., "A Power Pentode for Ultra High Frequencies," *Proc. IRE*, vol. 24, p. 1464, 1936.
SAMUEL, A. L., "A Negative Grid Triode Oscillator and Amplifier for Ultra-High-Frequencies," *Proc. IRE*, vol. 25, p. 1243, October, 1937.
SEELEY, S. W., and ANDERSON, E. I., "U.H.F. Oscillator Frequency Stability," *R.C.A. Rev.*, vol. 5, p. 77, July, 1940.
THOMAS, H. A., *Theory and Design of Valve Oscillators*, Chapman and Hall, 1939.

TONKS, L., "Space Charge as a Cause of Negative Resistance in a Triode and Its Bearing on Shortwave Generation," *Phys. Rev.*, vol. 30, p. 501, October, 1927.

VAN DER POL, B., "The Nonlinear Theory of Electric Oscillations," *Proc. IRE*, vol. 22, p. 1051, September, 1934.

WHITE, W. C., "The Pliotron Oscillator for Extreme Frequencies," *Gen. Elec. Rev.*, vol. 19, p. 771, September, 1916.

WINNER, L., "A 400 Megacycle Generator," *Communications*, vol. 21, p. 24, November, 1941.

Bibliography for Chapter 15

ALFVEN, H., "The Barkhausen-Kurz Generator," *Phil. Mag.*, vol. 19, p. 419, 1935.

ANDERSON, J. E., "Theory of Electron Oscillators," *Electronics*, vol. 9, p. 9, August, 1936.

BANERJEE, S. S., and RAO, A. S., "Production of Ultra-High-Frequency Radio Waves by Electronic Oscillations," *Indian J. Phys.*, vol. 14, pp. 93–100, April, 1940.

BARKHAUSEN, H., and KURZ, K., "The Shortest Waves Producible by Means of Vacuum Tubes," *Physik Z.*, vol. 21, 1, 1920.

CHIPMAN, R. A., "The Electron Oscillation Characteristics of an Experimental Plane-Electrode Triode," *Proc. Phys. Soc.*, vol. 47, p. 1042, 1935.

CHIPMAN, R. A., "Ultra High Frequency Resonances in the Positive Grid Triode," *Proc. Phys. Soc*, vol. 51, p. 566, 1939.

CLAVIER, A. G., "Production and Utilization of Micro-Rays," *Elec. Comm.*, vol. 12, p. 3, 1933.

COCKBURN, R., "Determination of the Fundamental Types of Electron Oscillations in a Triode Valve," *Proc. Phys. Soc.*, vol. 49, p. 38, 1937.

COPIN, H., "Action of a Magnetic Field on the Electrons of Oscillating Triodes," *Rev. gén. élec.*, vol. 46, p. 568, 1939.

DYTRT, L. F., "Barkhausen-Kurz Oscillator Operation with Positive Plate Potentials," *Proc. IRE*, vol. 23, p. 241, 1935.

ELLIOTT, W. S., and RATCLIFFE, J. A., "Barkhausen-Kurz Oscillations with Positive Ions," *Nature*, vol. 145, p. 265, 1940.

FAY, C. E., and SAMUEL, A. L., "Vacuum Tubes for Generating Frequencies above One Hundred Megacycles," *Proc. IRE*, vol. 23, p. 199, 1935.

GILL, E. W. B., and MORRELL, J. H., "Short Electric Waves Obtained by Valves," *Phil. Mag.*, vol. 44, p. 161, 1922.

GILL, E. W. B., and MORRELL, J. H., "Short Electric Waves Obtained by the Use of Secondary Emission," *Phil. Mag.*, vol. 49, p. 369, 1925.

GILL, E. W. B., "Electrical Oscillations of Very Short Wave-Length," *Phil. Mag.*, vol. 12, p. 843, 1931.

GILL, E. W. B., "The Theory of Ultra Short Wave Generators," *Phil. Mag.*, vol. 18, p. 832, 1934.

GILL, E. W. B., "A Short Wave Generator," *Phil. Mag.*, vol. 28, p. 203, 1939.

HAMBURGER, F., "Electron Oscillations with a Triple Grid Tube," *Proc. IRE*, vol. 22, p. 79, 1934.

HERSHBERGER, W. D., "Modes of Oscillation in Barkhausen-Kurz Tubes," *Proc. IRE*, vol. 24, p. 964, 1936.

HOLLMANN, H. E., "The Mechanism of Electron Oscillations in a Triode," *Proc. IRE*, vol. 17, p. 229, 1929.

HOLLMANN, H. E., "The Generation of Ultra Short Waves by Thermionic Valves," *Z. Hochf. Tech.*, vol. 35, pp. 21 and 76, 1930.

HOLLMANN, H. E., "The Retarding Field Tube as a Detector for Any Carrier Frequency," *Proc. IRE*, vol. 22, p. 630, 1934.

HOLLMANN, H. E., *Ultra Short Wave Technique* (2 vols.), 1937.

JONESCU, T. V., "A New Micro Wave Generator," *Compt. rend.*, vol. 204, p. 1411, 1937.

KLEINSTEUBER, W., "The Influence of Space Charge in Planar Retarding-Field Tubes," *Hochf. Tech. u. Elek. Akus.*, vol. 53, p. 139, 1939.

KLINGER, H., "Über die Erzeugung von Decimeterwellen mit Doppelgitterrohren nach der Bremsfeldmethode" (On the generation of decimeter waves with double-grid tubes by the retarded field method), *Funk. Tech. Monatshefte*, No. 8, pp. 121–124, August, 1940.

KOCKEL, B., and MROWKA, B., "On the Theory of the Barkhausen-Kurz Tube," *Z. tech. Physik.*, vol. 2, p. 42, 1938.

KOZANOWSKI, H. W., "Ultra-Short Wave Oscillations at 60 Cm," *Proc. IRE*, vol. 20, p. 957, June, 1932.

LLEWELLYN, F. B., and BOWEN, A. E., "Production of Ultra-High Frequency Oscillations by Means of Diodes," *Bell Sys. Tech. J.*, vol. 18, p. 280, 1939.

MATSUDAIRA, K., "Ultra High Frequency Oscillations of Diodes and Triodes," *Electrotech. J. Tokyo*, vol. 3, p. 19, 1939.

MCPETRIE, J. S., "A Graphical Method for Determining the Transit Time of Electrons in Triodes under Conditions of Space Charge Limitation," *Phil. Mag.*, vol. 16, p. 284, 1933.

MCPETRIE, J. S., "The Determination of the Inter-Electrode Times of Transit of Electrons in Triode Valves with Positive Grid Potentials," *Phil. Mag.*, vol. 16, p. 544, 1933.

MCPETRIE, J. S., "Experiments with Inverted Diodes Having Various Filament Cathodes," *Phil. Mag.*, vol. 19, p. 501, 1935.

MCPETRIE, J. S., "A Diode for Ultra High Frequency Oscillations," *Wireless Engineer*, vol. 11, p. 118, 1934.

MCPETRIE, J. S., "Electronic Oscillations in Positive Grid Triodes and Resonance Oscillations in Magnetron Generators," *Jour. I.E.E.*, vol. 80, p. 84, 1937.

MEGAW, E. C. S., "Electronic Oscillations," *Jour. I.E.E.*, vol. 72, p. 313, 1933.

MOORE, W. H., "Electron Oscillations without Tuned Circuits," *Proc. I.R.E.*, vol. 22, p. 1021, 1934.

MULLER, J., "Electron Oscillations in High Vacua," *Hochf. Tech. u. Elek. Akus.*, vol. 41, p. 156, 1933.

MULLER, J., "The Spiral Grid Electron Oscillator," *Ann. Physik*, vol. 21, p. 611, 1935.

NAKAMURA, S., "Electronic Oscillations Produced by a Thermionic Tube with Plane Electrodes Operated in Push Pull," *Nippon Elec. Comm. Eng.*, vol. 10, p. 134, 1938.

PFETSCHER, O., and MULLER, J., "Barkhausen-Kurz Oscillators," *Hochf. Tech. u. Elek. Akus.*, vol. 45, p. 1, 1935.

ROSTAGNI, A., "Barkhausen-Kurz Oscillators," *Atti accad. sci. Torino*, vol. 66, p. 124, 1931.

ROSTAGNI, A., "The Theory of Barkhausen-Kurz Electron Oscillators," *Radio e Televisione*, vol. 3, p. 28, 1939.

Scheibe, A., "The Generation of Ultra-Short Waves with Hot-Cathode Tubes," *Ann. Physik*, vol. 73, p. 54, 1924.

Thompson, B. J., and Zottu, P. D., "An Electron Oscillator with Plane Electrodes," *Proc. IRE*, vol. 22, p. 1374, 1934.

Tonks, L., "Electronic Oscillations," *Phys. Rev.*, 1927.

Wenstrom, W. H., "An Experimental Study of Regenerative Ultra-Short-Wave Oscillators," *Proc. IRE*, vol. 20, January, 1932.

Bibliography for Chapter 16

Ahrens, E., "Water-cooled Four Slit Magnetron for Decimetre Waves," *Hochf. Tech. u. Elek. Akus.*, vol. 50, p. 181, 1937.

Asai, S., "On the Characteristics of the Magnetron of a Symmetrical Type," *Nippon Elec. Commun. Eng.*, vol. 21, pp. 62–63, July, 1940.

Asai, S., "Tentative Proposition on the Mechanism of Electronic Oscillations," *Electrotech. J. Tokyo*, vol. 5, pp. 59–60, March, 1941.

Awender, H., "The Properties and Use of the Magnetic Field Tube," *Funktech. Monat.*, vol. 4, pp. 173, 201, 1938.

Bellustin, S. V., "Theory of Motion of Electrons in Crossed Electric and Magnetic Fields with Space Charge," *Physik. Z. Sowjet.*, vol. 10, p. 251, 1936.

Bellustin, S. V., "Study of a Whole Anode Magnetron Neglecting the Space Charge," *J. Tech. Phys.*, vol. 13, p. 1188, 1939.

Benham, W. E., "Electronic Theory and the Magnetron Oscillator," *Proc. Phys. Soc.*, vol. 47, p. 1, 1935.

Berger, H., "Fundamentals of the Magnetron," *Funktech. Monat.*, vol. 4, p. 97, 1938.

Bethenod, J., "Variation of Space Current in a Magnetron under the Action of the Magnetic Field," *Compt. rend.*, vol. 209, p. 832, 1939.

Biguenet, O., and Pierret, E., "On the Ultra Short Waves Obtained with a Magnetron," *J. phys. et rad.*, vol. 6, p. 67, 1935.

Biguenet, O., and Pierret, E., "On the Anomalous Increase of Anode Current in Magnetrons," *J. phys. et rad.*, vol. 7, p. 33, 1936.

Bovsheverov, V. M., and Grechowa, M. T., "Magnetrons for Decimetre Waves," *J. Tech. Phys.*, vol. 5, p. 69, 1935.

Bovsheverov, V. M., "Characteristic Curves of the Split Anode Magnetron," *Tech. Phys. U.S.S.R.*, vol. 2, p. 567, 1935.

Braude, S. J., "On the Cut-off in the Plane Magnetron with Space Charge," *Physik. Z. Sowjet.*, vol. 7, p. 565, 1935.

Braude, S. J., "Motion of Electrons in Electric and Magnetic Fields," *Physik. Z. Sowjet.*, vol. 8, p. 584, 1935.

Braude, S. J., "On the Wavelength of the Electronic Oscillations in the Plane Magnetron," *Physik. Z. Sowjet.*, vol. 8, p. 667, 1935.

Braude, S. J., "Motion of Electrons in Electric and Magnetic Fields with Space Charge," *Physik. Z. Sowjet.*, vol. 12, p. 1, 1937.

Brillouin, L., "The Theory of the Magnetron," *Phys. Rev.*, vol. 60, p. 385, September, 1941.

Chang, H., and Chaffee, E. L., "Characteristics of the Negative-Resistance Magnetron Oscillator," *Proc. IRE*, vol. 28, pp. 519–523, November, 1940.

Dehlinger, W., "Ultra High Frequency Oscillation of the Magnetostatic Vacuum Tube," *Physics*, vol. 2, p. 432, 1932.

Elder, F. R., "The Magnetron Amplifier and Power Oscillator," *Proc. IRE*, vol. 15, p. 159, 1925.

ERARD, J., and PIERRET, E., "Influence of Magnetic Field Tilt on the Oscillations in a Magnetron," *J. phys. et rad.*, vol. 9, p. 54, 1938.

FASSI, G., and SALOM, G., "The Magnetron as a Generator of Micro Waves," *Alta Freq.*, vol. 3, p. 396, 1934.

FISCHER and LUDI, "The Posthumous Oscillations in the Magnetron," *Bull. assoc. suisse élec.*, vol. 28, p. 277, 1937.

FRITZ, K., "Production of Oscillations with the Habann Valve," *Hochf. Tech. u. Elek. Akus.*, vol. 46, p. 16, 1935.

FRITZ, K., "Theory of Transit Time Oscillations of the Magnetron," *Telefunken Z.*, vol. 17, p. 31, 1936.

FRITZ, K., and LERBS, A., "Fremdsteuerung mit Magnetfeldröhren" (Separate control in magnetrons), *Telefunken Mitteilungen*, vol. 21, pp. 44–48, September, 1940.

FRITZ, K., and ENGBERT, W., "Schwingungsformer und Ordnungszahlen der Magnetfeldrohre" (Forms of oscillations and "orders" of magnetrons), *Telefunken Mitteilungen*, vol. 21, pp. 41–43, September, 1940.

GIACOMINI, A., "Anomalous Dispersion in the Magnetron," *Ricerca sci.*, vol. 1, p. 650, 1934.

GILL, E. W. B., and BRITTON, K. G., "The Action of the Split Anode Magnetron," *Jour. I.E.E.*, vol. 78, p. 461, 1936.

GROOS, O., "Production of Dwarf Waves with Magnetron Emitters," *Hochf. Tech. u. Elek. Akus.*, vol. 51, p. 37, 1939.

GROSZKOWSKI, J., and RYZKO S., "The Distribution of Electrostatic Field in Split Anode Magnetrons," *Hochf. Tech. u. Elek. Akus.*, vol. 47, p. 55, 1936.

GROSZKOWSKI, J., and RYZKO, S., "A New Method of Modulating the Magnetron Oscillator," *Proc. IRE*, vol. 24, p. 771, 1936.

GRÜNBERG, G., and LULASCHKOV, W., "On the Theory of the Split Anode Magnetron," *Tech. Phys. U.S.S.R.*, vol. 2, p. 484, 1935.

GRÜNBERG, G., and WOLENSTEIN, "The Influence of a Magnetic Field on Electron Motion between Coaxial Cylinders," *Tech. Phys. U.S.S.R.*, vol. 4, p. 179, 1937.

GUNDLACH, F. W., "Recent Investigations on Decimetre Wave Magnetron Transmitters," *Hochf. Tech. u. Elek. Akus.*, vol. 48, p. 201, 1936.

GUNDLACH, F. W., "The Habann Valve as a Generator of Decimetre Waves," *E.T.Z.*, vol. 58, p. 653, 1937.

GUNDLACH, F. W., "The Behaviour of the Habann Valve as a Negative Resistance," *E.N.T.*, vol. 15, p. 183, 1938.

GUNDLACH, F. W., "Modulation of Magnetron Emitters by Amplitude Limitation," *Hochf. Tech. u. Elek. Akus.*, vol. 53, p. 10, 1939.

GUTTON, H., and BERLINE, S., "Research on Magnetrons for Ultra Short Waves," *Bull. de la S.F.R.*, vol. 12, p. 30, 1938.

HABANN, E., "The Split Anode Magnetron," *Z. f. Hochfreq.*, vol. 24, p. 115, 1924.

HAELBIG, A., "A Grid-controlled Magnetic Field Valve," *Hochf. Tech. u. Elek. Akus.*, vol. 50, p. 96, 1937.

HARA, G., "Internal Capacity and Resistance of Magnetrons," *Electrotech. J. Tokyo*, vol. 4, p. 24, 1940.

HARA, G., and MITO, S., "Wavelength of Electron Oscillations in the Magnetron," *Electrotech. J. Tokyo*, vol. 3, p. 23, 1939.

HARVEY, A. F., "The Cut-off Characteristic of the Single Anode Magnetron," *Proc. Camb. Phil. Soc.*, vol. 35, p. 637, 1939.

HARVEY, A. F., "The Impedance of the Magnetron in Different Regions of the Frequency Spectrum," *Jour. I.E.E.*, vol. 86, p. 297, 1940.

HARVEY, A. F., "Output and Efficiency of the Split Anode Magnetron Oscillating in the Dynatron Regime," *Jour. I.E.E.*, vol. 84, p. 683, 1939.

HELLER, G., "The Magnetron as a Generator of Ultra Short Waves," *Philips Tech. Rev.*, vol. 4, p. 189, 1939.

HERRIGER, F., and HULSTER, F., "Oscillations in Tubes with Magnetic Fields," *Hochf. Tech. u. Elek. Akus.*, vol. 49, p. 123, 1937.

HISHIDA, M., "Theory of Asymmetrically Split Magnetrons," *Electrotech. J. Tokyo*, vol. 3, p. 144, 1939.

HOAG, J. B., "A Note on the Theory of Magnetron Oscillator," *Proc. IRE*, vol. 21, p. 1132, August, 1933.

HOFFMAN, F., "Bremsfeldrohren mit Magnetfeld; Statische Kennlinie und Kurzwellenerzeugung." (Retarded field tubes with magnetic fields; static characteristics and generation of short waves), *Hochf. Tech. u. Elek. Akus.*, vol. 56, pp. 137–149, November, 1940.

HOLLMANN, H. E., "The Behaviour of the Electron Oscillator in a Magnetic Field," *E.N.T.*, vol. 6, p. 377, 1929.

HOLLMANN, H. E., "The Generation of Ultra Short Waves by Thermionic Valves," *Z. Hochf. Tech.*, vol. 35, pp. 21 and 76, 1930.

HOLLMANN, H. E., "The Magnetron as a Negative Resistance," *Ann. Physik*, vol. 8, p. 956, 1931.

HOLLMANN, H. E., "Arrangement for Push-Pull Modulation of Split Anode Magnetrons," *Hochf. Tech. u. Elek. Akus.*, vol. 49, p. 105, 1937.

HOLLMANN, H. E., "Survey of New Magnetic Field Valves," *Funktech. Monat.*, vol. 3, p. 75, 1938.

HULL, A. W., "The Effect of a Uniform Magnetic Field on the Motion of Electrons between Co-axial Cylinders," *Phys. Rev.*, vol. 18, p. 13, 1921.

HULL, A. W., "The Magnetron," *Jour. A.I.E.E.*, vol. 40, p. 715, September, 1921.

HULL, A. W., "The Axially Controlled Magnetron," *Jour. A.I.E.E.*, vol. 42, p. 1013, October, 1923.

HÜLSTER, F., "Cathode Overheating in the Magnetic Field Valve," *Telefunken Rohr.*, vol. 14, p. 217, 1938.

ITO, Y., and KATSURAI, S., "The Three Slit Magnetron and Three Phase Oscillations," *Nippon Elec. Comm. Eng.*, vol. 1, p. 467, 1937.

KATSURAI, S., "Generation of Three Phase Oscillations by a Three Slit Magnetron," *Electrotech. J. Tokyo*, vol. 1, p. 152, 1937.

KATSURAI, S., "Theory of the Split Anode Magnetron Oscillator," *Nippon Elec. Comm. Eng.*, vol. 1, p. 129, 1937.

KATZMAN, Y. A., and RUBINA, T. F., "Calculation of the Static Characteristics of a Split Anode Magnetron," *J. Tech. Phys.*, vol. 9, p. 499, 1939.

KILGORE, G. R., "Magnetostatic Oscillators for Generation of Ultra Short Waves," *Proc. IRE*, vol. 20, p. 1741, 1932.

KILGORE, G. R., "Magnetron Oscillators for Generation of Frequencies from 300–600 mc/s," *Proc. IRE*, vol. 24, p. 1140, 1936.

KILGORE, G. R., "The Magnetron as a High Frequency Generator," *J. Applied Phys.*, vol. 8, p. 666, 1937.

KOHLER, H. W., "A Thermal Method for Measuring Efficiencies at Ultra-High Frequencies Applied to the Magnetron Oscillator," *Proc. IRE*, vol. 25, p. 1381, November, 1937.

LÄMMCHEN, K., "Differentiation of the Various Types of Oscillation in a Habann Valve," *Hochf. Tech. u. Elek. Akus.*, vol. 51, p. 87, 1938.

LÄMMCHEN, K., and LERBS, A., "First Order Transit Time Oscillations in Magnetrons," *Hochf. Tech. u. Elek. Akus.*, vol. 52, p. 186, 1938.

LÄMMCHEN, K., and MULLER, L., "Sinusoidal Oscillations in the Habann Valve," *E.N.T.*, vol. 16, p. 37, 1939.

LINDER, E. G., "Improved Magnetron Oscillator," *Phys. Rev.*, vol. 45, p. 656, 1934.

LINDER, E. G., "Description of the End Plate Magnetron," *Proc. IRE*, vol. 24, p. 633, 1936.

LINDER, E. G., "Excess Energy Electrons and Electron Motion in High Vacuum Tubes," *Proc. IRE*, vol. 26, p. 346, 1938.

LINDER, E. G., "Effect of High Energy Electron Random Motion on the Shape of the Magnetron Cut-off Curve," *J. Applied Phys.*, vol. 9, p. 331, 1938.

LINDER, E. G., "The Anode Tank Circuit Magnetron," *Proc. IRE*, vol. 27, p. 732, 1939.

MAIDANOV, A. P., "Influence of Gas Pressure on the Energy and Efficiency of Magnetron Oscillators," *Physik, Z. Sowjet.*, vol. 10, p. 718, 1936.

MCNAMARA, F. T., "A Note on Magnetron Theory," *Proc. IRE*, vol. 22, p. 1037, August, 1934.

MCPETRIE, J. S., "Electronic Oscillations in Triodes and Resonance Oscillations in Magnetrons," *Jour. I.E.E.*, vol. 80, p. 84, 1937.

MCPETRIE, J. S., "Experimental Investigation of Resonance and Electronic Oscillations in Magnetrons," *Jour. I.E.E.*, vol. 86, p. 283, 1940.

MEGAW, E. C. S., "A Magnetron Oscillator for Ultra Short Wavelengths," *Wireless Engineer*, vol. 10, p. 197, 1933.

MEGAW, E. C. S., "Magnetron Valves for Ultra Short Wavelengths," *G.E.C.J.*, vol. 7, p. 94, 1936.

MEGAW, E. C. S., "An Investigation of the Magnetron Short-Wave Oscillator," *Jour. I.E.E.*, vol. 72, p. 326, April, 1933.

NERGAARD, L. S., "Electrical Measurements at Wavelengths Less Than Two Meters," *Proc. IRE*, vol. 24, p. 1207, September, 1936.

OKABE, K., "The Short Wave Limit of Magnetron Oscillators," *Proc. IRE*, vol. 17, p. 652, 1929.

OKABE, K., "Magnetron Oscillations of a New Type," *Proc. IRE*, vol. 18, p. 1748, 1930.

OKABE, K., "Split Anode Magnetron of Special Type," *Electrotech. J. Tokyo*, vol. 1, p. 213, 1937.

OKABE, K., "Electron Beam Magnetrons and Type B Magnetron Oscillations," *Proc. IRE*, vol. 27, p. 24, January, 1939.

OKABE, K., and OWAKI, K., "A Sectionalised Magnetron," *Electrotech. J. Tokyo*, vol. 2, p. 257, 1938.

OKABE, K., *Magnetron Oscillations of Ultra Short Wavelengths and Electron Oscillations in General*, 57 pp., 1938.

OWAKI, K., "Specially Sectionalised Magnetron Tube," *Nippon Elec. Comm. Eng.*, vol. 16, p. 631, 1939.

PFETSCHER, O., and PUHLMANN, W., "High Power Habann Generator for Micro Waves," *Hochf. Tech. u. Elek. Akus.*, vol. 47, p. 105, 1936.

PIDDUCK, F. B., "Theory of Short Wave Oscillations with the Magnetron," *Wireless Engineer*, vol. 18, p. 404, October, 1940.

PONTE, M., "The Use of Magnetic Fields for the Production of Ultra Short Waves," *l'Onde élec.*, vol. 13, p. 493, 1934.

POSTHUMOUS, K., "Oscillations in a Split Anode Magnetron," *Wireless Engineer*, vol. 12, p. 126, 1935.

RICHTER, H., "Production of cm and mm Waves in Magnetrons," *Hochf. Tech. u. Elek. Akus.*, vol. 51, p. 10, 1938.

RUNGE, W., "The Generation of Oscillations with the Magnetron," *Telefunken Z.*, vol. 15, p. 5, 1934.

SCHWAYENBACH, H. A., "Space Charges and Electron Oscillations in Magnetron Triodes," *Helv. Phys. Acta*, vol. 9, p. 565, 1935.

SLUTZKIN, A. A., and others, "Factors influencing the Output and Efficiency of Magnetron Oscillations," *Physik. Z. Sowjet.*, vol. 5, p. 887, 1934.

SLUTZKIN, A. A., and STEINBERG, D. S., "Production of Short Wave Undamped Oscillations by the Use of a Magnetic Field," *Ann. Physik*, vol. 1, p. 658, 1929.

SLUTZKIN, A. A., and LELJAKOV, P. P., "A Special Type of Magnetron Oscillation," *Physik. Ber.*, vol. 14, p. 1266, 1933.

SLUTZKIN, A. A., and others, "The Generation of Electromagnetic Waves below 50 cm. with Magnetrons," *Physik Z. Sowjet.*, vol. 6, p. 150, 1934.

SPIWAK, G. V., and ZREBNY, P. E., "On the Magnetron Cut-off Curve," *Compt. rend. sci. U.R.S.S.*, vol. 24, p. 237, 1939.

TONKS, ., "Motion of Electrons in Crossed Electric and Magnetic Fields with Space Charge," *Physik. Z. Sowjet.*, vol. 8, p. 572, 1935.

UCHIDA, Y., and others, "Radiation-cooled Magnetron with Four Slit Anode," *Nippon Elec. Comm. Eng.*, vol. 15, p. 537, 1939; and *Electrotech. J. Tokyo*, vol. 2, p. 290, 1938.

WHITE, W. C., "Producing Very High Frequencies by Means of the Magnetron," *Electronics*, vol. 1, p. 376, November, 1930.

WIGDORTSCHIK, I. M., "Velocity Distribution of Electrons under the Influence of a Magnetic Field," *Physik. Z. Sowjet.*, vol. 10, p. 245, 1936.

ZUHRT, H., "The Static Characteristics of the Split Anode Magnetron," *Hochf. Tech. u. Elek. Akus.*, vol. 48, p. 91, 1936.

Bibliography for Chapter 17

ALLERDING, A., DALLENBACH, W., and KLEINSTEUBER, W., "The Resotank, A New Generator for Micro Waves," *Hochf. Tech. u. Elek. Akus.*, vol. 51, p. 96, 1938.

ARSENJEWA-HEIL, A., and HEIL, O., "A New Method of Producing Short Undamped Waves of Great Intensity," *Z. Physik*, vol. 95, p. 752, 1935. (Translated into English in, *Electronics*, vol. 16, p. 164, July, 1943.)

BENHAM, W. E., "Phase-Focusing in Velocity Modulated Beams," *Wireless Engineer*, vol. 17, p. 514, December, 1940.

BETHENOD, J., "The Electronic Valve with Velocity Modulation," *Compt. rend.*, vol. 210, p. 103, 1940.

BLACK, D. H., "Ultra Short Wave Oscillators," *Elec. Comm.*, vol. 17, p. 325, 1939.

BORGNIS, F., and LEDINEGG, E., "Zur Phasenfokussierung geradlinig bewegter Elekronenstrahlen" (On the phase focusing of electron rays moving in a straight line), *Z. tech. Physik*, vol. 21, No. 10, pp. 256–261, 1940.

COLEBROOK, F. M., "Ultra Short and Decimetre Waves Valves and Deflection of a Focussed Beam," *Wireless Engineer*, vol. 15, p. 198, 1938.

CONDON, E. U., "Electronic Generation of Electromagnetic Oscillations," *J. Applied Phys.*, vol. 11, p. 502, 1940.

GEIGER, M., "Current Flow Characteristics in Velocity Modulated Beams," *Physik. Ber.*, vol. 21, p. 710, 1940.

GULJAEV, V., "Theory of the Klystron," *J. Phys. U.S.S.R.*, vol. 4, No. 1–2, pp. 143–146, 1941.

HAEFF, A. V., and NERGAARD, L. S., "A Wide Band Inductive-Output Amplifier," *Proc. IRE*, vol. 28, p. 126, 1940.

HAEFF, A. V., "An Ultra High Frequency Power Amplifier of Novel Design," *Electronics*, vol. 12, p. 30, February, 1939.

HAHN, W. C., "Small Signal Theory of Velocity Modulated Electron Beams," *Gen. Elec. Rev.*, vol. 42, p. 258, 1939.

HAHN, W. C., "Wave Energy and Transconductance of Velocity Modulated Beams," *Gen. Elec. Rev.*, vol. 42, p. 497, 1939.

HAHN, W. C., and METCALF, G. F., "Velocity Modulated Tubes," *Proc. IRE*, vol. 27, p. 106, 1939.

HEIL, O., "Velocity Modulated Tube," French patent, 1936.

HOLLMANN, H. E., "New Transit-Time Devices," *Funktech. Monat.*, vol. 12, p. 1, 1940.

HOLLMANN, H. E., "Ultra Short Wave Generator Valves," *Funktech. Monat.*, vol. 10, p. 319, 1938.

HOLLMANN, H. E., and THOMA, A., "Zur Theorie der Triftröhren" (On the theory of drift tubes), *Hochfrequenz. u. Elek. Akus.*, vol. 56, pp. 181–186, December, 1940.

JOBST, G., "Charges Moving Periodically in Electron Tubes, Particularly in Shock Excited Ultra Short Wave Oscillators," *Telefunken-Hausmittel.*, vol. 20, p. 84, 1939.

JONESCU, T. V., "A Short Wave Oscillator," *Compt. rend.*, vol. 204, p. 1411, 1937.

KOCKEL, B., "Geschwindigkeitsgestuerte Laufzeitröhren. Beitrag zur Theorie" (Velocity-controlled transit-time tubes. Contribution to their theory), *Z. tech. Physik*, vol. 22, No. 2, pp. 77–85, 1941.

KOCKEL, B., and MAYER, L., "Velocity Modulated Electron Beams in Crossed Deflection Fields," *Jahrb. A.E.G.-Forschung.*, vol. 6, p. 72, 1939.

KOMPFNER, R., "Velocity Modulated Beams," *Wireless Engineer*, vol. 17, p. 110, 1940.

KOMPFNER, R., "Velocity Modulation-Results of Further Considerations," *Wireless Engineer*, vol. 17, p. 478, November, 1940.

LUDI, F., "Über einen neuartigen Ultrakurzwellengenerator mit Phasenfocussierung" (On a new type of ultra-short-wave generator with phase focusing), *Helv. Phys. Acta*, vol. 13, No. 6, pp. 498–522, 1940.

MAYER, L., "Experimental Demonstration of Phase Focussing," *Z. tech. Physik*, vol. 20, p. 38, 1939.

MULLER, J. J., and ROSTAS, E., "A Transit Time Generator Using a Single Rhumbatron," *Helv. Phys. Acta*, vol. 13, No. 6, pp. 435–450, 1940.

PIERCE, J. R., "Limiting Current Densities in Electron Beams," *J. Applied Phys.*, vol. 10, p. 715, 1939.

RAMO, S., "Space Charge and Field Waves in an Electron Beam," *Phys. Rev.*, vol. 56, p. 276, 1939.

RAMO, S., "The Electron-Wave Theory of Velocity Modulated Tubes," *Proc. IRE*, vol. 27, p. 757, 1939.

RAMO, S., "Traveling Waves in Electron Beams," *Communications*, vol. 20, p. 5, November, 1940.

SMITH, L. P., and HARTMAN, P. L., "Formation and Maintainance of Electron and Ion Beams," *J. Applied Phys.*, vol. 11, p. 220, 1940.

STRACHEY, C., "Velocity Modulated Beams," *Wireless Engineer*, vol. 17, p. 202, 1940.

TIBERIO, U., "Deflection Valves for Ultra Short Decimetre Waves," *Wireless Engineer*, vol. 15, p. 612, 1938.

TOMBS, D. M., " Velocity Modulated Beams," *Wireless Engineer*, vol. 17, p. 54, 1940.
VARIAN, R. H., and VARIAN, S. F., " A High Frequency Amplifier and Oscillator," *J. Applied Phys.*, vol. 10, pp. 140 and 321, 1939.
VARIAN, R. H., and HANSEN, W. W., " New Radio Apparatus Using 10 Cm. Waves," *Sci. News Letter*, vol. 35, p. 89, 1939.
WEBSTER, D. L., " Cathode Ray Bunching," *J. Applied Phys.*, vol. 10, p. 501, July, 1939.
WEBSTER, D. L., " The Theory of Klystron Generators," *J. Applied Phys.*, vol. 10, p. 864, 1939.
WHEELOCK, N., " Velocity Modulation of Electron Beams," *QST*, vol. 23, p. 37, 1939.

INDEX

ation f an	Text Books — 3 days Technical Books — 7 days General Books — 14 days	One mon

FROM THE DATE OF ISSUE